

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



# **Sistema Inteligente de Detecção de Pessoas para Robôs Móveis Autónomos de Desinfeção**

**Hugo Lima Mendonça**

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Orientador: Professor Doutor Paulo Costa

Co-orientador: Professor Doutor José Lima

31 de julho de 2021



# Resumo

A pandemia de COVID-19, provocada pela propagação descontrolada do vírus SARS-CoV-2, levou à necessidade de desenvolver sistemas de desinfecção inteligentes, não só para proteger as pessoas que frequentam espaços públicos, mas também para proteger aqueles que estejam sujeitos a situações de risco em zonas contaminadas.

Nesta dissertação, é desenvolvido um sistema de detecção de pessoas inteligente para robôs autônomos de desinfecção que utilizam radiação Ultravioleta do tipo C para esse propósito. A desinfecção é interrompida pelo sistema quando uma pessoa é detetada à volta do robô, em qualquer direção. A radiação UVC é perigosa para humanos e, como tal, o sistema de detecção proposto tem como objetivo proteger as pessoas, desativando o processo de desinfecção assim que uma pessoa seja detetada. O sistema utiliza três tipos de sensores para a detecção, uma câmera RGB Raspberry Pi, uma câmera térmica FLIR Lepton 3.5 e quatro sensores PIR, fundindo posteriormente os dados por eles recolhidos. As imagens RGB e térmicas são processadas com recurso a uma rede neuronal Single Shot Detector Mobilenet, para identificar e detetar pessoas com recurso a métodos de inteligência artificial. A câmera térmica é ainda utilizada para gerar um histograma de temperaturas para cada imagem processada, que contribui para o algoritmo de detecção desenvolvido.

Os resultados mostram que a fusão sensorial dos diferentes sensores melhora o desempenho do sistema, quando comparado a situações em que os sensores são utilizados individualmente. Um dos testes realizados mostra a capacidade do sistema em distinguir pessoas em fotografias de pessoas reais, através da fusão dos dados da câmera térmica com os dados da câmera RGB. Os resultados da detecção validam o sistema proposto.

O sistema tem ainda dois sensores extra que comunicam diretamente com o robô de desinfecção que são um LiDAR para cálculo de área na direção da detecção e um sonar para medir a distância do robô ao teto.



# Abstract

The COVID-19 virus outbreak led to the need of developing smart disinfection systems, not only to protect the people that usually frequent public spaces but also to protect those who have to subject themselves to the contaminated areas.

In this dissertation it is developed a human detector smart sensor for autonomous disinfection mobile robots that uses Ultraviolet C type light for the disinfection task. The disinfection is interrupted by the system when a human is detected around the robot in any direction. UVC light is dangerous for humans and thus the need for a human detection system that will protect them by disabling the disinfection process, as soon as a person is detected. This system uses different sensors, a RGB Raspberry Pi Camera as well as a FLIR Lepton 3.5 Thermal Camera and also four PIR sensors and combines the data gathered by them. The RGB and Thermal images are processed with a Single Shot Detector Mobilenet neural network to identify and detect persons. The thermal camera also generates a temperature histogram for each processed image.

Results show that the sensorial fusion of the different sensors data improves the system performance compared to when the sensors are used individually. One of the tests performed proves that the system is able to distinguish a person in a picture from a real person by fusing the thermal camera and the visible light camera data. The detection results validate the proposed system.

The system also possesses two extra sensors that communicate directly with the disinfection robot, a LiDAR to measure the area in the detection direction and a sonar to measure the distance from the robot to the ceiling.



# Agradecimentos

Gostaria de expressar a minha gratidão aos meus orientadores Dr. Paulo Costa e Dr. José Lima, por todo o apoio, confiança, disponibilidade e orientação que me deram e sem a qual não teria sido possível a realização desta tese.

A todos os membros da equipa do CRIIS, com quem tive o prazer de trabalhar e que sempre se mostraram disponíveis para me ajudar e com os quais desenvolvi uma relação de amizade.

À minha família, especialmente os meus pais e à minha irmã, por todo o amor e apoio que me deram, os quais tornaram possível a minha jornada de sucesso ao longo do meu percurso académico. Sem eles nada disto teria sido possível.

A todos os meus amigos, pelos momentos vividos e por todo o companheirismo durante este período de tempo.

Com tudo isto, estes últimos anos ajudaram-me não só a crescer e a tornar-me na pessoa que sou hoje, mas também a moldar a maneira como vou encarar os desafios que me aguardam no futuro.

Obrigado,

Hugo Lima Mendonça



*“The Pessimist Sees Difficulty In Every Opportunity.  
The Optimist Sees Opportunity In Every Difficulty.”*

Winston Churchill



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Enquadramento . . . . .	1
1.2	Motivação . . . . .	2
1.3	Objetivos . . . . .	2
1.4	Estrutura da Dissertação . . . . .	3
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>5</b>
2.1	Caraterização do Problema . . . . .	5
2.2	Descrição do Problema . . . . .	5
2.3	Sistemas de Detecção . . . . .	6
2.3.1	Sensores PIR . . . . .	7
2.3.2	Sensores Térmicos . . . . .	10
2.3.3	Sensores de Visão RGB . . . . .	11
2.4	Fusão Sensorial . . . . .	16
<b>3</b>	<b>Arquitetura Proposta</b>	<b>21</b>
3.1	Hardware . . . . .	21
3.2	Software . . . . .	23
<b>4</b>	<b>Implementação</b>	<b>27</b>
4.1	Hardware e Mecânica . . . . .	27
4.2	Software e Protocolos . . . . .	33
4.2.1	Câmara Térmica . . . . .	33
4.2.2	Rede Neuronal . . . . .	36
4.2.3	Microcontrolador . . . . .	37
4.2.4	Comunicação Assíncrona . . . . .	40
<b>5</b>	<b>Resultados e Discussão</b>	<b>43</b>
<b>6</b>	<b>Conclusões e Trabalho Futuro</b>	<b>55</b>
<b>A</b>	<b>Esquema elétrico do PCB</b>	<b>57</b>
	<b>Referências</b>	<b>61</b>



# Lista de Figuras

2.1	Esquema da descrição do problema . . . . .	6
2.2	Identificação de direção, velocidade e distância de movimento em [1] . . . . .	7
2.3	Disposição do sistema proposto em [2] . . . . .	8
2.4	Influência da temperatura na sensibilidade do sensor em [3] . . . . .	9
2.5	Resultados experimentais da detecção obtidos em [4] . . . . .	12
2.6	Resultados experimentais da odometria obtidos em [5] . . . . .	17
2.7	Exemplo do mapa de ocupação gerado em [6] . . . . .	18
3.1	Arquitetura principal do sistema . . . . .	22
3.2	Arquitetura do módulo rotativo 360° . . . . .	23
3.3	Arquitetura do software do sistema . . . . .	25
4.1	Peças que constituem a base da caixa superior . . . . .	28
4.2	Peças que constituem o topo da caixa superior . . . . .	29
4.3	Peças que constituem a caixa superior . . . . .	29
4.4	Resultado final da montagem elétrica do módulo superior . . . . .	30
4.5	Resultado final da montagem elétrica do módulo inferior . . . . .	30
4.6	Mockup do PCB desenvolvido . . . . .	31
4.7	Resultado final do PCB desenvolvido . . . . .	32
4.8	Resultado final da caixa do sensor de detecção de pessoas . . . . .	32
4.9	Exemplos de transmissão por VoSPI a diferentes velocidades . . . . .	34
4.10	Exemplo de transmissão SPI Mode 3 do valor 0x8C08 . . . . .	34
4.11	Organização dos segmentos e respectivos pacotes num frame, por defeito . . . . .	35
4.12	Exemplo de dessincronização . . . . .	36
4.13	Exemplo de imagem processada após detecção . . . . .	37
4.14	Diagrama de estados representativo do mecanismo de rotação . . . . .	38
4.15	Estrutura de decisão do sistema de detecção de pessoas . . . . .	40
4.16	Arquitetura de comunicação da framework ZeroMQ no sistema . . . . .	42
5.1	Exemplo das dez capturas de imagem numa rotação na câmera térmica . . . . .	44
5.2	Exemplo das dez capturas de imagem numa rotação na câmera RGB . . . . .	45
5.3	Falso Negativo na detecção com imagem térmica . . . . .	46
5.4	Falso Positivo na detecção com histograma da imagem térmica . . . . .	46
5.5	Falso Positivo na detecção com imagem RGB com um objeto semelhante a uma pessoa . . . . .	47
5.6	Falso Positivo em imagem RGB com uma fotografia de uma pessoa não presente . . . . .	48
5.7	Falso Negativo na detecção com imagem RGB em baixa luminosidade . . . . .	49
5.8	Falso Negativo na detecção com imagem RGB com uma pessoa fora do alcance do sistema . . . . .	50

5.9	Verdadeiro Positivo na deteção em imagem RGB com confirmação térmica . . .	51
5.10	Falso Positivo na deteção em imagem RGB com confirmação térmica . . . . .	52
5.11	Deteção positiva por confirmação com imagem RGB e térmica . . . . .	53
5.12	Deteção positiva por confirmação com imagem térmica . . . . .	53

# Lista de Tabelas

5.1	Resumo das vantagens e desvantagens dos métodos de detecção . . . . .	49
-----	---	----



# Abreviaturas e Símbolos

AGV	Automated Guided Vehicles
CNN	Convolutional Neural Network
CRC	Cyclic Redundancy Check
CRIIS	Centre for Robotics in Industry and Intelligent Systems
CS	Chip Select
FEUP	Faculdade de Engenharia da Universidade do Porto
I <sup>2</sup> C	Inter-Integrated Circuit
ID	Identity
IP	Internet Protocol
MISO	Master Input Slave Output
MOSI	Master Output Slave Input
INESC TEC	Institute for Systems and Computer Engineering, Technology and Science
PCB	Printed Circuit Board
PIR	Passive Infrared Sensor
R-CNN	Region Based Convolutional Neural Networks
SCK	Serial Clock
SSD	Single Shot Detector
TWI	Two Wire Interface
V	Volts
VoSPI	Video over Serial Peripheral Interface
YOLO	You Only Look Once



# Capítulo 1

## Introdução

Este capítulo está dividido em três secções. A primeira secção consiste num enquadramento do contexto em que a dissertação será desenvolvida assim como o seu tema. De seguida é referida a motivação por trás do desenvolvimento deste trabalho. E, finalmente, são expostos os objetivos que se pretende atingir com o desenvolvimento desta dissertação.

### 1.1 Enquadramento

A situação pandémica em que se vive, tem levado à necessidade de alocar grandes números de pessoas em espaços públicos que, muitas vezes, não possuem a capacidade para garantir condições de segurança, relativamente à exposição das pessoas que circulam nesses espaços ao vírus. Assim sendo, surge a necessidade de garantir que, quartos e zonas comuns de centros hospitalares, estejam constantemente a ser desinfetados, de modo a garantir que quem circule nestas áreas possa fazê-lo em segurança. A desinfeção dos espaços garante não só a segurança dos utentes, mas também de médicos e auxiliares de saúde cujo desempenho da sua função implica estar diariamente sujeito a frequentar ambientes perigosos, sob risco de estar exposto ao vírus. Para além da variante médica, surge ainda a necessidade de fazer a desinfeção de outras zonas públicas, que sejam frequentadas por grandes quantidades de pessoas de forma regular. Os supermercados, aeroportos e transportes públicos são alguns exemplos desses mesmos espaços.

Surge então a necessidade de procurar alternativas seguras e eficazes na eliminação do vírus nestes ambientes. A alternativa mais comum passa pela simples desinfeção, através de produtos químicos que sejam destinados para esse efeito. Contudo, cada vez mais se tem procurado desenvolver outros sistemas que permitam a desinfeção de áreas expostas ao vírus, sem necessidade de recorrer ao envolvimento de produtos químicos. Estes produtos para além de não serem totalmente eficazes, requerem ainda a ação direta de seres humanos para realizar a desinfeção expondo-os também ao vírus. Isto leva a que, cada vez mais, se procurem alternativas que não envolvam fisicamente a participação humana. A desinfeção com recurso a AGVs tem sido utilizada como uma alternativa segura e eficiente no que toca a combater o vírus em zonas fechadas e de relativamente fácil circulação.

Um dos possíveis métodos de desinfecção passa pela utilização de radiação ultravioleta. Quando se expõem zonas contaminadas, durante um certo intervalo de tempo a esta radiação, é possível verificar que, a carga viral nelas presente passa a ser praticamente nula. Pode-se então afirmar que, a desinfecção das superfícies contaminadas é atingida com níveis muito mais eficazes, quando comparado a métodos tradicionais, como a utilização de produtos químicos. Esta abordagem tem ainda a vantagem de não ser necessário recorrer à operação direta de pessoas no processo, podendo ser explorada a utilização de AGVs para o efeito. Contudo, é de realçar que, apesar de ser um método eficaz, apresenta também outro tipo de inconvenientes, nomeadamente os perigos que a exposição à radiação ultravioleta apresenta para organismos vivos, como o corpo humano. Assim, surge a necessidade de encontrar um método que possibilite a utilização deste processo de desinfecção com radiação ultravioleta protegendo humanos que se encontrem numa zona próxima.

## 1.2 Motivação

Com o objetivo de proteger seres humanos que se encontrem próximos da zona de atuação de sistemas de desinfecção por radiação ultravioleta, é necessário desenvolver um sistema que seja capaz de identificar a presença dos mesmos, de forma a interromper a operação de desinfecção. Para otimizar o processo e torná-lo mais robusto, estes sistemas deverão ser capazes de atuar em áreas onde haja movimentações de diferentes tipos, como pessoas, animais, máquinas e robôs, e conseguir diferenciar aquilo que são humanos daquilo que não são.

A simples utilização de um sensor de presença, não permitiria um conhecimento completo e robusto do ambiente em análise. A cooperação de vários AGVs num mesmo espaço seria também comprometida, uma vez que o movimento de um faria com que outros deixassem de atuar. Como tal, nesta tese será desenvolvido um sistema inteligente de deteção de pessoas em ambientes perigosos que, irá combinar a utilização de múltiplos sensores, através da fusão sensorial dos seus dados, de forma a fazer uma deteção precisa da presença de humanos, de forma a protegê-los.

## 1.3 Objetivos

Dentro do contexto referido, o trabalho apresentado nesta dissertação tem como objectivo geral a implementação e desenvolvimento de um sistema de deteção de presença de humanos para robôs autónomos de desinfecção que atuem em ambientes perigosos. Tem também, como base, o estudo de diferentes tipos de sensores de baixo custo, que serão posteriormente utilizados num único sistema que tirará partido da fusão sensorial dos mesmos, para reportar a eventual presença de um ser humano no espaço onde estiver a ser utilizado. O objetivo é prevenir a exposição direta das pessoas a métodos de desinfecção perigosos, como a radiação ultravioleta, que é nociva para o ser humano.

Para que a dissertação seja concluída com sucesso foram estabelecidos vários objectivos específicos:

1. Implementação do sistema de detecção de pessoas com base nos sensores de baixo custo estudados e escolhidos
  - Após o estudo dos diferentes tipos de métodos de sensorização, é possível ter uma ideia mais concreta da forma mais eficiente de desenvolver o sistema, otimizando-o em termos de relação custo-valor
  - O sistema desenvolvido tem de conseguir passar o máximo de informações úteis relativamente ao ambiente que o envolve de maneira a essa informação ser tratada com o objetivo de confirmar a presença ou não de pessoas
  - O pré-processamento dos dados de cada sensor deve ser feito de forma individual uns dos outros
2. Desenvolvimento de um programa de fusão sensorial que será processado por um sistema embebido
  - A fusão sensorial tem de permite combinar as informações obtidas pelos diferentes sensores do sistema de maneira a verificar a existência ou não de pessoas
  - O sistema deve ser capaz de detetar pessoas num ambiente fechado
  - O sistema deve saber distinguir a presença de pessoas da presença de outros objetos como robôs, por exemplo
  - O sistema deve conseguir informar o utilizador, assim como os sistemas onde está implementado das possíveis deteções de pessoas que são realizadas
3. Desenvolvimento de uma interface de visualização dos dados obtidos pelos diferentes sensores
  - A interface tem de permitir uma interação mais prática com o sistema permitindo visualizar o comportamento do sistema como um todo
  - A interface tem de permitir uma visualização do estado individual de cada um dos sensores a serem utilizados
  - A interface deve permitir a configuração do sensor por parte do utilizador

## 1.4 Estrutura da Dissertação

Para além da introdução, esta dissertação contém mais 5 capítulos. No capítulo 2, é descrito o estado da arte e são apresentados trabalhos relacionados. No capítulo 3, é apresentada a arquitetura do sistema a desenvolver em termos de hardware e software. No capítulo 4, é explicada a implementação do sistema com base na arquitetura apresentada no capítulo anterior. No capítulo 5, são apresentados os resultados obtidos pelo sistema desenvolvido. Finalmente, no capítulo 6, retiram-se as conclusões após a realização desta dissertação, assim como o trabalho a realizar no futuro.



## Capítulo 2

# Revisão Bibliográfica

Este capítulo está dividido em quatro secções. A primeira e a segunda secção servem para fazer uma breve caracterização e descrição, respetivamente do problema que se pretende resolver com o desenvolvimento desta dissertação. A terceira e a quarta secção consistem numa revisão bibliográfica dos assuntos principais desta dissertação, que são eles Sistemas de Detecção e Fusão Sensorial.

### 2.1 Caracterização do Problema

O surgimento da pandemia no último ano levou a um aumento da concentração de pessoas em ambientes públicos como hospitais, aeroportos, lares, supermercados, entre outros. É também conhecido que, todos os espaços sujeitos à criação de ajuntamentos de grandes grupos de pessoas, constituem um risco elevado para a saúde pública e, como tal, é necessário desenvolver formas de neutralizar o vírus. A principal forma de realizar esta tarefa, passa pela desinfeção desses locais. A desinfeção pode ser realizada de diferentes formas, quer seja através da utilização de produtos químicos, quer seja através de métodos alternativos como a utilização de radiação ultravioleta. Este último método é aquele que é tido em conta, durante o desenvolvimento desta dissertação.

### 2.2 Descrição do Problema

É possível distinguir um conjunto de causas e consequências resultantes da utilização de radiação ultravioleta para a desinfeção de espaços fechados e para os quais se pretende oferecer uma solução, de maneira a tornar a sua utilização o mais segura possível, para as pessoas que frequentam os ditos espaços.

A nível de causas podem ser distinguidos os fatores que levam à necessidade de utilizar os métodos de desinfeção de uma forma geral, mas também aqueles que levam à necessidade de aplicação do método em foco, que é a radiação ultravioleta:

- Existência do vírus em ambientes fechados frequentados por muitas pessoas

- Facilidade de propagação do vírus em ambientes fechados
- Resistência do vírus a determinados métodos de desinfecção, como a utilização de sprays desinfetantes

Além disso, é possível ainda destacar as consequências que este problema tem a nível de impacto na sociedade e na saúde pública, nomeadamente:

- Problemas de saúde em grupos de risco, nomeadamente problemas respiratórios
- Desinfecção dos espaços através de métodos alternativos, pode ser nociva para o corpo humano, quando expostos diretamente

O seguinte esquema ajuda a compreender a relação entre os pontos anteriormente mencionados, no que diz respeito ao problema e às suas causas e consequências, à esquerda e à direita, respetivamente.

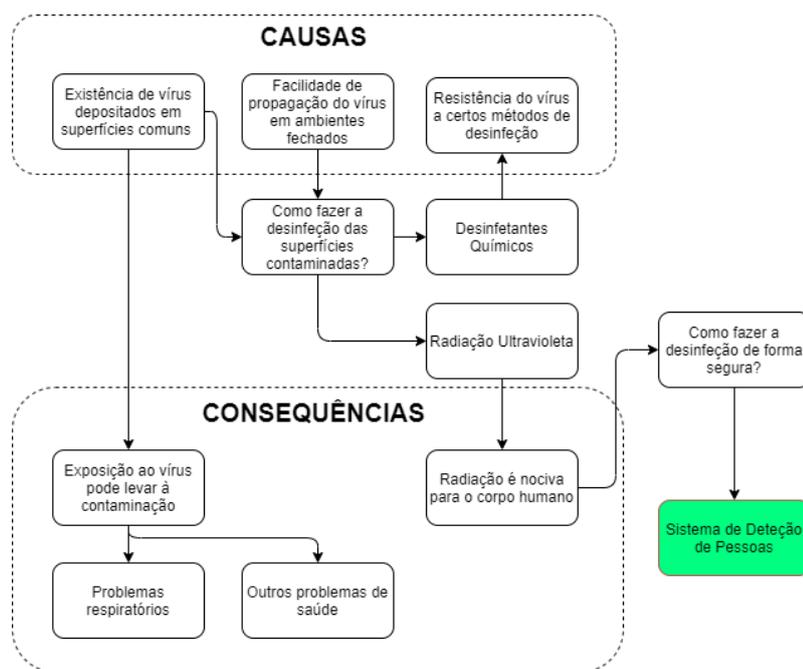


Figura 2.1: Esquema da descrição do problema

## 2.3 Sistemas de Detecção

Os sistemas de deteção escolhidos tiveram por base um conhecimento prévio das necessidades do sistema, assim como os requisitos do sistema a nível de desempenho e funcionalidades. A divisão é feita em três subsecções, uma primeira onde é feito o estudo de uma família de sensores

passivos, muitas vezes utilizados em sistemas de deteção de movimento. A segunda subsecção explora aplicações onde sejam utilizados sensores de visão baseados em câmeras RGB, com um foco especial na utilização de redes neuronais para deteção de objetos. Finalmente, a última subsecção centra-se na utilização de sensores térmicos baseados em câmeras térmicas para realizar a deteção e identificação de pessoas, motivado por situações onde os métodos anteriores não sejam tão eficazes, como em cenários de baixa luminosidade, por exemplo, onde os outros métodos de deteção não terão um desempenho ideal.

