Universidade do Minho
Escola de Engenharia

António Miguel Fernandes Martins, 65369

Deteção de pessoas para *Smart Autonomous Mobile Units*

Dissertação de Mestrado

Mestrado Integrado em Engenharia Eletrónica Industrial e de Computadores

Trabalho efetuado sob a orientação de

Professor Doutor Sérgio Monteiro

Novembro de 2017

DECLARAÇÃO

Nome: António Miguel Fernandes Martins

Endereço eletrónico: a65369@alunos.uminho.pt Telefone: 919631475

Cartão do Cidadão: 14221778

Título da dissertação: Deteção de pessoas para *Smart Autonomous Mobile Units*

Orientadores:

Professor Doutor Sérgio Monteiro

Ano de conclusão: 2017

Mestrado em Engenharia Eletrónica Industrial e Computadores

É AUTORIZADA A REPRODUÇÃO PARCIAL DESTA DISSERTAÇÃO (indicar, caso tal seja necessário, nº máximo de páginas, ilustrações, gráficos, etc.), APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, _____/_____/_____

Assinatura:

AGRADECIMENTOS

Durante o decorrer desta dissertação, foram várias as pessoas que de alguma forma tiveram um papel fundamental no sentido de tornar possível a sua realização e conclusão. Em primeiro lugar, gostaria de agradecer à minha família especialmente aos meus pais e irmãos pelo apoio constante ao longo desta longa caminhada.

A minha namorada deixar um agradecimento especial por todo apoio, compreensão, dedicação e companheirismo em todos os momentos ao longo deste percurso.

Deixar também uma palavra de apreço a todos os meus amigos, em especial ao Pedro Silva, Marco Cruz, Vítor Pacheco, Romeu Gonçalves, João Cunha e Tiago Teixeira, por estarem sempre presentes, e por toda a entreajuda partilhada nos momentos que assim o exigiam.

Agradecer ao meu orientador, Dr. Sérgio Monteiro, pela confiança depositada em mim para realizar este projeto. Agradecer também toda a motivação e força em momentos de maior ceticismo, a sua orientação e ajuda prestada.

A todas estas pessoas que de alguma forma facilitaram e tornaram possível a realização deste trabalho, deixo um profundo agradecimento.

RESUMO

O interesse por veículos autónomos tem aumentado nos últimos tempos. São veículos dotados com alguma inteligência que lhes permite decidir sempre qual o percurso a tomar para atingir o alvo, sem requerer ajuda de operadores ou de uma marcação explícita do caminho a seguir. Para terem sucesso nestas tarefas, estes veículos devem possuir sistemas de perceção do ambiente circundante que sejam robustos e precisos, de modo a otimizar as rotas e a evitar potenciais obstáculos que se encontrem ao seu redor.

O caso de estudo para o tema desta dissertação é a perceção do ambiente em redor de um veículo autónomo de movimentação interna de materiais num chão de fábrica. Esta tarefa é parte integrante do projeto da IFactory que resulta de uma parceria entre a Universidade do Minho e a Bosh car Multimedia Portugal, S.A.. O principal objetivo é a aplicação deste veículo numa zona de produção da Bosh Car Multimédia Portugal, S.A..

Neste tipo de ambientes industriais existem inúmeros obstáculos que se podem opor à rota do veículo. Para além de ser necessário a deteção dos mesmos é ainda requerida a sua identificação, pois mediante o tipo de obstáculo o comportamento a exibir pelo veículo poderá ter de ser diferente.

São várias as tecnologias que têm vindo a ser desenvolvidas que podem ser usadas neste tipo de aplicações. Apesar das inúmeras vantagens que cada uma das tecnologias possui, estas mesmas por si só não são suficientes para garantir segurança e robustez, sendo então necessário usar várias em simultâneo e recorrer a métodos de fusão sensorial. Outro fator também muito importante são os algoritmos e os métodos a utilizar. Estes vão permitir a análise ou o tratamento da informação obtida dos sensores, para serem retiradas as devidas informações.

Para a comunicação entre os vários algoritmos e os sensores do veículo recorreu-se ao *middleware* ROS, fornecendo este bibliotecas e ferramentas de suporte de modo a simplificar todo o desenvolvimento de software.

O sistema de perceção proposto nesta dissertação foca-se na deteção de pessoas. Este recorre a sensores LiDAR 2D para a deteção de um padrão de pernas e a uma câmara 3D para a deteção da parte superior do corpo humano. Neste documento, é realizada toda a implementação dos algoritmos assim como a descrição do funcionamento. Por fim são apresentados todos os resultados e conclusões dos métodos utilizados.

Palavras Chave: Perceção ambiente, Sistemas de visão, LiDAR, Deteção de Pessoas.

ABSTRACT

In the last years the interest for autonomous vehicles has increased. They are vehicles with some intelligence that allows them to always decide which route to take to reach the target without requiring the help of operators or any explicit marking on the way to follow.

In order to be successful in these tasks, these vehicles must have environment perception systems that are robust and accurate, so that their navigation system always choose the best route and avoid potential obstacles that can be found around them.

The case study for the topic of this dissertation is the environment perception around an autonomous vehicle for material transport in a factory floor. This task is an integral part of the IFactory project that results from a partnership between the University of Minho and Bosh car Multimedia Portugal, S.A.. The main objective is to apply this vehicle in a production area of Bosh car Multimedia Portugal, S.A..

In this type of environment there are numerous obstacles that can oppose the route of the vehicle. In addition to the need of detecting them, it is still required to identify them, because the behavior of the vehicle may change, depending on the type of obstacle.

There are several technologies that can be used in this type of applications. Despite the many advantages that each technology has, these alone are not enough to guarantee security and robustness, so it is necessary to use several technologies simultaneously and use sensor fusion methods. Other important factors are the algorithms and methods to implement. They will allow obtaining all sensor data and perform all the analysis and processing required for that information.

The communication between the algorithms and the sensors of the vehicle is done by the middleware ROS. This middleware provides libraries and support tools that will allow simplifying all the software development.

The perception system proposed in this dissertation focused on people detection. This uses 2D LiDAR sensors for the detection of a leg pattern and a 3D camera for the detection of the human upper body. All the implementation of the algorithms, as well as the description of the operation are carried out in this document. Finally, all the results and conclusions for the methods used are presented.

Key Words: Environment Perception, Vision systems, LiDAR, People Detection.

ÍNDICE

LISTA DE FIGURAS

LISTA DE TABELAS

Lista de Abreviaturas, Siglas e Acrónimos

RGB	Red, Green and Blue
RGB-D	Red, Green, Blue and Depth
SAMU	Smart Autonomous Mobile Units
LiDAR	Light Detection and Ranging
Sonar	Sound Navigation and ranging
Radar	Radio Detection and Ranging
ICP	Iterative Closest Point
DOF	Degrees of freedom
ROI	Region of interest
TOF	Time of Flight
SVM	Support vector machine
HOG	Histograms of Oriented Gradients
SIFT	Scale-invariant feature transform
SNR	Signal-to-Noise rate
RANSAC	Random Sample Consensus
NMS	Non-maximum suppression
VGV	Vision Guided Vehicles
DOG	Difference of Gaussian

1. INTRODUÇÃO

1.1 Motivações e Enquadramento

Com os avanços na tecnologia, a procura e o interesse por veículos e sistemas autónomos tem aumentado. As aplicações para estes veículos são inúmeras, e podem ir desde veículos de reconhecimento/exploração do espaço, terra ou no oceano, veículos de cargas em ambientes industriais ou até mesmo cadeiras de rodas inteligentes.

Para que estes veículos autónomos sejam capazes de uma navegação segura devem estar dotados de sistemas que lhes permitam perceber o ambiente circundante. Estes sistemas de perceção do meio ambiente permitem dotar um veículo autónomo ou outros sistemas com a capacidade de ao reconhecer o meio, identificar e caracterizar todos os obstáculos e/ou marcadores de interesse, sendo isto, uma necessidade essencial principalmente em ambientes em constante mudança, como é o caso de ambientes industriais. Outras funcionalidades importantes de um sistema de perceção de ambiente são, por exemplo, a capacidade de auxiliar na determinação da localização do veículo, a deteção e reconhecimento de um alvo, o cálculo de distâncias a obstáculos ou até mesmo a auxiliar na interação com humanos ou outros sistemas.

Este tipo de sistemas de perceção quando aplicado a um determinado veículo traduzir-se-á num aumento de automação, eficiência no planeamento de rotas e acima de tudo segurança, tanto para humanos como bens ou estruturas.

1.2 Contexto

O caso de estudo nesta dissertação é a perceção ambiente de um veículo autónomo de movimentação interna de materiais num ambiente industrial, sendo o foco principal a deteção de pessoas. Esta tarefa é parte integrante do projeto da IFactory que resulta de uma parceria entre a Universidade do Minho e a Bosh car Multimedia Portugal, S.A..

A tarefa principal é o transporte de materiais do armazém até as linhas de produção e o transporte de paletes com os produtos acabados dessas mesmas linhas até ao local de armazenamento dos mesmos. Este transporte será efetuado por dois veículos autónomos distintos: o veículo rebocador (Tugger) e o veículo empilhador (Stacker). Na Figura 1, podemos visualizar estes mesmos veículos.

Figura 1- Veículo rebocador (Tugger) à esquerda e o veículo empilhador (Stacker) à direita.

1.3 Objetivos

O caso de estudo é a perceção ambiente de um veículo autónomo de movimentação interna de materiais, estando este mesmo inserido num ambiente industrial.

Existem 4 potenciais obstáculos neste tipo de ambientes. Estes podem ser objetos estáticos, operadores humanos, veículos autónomos ou outros veículos dirigidos por operadores humanos. O objetivo principal desta dissertação é o desenvolvimento de um sistema com a capacidade de deteção de pessoas, pois mediante o tipo de obstáculo o comportamento do veículo poderá ser diferente. Com este objetivo, o sistema recorrerá a várias tecnologias de sensorização de modo a providenciar um bom reconhecimento do meio e uma deteção robusta.

Desta forma, quando aplicado a um veículo, o sistema deverá garantir uma deteção de pessoas eficaz e em tempo real, de modo a todas as operações realizadas pelo veículo sejam efetuadas em segurança.

1.4 Estrutura da dissertação

Este documento encontra-se organizado da seguinte forma:

- No capítulo 1 é realizada uma introdução ao tema, onde é apresentado o problema relativo a sistemas de perceção do meio, assim como a sua importância na atualidade.
- O capítulo 2 é composto pelo estado da arte, onde são referidas um conjunto de tecnologias de sensorização atuais importantes para sistemas de perceção ambiente. É ainda apresentado alguns estudos e propostas de solução atuais sobre veículos autónomos ou sistemas de

perceção que incorporam essas tecnologias, assim como alguns métodos utilizados para a deteção de pessoas. Todos estes encontram-se divididos por áreas de aplicação.

- No capítulo 3, são apresentados alguns fundamentos teóricos sobre as diferentes tecnologias de sensorização atuais, como os princípios de funcionamento. É ainda abordado alguns métodos de processamento de imagens, assim como algoritmos utilizados na deteção de pessoas.
- A implementação do sistema é apresentada no capítulo 4. Este possuirá todo o desenvolvimento do sistema, como por exemplo as tecnologias de sensorização escolhidas, os algoritmos desenvolvidos e a configuração de todo o sistema
- No capítulo 5 são apresentados os resultados obtidos do modelo desenvolvido no capítulo 4.
- Por último, no capítulo 6, são apresentadas todas as conclusões do sistema implementado, assim como apresentadas algumas propostas de evolução do sistema.

2. ESTADO DA ARTE

2.1 Sistemas de sensorização

São várias as tecnologias de sensorização que têm vindo a ser desenvolvidas e melhoradas e que podem ser usadas neste tipo de aplicações. Segundo [1], sistemas de visão, são uma das técnicas mais comuns utilizadas em navegação de veículos e em vigilância do tráfego, pois providenciam uma grande quantidade de informação do ambiente em redor. Por outro lado, dependendo da aplicação, podem requerer um esforço computacional elevado e com alguma sensibilidade a variações de luminosidade.

Outra tecnologia importante neste tipo de aplicações é o LiDAR (*Light Detection and Ranging),* em que são emitidos um conjunto de pulsos pelo LiDAR e mediante a presença de um obstáculo eles são refletidos de volta. Uma vez medido esse tempo de voo, a distância do objeto ao sensor é estimada. Apresenta como vantagens uma elevada precisão, baixo processamento de dados e um campo de visão amplo [2]. Existem ainda outros sistemas de sensorização como o Sonar (*Sound Navigation and Raging*), em que um conjunto de pulsos acústicos são emitidos pelo sensor e no caso da presença de algum obstáculo eles são refletidos de volta. Esses pulsos são então capturados e processados para obter informações do obstáculo que os refletiu de volta. Esta tecnologia é maioritariamente utilizada para uma deteção de obstáculos próxima. Para além desta, e com o mesmo principio de funcionamento, e o Radar (*Radio Detection and Ranging*) com a diferença que as ondas emitas são do tipo rádio, permitindo um alcance superior e respostas mais rápidas. O resto da secção irá aprofundar os sistemas de visão 3D: Luz Estruturada, TOF (*Time of Flight*) e estéreo. A principal vantagem da utilização deste tipo de sistemas de visão é a possibilidade de construção de mapas de profundidade, providenciando assim com boa precisão a distância aos diferentes tipos de obstáculos.

Uma das câmaras mais utilizadas para este efeito é a Microsoft Kinect e a sua escolha deve-se ao seu preço bastante competitivo em relação às restantes tecnologias e a sua capacidade de fornecer uma boa informação do ambiente em redor. A Kinect v1 é uma câmara 3D, desenvolvida pela Microsoft para a consola XBOX 360, para aplicações que permitam o jogador utilizar o seu corpo como controlador. Ela consiste numa câmara a cores associada com um projetor de um padrão de luz estruturada e uma câmara para observar os desvios do padrão causados pelos contornos dos objetos no cenário, possuindo assim a capacidade de estimar a distância dos elementos do ambiente, mesmo com uma baixa luminosidade. O mapa de profundidade fornecido pela Kinect v1 possui uma resolução de 640 x 480 pixels e é composto por uma imagem a preto e branco, em que cada pixel guarda a distância dos

elementos do ambiente ao sensor. Existem ainda outras câmaras, do tipo TOF, como a Visionary-T da SICK ou a Kinect v2, que recorrem a um emissor e um recetor infravermelhos, e através do tempo de voo de um sinal conseguem estimar a distância de um ponto do cenário a câmara. Estas tipicamente operam em alcances superiores a câmaras de luz estruturada como a Kinect v1 e a Asus XtionPro Live.

A imagem fornecida por câmaras do tipo TOF e de Luz estruturada é composta por uma imagem a preto e branco, em que cada pixel guarda a distância dos elementos do ambiente ao sensor. Este tipo de câmaras recorre a um sensor infravermelho para obter a perceção em profundidade, possuindo a capacidade de identificar obstáculos em condições de visibilidade limitada [3], como por exemplo o período da noite ou na presença de pó ou fumo, e a distinção entre humanos e objetos [4].

Contudo, é possível obter uma perceção em profundidade sem recorrer a câmaras com sensores infravermelhos, através de câmaras do tipo estéreo que recorrem a pares de câmaras para obter duas imagens de pontos de vista diferentes simulando assim a visão binocular do ser humano [5] [6]. Contudo é necessário realizar a correspondência de imagens das diferentes câmaras para obter o mapa de profundidade, sendo este um processo complicado [7].

Apesar das inúmeras vantagens que cada um destes sensores dispõe, cada um deles por si só não é suficiente para garantir segurança, robustez e um bom reconhecimento do meio. Para a resolução deste problema, a maior parte dos sistemas que incorporam navegação autónoma, recorrem a métodos de fusão sensorial.

2.2 Veículos Autónomos

É expectável que num futuro próximo os veículos autónomos dominem o mercado automóvel, pois garantem comodidade e segurança a todos passageiros. Para navegarem eficazmente, estes veículos são dotados por sistemas de perceção que recorrem sempre a uma fusão sensorial, tirando assim partido dos benefícios de cada uma das tecnologias sensoriais de que dispõem. Enquanto um condutor possui um campo de visão limitado, estes tipos de sistemas dispõem de uma perceção total em redor do veículo, sendo isto uma mais valia para uma navegação segura e robusta. Um exemplo destes sistemas é o apresentado em [2], em que é proposto um método de deteção dos limites laterais do percurso e dos obstáculos para veículos autónomos, recorrendo a um sensor LiDAR 2D, colocado no topo do veículo (1.75m do chão) com uma inclinação θ relativa a superfície de varrimento do sensor ficando assim direcionado para baixo (demonstrado na Figura 2), de modo a permitir a deteção

obstáculos mais abaixo da altura do veículo como por exemplo curvas ou cones de sinalização, otimizando assim o desempenho da leitura.

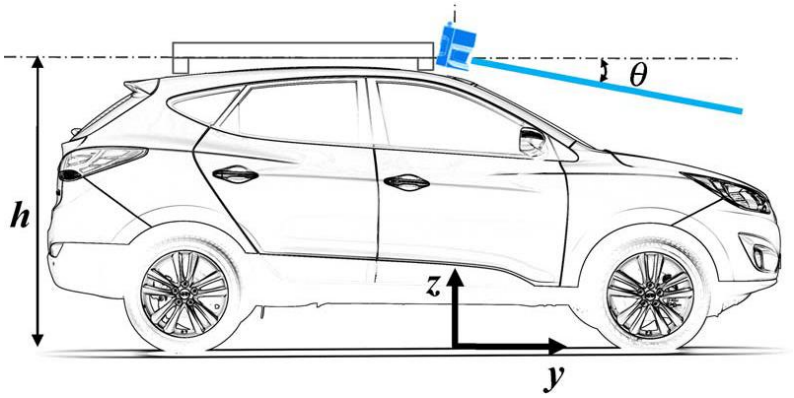

Figura 2- Configuração do LiDAR no veículo, [2].

Os limites laterais do percurso são definidos para o veículo com recurso a um filtro probabilístico de dados [8], que fornece a probabilidade de o caminho existir e quais os alvos existentes. Este mesmo filtro foi também usado para estimar qual o melhor ângulo de inclinação do sensor LiDAR. Todo este sistema foi validado num veículo autónomo, denominado de A1, que venceu um concurso de veículos autónomos em 2010 na Coreia do Sul, organizado pelo grupo automóvel Hyunday-Kia.

Já em [9] é apresentada a evolução do veículo autónomo do estudo anterior (A1) para o mesmo concurso, desta vez do ano 2012, sendo mais uma vez este o vencedor. Desta vez, o ambiente de condução em redor do A1 é determinado através de uma fusão sensorial, com recurso a LiDARs e a sistemas de visão (câmaras). Num total de 8 LiDARs (4 LMS151 com um alcance entre 0,5 m a 50 m e um campo de visão de 270° posicionados nos quatro cantos do veículo, 2 IBEO com um alcance de 150 m e um campo de visão de 180° posicionados na frente do veículo e 2 LMS291 com alcance de 80 metros e um campo de visão de 180° colocados no topo do veículo e direcionados para a frente do veículo), foram utilizados para determinar o tamanho, direção e a posição dos objetos. A inclusão de um sistema de visão permitiu um reconhecimento do ambiente em redor mais preciso, através da identificação de alguns obstáculos como por exemplo sinais de trânsito, semáforos ou até mesmo humanos. Este sistema é composto por duas câmaras monocromáticas, para identificar objetos no percurso (humanos, semáforos e zonas de estacionamento), e uma câmara a cores para analisar a cor desses objetos. Essa análise de cor é realizada para determinar o estado dos semáforos, identificar alguns sinais de trânsito e, no caso do concurso, para recolher um passageiro com uma camisola laranja. Os algoritmos de visão recorrem a métodos de aprendizagem baseada em reconhecimento, a

segmentação de cores e a fusão de informação. Exemplo disso pode ser a identificação de semáforos, em que recorrem ao Adaboost (algoritmo de aprendizagem) [10] e a segmentação de cores para uma rápida detenção. Depois da detenção recolhem a informação da distância a que se encontram do semáforo com recurso aos LiDAR e fazem a comparação com vários *templates* apropriados a distância obtida para fazer a respetiva identificação do estado do semáforo.

Figura 3 - Deteção e identificação de obstáculos veículo A1, [9]: Á esquerda a deteção do passageiro; Ao centro a deteção da passadeira e do semáforo; À direita a deteção do sinal de trânsito.

No caso de objetos dinâmicos, os sensores LiDAR devem detetar e seguir o objeto sempre que ele se encontre próximo do veículo autónomo.

Para além destes, inúmeros estudos têm sido realizados por construtoras automóveis ou empresas de renome internacional, dos quais alguns deles já se encontram a circular nas estradas, ainda numa fase de teste. Exemplos esses podem ser a Google, Uber ou a Tesla.

Começando pelo caso da Google, desenvolveu um sistema de condução autónoma denominado Waymo [11]. O sistema de perceção recorre a um Radar, câmaras e a um LiDAR com características especiais desenvolvido pela empresa. A principal diferença para os principais rivais esta no sensor LiDAR, pois este é constituído por 3 sensores, dos quais 2 são novas categorias de LiDAR: um sensor LiDAR de curto alcance, que permite visualizar toda a área próxima do veículo, não importando quão próximo o obstáculo se encontra; um LiDAR de longo alcance com a capacidade de fazer zoom sobre objetos que se encontrem na estrada, tornando assim possível a perceção de uma bola a uma distância de 2 campos de futebol; um LiDAR de alta resolução, que se encontra acoplado com os dois anteriores, que permite uma reconstrução de uma imagem virtual 3D com elevada resolução, sendo este um ponto fulcral no sistema de navegação, principalmente em ambientes complexos como cidades (estradas repletas de ciclistas, pessoas e objetos inesperados). Para além dos LiDAR, utiliza os sensores Radar pois permitem detetar objetos em condições atmosféricas adversas tais como chuva, neve ou nevoeiro e um sistema de visão para efetuar tarefas tais como identificação de sinais de trânsito.

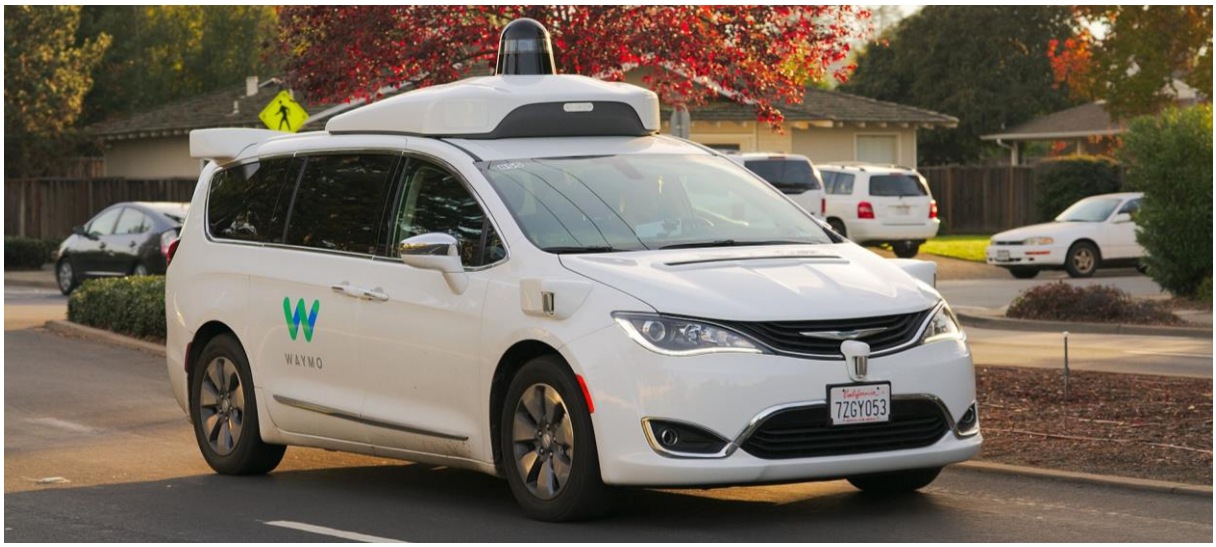

Figura 4- Waymo: Carro autónomo da Google [12].

Já a Tesla [13], um dos maiores fabricantes automóveis elétricos, desenvolveu o sistema *auto-pilot* (piloto automático) que recorre a 8 câmaras, um sistema Radar e 12 sensores ultrasónicos, e que ao contrário das suas rivais não incorporara nenhum LiDAR. As câmaras encontram-se espalhadas pelo veículo para permitir uma visibilidade em todo o seu redor e possuindo um alcance máximo de 250 m. A configuração utilizada é a seguinte: três câmaras à frente montadas atrás do para-brisas, em que cada uma delas possui capacidades distintas. Uma para providenciar uma visão de longas distâncias, que é utilizada principalmente em velocidades mais elevadas; outra com uma lente do tipo olho de peixe que permite uma visão ampla de 120°, com a função de captar sinais de trânsito, obstáculos que se oponham a rota do veículo e objetos próximos, sendo esta uma mais valia em ambientes urbanos ou em manobras a baixa velocidade. E uma ultima câmara com capacidade de médio alcance que cobre a área intermedia entre estas duas. Para além destas, possui ainda duas câmaras laterais em cada um dos lados do veículo, cada uma delas com uma visão de 90° que permite alguma redundância na deteção de carros inesperados que possam entrar na mesma via que o veículo e ao mesmo tempo providenciar segurança adicional quando este entra em interseções de visibilidade limitada. Por fim, uma câmara ao centro atrás, de modo a efetuar manobras de estacionamento mais complexas, e duas câmaras com vista para trás em cada um dos lados do carro, que cobrem uma zona tipicamente cega para os condutores, permitindo assim a mudança de via e misturar-se no trânsito. O sistema radar utiliza um comprimento de onda que permite detetar obstáculos mesmo em condições de nevoeiro, pó, chuva, neve e debaixo dos carros, permitindo a previsão e a deteção de possíveis obstáculos que possam opor-se á trajetória do veículo.

Por último, os sistemas Sonar possuem um alcance de 5 metros e permitem detetar carros próximos, especialmente aqueles que invadem a linha do veículo. Para além disto, providencia orientação nas manobras de estacionamento.

A UBER [14], uma das maiores prestadoras de serviços eletrónicos na área dos transportes, dispõe já de vários veículos a circular autonomamente por algumas capitais. Os sistemas de sensorização utilizados são semelhantes aos usados pelas duas rivais: Câmaras óticas, Radar, Sonar e LiDAR.

Figura 5 - Veículo autónomo UBER, [15].

