



INSTITUTO TECNOLÓGICO VALE



**Programa de Pós-Graduação em Instrumentação, Controle e  
Automação de Processos de Mineração - PROFICAM  
Universidade Federal De Ouro Preto - Escola de Minas  
Associação Instituto Tecnológico Vale**

**Dissertação**

**Sistema para detecção de estradas e obstáculos baseado em  
imagens RGB e nuvem de pontos para equipamentos de mineração**

**Bruno Peres Klafke**

**Ouro Preto  
Minas Gerais, Brasil  
2018**

**Bruno Peres Klafke**

**Sistema para detecção de estradas e obstáculos baseado em  
imagens RGB e nuvem de pontos para equipamentos de mineração**

Dissertação apresentada ao curso de Mestrado Profissional em Instrumentação, Controle e Automação de Processos de Mineração da Universidade Federal de Ouro Preto e do Instituto Tecnológico Vale, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Controle e Automação.

Orientador: Prof. Gustavo Medeiros Freitas, D.Sc – Pesquisador Colaborador do ITV MI, Professor da UFMG.

Coorientadora: Prof. Andrea C. Bianchi, D.Sc. – Professora da UFOP.

Ouro Preto

2018

K631s

Klafke, Bruno Peres.

Sistema para detecção de estradas e obstáculos baseado em imagens RGB e nuvem de pontos para equipamentos de mineração [manuscrito] / Bruno Peres Klafke. - 2018.

99f.: il.: color; grafs; tabs.

Orientador: Prof. Dr. Gustavo Medeiros Freitas.

Dissertação (Mestrado) - Universidade Federal de Ouro Preto. Escola de Minas. Departamento de Engenharia de Controle e Automação e Técnicas Fundamentais. Programa de Pós Graduação em Instrumentação, Controle e Automação de Processos de Mineração.

Área de Concentração: Engenharia de Controle e Automação de Processos Minerais.

1. Fusão de imagens. 2. Câmera RGB. 3. Laser. 4. Nuvem de pontos. I. Freitas, Gustavo Medeiros. II. Universidade Federal de Ouro Preto. III. Título.

CDU: 681.5:622.2

Catálogo: [www.sisbin.ufop.br](http://www.sisbin.ufop.br)

Mestrado Profissional em Instrumentação, Controle e Automação de Processos  
de Mineração - PROFICAM

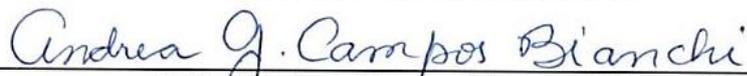
Sistema para Detecção de Estradas e Obstáculos Baseado em Imagens RGB  
e Nuvem de Pontos para Equipamentos de Mineração

Bruno Peres Klafke

Dissertação defendida e aprovada em 09 de outubro de 2018 pela banca  
examinadora constituída pelos professores:

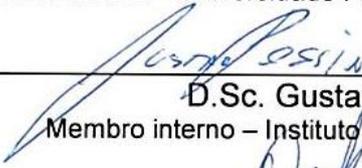
---

D.Sc. Gustavo Medeiros Freitas  
Orientador – Universidade Federal de Minas Gerais (UFMG)  
Participação por Videoconferência



---

D.Sc. Andrea Gomes Campos Bianchi  
Coorientadora – Universidade Federal de Ouro Preto (UFOP)

  
D.Sc. Gustavo Pessin  
Membro interno – Instituto Tecnológico Vale (ITV)

---

  
D.Sc. Bruno Nazário Coelho  
Membro externo – Universidade Federal de São João del-Rei (UFSJ)

## **AGRADCIMENTOS**

Agradeço primeiramente aos meus pais por todo apoio e empenho.

Agradeço a minha esposa que me apoiou sempre e soube compreender os momentos de ausência, porem sempre se empenhando para me motivar.

Agradeço ao Prof. Dr. Gustavo Medeiros Freitas, meu orientador, que durante todo o meu curso me orientou precisamente para a elaboração deste trabalho e por ter sido atencioso e prestativo em relação às necessidades de pesquisas acadêmicas.

Agradeço a Prof. Dr. Andrea Gomes Bianchi por coorientar este trabalho e direcionar a pesquisa com seus conhecimentos acadêmicos.

Agradeço ao Instituto de Tecnologia Vale (ITV) pela oportunidade de concluir este trabalho usando suas ferramentas e laboratórios e também a todos os pesquisadores que agregaram valor em conhecimento para essa pesquisa.

Por fim, agradeço a todos meus amigos que sempre apoiaram e souberam ouvir em momentos difíceis e a todos acadêmicos da primeira turma de mestrado da qual faço parte.