### 2.3.1 Sensores PIR

São uma família de sensores muito utilizados nas áreas de deteção de presença e movimento, estando presentes em muitos sistemas de alarme por constituírem uma solução eficiente e de baixo custo. É ainda importante notar que estes sensores não só são utilizados como trigger de presença, como também se pode tirar partido da sua saída analógica, que depende de outros aspetos, como a distância do corpo ao sensor, da direção e velocidade do movimento, do formato do corpo e da maneira de andar [1].

Em [1] é descrito um método de deteção e identificação do movimento de pessoas, através do desenvolvimento de um módulo de recolha de dados que consiste em dois pares de sensores PIR alinhados ortogonalmente como na Fig. 2.2 e lentes de Fresnel, uma bateria recarregável, um data logger e circuitos op-amp. Neste artigo é realizada uma experiência com três destes módulos, dispondo-os em diferentes posições ao longo de um corredor, um no teto e os outros dois em duas paredes opostas viradas uma para a outra. Verifica-se que os resultados obtidos foram bastante positivos tanto em termos de identificação, como em termos de deteção de movimento, tendo sido possível distinguir a velocidade e direção dos movimentos.

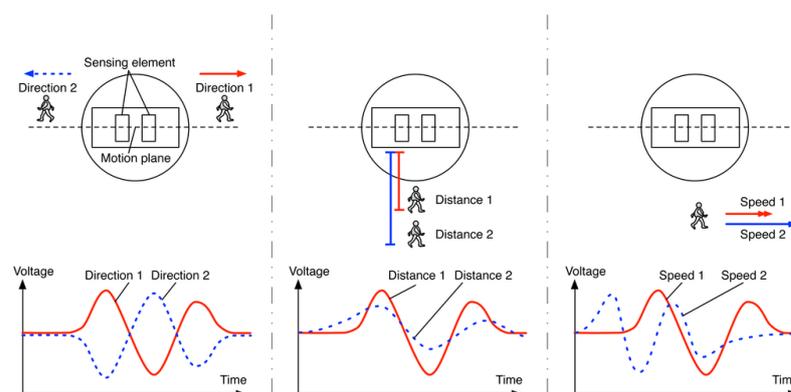


Figura 2.2: Identificação de direção, velocidade e distância de movimento em [1]

A metodologia apresentada em [7], explora a deteção e localização de alvos em movimento com uma rede de sensores PIR, em ambientes interiores e exteriores como na Fig. 2.3. O algoritmo de localização proposto utiliza um *threshold* adaptativo, com uma razão de falsos alarmes

constante e consiste em 3 fases: eliminação de ruído, computação da energia do sinal e detecção do rácio de falsos alarmes. É ainda desenvolvido um algoritmo de localização baseado em *direction search* na grelha do espaço da rede. Os resultados experimentais obtidos permitem detetar e localizar com alguma precisão os alvos em movimento, como proposto.

Com o crescimento da indústria 4.0, cada vez existe uma maior variedade de sensores que são usados em *smart homes*, um dos quais são os sensores PIR. Em [2] é desenvolvido um sistema de localização de pessoas, numa habitação onde os sensores PIR desempenham um papel crucial para o efeito. São estudadas duas abordagens, uma baseada em terminais, isto significa que a tecnologia é transportada pelo utilizador. E outra não baseada em terminais onde, pelo contrário, não é necessária a utilização de qualquer tipo de dispositivo. O foco deste artigo é a segunda abordagem mencionada. O sistema é desenvolvido com sensores PIR tendo em conta quatro requisitos: ser relativamente barato, por se ter que utilizar vários sensores; ser flexível, devido à diferença entre as diferentes divisões de uma casa; ser robusto a ruído e ao ambiente envolvente, uma vez que podem coexistir com outros sensores que produzem ruído eletromagnético ou devido a variações das condições ambientais. Relativamente à distribuição dos sensores, estes deverão formar uma matriz e as suas áreas de deteção devem estar parcialmente sobrepostas. Com base nos testes realizados em diversas condições diferentes, verifica-se que o sistema tem um desempenho relativamente aceitável. É de destacar que o desempenho do sistema está associado à disposição espacial dos sensores e, portanto, é necessário otimizar a localização dos sensores. Para além disso, o método de localização está limitado à presença de uma pessoa por divisão.

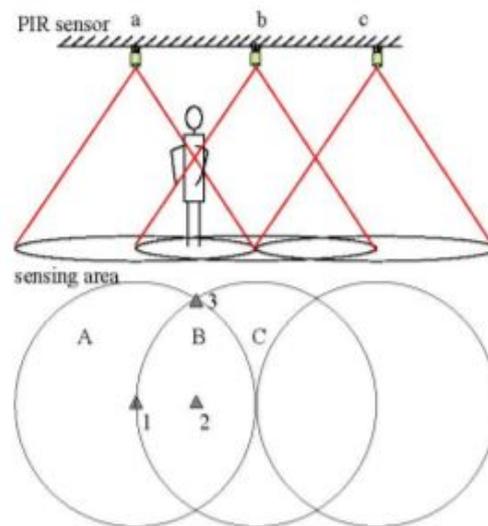


Figura 2.3: Disposição do sistema proposto em [2]

Uma outra aplicação que muitas vezes tira partido da tecnologia utilizada neste tipo de sensores, é na segurança e vigilância de propriedades privadas. Em [8], é desenvolvido um sistema de vigilância de baixo custo, que utiliza sensores PIR para realizar a deteção de movimento. As suas dimensões discretas e o seu funcionamento independente da luminosidade, constituem uma van-

tagem de utilização quando comparados a outros métodos de vigilância. O sistema desenvolvido consiste num sensor PIR, num microcontrolador ESP8266 que possui capacidade de comunicação Wifi e um GSM (Global System for Mobile communication) para enviar mensagens de texto com alertas para o utilizador. Conclui-se que a solução faz a deteção com sucesso de pessoas, animais ou outros objetos e tem custos baixos de implementação.

A utilização de sensores PIR em sistemas de segurança, está ainda associada à sua grande sensibilidade e área de deteção. O output do sensor é influenciado pela relação temporal entre o objeto no campo de visão do sensor, a sensibilidade do sensor, as características da lente do sensor e a temperatura ambiente [3]. Conclui-se que a sensibilidade do sensor reduz com o aumento da temperatura como se vê no gráfico da Fig. 2.4 e ainda que a sensibilidade e resolução do sensor reduz com o aumento da distância. É também referido que os sensores PIR não só são usados para detetar movimento, como também para deteção de temperatura de espaços remotos, com algumas limitações. Isto deve-se ao facto de o seu método de atuação ser baseado na deteção de radiação infravermelha, radiação essa que é emitida por todos os objetos e seres vivos.

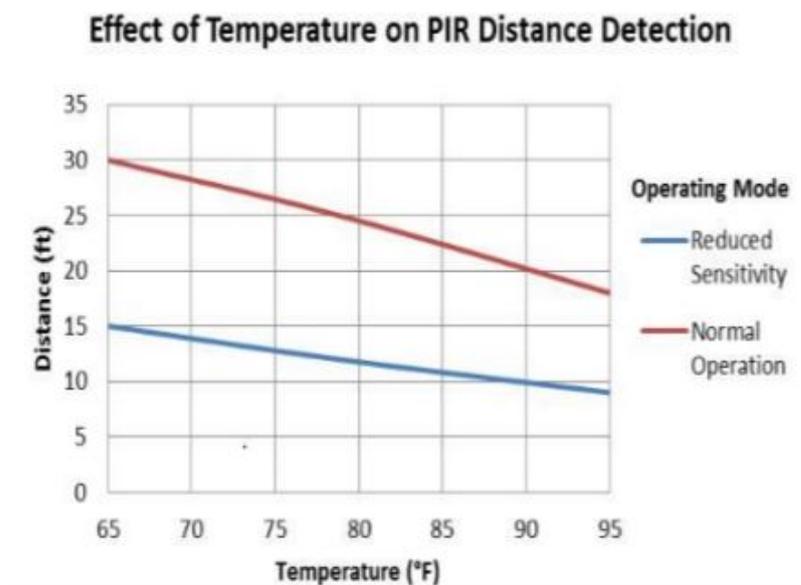


Figura 2.4: Influência da temperatura na sensibilidade do sensor em [3]

Como alternativa ao uso de câmeras de circuito fechado para vigilância por serem demasiado dispendiosas e as respetivas gravações ocuparem demasiado espaço em memória, em [9] é desenvolvida uma solução de baixo custo e baixo consumo. O sistema consiste num Raspberry Pi, para operações de processamento e é usado juntamente com os sensores PIR para fazer a deteção de movimentos. Esta solução é particularmente interessante para vigilância de espaços com poucas pessoas, como, por exemplo, cofres de bancos. Quando um movimento é detetado pelo sistema, é utilizada uma câmara para capturar uma imagem do local e é armazenada temporariamente no Raspberry Pi. Conclui-se que o sistema é desenvolvido para situações particulares, como vigilância de habitações e locais de trabalho de forma remota e consulta do interior de cofres e armazéns.

### 2.3.2 Sensores Térmicos

A detecção de humanos é particularmente eficaz através da utilização de sensores térmicos, uma vez que utilizam o conceito de emissão da radiação infravermelha dos objetos. Fazendo a associação ao espectro de temperaturas conhecidas para o ser humano, estes sensores conseguem tirar partido dessa característica, para fazer a detecção de pessoas até mesmo em ambientes escuros, ao contrário de sistemas de visão convencionais.

Tal como se refere em [10], este tipo de sensores são geralmente pequenos, discretos e baratos. Aqui é proposto um método de detecção de humanos em ambientes fechados, utilizando uma imagem térmica de baixa resolução. Os dados obtidos são utilizados para fazer a monitorização de idosos, enquanto realizam atividades do seu dia-a-dia em casa e detetar situações críticas como quedas. Para o processamento dos dados, é ainda tido em conta as variações de temperatura ambiente, através da utilização de uma estimação do fundo da imagem e remoção de ruído com um filtro de Kalman. O sistema utiliza um único sensor montado no teto de uma divisão, no entanto, são apontados alguns problemas, tais como a obstrução do campo de visão do sensor por parte da mobília dos quartos e outros objetos, o movimento das pessoas em relação à câmara alterar o número de pixels associados a ela e, também, a distribuição da temperatura. Conclui-se que o método apresentado tem uma precisão de 97%, apresentando a limitação de não conseguir distinguir emissões de calor humanas de outras emissões não humanas semelhantes.

Os sensores térmicos apresentam claras vantagens, em situações de baixa luminosidade, quando comparados a sistemas de visão RGB. No entanto, estes apresentam também algumas limitações, no que diz respeito a situações onde a diferença de temperatura entre as pessoas e o ambiente não são significativas. Em [11], é proposto um método de detecção de pessoas que visa superar estas adversidades, através da utilização de imagens obtidas com uma câmara térmica. A diferença binária entre a imagem de entrada e a imagem de fundo pode ser adaptada com base em sistemas difusos, com o uso de informação proveniente do fundo da imagem e a diferença entre os valores de fundo e de entrada. Com este método, é possível detetar formas humanas, independentemente das diversas condições de detecção da imagem. Conclui-se que, este método é capaz de realizar a identificação de pessoas, com recurso a operações de processamento da imagem detetada, gerando a imagem de fundo, através da mediana da imagem original e, de seguida, removendo esse fundo à imagem de entrada. A imagem resultante é então segmentada em regiões de interesse, com base na informação do histograma da imagem e nas proporções conhecidas do corpo humano. Os resultados obtidos em diversas condições, permitem validar a solução proposta como sendo precisa e eficaz.

Os algoritmos de detecção poderão ainda recorrer à combinação de sensores térmicos com outro tipo de sensores. Em [12], é apresentado um sistema que permite a detecção e identificação de pessoas por robôs em ambientes domésticos, através da fusão de dados provenientes de sensores térmicos e de visão. Estes dados são processados de forma a identificar pessoas, com base na identificação dos seus rostos. Dado que a identificação de pessoas em ambientes domésticos pode ser complicada, quando são utilizados sistemas de visão, devido às condições variáveis de luminosidade

dade e também dadas as diferentes poses do corpo humano, a combinação com sensores térmicos permite uma deteção mais robusta, independentemente das condições da deteção. Para além disso, este método de deteção consegue um alcance relativamente considerável, o que o torna eficiente em aplicações em espaços reduzidos. A deteção dos rostos também é possível pela análise do espectro térmico, daí ter sido aplicado neste método. O sistema é composto por cinco módulos distintos: de deteção da pele humana, de deteção do corpo humano, de tomada de decisão relativamente à posição das caras das pessoas, de deteção de caras e de reconhecimento das caras. Os resultados obtidos permitem concluir que esta abordagem de deteção é capaz de realizar a identificação e deteção de pessoas em ambientes domésticos, em diferentes condições de iluminação e ângulos de visualização, até uma distância de 6 metros. A combinação dos dois sistemas de deteção permite uma maior precisão do método, uma vez que o sistema de visão permite reduzir falsos positivos do sistema térmico e este permite a redução de ruídos de fundo, que perturbem a análise do sistema de visão.

Ainda relativamente a aplicações de deteção ativa de pessoas utilizando câmeras/imagem, sabe-se que a utilização de sensores térmicos é menos indicativa, no que diz respeito a informação visual, tal como cor, detalhe ou definição. Isto deve-se ao facto da sua função ser simplesmente detetar temperaturas, baseando-se no mesmo princípio dos sensores infravermelhos. Contudo, como é descrito em [13] e [14], em situações de baixa luminosidade, este tipo de câmara ainda consegue obter resultados de deteção mais precisos quando comparado a câmeras RGB, sem sensores infravermelhos. Em [15], é feita uma comparação entre duas câmeras, uma com e outra sem sensores infravermelhos e os resultados revelam que a primeira supera o desempenho da segunda. Isto leva-nos a acreditar que, em situações de baixa luminosidade, as câmeras térmicas são mais informativas que as câmeras de luz visível.

Dado o facto das câmeras térmicas serem relativamente caras, é necessário encontrar uma que vá de encontro aos nossos requisitos sem comprometer o orçamento disponível. A FLIR apresenta uma grande variedade de soluções, da qual se destaca a família FLIR Lepton que representa uma opção de custo relativamente baixo, como é referido em [16]. Esta câmara tem a capacidade de detetar temperaturas específicas em objetos mais pequenos que pessoas, a longas distâncias, apresentando ainda um design compacto, leve e compatível com sistemas embebidos como o Raspberry Pi. Este modelo tem duas versões 2.x e 3.x onde, entre outras diferenças, se destaca a diferença de resolução entre as duas, apresentando a primeira uma resolução de 80 x 60 pixels e a segunda uma resolução de 160 x 120 pixels.

### 2.3.3 Sensores de Visão RGB

Para além dos métodos anteriormente indicados, existem ainda maneiras mais sofisticadas de fazer a deteção de pessoas, em determinados ambientes, embora sejam soluções que acarretam custos operacionais muito mais elevados e são computacionalmente mais pesados. Um desses métodos é a utilização de sistemas de visão RGB como em [17], onde é apresentado um sistema de deteção e rastreamento de pessoas em ambientes fechados, usando uma câmara estática. A solução apresentada, combina o reconhecimento de objetos em imagens paradas, com métodos que

utilizam a informação temporal do vídeo captado e consiste em três diferentes passos: detecção de mudança, rastreamento e classificação. O primeiro passo reduz o espaço de procura dos *classifiers*, utilizando *Background Subtraction*. Isto ajuda a localizar a região de interesse da imagem e também a reduzir o número de falsos positivos. Uma vez caracterizada a imagem, passa-se para o segundo passo, onde é feita a segmentação da imagem do fundo, através de um model gaussiano. A imagem é depois analisada e os componentes ligados, que correspondam a potenciais pontos de interesse, são rastreados e analisados em diferentes instâncias temporais, com base na história da imagem. Finalmente, no terceiro passo, tendo os objetos de interesse sido detetados e rastreados, determina-se a origem dos objetos e associa-se, ou não, a uma classe de humanos. Os resultados experimentais são satisfatórios obtendo um rácio de deteção de 97%, em condições reais.

De forma a proteger pessoas em zonas de operação de robôs tratores, é proposto em [4] um método de deteção que utiliza OSV, para identificar e localizar os obstáculos na proximidade do veículo. O sistema posicionado no topo do veículo consiste em duas câmeras omnidirecionais HD, com múltiplas lentes ligadas a um computador, para efeitos de processamento. É desenvolvida então uma imagem 3D panorâmica, como na Fig. 2.5, através da reconstrução de imagens *stereo* retificadas, utilizando um método de soma da diferença dos quadrados, para detetar a posição de humanos. O objetivo era detetar um humano numa imagem panorâmica 360 graus, utilizando um método deteção ótica de movimento ou *optical flow detection*. A imagem panorâmica é agregada, utilizando um método baseado nas relações geométricas das câmeras. Os resultados experimentais concluem que o sistema consegue fazer a deteção com um alcance entre 4 a 11 metros, mas apenas durante o dia.

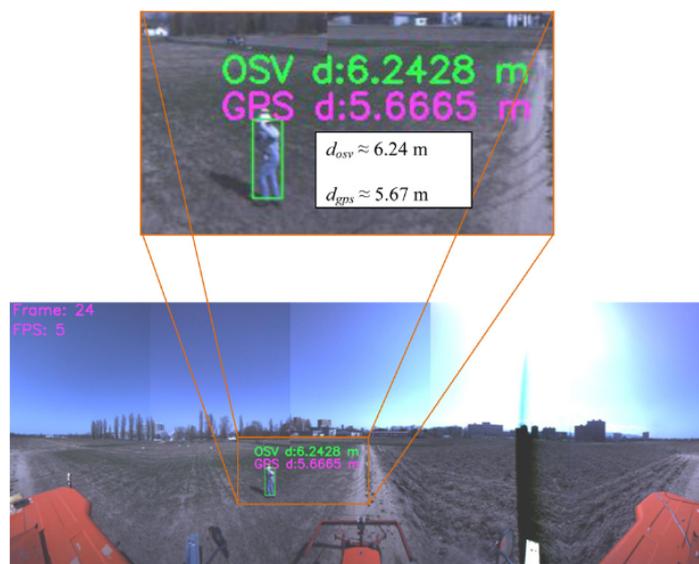


Figura 2.5: Resultados experimentais da deteção obtidos em [4]

Um outro método utilizado para deteção de humanos na zona de operação de robôs, é a utilização de lasers, como em [18], onde é proposto um sistema de deteção e localização de pessoas

que usa LRF de múltiplas camadas num robô móvel. Em cada camada existem sensores laser que apontam em direções opostas, de forma a produzir uma representação 360 graus do ambiente em volta do robô. A informação é depois fundida e dividida em diferentes segmentos, utilizando um *threshold* adaptativo para encontrar *breakpoints*. Para cada segmento é processado um conjunto de características, para determinar se o segmento corresponde ou não a uma parte do corpo. É depois utilizado um algoritmo, que treina o classificador para reconhecer as partes do corpo. Os segmentos que sejam potenciais pontos de interesse, são combinados para realizar a estimação da posição e deteção de pessoas. A variação e uniformização de intensidade foi introduzida de maneira a melhorar a deteção, já a intensidade de reflexão do laser foi introduzida para evitar o problema de fusão de segmentos diferentes.