2.3 Veículos Autónomos aplicados à indústria.

Este tipo de veículos surgiu para, entre outras aplicações, cobrir as necessidades de uma distribuição de materiais flexível, pois permite uma fácil reconfiguração da área de trabalho, o que é uma vantagem em ambientes industriais que mudem frequentemente. Tipicamente, o transporte interno de materiais é realizado por veículos operados manualmente, sendo esta uma tarefa rotineira, tediosa e umas das maiores causas de acidentes na indústria. Desta forma, este tipo de veículos autónomos garantem uma maior segurança, pois possuem sistemas de perceção robustos e com alguma redundância, e quando dotados de sistemas de gestão logística eficazes garantem uma eficiência na distribuição de matérias e um acompanhamento em tempo real. Para além disto, dispõem de sistemas de navegação flexíveis e não necessitam de operadores, sendo várias as potenciais aplicações. Os sistemas de perceção destes

veículos recorrem maioritariamente a sensores LiDAR para detetar eventuais obstáculos e localizar se a si próprio no ambiente. Algumas propostas de soluções atuais, são apresentadas em baixo:

O KARIS PRO [16], desenvolvido pelo instituto de pesquisa e educação Kit na Alemanha, é um pequeno veículo autónomo com a capacidade de transportar materiais sozinho ou conectado a outros veículos. Possui a capacidade de interação com o ambiente, especialmente com outros veículos ou pessoas, pois quando se deparam com uma rota com algum trânsito intenso, procuram uma nova rota ou alinham-se com outros veículos como alternativa. Além disso, a ordem das tarefas a executar por estes veículos pode ser alterada por qualquer interveniente. O sistema de perceção do meio recorre a sensorização LiDAR de modo a detetar obstáculos que se possam opor à trajetória do veículo e ao mesmo tempo fornece informação ao sistema de navegação sobre os limites do ambiente em redor. Com isto, o sistema de navegação compara a informação recebida com o mapa do ambiente previamente adquirido e localiza-se e navega ao mesmo tempo em direção ao alvo.

O MIR 200 (Mobile Industrial Robots) [17] possui a capacidade de transportar materiais autonomamente até um total de 200 kg. Possui a capacidade de ser customizado mediante as tarefas a executar, como por exemplo aplicação de estantes, caixas ou até mesmo um braço robótico cooperativo. Este veículo, pode deslocar-se autonomamente adquirindo o seu próprio mapa ou pode importar o traçado do ambiente em redor. O sistema de perceção recorre a várias tecnologias sensoriais: dois sensores LiDAR (SICK S300) colocados em dois cantos opostos do veículo para uma perceção de 360°, uma câmara 3D (Intel Realsense) para a deteção de objetos localizados a frente do veículo entre 5 cm a 50 cm do chão e 4 sensores ultrassónicos para a deteção de objetos transparentes como por exemplo uma porta de vidro.

Já a SEEGRID [18] apresenta uma solução para a automação do transporte e armazenamento de materiais baseada em VGV (veículos guiados através da visão), que consistem em veículos empilhadores e transportadores que recorrem unicamente a sistemas de visão 3D para a navegação e perceção. Podem operar em modo manual e automático, e o modo de treino para a operação automática é realizado pelo operador a caminhar ou a conduzir pela rota desejada. Durante esta fase, o sistema de visão captura imagens e constrói um mapa 3D do ambiente em redor. Para além disto, possui ainda a capacidade de posicionamento automático dos garfos para o carregamento de paletes.

Outra solução diferente das até agora aqui apresentadas é a da BALYO [19], que desenvolveu um kit de condução autónomo denominado "*Driven*" para aplicar em veículos de transporte de materiais, sem qualquer alteração mecânica. Este kit permite converter um veículo elétrico manual num veículo autónomo, podendo o utilizador escolher o modo em que pretende operar, manual ou automático.

O sistema de perceção recorre a câmaras 3D e a sensores LiDAR em que este último, fornece toda a informação ao sistema de navegação. Possui ainda um sensor LiDAR a altura dos garfos para garantir que as paletes são recolhidas e entregues devidamente.

2.4 Robôs de exploração

Para além das soluções mais vocacionadas para o transporte de pessoas ou materiais, existe ainda outras aplicações de sistemas de perceção. A necessidade de uma rápida caracterização e quantificação de ambientes mais complexos levam a que robôs com capacidade de digitalizar modelos de ambiente em 3D sejam utilizados para aplicações como missões resgate, inspeção robótica ou simplesmente a navegação em ambientes desconhecidos. Em [20] é apresentado o desenvolvimento de um modelo de um robô móvel autônomo para exploração 3D e digitalização de ambientes interiores. Este recorre a um sensor LiDAR 3D, para a aquisição de toda a informação relativa ao ambiente, e a três módulos de *software* para o tratamento dos dados e para o controlo do mesmo.

Um primeiro módulo recebe os varrimentos 3D do LiDAR, e com base num algoritmo ICP (*Iterative closest Point)* esses varrimentos 3D são registados num sistema de coordenadas comum, ao mesmo tempo que o robô é relocalizado. Este tipo de algoritmo é usado para reconstruir a superfície em 3D a partir dos diferentes varrimentos efetuados, que neste caso são efetuados pelo LiDAR.

Um segundo módulo é responsável por identificar sempre qual o seguinte melhor ponto para o robô efetuar o varrimento do ambiente. Este é um dos pontos fundamentais para a geração de modelos completos, tendo em conta que os modelos 3D são o resultado de vários varrimentos e o robô tem de planear e mover se até estes pontos. A Figura 6 ilustra esse mesmo planeamento em 6 diferentes imagens. No topo à esquerda podemos ver os dados obtidos do varrimento. No topo ao centro podemos ver a segunda etapa deste módulo que é a deteção de linhas no plano com recurso a uma transformação de Hough[1]. No topo à direita as linhas são transformadas em polígonos, onde as arestas são classificadas como vistas ou não vistas. As linhas vistas, são as linhas que foram detetadas, que estão ligadas por linhas não vistas. Em baixo à esquerda podemos observar como a linhas 1 e a linha 2 são conectadas. O critério utilizado é ligar sempre as extremidades das linhas com o menor ângulo entre eles. Neste caso

[1] - A transformada de Hough é uma técnica matemática que realiza a deteção de formas geométricas em imagens digitais. No início do seu desenvolvimento permitiu a deteção de retas, mas mais tarde foi estendida a outras formas geométricas tais como círculos e elipses [71].

uma extremidade da linha 1 é ligada à extremidade da linha dois com menor ângulo. Em baixo ao centro são geradas posições candidatas a ser o próximo ponto a efetuar o varrimento dentro do polígono. E por fim, em baixo à direita, é selecionado o melhor ponto onde existe maior ganho de informação de acordo com a posição das linhas não vistas.

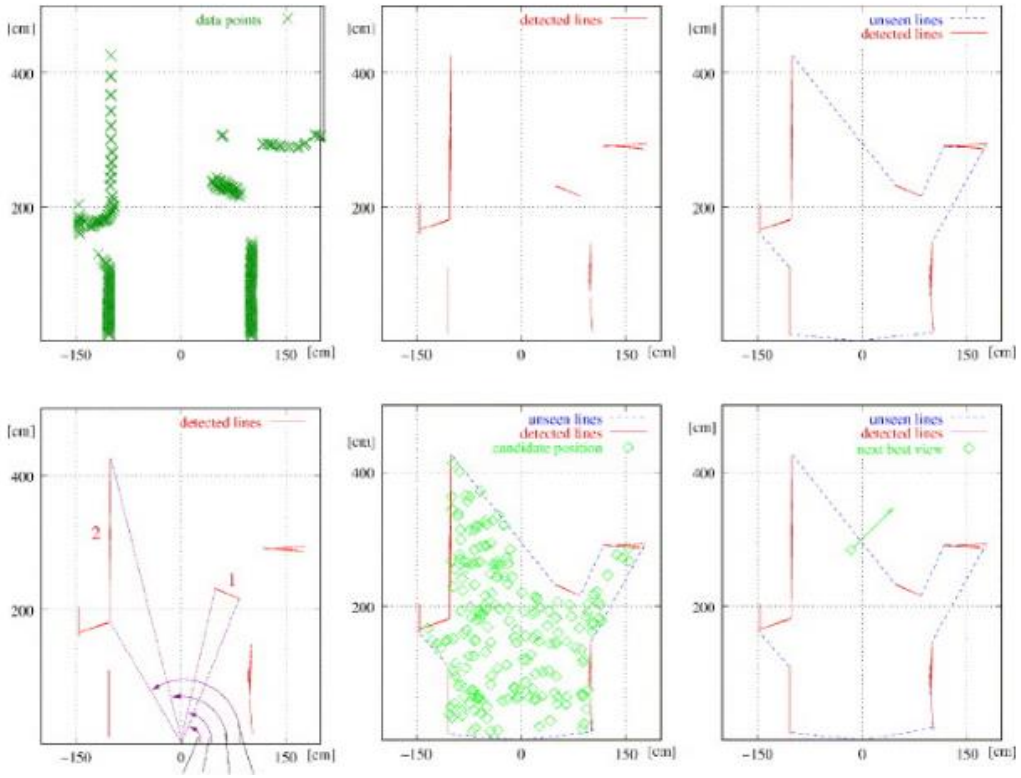

Figura 6 - Cálculo do próximo ponto de varrimento, [20].

Por fim, um terceiro módulo responsável pelo controlo dos motores, navegando assim o robô para o próximo ponto, com base em odometria[2], e evitando ainda colisões com obstáculos dinâmicos. Os modelos 3D obtidos pelo robô são o resultado de múltiplos varrimentos do LiDAR 3D e como tal objetos dinâmicos podem levar a erros nesses mesmos modelos, como por exemplo desalinhamentos. A solução encontrada foi a utilização um LiDAR 2d para a deteção desses mesmos obstáculos, em que caso se encontre presente um objeto dinâmico o robô simplesmente repete o varrimento 3D.

[2] - Odometria é o uso dos dados adquiridos pelos sensores de movimentos para estimar mudanças na posição ao longo do tempo. É usado por alguns robôs para estimar a sua posição relativamente ao ponto de começo, [5].

Apesar deste último estudo recorrer a um LiDAR 3D, em [21] e em [22], recorrem a dois LiDAR 2D para fazer a representação do ambiente em 3D. A configuração utilizada é um montado horizontalmente e o outro verticalmente, em que este último efetua uma linha de varrimento vertical que é transformada em pontos 3D mediante a pose do robô. Esta configuração apresenta algumas desvantagens pois a precisão dos pontos 3D depende da pose do robô e da precisão do LiDAR, enquanto que um LiDAR 3D consegue gerar um mapa de pontos 3D consistentes só com um único varrimento. Para além disto, segundo [20], todas estas abordagens têm dificuldade em navegar à volta de obstáculos 3D com beiras sobressaídas.

Todos os casos apresentados anteriormente recorrem sobretudo a sistemas LiDAR para obter a perceção do meio, mas segundo [23] e [24] câmaras de profundidade podem também fornecer uma boa perceção do meio sobretudo em ambientes interiores.

Em [23] é apresentado o estudo da aplicação de uma Kinect no topo de um robô, de modo a navegar autonomamente num ambiente interior. Neste estudo, a câmara é acedida através da biblioteca *freenect* [25], fornecendo assim uma visão do ambiente em profundidade e outra imagem a cores. Apesar da câmara fornecer imagens com uma resolução de 640 x 480 pixels, optaram por redimensionar as imagens por um fator de 5 em todas as dimensões, ficando assim com imagens de 128 x 96 pixéis. A abordagem utilizada para a deteção das paredes no ambiente é a seleção de todos os pontos com altura inferior ao teto, relativamente ao chão, e a pesquisa por colunas de qual dos pontos se encontra mais afastado do robô. Para a deteção dos obstáculos a abordagem utilizada é igual à das paredes, com a diferença que todos os pontos com altura superior à do robô, relativamente ao chão, e que não se encontrem no chão são rejeitados.

A deteção de obstáculos utilizando a abordagem explicada anteriormente é realizada com sucesso, mas em contrapartida este método de deteção apresenta algumas desvantagens como a necessidade de calibração mediante o tipo de ambiente e a localização da câmara. Apesar da imagem ter sido redimensionada a informação providenciada pela mesma foi suficiente para a deteção de obstáculos e paredes.

Em [24] é apresentado também o desenvolvimento de um sistema de perceção para a navegação autónoma de um robô num ambiente interior com recurso a uma Kinect. O sistema de perceção recebe a informação da Kinect e em conjunto com a biblioteca OpenCV [26] cria o seu mapa de obstáculos, para permitir que o robô se mova no ambiente enquanto evita colisões com obstáculos. Para a deteção de obstáculos a imagem recebida é dividida verticalmente em 5 partes, e em cada uma delas é analisada a intensidade dos pixéis de modo a determinar o mínimo e o máximo absoluto de

distância entre o sensor e os obstáculos. Quando esta distância é inferior de 60 cm então o robô considera que tem um obstáculo no seu percurso.

Neste estudo a deteção de obstáculos foi sempre bem-sucedida, tendo obtido sempre os mesmos resultados em ambientes interiores iluminados e não iluminados graças ao sensor infravermelhos. A principal desvantagem em relação às restantes tecnologias é o campo de visão inferior.

2.5 Robôs de serviços

Todos os estudos até agora demonstrados, focaram-se na deteção de obstáculos sobretudo para providenciar uma navegação autônoma segura e robusta, mas para a robótica de serviços isto por si só não é suficiente. Um dos problemas fundamentais da robótica de serviços é a interação humano-robô. Acima de tudo este tipo de sistema deve possuir a capacidade de detetar, seguir e mostrar reconhecimento da presença de humanos.

Em [27] foi proposto um método para a deteção e o seguimento de humanos. Segundo este estudo, as tecnologias mais utilizadas para este mesmo efeito são sensores LiDAR ou câmaras. Devido às limitações de cada uma das tecnologias, decidiram recorrer a uma fusão sensorial para obter melhores resultados. O método utilizado foi um sensor LiDAR para a deteção de pernas e uma câmara monocular para a deteção de rostos.

Para a deteção de pernas, o sistema, recorre a um algoritmo baseado no reconhecimento de um padrão de pernas para efetuar a deteção das mesmas. Este padrão pode ser usado tanto para localizar pessoas paradas, como pessoas a caminhar, e pode ser usado tanto em espaços largos e vazios como em espaços pequenos e desordenados. Existem 3 padrões possíveis: pernas afastadas (LA), pernas juntas ou só uma perna (FS) e a posição de caminhada, ou seja, uma perna para a frente e outra para trás (SL). Esta abordagem permite assim distinguir posturas de pernas diferentes e reduzir o erro de falsos positivos.

Para a deteção de rostos, é utilizada uma câmara monocular para a aquisição das imagens em conjunto com o algoritmo *Viola and Jones* [28], com uma pequena extensão proposta por [29]. Este algoritmo utiliza um conjunto de características visuais, com formatos específicos (Figura 7), que podem ter dimensões e posições arbitrarias dentro de uma determinada imagem.

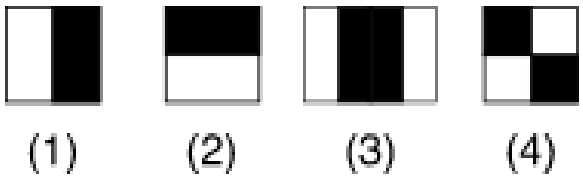

Figura 7 - Conjunto características visuais propostas por Viola and Jones, [30] .

O valor de cada uma das características sobre uma imagem é calculada pela subtração do somatório de todos os pixéis na região preta, pelo somatório de todos os pixéis na região branca.

Para além disto, o sistema de deteção consiste numa cascata de classificadores, sendo cada um relativo a uma determinada característica. Isto permite a deteção de casos negativos rapidamente, pois estes são rejeitados muito antes de chegar ao último nível da cascata. Em cada etapa, o classificador torna-se mais preciso. Desta forma, para ser detetado um rosto de uma sub-região de uma imagem, esta deve percorrer todas as camadas da cascata. Caso seja rejeitado por um classificador, este é logo descartado. O treino dos classificadores é realizado com recurso ao algoritmo AdaBoost [31].

O sistema de seguimento de humanos recorre a um filtro Kalman, que une a informação proveniente da câmara e do LiDAR de forma a fornecer uma deteção e um seguimento robusto. Na Figura 8, em baixo, podemos ver nas linhas pretas a informação adquirida do sensor LiDAR relativamente a deteção 3 humanos com diferentes padrões (LA, FS e SL). Em cima podemos ver a deteção de rosto de cada uma das pessoas.

Este sistema foi testado em ambientes interiores desordenados, onde a deteção de humanos é difícil devido a presença de objetos ou até mesmo ao tamanho reduzido do espaço, e os resultados indicam ser robusto o suficiente para seguir humanos mesmo quando o robô se movia a uma velocidade aproximada à velocidade de caminhar dos humanos. Neste estudo a deteção de pessoas poderia ainda ser melhorada, pois nem todas as orientações dos humanos em relação ao robô permite a deteção de rostos.

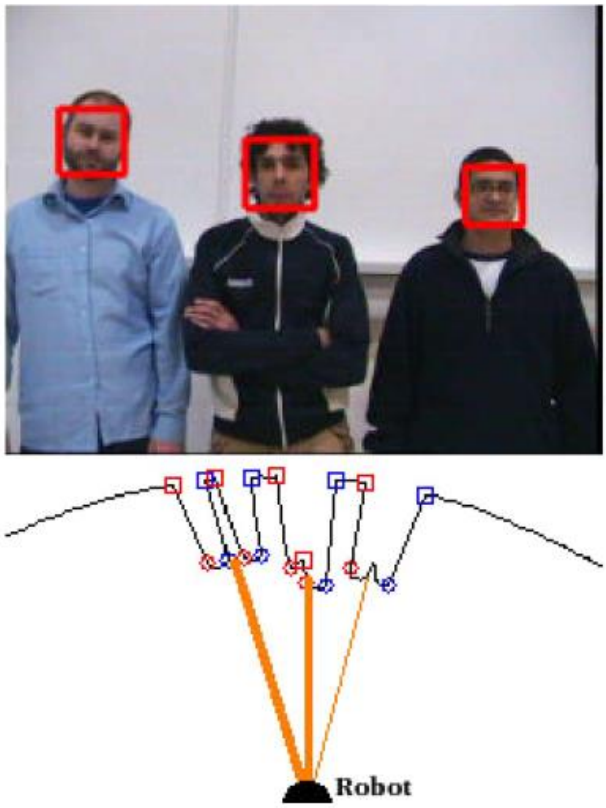

Figura 8 - Deteção de humanos. Em cima a deteção de rostos e em baixo deteção de pernas, [27].

Em [32] é também apresentado um método para efetuar o seguimento de humanos, desta vez com uma abordagem diferente, que foi a utilização de uma câmara térmica para a deteção de humanos através do calor corporal. Esta abordagem permite a deteção de pessoas independentemente das condições de luminosidade ou visibilidade da cor da pele. A câmara utilizada foi uma Tracer TS7302 NEC, que possui capacidade de detetar radiação infravermelha e converter esta informação numa imagem, onde cada pixel corresponde a um valor de temperatura. Depois de a imagem capturada, a deteção de pessoas é efetuada com recurso a dois modelos: *Elliptic contour model* e *Feature model*. O primeiro modelo é utilizado para estimar a posição de uma pessoa na imagem, utilizado duas metades de elipses: uma elipse maior para estimar a posição da região dos ombros e uma mais pequena para estimar a posição da cabeça. Como podemos visualizar na Figura 9, este modelo permite detetar pessoas mesmo quando estas se encontrem de lado para a câmara.

O segundo modelo, *Feature model*, é utilizado o algoritmo *Viola and Jonas* para encontrar características na imagem que permitam detetar pessoas, baseando-se nas características dos valores de cinza. Como este apresenta um cálculo mais rápido, este mesmo foi utilizado para melhorar o desempenho do primeiro modelo. Este começa por tentar localizar regiões na imagem que apresentem

as mesmas características dos valores de cinza pertencentes a uma pessoa. Uma vez encontradas essas regiões, é aplicado o primeiro modelo de modo a confirmar se era uma pessoa ou não.

Figura 9- Deteção de humanos com uma câmara térmica, retirado de [32].

Este estudo obteve bons resultados na deteção e no seguimento no caso de uma pessoa. Quando se encontra na presença de múltiplas pessoas, o desempenho do sistema dependia da interação das mesmas. Neste tipo de sistemas é difícil seguir ou detetar pessoas quando estas se cruzam entre si, pois a imagem obtida é quase como se fosse uma só. Contudo, o sistema recupera bem a uma falha de seguimento.

Para além destes estudos existe ainda algumas soluções comerciais disponíveis para venda ao publico em geral. Um desses exemplos é o robô Care-O-Bot 3 da Fraunhofer [33], que disponibiliza diferentes configurações de robôs mediante o tipo de aplicação. As aplicações podem ser domésticas (executar varias tarefas domésticas), serviços (entregar pedidos em restaurante ou providenciar um serviço de quartos num hotel) ou indústria (tarefas de recolha e entrega de materiais ou até mesmo de carga e descarga).

Na Figura 10 podemos visualizar a configuração atual do robô. O sistema de visão é composto por duas câmaras 3D com princípios de funcionamento diferentes, uma utiliza tecnologia tempo de voo (TOF) e outra visão-estéreo. Apesar de as duas possuírem diferentes tecnologias, ambas permitem a perceção em profundidade do ambiente em redor permitindo assim ao robô a identificação, localização e o seguimento de objetos e pessoas. A principal vantagem da fusão dos dados destas duas tecnologias é a robustez do sistema, pois cada uma das tecnologias apresenta um maior erro para uma determinada superfície. Estes sensores estavam montados numa unidade com 5 *DOF* (graus de liberdade),

permitindo assim ao robô visualizar qualquer área do seu interesse. Este sistema de visão permite ao robô identificar, localizar e seguir objetos e pessoas em 3D permitindo uma supervisão do ambiente 3D.

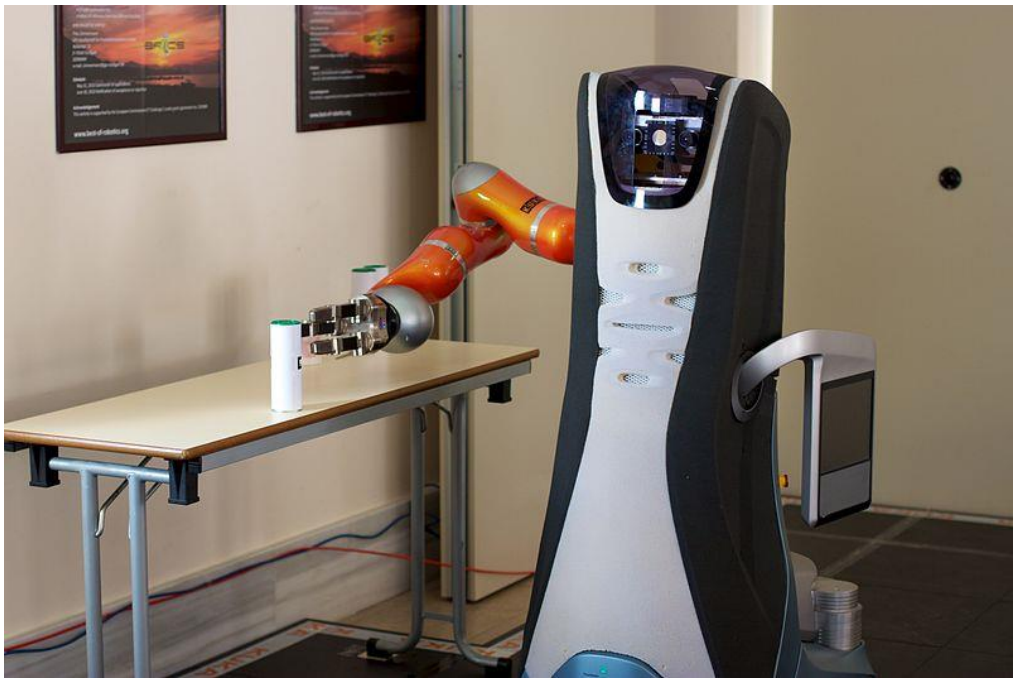

Figura 10- Care-o-Bot, [34].

A deteção de obstáculos para o Care-o-Bot 3 é baseada na criação de uma "imagem partilhada", que integra informação de profundidade e cor. Através da determinação do pixel correspondente entre a imagem de profundidade e a imagem de cor, é possível calcular o mapeamento de coordenadas entre a imagem do sensor de profundidade e do sensor a cores. Esta informação 3D resultante é guardada na imagem partilhada, sendo esta a base do da deteção de objetos. Este processo pode ser dividido em duas etapas essenciais: treino de objetos e deteção de objetos.

A primeira etapa é o treino de objetos, pois o robô necessita de uma representação concreta do mesmo para a deteção de objetos num ambiente desconhecido. Com esse objetivo, o objeto é colocado na pinça do robô, de forma a que este rode o pulso e retire um conjunto de imagens do objeto, retirando toda a informação necessária para o procedimento de treino. Uma vez retirada a informação do objeto, é aplicada uma segmentação das imagens. Uma vez que a posição aproximada do objeto é conhecida na imagem 3D, é definida uma esfera 3D a volta do mesmo, de modo a aplicar uma segmentação ao objeto mais rápida, pois só os pontos localizados dentro da esfera são considerados parte do objeto.

Depois da etapa de segmentação é realizada a extração de características relevantes do objeto na imagem. O método proposto é baseado numa aproximação rápida do algoritmo SIFT [35], em que este aplica descritores distintivos invariantes a mudanças de iluminação, escala, rotação, distorção e no ponto

de vista. Através da aplicação de um filtro DOG (*Difference of Gaussian)*[3] nos dados da imagem de cor e profundidade, os pontos característicos que apresentam um alto contraste com o fundo são extraídos.

Uma vez extraídas todas as características, é realizada a etapa final do processamento da imagem em que associa uma imagem 6D para representar a posição e orientação do objeto. Para cada ponto de características é estabelecido um quadro de coordenadas cartesianas. A combinação de todos os quadros de todas as imagens retiradas, permite a criação de uma nuvem de pontos 6D de características do objeto, fornecendo a representação aproximada do formato do objeto.

O reconhecimento de um objeto num ambiente desconhecido é realizado através de uma abordagem baseada em votação. Em primeiro são extraídas as características distintivas de uma determinada imagem ou uma região de interesse. Depois esses descritores são comparados com descritores de objetos previamente treinados, em que a cada correspondência é incrementado um contador para um determinado objeto. Assim que esse contador atinge um determinado nível, o objeto é considerado como detetado.