## RESUMO

O Brasil é um país em que predomina o clima tropical. Todas as minas da Vale, independentemente da *commodity* (ferro, manganês e ou cobre), sofrem o impacto das condições climáticas adversas, principalmente nos períodos chuvosos, com neblinas intensas, que interferem fortemente nas operações, reduzindo e/ou interrompendo os processos de lavra em determinados períodos do dia. Até o presente momento não foi desenvolvida uma tecnologia que viabilize com segurança a operação dos caminhões nos períodos de neblina intensa, bem como diminuir os riscos de perdas materiais significativas nos equipamentos em operação. As câmeras convencionais, já usadas para auxílio em manobras, geram imagens no formato RGB e possuem alcance elevado. Porém, em condições adversas, a visibilidade não é diferente do olhar humano, dificultando a identificação de obstáculos e até mesmo a pista de rolagem dos equipamentos. A possibilidade de incluir o sistema de laser pode auxiliar a operação em condições adversas, incluindo neblina, chuva e pouca luminosidade. A fusão dos dados da câmera RGB com a varredura laser deve auxiliar a operação dos equipamentos, identificando obstáculos e estradas com maior confiabilidade. O estudo proposto consiste em investigar técnicas capazes de fornecer informações sobre a estrada e obstáculos considerando imagens RGB e uma nuvem de pontos de varredura laser, assim como sua fusão. No processamento são utilizados filtros e redes neurais para o processamento de imagens, e algoritmos de identificação de aglomerados ou planos para processar nuvens de pontos. Para esse estudo as estratégias propostas para identificação de objetos por imagem atingiram 90% de acerto e as estratégias de identificação através da nuvem de pontos alcançaram 91% de acerto. Porém, quando o processamento é realizado com a fusão dos dados dos sensores 95% dos obstáculos são identificados. Apesar desta pesquisa consistir numa etapa preliminar inserida num projeto de escopo maior, os resultados obtidos com o processamento de dados podem ser considerados satisfatórios.

**Palavras-chave:** Fusão de imagens, câmeras RGB, laser, nuvem de pontos.

## **ABSTRACT**

Brazil is a country where the tropical weather is predominant. All Vale's mines, regardless of the commodity (iron, manganese, nickel, phosphate and copper), suffer impact of the adverse weather conditions, especially on the rainy season, with heavy fog, which strongly affects operations, reducing and/or interrupting the processes of mining outdoor in specific periods of the day. Up to the present moment it has not been developed a technology which could make feasible with security the off road operations on periods of heavy fog, as well as to reduce risks of significant material loss in equipments during operation. The conventional cameras, already used to assist maneuvers, generate images in RGB format and have increased range. However, in adverse conditions, the visibility is not different from the human eye, making difficult the identification of obstacles and even the equipment's rolling track. The possibility of including a laser system can support the operation in adverse conditions including fog, rain and poor visibility. The data fusion of RGB camera with sweep laser should support the equipments operation indentifying obstacles and roads with more reliability. The proposed study evaluates techniques capable to provide information regarding the road and obstacles considering RGB and a sweep laser cloud of points, such as its fusion. The strategy use filter and neural nets for images process, and algorithms of clusters identification or plans to process clouds of points. For this study the identification of objects by image reached 90%, through the cloud of points reached 91%. However, using data fusion allows to correctly identify 95% of the objects in the environment. Although this research consists in a preliminary stage inserted in a project of greater scope, the results obtained with data process can be considered satisfactory.

**Keywords:** Picture fusion, RGB cameras, laser, cloud points.

## LISTA DE FIGURAS

Figura 1: Perfil de perdas anuais nas minas de ferrosos da Vale por condições adversas de clima (Relatório de Acompanhamento Proj. Direção Assistida 2017).....	13
Figura 2: Condições de visibilidade em situações adversas (Treinamento de Direção de Mina - Vale, 2012). .....	13
Figura 3: Visão conceitual: Disposição do hardware para exibição da interface gráfica (Relatório de Acompanhamento Proj. Direção Assistida 2017).....	15
Figura 4: Interface para o painel do operador (Relatório de Acompanhamento Proj. Direção Assistida 2017). .....	16
Figura 5: Sistemas principais para um veículo inteligente (HABERMANN, 2016).....	24
Figura 6: (a) Ondas curtas com alta frequência e (b) Ondas longas com baixa frequência. (SCURI, 2002). .....	25
Figura 7: Comprimento de Onda e Espectro visível (SCURI, 2002). .....	26
Figura 8: Sistema de cores RGB e Representação do cubo RGB. ....	26
Figura 9: Convenção dos eixos para representação de imagens digitais (GONZALES et al., 2007). .....	27
Figura 10: Sequência de etapas do Processamento de Imagens (SCURI, 2002). .....	28
Figura 11: Máscara detectores de bordas e derivadas de primeira ordem (GONZALES et al., 2007). .	30
Figura 12: Princípio de <i>time os flight</i> (MAIA, 1999). .....	31
Figura 13: Sistema de Coordenadas do Sensor VLP-16, adaptado de velodyne.com manual do usuário. ....	32
Figura 14: Nuvem de pontos tridimensional e seleção do conjunto de pontos situados a uma distância $r$ do ponto central P (CESAR, 2016). .....	35
Figura 15: Imagem de varredura de sensor de laser (velodynelidar.com). .....	37
Figura 16: Hiperplanos separando pontos no treinamento. Hiperplano H3 possui maior margem (HABERMANN, 2016). .....	41
Figura 17: Separação de pontos por hiperplanos. Em (a) caso linear e (b) caso não linear (HABERMANN, 2016). .....	42
Figura 18: Ilustração de uma rede neural artificial (Haykin, 1998). .....	43
Figura 19: Configuração da rede neural <i>Convnet</i> classificadora de imagens e suas principais camadas. Adaptado de (Ren et al., 2017). .....	44