Também existem métodos onde as duas tecnologias antes abordadas são usadas em conjunto, como é feito em [19]. Aqui é apresentado um sistema que procura detetar e localizar, de forma fiável, pessoas numa área aberta. Este sistema consiste em vários lasers dispostos numa única fila e uma câmara. A deteção baseada em laser resulta em medições de humanos de forma automática e de confiança, em vários cenários, uma vez que não são afetados pelas condições de luminosidade e o processamento das informações do laser não requer muito tempo. No entanto, tem limitações relativas à identificação de cores. A deteção baseada em visão está limitada a situações simples, onde as pessoas se encontram isoladas, havendo também métodos que recorrem à identificação de rostos e cabeças para fazer a deteção. Contudo, estes métodos são computacionalmente pesados e requerem muito tempo de processamento. A solução apresentada recorre ao rastreamento da trajetória de pés, registando informações de distância provenientes de diferentes lasers e combinando-as com as informações captadas pela câmara, relativamente ao histograma das cores da região do corpo. Embora apresente algumas limitações, conclui-se que, o sistema ao tirar partido das informações de dois tipos diferentes de sensores, consegue ultrapassar os problemas associados individualmente a cada um deles.

Em [20], tal como em [19], é apresentado um sistema de deteção de pessoas com base em sensores laser e sensores de visão RGB, para aplicação em sistemas autónomos de vigilância de infraestruturas ao ar livre. A fase de previsão do algoritmo utiliza a fusão de dois módulos, um para os dados do laser e outro para os dados da câmara. O módulo laser tem duas funções, uma delas é gerar a primeira estimação da presença de pessoas e a outra é dar informação relativa ao alcance de valores do sensor, no que diz respeito às distâncias das estruturas no plano da imagem. Isto permite ao algoritmo reduzir a sua zona de procura significativamente. O módulo de visão serve como uma segunda opinião para o algoritmo, o que faz com que o seu desempenho seja melhor. A deteção é feita com base num método de histograma de gradientes orientados. A fusão dos dados dos sensores é realizada através de uma ferramenta probabilística, que combina o *output* dos classificadores de cada um dos módulos dos sensores. Estes dados dos sensores são tidos em conta como probabilidades condicionais, que dependem do estado atual da avaliação. No cálculo da probabilidade final, ambos os sensores têm o mesmo peso. Conclui-se, com os resultados experimentais, que o método proposto consegue fazer a deteção de forma precisa e que a fusão das informações dos dois sensores aumenta a eficácia do sistema quando comparada à utilização

individual de cada um dos sensores.

Uma pequena variação dos métodos de detecção baseados em sensores laser e sensores de visão é o apresentado em [21], onde é proposto um método de detecção de pessoas em ambientes urbanos, que utiliza a informação gerada por um sensor LiDAR e uma câmera. É ainda descrita uma arquitetura de fusão dos dados dos dois sensores, de forma centralizada e descentralizada. No primeiro, o processo de fusão ocorre ao nível das características de cada sensor, que são agregadas num único vetor para, posteriormente, serem classificadas de forma única. No segundo, os sensores são classificados de forma individual e fundidos através de um método de treino, que leva a um melhoramento do desempenho do método de detecção e, ainda, a um controlo do número de falsos alarmes a um nível tolerável, o que não acontece no primeiro processo. Conclui-se, que as falhas do LiDAR devem-se ao alcance limitado em termos de distância, quanto mais afastado está o objeto menos informação é obtida e, conseqüentemente, o sistema torna-se mais propenso a detecção de falsos negativos. Para além disso, situações onde haja proximidade entre pessoas ou pessoas e objetos leva a falhas de detecção. No caso do sistema de visão, as falhas devem-se sobretudo a situações onde as imagens detetadas têm baixo contraste ou as texturas dos fundos da imagem têm altos valores de intensidade. Em relação às estratégias de fusão, conclui-se que o processo descentralizado consegue obter melhores desempenhos que o processo centralizado.

Outra forma de realizar a detecção de objetos em imagens captadas por sensores de visão é apresentada em [22] e consiste em reformular a detecção, como um único problema de regressão, passando diretamente os píxeis da imagem para coordenadas da *bounding box*. Este método denomina-se YOLO (You Only Look Once), dado que o sistema numa única iteração faz a previsão do número de objetos e da sua localização. O processo é bastante simples, através de uma única rede de convolução é realizada a previsão das *bounding boxes* e a sua classe de probabilidades. A aprendizagem da rede é feita em imagens completas e otimiza diretamente o desempenho da detecção, assim pode-se obter vantagens comparadamente a métodos tradicionais de detecção. A velocidade é também uma vantagem deste método, uma vez que não possui uma *pipeline* complexa. Ao contrário de métodos de detecção, como processamento baseado em regiões e deslizamento da janela, este utiliza a imagem completa para treinar e para testar e, portanto, inevitavelmente utiliza informação contextual sobre as classes, bem como o seu aspeto. É ainda de realçar que este método tem as suas limitações, nomeadamente na restrição espacial das *bounding boxes* devido à natureza de cada célula da grelha de detecção só poder identificar duas caixas e uma classe. Isto faz com que o número de objetos detetados seja limitado. Outra limitação é a generalização de objetos novos ou de proporções desconhecidas, visto que o processo de aprendizagem prevê as regiões com base em informação conhecida *a priori*.

A utilização de redes neuronais pode permitir, em algumas situações, fazer o tratamento das imagens capturadas pelos sensores de visão e detetar objetos, como é proposto em [23]. O método obtém a localização dos objetos numa imagem, através de redes neuronais, que são usadas para o processo de extração de características e aprendizagem, fazendo a previsão de várias *bounding boxes* ao mesmo tempo. A função de custo das diversas localizações é formulada e tira partido do diferente número de localizações de interesse do *groundtruth*, numa dada imagem, para apren-

der a prever a localização dos objetos em imagens ainda por analisar. Este método, embora não atinja o nível de desempenho de outros métodos de deteção, tem um custo computacional bastante reduzido, o que o torna interessante quando combinado com outros sensores complementares.

A deteção em câmeras visíveis acenta em modelos previamente treinados, para identificar objetos e, mais especificamente, pessoas. Isto é relativamente comum em aplicações de IoT e sistemas de vigilância [15] que, por vezes, utilizam as câmeras para complementar a utilização de sensores passivos, como é o caso dos sensores PIR. Em outros casos, os sensores passivos até acabam por se tornar obsoletos, tendo em conta as possibilidades existentes quando utilizamos câmeras, como o uso de visão computacional, processamento de imagem, machine learning, entre outros, que garantem resultados de identificação mais completos e sofisticados.

Embora os sensores PIR, em alguns casos, sejam uma boa opção como, por exemplo, em sistemas de deteção de movimento, por serem uma solução barata e fácil de usar, por vezes resultam em falsos resultados que levam a uma performance pouco robusta do sistema. Existem dois tipos de erros que são normalmente associados a este tipo de sensores que são o facto de não poderem distinguir o movimento de uma pessoa do movimento de outro objeto e, portanto, seria considerado um falso positivo. O outro é quando existe uma pessoa parada na zona de ação do sistema e este não é capaz de a detetar, o que é considerado um falso negativo. A deteção de pessoas pode ser então obtida através do uso de processamento de imagem com o método da subtração do fundo da imagem, juntamente com métodos de deteção de caras ou cabeça e ombros [15]. Outra opção passa pelo uso de inteligência artificial, através de redes neuronais treinadas para a tarefa de deteção e identificação de objetos, como as CNN [24], DarkNET Deep Learning NN [25] e outras como a YOLO, SSD e CNN e as suas variantes [26].

De uma forma geral, existem dois tipos de algoritmos de deteção de objetos. O primeiro é um algoritmo baseado numa abordagem em dois passos, onde o primeiro passo é identificar as regiões onde se espera que os objetos estejam localizados e o segundo passo é detetar os objetos exclusivamente nessas regiões utilizando redes neuronais convolucionais como, por exemplo, a R-CNN [27], Fast R-CNN [28] e Faster R-CNN [29]. O outro é um algoritmo que utiliza uma abordagem totalmente convolucional, na qual a rede é capaz de descobrir todos os objetos presentes numa imagem numa só passagem como, por exemplo, o YOLO [30] e o Single Shot Detector [31]. O SSD, em particular, é um dos algoritmos mais comuns de deteção de objetos de apenas um nível. Os algoritmos baseados em identificação das regiões têm normalmente uma maior precisão, mas são mais lentos, uma vez que exigem maiores cargas computacionais. Os algoritmos de deteção baseados numa só passagem são mais eficientes e apresentam uma boa precisão e, como tal, foram alvo de um estudo mais aprofundado.

O SSD faz uso da ideia de regressão do YOLO, para transformar a deteção de objetos num simples problema de regressão para processamento. Ao mesmo tempo, o SSD adota o método de deteção em pirâmide (diferentes níveis de características) e adota a classificação softmax [32], bem como a regressão da posição nas *feature images* de múltipla escala.

O SSD é um modelo de deteção diferente do Region-CNN ou o Faster R-CNN, visto que em vez de gerar regiões de interesse e classificá-las separadamente, fá-lo simultaneamente de uma só

vez. Esta rede está desenhada para ser independente da rede base e, portanto, pode ser executada "em cima" de outras CNNs (Convolutional Neural Networks), incluindo a Mobilenet [33].

Este modelo é uma rede convolucional retroalimentada, que produz um conjunto de bounding boxes com tamanho fixo e cada uma tem um resultado associado para a presença de objetos de uma determinada classe, seguido de um degrau de supressão de não-máximos, para produzir a detecção final. Uma estrutura auxiliar é adicionada nas primeiras camadas da rede, que são baseadas numa arquitetura standard utilizada em classificação de imagens de alta resolução. Esta estrutura é adicionada à rede, para produzir detecções com características chave, como mapas de detecção de características multi-escala, preditores convolucionais para detecção e boxes e resoluções, por defeito.

A Mobilenet é uma nova geração de CNNs, proposta pela Google. O modelo mantém um equilíbrio entre o desempenho e fluidez, que é adequado no desenvolvimento em plataformas móveis, como sistemas embebidos com poucas configurações de hardware e baixo custo computacional. O Raspberry Pi é um exemplo deste tipo de sistemas.

Basicamente, a Mobilenet é um conjunto de CNNs que são amplamente mais pequenas em tamanho e rápidas em desempenho quando comparadas a outros modelos. Em termos de profundidade, convoluções separadas aplicam um único filtro em cada entrada, para filtrar os dados de entrada, seguido por convoluções 1x1 que combinam estes filtros num conjunto de características de saída. Estas camadas separáveis, em termos de profundidade, replicam o funcionamento típico de camadas de convolução, mas com velocidades muito maiores e com uma pequena diferença. Tipicamente os filtros de convolução são combinados em características de saída, mas a convolução separável em profundidade é dividida em duas camadas separadas, uma para filtragem e outra para combinar as saídas. Isto minimiza o tamanho do modelo e reduz o custo computacional do sistema. Todas as camadas são seguidas por um *batchnorm* e uma não-linearidade ReLU, à exceção da última camada completamente ligada, que alimenta a camada *softmax*, para classificação sem não-linearidade [34]. A Mobilenet tem 28 camadas sem contar com as convoluções *depthwise* e *pointwise*.

## 2.4 Fusão Sensorial

Como já foi sendo referido em [12], [19], [20] e [21] a fusão de dados de diferentes sensores permite obter melhores resultados, comparativamente a situações em que os sensores são utilizados de forma individual num mesmo sistema. Assim, neste capítulo são explorados diferentes abordagens de fusão sensorial.

Em robótica móvel é muitas vezes utilizada odometria, baseada no movimento das rodas, contudo, este mecanismo apresenta alguns inconvenientes, devido à irregularidade dos pavimentos e imprecisão crescente ao longo do tempo. Para minimizar esses problemas, é sugerida em [5] uma solução que consiste em fundir a informação de outros sensores de deslocamento e rotação, com o intuito de otimizar o posicionamento do robô. A utilização de diversos sensores pode dificultar o

tratamento das incertezas e imprecisões associados a cada um, assim como a eliminação das perturbações provocadas ocasionalmente pelos sensores e a determinação do modelo de observação e movimentação mais apropriado. Para minimizar estes problemas é utilizado um Filtro Estendido de Kalman, desta forma os dados dos sensores da IMU, dos encoders e do Kinect são fundidos e utilizados como dados de entrada do EKF, para produzir informações de posicionamento e orientação mais precisas. Conclui-se que a utilização simultânea dos três sensores, realizando a fusão sensorial melhora significativamente o rastreamento do posicionamento real do robô. Na Fig. 2.6 demonstra-se a trajetória desejada, assim como alguns resultados de odometria do robô em diferentes configurações. O verde representa a configuração utilizando a fusão sensorial de todos os sensores, o azul, apenas os encoders e a IMU e o vermelho, somente os encoders.

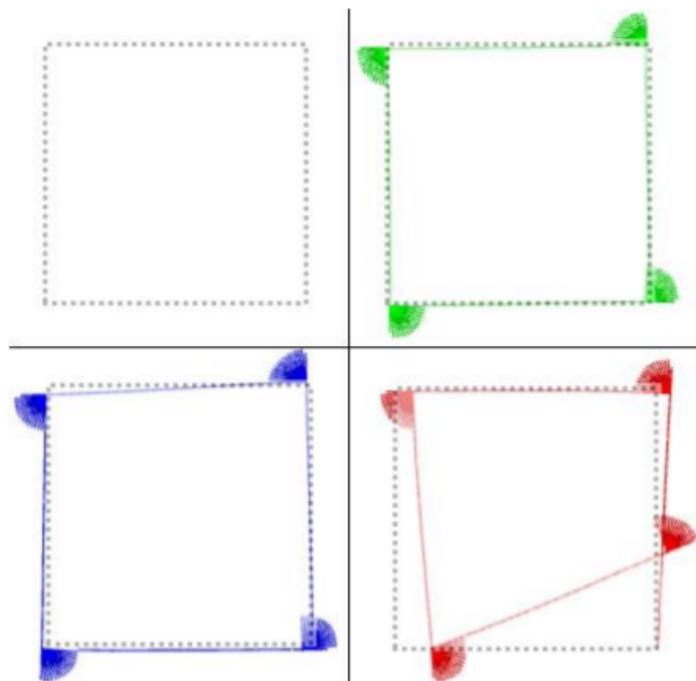


Figura 2.6: Resultados experimentais da odometria obtidos em [5]

Com o objetivo de reconhecer o ambiente em volta de AGVs, como é o caso desta dissertação, é implementado em [6] um sistema de fusão sensorial, composto por dois subsistemas, que oferecem uma resposta a esse problema. Um subsistema é responsável pela estimação da posição do robô e o outro pelo estado do ambiente em volta do robô, fazendo um mapeamento do mesmo. A utilização de diversos sensores é importante, para garantir a fiabilidade do sistema de percepção e aumentar a velocidade de processamento do sistema, ao mesmo tempo que reduz o seu custo. A informação dos diferentes sensores é processada e fundida, com o objetivo de localizar o robô no espaço, ao mesmo tempo que deteta os obstáculos ao seu redor. A informação recolhida pelos sensores é armazenada e utilizada para gerar uma grelha multidimensional, onde cada célula corresponde a uma região do ambiente, como na Fig. 2.7. Neste método o sistema de fusão sensorial

é utilizado em conjunto com redes *Bayesian* dinâmicas e é construído um mapa do ambiente que armazena a probabilidade de cada célula estar ocupada ou não.

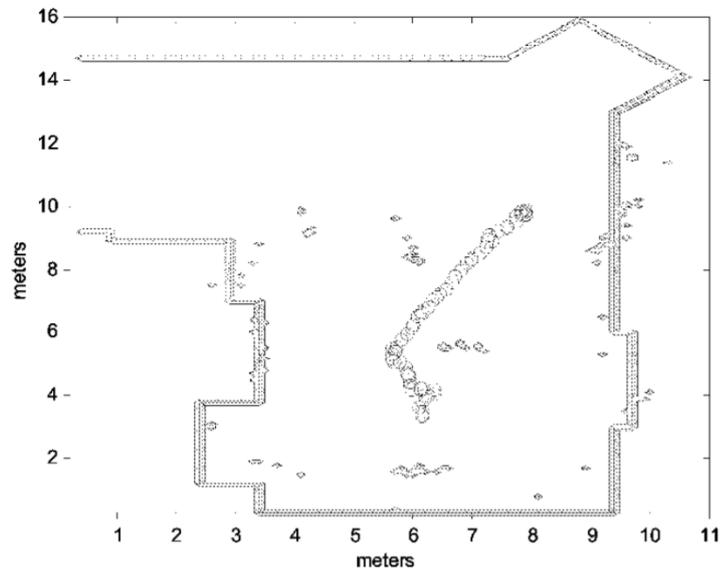


Figura 2.7: Exemplo do mapa de ocupação gerado em [6]

Dada a necessidade de fazer uma detecção eficiente de obstáculos em trajetos rodoviários, é proposto em [35] um método anticolisão, através da gestão de distâncias que funde dados 3D de um sensor LiDAR e dados de imagem 2D. Os obstáculos são primeiro detetados através do sensor LiDAR e entre os seus scans relativamente lentos, os dados da câmara 2D de alta frequência permitem auxiliar a detetar a distância entre os potenciais obstáculos e o veículo a ser conduzido. Esta técnica de fusão tira partido da informação geométrica detetada, para extrair informação relativa à profundidade, em cada instante, e é depois corrigida com a informação do LiDAR. A solução obtida permite gerir a interdistância em autoestradas, especialmente a altas velocidades. Através de experiências realizadas, conclui-se que o método permite obter estimações rápidas e precisas de potenciais ameaças, sendo particularmente útil para evitar colisões entre veículos e obstáculos.

A fusão sensorial é também usada em veículos inteligentes, para questões de segurança, como é referido em [35] e otimização da mobilidade. Em [36] é discutida a realização de múltiplas tarefas de perceção em simultâneo, de forma a obter melhores representações das características do ambiente e através de fusão sensorial otimizar o processo de detecção. É então desenvolvido um sistema de detecção, composto por múltiplos sensores, que retira vantagem de perspectivas de visão por ponto e por região de interesse, resultando numa representação mais completa. O processo de aprendizagem multitarefa confere dados geométricos e do mapa, permitindo melhorar o processo de aprendizagem e tornar o processo de fusão mais robusto.

Em [37], é proposto um outro processo de detecção e classificação de objetos em movimento que, através das informações de diferentes sensores, é capaz de melhorar a percepção da representação do ambiente, com recurso à fusão sensorial. A tarefa de percepção do algoritmo é composta por duas componentes, a primeira define uma representação do objeto que inclui informação sobre a sua classe e a segunda consiste numa arquitetura de percepção completa, baseada numa ferramenta que resolve o problema de detecção e localização de objetos em movimento, através da integração da representação do objeto e da gestão de incerteza. Para a fase de validação, o sistema desenvolvido é integrado num veículo de demonstração que inclui três sensores principais: LiDAR, radar e câmara. O algoritmo é testado em vários cenários e com vários objetos e conclui-se com os resultados obtidos que é possível reduzir o número de falsas deteções e melhorar a precisão e velocidade do rastreamento.

A fusão sensorial tem também aplicação em cenários industriais, para garantir que a separação física entre operadores e robôs é feita de forma segura. Em [38], é feita uma abordagem baseada em sensores em tempo real, para garantir que em situações de proximidade possa ser garantida a segurança do operador humano. O sistema funde os dados adquiridos por diferentes sensores de imagem 3D numa grelha volumétrica e segmenta o volume em regiões correspondentes ao fundo, às pessoas e aos robôs. Em volta de cada robô é identificada uma zona de perigo, que é atualizada de forma dinâmica de acordo com a posição e trajetória do robô. Da mesma forma, é ainda identificada uma zona de segurança em torno do operador. Uma eventual colisão entre as duas zonas referidas, indica uma situação de perigo iminente e, portanto, o robô é imediatamente parado até o problema ser resolvido. O sistema desenvolvido é testado com e sem a utilização da fusão dos dados dos sensores, o primeiro para determinar a capacidade individual de cada sensor fazer a deteção de pessoas em diferentes condições e o segundo para verificar o desempenho global do sistema. Conclui-se que a utilização dos diferentes sensores em simultâneo, de forma uniformizada, resulta em deteções mais eficazes, garantindo assim a operação de forma segura em diferentes condições para o utilizador.

A contagem de pessoas, num ambiente fechado, está intrinsecamente ligada à deteção das mesmas. Assim sendo, em [39], é apresentado um sistema de deteção e contagem de pessoas, que tira partido da fusão de dados de visão e profundidade, para superar problemas comuns de sistemas de deteção, como variações de luminosidade e ambientes sobrelotados. Este sistema é montado no teto da zona que se pretende implementar e começa por processar a informação de profundidade do sensor. Através de um operador morfológico, é possível corrigir problemas relacionados com a profundidade como ruído ótico e perda de dados. Assim que este passo é completo, a silhueta humana é extraída utilizando um modelo humano da imagem de profundidade processada. Finalmente, a trajetória do objeto detetado é estabelecida por aplicação de um algoritmo de correspondência bidirecional. A deteção, uma vez que é feita a partir do teto, explora as proporções, em média, entre a cabeça e torso de um humano, como base do algoritmo. Essa proporção é utilizada por ser mais consistente do que outros métodos abordados, como a utilização de apenas as dimensões da cabeça humana. Os resultados experimentais mostram que o sistema de deteção e contagem baseado em fusão sensorial desenvolvido é robusto a diferentes situações de ilumina-

ção, combinando os dados de profundidade e cor para recuperar informações visuais perdidas e obtendo uma eficácia de 98%.