A localização do robô é baseada em odometria visual, contudo pequenos erros são inevitáveis e acumuláveis ao longo do tempo. Para isto, são utilizadas as informações dos sensores LiDAR (colocados à frente e atrás), de modo a detetar características significantes como por exemplo paredes. Estas características são comparadas com as suas posições de referência e a posição do robô é determinada em relação as mesmas.

Apesar desta solução comercial aqui apresentada possuir a capacidade de executar tarefas do dia-a-dia de modo a trazer alguma comodidade às pessoas, existem ainda outras soluções com diferentes propósitos que não possuem essa capacidade, mas que requerem a interação com humanos. Exemplo disto, é o robô SPENCER [36], que é o resultado de um projeto europeu de investigação denominado FP7 na área da robótica com uma parceria de 6 universidades, em que o seu propósito é assistir, informar e guiar grupos de pessoas através de ambientes dinâmicos e sobrelotados como aeroportos. Um dos principais objetivos é conduzir eficientemente passageiros de voos com escalas do seu terminal de chegada até a zona de controlo de passaportes.

Tendo em conta o propósito do robô, um dos componentes cruciais é uma deteção e um seguimento robusto de humanos no ambiente em redor. O Spencer foi dotado com 2 sensores LiDAR 2D "SICK LMS 500", um sistema de câmaras estéreo à altura dos ombros e 4 câmaras Kinect v1do tipo

[3] DOG (*Difference of Gaussian)* é um algoritmo que realiza o realce de características numa imagem.

RGB-D (*Red Green Blue – Depth*), em que cada uma delas providencia uma imagem a cores (RGB) e outra de profundidade (D). Cada um dos sensores LiDAR efetua um varrimento de 190° e a junção dos dois permite uma perceção 360° em todo o seu redor a uma altura de 0,65 m. No entanto cada uma das tecnologias apresenta vantagens e desvantagens, pois enquanto o LiDAR 2D providencia uma perceção em todo o seu redor e é mais robusto a mudanças na iluminação este não consegue fornecer nenhuma informação sobre a aparência do obstáculo, ficando ainda limitada a altura a que este se encontra. Já a Kinect possui um sensor de profundidade do tipo Luz estruturada com um alcance máximo de 4 m, sendo desta forma obrigatório recorrer a imagem RGB para a deteção de pessoas a distâncias superiores. Assim sendo, foram utilizados múltiplos algoritmos de deteção e seguimento que operam em diferentes sensores, criando assim alguma redundância no sistema.

Para a deteção de pessoas com o LiDAR 2D, os pontos de dados recebidos são primeiramente segmentados recorrendo a um método de *agglomerative hierarchical clustering* [37], em que os pontos são agrupados em hierarquia mediante a proximidade de cada um, começando por agrupar sempre os mais próximos, durante as várias etapas.. Depois disso, é aplicado um classificador otimizado que foi previamente treinado manualmente em 9535 *frames* de dados. As deteções resultantes são seguidas recorrendo a um filtro Kalman com um modelo de movimento de velocidade constante para efetuar o seguimento e prever o estado das pessoas.

A deteção e o seguimento de pessoas com recurso ao sensor RGB-D pode ser dividido em duas partes. Para distâncias próximas, foi desenvolvido um sistema que recorre unicamente a informação de profundidade de modo a acelerar a computação e a providenciar um sistema em tempo-real. Este sistema classifica os pontos 3D observados em objetos candidatos, chão e estruturas fixas, como por exemplo paredes. Os pontos do chão são utilizados para estimar o plano do chão, e os objetos candidatos são analisados por um detetor da parte superior do corpo [38], que recorre a um *template* de um modelo normalizado de profundidade com aprendizagem para encontrar a região da cabeça-ombro. Este sistema fica limitado ao alcance máximo do sensor de profundidade da Kinect. Apesar de este sensor apresentar um alcance de 4 m, na prática quanto colocado a uma distância superior de 2 m de uma pessoa, a perceção da cabeça não é tão eficaz, dificultando o processo de deteção por parte do algoritmo.

Para a deteção de pessoas a distâncias superiores foi implementado um detetor de pessoas baseado no algoritmo HOG (*Histogram of Oriented Gradients*). Segundo [39] e [40], a deteção pode ser dividida em quatro etapas. Cálculo do gradiente de cada pixel, ou seja, da orientação e magnitude das arestas da imagem; Divisão da imagem em blocos e células em que os pixéis são agrupados em células e as células agrupadas em blocos; Cálculo do histograma de gradientes em cada célula, sendo

depois agrupado em blocos; Junção destes histogramas, originando o descritor. Para o reconhecimento de objetos é ainda necessário um sistema de reconhecimento baseado em aprendizagem supervisionada, como por exemplo o SVM (Support Vector Machines) [41], que recebe o descritor e que classifica como sendo uma pessoa ou não pessoa.

Este segundo detetor opera eficientemente sobre a unidade de processamento gráfico e usa o plano do chão estimado para restringir a pesquisa por objetos em regiões geometricamente validas, diminuindo assim o tempo de processamento

Para a integração da camada de hardware com o sistema foi utilizado o *middleware* ROS [42].

3. METODOLOGIAS

3.1 Ferramentas

Para a realização deste estudo recorreu-se a algumas ferramentas de apoio. Toda a comunicação com os sensores e atuadores, tanto da parte do controlador como do simulador do modelo real, será realizada com recurso ao *middleware* ROS. Já o tratamento de imagens é realizado com recurso a ferramenta OpenCV.

3.1.1 ROS

A plataforma ROS (Robot Operating System) surgiu da necessidade de integrar soluções comuns na área da robótica. Diferentes tipos de robôs utilizam diferentes tipos de *hardware,* e isto dificulta o processo de desenvolvimento de software. Desta forma, o ROS providência uma camada de abstração de hardware, fornecendo ainda bibliotecas e ferramentas para ajudar a criação de aplicações robóticas, como por exemplo, *drivers,* algoritmos e a comunicação por mensagens. Para além disto, o ROS é compatível com várias linguagens de programação já existentes, dando assim uma maior flexibilidade ao desenvolvimento de *software*.

As comunicações no ROS baseiam-se no modelo *peer-to-peer* (P2P), pois em casos que hajam múltiplos nós a propagar informação este modelo apresenta melhores resultados do que os modelos centralizados. Apesar das comunicações basearem-se no modelo P2P, existe um nó principal responsável pelo registo de novos nós e informar quais os que se encontram ativos, seguindo sempre o modelo centralizado [42]. Desta forma, um nó que se registe na rede pode fornecer ou receber mensagens/dados de outros nós, podendo estes ser algoritmos como drivers para a camada de hardware.

Cada um destes nós que se regista na rede pode publicar vários tipos de dados, publicando um tópico para cada um deles. Estas ligações em que qualquer nó pode publicar e fornecer dados não são apropriadas para interações entre nós de pedido/resposta. Para isto, cada nó dispõe de serviços, que são definidos por um par de mensagens: uma para o pedido e outra para a resposta. Um nó que forneça um serviço sob uma *string* nome, pode ser acedido por um cliente que chama o serviço enviando uma mensagem de pedido e aguardando a sua resposta. Para além disto, cada nó pode possuir ainda parâmetros, que são visíveis globalmente e permitem algumas configurações dos nós.

A Figura 11, representa um exemplo dessa mesma comunicação. Um nó chamado de "LiDAR 2D" publicou um tópico "scan". Este nó começa por registar o seu tópico no nó principal e indicar qual é o seu ponto de acesso. Outro nó chamado "Aplicação" pergunta ao nó principal como aceder ao tópico "scan", obtendo como resposta o ponto de acesso desse mesmo serviço. A comunicação entre os dois nós ("LiDAR 2D" e "Aplicação") é estabelecida através de uma comunicação TCP ou UDP, sem que o nó principal faça parte dessa mesma comunicação.

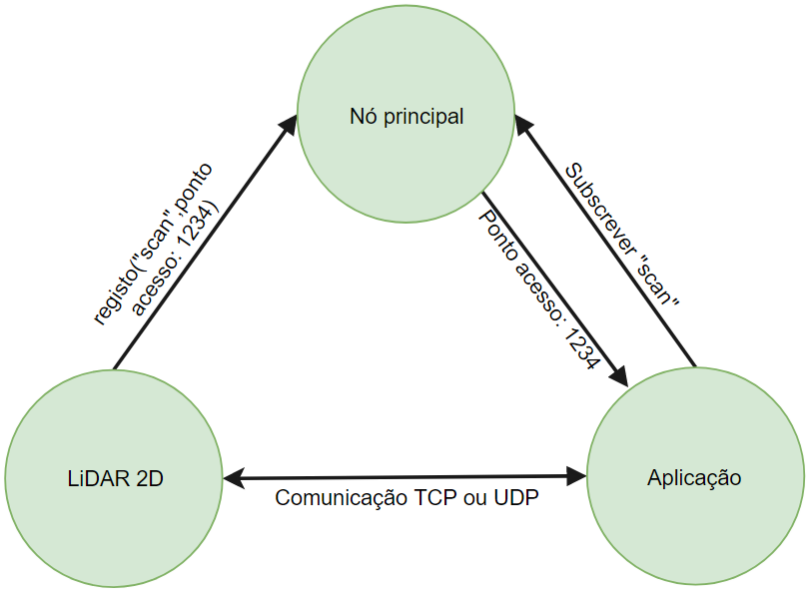

Figura 11 - Modelo comunicação ROS.

Desta forma, este *middleware* permite a comunicação com toda a camada de hardware assim como o desenvolvimento dos algoritmos de processamento e atuação.

Configuração de transformadas ROS:

Um sistema robótico tipicamente é constituído por várias tecnologias de sensorização ou outros tipos de mecanismos de atuação. Como todos partilham o mesmo espaço e operam em conjunto, há a necessidade de representar a posição e a orientação de cada um deles num mapa de coordenadas relativo ao sistema. No caso de sistemas de perceção ambiente, em que recorrem a uma fusão sensorial, é um ponto fulcral pois o sistema necessita da posição e orientação de cada um dos sensores para a determinação da área que cada um opera. Exemplo disto, pode ser a deteção de uma pessoa por dois sensores com diferentes orientações, em que caso não haja este sincronismo a determinação da posição exata da pessoa pode não ser possível.

De modo a simplificar a integração dessas várias tecnologias, o ROS já dispõe de uma biblioteca denominada "tf", que permite manter a relação de coordenadas entre os vários sensores e atuadores, e a configuração de novos sensores e atuadores ao sistema. A relação dos sensores é publicada numa estrutura em árvore e dela podemos retirar todas as informações como pontos de transformadas ou vetores como por exemplo a relação de coordenadas entre dois sensores a qualquer instante de tempo.

A utilização desta biblioteca permite a execução de duas tarefas importantes na implementação de sistema interno de coordenadas: emitir e receber transformadas.

Emissão – a pose relativa de quadros de coordenadas é enviada para o resto do sistema.

Receção - é realizada a receção e o armazenamento de todos os quadros de coordenadas que são transmitidos no sistema, podendo ainda solicitar transformadas específicas entre cada um dos quadros.

Apesar de ser principalmente uma biblioteca de código para ser utilizada com os nós ROS, ela já providencia um conjunto de comandos, que podem ser utilizados na linha de comandos do sistema operativo, de modo a monitorizar e a ajudar a criação de novos mapas de coordenadas.

Na Figura 12, podemos visualizar um exemplo de uma estrutura em árvore das relações entre coordenadas de um sistema com uma câmara Kinect colocada sobre este. A raíz (base_footprint) é uma projeção no plano do chão ($z = 0$) da posição central do sistema, em que neste caso deriva a posição central de uma câmara kinect (rgbd_front_top_link). Como esta câmara possui um sensor de profundidade e outro sensor a cores com diferentes posições, derivam duas transformadas da posição central da câmara (rgbd_front_top_rgb_frame e rgbd_front_top_depth_frame). De cada uma das transformadas da posição dos sensores, deriva ainda a posição do ponto de captura da imagem em cada um dos sensores (rgbd_front_top_rgb_optical_frame e rgbd_front_top_depth_optical_frame).

É ainda possível visualizar a taxa de publicação de cada uma das transformadas, quem está a publicar, a que tempo foi publicada a transformada mais recente e o tamanho do buffer de armazenamento de transformadas.

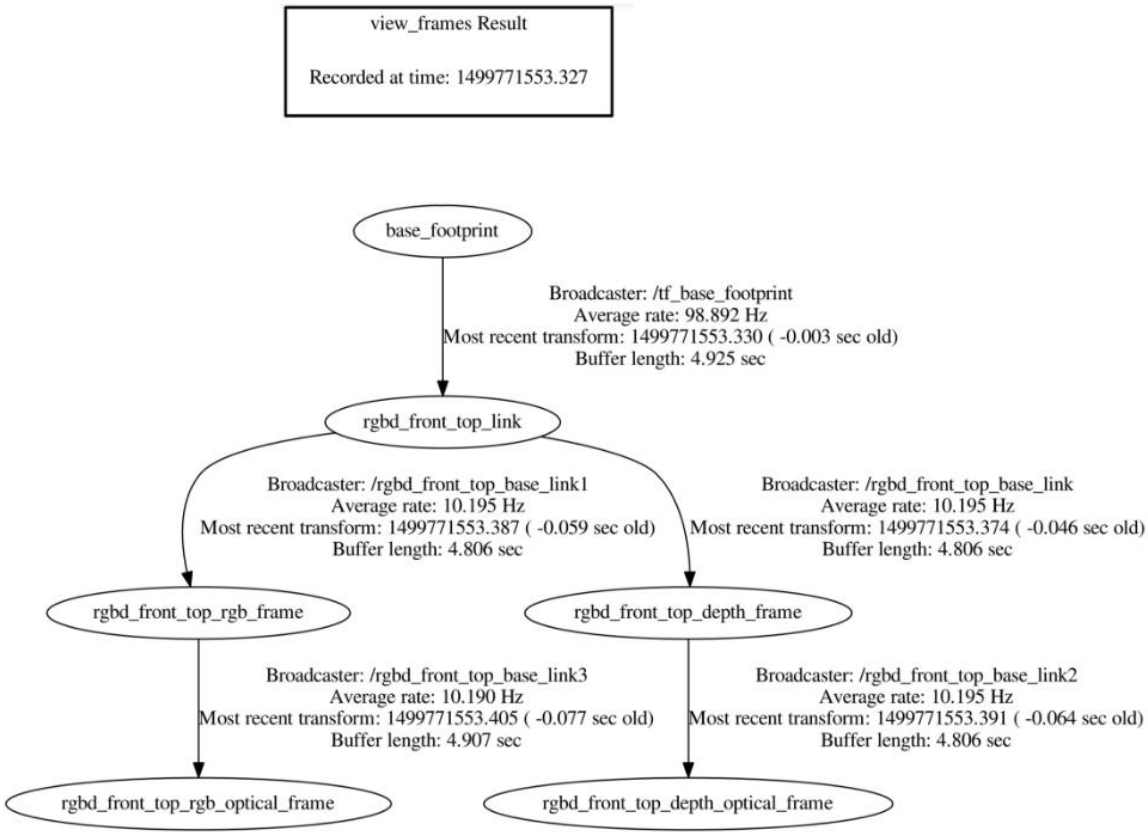

Figura 12 - Exemplo de publicação de transformadas internas de uma Kinect v1 montada num sistema.

Supondo que o ponto central da câmara se encontra com a mesma orientação e apenas com uma translação em z de 1.6 m em relação ao pornto central do sistema "base_footprint". Quando evocado o comando tf_echo na linha de comandos podemos visualizar a translação, e a rotação entre o ponto central da câmara ao ponto central do sistema, Figura 13.

```
Ubunto@Ubunto:~$ rosrun tf tf_echo /base_footprint /rgbd_front_top_link
At time 1499771553.327
- Translation: [0.000, 0.000, 1.600]
- Rotation: in Quaternion [0.000, 0.000, 0.000, 1.000]
in RPY [0.000, 0.000, 0.000]
```

Figura 13 – Visualização da relação entre duas coordenadas com recurso a biblioteca tf do ROS.

Packages ROS

O *software* no ROS encontra se organizado por *packages*. Estes podem conter nós ROS, bibliotecas independentes do ROS, ficheiros de configuração, software desenvolvidos por terceiros ou algo que constitua um módulo lógico e útil. O principal objetivo desta organização por *packages* é permitir uma fácil reutilização de *software*.

Os ROS *packages* seguem tipicamente uma estrutura comum. Estes são os diretórios ou ficheiros que podemos encontrar:

- include/package_name: diretório onde podemos encontrar os *header files c++;*
- Msg: Diretório que contêm tipo de mensagens (msg). Isto permite ao ROS saber o tipo de mensagens que têm que ser geradas, para serem publicadas por cada um dos nós;
- src/package_name: diretório que contêm os *Source files*;
- srv/: Diretório que contêm os tipos de serviços ROS. É compilado diretamente sobre o ROS *msg format* para permitir a comunicação pedidos/respostas entre os nós;
- scripts: pode conter algumas scripts executáveis;
- CMakeLists.txt: este ficheiro TXT descreve como compilar o código, como por exemplos todas as dependências necessárias, testes ou regras;
- Package.xml: é um ficheiro no formato XML que é obrigatório ser incluído no diretório raiz. Ele define as propriedades sobre o package, como por exemplo o nome do package, autor, numero de versão, e dependências em outros packages;
- CHANGELOG.rst: alguns packages incluem este ficheiro que contém algumas informações relativas aos ficheiros, como lista de atualizações realizadas;

3.1.2 Biblioteca OpenCV

A biblioteca OpenCV *(Open Source Computer Vision Library)* [26], é uma biblioteca gratuita que providencia um conjunto de classes e funções, permitindo a criação de aplicações de forma simples na área de visão computacional. Para além disto, é compatível com múltiplos sistemas operativos, como por exemplo o Linux, Windows e Mac OS X.

Apesar desta se encontrar dividida em cinco componentes principais [43], nesta dissertação só irá ser utilizado o componente CV, pois este é constituído por algoritmos de processamento de imagem e visão computacional de alto nível.

3.2 Sensorização

Existe uma vasta gama de sensores utilizados em robótica móvel. Alguns sensores são utilizados para a leitura de valores simples como a temperatura interna de um robô ou a velocidade rotacional do motor. Outros, sensores mais sofisticados podem ser utilizados para adquirir informação sobre o ambiente em que o robô opera ou até mesmo para medir diretamente a posição global do robô. Neste subcapítulo irá ser focado primeiramente em sensores com a capacidade de extrair informação do ambiente em redor do veículo, para uma deteção e identificação de obstáculos robusta. Como os veículos autónomos operam maioritariamente sobre ambientes dinâmicos, em que as características do ambiente são na maior parte das vezes desconhecidas, a sensorização torna-se um dos pontos fundamentais. Primeiramente irá ser feita uma classificação funcional dos sensores e em seguida será apresentada uma seleção de tecnologias, tendo em conta o tema desta dissertação e a capacidade de uma boa perceção do meio.

Apesar das inúmeras vantagens que cada uma das tecnologias apresentadas neste subcapítulo dispõem, por si só não são suficientes para garantir segurança, robustez e um bom reconhecimento do meio. Para a resolução desse problema muitas vezes recorre-se a uma fusão sensorial, pois cada uma das tecnologias escolhidas têm capacidades distintas, que em conjunto, permitem uma reconstrução precisa do ambiente em redor do veículo.

3.2.1 Classificação de Sensores

Podemos classificar os sensores como sendo Propriocetivos ou Exterocetivos e Ativos ou Passivos.

Sensores Propriocetivos fornecem medidas dos valores internos do sistema (veículo). Como por exemplo o estado da bateria, a velocidade do motor, ângulos de juntas de rotação ou até mesmo informação sobre a posição do veículo (encoders).

Sensores Exterocetivos [44] adquirem informação do ambiente em redor do veículo ou outro sistema. Como por exemplo, medidas de distância intensidades de luz, amplitude do som. Consequentemente as medidas dos sensores são interpretadas pelo sistema para extrair características do ambiente relevantes. Estes são geralmente os mais utilizados para a deteção de obstáculos.

Sensores Passivos são caracterizados por utilizar a energia já presente no ambiente para obter as suas medidas [45]. Este tipo de sensores são utilizados preferencialmente em aplicações militares pois evita serem detetados facilmente. Exemplos deste tipo de sensores pode ser sondas de temperatura, microfones, e câmaras CCD ou CMOS.

Sensores Ativos emitem energia para o ambiente, que tipicamente é refletida de volta para o sensor ou então é medida a reação no ambiente. Como este tipo de sensores interage de uma forma controlada com o ambiente oferecem um desempenho superior. No entanto, os sensores ativos introduzem sérios riscos: a energia emitida pode afetar as características que o sensor está a tentar medir do ambiente. Além disso, um sensor ativo pode sofrer de interferência entre o seu sinal e outros além do seu controlo, como por exemplo, sinais emitidos por outros sistemas próximos, ou sensores similares no mesmo sistema, influenciando as medidas adquiridas. Exemplos de sensores ativos podem ser sensores ultrassónicos, LiDAR e Radar.

3.2.2 LiDAR

Os sensores LiDAR são sensores ativos muito utilizados em veículos autónomos para a detenção de obstáculos [1]. O princípio de funcionamento é similar tanto para o LiDAR 2D como para o 3D: um conjunto de pulsos é emitido pelo LiDAR e no caso da presença de algum obstáculo o pulso é refletido de volta. Uma vez recebido esse pulso refletido, a distância é estimada (Equação (7)) baseando-se no intervalo de tempo que o pulso demorou a atingir o obstáculo e a ser refletido de volta, Figura 14.

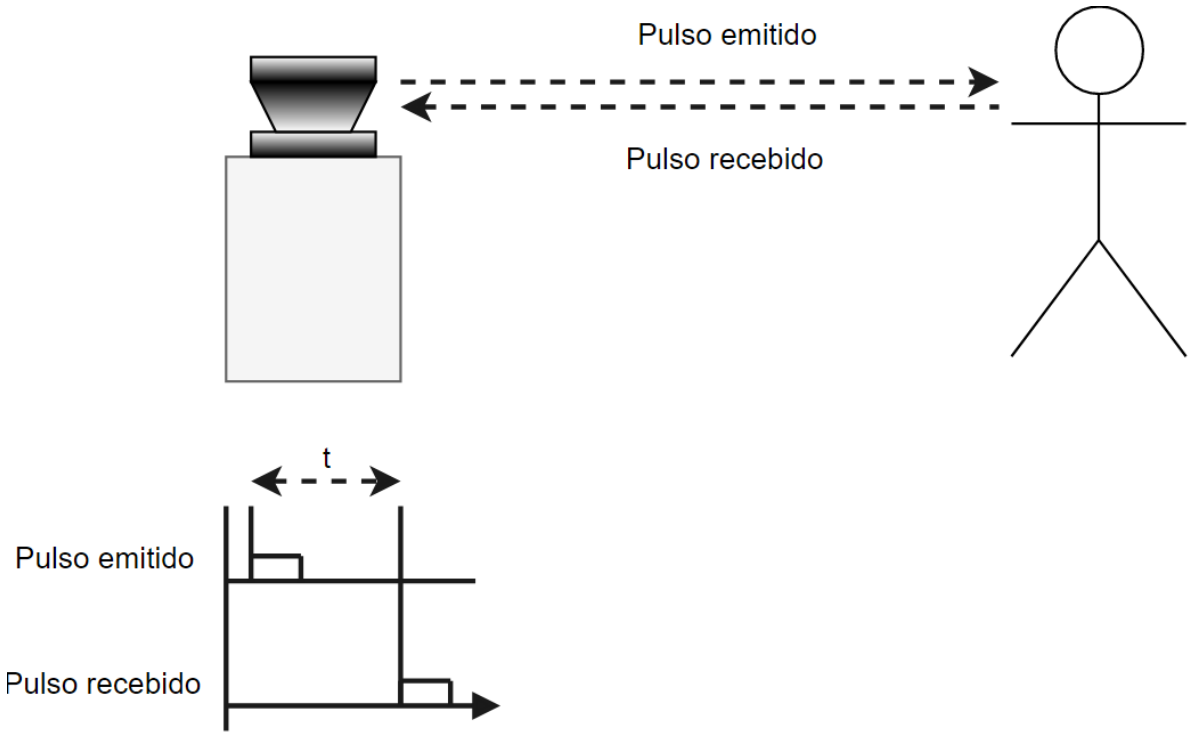

Figura 14- Principio funcionamento sensor LiDAR.

De acordo com [1]:

$$d = v_{luz}\ \frac{t}{2}, onde\ v_{luz} = 3{,}0 \times 10^{8} m/s \qquad (1)$$

Onde:

d é a distância entre o LiDAR e o objeto.

v_{luz} é a velocidade da luz.

t é o tempo de voo.

Para além da distância estimada, baseada no método tempo de voo, a amplitude do sinal recebido pode ser utilizada para caracterizar a refletividade da superfície do objeto [1]. A Figura 15 apresenta o exemplo de um LiDAR 2D. Este é um PLS 101/201 da SICK, que é uma solução muito utilizada no que toca a deteção de obstáculos em ambientes interiores. Possui um campo de visão de 180° e um alcance máximo de 50 m.

Figura 15 - LiDAR 2D, SICK PLS 101/201, [46].

Esta tecnologia apresenta como principais vantagens uma baixa sensibilidade a condições de visibilidade reduzida (nevoeiro, chuva ou noite), não requer muito processamento de dados, possui uma elevada precisão de leituras e, segundo [1], o campo de visão do LiDAR é tipicamente superior (geralmente 180°) quando comparado a outras tecnologias como sistemas de visão ou sistemas Radar.

Por outro lado, este tipo de sistemas apresenta uma distância de deteção limitada e no caso do LiDAR 2d existe a dificuldade de caracterização de objetos. É ainda importante referir que este tipo de

sistemas é geralmente cego em superfícies transparentes como o vidro e pode ser ofuscado pelo sol direto. Existe ainda a possibilidade de interferência mútua com outros sistemas de sensorização com o mesmo princípio de funcionamento (como por exemplo, a receção de outros pulsos provenientes de outros LiDAR presentes no mesmo ambiente), podendo levar a um rácio *signal-to-noise*[4] *(SNR)* reduzido e a alvos inexistentes [47]. Contudo existem alguns sensores LiDAR, como o S300 da SICK, que modulam os pulsos emitidos para o ambiente de forma a evitar interferências com outros sensores.

As ofertas de mercado são inúmeras, podendo ir de uma perceção de 2 a 3 dimensões. A vantagem do LiDAR 3D em relação ao 2D é para além de incidir sobre um plano horizontal como o 2D, incide também sobre um plano Vertical, permitindo desta forma efetuar a deteção e a distinção de obstáculos, através do mapeamento de profundidades na zona de visualização. Existem algumas configurações que permitem uma visão ampla de 360°, pois recorrem a um suporte rotativo debaixo do sistema sensorial, permitindo assim efetuar um varrimento de toda a área em redor (Figura 16 e Figura 17Figura 16).

Figura 16- RPLiDAR - LiDAR 2D com perceção de 360º.

[4] Signal-to-noise ratio (SNR) – É uma medida utilizada em ciência e engenharia que compara o nível de um determinado sinal em relação ao nível de ruído presente no mesmo.

Figura 17- FARO LS, LiDAR 3D com percepção 360º, [48].

3.2.3 Sonar

Os sensores Sonar *(Sound navegation and ranging)* são caracterizados por serem sensores ativos que emitem pulsos acústicos [49], em frequências que não são percetíveis ao ouvido humano. Uma vez que esse pulso colide com o obstáculo, é refletido de volta, como ilustrado na Figura 18. Como é o caso do LiDAR, o sistema de sensorização captura esse sinal refletido e processa o sinal para obter informação sobre o objeto que o refletiu de volta.

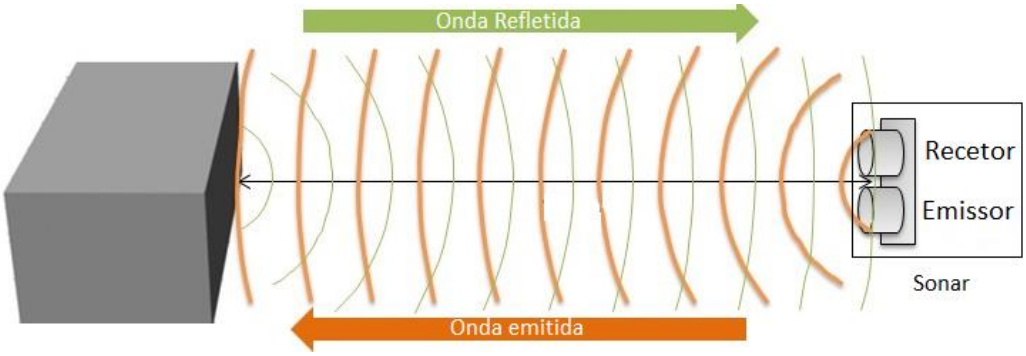

Figura 18- Principio funcionamento Sonar.

A sua aplicação mais comum é a medição de distâncias, pois uma vez obtida a diferença de tempo entre a emissão e a receção de um pulso é possível obter a distância entre o sensor e o obstáculo detetado. A distância pode ser obtida de acordo com a Equação (2):

$$d = v_{som}\ \frac{t}{2} \tag{2}$$

Onde:

d é a distância entre o Sonar e o obstáculo detetado.

v_{som} é a velocidade do som.

t é o tempo de voo.

A velocidade do som depende não só apenas do meio, mas também da sua temperatura. Desta forma, se a temperatura não for levada em conta no cálculo da distância, esta poderá conter um erro associado. Segundo [49], a velocidade exata do som pode ser calculada, tendo em consideração a temperatura do meio, da seguinte forma, Equação (3):

$$v_{som} = v_{som\ 0^{o}C} + 0.6T \tag{3}$$

Onde:

v_{som} é a velocidade do som no ar de acordo com a temperatura T.

$v_{som\ 0^{o}C}$ é a velocidade do som no ar a 0°C, sendo esta 331 m/s.

T é a temperatura do meio.

Apesar da sua aplicação mais comum ser para medição de distâncias, é possível usar o sensor para medição de velocidades de um alvo. O cálculo da velocidade é efetuado através do efeito de Doppler, em que é observada a diferença de frequência entre uma onda emitida e recebida. Uma explicação mais detalhada pode ser encontrada em [50]. Este tipo de tecnologia tem uma precisão na ordem dos 3 cm sobre um alcance máximo de 10 m. Apresenta algumas vantagens, como por exemplo, não requerer um grande esforço computacional pois os dados recebidos não requerem grande tratamento e apresenta um baixo custo. Por outro lado, este tipo de tecnologia não providencia capacidade discriminatória do obstáculo e é muito sensível a variações das condições ambientais (levando a problemas de ruído e distorção). Para além disto, quando o som é emitido sobre uma superfície com uma pequena inclinação, parte do sinal será refletido para longe, e dependendo do tipo de material só uma pequena quantidade de energia poderá ser refletida de volta, podendo causar falhas na deteção de obstáculos. A presença de múltiplos Sonar no mesmo ambiente, poderá causar interferências, originando um baixo rácio entre sinal

e ruído (SNR). A sua principal aplicação na indústria é a deteção de obstáculos. Na Figura 19 podemos visualizar um exemplo de um sensor Sonar da TOUGHSONIC que apresenta um alcance máximo de 9.1 m.

Figura 19- Sensor Sonar TOUGSONIC 30, [51].

3.2.4 Radar

Os sensores Radar apresentam o mesmo princípio de funcionamento dos sensores Sonar (como se pode ver na Figura 20), sendo a principal diferença o tipo de onda emitida. As ondas emitidas são ondas do espectro eletromagnéticas do tipo rádio, e quando uma onda é transmitida só uma pequena porção da energia dessa onda retorna ao sensor Radar [44].

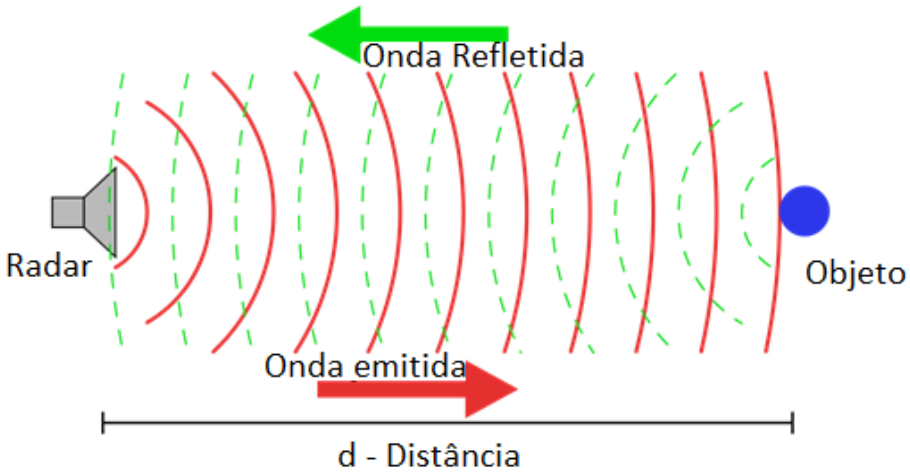

Figura 20 - Principio de Funcionamento de um Sistema Radar.

A distância entre sensor e o objeto refletido pode ser calculada de acordo com a equação (4):

$$d = v_{luz}\ \frac{t}{2} \tag{4}$$

Onde:

d é a distância entre o Radar e o objeto refletido.

v_{luz} é a velocidade da luz.

t é o tempo de voo.

A diferenças nas propriedades destas ondas em relação as ondas acústicas tornam o Radar com respostas muito mais rápidas (as ondas rádio viajam a velocidade da luz) e um alcance superior. Segundo [52], a frequência utilizada neste tipo de sistemas varia entre os 5 MHZ e os 95 GHZ, contudo, poderá utilizar frequências fora desta gama com sucesso. As características do sistema Radar dependem da frequência que é utilizada. Uma frequência baixa permite uma deteção de alvos a longas distâncias enquanto que uma frequência alta permite uma precisão superior na deteção e localização devido a uma maior largura de banda. Como é o caso do Sonar, é possível utilizar a tecnologia Radar para medir a velocidade de um determinado alvo, considerando o efeito de Doppler.

Este tipo de tecnologia pode ser encontrado, não só em aplicações militares, mas também no nosso dia a dia, como por exemplo: controlo de tráfego (aéreo, marítimo ou terrestre), sistemas de vigilância de pequena e grande escala, previsões meteorológicas, sistemas de controlo adaptativo de velocidade, medições geológicas, entre outros.

Apresenta a vantagem de uma baixa sensibilidade a condições ambientais (como por exemplo nevoeiro, poeira) e providencia informações sobre a superfície como por exemplo à geometria, mas por outro lado para aplicações robóticas apresenta um custo relativamente elevado e apresenta alguns problemas relacionados com a reflexão especular [44].

3.2.5 Câmaras

Um sistema de visão, com recurso a câmaras, é um ponto chave neste tipo de sistemas que requerem uma navegação autónoma [44], [1]. Através do mesmo, podemos reconhecer os obstáculos que se encontram no percurso assim como fazer distinções dos mesmos, como por exemplo o reconhecimento de sinais de trânsito ou luzes de trânsito [53]. Existem vários tipos de câmaras a operar em diferentes comprimentos de onda do espectro eletromagnético, Figura 21. Os sistemas de visão

surgem normalmente associados ao processamento de imagens digitalizadas, obtidas de câmaras que são sensíveis à mesma gama de radiação eletromagnética do olho humano (comprimento de onda aproximado entre 400 nm a 700 nm). Apesar disto, existem outros tipos de câmaras, que operam em diferentes comprimentos de onda do espectro eletromagnético com capacidade de providenciar uma boa perceção do meio.

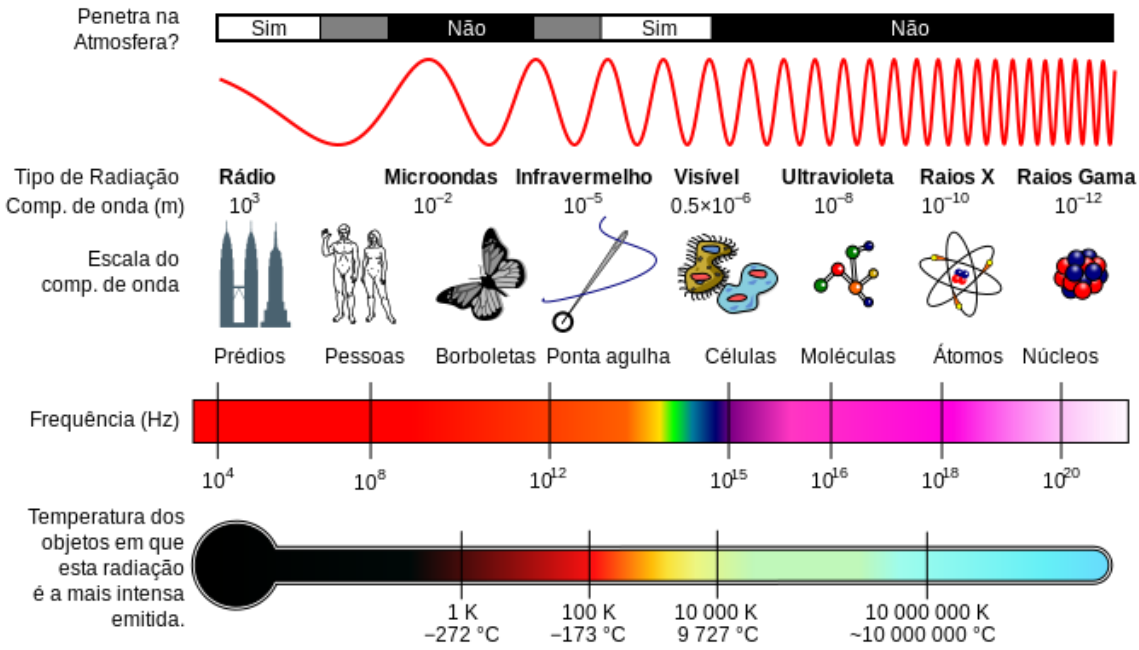

Figura 21 - Comparação das ondas do espectro eletromagnético, [54].

Câmaras Mono e RGB (preto e branco e a cores):

Uma câmara de vídeo é um dispositivo tipicamente composto por uma lente (fixa ou zoom) colocada em frente a um sensor CCD (*Charged coupled device*) ou CMOS (*Complementary metal oxide semicondutor*), que obtém uma matriz retângular de intensidades ou medidas de cor num espaço temporal de amostras fixo. Cada ponto dessa matriz representa um pixel da imagem, em que cada um indica o valor da intensidade de luz que atravessa a lente de um ponto particular da cena. Estas medidas são então convertidas para uma forma que possam ser lidas pelo sistema, de acordo com a taxa de amostragem da câmara, resultando numa gama de valores com uma precisão típica de 8 ou 16 bits.

Os sensores CCD são tipicamente os mais utilizados e os que apresentam melhores resultados, com a desvantagem que requerem um maior consumo de energia, mas apesar de cada um dos sensores

ter as suas próprias vantagens ou desvantagens, ambos têm como função de converter luz em eletricidade.

Nos últimos anos, com o aparecimento das câmaras digitais, a qualidade de imagem tornou se superior assim como a transmissão de informação tonou-se simplificada, facilitando a utilização das mesmas.

Há um conjunto de aplicações robóticas em que é suficiente usar uma imagem a preto e branco, como por exemplo robô a seguir uma linha ou detetar outros robôs, pois essa imagem é suficiente para obter a perceção do meio necessária. Um dos benefícios das câmaras a preto e branco é que requerem 3 vezes menos esforço computacional em relação às cores, pois as câmaras a cores requerem em cada pixel 3 sensores sensíveis a comprimentos de onda para cada uma das cores principais vermelho, verde e azul [53], enquanto que a câmara a preto e branco apenas precisa de 1. Para além disto, as câmaras a cores requerem um controlo adicional para o balanço do branco, pois dependendo da fonte de iluminação no cenário as medidas de cor vermelho, verde e azul podem mudar drasticamente.

Existem varias soluções comerciais de câmaras que podem ser integradas em plataformas robóticas móveis. A utilização de câmaras tem como principal vantagem um baixo custo associado. Por outro lado, a utilização deste tipo de tecnologia requer bastante processamento (dependendo da aplicação pode ser elevado) levando assim a um maior esforço computacional e apresenta uma grande sensibilidade a variação da luminosidade. Quando comparado com outras tecnologias apresentadas anteriormente (LiDAR, Sonar e Radar), tem como vantagem a capacidade de caracterização de objetos, mas por outro lado apresenta um campo de visão limitado.

Câmaras Térmicas:

Uma câmara infravermelha é um dispositivo que deteta energia infravermelha (calor) e converte num sinal eletrónico, que é então processado de modo a obter uma imagem ou vídeo térmico, em que podemos executar cálculos de temperaturas.

Como podemos observar na Figura 21, com imagens térmicas, a porção do espectro magnético que podemos obter perceção é superior à luz visível, pois tudo com a temperatura acima do zero absoluto emite energia infravermelha, mesmo quando se trata de gelo. Desta forma a luz visível não afeta uma imagem térmica. Apesar disto, quanto maior for a temperatura do objeto, maior será a energia infravermelha emitida [55].

Apresenta algumas vantagens, como por exemplo: a imagem obtida não depende das condições de luminosidade, mas sim, da respetiva quantidade de calor que o objeto emite e é mais robusta a

condições ambientais como por exemplo o pó. Por outro lado, apresenta um custo elevado e numa pequena análise aos fabricantes deste tipo de tecnologia permite concluir que oferecem uma menor resolução em relação às restantes. Exemplos de aplicações das câmaras infravermelhas são reconhecimento de destino, vigilância, inspeção industrial e manutenção preditiva. Um exemplo de uma imagem térmica pode ser visualizado na Figura 22.

Figura 22- Exemplo de imagem capturada por uma câmara térmica, [56].

Câmaras 3D:

Câmaras que permitam a perceção em profundidade são um dos pontos fulcrais em sistemas de visão e as aplicações vão de navegação robótica, a reconstrução de objetos e a interação entre humanos e robôs [57]. Nos últimos anos foram desenvolvidas varias câmaras com esta capacidade de perceção, e exemplo disso é a Kinect que é um dos dispositivos mais usados em aplicações robóticas devido ao sua excelente entre preço e qualidade de perceção do ambiente em redor. É uma câmara 3D desenvolvida pela Microsoft para a consola XBOX 360 que permite ao jogador usar o seu corpo como controlador. O sensor de profundidade é do tipo de luz estruturada.

Relativamente a sistemas 3D baseados em visão por computador, as principais tecnologias utilizadas para obter uma perceção em profundidade do meio são:

Visão *estereo* - Este tipo de sistema providência uma perceção de profundidade, com recurso a pares de câmaras que obtêm duas ou mais imagens de pontos de vista diferentes simulando assim a visão binocular do ser humano, permitindo desta forma estimar a distância a um ponto

no cenário [58], [45]. O conceito base utilizado é a triangulação [59], pois um ponto no cenário e as duas câmaras formam um triângulo, e sabendo a distância de uma câmara a outra conseguimos obter a linha base entre as duas (Figura 23). Sabendo isto e o ângulo que ambas as câmaras formam conseguimos obter a distância a um ponto de um objeto no cenário. Na prática, existe uma dificuldade enorme em encontrar uma correspondência entre as imagens obtidas, pois quando um cenário é visualizado por duas ou mais câmaras, os pontos do cenário serão mapeados em diferentes pontos de imagem em cada uma das câmaras, pois em cada uma delas, cada ponto terá diferentes posições (vertical e horizontal). A técnica mais comum é considerar todas as possíveis correspondências de pontos entre as imagens das diferentes câmaras e descartar todos aqueles que não cumpre os critérios de correspondência estabelecidos. O problema é a possível existência de uma quantidade enorme de possíveis correspondências, e uma dificuldade da definição dos critérios que são estáveis e não descartam possíveis boas correspondências.

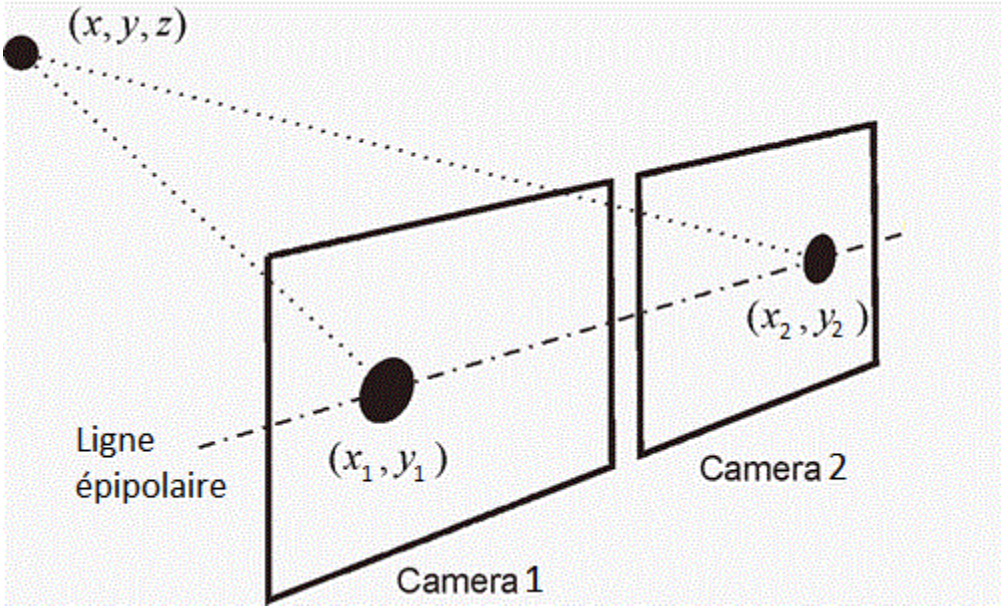

Figura 23- Modelo de um sistema Visão Stereo, [60].

Segundo [44], existe um vasto número de algoritmos na literatura que fazem essa correspondência e processam as varias imagens, em que o resultado é apresentado como um mapa de disparidades.

Apesar desta técnica de perceção ser atrativa, pois teoricamente o conceito é simples e fornece uma perceção em profundidade, esta apresenta várias desvantagens, [44] :

corresponder a imagem de um objeto com outra imagem do mesmo objeto proveniente de outra câmara pode se tornar complicado, apesar das câmaras poderem estar paralelas ou a convergir, idealmente estas devem visualizar a mesma região de interesse, apresentando desta forma uma calibração difícil; sensíveis a luminosidade; a precisão do processo de triangulação é limitada pelo tamanho do pixel e pela geometria da câmara. Se as câmaras forem colocadas muito próximas uma da outra a precisão será baixa, mas o processo de correspondência de imagens será simplificado pois o objeto terá uma visualização muito similar na imagem das duas câmaras. Quanto mais as câmaras estiverem afastadas, a precisão será superior (quanto mais afastadas estiverem as câmaras maior será o efeito 3D), mas o processo de correspondência tornar-se-á mais árduo.

Para facilitar a correspondência de imagens de diferentes câmaras, alguns sistemas como é o caso da BIRIS [44], recorrem a projeção de um laser no ambiente para que a correspondência entre as duas imagens tenha só a iluminação laser como ponto alvo. Desta forma a correspondência é simplificada ao máximo pois só existe um ponto de referência para a junção das imagens.

Luz Estruturada – Esta tecnologia é uma das mais utilizadas para obter uma perceção em profundidade do meio e é tipicamente composta por um projetor e uma câmara. Consiste na projeção de um padrão conhecido numa superfície desconhecida através de um projetor especial (tratam-se de sensores ativos) de luz estruturada e na captação da imagem dessa mesma superfície que o padrão incide com recurso a uma câmara, Figura 24.

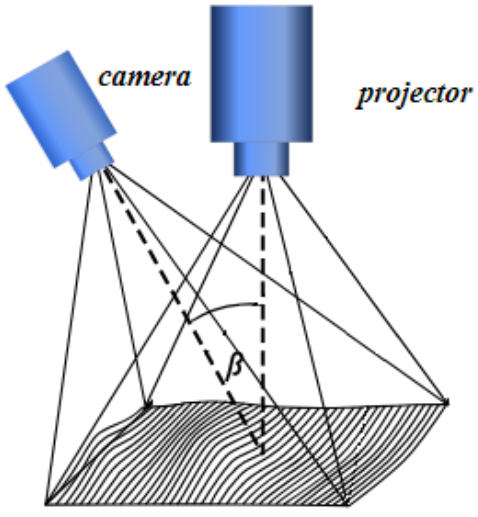

Figura 24- Principio funcionamento da Luz Estruturada, imagem adaptada de [61].

Se a superfície onde o padrão é projetado for planar, sem qualquer variação 3d da superfície, o padrão de luz capturado pela câmara será aproximadamente igual ao emitido pelo projetor. No entanto, se a superfície for não planar, a forma geométrica da superfície irá distorcer o padrão de luz estruturada vista pela câmara. Desta forma, a perceção em profundidade do meio é extraída baseada na informação de distorção do padrão de luz estruturada projetado na superfície.

Segundo [62] , a relação entre a câmara, o projetor de luz estruturada e o ponto de um objeto na superfície pode ser expressa segundo o principio da triangulação da seguinte forma:

$$R = B\frac{\sin(\theta)}{\sin(\alpha + \theta)} \tag{5}$$

Onde:

θ é o ângulo do projetor entre o ponto do objeto e a câmara.

α é o angulo da câmara entre o ponto do objeto e o projetor.

B é a distância entre o projetor e a câmara.

R é a distância da câmara ao ponto do objeto.

Apesar do sistema apresentar bons resultados na perceção de profundidade, esta tecnologia apresenta algumas limitações, tais como, a calibração do sistema, a precisão desta depende do sistema ótico e da resolução da câmara e a dificuldade de fornecer uma imagem em tempo real devido aos varrimentos sucessivos que tem que ser efetuados ao cenário. Contudo, segundo [36], em que a solução utilizada é um Kinect v1, esta apresenta bons resultados, providenciando uma imagem em tempo real e sem a necessidade de qualquer tipo de calibração. Outros exemplo de câmaras que incorporam este tipo de sensor é a Asus Xtion Pro Live e a ENSENSO.

TOF- É emitido um sinal de luz e mediante o tempo de voo (tempo que o sinal depois de emitido demora a ser refletido e a chegar a câmara) é calculada a distância para cada ponto da imagem. Ao contrário do LiDAR, os pulsos laser emitidos tem de cobrir toda a porção do cenário que está a ser observada, sendo tipicamente constituídos por uma matriz de projetores de luz laser [59], [63]. Isto permite o cálculo do valor de distâncias por cada pixel na imagem, pois mediante a distância, a intensidade do mesmo varia. Na Figura 25, é apresentado um esquema de uma

câmara TOF, que pode ser dividido em vários blocos: (a) uma fonte de luz modulada ou pulsada, normalmente baseada em laser ou Led na parte infravermelha do espectro magnético para que a iluminação seja discreta, (b) um difusor ótico para espalhar a luz emitida pelo cenário a visualizar, (c) um conjunto de lentes para aglomerar a luz refletida de volta pelo alvo. Utiliza-se também um filtro ótico passa banda, configurado para o comprimento de onda da fonte de luz, permitindo a rejeição de algum ruído presente no ambiente e (d) composto por uma matriz de pixéis capazes de medir, de forma direta ou indireta, o tempo de voo (TOF) necessário pelo pulso de luz da fonte de luz ao alvo e de volta ao sensor. Para além disto, o sistema requer uma interface de controlo que além de ser responsável pela comunicação com o mundo externo, é também responsável por fornecer a alimentação correta ao sensor, emitir os sinais de controlo e fazer a leitura dos dados obtidos pelo sensor.

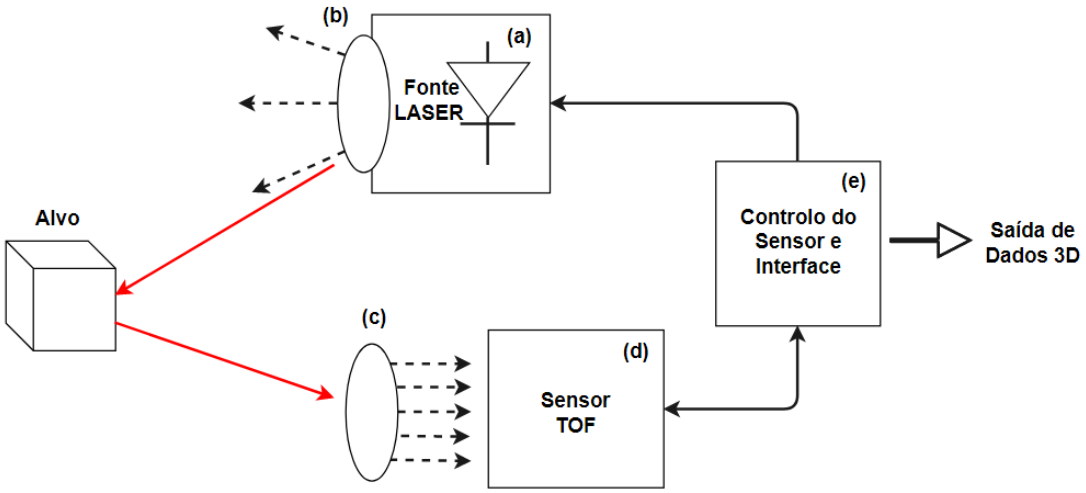

Figura 25- Modelo câmara TOF.