## Capítulo 3

# Arquitetura Proposta

Neste capítulo é apresentada a arquitetura proposta para desenvolvimento. São mencionados os aspetos mais relevantes que foram tidos em conta, antes de passar para a fase de Implementação. A primeira secção refere-se à arquitetura proposta a nível de hardware, enquanto que a segunda faz referência à arquitetura do software a desenvolver.

### 3.1 Hardware

O objetivo desta dissertação é desenvolver um sistema de deteção de pessoas que possa ser utilizado em robôs de desinfeção autónomos. O sistema foi concebido para ser uma solução de baixo custo e baixo peso, de forma a tornar prática a sua integração, sem comprometer o desempenho do sistema. A ideia passa por montar o sistema no topo de um robô, equipado com uma torre que possui um conjunto de luzes Ultravioleta, responsáveis por realizar as tarefas de desinfeção. O detetor de pessoas é um módulo à parte do robô, que opera de forma autónoma. É alimentado a 24V e comunica sem fios, através de WiFi, ou com fios, através de uma porta série. Nesta dissertação é utilizado essencialmente o modo de comunicação com fios, através de porta série, entre o módulo de deteção de pessoas e o robô de desinfeção. No entanto, também é possível estabelecer uma comunicação sem fios com o módulo, através de WiFi. A Fig. 3.1 representa a arquitetura principal do sistema de deteção de pessoas, bem como a sua integração.

O hardware a desenvolver deve estar dividido em duas partes, uma parte estática que serve como base do módulo de deteção, mas também serve de suporte aos sensores PIR, que vão ser usados para fazer a deteção de movimento, assim como os diferentes componentes que permitem a comunicação entre o módulo de deteção e o robô de desinfeção. A ideia passa por desenvolver um PCB especificamente para este módulo, de forma a poupar espaço e ligações elétricas, onde está presente um microcontrolador responsável pelas tarefas de controlo do sistema, bem como um conjunto de outros circuitos integrados, que irão desempenhar diferentes tarefas, desde controlo de um motor de passo, até à conversão das entradas e saídas lógicas do microcontrolador de 24V para 5V e vice-versa.

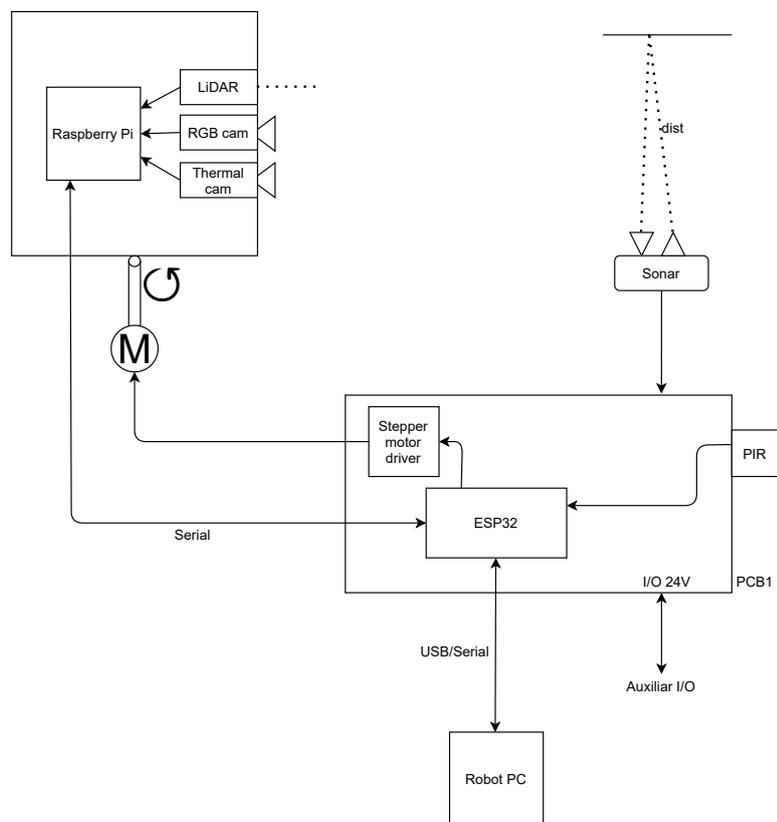


Figura 3.1: Arquitetura principal do sistema

A segunda parte consiste numa peça móvel, cuja arquitetura está apresentada na Fig. 3.2, que serve de suporte aos componentes de detecção baseada em visão, que são uma câmera térmica e uma câmera RGB orientadas de forma vertical, de maneira a maximizar o campo de visão vertical, em detrimento do campo de visão horizontal. O sistema consegue garantir que são analisados os 360 graus à volta do robô, movendo a parte superior de forma rotativa e captando os 360 graus em redor do módulo de detecção. Para além das câmeras, existe ainda um LiDAR que será responsável por fazer o cálculo da área em torno do módulo e um computador de dimensões reduzidas, cujo objetivo será processar os dados dos sensores da parte superior e comunicá-los ao microcontrolador da parte inferior.

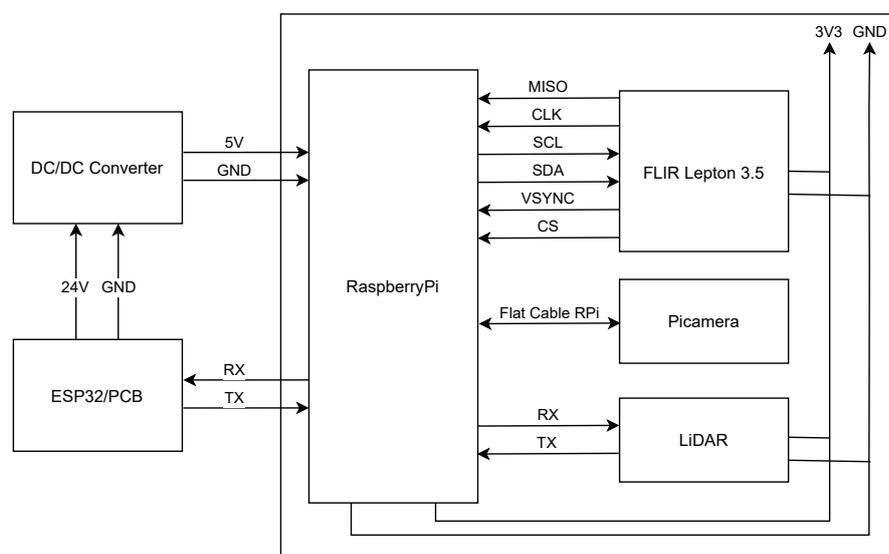


Figura 3.2: Arquitetura do módulo rotativo 360°

Para garantir o movimento de rotação e tendo em conta que o a zona central do módulo tem um espaço dedicado para a passagem de fios entre a parte móvel e a parte estática, a abordagem proposta consiste em utilizar duas polias com uma correia a envolvê-las. Uma das polias estará associada a um motor de passo, controlado pelo microcontrolador, a outra estará na base da zona central do módulo e o seu movimento é o que garante que a parte superior do módulo rode.

## 3.2 Software

A nível de software, à semelhança da arquitetura do hardware, existem duas partes, uma mais associada à parte física inferior, cujo objetivo é servir como um centro de controlo, no sentido em que é nesta zona que serão implementadas as tarefas de controlo da rotação, bem como as tarefas de comunicação entre o sistema e o computador do robô.

Dentro do sistema de detecção, também existe uma comunicação contínua entre o microcontrolador e o computador da parte superior do módulo. No sentido do microcontrolador para o

computador do módulo, é necessário transmitir informações relativas ao estado da rotação, de maneira a que seja possível sincronizar a aquisição das imagens durante as paragens. Esta abordagem garante que as imagens não fiquem distorcidas durante os movimentos de rotação do módulo e que a aquisição de ambas as câmeras, bem como a medição da área por parte do LiDAR sejam feitas simultaneamente. No sentido contrário, o computador da parte superior do módulo fica responsável por comunicar via porta série os dados de deteção obtidos pelo processamento das imagens das câmeras, retornando um sinal de deteção que consiste num único bit que será 0 ou 1. Eventualmente também poderão ser retornadas as coordenadas das bounding boxes da deteção, para posteriormente comparar os dois resultados das câmeras. A área calculada com os dados do LiDAR também é transmitida via porta série para o microcontrolador.

Já o programa de deteção de pessoas, que corre no computador do módulo superior do sistema, uma vez que executa três tarefas em simultâneo, em programas diferentes, que são elas a captura de imagens térmicas, a captura de imagens RGB e a medição de distâncias com o LiDAR, tem de disponibilizar os dados adquiridos a um programa central. Esse programa fica encarregue não só das tarefas de comunicação com o microcontrolador e sincronização das aquisições dos sensores, mas também do processamento das imagens e das distâncias do LiDAR. Assim sendo, é necessário recorrer a um programa de comunicação dentro do próprio dispositivo que seja compatível com diferentes linguagens de programação, como C++ e Python, e não comprometa o desempenho do sistema.

Relativamente à comunicação de saída do microcontrolador, para o computador do robô de desinfeção, uma vez recebidos os dados da deteção, esses são inicialmente associados a um conjunto de pinos específicos do microcontrolador que estão conetados às saídas de 24V, compatíveis com o computador. Os dados transmitidos consistem num único bit de deteção do sistema, que é baseado nos dados de deteção dos três métodos diferentes existentes, deteção baseada em imagens térmicas, deteção baseada em imagens RGB e deteção baseada em sensores PIR. Para além desse bit de deteção, são ainda enviados outros dados, que correspondem diretamente ao valor da área calculada pelo computador da parte superior do módulo e o histograma de temperaturas, calculado com a câmara térmica.

No sentido contrário, o robô envia diferentes pedidos para o microcontrolador. Um dos sinais é um pedido de paragem da rotação, isto é particularmente útil em situações em que o sistema não seja necessário e, portanto, o movimento de rotação controlado pelo microcontrolador é cessado, mantendo-se no entanto o sistema ativo em caso de necessidade de reativação do movimento. Caso o robô deixe de ser necessário, assim como o sistema de deteção, cabe ao robô comunicar essa informação enviando um sinal de *shutdown* para o microcontrolador que, por sua vez, irá comunicar ao computador da parte superior do módulo que deverá dar início ao procedimento de encerramento. Este passo evita que a alimentação seja cortada de forma abrupta, o que poderia corromper ou danificar o sistema de deteção de pessoas.

Existe ainda uma comunicação direta entre um sensor de medição de distância, direcionado para o teto, e o computador do robô de desinfeção, que informa sobre a altura a que o módulo se encontra. Este sensor é necessário, uma vez que se pretende que a altura da torre de desinfe-

ção seja variável e, tendo em conta que o módulo de detecção se encontra no topo da torre, fica salvaguardado algum tipo de dano a que o mesmo poderia estar sujeito.

Na Fig. 3.3 é apresentada a arquitetura do sistema a nível de software, nos três níveis anteriormente falados, computador do robô, microcontrolador e computador do módulo de detecção.

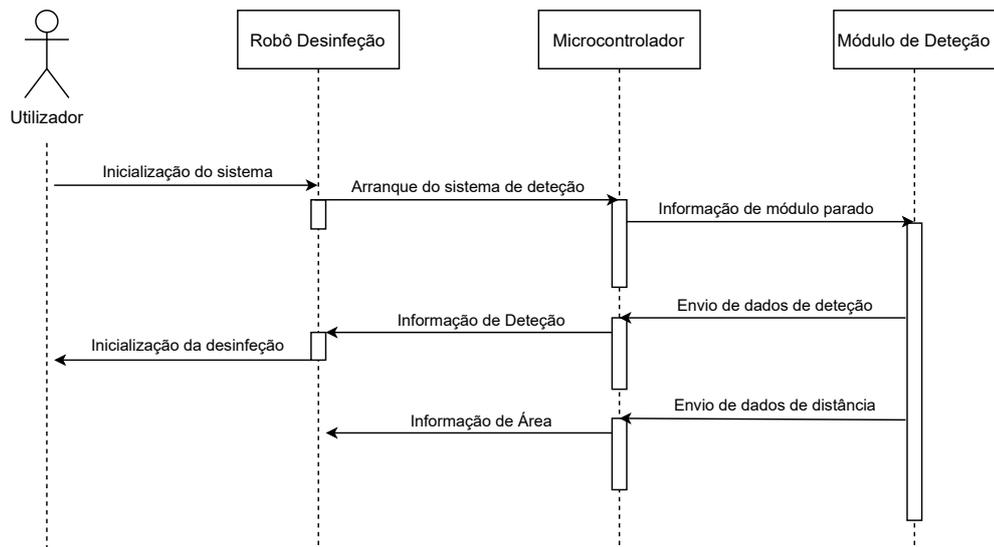


Figura 3.3: Arquitetura do software do sistema



## Capítulo 4

# Implementação

Neste capítulo é apresentado todo o trabalho realizado de acordo com a arquitetura proposta. A primeira secção apresenta todo o hardware desenvolvido para o sensor de deteção. A segunda secção contém todos os componentes desenvolvido a nível de software.

### 4.1 Hardware e Mecânica

O sensor desenvolvido pode ser decomposto em dois grandes grupos. O grupo superior, constituído pelos principais sensores de deteção que são a câmara térmica e a câmara RGB, bem como o sistema embebido de processamento e tratamento dos dados adquiridos pelas câmaras, que é um Raspberry Pi 4B. Para além desses componentes, a parte superior possui ainda um LiDAR e um conversor DC-DC. A nível de componentes, o grupo inferior é constituído por um PCB, desenvolvido especificamente para este projeto, onde reúne todos os circuitos integrados responsáveis por fazer a comunicação entre o Raspberry Pi e o PC do robô que realiza a desinfeção. Este grupo inferior possui ainda um motor de passo.

Os componentes usados na parte superior são um Raspberry Pi 4B, para tarefas de processamento dos sensores e comunicação desses dados com o microcontrolador da parte inferior, um conversor DC-DC LM2596, para conversão dos 24V (ou um pouco menos devido à queda de tensão provocada pelas dimensões dos fios do slip ring) para 5V, uma câmara RGB Raspberry Pi Camera Module v2, uma câmara térmica FLIR Lepton 3.5, com uma placa de desenvolvimento FLIR Lepton Breakout Board v2 e um LiDAR TFMini Plus UART, para medições de distância de objetos na direção em que o sensor estiver apontado. Para além desses componentes na zona superior, são ainda usados um rolamento 20 x 40 x 12 mm e ainda um Slip Ring de 6 fios. Na parte inferior, os componentes usados são quatro sensores PIR HC-SR501 para deteção de movimento, um sonar US-100 para medição da distância do sensor ao teto na vertical, um motor de passo NEMA 17HS4401 e um PCB com um microcontrolador Firebeetle ESP32-WROOM-32 Development Board, um driver de motor DRV8825, um conversor DC-DC LM2596, igual ao da parte superior e um conjunto de integrados ULN2003, expansores de pinos PCF8574 e optoacopladores TIL199.

A caixa que contém os componentes da parte superior é constituída por cinco peças, todas elas desenvolvidas no programa Fusion 360 e impressas em 3D. Três delas, como se pode ver na Fig. 4.1, constituem a base da caixa superior, uma peça está fixa à parte inferior e serve também para segurar o rolamento, que irá facilitar a rotação da parte superior em torno do eixo central. Esta peça possui ainda na sua base um slip ring que permite a passagem de fios, da parte inferior para a superior, sem haver o perigo de se entrelaçarem com o movimento rotativo do sistema. Uma segunda peça que envolve o rolamento e por onde são transmitidos os fios provenientes do slip ring. Esta peça serve como uma espécie de polia e é uma das duas polias que são envoltas por uma correia dentada que garante a rotação do sistema com a ação do motor de passo. A terceira peça serve para fixar a polia central às caixas que envolvem as câmeras, o conversor DC-DC e o Raspberry Pi. As duas últimas peças, representadas na Fig. 4.2, são a caixa que segura o Raspberry Pi e o conversor DC-DC e uma tampa que é responsável pela fixação das câmeras térmica e RGB e do LiDAR. Para facilitar a comunicação com os periféricos do Raspberry Pi foram abertos orifícios para as portas USB, Ethernet e HDMI.

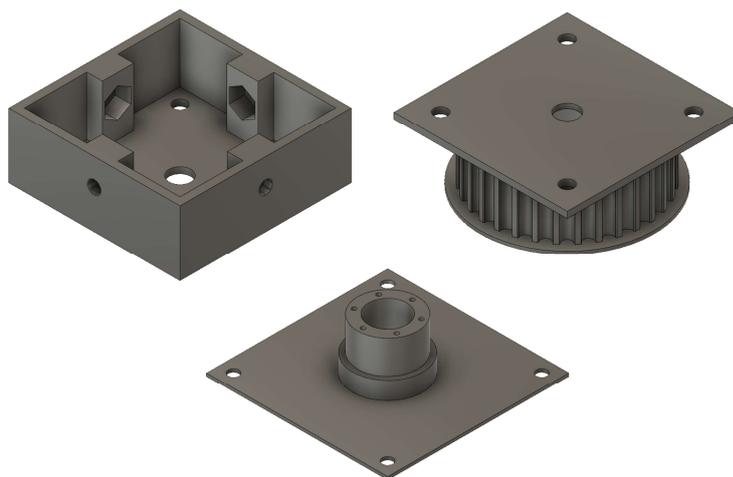


Figura 4.1: Peças que constituem a base da caixa superior

A parte inferior consiste numa caixa dividida em três partes, como se pode ver na Fig. 4.3. Para além disso, esta parte inferior possui ainda uma polia mais pequena, comparativamente à polia da caixa superior, conetada ao eixo do motor de passo e envolvida também pela mesma correia da outra polia. Estas peças também foram todas elas desenvolvidas utilizando o programa Fusion 360 e impressas em 3D.

O motor de passo, assim como a caixa superior e o sonar são fixos a uma primeira peça, que serve como tampa superior da caixa. Esta tampa tem ainda quatro pontos de fixação que são usados para juntar à segunda parte da caixa. A segunda parte garante a integridade estrutural da base e serve ainda para fixar, em quatro das suas paredes, quatro sensores PIR responsáveis pela deteção de movimento. Finalmente, a terceira parte da caixa inferior é muito semelhante à primeira parte,

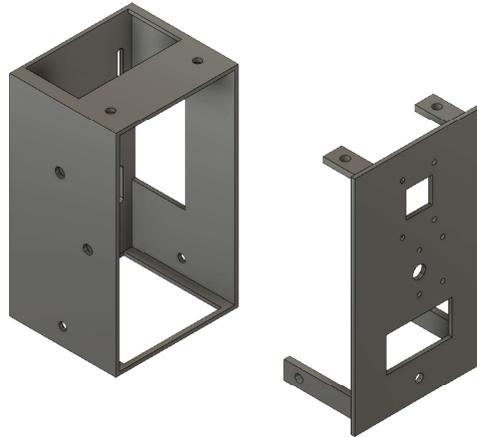


Figura 4.2: Peças que constituem o topo da caixa superior

no entanto, apenas possui furação para o PCB de comunicação entre o Raspberry Pi e o PC do robô e tem também uma outra abertura para a saída das ligações elétricas necessárias entre o Raspberry Pi e o PC.

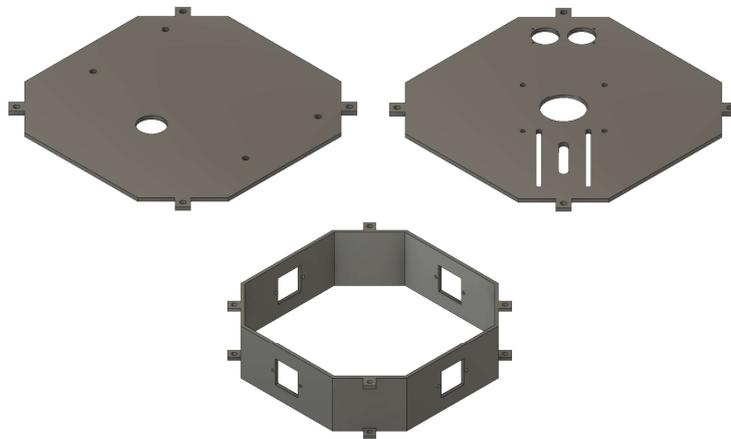


Figura 4.3: Peças que constituem a caixa superior

Nas Fig. 4.4 e 4.5, respectivamente, podemos observar o resultado final das ligações elétricas do módulo superior e do módulo inferior do sensor de detecção de pessoas.