As câmaras TOF mais comuns dispõem de uma resolução por pixel de 8 bits (0 a 255). Esta tecnologia é considerada ativa e apresenta resoluções inferiores quando comparada as câmaras atuais (ex. 60x24, 128x128, 174x144). A taxa de vídeo varia entre 30 a 50 fps, o que fornece informação suficiente para algumas aplicações robóticas. Estes sistemas, quando comparados a sistemas que possuem um sensor de luz estruturada, conseguem facilmente providenciar imagens em tempo real e apresentam alcances tipicamente superior a 5 m. Quando comparados a sistema de *visão estéreo*, estas providenciam estimativas de profundidade em tempo real em condições que as câmaras estéreo não funcionam muito bem, como por exemplo

na presença de paredes brancas. Em contrapartida, estes sensores por vezes não apresentam bons resultados em cenários texturizados no qual a visão stereo apresenta bons resultados, [6].

3.3 Sistemas de visão

Um dos pontos fulcrais em sistemas de deteção de humanos são os sistemas de visão. Tipicamente são constituídos por câmaras para a aquisição de imagens (apresentadas em 3.2.5) e um sistema que recorre a várias técnicas ou algoritmos para fazer a classificação e o reconhecimento de objetos na imagem. Tipicamente essas técnicas passam pelo pré-processamento da imagem, segmentação, deteção e reconhecimento de características.

3.3.1 Pré-Processamento

Depois de adquiridas as imagens, o pré-processamento da imagem é uma das primeiras etapas em muitos sistemas de visão por computador. Durante esta fase recorre se a técnicas de correção de imagem para favorecer certas características de interesse, permitindo assim uma melhor análise da mesma nas próximas etapas.

Um dos operadores[5] mais simples que é utilizado nesta fase é o *Point Operator*, que manipula cada pixel independentemente da sua vizinhança, dependendo só do seu valor. Este é maioritariamente utilizado para ajustes de brilho e contraste assim como correção de cores e transformações.

Para além deste, existem os operadores denominados *neighborhood operator*, em que o valor de um novo pixel depende de um pequeno numero de valores de pixéis na sua vizinhança. Exemplo deste operador pode ser o *Histogram equalization*, utilizado em vários sistemas para realçar características importantes através da redução de ruído, modificação do contraste ou melhoramento da qualidade de imagem.

Em imagens adquiridas por câmaras de profundidade, em que as imagens podem conter alguns pixéis em falta devido ao alcance limitado da câmara, tipicamente aplica-se o algoritmo *nearest neighbor interpolation*. Este converte a informação da câmara discreta em contínua e recupera o valor dos pixéis em falta, mas por outro lado pode introduzir algum ruído na imagem, sendo por vezes necessário a aplicação de outros filtros. A diferença deste algoritmo para os restantes algoritmos de

[5] Um operador de processamento de imagem é uma função que recebe uma ou mais imagens de entrada e produz uma imagem de saída.

interpolação é que este seleciona o pixel que pretende aproximar ao valor real e escolhe o ponto mais próximo desse pixel como valor de comparação, ao contrário de outros que selecionam toda a vizinhança.

3.3.2 Segmentação

A segmentação de imagens é uma técnica que permite dividir uma imagem em várias regiões, de forma que os pixéis de cada uma dessas regiões possua características específicas, como por exemplo, níveis de amarelo com valor semelhante. Como estes apresentam as mesmas propriedades, no final acabam por possuir uma enorme correlação com objetos nessa imagem, tornando esta técnica uma das mais utilizadas para a separação de objetos de uma dada imagem. São várias as técnicas que podemos utilizar, e cada uma delas apresenta desempenhos diferentes mediante a aplicação. Em [64] são apresentadas algumas das técnicas desenvolvidas para a segmentação de imagens:

Region based on active contours: a região de cada grupo é definida baseada nos limites dos objetos presentes imagem. É feito um contorno inicial inserido dentro do objeto, em que este se deforma e se expande de modo a atingir a estrutura total do objeto.

Regions spliting and merging: esta técnica divide a imagem em várias regiões. A representação básica é em pirâmide. Inicialmente, o algoritmo considera a imagem como uma só e depois aplica um critério de homogeneidade para verificar se este resulta positivo. Se resultar negativo volta a dividir essa região em quatro pequenas regiões, repetindo continuamente todo o processo em todas as sub-regiões até não ser necessária mais nenhuma divisão.

Mean Shift: é uma técnica que suaviza a distribuição de valores e encontra os seus picos, como as regiões de características no espaço a que corresponde cada pico. Como os picos de valores são sabidos, a cada iteração cada um dos restantes pontos desloca-se para o pico mais próximo.

Normalized cuts: esta técnica efetua grupos de pixéis através das afinidades ou similaridade entre pixéis próximos, e tenta separar grupos que estão ligados por uma afinidade baixa.

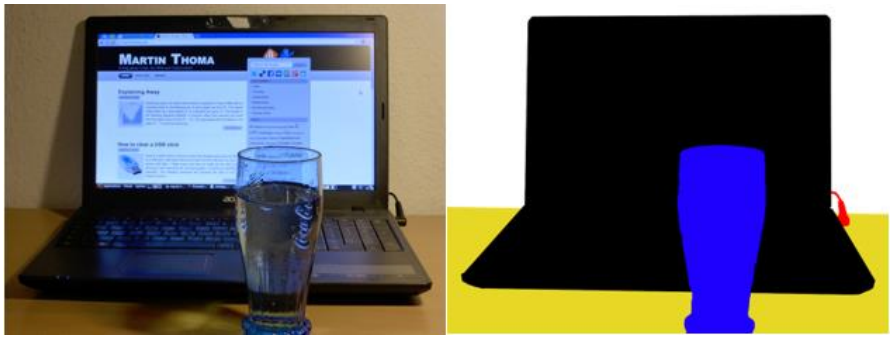

Figura 26- Exemplo segmentação de uma imagem, [65] . À esquerda imagem original, à direita a imagem segmentada.

3.3.3 Deteção de características e Reconhecimento

As tarefas de reconhecimento e deteção de objetos em imagem são umas das tarefas mais complicadas na análise de imagens, requerendo algoritmos por vezes complexos. A deteção consiste em estabelecer critérios que possam ser utilizados para identificar ou distinguir um determinado grupo de objetos que possam estar presentes numa imagem, descartando os restantes grupos. O reconhecimento é um processo utilizado para encontrar um objeto específico ou uma estrutura de um determinado grupo numa imagem.

A deteção de um grupo de objetos numa imagem permite ainda que o algoritmo de reconhecimento processe somente as regiões candidatas, acelerando assim todo o processo de reconhecimento e reduzindo os falsos positivos. O reconhecimento funciona em vários níveis, incluindo diferentes processos como encontrar um rosto numa imagem ou fazer corresponder esse rosto com um indivíduo específico.

Nesta fase o foco será a deteção de humanos, apresentando assim métodos para deteção de rostos e métodos para deteção de pessoas.

Deteção de rostos: umas das primeiras etapas nesta fase é encontrar os tamanhos e as localizações de rostos na imagem. Apesar de podermos aplicar um algoritmo de reconhecimento de rostos a cada pixel e escala, este processo seria bastante lento na prática. Segundo [64], as técnicas de deteção de rostos podem ser classificadas em 3 tipos: *template-based*, *feature-based* ou *appearance-based*.

Feature-based é uma técnica que procura encontrar a localização de características distintivas na imagem como olhos, nariz, boca e verificar se essas características são geometricamente plausíveis para se considerar um rosto.

Template-based é uma técnica que recorre a um modelo estático de um rosto real, onde a textura e a forma são capturadas durante uma fase treino. A deteção de um rosto é efetuada quando um objeto na imagem corresponde a esse modelo.

Appearance-based é uma técnica que procura por prováveis rostos candidatos, em pequenos fragmentos retangulares sobrepostos da imagem, que são mais tarde verificados por classificadores que requerem aprendizagem. Estes dispõem de conjuntos de imagens de rostos positivos e de falsos positivos. De modo a LiDAR com diferentes escalas geralmente a imagem é convertida numa pirâmide sub-octava, onde é realizada uma procura em cada uma das camadas.

Um dos algoritmos mais conhecido e utilizado do tipo *appearence-based* é o Viola and Jones [64]. Embora possa ser utilizado para detetar uma variedade de objetos, inicialmente foi projetado para a

deteção de rostos. Este algoritmo utiliza um conjunto de características visuais, com formatos específicos, como podemos visualizar na Figura 27, que podem ter dimensões e posições arbitrarias dentro de uma determinada imagem. O valor de cada uma das características sobre uma imagem é calculada pela subtração do somatório de todos os pixéis na região preta, pelo somatório de todos os pixéis na região branca.

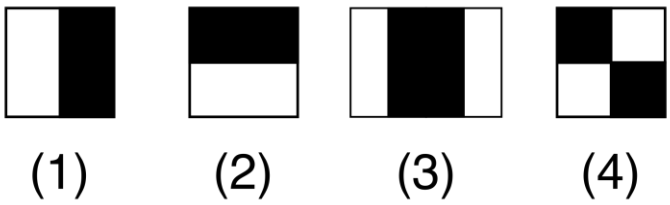

Figura 27- Características visuais Viola and Jonas, [30].

O sistema de deteção recorre a vários classificadores em cascata. Cada classificador é relativo a uma determinada característica e para a deteção de um rosto numa sub-região de uma imagem, este deve percorrer todas as camadas da cascata de classificadores. A cada nível que passos os classificadores tornam-se mais precisos. Para além disto a deteção de objetos proposta por Viola and Jonas [64] utiliza uma variante do algoritmo de aprendizagem AdaBoost [28], para a seleção das melhores características e para o treino de todos os classificadores.

Deteção de pessoas: um dos exemplos de algoritmos de deteção de pessoas mais conhecido é o descritor HOG (*Histogram of Oriented Gradients*), com recurso a um SVM (*Support vector machine*) para classificar o descritor recebido como sendo ou não sendo uma pessoa.

O HOG é um descritor de características, que tem por objetivo generalizar um objeto de tal forma, que quando esse mesmo objeto é visto sobre condições diferentes, apresente as mesmas características, tornando a classificação mais simples.

Segundo [39] e [40], a deteção pode ser dividida em quatro etapas. Cálculo do gradiente de cada pixel, ou seja, da orientação e magnitude das arestas da imagem; divisão da imagem em blocos e células em que os pixéis são agrupados em células e as células agrupadas em blocos; cálculo do histograma de gradientes em cada célula, sendo depois agrupado em blocos; junção destes histogramas, originando o descritor.

A primeira etapa, cálculo da orientação e magnitude das arestas, é realizada com recurso a métodos de deteção de arestas. Segundo [39], uma das técnicas que apresenta bons resultados é a *Sobel* [66] a uma dimensão.

A segunda etapa, divisão de imagem em blocos e células, é realizada uma divisão da imagem original em várias janelas denominadas células, Figura 28. Já os blocos são conjuntos de células, adjacentes a outros blocos, em que existe a sobreposição de uma célula em qualquer uma das direções, para os blocos circundantes. Estas divisões para além de permitirem a observação direta da magnitude, permitem ainda haver uma estatística associada aos descritores, maximizando assim o ganho de informação. Exemplo do agrupamento de pixéis em células e das células em blocos pode ser visualizada na Figura 28.

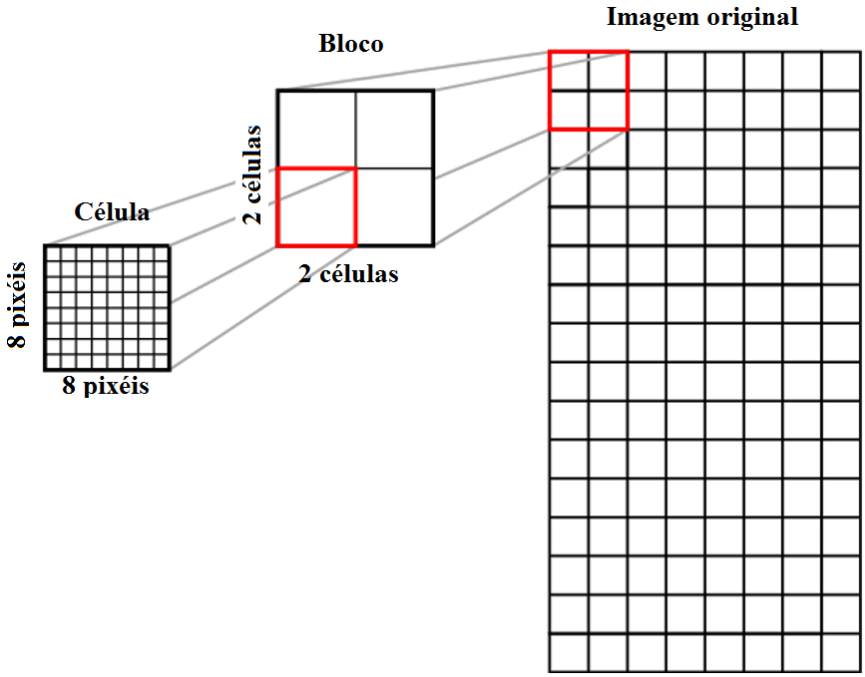

Figura 28 - Algoritmo HOG – Agrupamento dos pixéis em células e das células em blocos.

A terceira etapa consiste no agrupamento das direções das arestas num histograma. Antes do agrupamento recorre-se a uma máscara Gaussiana para efetuar uma ponderação das magnitudes em cada pixel de cada bloco, dando uma maior importância aos valores centrais. O agrupamento é realizado por células dentro de um determinado bloco, sendo depois esse processo repetido para todos os restantes blocos.

A última etapa corresponde ao cálculo do histograma da orientação dos gradientes, por bloco, em cada uma das células separadamente. Cada pixel possui um peso na orientação do histograma da sua célula.

Para o reconhecimento de objetos é ainda necessário um sistema de reconhecimento baseado em aprendizagem supervisionada, como por exemplo o SVM (Support Vector Machines), que recebe o descritor HOG e que classifica como sendo uma pessoa ou não pessoa.

Apesar de este método anterior apresentar bons resultados, há ainda outros algoritmos (*templated-based*) que recorrem a um modelo normalizado de uma pessoa, para efetuar a deteção de pessoas numa imagem. Esse modelo é deslizado sobre a imagem de modo a procurar semelhanças entre o modelo da pessoa e a imagem real. Caso a imagem real possua pessoas com diferentes alturas ou distâncias do modelo normalizado, torna-se necessário a realização de redimensionamentos na imagem para compensar as diferentes escalas.

4. IMPLEMENTAÇÃO

Neste capítulo são apresentadas as tecnologias de sensorização a implementar no sistema de perceção ambiente proposto, assim como todas as configurações necessárias do sistema e os métodos de deteção utilizados.

4.1 Sensores

O sistema de perceção do ambiente irá ter como recurso diferentes tecnologias sensoriais, pois, apesar das inúmeras vantagens que cada sensor dispõe, por si só não é suficiente para garantir segurança, robustez e um bom reconhecimento do meio. Cada uma das tecnologias escolhidas têm capacidades distintas, que em conjunto, permitem uma reconstrução precisa do ambiente em redor do veículo.

Após um estudo e análise das tecnologias sensoriais disponíveis e das soluções até agora implementadas, o veículo autónomo será dotado de sensores do tipo LiDAR (*Light Detection And Ranging*), Sonar (*Sound Navigation and ranging*) e câmaras para visão por computador. Os sistemas de visão irão ser constituídos por câmaras a cores com elevada resolução e uma câmara de profundidade.

Os sensores a implementar na solução final são caracterizados por serem Exteroceptivos [44], pois além de serem os mais utilizados para a deteção de obstáculos, são sensores que permitem a aquisição de informação sobre a área de atuação do veículo. Os sensores exteroceptivos podem ainda ser ativos, no caso de requererem o envio de energia para o ambiente, ou passivos, em que utilizam a energia já presente no ambiente [45].

4.1.1 LiDAR

O sensor LiDAR escolhido para uma deteção de obstáculos robusta é o SICK S300. Este é um dos modelos mais utilizados em ambientes industriais, pois apresenta uma boa precisão e um alcance de medição de cerca de 30m. Exemplo de aplicações que integram este modelo podem ser o MIR 200 [17], KARIS PRO [16] e o sistema DRIVEN da BALYO [19], apresentados no capítulo 2.3. Este modelo segundo as normas do fabricante só pode ser utilizado em ambientes interiores. Para além das medições fornecidas entre os seus 270° (-45° a 225°), possui ainda um modo de aviso e um de proteção, quando

se depara com obstáculos a 10 m e a 3 m (estes valores podem ser configurados). Alguns dados técnicos deste modelo são apresentados na Tabela 1.

Figura 29- Sensor LiDAR SICK S300, [67].

Tabela 1 - Dados técnicos do sensor LiDAR SICK S300

Ângulo máximo de leitura	270°
Distância máxima de medição	30 m
Resolução ângular	0.5°
Modo de Proteção	3 m
Modo de Aviso	8 m
Tempo de Resposta	80 ms
Comprimento de onda	905 nm
Class Laser	1 (21 CFR 1040.10 e 1040.11, IEC 60825-1)
Tensão Alimentação Vs	(16.8 a 30 V) DC

Para uma perceção total em redor do veículo são utilizados 2 sensores LiDAR, colocados na dianteira do veículo como se pode visualizar na Figura 30. Contudo, um dos sensores será montado

inversamente, rodado 180° em relação ao eixo dos *x*, de forma a que os pulsos emitidos por ambos os sensores não interfiram entre si. Desta forma, um dos sensores obterá uma perceção a uma altura de 20 cm do chão, enquanto o outro sensor obterá uma perceção a uma altura de 17 cm. Para além de não haver interferências entre ambos, esta configuração apresenta ainda o benefício de a estrutura do veículo não afetar a área de medição do LiDAR, mantendo-o sempre relativamente perto do chão. A sua principal função será a deteção de obstáculos e o possível reconhecimento de pessoas através de um padrão de pernas.

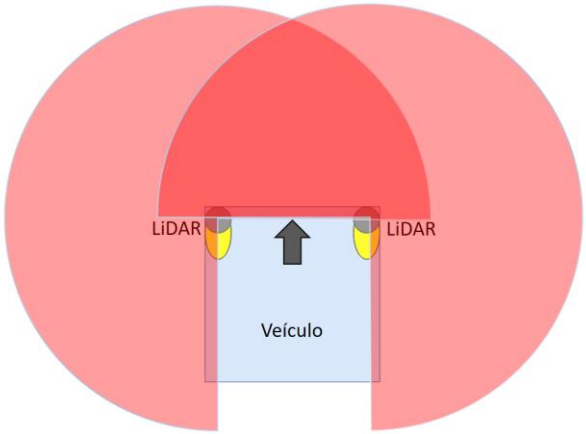

Figura 30- Configuração dos sensores LiDAR no veículo.

Este dispositivo utiliza uma ligação RS422 para a comunicação com o *middleware* ROS, fornecendo este algum suporte e os drivers necessários para a receção de todos os dados do sensor. O formato das mensagens, assim como os tópicos publicados e os parâmetros que pode sem ajustados são os seguintes:

Tópicos publicados:

-/laserscan(sensor_msgs/LaserScan): Dados das medições do sensor.

Parâmetros:

-frame (String, default: "base_laser_link"): frame do sensor.

-send_transform (int, default: 1): Envia a transformada de base_link to base_laser_link.

-tf_x (double, default: 0.115): Transformada ao longo do eixo dos x.

-tf_y (double, default: 0.0): Transformada ao longo do eixo dos y.

-tf_z (double, default: 0.21): Transformada ao longo do eixo dos z.

-reduced_fov (int, default: 0): Permite Reduzir o ângulo de leitura.

-devicename (String, default: "/dev/SICK300") Define a porta conectada ao sensor.

-baudrate (int, default: 500000): Define o baudrate.

4.1.2 Câmara 3D

Câmaras que permitam a perceção em profundidade são um dos pontos fulcrais em sistemas de visão. No nosso sistema a câmara terá como função a caracterização de alguns obstáculos, como a deteção de pessoas em redor do veículo e a deteção de possíveis obstáculos. Um dos principais requisitos na escolha da câmara a utilizar é o alcance em que esta consegue operar, pois uma deteção obstáculos antecipada permite um planeamento de trajetórias muito mais eficaz, por parte do sistema de navegação. Tendo isto em consideração o sensor da câmara a utilizar deve ser do tipo TOF, pois são os que apresentam melhores resultados a distâncias superiores a 4 m e não requerem grandes calibrações nem processamentos.

A câmara 3D escolhida foi a Visionary-T da SICK (Figura 31), pois apresenta um alcance de 7 m, uma informação de profundidade por cada pixel em tempo-real e robustez suficiente para ser aplicada em veículos industriais. A informação de profundidade de uma cena é fornecida em uma matriz de pixéis a duas dimensões, denominada imagem ou mapa de profundidade. Cada valor de cada pixel do mapa de profundidade representa a distância (neste caso em mm) de um ponto do cenário à câmara. Por questões de visualização e compressão de imagens este valor possui um tamanho de 8 bits (0 a 255). Se um pixel capturar um ponto do cenário que se encontra fora do alcance da câmara é preenchido com o valor 0.

Figura 31 - Câmara 3D - SICK Visionary-T.

Tabela 2 - Características da câmara 3D SICK Visionary-T.

Distância de operação	0,5 m a 7.2 m
Ângulo de deteção	69° x 56°
Resolução de imagem	176 x 144
Taxa de captura de imagens	0.03 s por imagem 3D (30 fps)
Precisão	(Distância de 0,5m e 7,0 m) ±10 mm e 40 mm
Emissor de luz	Luz infravermelha invisível (LED, 850 nm)
Classe do LED	Grupo de risco 0 em acordo com EN 62471
Calibração de fábrica	Sim
Tensão Alimentação	24V DC

A câmara estará localizada no veículo a uma altura de 1.6 m, sobre um *pan-til* (capacidade rotacional sobre dois eixos), podendo desta forma visualizar diferentes cenários do ambiente em redor.

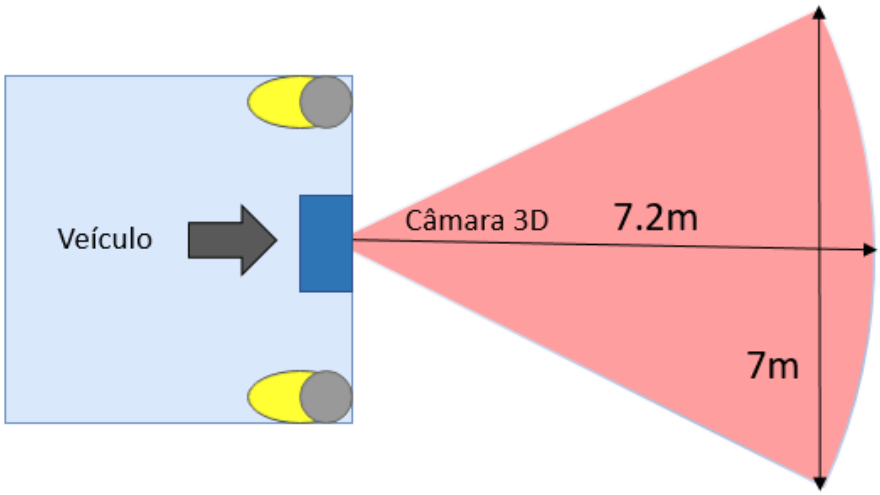

Figura 32 - Configuração da câmara 3D no veículo.

Todos os parâmetros iniciais da câmara podem ser ajustados, com recurso ao software SOPAS ET da SICK, como por exemplo o endereço IP. A transmissão de dados é realizada por ethernet e o ROS dispõem de um pacote de drivers *open source* necessários para a comunicação com o mesmo. O formato das mensagens, assim como os tópicos publicados, serviços ou parâmetros são os seguintes:

Tópicos publicados:

-camera/camera_info (sensor_msgs/CameraInfo): Calibração da câmara assim como os metadados;

-camera/depth (sensor_msgs/Image): Imagem de profundidade da câmara. Esta imagem encontra-se no formato uint16 em mm;

-camera/confidence (sensor_msgs/Image): Imagem de "confiança" da câmara. Contêm valores de confiança uint16;

-camera/intensity (sensor_msgs/Image): Imagem de intensidade da câmara. Contêm valores de intensidade uint16;

-camera/ios (std_msgs/ByteMultiArray): Estado da câmara IO;

-camera/diagnostics (diagnostic_msgs/DiagnosticStatus): Estado da câmara, incluindo as variáveis da câmara;

-camera/scan (sensor_msgs/LaserScan): Medições reduzidas a um *laser scan*;

-camera/cartesian (sensor_msgs/PointCloud2): Nuvem de pontos 2D;

Serviços:

-enable_depth_map (std_srvs/Trigger): Habilita a publicação de nuvens de pontos e desabilita a publicação do mapa de alturas e o varrimentos polares;

-enable_height_map (std_srvs/Trigger): Habilita a publicação do mapa de alturas e desabilita a publicação do nuvens de pontos e o varrimentos polares;

-enable_polar_scan (std_srvs/Trigger): Habilita a publicação de varrimentos polares e desabilita a publicação do mapa de alturas e o nuvens de pontos;

Parâmetros:

–remote_device_ip (string, default: 192.168.1.10): Especifica qual o endereço IP da câmara a aceder;

–frame_id (string, default: camera): Especifica o nome do frame;

–prevent_frame_skipping (bool, default: false): No caso de o computador ser muito lento e em caso de atraso, especifica se avança para a próxima imagem ou processa todas as imagens com atraso;

–channel (string, default: NO_CHANGE): Especifica o canal inicial a publicar;

-user (string, default: SERVICE): Especifica o nível utilizado para a escrita de parâmetros, exemplo: RUN, AUTHORIZEDCLIENT, MAINTENANCE, SERVICE;

4.1.3 Sonar

Para uma redundância por parte do sistema de navegação, o veículo irá ser equipado com 3 sensores sonar UM18-2 Pro da SICK. As suas funções serão deteção de obstáculos próximos do veículo que possam interferir com a trajetória do veículo e causar danos. Estes sensores atuam com precisão para distâncias entre 6.5 cm a 35 cm, mas possuem um alcance máximo de 60 cm.