O PCB da Fig. 4.6 foi desenvolvido para facilitar a integração dos componentes que permitem a comunicação entre o microcontrolador e o Raspberry Pi, assim como a comunicação entre o microcontrolador e o PC do robô de desinfecção. É importante destacar que de um lado da placa



Figura 4.4: Resultado final da montagem elétrica do módulo superior

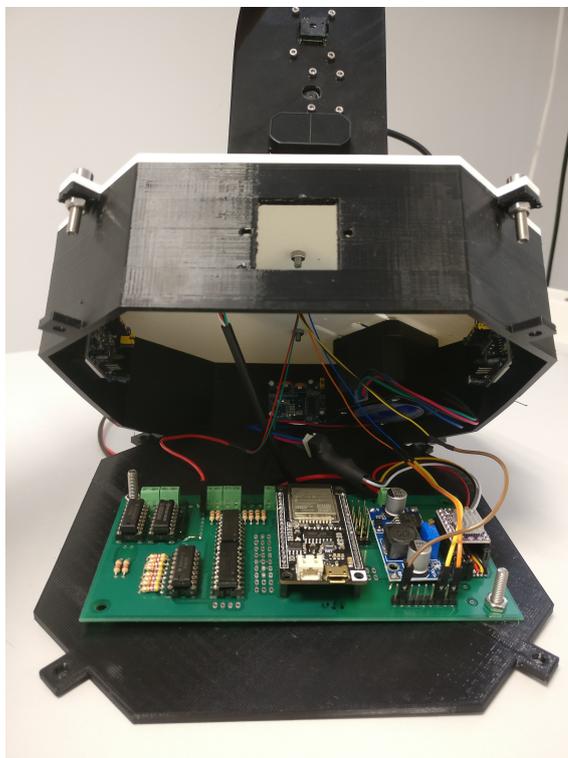


Figura 4.5: Resultado final da montagem elétrica do módulo inferior

está uma camada de alimentação 3.3V e do outro lado da placa está uma camada de massa. No Anexo A estão presentes o esquema elétrico do PCB, assim como o seu esquemático.

Os pinos associados ao microcontrolador têm um conjunto de três *headers* que correspondem à massa, alimentação e sinal sempre nessa ordem, de forma a ser compatível com dispositivos que necessitem de recorrer à alimentação da placa, para além do sinal, como é o caso dos sensores PIR.

O conversor DC-DC tem associado aos seus pinos de entrada, positivo e negativo, um terminal onde é ligada a alimentação de 24V proveniente do exterior. Por outro lado, os seus pinos de saída de 5V e Ground estão ligados ao microcontrolador, assim como outros componentes que necessitem de alimentação de 5V.

O driver do motor tem três conjuntos de pinos em destaque. O primeiro são três pinos associados ao microcontrolador: STEP, EN e DIR, para controlar as configurações do mesmo, assim como velocidade de rotação e ativação. O segundo são três pinos associados a um conjunto de jumpers: M0, M1, M2, que permitem seleccionar um dos oito modos do motor como, por exemplo, full-step ou half-step. O último conjunto de pinos em destaque são os de comunicação entre o driver e o motor de passo: A1, A2, B1 e B2.

Os optoacopladores, juntamente com a matriz de transistores e os expansores de pinos, permitem compatibilizar os sinais do microcontrolador e do PC do robô de desinfeção. Assim sendo, os sinais de saída de 5V do microcontrolador são convertidos para 24V antes de serem lidos pelo PC e, por outro lado, os sinais de 24V do PC são convertidos para 5V antes de chegar aos pinos do microcontrolador sem comprometer a integridade elétrica dos sistemas.

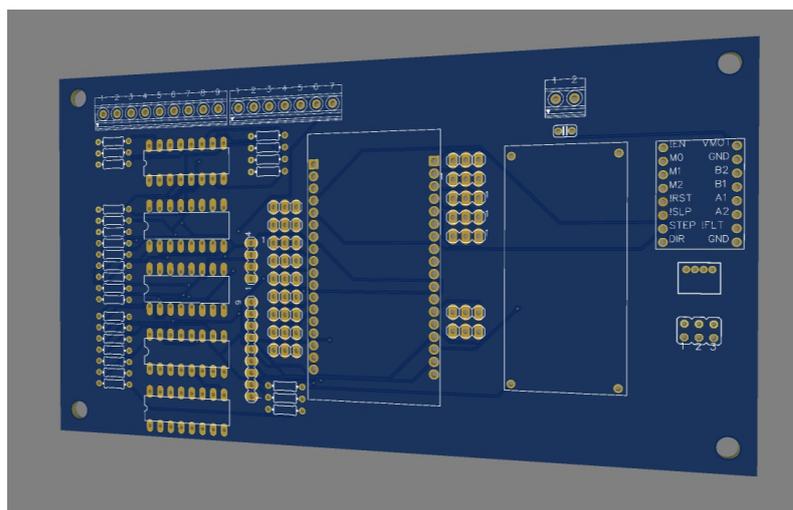


Figura 4.6: Mockup do PCB desenvolvido

Finalmente, nas Fig. 4.7 e 4.8, é apresentado o protótipo final do PCB desenvolvido, assim como a caixa final do sensor de detecção de pessoas já montada com todos os seus componentes.

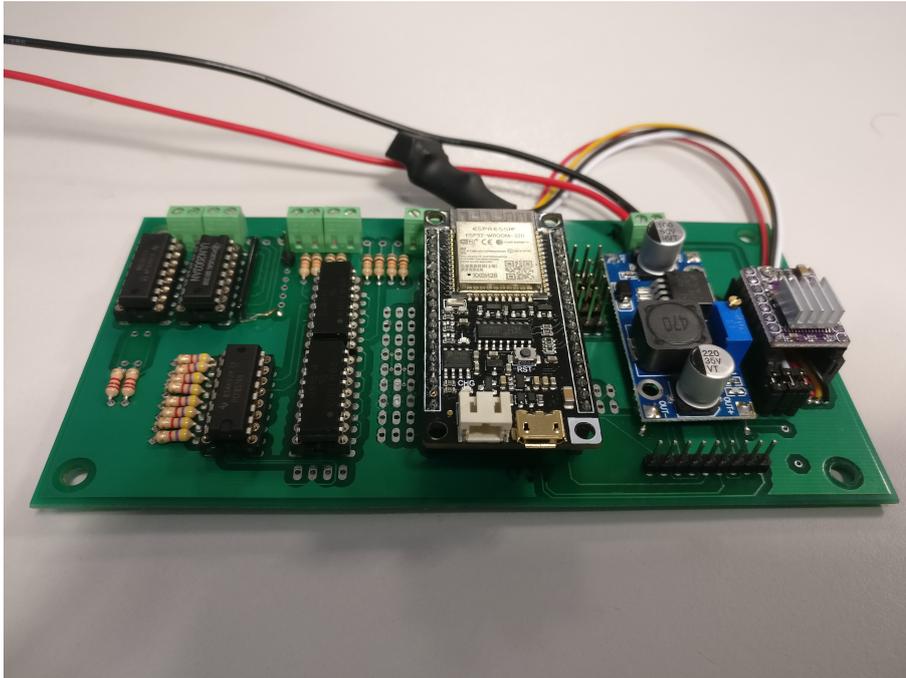


Figura 4.7: Resultado final do PCB desenvolvido

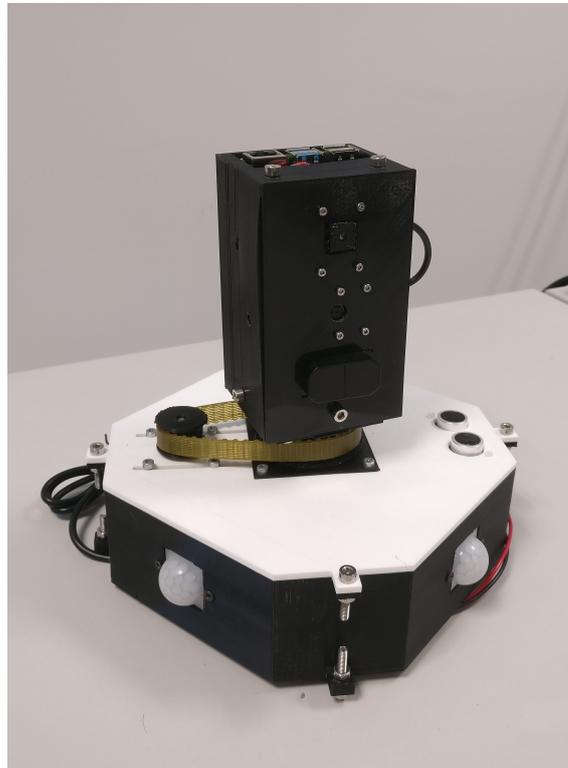


Figura 4.8: Resultado final da caixa do sensor de detecção de pessoas

## 4.2 Software e Protocolos

Esta secção está dividida em quatro subsecções, que correspondem aos principais passos da implementação do software do sistema de detecção de pessoas. A primeira diz respeito à câmara térmica onde é explicado todo processo, desde a aquisição dos dados *raw* e a sua estrutura, até à transformação desses dados numa imagem final térmica. De seguida, é apresentado o método de detecção e identificação de pessoas em imagens com base numa rede neuronal treinada. A terceira parte, corresponde às tarefas desempenhadas pelo microcontrolador, que desempenha um papel fundamental de sincronização e comunicação entre todas as partes envolvidas no sistema. Finalmente, a última subsecção refere-se ao protocolo implementado para a comunicação entre os diferentes programas de detecção, que são executados de forma simultânea no Raspberry Pi.

### 4.2.1 Câmera Térmica

A Lepton disponibiliza um CCI via Two Wire Interface, semelhante ao I2C, onde a única diferença está no facto de todos os registos da Lepton serem de 16 bits e, conseqüentemente, só são permitidas transferências de 16 bits. O endereço CCI é 0x2A. De uma forma geral, todos os comandos enviados pelo CCI são de uma das seguintes formas: GET para leituras, SET para escritas ou RUN para execução de funções.

Existe um protocolo VoSPI, que permite a transferência de vídeo, através de canais SPI, de forma eficiente e verificável. O protocolo é baseado em pacotes sem sinais temporais incorporados e sem necessidade de controlo de fluxo. O master inicia todas as transações e controla a velocidade de clock. Os dados podem ser extraídos do Lepton, que é o slave, a um ritmo flexível. Esta flexibilidade pode ser melhor entendida na Fig. 4.9, onde é possível observar a utilização de um clock mais lento e, portanto, utiliza quase completamente o período de um frame. Por outro lado, vemos também um clock mais rápido, onde apenas é utilizada uma pequena fração do período de um frame. O master tem a possibilidade de parar o clock e desseleccionar o CS até ao próximo frame disponível, assim que todos os dados de um dado frame tenham sido lidos. É ainda possível que o master mantenha o clock e o CS ativo e, nesse caso, o Lepton transmite pacotes descartáveis, até chegarem os próximos dados de vídeo disponíveis.

O VoSPI utiliza três das quatro linhas normalmente utilizadas nos canais SPI, sendo elas o SCK, o CS e o MISO. O sinal MOSI não é atualmente utilizado e deve ser ligado à massa ou definido com o nível lógico baixo. As implementações com o Lepton estão limitadas a um único master e um único slave. É utilizado o SPI Mode 3, o que significa que o sinal de clock está normalmente com o nível lógico alto e os dados são definidos no falling edge do SCK e devem ser amostrados no rising edge do controlador do utilizador, como se pode observar na Fig. 4.10.

A frequência máxima da transmissão é 20 MHz, enquanto que a frequência mínima é dada em função do número de bits de dados por frame que são necessários recolher. O número de bits depende de definições do utilizador, como o formato de vídeo, Raw14 (por defeito) ou RGB888

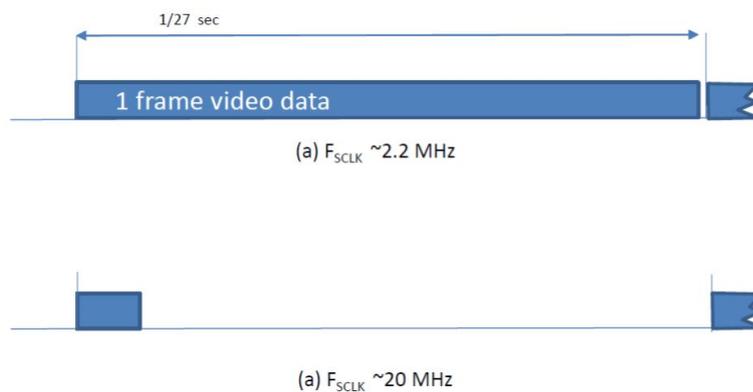


Figura 4.9: Exemplos de transmissão por VoSPI a diferentes velocidades

e o modo da telemetria, ativo ou inativo (por defeito). No nosso caso, como ambos os valores usados são aqueles definidos por defeito, o comprimento dos pacotes é 164 bytes e o número de pacotes por frame são 60. Por defeito, cada pacote contém um header de 4 bytes e um payload de 160 bytes.

No nosso caso é utilizado um Lepton 3.5 e, assim sendo, utiliza o protocolo Lepton 3 VoSPI que é construído de forma hierárquica com diferentes tipos de objetos.

- **VoSPI Packet:** O protocolo Lepton 3 VoSPI é baseado num único pacote VoSPI standard que é a transação mínima entre master e slave. Cada pacote de vídeo contém dados de uma única linha de vídeo ou telemetria. Para além disso, o protocolo inclui pacotes descartáveis que são transmitidos quando não existem pacotes de video disponíveis.
- **VoSPI Segment:** Um segmento VoSPI é definido como uma sequência contínua de pacotes VoSPI que consistem num quarto dos dados dos pixéis de um frame. Para manter a sincronização, é necessário ler cada segmento antes do seguinte ser disponibilizado.
- **VoSPI Stream:** É definida como uma sequência contínua de segmentos VoSPI.

No caso de pacotes de vídeo, o header inclui um ID de 2 bytes e um CRC de 2 bytes. O ID codifica o número do segmento, que pode ser um inteiro entre 1 e 4, e o número do pacote que é necessário, para saber qual a posição do pacote em relação à matriz de 160 x 120 que constitui a

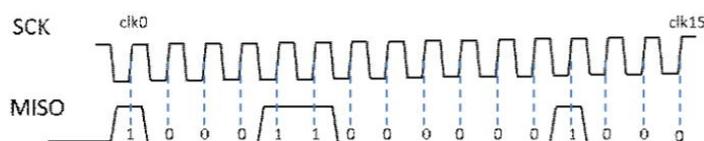


Figura 4.10: Exemplo de transmissão SPI Mode 3 do valor 0x8C08

imagem. Na Fig. 4.11 é possível compreender melhor como é composto um frame, com base nos seus segmentos e a localização dos pacotes. O primeiro bit do ID é sempre um zero e os três bits seguintes, chamados TTT bits, apenas são relevantes no pacote número 20, uma vez que codificam o número do segmento. Os 12 bits seguintes são o número do pacote. Caso o número do segmento no pacote 20 seja zero, significa que o pacote deve ser descartado.

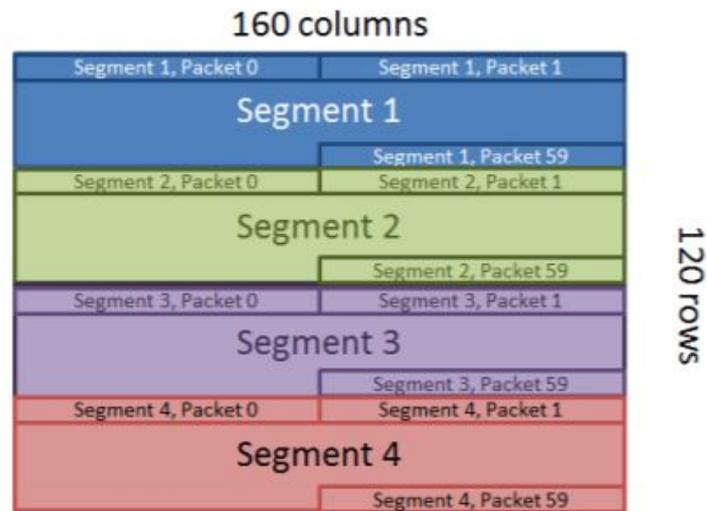


Figura 4.11: Organização dos segmentos e respectivos pacotes num frame, por defeito

O processo de sincronização é feito em quatro partes.

- Deseleccionar o CS e desativar o SCK por, pelo menos, 5 períodos de um frame (equivalente a cerca de 185 milisegundos). Este passo garante o timeout da interface VoSPI e coloca o Lepton no estado adequado para estabelecer a sincronização.
- Seleccionar o CS e ativar o SCK para iniciar a transmissão do primeiro pacote.
- Analisar o ID do pacote, identificar os pacotes descartáveis e ler os dados do pacote inteiro.
- Continuar a ler os pacotes e, quando um novo pacote está disponível, 10 milisegundos depois de desseleccionar o CS e ler o primeiro pacote, o primeiro pacote de vídeo é transmitido e o master e o slave estão sincronizados.

A dessincronização da comunicação ocorre quando se verifica uma das seguintes violações.

- Quando um pacote inicia a transmissão deve ser completado dentro de um período de 3 linhas. Assumindo que a frequência do VoSPI está corretamente seleccionada e que o CS não se encontra desseleccionado a meio da transferência de um pacote, esta violação não é expectável.

- Caso falhe a leitura de todos os pacotes de um dado frame antes do frame seguinte ficar disponível. Na Fig. 4.12 encontra-se um exemplo de dessincronização provocada pela falha da leitura de um frame completo antes da chegada do seguinte.
- Caso falhe a leitura de todos os frames disponíveis.

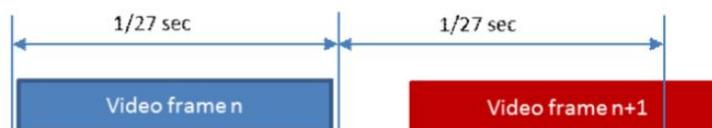


Figura 4.12: Exemplo de dessincronização

#### 4.2.2 Rede Neuronal

No programa *TFLite\_detection\_image.py* é usado uma rede neuronal treinada para detecção de objetos, com a framework TensorFlow Lite, para realizar a detecção de pessoas numa determinada imagem.

O processamento das imagens é feito utilizando a ferramenta OpenCV, começando por ler as imagens capturadas para a pasta *capture* com a função `cv2.imread()` e, posteriormente, é feita a sua redimensão com a função `cv2.imresize()`. Neste passo, as imagens provenientes da câmera RGB são redimensionadas para a resolução de 640 x 480 pixels e as da câmera térmica para 320 x 240 pixels. De seguida, as imagens são rodadas 90 graus no sentido contrário aos ponteiros do relógio com a função `cv2.rotate()`. Este passo é motivado por as câmeras estarem numa disposição, onde as imagens são capturadas com resolução vertical. Finalmente, as imagens são novamente redimensionadas para uma resolução compatível com o modelo da rede neuronal e, sem comprometer a imagem original, são guardadas numa nova variável.

Uma vez processadas, seguem-se os passos de análise das imagens de ambas as câmeras, RGB e térmica, utilizando para isso o modelo baseado numa rede neuronal SSD-MobilenetV2 quantizada. O modelo utilizado é previamente treinado com um dataset da COCO para detetar e identificar objetos que, no nosso sistema são exclusivamente pessoas. Relativamente à rede, esta é quantizada para uma resolução mais baixa, isto é, os pesos internos da rede são convertidos de floats para inteiros, utilizando a framework TensorFlow, para reduzir o custo computacional da operação.

Depois de passar uma imagem pela rede, esta retorna quatro parâmetros de saída relativamente à detecção de objetos na mesma.

- **boxes**, este parâmetro consiste num array com as coordenadas das bounding boxes dos objetos detetados na imagem

- **classes**, este parâmetro consiste num array com as classes detetadas na imagem, no nosso caso apenas interessa a classe correspondente às pessoas
- **scores**, este parâmetro representa o grau de confiança do modelo nas deteções efetuadas, no nosso programa apenas são consideradas deteções com um grau de confiança superior a 50%
- **num**, este parâmetro representa o número de deteções efetuadas e não está a ser utilizado, atualmente

Finalmente, os parâmetros de saída são utilizados para restringir a classe de interesse e o grau de confiança pretendido e, com as coordenadas das bounding boxes, é possível desenhar na imagem a localização da deteção com a função `cv2.rectangle()`. É ainda possível acrescentar junto à caixa de deteção, uma etiqueta com a classe detetada e o grau de confiança da deteção, como na Fig. 4.13, com a função `cv2.putText()`.