Tabela 3 - Dados técnicos sensor Sonar UM18-2pro

Distância de operação	0.65 cm a 35cm
Distância limite	60cm
Precisão	±1%
Compensação de temperatura	Sim
Tempo de Resposta	64 ms
Frequência típica	400KHz
Resolução Analógica	12 bit
Tensão Alimentação	(10V a 30V) DC

Os sensores irão ser colocados na frente e nas laterais de veículo, como podemos visualizar na Figura 33, visto que as leituras destes sensores na retaguarda do veículo iriam ser influenciadas pelos garfos ou pelo transporte de materiais.

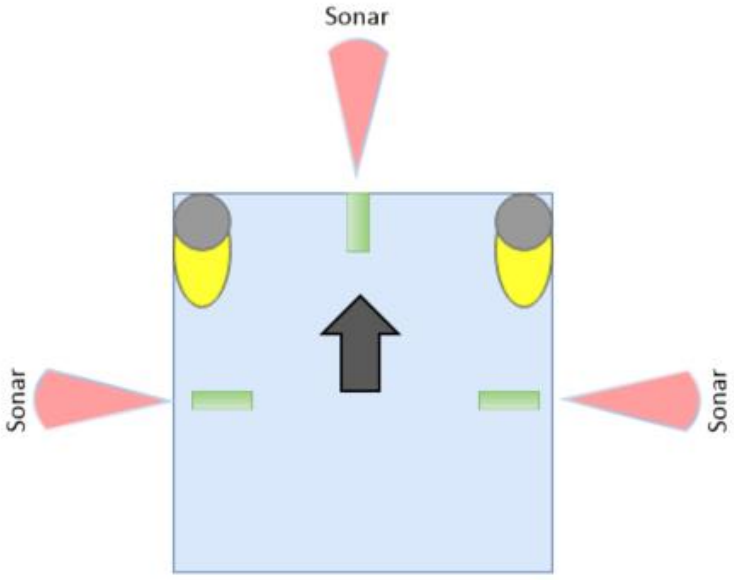

Figura 33 - Configuração dos sensores no veículo.

Como estes sensores providenciam apenas alguma redundância ao sistema na deteção de obstáculos, neste documento não será descrita a implementação desta mesma tecnologia, pois não providencia informação suficiente para uma deteção e caracterização de obstáculos.

4.1.4 Câmara RGB

Para além de uma câmara 3D, o veículo será equipado com duas câmaras RGB uEye 3013XC. Esta câmara apresenta uma resolução de imagem de 4216x3160 em modo de imagem e uma resolução de 1920x1080 em modo de vídeo.

Tabela 4 - Dados técnicos câmara uEye 3013XC.

Sensor	Sony CMOS 13 Megapixéis
Resolução máxima	4216x3160 pixéis
Vídeo	1920x1080 pixéis @30fps
Compensação de temperatura	Sim
Tempo de Resposta	64 ms
Frequência típica	400KHz
Resolução Analógica	12 bit
Tensão Alimentação	(10V a 30V) DC

As câmaras serão posicionadas na frente do veículo. A câmara da direita encontrar-se-à orientada a 90° e a da esquerda a 110° como se pode visualizar na Figura 34.

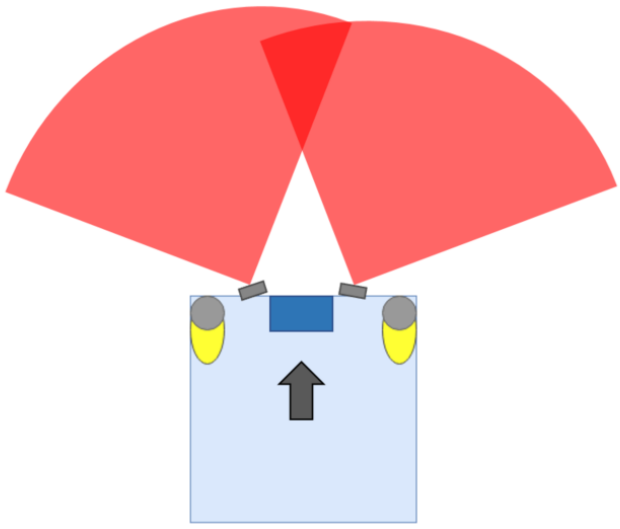

Figura 34 - Posicionamento das duas câmaras uEye 3013XC no veículo.

Neste documento não será descrita a implementação desta mesma tecnologia, pois a abordagem utilizada é a deteção de pessoas com recurso a câmara 3D "SICK Visionary-T e ao LiDAR 2D "SICK S300".

4.2 Configuração do *middleware* ROS

Neste subcapítulo é apresentado todas as configurações efetuadas durante a instalação da ferramenta ROS, assim como a configuração de todas as bibliotecas ROS necessárias para a implementação da abordagem proposta.

4.2.1 Inicialização do espaço de trabalho

A configuração do espaço de trabalho inicial e a criação de novos *packages ROS* é realizada seguindo todas as etapas recomendadas pelo ROS. Numa primeira etapa é criado o espaço de trabalho, onde mais tarde são instaladas todas as dependências necessárias e é realizada a compilação de todos os *packages* ROS. Uma das dependências essenciais instaladas é o cv_bridge que permite a conversão de mensagens de imagem ROS em imagens OpenCV, permitindo a utilização da biblioteca OpenCV para o processamento de imagens. Ainda nesta fase é necessário a instalação de todos os packages de comunicação entre os sensores e o ROS.

4.2.2 Sistema de Transformadas

Como referido no subcapítulo 3.1.1, um sistema robótico tipicamente é constituído por várias tecnologias de sensorização ou outros tipos de mecanismos de atuação. Como todos partilham o mesmo espaço e operam em conjunto há a necessidade de representar a posição e orientação de cada um deles num mapa de coordenadas relativo ao sistema, pois só assim é sabida a área de atuação de cada um. O sistema proposto de deteção de pessoas recorrerá a dois sensores LiDAR 2D e a uma câmara 3D. Para além dos sensores, é ainda necessário representar a posição central do sistema "base_footprint" e a posição do sistema baseada em odometria "odom".

Para a construção deste mapa de coordenadas internas do sistema, recorreu-se a biblioteca "tf" do ROS. Todas as relações entre transformadas seguem uma estrutura em árvore, definindo uma raíz e desenvolvendo-se até ao nó desejado. Para esta dissertação foi considerado o modelo da Figura 35, para o desenvolvimento do sistema de transformadas.

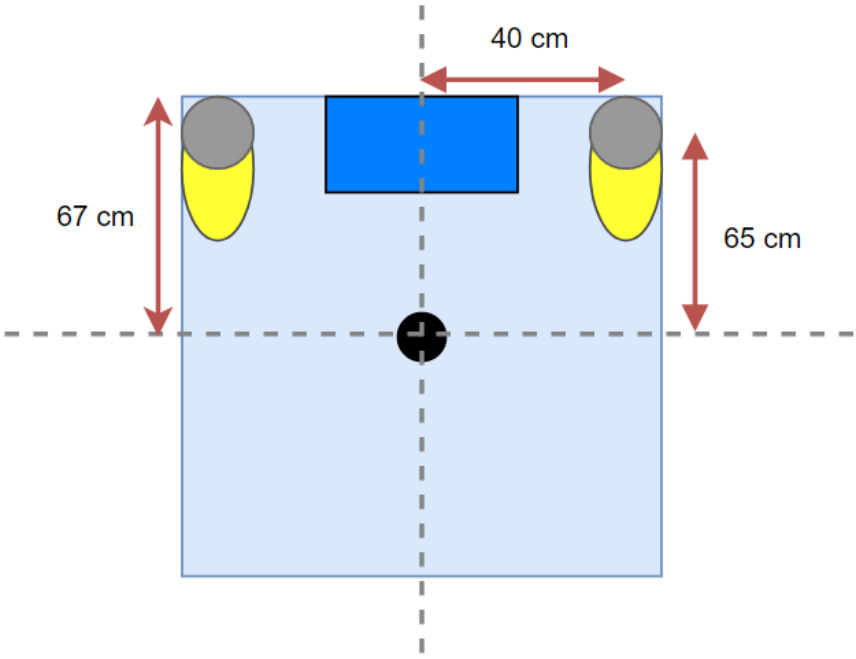

Figura 35 - Modelo de veículo considerado para as transformadas.

Na Figura 36 podemos visualizar toda a estrutura em árvore de todas as transformadas do sistema proposto para a deteção de pessoas. A raiz denominada "odom" é a posição estimada do sistema baseada em odometria, derivando o ponto central do sistema "base_footprint" no plano do chão. Deste último ponto derivam 3 transformadas: "base_laser_left_link" que representa a posição central do sensor LiDAR colocado á esquerda no veículo, onde ocorre uma translação em *x* de 0,65 m, *y* de -0,4 m e em *z* de 0,17 m; "base_laser_right_link" que representa a posição central do sensor LiDAR colocado á direita no veículo onde ocorre uma translação em *x* de 0,65 m, *y* de 0,4 m e em *z* de 0,2 m; e por fim "rgbd_front_top_link" onde foi considerado como a posição central do suporte da câmara, onde ocorre uma translação em *x* de 0,65 m e em *z* de 1,55 m. Como o sensor da câmara encontra-se deslocado da posição central do suporte da câmara foi ainda necessário derivar mais uma transformada "rgbd_front_top_depth_optical_frame", onde ocorre uma translação em *x* de 0,02 m e em *z* de 0,05 m.

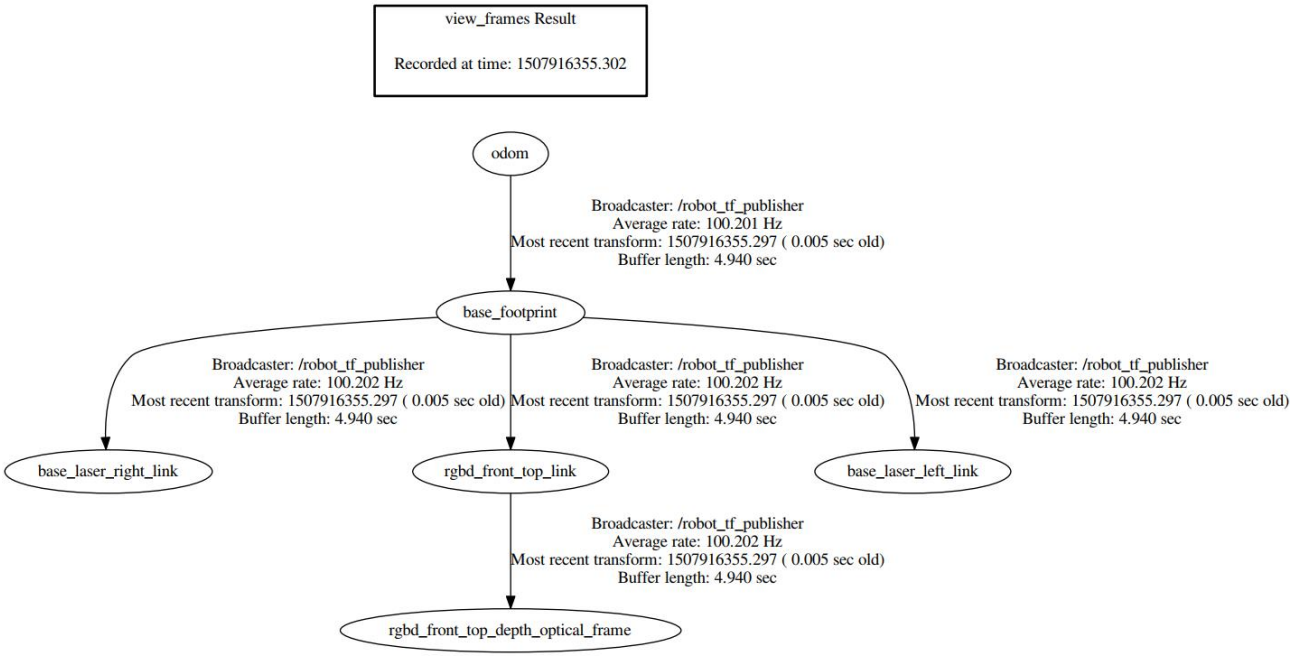

Figura 36 – Estrutura em árvore de todas as transformadas publicadas para o veículo considerado.

4.2.3 Configuração dos sistemas de sensorização

Uma vez feito o download de todos os drivers para o ambiente ROS da pagina de suporte de cada um dos sensores, é necessário realizar uma pequena configuração de cada um deles:

Configuração dos sensores LiDAR 2D:

- *Baud rate*: valor por defeito de 500000.
- Tópico onde são publicadas as leituras do sensor no formato de mensagens: /laser_right e /laser_left.
- Nome do quadro de coordenadas: /base_laser_right_link e /base_laser_left_link.

Configuração da câmara 3D SICK Visionary-T:

- Endereço IP para a comunicação: 192.168.1.10.
- Tópico onde são publicadas as leituras do sensor no formato de mensagens: /camera.
- Nome do quadro de coordenadas: /rgbd_front_top_depth_optical_frame.

4.3 Algoritmos de deteção

Após uma análise das soluções atuais para a deteção de pessoas e da escolha das tecnologias de sensorização a incorporar no modelo final, decidiu-se utilizar duas abordagens diferentes para a deteção de pessoas. Uma primeira abordagem em que a deteção de humanos é efetuada com recurso a imagem de profundidade fornecida pelo sistema de visão 3D adquirido e uma segunda recorrendo aos sensores LiDAR 2D para a deteção de um padrão de pernas.

4.3.1 Deteção de pessoas com recurso a câmara 3D

A deteção de objetos deverá exigir um processamento elevado, levando a atrasos na deteção e a falhas caso o sistema não consiga suportar tal processamento. Para além disto, o sistema deve possuir a capacidade de detetar pessoas, quando se encontra em movimento, dificultando assim esta tarefa. Nesta dissertação é pretendido retirar o máximo partido da informação de profundidade fornecida pela câmara 3D.

No capítulo 2.5 foram apresentadas algumas soluções atuais, que recorrem a informação de profundidade, como o exemplo do robô Care-O-Bot [33] ou o SPENCER [36], em que ambos disponibilizam as *frameworks* dos seus sistemas de forma livre e gratuita. Apesar de ambos apresentarem bons resultados na realização das suas tarefas, o sistema do SPENCER é o que mais se adequa ao âmbito desta dissertação. Este detetor utiliza ao máximo a informação de profundidade do sensor RGB-D de modo a acelerar a computação e a providenciar um sistema em tempo-real, mas por outro lado o alcance de deteção encontra-se limitado pelo alcance em profundidade da câmara, que no caso da Kinect, revela limitações a distâncias superiores a 2 m, nomeadamente ao nível da perceção da cabeça do ser humano. O sistema de deteção que recorre a informação de profundidade pode ser dividido em 3 etapas: Identificação de estruturas e objetos, Processamento da ROI (Região de Interesse) e a deteção de Pessoas.

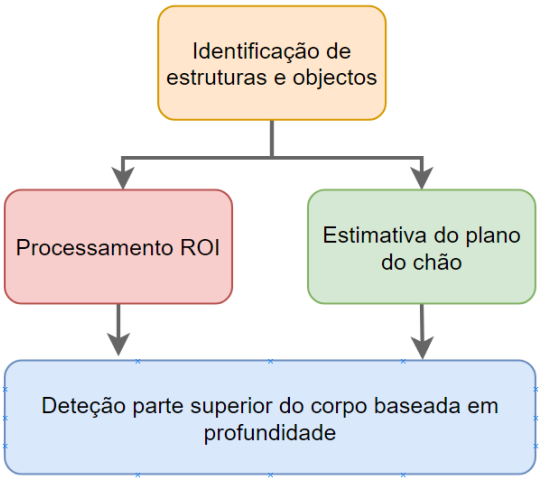

Figura 37 - Esquema do detetor recorrendo ao sensor RGB-D.

Na primeira etapa de identificação de estruturas e objetos, os pontos 3D recebidos da câmara 3D ao longo do tempo são unidos sobre um período de tempo e projetados num referencial utilizando os movimentos estimados pelo robô. Esses pontos 3D são então segmentados em três diferentes classes: objetos, plano do chão e estruturas fixas, como por exemplo paredes. Esta segmentação permite excluir todos os pontos 3D classificados como estruturas fixas antes da extração da ROI, diminuindo o processamento e reduzindo os falsos positivos. O próximo passo é estimar o plano do chão, através do cálculo de um mapa de ocupação, em que todos os pontos 3D dentro de um corredor de 2 m de altura são projetados para obter uma estimativa aproximada do plano do chão, baseando-se sempre na altura da câmara. Todos os pontos 3D dentro de classes com maior com maior densidade são excluídos e os restantes pontos são percorridos por um algoritmo RANSAC (*Random Sample Consensus*) [68] de modo a estimar quais os pontos que pertencem ao plano do chão. Este algoritmo (RANSAC) é um método que permite extrair os *inliers*[6] e *outliers*[7] de um conjunto de pontos. Neste caso todos os pontos *inliers* correspondem ao plano do chão.

O calculo do mapa de ocupação é calculado com recurso a um histograma 2D, em que para cada classe do histograma é calculado um histograma de alturas composto por quatro bandas de altura: a banda do plano do chão, a banda dos objetos, a banda do espaço livre e a banda de estruturas elevadas. A banda de espaço livre tem bastante importância no desempenho da classificação, pois

[6] Inliers – São pontos de dados que se ajustam com um determinado modelo desejado, dentro de uma tolerância de erro.
[7] Outliers – São pontos de dados que não se ajustam ao modelo de objeto desejado, estando fora dos valores de tolerância.

situações como zonas salientes de edifícios ou árvores criavam erros de classificação. Desta forma é assumido que irá haver sempre um corredor livre (2 m a 2.3 m) entre a cabeça das pessoas e edifícios com zonas salientes ou árvores. Isto significa que em cada classe de um histograma 2D a densidade dos pontos na banda de espaço livre é baixa, os pontos na banda dos objetos são classificados como objetos e os pontos dentro das bandas de espaço livre e estruturas elevadas são classificadas com estruturas fixas. Devido aos ruídos da imagem de profundidade e a odometria do robô o corredor livre pode conter alguns pontos falsos. Neste caso, esses pontos no corredor livre só são classificados como estruturas fixas caso possuam uma densidade elevada. Os pontos classificados como estruturas fixas são excluídos de qualquer processamento adicional e os pontos classificados como objetos passam ao próximo modulo de processamento de ROI.

A segunda etapa consiste no processamento da ROI. Este pode ser dividido em duas etapas distintas: Extração da ROI e a Segmentação da ROI.

Na extração do ROI todos os pontos definidos como objetos no plano do chão são recolhidos num histograma 2D e são votados de acordo com a distância quadrática à câmara de modo a compensar os objetos que se encontram mais afastados da câmara, pois esses são constituídos por menos pontos. Cada classe dos histogramas é suavizada com recurso a um filtro gaussiano [69] e os valores irrelevantes causados pelo ruído são eliminados através de um valor de *threshold*. As restantes classes são então agrupadas em componentes conectados utilizando o método *8-neighborhood* [70], em que os pixéis em redor de um dado pixel ficam conectados a este caso apresentem alguma similaridade. Este processo é então repetido para todos os restantes pixéis. Para cada ROI é então atribuído um retângulo ao centro do ROI, com o comprimento do ROI e com a altura do ponto mais alto da classe correspondente, orientado paralelamente a câmara. Desta forma a imagem a ser avaliada pelo detetor é já obtida.

Na segmentação dos ROIs os componentes são segmentados em regiões distintivas. Como por vezes as pessoas podem encontrar-se próximas umas das outras na imagem (Figura 38 (b)), ao aplicar a deteção da parte superior do corpo será necessário o redimensionamento do modelo normalizado de forma a compensar as diferentes alturas de cada uma delas. De forma a evitar o redimensionamento, estes métodos propõem segmentar os componentes conectados em regiões distintivas. Com este objetivo, recorrem ao algoritmo Quick Shift, sendo esta uma variante mais rápida do Mean Shift apresentado no capítulo 3.3.3. Este algoritmo encontra a moda[8] de uma densidade através do deslocamento de cada ponto para o mais próximo na vizinhança com o valor de densidade maior. Cada

[8] - É uma medida de tendência central de um conjunto de dados.

ponto do histograma é ligado à sua respetiva vizinhança com um valor de densidade mais alta até atingir a moda. Os pontos a caminho da moda são automaticamente associados a esta moda. Na Figura 38 (c), podemos visualizar esta mesma segmentação das ROIs em objetos individuais.

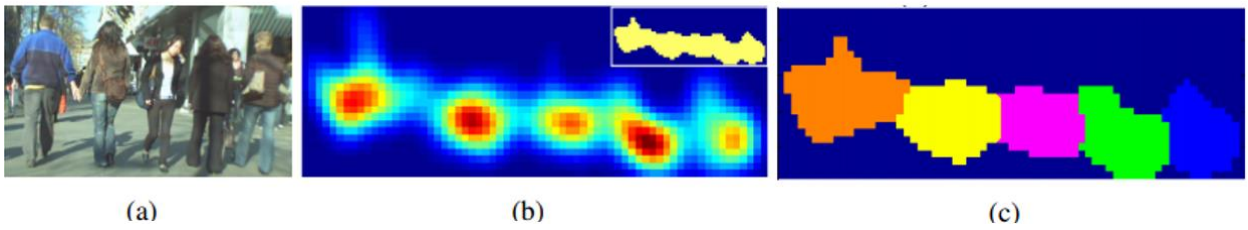

Figura 38- Segmentação e extração da região de interesse, imagem adaptada de [38]. (a) Imagem original. (b) Projeção das 5 pessoas no plano do chão. No canto superior a direita podemos visualizar o componente conectado (c) Segmentação da ROI em objetos individuais.

A terceira e última etapa é a deteção de pessoas através de um detetor da parte superior de corpo (formato dos ombros, pescoço e cabeça). A ideia principal é a utilização de um modelo normalizado de profundidade da parte superior do corpo (Figura 39), em que este é deslizado sobre a ROI ao mesmo tempo que a área sobreposta é comparada para verificar se o ROI apresenta semelhanças nos contornos em relação ao modelo. A escala do modelo inicial é estimada através da altura do ROI 2D (determinado pela pessoa mais alta do grupo). Como os ROIs podem conter várias pessoas próximas umas das outras o modelo é então redimensionado para compensar as diferentes alturas de cada uma delas. Como estes redimensionamentos podem gerar várias deteções positivas sobre a mesma pessoa, é utilizado o algoritmo NMS (*non-maximum-suppression*) para reduzir a uma só deteção. Assim sendo a avaliação do modelo é efetuada só na máxima local dos contornos extraídos do ROI. Os contornos são extraídos da imagem de profundidade de um determinado ROI e são representados por um valor y máximo a cada posição x. Com o objetivo de reduzir o numero de máximas locais na imagem de profundidade que são causadas pelo ruído, os contornos são suavizados através de um filtro gaussiano, depois da extração das mesmas. Como por vezes a máxima local não se encontra exatamente ao centro da cabeça das pessoas, o modelo é avaliado para cada escala na máxima local e ainda na sua vizinhança.

Figura 39 - Modelo da parte superior do corpo humano utilizado para a deteção de pessoas.

Como foi mencionado anteriormente, a utilização da máxima local para encontrar a escala correta da pessoa na imagem requer o redimensionamento do modelo. Contudo, a segmentação do ROI permite produzir ROIs que contém unicamente uma pessoa, permitindo ajustar a escala do modelo e aplica lo somente uma vez. Desta forma, foi combinada a utilização da máxima local juntamente com a segmentação do ROI para obter um melhor desempenho de deteção, quando comparado a utilização só da máxima local nos ROIs originais.

Todo este algoritmo foi testado e implementado no Robô Spencer apresentado anteriormente. Apesar de este possuir uma boa capacidade de deteção de pessoas com recurso as câmaras de profundidade, todo o algoritmo foi desenvolvido para uma câmara específica neste caso a Kinect v1 ou a Asus Xtion Pro, encontrando-se limitado ao alcance das mesmas. A biblioteca utilizada para aceder a imagem destas câmaras através do *middleware* ROS é o Openni, em que este fornece já alguns tópicos de imagem com diferentes formatos. O detetor da parte superior do corpo do SPENCER necessita de dois tópicos da câmara 3D: um denominado "*camera_info*", que contém a calibração da câmara e alguns metadados e outro denominado "*image_rec*", que contém uma imagem de profundidade já retificada. As profundidades são tipicamente fornecidas em metros e os valores são do tipo *float*.

De modo a obter um alcance superior, como referido anteriormente, a câmara escolhida para integrar no veículo de estudo desta dissertação foi a Visionary-T. Esta quando comparada a Kinect v1 ou a ASUS, apresenta um alcance bastante superior (7 m) mas por outro lado apresenta uma resolução inferior (176 x 144 pixéis). Para além disto, o formato de imagem fornecido é 16UC (16 bits *unsigned*), não sendo compatível com formato de imagem utilizada pelo algoritmo de deteção 32FC (32 bits *float*).

Uma das possíveis soluções seria a alteração de todo o algoritmo para uma nova resolução de imagem e um novo formato, mas esta seria uma tarefa bastante complexa e morosa. Desta forma, como

demonstrado na Figura 40, a abordagem utilizada é o desenvolvimento de um novo *package* no ROS que efetue um pré-processamento da imagem. Este pré-processamento consiste em receber a imagem de profundidade de uma câmara e caso a resolução e os valores desta não se encontrem de acordo com o formato pretendido, efetue o devido redimensionamento e conversão. Para além disto, o tópico "*camera_info*" terá também de ser atualizado para os novos valores, sendo depois publicado como "*new_camera_info*".

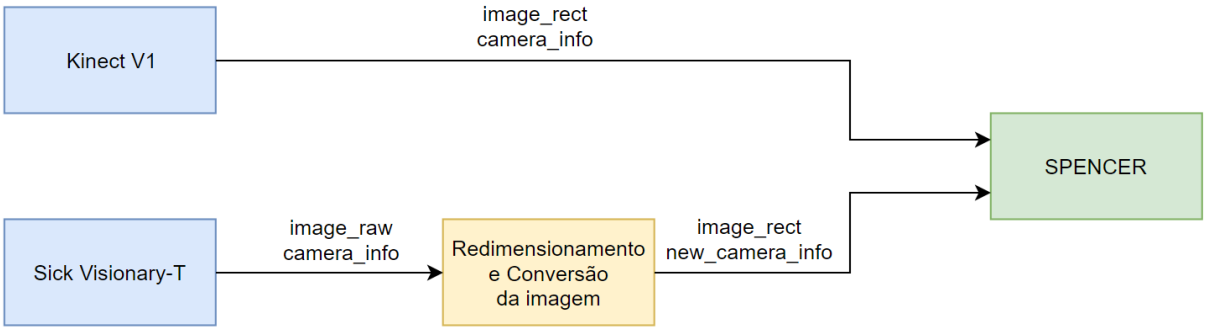

Figura 40 - Modelo de comunicação da Kinect V1 e a proposta de solução para a nova câmara SICK Visionary-T.