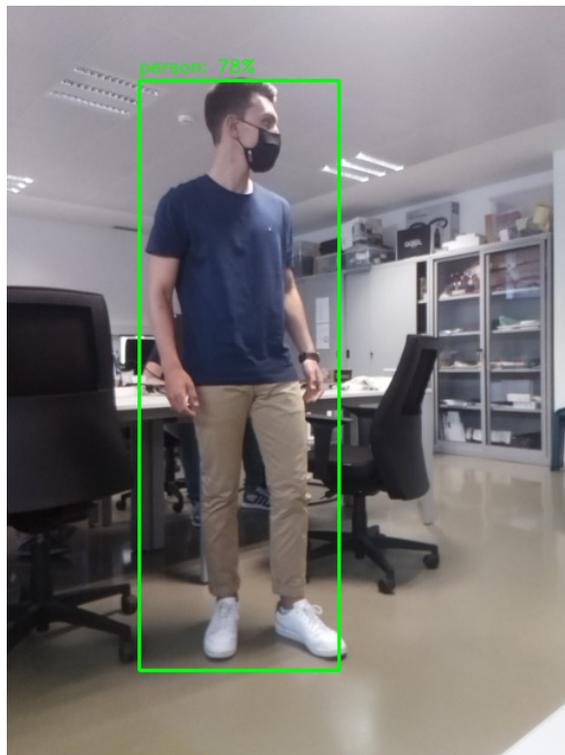


Figura 4.13: Exemplo de imagem processada após deteção

### 4.2.3 Microcontrolador

O microcontrolador ESP32 implementa uma máquina de estados, como é apresentado na Fig. 4.14, que permite controlar o movimento de rotação da caixa que aloca as duas câmeras, térmica e RGB, que realizam a deteção de pessoas, assim como o LiDAR que mede a área na zona em redor do sensor. Esta máquina de estados possui quatro diferentes estados com diferentes funções.

- *ps\_init*, um simples estado de inicialização que ativa imediatamente o estado seguinte de sincronização.
- *ps\_sync*, é o estado responsável por verificar a funcionalidade do motor e do sistema de rotação em geral, inicializa a contagem de passos do motor e transita para o estado seguinte de rotação ao fim de três segundos ou caso exista um sensor de fim de rotação, como por exemplo um sensor de efeito de Hall.
- *ps\_sweep*, é o estado de rotação onde é ativo o motor de passo e transita para o estado seguinte de paragem quando atinge um número de passos predefinido como equivalente ao número de paragens necessárias, neste caso o valor escolhido foi  $640 \times 4.7$  que corresponde a 10 paragens numa rotação. O sistema também pode ser forçado a parar quando recebe um comando de paragem por parte do Raspberry Pi via porta série.
- *ps\_stop*, no estado de paragem é decrementado o número de passos de forma a garantir que as paragens entre posições são realizadas de forma regular e igualmente espaçada e ocorre a transição de estado para uma novo movimento de semirotação quando se verifica uma de três situações. A primeira é quando recebe um sinal de fim de processamento por parte do sistema de deteção via porta série. A segunda é quando recebe um sinal do Raspberry Pi via porta série que força o sistema a retomar o movimento de rotação. A última é quando o tempo entre paragens excede os dois segundos, neste caso o sistema avança de forma a evitar que fique preso nesse estado.

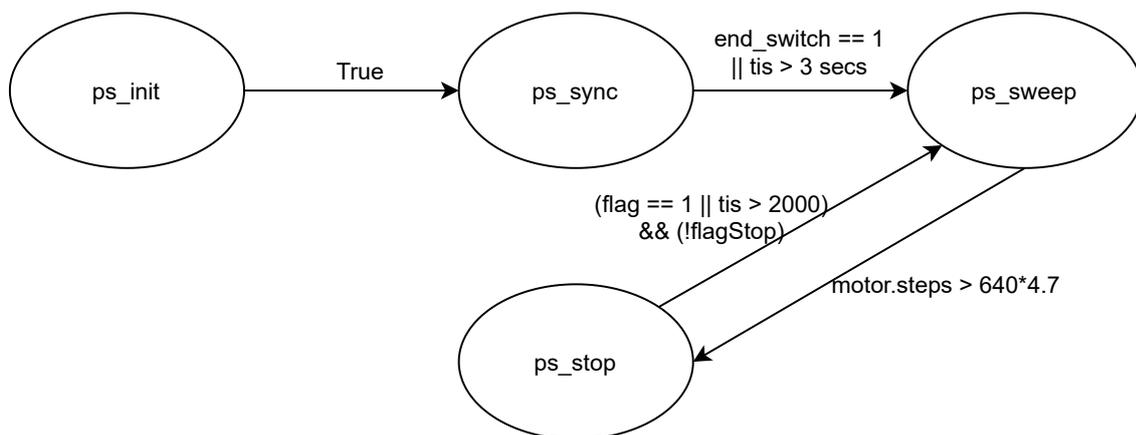


Figura 4.14: Diagrama de estados representativo do mecanismo de rotação

A definição dos estados e da velocidade de rotação do motor é feita, respetivamente, utilizando as funções `set_state()` com o parâmetro correspondente ao estado e `set_motor_speed()` com o parâmetro da velocidade que é, neste caso, zero ou um valor predefinido na variável `sweep_speed()`.

A leitura dos quatro sensores PIR é também da responsabilidade do microcontrolador, para tal é utilizada a função `pinMode()` com dois parâmetros, o primeiro é o pino do microcontrolador associado a um determinado sensor e o outro é o modo propriamente dito. Neste caso é escolhido o modo `INPUT_PULLUP` de forma a considerar o pino como entrada, tendo em conta que apenas vamos receber informações do sensor, e ativando a resistência de pull-up.

Por outro lado, a leitura dos sinais de entrada provenientes do robô de desinfecção são lidas com a função `PCF_IN.READ8()` e guardadas na variável `pcf_in_data`. Consoante o valor dos 8 bits lidos, o sistema pode realizar diferentes funções.

- **Bit 0 (menos significativo)**, envia um sinal de paragem da rotação do sistema de deteção
- **Bit 1**, envia um sinal de arranque da rotação do sistema de deteção
- **Bit 2**, envia um sinal de *shutdown* para o sistema de deteção

Existem ainda dois canais de comunicação série, um para comunicar com o PC em operações de debug, inicializado com o comando `Serial.begin(115200)` e outro para comunicar com o Raspberry Pi, inicializado com o comando `Serial.begin(115200, SERIAL_8N1, 19, 18)`. Ambas as comunicações são feitas a um *baud rate* de 115200 bits por segundo, no entanto, a segunda comunicação, por não ser realizada nas portas de comunicação nativas do microcontrolador, requer que seja definido o modo de comunicação. Neste caso é feita a transmissão de 8 bits sem paridade e com um stop bit e são utilizados os pinos 19 e 18 como recetor e transmissor, respetivamente.

É também a nível do microcontrolador que é tomada a decisão final em relação à deteção de pessoas, utilizando para esse efeito o conceito de fusão sensorial, como está exemplificado na Fig. 4.15. Para tal, são atribuídos pesos aos diferentes dados dos sensores, de forma arbitrária, com base no seu impacto na deteção. Caso a sua média seja superior ao limiar definido pelo utilizador, é então enviado um sinal para o computador do robô de desinfecção, sinalizando-o da deteção realizada. Os pesos estão distribuídos da seguinte forma:

- **Deteção em imagem RGB**, tem um peso de 30% uma vez que esta metodologia é das que apresenta resultados mais fiáveis e consistentes
- **Deteção em imagem térmica**, tem um peso de 20% visto que muito raramente faz deteções erradas utilizando a rede neuronal, no entanto, é mais suscetível a não realizar deteções quando comparado à deteção RGB
- **Histograma térmico**, tem um peso de 10% dado que é um método de deteção que não nos dá informação sobre a que objeto correspondem as temperaturas detetadas
- **Sensores PIR**, tem um peso de 40% uma vez que é um método de deteção ativado por qualquer tipo de movimento e cobre os 360 graus em torno do robô, no entanto, nem sempre é necessariamente ativado por movimentos humanos

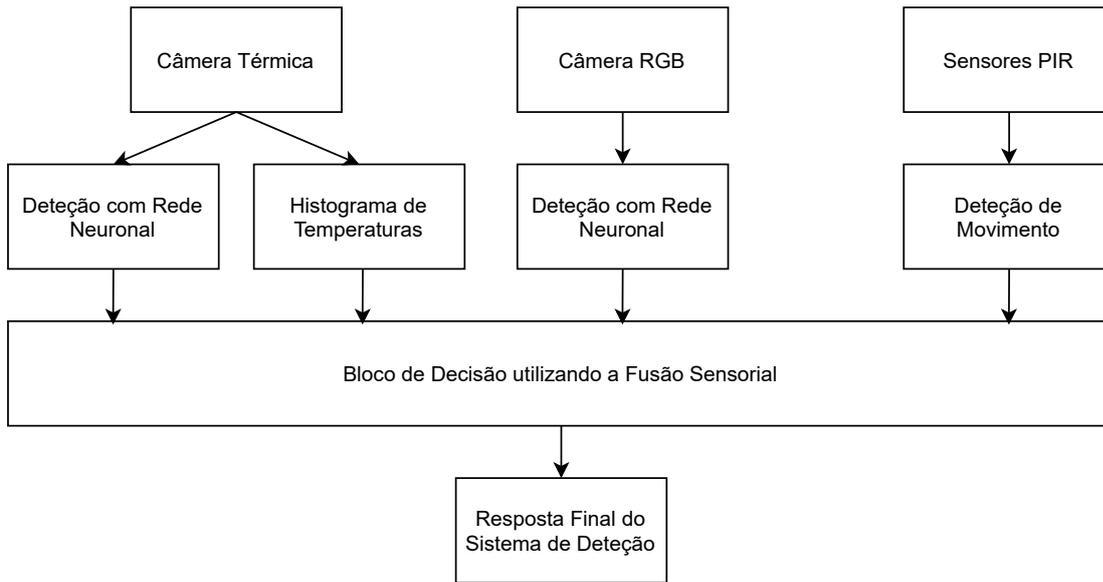


Figura 4.15: Estrutura de decisão do sistema de detecção de pessoas

- **Threshold de Detecção** está definido por defeito para 50%

Finalmente, os sinais para o robô de desinfecção são enviados com auxílio de um expensor de pinos, através da função `PCF_OUT.WRITE8(PCF_OUT_DATA)`, onde o parâmetro `pcf_out_data` corresponde a um valor de 8 bits e cada bit corresponde a uma informação de detecção.

- **Bit 0 (menos significativo)**, corresponde a uma detecção positiva com uma imagem térmica
- **Bit 1**, corresponde a uma detecção positiva com uma imagem RGB
- **Bit 2**, corresponde a uma detecção positiva com os sensores PIR
- **Bit 3**, corresponde a uma detecção positiva com base na fusão sensorial de todos os sensores

#### 4.2.4 Comunicação Assíncrona

De forma a realizar uma comunicação eficiente e rápida entre os programas que correm simultaneamente no Raspberry Pi, foi adotada uma biblioteca de mensagens assíncronas. Esta framework, chamada ZeroMQ ou 0MQ, é majoritariamente aplicada em sistemas distribuídos, fornecendo uma fila de mensagens. Ao contrário de middleware orientado a mensagens, um sistema ZeroMQ pode ser executado sem um agente de mensagens dedicado.

A arquitetura de comunicação da Fig. 4.16 representa a relação hierárquica entre os diferentes programas que correm em simultâneo no sistema de detecção. Para estabelecer a ligação entre os programas é necessário seguir um conjunto de passos:

- Criar um context através do comando `zmq.Context()`

- Criar um socket com esse context e definir o tipo de socket a utilizar com o comando `context.socket(zmq.TYPE)`
- De seguida, caso o socket seja um publisher é necessário fazer bind do socket, definindo o protocolo a utilizar, o IP do dispositivo com que se quer comunicar e a porta onde será feita a comunicação com o comando `socket.bind("protocol://ipaddress:port")`
- Caso o socket seja um subscriber é necessário fazer connect do socket, definindo o protocolo a utilizar, o IP do dispositivo com que se quer comunicar e a porta onde será feita a comunicação com o comando `socket.connect("protocol://ipaddress:port")`

Uma vez estabelecida a ligação, os sockets estão prontos a comunicar entre si. Dos cinco programas que correm simultaneamente no Raspberry Pi, podemos destacar dois como subscritores, dois como publicadores e um como publicador e subscritor, simultaneamente. Para além do modo utilizado pub-sub, existem ainda mais tipos de ligação entre sockets disponibilizados pela framework como fan-out, task distribution e request-reply. Relativamente aos protocolos de comunicação podem ser in-process, inter-process, multicast e TCP que foi a opção utilizada com o IP local da máquina. As porta utilizadas são a 5555, 5556 e a 5557.

As publicações são realizadas com o comando `socket.send_string()` para enviar mensagens que serão posteriormente subscritas pelo programa de subscrição respetivo, utilizando o comando `socket.recv()`. Do lado do publicador, a mensagem ZeroMQ é composta por um tópico e o conteúdo da mensagem que serão enviados sempre de forma conjunta. Do lado do subscritor pode-se subscrever todos os tópicos existentes ou restringir àqueles que são de interesse com a função `socket.setsockopt_string(zmq.SUBSCRIBE, topicfilter)`, onde o `topicfilter` é o tópico que se pretende escutar. No programa *sub.py* foram definidos quatro tópicos para quatro tipos de dados diferentes:

- **Tópico 5001**, refere-se ao histograma calculado a partir das imagens térmicas
- **Tópico 5002**, refere-se à deteção de pessoas por parte da rede neuronal utilizando imagens térmicas
- **Tópico 5003**, refere-se à deteção de pessoas por parte da rede neuronal utilizando imagens RGB
- **Tópico 5004**, refere-se à área calculada em metros quadrados pelo LiDAR

Os programas *webcam.py* e *LeptonThread.cpp* subscrevem todos os tópicos e quando recebem a mensagem *capture*, proveniente do programa *TFLite\_detection\_image.py*, fazem a captura de uma imagem em cada uma das câmeras e guardam-nas numa pasta própria, servindo a mensagem como um sinal de sincronização para todo o sistema. Para além da captura da imagem, o programa da câmara térmica publica ainda o histograma das temperaturas calculado para o mesmo frame.

O programa *tfmini.py*, referente às tarefas do LiDAR, também subescreve todos os tópicos e, aquando da receção da mensagem de captura sincronizada, recolhe a informação relativa à distância medida naquele momento e converte-a para uma medição de área.

$$A = \frac{(\pi \times r^2)}{N} \quad (4.1)$$

A área calculada, *A*, diz respeito a uma de *N* parcelas da área de uma circunferência de raio, *r*, e é igual à soma da distância medida pelo LiDAR e a distância do LiDAR ao centro do sensor. As *N* parcelas correspondem ao número de paragens do sensor, numa volta.

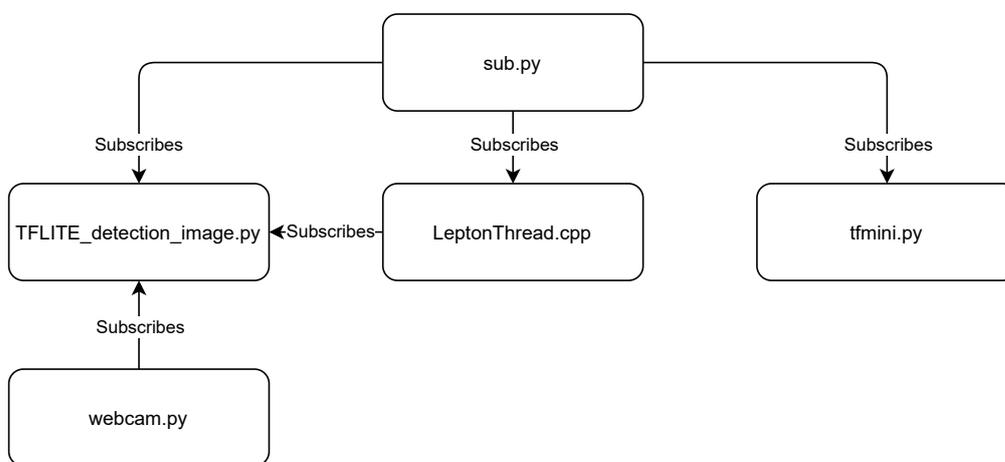


Figura 4.16: Arquitetura de comunicação da framework ZeroMQ no sistema

Após a subscrição de todas as mensagens relevantes para o sistema, por parte do programa *sub.py*, segue-se o passo final que consiste em enviar, via porta série, para o microcontrolador, os dados obtidos por cada um dos programas de deteção e, ainda os dados do LiDAR.

## Capítulo 5

# Resultados e Discussão

Dada a impossibilidade de realizar os testes necessários para validar o sistema desenvolvido em ambiente hospitalar, estes foram efetuados em dois ambientes distintos. O primeiro no Departamento de Engenharia Eletrotécnica e Computador da Faculdade de Engenharia da Universidade do Porto, mais especificamente no laboratório do Centro de Robótica Industrial e Sistemas Inteligentes, enquanto que o segundo local de testes foi no iiLab, Laboratório de Indústria e Inovação pertencente ao Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência.

Os testes tinham como objetivo validar a fiabilidade do sistema de deteção como um todo, ao mesmo tempo que se verificava o comportamento dos diferentes métodos de deteção desenvolvidos, nas mesmas situações. Idealmente, os sistemas de deteção baseados em visão iriam fazer deteções com maior precisão, quando comparados à deteção feita pelos sensores PIR, dada a maior disponibilidade de recursos e poder computacional envolvidos no processo.

A avaliação dos resultados experimentais foi feita com base num método de classificação de deteção de objetos, onde são propostos quatro níveis de avaliação. O primeiro, chamado Verdadeiro Positivo, representa um cenário onde exista pelo menos uma pessoa na zona de deteção e o sistema realiza uma deteção positiva, confirmando de forma correta a existência de pessoas nas proximidades do robô. Por outro lado, um Falso Positivo representa um cenário onde não existam pessoas na zona de deteção e o sistema, erradamente, indica que foi detetada uma pessoa. Existem ainda mais dois cenários considerados para este método de deteção. O Verdadeiro Negativo que representa uma situação onde não existam pessoas presentes na zona de deteção e o sistema não realiza qualquer tipo de deteção, como seria expectável. E, finalmente, o Falso Negativo que é um cenário onde pelo menos uma pessoa se encontra presente na zona de deteção do sistema e não é realizada nenhuma deteção, verificando-se portanto, uma falha do sistema ao não detetar qualquer pessoa.

Para testar o comportamento do sistema em ambientes próximos da realidade, começou por se realizar um teste no laboratório do CRIIS ainda sem o robô de desinfeção. A ideia destes primeiros testes passava por compreender o comportamento dos diferentes sistemas de deteção, nomeadamente o comportamento da rede de deteção de objetos com a câmara RGB e com a câmara térmica. O sistema realiza a aquisição de uma imagem a cada paragem e realiza dez paragens por

cada rotação completa, igualmente espaçadas de 36 graus. Na Fig. 5.1 e na Fig. 5.2, podemos observar um exemplo das dez imagens obtidas pela câmara térmica e pela câmara RGB durante uma rotação, respetivamente, de forma sequencial e de forma coincidente entre as duas câmeras.



Figura 5.1: Exemplo das dez capturas de imagem numa rotação na câmara térmica

Relativamente ao comportamento dos quatro sensores PIR presentes na zona estática da caixa de deteção, foi possível identificar algumas falhas no seu desempenho. Este conjunto de sensores resulta em Falsos Negativos, com alguma frequência, em cenários algo específicos em que existe uma pessoa junto ao sensor, mas esta permanece completamente imóvel. A nível de Verdadeiros Positivos o sistema é bastante eficiente a identificar movimentos de pessoas que se encontrem num alcance considerável, entre cinco e seis metros do sensor. Os Verdadeiros Negativos foram comprovados, deixando o sistema durante algum tempo sem qualquer pessoa junto a ele e, como seria de esperar, não foi realizada qualquer tipo de deteção. Finalmente, os Falsos Positivos não tiveram grande expressão nos testes realizados, uma vez que não existiam corpos móveis não humanos que pudessem confundir o sistema.

Em relação ao comportamento da deteção com base nas imagens térmicas, o desempenho da mesma, como era expectável, foi significativamente melhor, quando comparado com o dos sensores PIR. A principal falha detetada neste método, prende-se com o facto de a rede neuronal de deteção, uma vez que não foi treinada com imagens térmicas, não ter uma robustez de identificação tão grande como quando comparado à câmara RGB. Verifica-se em algumas das imagens térmicas que, embora seja perceptível ao olho humano a existência de uma pessoa, como podemos observar na Fig. 5.3, a rede não é capaz de realizar a deteção da mesma. No entanto, uma vez que a deteção realizada com as imagens térmicas tem duas camadas de deteção, podemos sempre recorrer ao histograma de temperaturas das mesmas, para verificar se existem valores de temperatura que se encontrem dentro da gama considerada como aceitável para a temperatura do corpo

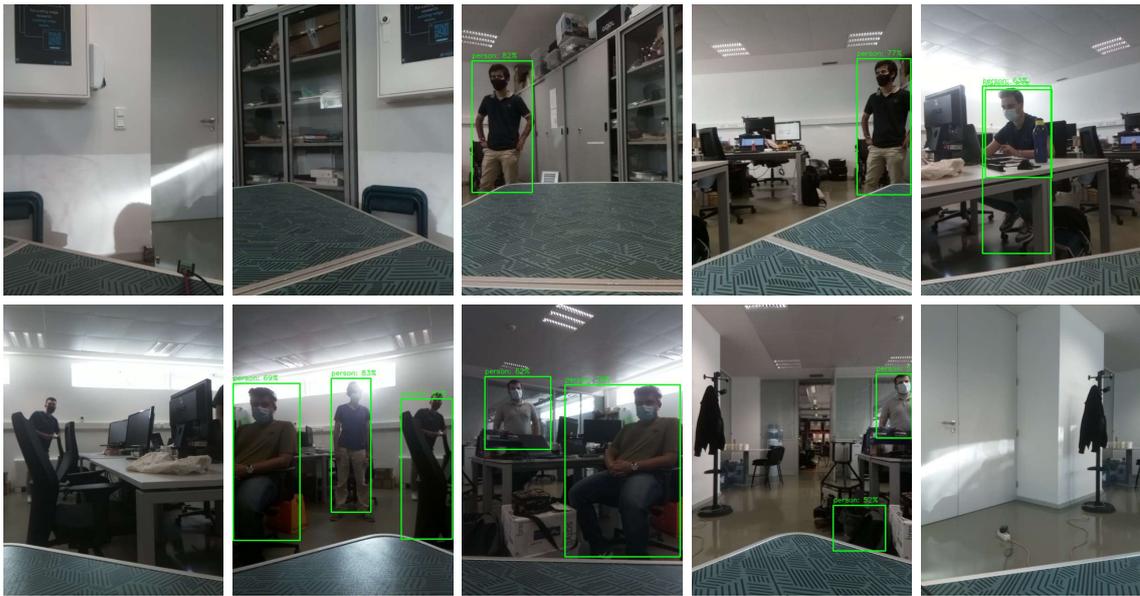


Figura 5.2: Exemplo das dez capturas de imagem numa rotação na câmera RGB

humano. Obviamente, este método também está sujeito a Falsos Positivos, uma vez que podem existir corpos que estejam a uma temperatura que coincida com a gama de temperaturas humanas considerada, Fig. 5.4, mas não sejam corpos humanos.

Assim sendo, de forma a prever os erros identificados por cada um dos métodos de deteção, deve-se considerar a resposta do sistema através da combinação de ambos os métodos. Nunca se tendo verificado um cenário em que a rede tenha identificado uma pessoa numa imagem térmica e o histograma de temperaturas correspondente não tenha identificado temperaturas dentro da gama considerada. De igual forma, nunca se verificou um cenário em que a rede tenha detetado uma pessoa numa imagem térmica que não estivesse presente ou onde o histograma de temperaturas fosse zero. Os valores seguintes representam os graus de confiança relativos apenas à deteção utilizando a câmera térmica.

- Deteção da rede e deteção do histograma, aproximadamente 100% certeza de deteção
- Deteção do histograma e não deteção da rede, aproximadamente 33% certeza de deteção
- Não deteção do histograma e não deteção da rede, 0% certeza de deteção

No que diz respeito à deteção com base nas imagens RGB, verifica-se que o desempenho da rede de deteção é superior, comparativamente ao uso da rede com imagens térmicas. No entanto, este método de deteção tem apenas essa camada de resultados e, portanto, não há maneira de comprovar a fiabilidade da deteção, que não seja através da combinação dos resultados deste método com os de outro método. Algumas falhas foram identificadas como Falsos Positivos, para algumas situações específicas, como por exemplo objetos semelhantes a pessoas, como casacos ou calças que estivessem pendurados em cabides Fig. 5.5. Neste caso é possível verificar que,



Figura 5.3: Falso Negativo na detecção com imagem térmica



Figura 5.4: Falso Positivo na detecção com histograma da imagem térmica

um objeto que não representa um humano, é erradamente detetado como tal. A falha encontrada no sistema está relacionada com o facto da rede neuronal estar treinada para detetar pessoas de uma forma, que não exija necessariamente uma imagem completa da pessoa da cabeça aos pés. A rede interpreta as imagens de peças de roupa, como se de uma imagem parcial de uma pessoa se estivesse a tratar.

Outra situação considerada como Falso Positivo ocorre quando o sistema capta uma imagem, onde esteja presente um cartaz ou uma fotografia com uma pessoa Fig. 5.6 e a rede não possui capacidade, nessas situações, de distinguir uma pessoa que esteja fisicamente presente, de uma pessoa que não esteja fisicamente presente.

Relativamente aos resultados identificados como Falsos Negativos estão associados geralmente a dois fatores. O primeiro é essencialmente em situações de baixa luminosidade, onde a câmara não é capaz de captar uma imagem suficientemente nítida para ser identificada por parte da rede Fig. 5.7. O segundo fator que motiva os Falsos Negativos relaciona-se com situações em que as pessoas na imagem, não estão próximas o suficiente do sistema Fig. 5.8 e, portanto, a rede não tem capacidade de identificar as pessoas. Esta distância limite situa-se geralmente entre dez a quinze metros do robô consoante as condições de luminosidade. Finalmente, os Verdadeiros Positivos foram obtidos numa grande variedade de situações com diferentes condições de luminosidade, distância, quantidade de pessoas, em movimento e estáticas, de forma parcial e completa, entre outros fatores. Estes resultados comprovam a fiabilidade deste método de deteção em diferentes cenários e, como tal, este método representa a principal fonte de informação do sistema.



Figura 5.5: Falso Positivo na deteção com imagem RGB com um objeto semelhante a uma pessoa

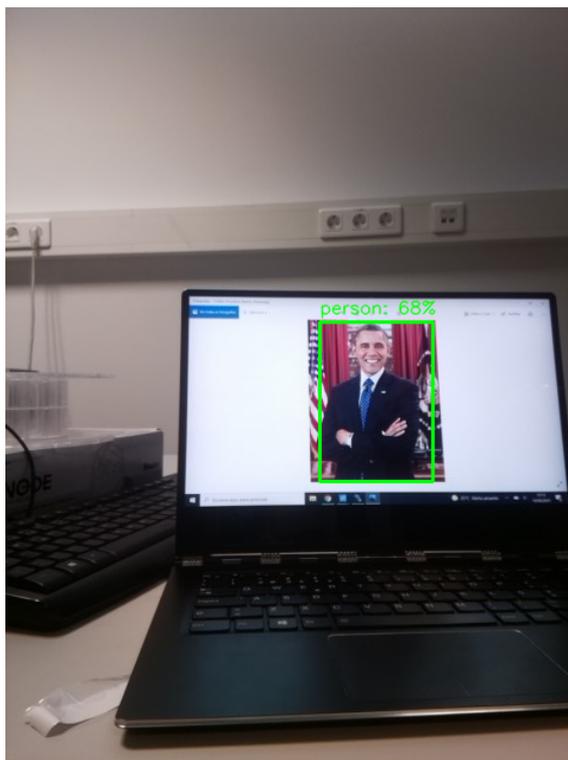


Figura 5.6: Falso Positivo em imagem RGB com uma fotografia de uma pessoa não presente

Por fim, uma vez compreendido o funcionamento individual dos diferentes métodos de detecção implementados, foi então testado o algoritmo de fusão sensorial dos diferentes métodos de detecção, em simultâneo. Este algoritmo resulta da combinação dos resultados dos diferentes subsistemas, de maneira a obter uma resposta única e final por parte do sistema. Esta decisão final é realizada a nível do microcontrolador, com base nos resultados obtidos pelos quatro sensores PIR associados, bem como os dados das câmeras térmica e RGB, provenientes do Raspberry Pi. O algoritmo de fusão sensorial calcula o grau de confiança da detecção, com base nos pesos atribuídos a cada método de detecção, o que corresponde à fusão dos dados provenientes das diferentes metodologias. Caso esse nível seja superior ao limiar estabelecido pelo utilizador, a detecção é considerada positiva.

A Tabela 5.1 resume os resultados obtidos por cada um dos métodos de detecção, identificando não só os seus pontos fortes, mas também os seus pontos fracos.

Assim sendo, o sistema foi colocado nas mesmas condições em que foram realizados os testes anteriores, de maneira a que seja possível verificar que a resposta combinada dos diferentes métodos de detecção apresenta vantagens, quando comparado à utilização individual dos mesmos. Os pesos associados a cada método de detecção, assim como o limiar de detecção utilizado, foram os valores definidos por defeito.

Em situações de resultado positivo, por parte da detecção em imagens RGB, podemos recorrer à utilização do método de detecção em imagens térmicas, para verificar que o resultado obtido se trata, de facto, de um Verdadeiro Positivo como na Fig. 5.9. Por outro lado, caso se trate de um

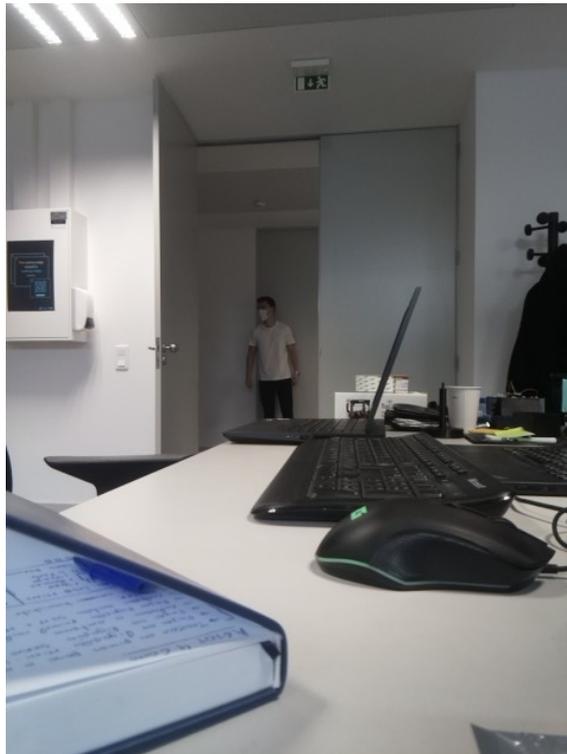


Figura 5.7: Falso Negativo na deteção com imagem RGB em baixa luminosidade

Método de Deteção	Vantagens	Desvantagens
Deteção em imagens RGB	Facilidade de deteção de pessoas	Não é capaz de distinguir pessoas reais de não reais
	Pode localizar a deteção numa imagem	Deteta peças de roupa como sendo uma pessoa com frequência
Deteção em imagens Térmicas	Ideal para situações de baixa luminosidade	Falha mais deteções que nas imagens RGB
	Pode localizar a deteção numa imagem	
Histograma de Temperaturas	Deteta temperaturas de corpos dentro de uma gama predefinida	Não pode localizar uma deteção numa imagem
		Não é capaz de identificar a origem da temperatura detetada
Deteção de Movimento	Amplitude do campo de deteção de 360°	Não deteta objetos imóveis
	Baixo custo computacional e energético	Deteta movimentos não humanos

Tabela 5.1: Resumo das vantagens e desvantagens dos métodos de deteção



Figura 5.8: Falso Negativo na detecção com imagem RGB com uma pessoa fora do alcance do sistema

Falso Positivo por parte da deteção na imagem RGB e não existam pessoas na zona de deteção do sistema, como podemos comprovar na Fig. 5.10, a deteção na imagem térmica não iria ser sinalizada e o sistema não resultaria numa deteção verdadeira.

De forma contrária, a deteção em imagens RGB também poderá ajudar a comprovar a veracidade dos resultados obtidos pela deteção em imagens térmicas, quer seja em caso de deteção pela rede neuronal, quer seja em caso de deteção através do histograma ou até mesmo em casos em que ambos os métodos sinalizem a deteção. No entanto, como já tinha sido comprovado, quando ambas as camadas de deteção em imagens térmicas resultam num positivo, a probabilidade do resultado da deteção estar correto é elevada. O complemento com os resultados da deteção em imagens RGB simplesmente acrescentam uma camada extra de segurança e fiabilidade ao sistema.



Figura 5.9: Verdadeiro Positivo na deteção em imagem RGB com confirmação térmica

Um outro cenário onde se verificou que a utilização combinada dos diferentes subsistemas torna a deteção por parte do sistema mais robusta, foi ao testar num ambiente com pessoas imóveis. Neste caso, os sensores PIR não identificaram qualquer presença de pessoas na zona de deteção, quando na verdade existia mais que uma pessoa presente. No entanto, devido à ação combinada do sistema, tanto a deteção em imagens RGB Fig. 5.11, como a deteção em imagens térmicas Fig. 5.12 resultaram em deteções positivas. Como tal, o sistema foi capaz de sinalizar este cenário como estando na presença de pessoas. Tal desfecho não teria sido possível utilizando apenas os sensores PIR e teria sido certamente menos fiável, caso tivessem sido simplesmente utilizados os sensores PIR em conjunto com o método de deteção em imagens RGB ou os sensores PIR em conjunto com o método de deteção em imagens térmicas. Fica assim uma vez mais comprovado

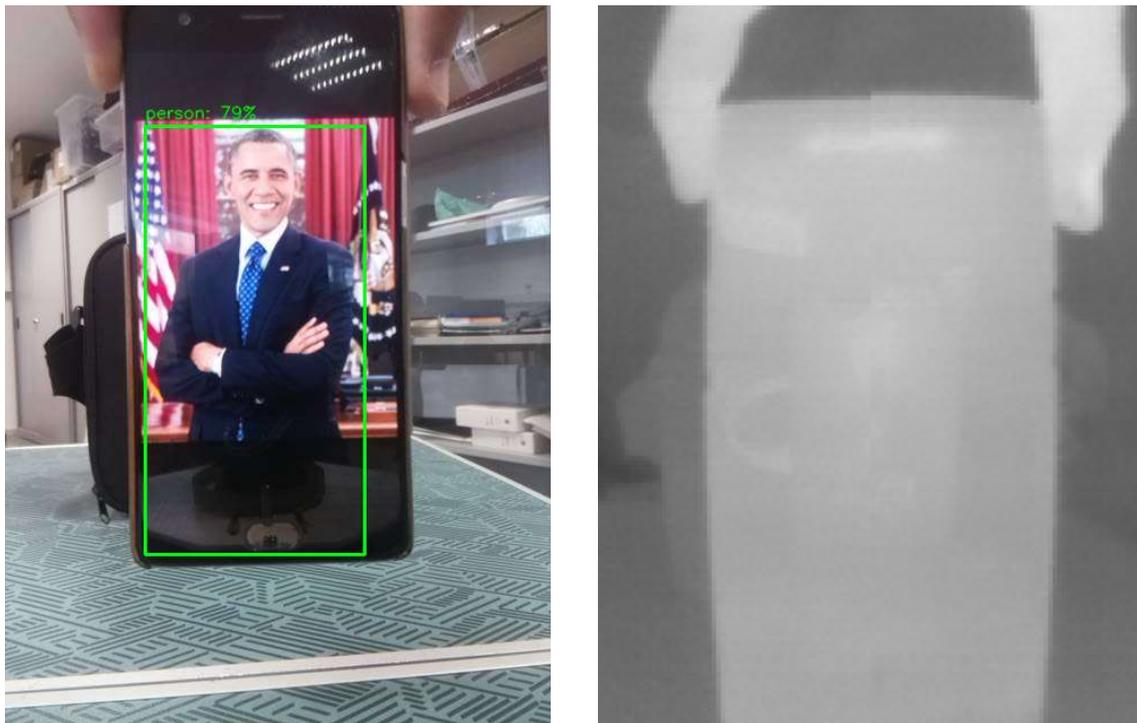


Figura 5.10: Falso Positivo na detecção em imagem RGB com confirmação térmica

que a ação combinada dos três métodos de detecção é sempre mais robusta que a ação individual dos mesmos ou até mesmo quando são utilizados apenas dois deles.

Por outro lado, em situações onde o movimento das pessoas perto do robô de desinfecção coincida com o movimento de rotação do sensor de detecção, constata-se que podem ocorrer falhas de detecção a nível das câmeras do módulo superior. Ao contrário das falhas mencionadas anteriormente, estas são provocadas pela ausência de pessoas nas imagens captadas pelas câmeras. Uma vez que as câmeras, RGB e térmica, apenas têm um ângulo de visão de 36 graus em cada paragem, num cenário em que as pessoas se mantenham constantemente em movimento na zona morta do sensor, não é possível recorrer aos métodos de detecção baseados em visão ou em temperaturas. É maioritariamente nestas situações que se recorre aos sensores PIR para realizar a detecção de pessoas. Estes garantem que, embora o sistema possua ângulos mortos, continua a ser capaz de realizar a detecção, em qualquer direção à volta do robô de desinfecção.

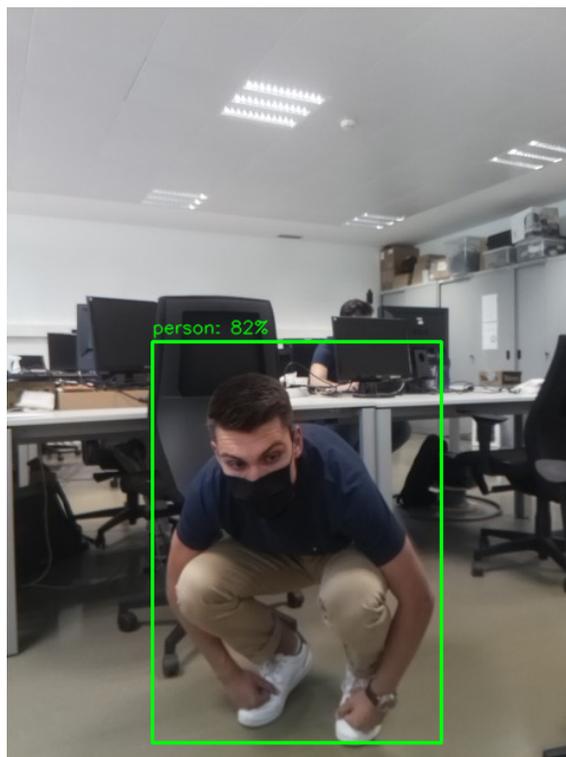


Figura 5.11: Detecção positiva por confirmação com imagem RGB e térmica



Figura 5.12: Detecção positiva por confirmação com imagem térmica



## Capítulo 6

# Conclusões e Trabalho Futuro

Nesta Dissertação de Mestrado foi desenvolvido um sistema inteligente de detecção de pessoas para robôs de desinfecção. O sistema previne a exposição de humanos a radiação ultravioleta nociva, alertando o robô sempre que exista a possibilidade de existir uma pessoa na zona de atuação do mesmo. Foi realizada a fusão sensorial de diferentes sensores de modo a garantir uma detecção mais fiável e robusta, atingindo níveis de confiança superiores quando comparado à utilização individual dos mesmos sensores.

A arquitetura de hardware e software desenvolvida permitiu isolar os três grandes níveis do sistema, sem comprometer a sua utilização em simultâneo e com capacidade de comunicação entre as diferentes camadas do mesmo.

A primeira camada do sistema consiste num módulo rotativo, composto por um computador de dimensões reduzidas, duas câmaras, uma térmica e uma RGB e ainda um LiDAR. Foi desenvolvido um sistema controlado por um Raspberry Pi que gere situações de recolha de dados dos sensores, processamento dos mesmos e, finalmente, o envio desses dados para a segunda camada do sistema. Através da utilização de uma câmara térmica e uma câmara RGB, foi possível utilizar uma rede de identificação e detecção de objetos, com o objetivo de detetar pessoas. Os testes inicialmente realizados comprovaram que a detecção nas imagens de cada uma das câmaras apresentava algumas falhas, nomeadamente falta de resolução na câmara térmica e fraco desempenho em baixa luminosidade na câmara RGB. Além disso, o LiDAR foi implementado para realizar medições da área onde se encontra o robô, na direção onde o sistema está a executar a detecção de pessoas.

A segunda camada do sistema consiste num módulo de controlo, composto por um PCB com componentes que gerem diferentes tarefas, desde o controlo do motor de passo, até a comunicação com o Raspberry Pi e o robô de desinfecção. Este módulo possui ainda o motor de passo que permite o movimento rotativo do módulo superior e um sonar para medições da distância do sensor ao teto, por motivos de segurança e um conjunto de sensores PIR para detecção de movimento. Estes últimos sensores são o nível mais básico do sistema de detecção de pessoas.

A nível de hardware, dada a disponibilidade por parte do laboratório do CRIIS em realizar a impressão 3D, todos os componentes estruturais do sistema são compostos por peças impressas

em materiais de impressão 3D. Eventualmente, de forma a garantir uma maior rigidez e durabilidade, a estrutura poderia ser desenvolvida em alumínio ou outro material de maior resistência que não comprometa a premissa da portabilidade e facilidade de integração do sistema de detecção. Também a utilização de um outro computador de dimensões reduzidas, em alternativa ao Raspberry Pi, ou a utilização de um acelerador TPU permitiria um desempenho mais rápido ao sistema em termos de processamento das imagens passadas à rede de detecção. As câmeras utilizadas, embora apresentem resultados satisfatórios, também poderão ser substituídas por outras com maior resolução para melhorar o desempenho da detecção.