Para efetuar o pré-processamento da imagem (redimensionamento e conversão do formato) recorreu-se ao OpenCV [26]. A primeira etapa deste processo é a conversão da imagem do ROS para o OpenCV, pois o ROS possui o seu próprio formato de mensagens de imagem. Contudo, este possui já uma biblioteca denominada *cv_bridge* que providencia uma interface entre o ROS e o OpenCV, com funções próprias para efetuar a conversão de mensagens de imagem do ROS para o OpenCV. O modelo de comunicação entre o ROS e o OpenCV através da biblioteca *cv_bridge* pode ser visualizada na Figura 41.

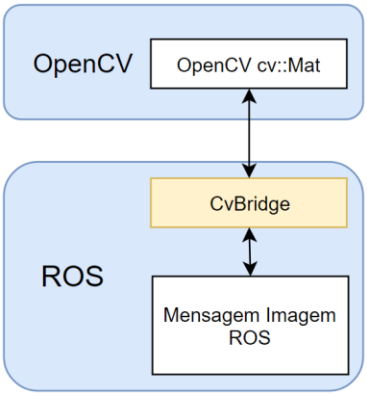

Figura 41 - Modelo comunicação entre ROS e OpenCV, recorrendo a biblioteca cv_bridge.

Na Figura 42, podemos visualizar a ordem de todas as etapas de pré-processamento da imagem.

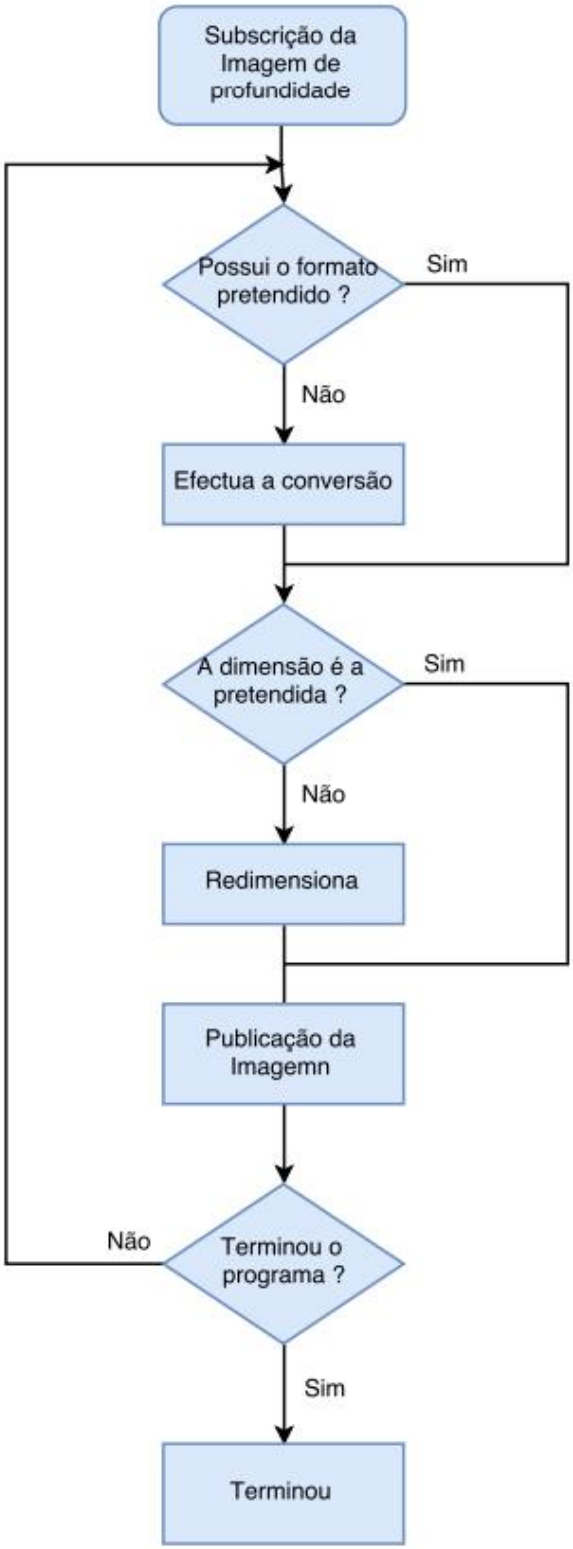

Figura 42 - Fluxograma do package ROS desenvolvido para o pré-processamento da imagem (Conversão e Redimensionamento).

A função utilizada, toCvCopy, em que cria uma cópia dos dados da mensagem da imagem do ROS, convertendo-a para o formato do OpenCV. Esta função requer dois argumentos de entrada: um apontador para a mensagem de imagem a converter para OpenCV; um argumento opcional de qual é o tipo de formato que pretende que a imagem seja depois de convertida para o OpenCV. Este último argumento opcional será utilizado para converter o formato da imagem da câmara SICK (16UC1) para o formato pretendido pelo detetor (32FC1).

Depois da conversão, resta o redimensionamento da imagem caso este seja necessário. Para isto, é utilizada uma função do OpenCV, que recebe sempre como argumento de entrada a imagem a redimensionar, a imagem de saída, o tamanho pretendido da nova imagem e o método de interpolação pretendido. Esta função permite redimensionar uma imagem quer para um tamanho superior como inferior a original. Quanto ao método de interpolação a utilizar estes podem ser:

- INTER_NEAREST – Neste método de interpolação, para cada pixel de saída interpolado, é associado o valor do pixel mais próximo na imagem de entrada.
- INTER_LINEAR – Na Interpolação Bilinear, o valor de um novo pixel é determinado a partir da média ponderada dos 4 pixéis adjacentes a este.
- INTER_AREA - O redimensionamento é efetuado utilizando a relação entre a área e os pixéis;
- INTER_CUBIC – Interpolação Bicúbica, o valor de um novo pixel é determinado a partir da média ponderada de uma matriz de 16 pixéis da sua vizinhança.
- INTER_LANCZOS4 – A Interpolação Lanczos é realizada sobre 8x8 pixéis na sua vizinhança.

Nesta dissertação, estes métodos de interpolação, vão permitem estimar os novos pixéis a partir da imagem original, de modo a aumentar a resolução da imagem. O método de interpolação utilizado foi o INTER_NEAREST, pois apresentou sempre uma deteção estável. Uma vez feito o processamento de toda a imagem para o formato pretendido, é necessário atualizar os novos parâmetros da câmara. Os parâmetros de calibração da câmara podem ser divididos em 2 grupos: Parâmetros Intrínsecos e Extrínsecos. Os parâmetros extrínsecos definem a localização e orientação do sistema de coordenadas da câmara ao sistema de coordenadas do ambiente de captura da imagem. Os parâmetros intrínsecos são necessários para ligar as coordenadas de um pixel de um ponto de uma imagem com a coordenada correspondente no sistema de coordenadas da câmara.

Como é realizado um redimensionamento da imagem, torna-se necessário atualizar os parâmetros intrínsecos da câmara. Na matriz abaixo Equação (6) podemos visualizar todos os parâmetros intrínsecos que são necessários atualizar para a nova dimensão de imagem.

$$K = \begin{bmatrix} fx & 0 & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

Onde:

- fx e fy representam a distância focal em pixéis.
- cx e cy representam o centro ótico, também conhecido como $principal\ point$.

De modo a calcular estes novos valores, é calculada a razão entre a resolução da imagem pretendida e a atual, como podemos visualizar na Equação (7), em que é calculada a razão para a altura e a largura da imagem. Depois disto, os valores da distância focal (fx e fy) e o centro ótico (cx e cy) são atualizados, através da multiplicação da razão obtida pelos valores da imagem original Equação (8).

$$\begin{aligned} Razão\ x &= Nova\ largura \div Largura\ da\ imagem\ atual \\ Razão\ y &= Nova\ altura \div Altura\ da\ imagem\ atual \end{aligned} \quad (7)$$

$$\begin{aligned} fx &= fx \times Razão\ x \\ fy &= fy \times Razão\ y \\ cx &= cx \times Razão\ x \\ cy &= cy \times Razão\ y \end{aligned} \quad (8)$$

Uma vez que os parâmetros da imagem se encontram atualizados para a nova dimensão de imagem, é realizada a publicação destes novos valores no tópico ROS "…/camera _info_new", para serem acedidos pelo detetor.

4.3.2 Deteção de um padrão de perna

Apesar da $framework$ SPENCER incorporar o seu algoritmo para a deteção de pernas, este foi desenvolvido para um padrão obtido pelo LiDAR 2D a uma altura de 70 cm do chão. Como a perceção

do ambiente pelo LiDAR é realizada próxima da cintura do ser humano por vezes não é possível identificar esse padrão, principalmente em ambientes industriais ou laboratórios, onde o uso de batas é obrigatório.

No entanto, o ROS dispõe já de um *package* denominada "leg_detetor", que recebe como entrada as leituras do LiDAR 2D e recorre a uma técnica de aprendizagem máquina para identificar possíveis padrões de pernas presentes nessas mesmas leituras.

Este possui um ficheiro de configurações "LegDetection.cfg", onde todos os parâmetros são definidos (como por exemplo o numero de pontos num grupo), de forma a definir os valores padrão de uma perna. Se duas pernas são encontradas próximas uma da outra e a separação entre elas satisfaz os parâmetros definidos (normalmente, a distância de separação entre duas pernas encontra se entre os 10 cm a 50 cm), é então detetada uma pessoa.

Para o seguimento de humano recorreram a um filtro Kalman, em que este recebe a posição das pessoas detetadas, de forma a manter uma estimativa da posição da pessoa ao mesmo tempo que verifica a veracidade da mesma. Desta forma, este filtro permite reduzir a inexatidão da posição de humanos obtida pelos varrimentos do sensor LiDAR 2D.

Se as estimativas satisfazerem os parâmetros de configuração do filtro, a informação da pessoa é então convertida para o formato de mensagens do ROS e publicada em tópicos. São 3 os tópicos que este package publica:

- Leg_tracker_measurements - fornece uma estimativa da posição de pernas individuais;
- People_tracker_measurements - fornece uma estimativa da posição de pessoas através da junção de pernas, em que a posição é dada pela média entre ambas.
- Visualization_marker – Representa a posição das pernas e pessoas através de marcadores.

5. RESULTADOS

Uma vez implementado o sistema descrito no capítulo anterior, foram realizados vários testes para verificar o desempenho do mesmo. Tanto para o LiDAR 2D como para a câmara foram definidos vários casos de teste. Neste capítulo são apresentados todos os resultados obtidos.

5.1 Deteção da parte superior do corpo com a câmara SICK Visionary-T

Para a câmara 3D foram definidas duas plataformas de teste para avaliar o desempenho da deteção da parte superior do corpo em diferentes condições, como se pode visualizar na Figura 43. A primeira plataforma consiste numa câmara 3D montada num tripé fixo (Figura 43 a) e a segunda plataforma consiste nessa mesma câmara montada sobre um *trolley* para a realização de testes em movimento (Figura 43 b). Em ambos os casos a câmara encontra-se a 1.6 m do plano do chão.

(a) (b)

Figura 43 - Plataforma de testes com a câmara SICK Visionary-T: (a) Tripé; (b) Trolley;

O ambiente de teste utilizado é um corredor de uma oficina de forma a simular um ambiente tipicamente industrial. Foram efetuadas marcações no chão para verificar o alcance de deteção em

diferentes situações, como se pode visualizar na Figura 44, em que as marcações se encontravam 50cm distanciadas entre si.

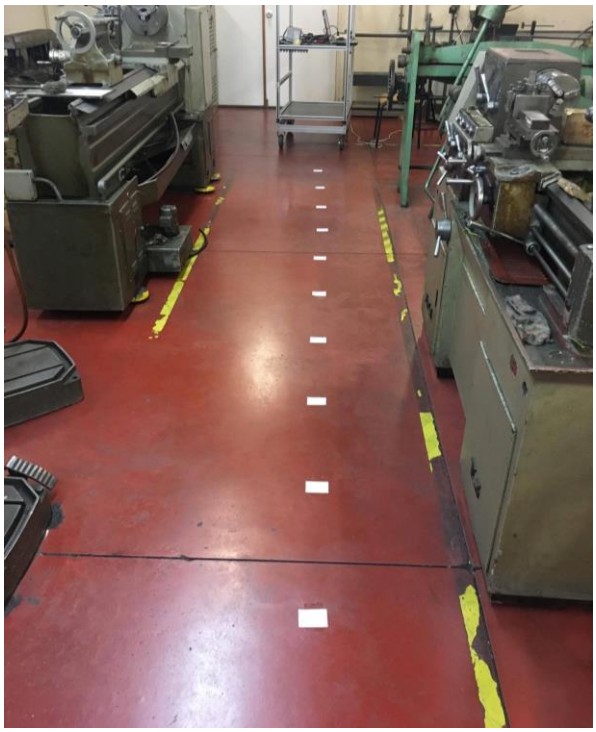

Figura 44 – Cenário de teste.

Uma vez montado o cenário, foram definidos vários casos de teste, para verificar o comportamento do sistema em diferentes situações. Em todos os casos foram verificados o alcance e a estabilidade de deteção de pessoas. Como o sistema proposto utiliza métodos de interpolação para efetuar o redimensionamento da imagem, o primeiro teste realizado foi o alcance de cada um dos métodos de interpolação. Este teste foi realizado com a câmara montada no tripé e com uma pessoa a aproximar-se do fundo. Na Tabela 5 podemos visualizar o alcance de cada um dos métodos. Como o INTER_NEAREST apresentou melhores resultados, este foi o método escolhido a aplicar no sistema final. Na Figura 45, podemos visualizar a deteção de uma pessoa a aproximar se do fundo a uma distância de 4 m.

Tabela 5 - Alcance dos diferentes métodos de interpolação.

Método de interpolação	Alcance
INTER_NEAREST	0,5m a 4 m
INTER_LINEAR	0,5 m a 3,5 m
INTER_AREA	0,5m a 3,5 m
INTER_CUBIC	Sem deteção
INTER_LANCZOS4	Sem deteção

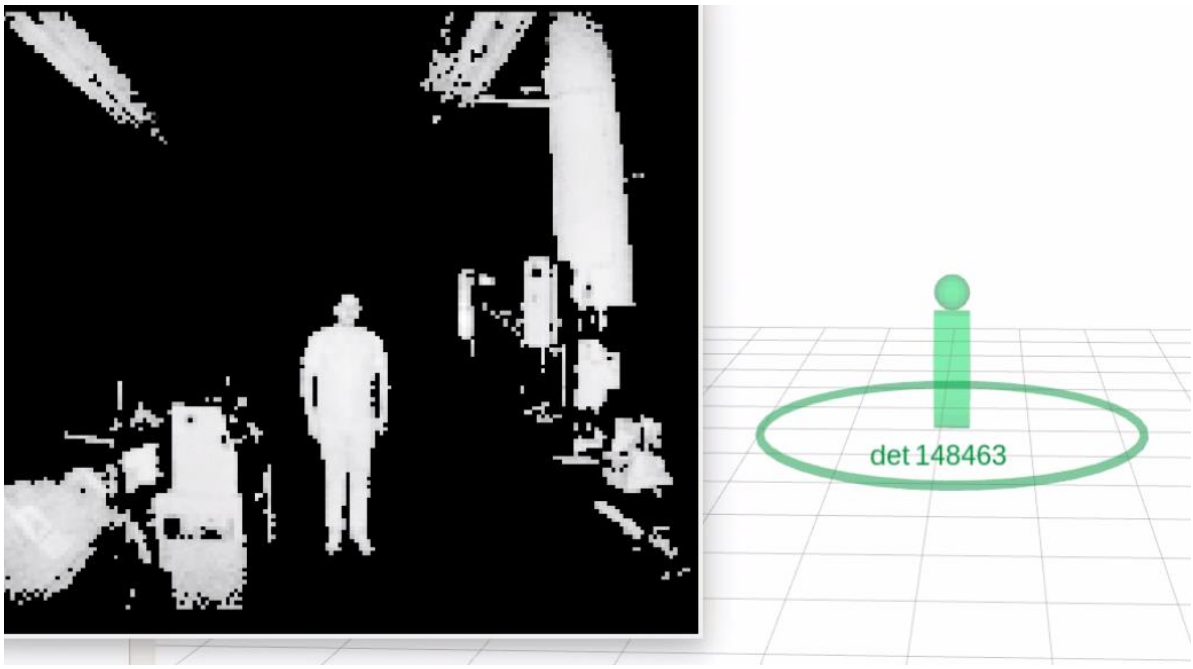

Figura 45 - Deteção de uma pessoa a 4 metros a aproximar-se da câmara montada num tripé.

Apesar da câmara SICK Visionary-T apesar de possuir um alcance de 7 m, esta não providencia uma imagem de profundidade com uma boa perceção da parte superior do corpo humano a distâncias superiores a 4,5 m, como podemos visualizar na Figura 46.

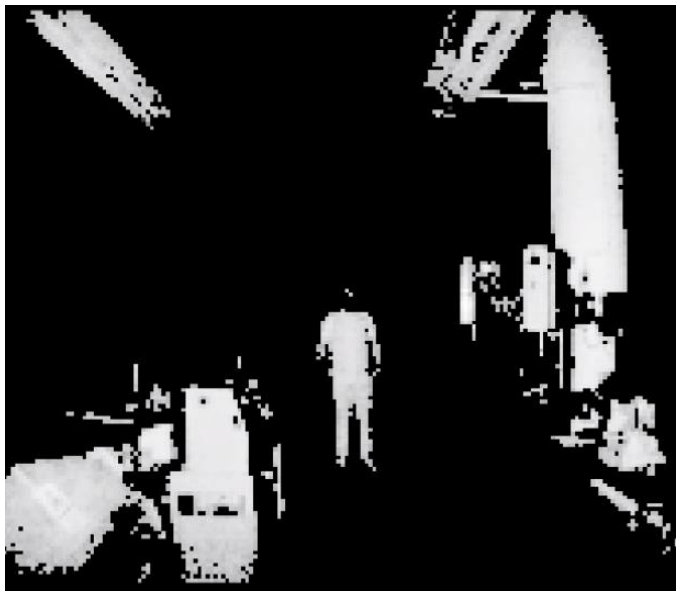

Figura 46 - Imagem de profundidade a uma distância de 5m.

Esta distância de 4,5 m fica reduzida para os 4 m na imagem pré-processada devido a aplicação dos métodos de interpolação, pois estes tem de estimar pixéis de modo a redimensionar a imagem para uma resolução superior. Este processo coloca um pequeno ruído na imagem.

Os casos de teste realizados com a câmara montada no tripé, de modo a simular o veículo parado foram os seguintes:

- Uma pessoa a aproximar se do veículo vindo do fundo do corredor;
- Duas pessoas a caminhar no mesmo corredor, com sentidos opostos;
- Duas pessoas a caminhar lado a lado, no corredor, em direção ao veículo;
- Três pessoas a caminhar, no corredor, na direção do veículo, em que duas vão a frente lado a lado e uma vai atrás;
- Uma pessoa a cruzar-se de perfil com o veículo;

No caso do veículo parado, o alcance de deteção encontra-se entre os 0,5 m e os 4 m, para uma pessoa a aproximar-se da câmara vinda do fundo a passo normal. Caso esta se aproxime em passo de corrida a deteção é efetuada aos 3,5 m. Em situações de duas ou mais pessoas em movimento a deteção é efetuada aos 3,5 m. Um exemplo de deteção pode ser visualizado na Figura 47.

Figura 47- Deteção de duas pessoas.

Já na presença de 3 pessoas em que uma se encontre sobreposta por outras na imagem esta não é detetada pois o sistema não obtém a perceção da mesma, um exemplo pode ser visualizado na Figura 48.

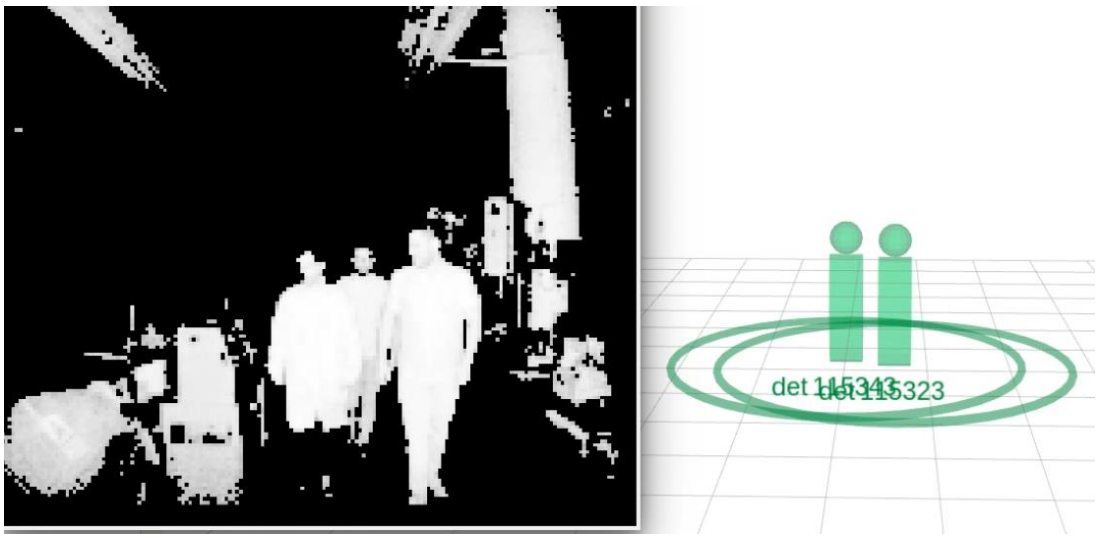

Figura 48- Deteção de grupo de 3 pessoas, em que uma se encontra sobreposta.

No caso de uma pessoa a cruzar-se de perfil o sistema apresenta alguma dificuldade de deteção principalmente a distâncias superiores a 2,5 m, pois o modelo da parte superior do corpo humano apresenta diferentes características em relação ao modelo utilizado. Na Figura 49 podemos visualizar a deteção de uma pessoa de perfil a 2 m, mas, no entanto, esta deteção não é estável.

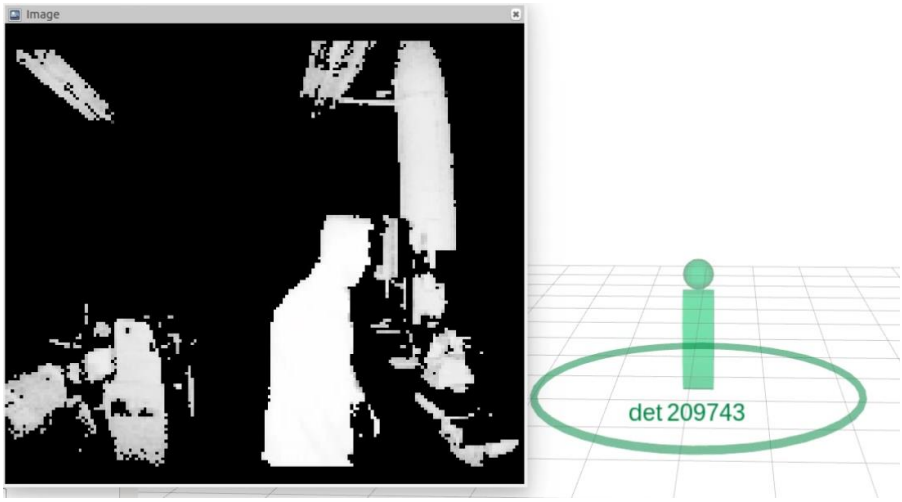

Figura 49- Deteção de pessoas de perfil.

Os casos de teste realizados com a câmara montada no trolley, de modo a simular o veículo em movimento foram:

- Veículo e uma pessoa a caminharem na mesma direção e sentidos opostos;
- Veículo a percorrer o corredor da esquerda para a direita com uma pessoa a aproximar-se do fundo;
- Veículo a sair de um cruzamento com uma pessoa a aproximar-se;
- Uma pessoa a cruzar-se de perfil com o veículo;

Para o veículo em movimento, o sistema de deteção possui um alcance entre os 0,5 m e os 3,5 m. Mesmo em situações em que o veículo sai de um corredor para outro, desde que as pessoas se encontrem no campo de visão da câmara e a parte superior do corpo humano seja percetível, a deteção é efetuada com sucesso.

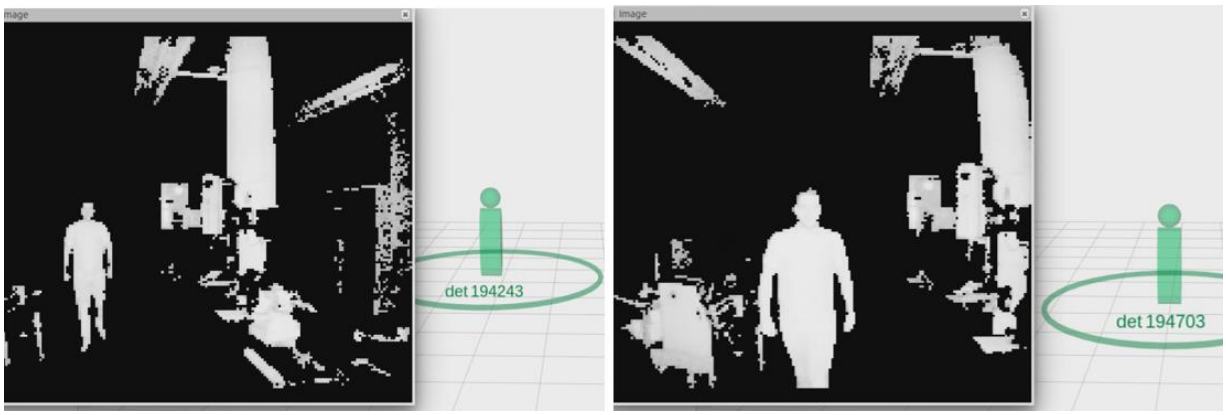

Figura 50- Sequencia de duas imagens de deteção de uma pessoa com a câmara em movimentos aleatórios.

Na deteção de duas ou mais pessoas, em que estas se encontrem sobrepostas, a deteção é efetuada somente nas que se encontrem mais próximas do veículo, pois são as únicas que o sistema de deteção obtém perceção da região superior do corpo humano. Na deteção de perfil o sistema a deteção não é estável pois, como referido anteriormente, o modelo escolhido não permite a correspondência com uma pessoa de perfil. No caso de pessoas escondidas atrás de objetos ou encostadas a estes, a deteção pode não é efetuada pois não existe uma boa perceção de profundidade. Exemplo disto pode ser visualizado na Figura 51 , onde não é percetível os contornos do corpo humano, pois este encontra-se encostado a uma parede.

Figura 51 - Imagem de profundidade de pessoa junto a uma parede.

5.2 Deteção de um padrão de pernas com o sensor LiDAR SICK S300

Já o LiDAR 2D, foi testado no veículo transportador pois este sensor possui uma saída de dados RS-422, não sendo possível a ligação com um computador convencional. Para a realização dos testes foi utilizado um dos LiDAR 2D do veículo a uma altura de 20 cm do chão, Figura 52. Como estes veículos encontram-se em desenvolvimento, não foi possível testar este sensor em movimento. Desta forma recorreu-se a outro sensor LiDAR 2D denominado RPLIDAR para a realização de alguns testes de forma a validar o conceito noutro tipo de ambientes.

Figura 52- Posição de um dos sensores LiDAR no veículo.

Esta abordagem foi testada em dois ambientes diferentes: um ambiente livre, em que se recorreu ao RPLIDAR; um ambiente lotado, em que se utilizou o SICK S300 montado no veículo.

Todo o processo de deteção pode ser visualizado na Figura 53 e Figura 54, ambos realizados num ambiente livre. Na Figura 53 é efetuada a deteção de o padrão de uma perna individualmente, em que os pontos a azul representam a deteção de cada uma das pernas. No caso da presença de duas pernas próximas uma da outra (distância entre pernas entre os 10 cm e os 50 cm) é então detetada uma pessoa, Figura 54, em que a posição desta é dada como a média entre as duas pernas e representada através de um ponto verde.

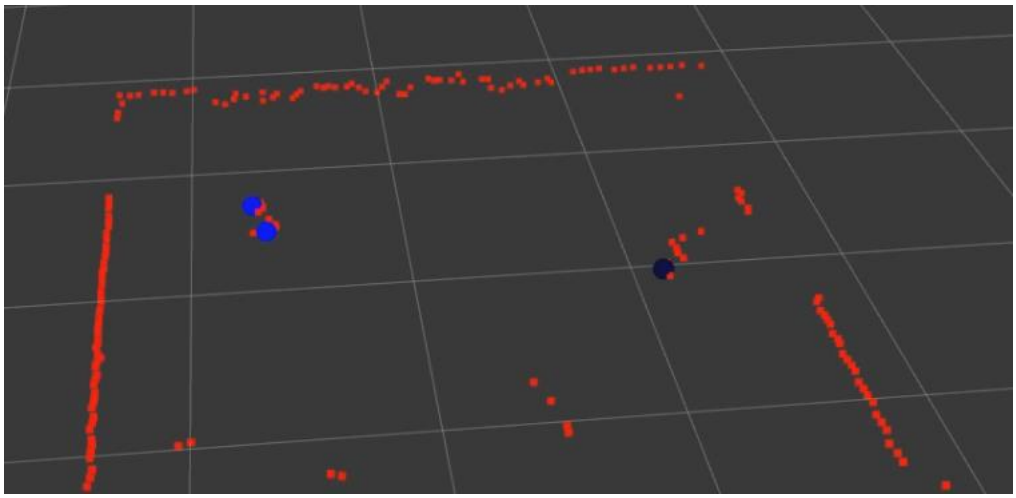

Figura 53 - Deteção de um padrão de uma perna, onde cada ponto azul representa uma deteção.

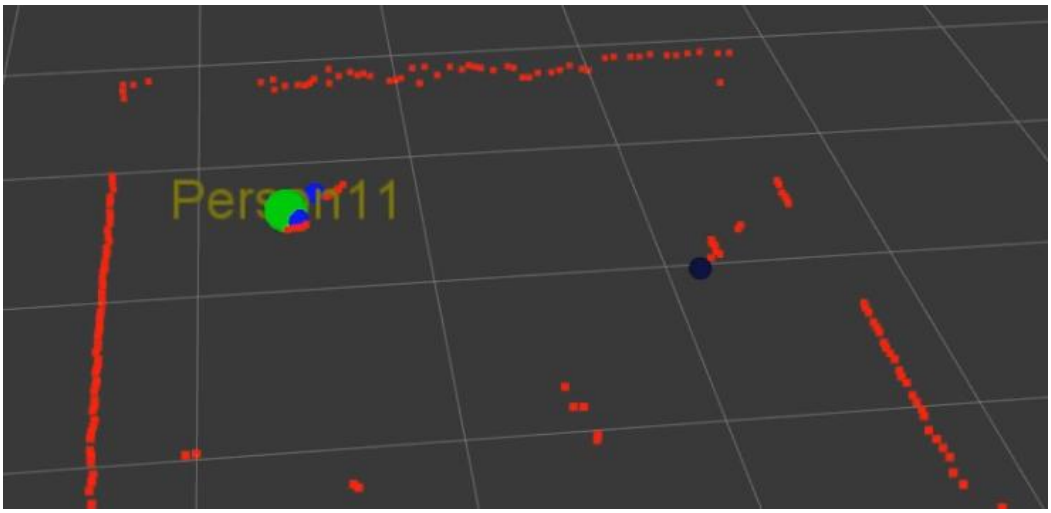

Figura 54 - Deteção de uma pessoa (ponto verde) através da associação de duas pernas próximas uma da outra (pontos a azul).

No caso de um ambiente lotado de objetos, podem ocorrer alguns falsos positivos principalmente em ambientes que possam conter objetos semelhantes a um padrão de pernas, pois o LiDAR 2D não possui grande capacidade discriminatória. Na Figura 55 podemos visualizar um ambiente lotado, em que foi utilizado o sensor SICK S300 montado no veículo. Como podemos visualizar apesar de se verificar a deteção três padrões de pernas, dois são falsos positivos.

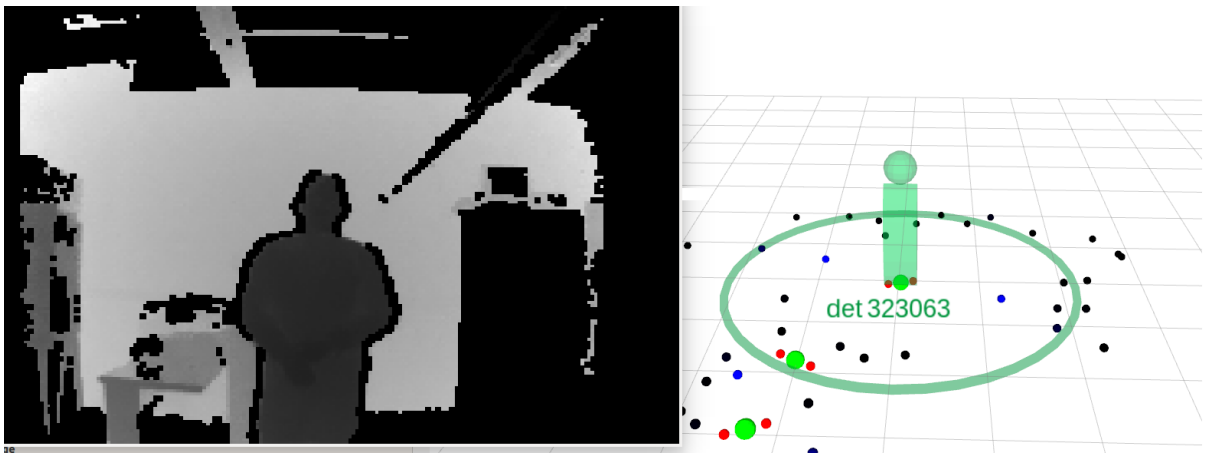

Figura 55- Deteção de um padrão de pernas com alguns falsos positivos.

Apesar de este método apresentar bons resultados para a deteção de um padrão de pernas, não possui a capacidade de detetar pessoas quando estas possuem as pernas juntas (de frente e perfil) para o sensor. Por outro lado, possui um alcance e um campo de visão bastante superior em comparação a câmara de profundidade.

5.3 Validação do sistema de Transformadas

De forma a validar o sistema de transformadas implementado no capítulo 4.2.2, foram realizados testes com a câmara 3D e o LiDAR 2D em simultâneo. Como podemos visualizar na Figura 56 (imagem à direita), onde são representadas todas as deteções de ambos os sensores, a orientação e a posição de cada um encontra-se sincronizada. Os pontos verdes representam a deteção de pessoas através de um padrão de pernas pelo LiDAR 2D, onde apresenta alguns falsos positivos. Cada boneco verde representa a deteção de uma pessoa pela câmara 3D. Apesar de estarem 3 pessoas presentes no ambiente a câmara só consegue detetar duas pois uma das pessoas encontra-se fora do seu campo de visão.

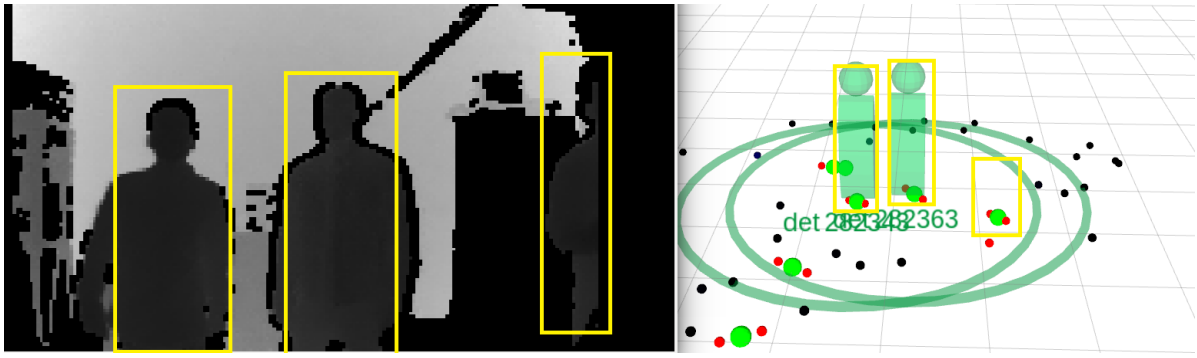

Figura 56 - Deteção de pessoas num ambiente lotado, onde os retângulos amarelos assinalam a presença de pessoas: Imagem à esquerda é a imagem de profundidade do ambiente de teste; Imagem à direita é a representação de todas as deteções num mapa de coordenadas relativo ao sistema.

6. CONCLUSÕES E TRABALHO FUTURO

6.1 Conclusões

Pelos resultados apresentados, conclui-se que o objetivo desta dissertação foi alcançado, apresentando uma abordagem para uma deteção de pessoas em tempo real, a aplicar num SAMU (*Smart Autonomous Mobile Unit*).

O sistema proposto recorre a combinação de uma câmara de profundidade para a deteção da parte superior do corpo humano, juntamente com dois sensores LiDAR para uma deteção de um padrão de pernas.

Para a deteção de pessoas com recurso a uma câmara de profundidade recorreu-se a *framework* SPENCER. Este recorre a um modelo de profundidade da parte superior do corpo humano para a deteção de pessoas numa imagem. Apesar do excelente alcance que a câmara SICK Visionary-T possui, esta apresenta uma baixa resolução, não sendo compatível com a *framework* SPENCER. A abordagem proposta é a implementação de um package ROS que realize todo o pré-processamento da imagem, como o redimensionamento, a conversão de formatos e a atualização de todos os parâmetros de modo a ser compatível com esta *framework*.

Apesar de a câmara possuir um alcance de 7 m, a imagem fornecida pela mesma não permite a perceção da parte superior do corpo humano a uma distância superior a 4,75 m, sendo esta distância diminuída para 4 m devido ao uso de métodos de interpolação para estimar novos pixéis durante o processo de redimensionamento da imagem original para uma resolução superior. Apesar de apresentar algumas dificuldades de deteção de pessoas de perfil, para distâncias entre os 0,5 m a 3,5 m apresenta bons resultados de deteção, mesmo quando a câmara se encontra em movimento.

Já a deteção de pessoas através de um padrão de pernas é realizada com recurso a um sensor LiDAR 2D. Este sensor possui um alcance e um campo de visão superior em relação as restantes tecnologias, apresentando bons resultados na deteção de um padrão de pernas. No entanto, este sensor não possui grande capacidade discriminatória, podendo apresentar alguns falsos positivos, principalmente em ambientes lotados ou na presença de objetos semelhantes ao padrão de pernas humano. Apresenta ainda alguma dificuldade de deteção no caso de um padrão de pernas juntas.

Apesar das vantagens e desvantagem de ambas as tecnologias, quando utilizadas em conjunto estas complementam-se. Ficou demonstrado que o sistema apresenta bons resultados na deteção de pessoas em diferentes cenários.

6.2 Trabalho Futuro

No seguimento desta dissertação, existem algumas propostas de trabalho futuro que poderiam complementar a abordagem proposta.

Por exemplo, a implementação de toda esta abordagem proposta no modelo Real, assim como a realização de todos os testes de modo a verificar a robustez do sistema.

No trabalho desenvolvido, foram utilizados métodos de interpolação para o redimensionamento da imagem. Estes métodos inserem algum ruído na imagem redimensionada, ficando para trabalho futuro a utilização de filtros de suavização de imagem que permitissem compensar o ruído inserido, de modo a aumentar o alcance de deteção.

Apesar do sistema proposto recorrer a várias tecnologias de sensorização para obter uma perceção do meio, este não utiliza métodos de fusão sensorial. A implementação desses métodos na abordagem proposta permitiria a redução de alguns falsos positivos e uma deteção mais robusta.

Uma outra tarefa seria desenvolver um filtro de seguimento de pessoas, que recebesse as deteções do LiDAR 2D e da câmara 3D, de modo a estimar a velocidade, posição e orientação de cada uma das pessoas, permitindo um planeamento de rota mais eficaz por parte do sistema de navegação.

Para além disto, poderiam ser implementados outros modelos da parte superior do corpo humano de modo a compensar as lacunas do modelo atual, como por exemplo a deteção de pessoas de perfil.

BIBLIOGRAFIA

[1] L. Huang, LIDAR, Camera and Inertial Sensors Based Navigation Techniques for Advanced Intelligent Transportation Systems, University of California, Riverside, 2010.

[2] J. Han, D. Kim, M. Lee e M. Sunwoo, "Enhanced Road Boundary and Obstacle Detection Using a Downward-Looking LIDAR Sensor," *IEEE Transactions on Vehicular Technology,* vol. 61.3, pp. 971-985, 2012.

[3] W. Fehlman e M. Hinders, Mobile robot navigation with intelligent infrared image interpretation, Springer Science & Business Media, 2009.

[4] L. Zhang, B. Wu e R. Nevatia, "Pedestrian Detection in Infrared Images based on Local Shape Features," em *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, Minneapolis, USA, IEEE, June 2007, pp. 1-8.

[5] T. Oskiper, Z. Zhu, S. Samarasekera e R. Kumar, "Visual odometry system using multiple stereo cameras and inertial measurement unit," em *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, Minneapolis, USA, IEEE, June 2007, pp. 1-8.

[6] J. Zhu, L. Wang, R. Yang e J. Davis, "Fusion of time-of-flight depth and stereo for high accuracy depth maps," em *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, Anchorage, USA, IEEE, June 2008, pp. 1-8.

[7] D. Scharstein e H. Hirschmuller, "Evaluation of Cost Functions for Stereo Matching," em *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, Minneapolis, USA, IEEE, June, 2007, pp. 1-8.

[8] Y. Bar-Shalom, F. Daum e J. Huang, "The probabilistic data association filter," *IEEE Control Systems,* vol. 29.6, pp. 82-100, 2009.

[9] K. Jo, M. Lee, D. Kim, J. Kim, C. Jang, E. Kim e K. Huh, "Overall reviews of autonomous vehicle a1-system architecture and algorithms," *IFAC Proceedings Volumes,* vol. 46.10, pp. 114-119, 2013.

[10] X. Baro, S. Escalera, J. Vitria, O. Pujol e P. Radeva, "Traffic Sign Recognition Using Evolutionary Adaboost Detection and Forest-ECOC Classification," *IEEE Transactions on Intelligent Transportation Systems,* vol. 10.1, pp. 113-126, 2009.

[11] "Waymo," [Online]. Available: https://waymo.com/. [Acedido em 2 05 2017].

[12] Dllu, "Autonomous Waymo Chrysler Pacifica Hybrid minivan undergoing testing in Los Altos, California.," Wikimedia Commons, under the Creative Commons Attribution-Share Alike 4.0 International license,, 19/11/2017.

[13] "Piloto automático Tesla," [Online]. Available: https://www.tesla.com/pt_PT/autopilot. [Acedido em 1 05 2017].

[14] "UBER," [Online]. Available: www.uber.com. [Acedido em 22 05 2017].

[15] Dllu, "Uber self driving car prototype testing in San Francisco.," Creative Commons, under the Creative Commons Attribution-Share Alike 4.0 International license., 7/10/2016.

[16] K. Furmans, "KARIS PRO," [Online]. Available: http://www.karispro.com/. [Acedido em 2 5 2017].

[17] "Mobile Industrial Robots," [Online]. Available: http://www.mobile-industrial-robots.com. [Acedido em 2 5 2017].

[18] "Seegrid," [Online]. Available: https://seegrid.com/. [Acedido em 10 5 2017].

[19] "Robotic trucks for Smart Industry (AGV) - Balyo," [Online]. Available: https://www.balyo.com. [Acedido em 10 5 2017].

[20] H. Surmann, N. Andreas e J. Hertzberg, "An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments," *Robotics and Autonomous Systems,* vol. 45.3, pp. 181-198, 2003.

[21] S. Thrun, D. Fox e W. Burgard, "A real-time algorithm for mobile robot mapping with application to multi-robot and 3D mapping," em *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, San Francisco, USA, IEEE, April 2000, vol.1, pp. 321-328.

[22] D. Hähnel, W. Burgard e S. Thrun, "Learning compact 3D models of indoor and outdoor environments with a mobile robot," *Robotics and Autonomous Systems,* vol. 44.1, pp. 15-27, 2001.

[23] J. Cunha, E. Pedrosa, C. Cruz, A. Neves e N. Lau, "Using a depth camera for indoor robot localization and navigation," em *DETI/IEETA-University of Aveiro, Portugal*, 2011.

[24] D. Correa, D. Sciotti, M. Prado, D. Sales, D. Wolf e F. Osorio, "Mobile Robots Navigation in Indoor Environments Using Kinect Sensor," em *Critical Embedded Systems (CBSEC), 2012 Second Brazilian Conference on*, Campinas, Brazil, IEEE, May 2012, pp. 36-41.

[25] J. Rocha, "OpenKinect," [Online]. Available: "http://openkinect.org/wiki/Main_Page", 2012. [Acedido em 06 02 2017].

[26] "Opencv homepage," 2017. [Online]. Available: http://opencv.org/, Acedida em Janeiro de 2017. [Acedido em 15 05 2017].

[27] N. Bellotto e H. Huosheng, "Multisensor-based human detection and tracking for mobile service robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 39.1, pp. 167-181, 2009.

[28] P. Viola e M. Jones, "Rapid object detection using a boosted cascade of simple features," em *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, Kauai, HI, IEEE, December 2001, pp. I-I.

[29] R. Lienhart e J. Maydt, "An extended set of haar-like features for rapid object detection," em *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, Rochester, NY, IEEE, September 2002, pp. I-I.

[30] Indif, "Viola Jones feature types," Wikimedia Commons, under the Creative Commons Attribution-Share Alike 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license, 4/10/2010.

[31] Z. Niu, S. Shan, S. Yan, X. Chen e W. Gao, "2d cascaded adaboost for eye localization," em *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on IEEE*, Hong Kong, China, IEEE, August 2006, pp. 1216-1219.

[32] A. Treptow, G. Cielniak e T. Duckett, "Real-time people tracking for mobile robots using thermal vision," *Robotics and Autonomous Systems,* vol. 54.9, pp. 729-739, 2006.

[33] U. Reiser, C. Connette, J. Fischer, J. Kubacki, A. Bubeck, F. Weisshardt e A. Verl, "Care-O-bot® 3-creating a product vision for service robot applications by integrating design and technology," em *{Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, St. Louis, MO, IEEE, Outubro 2009, pp. 1992-1998.

[34] J. Wang, "Care-O-Bot grasping an object on the table," wikimedia commons, under the Creative Commons Attribution-Share Alike 2.0 Generic license., 26/10/2010.

[35] M. Grabner, H. Grabner e H. Bischof, "Fast approximated SIFT," *Computer Vision–ACCV 2006,* pp. 918-927, 2006.

[36] R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila e H. Hung, "Spencer: A socially aware service robot for passenger guidance and help in busy airports," em *Field and Service Robotics*, Springer, 2016, pp. 607-622.

[37] W. Day e H. Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *Journal of classification,* vol. 1.1, pp. 7-24, 1984.

[38] D. Mitzel e B. Leibe, "Close-Range Human Detection for Head-Mounted Cameras," em *British Machine Vision Conference (BMVC)*, 2012.

[39] N. Dalal e B. Triggs, "Histograms of oriented gradients for human detection," em *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA., June, 2005.

[40] F. Reis, *Detecção de pedestres em imagens,* Lisboa, 2011.

[41] A. Hearst, S. Dumais, E. Osuna, J. Platt e B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications,* vol. 13.4, pp. 18-28, 1998.

[42] V. Hax, N. Filho, S. Botelho e O. Mendizabal, "ROS as a middleware to Internet Of Things," *Journal of Applied Computing Research ,* vol. 2.2, pp. 91-97, 2013.

[43] B. Gary e A. Kaehler, Learning OpenCV: Computer vision with the OpenCV library, O'Reilly Media, 2008.

[44] G. Dudek e M. Jenkin, Computational principles of mobile robotics, Cambridge university: Cambridge University Press, 2010.

[45] D. Ionescu, V. Suse, C. Gadea, B. Solomon, B. Ionescu e S. Islam, "A new infrared 3D camera for Gesture Control," em *Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International*, Minneapolis, USA, IEEE, May 2013, p. 629–634.

[46] S. Winkvist, "This mobile robot is equipped with a LIDAR sensor, allowing it to map the surrounding area and avoid obstacles," wikimedia commons, under the public domain license, 3/3/2008.

[47] G. KIM, J. Eom, S. Park e Y. Park, "Occurrence and characteristics of mutual interference between LIDAR scanners," *SPIE Optics+ Optoelectronics. International Society for Optics and Photonics,* pp. 95040K-1, 2015.

[48] Schorsch, "3D-laser-scanner mounted on a tripod," Wikimedia Commons, under the Creative Commons Attribution-Share Alike 3.0 Unported license, 6/12/2006.

[49] B. Wong, B. Morrissey e Z. Li, "Phased array pulsing : A simple digital sonar system," MIT, 2007.

[50] C. Silva, Sensors and Actuators: Engineering System Instrumentation, CRC Press, 2015.

[51] Bpducharme, "Senix ToughSonic 30 Ultrasonic Sensor," Wikimedia commons, under the Creative Commons Attribution-Share Alike 4.0 International license, 10/05/2015.

[52] A. Örstadius e E. Ali, "Passive Radar Detection of Aerial Targets," 2017.

[53] P. Smith, Active sensors for local planning in mobile robotics, vol. 26, World Scientific, 2001.

[54] N. Inductiveload, "EM Spectrum Properties," wikimidia Commons, under the Creative Commons Attribution-Share Alike 3.0 Unported license, 26/10/2007.

[55] M. Yasuno, N. Yasuda e M. Aoki, "Pedestrian detection and tracking in far infrared images," em *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*, Washington, USA, IEEE, June 2004, pp. 125-125.

[56] B. H. T. Imaging, "A thermal Image of a home in the Black Hills of South dakota," Wikimedia Commons, under the Creative Commons Attribution-Share Alike 3.0 Unported license, 27/11/2013.

[57] J. Gall, H. Grabner, X. Ren, K. Konolige e A. Fossati, Consumer depth cameras for computer vision: research topics and applications, Springer Science & Business Media, 2012.

[58] T. Teixeira, G. Dublon e A. Savvides, "A survey of human-sensing: Methods for detecting presence, count, location, track, and identity," *ACM Computing Surveys,* vol. 5.1, pp. 59-69, 2010.

[59] B. Siciliano e O. Khatib, Springer Handbook of Robotics, Springer, 2016.

[60] Bibliovision3D, "Principe de la stéréovision," Wikimedia commons, under the Creative Commons Attribution-Share Alike 4.0 International license., 2/12/2016.

[61] Vanessaezekowitz, "Structured-light 3D scanner," Wikimedia Commons, under the Creative Commons Attribution 3.0 Unported license, 5/08/2008.

[62] J. Geng, "Structured-light 3D surface imaging," *Advances in Optics and Photonics,* vol. 3.2, pp. 128-160, 2011.

[63] F. Remondino e D. Stoppa, TOF Range-Imaging Cameras, Springer, 2013.

[64] R. Szeliski, Computer Vision Algorithms and Applications, Springer Science & Business Media, 2011.

[65] M. Thoma, "Beispiel-Szene zur Semantischen Segmentierung," Creative commons, under the Creative Commons CC0 1.0 Universal Public Domain Dedication, 8/1/2016.

[66] S. Gupta e S. Mazumdar, "Sobel edge detection algorithm," *International journal of computer science and management Research,* vol. 2.2, pp. 1578-1583, 2013.

[67] Robodoc9, "Robotino with Sick laser range finder," Creative Commons, under the Creative Commons Attribution-Share Alike 3.0 Unported license, 24/10/2008.

[68] M. Fischler e R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM,* vol. 24.6, pp. 381-395, 1981.

[69] G. Deng e L. Cahill, "An adaptive Gaussian filter for noise reduction and edge detection," em *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record*, San Francisco, USA, IEEE, November 1993, pp. 1615-1619.

[70] L. Di Stefano e A. Bulgarelli, “A simple and efficient connected components labeling algorithm,” em *Image Analysis and Processing, 1999. Proceedings. International Conference on*, Venice, Italy, IEEE, September 1999, pp. 322-327.

[71] R. Duda e P. Hart, “Use of the Hough transformation to detect lines and curves in pictures,” *Communications of the ACM,* vol. 15.1, pp. 11-15, 1972.

[72] S. Savarese, “A visual representation of the variables used in image rectification example.,” Wikimedia commons, 11/12/2014.

[73] Indif, “Feature prototypes of simple Haar-like and center-surround features,” Creative Commons, under the Creative Commons Attribution-Share Alike 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license., 19/09/2010.