A nível de software foi possível concluir, através dos testes efetuados ao sistema em diferentes condições, que a fusão sensorial dos diferentes dados dos sensores disponíveis no sistema permite um desempenho na detecção de pessoas muito superior, quando comparado com métodos que utilizem apenas um dos sensores para essa tarefa. Verificou-se ainda que os valores dos pesos atribuídos a cada método de detecção, assim como o valor do *threshold* de detecção, devem ser ajustados de acordo com a situação em que o robô de desinfeção estiver a ser utilizado.

Em situações onde o robô de desinfeção atua em áreas em que seja previsível a existência de um maior volume de pessoas em movimento, como por exemplo num aeroporto ou um hospital, os sensores PIR devem ter um peso maior na detecção, quando comparados aos outros métodos, uma vez que este é o cenário ideal para a sua utilização. Nesta situação, o *threshold* de detecção pode ser, inclusivamente, ajustado para o mesmo valor do peso dos sensores PIR.

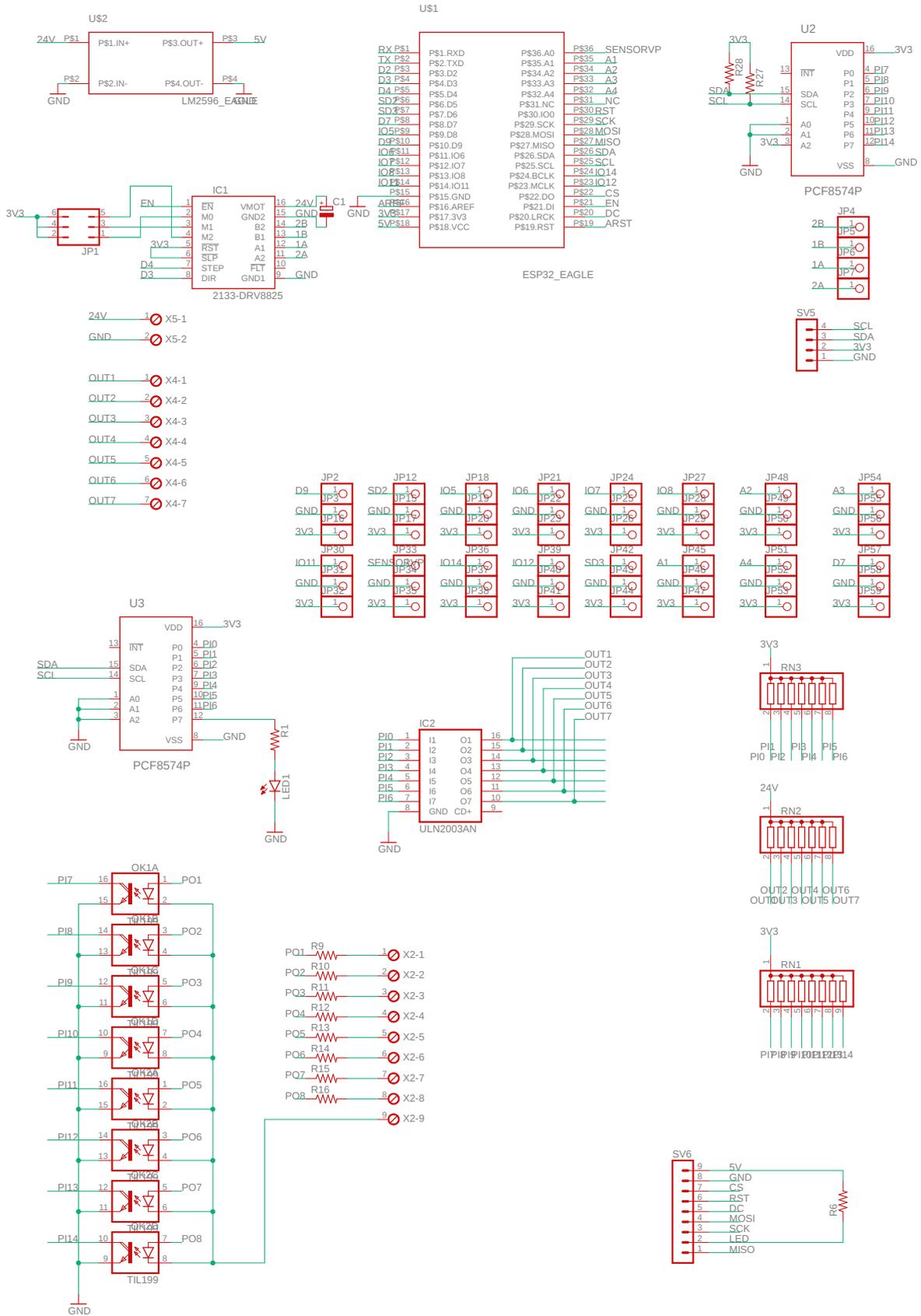
Por outro lado, em situações onde não exista grande movimentação de pessoas, o valor do *threshold* de detecção deve ser ajustado para um valor mais elevado, que tenha em consideração não só os sensores PIR, mas também pelo menos um dos outros métodos de detecção. O mesmo poderá ser dito de situações de baixa luminosidade, como por exemplo desinfeções noturnas, onde não só deverá ser ajustado o valor do *threshold*, mas também deverá ser atribuído um maior peso aos métodos de detecção térmicos e menos peso aos métodos de detecção RGB e aos sensores PIR.

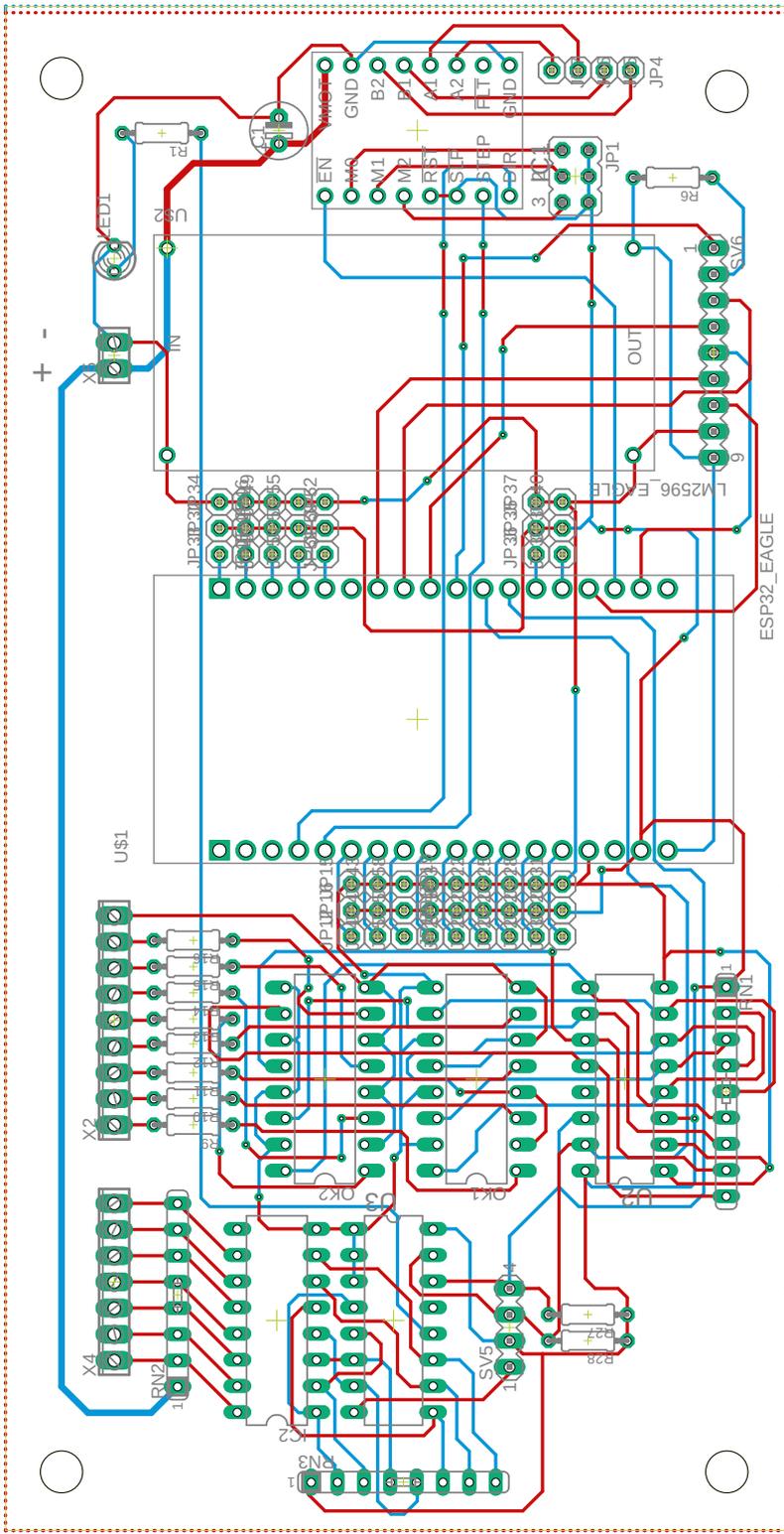
Conclui-se, portanto, que os objetivos principais da dissertação foram cumpridos, na medida em que foi desenvolvida um sistema inteligente de detecção de pessoas, com a capacidade de garantir a segurança de pessoas num ambiente de atuação de um robô de desinfeção. A fusão sensorial dos dados dos sensores ficou comprovada como uma vantagem, quando comparada à detecção individual por parte de cada um dos sensores.

O trabalho futuro consiste em implementar e testar a solução proposta em ambientes reais de desinfeção como hospitais, por exemplo, tendo em conta que tal não foi possível dadas as condições atuais. Em termos de software, seria também interessante treinar uma rede neuronal, exclusivamente para a tarefa de detecção de pessoas em imagens térmicas, visto que a rede utilizada para ambas as imagens, foi treinada para imagens RGB. A combinação e apresentação das imagens térmicas e das imagens RGB também foi uma tarefa parcialmente concluída e a qual seria interessante concluir na totalidade.

## **Anexo A**

# **Esquema elétrico do PCB**







# Referências

- [1] Jaeseok Yun e Sang Shin Lee. Human movement detection and identification using pyroelectric infrared sensors. *Sensors (Switzerland)*, 14(5):8057–8081, 2014. doi:10.3390/s140508057.
- [2] IEEE Xplore Full-Text PDF:. URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4109046>.
- [3] Yogesh Pawar, Abhay Chopde, e Mandar Nandre. Motion Detection Using PIR Sensor. *International Research Journal of Engineering and Technology*, 2018. URL: [www.irjet.net](http://www.irjet.net).
- [4] Liangliang Yang e Noboru Noguchi. Human detection for a robot tractor using omni-directional stereo vision. *Computers and Electronics in Agriculture*, 89:116–125, 11 2012. URL: <http://dx.doi.org/10.1016/j.compag.2012.08.011>, doi:10.1016/j.compag.2012.08.011.
- [5] Robson Marinho e Da Silva. Evaluation of mobile robot localization system’s accuracy based on sensor data fusion technique Balanced Automation Systems View project Intelligent Systems and Robotics View project. URL: <https://www.researchgate.net/publication/343646784>, doi:10.29327/118637.1-1.
- [6] Eva Besada-Portas, Jose A Lopez-Orozco, e Jesus M De La Cruz. Unified Fusion System Based on Bayesian Networks for Autonomous Mobile Robots. Relatório técnico, 2002.
- [7] IEEE Xplore Full-Text PDF:. URL: [https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4408178&casa\\_token=\\_Wz2Cg41t88AAAAA:8TP5C\\_elHYCRg3HSmO8AmdnJsI9jxyIc3kHi2LVw8KUtN6\\_htJI6HSW1Bu0xndR9zJc2k0G8xQ](https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4408178&casa_token=_Wz2Cg41t88AAAAA:8TP5C_elHYCRg3HSmO8AmdnJsI9jxyIc3kHi2LVw8KUtN6_htJI6HSW1Bu0xndR9zJc2k0G8xQ).
- [8] Khirod Chandra Sahoo e Umesh Chandra Pati. IoT based intrusion detection system using PIR sensor. Em *RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings*, volume 2018-January, páginas 1641–1645. Institute of Electrical and Electronics Engineers Inc., jul 2017. doi:10.1109/RTEICT.2017.8256877.
- [9] Priya B Patel, M Tech Student, Viraj M Choksi, Swapna Jadhav, e M B Potdar. ISSN : 2249-0868 Foundation of Computer Science FCS. Relatório técnico 5, 2016. URL: [www.ijais.org](http://www.ijais.org).
- [10] Anna A Trofimova, Andrea Masciadri, Fabio Veronese, e Fabio Salice. Indoor human detection based on thermal array sensor data and adaptive background estimation. *Journal of Computer and Communications*, 5:16–28, 2017. URL: <http://www.scirp.org/journal/jcc>, doi:10.4236/jcc.2017.54002.

- [11] Eun Som Jeon, Jong Hyun Kim, Hyung Gil Hong, Ganbayar Batchuluun, e Kang Ryoung Park. Human detection based on the generation of a background image and fuzzy system by using a thermal camera. 2016. URL: [www.mdpi.com/journal/sensors](http://www.mdpi.com/journal/sensors), doi:10.3390/s16040453.
- [12] Mauricio Correa, Gabriel Hermosilla, Rodrigo Verschae, Javier Ruiz-Del-Solar, M Correa, · G Hermosilla, · R Verschae, · J Ruiz-Del-Solar, G Hermosilla, e J Ruiz. Human detection and identification by robots using thermal and visual information in domestic environments. *J Intell Robot Syst*, 66:223–243, 2012. doi:10.1007/s10846-011-9612-2.
- [13] Petar Andraši, Tomislav Radišić, Mario Muštra, e Jurica Ivošević. Night-time detection of uavs using thermal infrared camera. volume 28, páginas 183–190. Elsevier B.V., 1 2017. doi:10.1016/j.trpro.2017.12.184.
- [14] Widodo Setyo Yuwono, Dodi Wisaksono Sudiharto, e Catur Wirawan Wijutomo. Design and implementation of human detection feature on surveillance embedded ip camera. páginas 42–47. Institute of Electrical and Electronics Engineers Inc., 7 2018. doi:10.1109/SIET.2018.8693180.
- [15] Marina Ivašić-Kos e Miran Pobar. Human detection in thermal imaging using yolo. ACM, 2019. URL: <https://doi.org/10.1145/3323933.3324076>.
- [16] Daniel Garigali Pestana, Fabio Mendonca, e F. Morgado-Dias. A low cost fpga based thermal imaging camera for fault detection in pv panels. Institute of Electrical and Electronics Engineers Inc., 8 2017. doi:10.1109/IoTGC.2017.8008976.
- [17] Y. Benezeth, B. Emile, H. Laurent, e C. Rosenberger. Vision-based system for human detection and tracking in indoor environment. *International Journal of Social Robotics*, 2(1):41–52, 2010. doi:10.1007/s12369-009-0040-4.
- [18] Alexander Carballo, Akihisa Ohya, e Shin'ichi Yuta. People detection using range and intensity data from multi-layered laser range finders. páginas 5849–5854, 2010. doi:10.1109/IROS.2010.5649769.
- [19] Jinshi Cui, Hongbin Zha, Huijing Zhao, e Ryosuke Shibasaki. Tracking multiple people using laser and vision. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, páginas 2116–2121, 2005. doi:10.1109/IROS.2005.1545159.
- [20] Efstathios P. Fotiadis, Mario Garzón, e Antonio Barrientos. Human detection from a mobile robot using fusion of laser and vision information. *Sensors (Switzerland)*, 13(9):11603–11635, 2013. doi:10.3390/s130911603.
- [21] Cristiano Premebida, Oswaldo Ludwig, e Urbano Nunes. LIDAR and vision-based pedestrian detection system. *Journal of Field Robotics*, 26(9):696–711, sep 2009. URL: <http://doi.wiley.com/10.1002/rob.20312>, doi:10.1002/rob.20312.
- [22] Joseph Redmon, Santosh Divvala, Ross Girshick, e Ali Farhadi. You only look once: Unified, real-time object detection. URL: <http://pjreddie.com/yolo/>.
- [23] Dumitru Erhan, Christian Szegedy, Alexander Toshev, e Dragomir Anguelov. Scalable object detection using deep neural networks.

- [24] H. Wei, M. Laszewski, e N. Kehtarnavaz. Deep learning-based person detection and classification for far field video surveillance. Institute of Electrical and Electronics Engineers Inc., 1 2019. doi:10.1109/DCAS.2018.8620111.
- [25] Catalin Alexandru Mitrea, Mihai-Gabriel Constantin, Liviu-Daniel Stefan, Marian Ghencescu, e Bogdan Ionescu. Little-big deep neural networks for embedded video surveillance. páginas 493–496. Institute of Electrical and Electronics Engineers (IEEE), 10 2018. doi:10.1109/iccomm.2018.8484765.
- [26] Oscar Alsing. Mobile object detection using tensorflow lite and transfer learning, 2018. URL: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-233775>.
- [27] Gulraiz Khan, Zeeshan Tariq, e Muhammad Usman Ghani Khan. Multi-person tracking based on faster r-cnn and deep appearance features, 12 2019. URL: [www.intechopen.com](http://www.intechopen.com), doi:10.5772/intechopen.85215.
- [28] Jianan Li, Xiaodan Liang, Shengmei Shen, Tingfa Xu, Jiashi Feng, e Shuicheng Yan. Scale-aware fast r-cnn for pedestrian detection. *IEEE Transactions on Multimedia*, 20:985–996, 4 2018. doi:10.1109/TMM.2017.2759508.
- [29] Liliang Zhang, Liang Lin, Xiaodan Liang, e Kaiming He. Is faster r-cnn doing well for pedestrian detection? volume 9906 LNCS, páginas 443–457. Springer Verlag, 2016. URL: [https://link.springer.com/chapter/10.1007/978-3-319-46475-6\\_28](https://link.springer.com/chapter/10.1007/978-3-319-46475-6_28), doi:10.1007/978-3-319-46475-6\_28.
- [30] Joseph Redmon, Santosh Divvala, Ross Girshick, e Ali Farhadi. You only look once: Unified, real-time object detection. URL: <http://pjreddie.com/yolo/>.
- [31] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, e Alexander C Berg. Ssd: Single shot multibox detector. URL: <https://github.com/weiliu89/caffe/tree/ssd>.
- [32] Debojit Biswas, Hongbo Su, Chengyi Wang, Aleksandar Stevanovic, e Weimin Wang. An automatic traffic density estimation using single shot detection (ssd)and mobilenet-ssd. *Physics and Chemistry of the Earth*, 110:176–184, 4 2019. doi:10.1016/j.pce.2018.12.001.
- [33] Cemil Sungur e Akif Durdu. Real-time diseases detection of grape and grape leaves using faster r-cnn and ssd mobilenet architectures rac-lab autonomous systems view project the extension of battery lives used in a solar irrigation system with regards to charge-discharge numbers view project shekofa ghoury, 2019. URL: <https://www.researchgate.net/publication/334987612>.
- [34] Wahyu Rahmaniari e Ari Hernawan. Real-time human detection using deep learning on embedded platforms: A review. *Journal of Robotics and Control (JRC)*, 2:462–468, 11 2021. URL: <https://journal.umy.ac.id/index.php/jrc/article/view/10558>, doi:10.18196/jrc.26123.
- [35] Ahmad K. Aijazi, Paul Checchin, e Laurent Trassoudaine. Multi sensorial data fusion for efficient detection and tracking of road obstacles for inter-distance and anti-collision safety management. Em *2017 3rd International Conference on Control, Automation and Robotics, ICCAR 2017*, páginas 617–621. Institute of Electrical and Electronics Engineers Inc., jun 2017. doi:10.1109/ICCAR.2017.7942771.

- [36] Ming Liang, Bin Yang, Yun Chen, Rui Hu, e Raquel Urtasun. Multi-task multi-sensor fusion for 3d object detection.
- [37] Ricardo Omar Chavez-Garcia e Olivier Aycard. Multiple sensor fusion and classification for moving object detection and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 17:525–534, 2 2016. doi:10.1109/TITS.2015.2479925.
- [38] Paul Rybski, Peter Anderson-Sprecher, Daniel Huber, Chris Niessl, e Reid Simmons. Sensor fusion for human safety in industrial workcells. páginas 3612–3619, 2012. doi:10.1109/IROS.2012.6386034.
- [39] Byoung Kyu Dan, You Sun Kim, Suryanto, June Young Jung, e Sung Jea Ko. Robust people counting system based on sensor fusion. *IEEE Transactions on Consumer Electronics*, 58:1013–1021, 2012. doi:10.1109/TCE.2012.6311350.