Figura 20: Representação de uma convolução de filtro de ordem 3 e o volume de entrada (CARNEIRO e SILVA, 2017).....	45
Figura 21: Representação da atuação do passo no deslocamento do filtro de ordem 3 em duas etapas seguidas de convolução. Em (a) e (b) com passo 1 e em (c) e (d) com passo 2 (CARNEIRO e SILVA, 2017). .....	46
Figura 22: Representação de uma operação <i>max pooling</i> em uma imagem de ordem 4 utilizando filtro de ordem 2. Reduz o tamanho da imagem e o processamento para camadas seguintes (FERREIRA, Santos 2017).....	47
Figura 23: Representação de uma extração de padrões da imagem de entrada por uma rede neural <i>Convnet</i> e sua classificação. Adaptado de (SALIMANS et al., 2017). .....	48
Figura 24: Câmera AXIS P1224-E ( <a href="http://axis.com/br">axis.com/br</a> ).....	51
Figura 25: Sensor Velodyne LiDAR VLP-16 ( <a href="http://velodyne.com">velodyne.com</a> ). .....	52
Figura 26: Unidade para controle de movimentos PTU-D47. ....	52
Figura 27: Bateria portátil recarregável BT-70791.....	53
Figura 28: Conversor de tensão DC/DC do fabricante VICOR. ....	53
Figura 29: Fluxograma do processo de detecção de estradas e obstáculos. ....	55
Figura 30: Suporte para os sensores e acessórios. ....	56
Figura 31: Fluxograma do processo de calibração. ....	56
Figura 32: Calibração do laser com câmera RGB através dos vetores R e t. (VELAS et al., 2014).....	57
Figura 33: Marcador plano com furos circulares em (a) e arestas detectadas pelo laser (b). Adaptado de (VELAS et al., 2014). ....	57
Figura 34: Detecção dos quatro círculos e verificação através das distâncias dos centros. Adaptado de (VELAS e colab., 2014).....	58
Figura 35: Localização da Mina de Abóbora no Complexo Vargem Grande em Minas Gerais (Imagens Google). ....	61
Figura 36: Sensores fixados no veículo de apoio para coleta de dados.....	62
Figura 37: (a) Imagem original e (b) com pré-processamento por classificação de pixel.....	62
Figura 38: Operações da fase de processamento de imagens.....	63
Figura 39: (a) Imagem com pré-processamento e (b) com eliminação dos maiores ruídos no Matlab. ....	64
Figura 40: (a) Imagem com eliminação dos maiores ruídos e (b) com preenchimento de espaços no Matlab. ....	64

Figura 41: (a) Imagem original (b) detecção da estrada no Matlab.....	65
Figura 42: Arquitetura geral da rede neural de 16 camadas usada. Adaptado de (NOH et al., 2015).	68
Figura 43: Identificação de objetos por rede neural.....	69
Figura 44: Distância entre feixes de laser para determinar valor de t.....	70
Figura 45: (a) Cenário de coleta da nuvem de pontos, (b) nuvem de pontos original, (c) nuvem de pontos com identificação de planos. ....	71
Figura 46: Funcionamento do DBSCAN para encontrar agrupamentos (VALÊNCIO, 2013). ....	72
Figura 47: (a) Cenário de coleta da nuvem de pontos, (b) nuvem de pontos original, (c) nuvem de pontos com identificação de plano e aglomerado de pontos do obstáculo.....	73
Figura 48: Plano tridimensional considerado para referência.....	74
Figura 49: Modelo simplificado do algoritmo. ....	75
Figura 50: Identificação por fusão de sensores.....	76
Figura 51: Resultado da calibração dos sensores de imagem e laser. ....	78
Figura 52: Identificação da estrada por imagem por detecção de bordas em ambiente urbano. ....	79
Figura 53: Identificação da estrada por imagem por detecção de bordas em ambiente de mina.....	79
Figura 54: identificação de objetos por imagem: (a) durante o dia, (b) durante a noite e (c) durante neblina ( <a href="https://br.123rf.com">https://br.123rf.com</a> ).....	81
Figura 55: Identificação de obstáculos em ambiente urbano.....	82
Figura 56: Identificação de obstáculos em ambiente de mina. ....	82
Figura 57: Erro na identificação de objetos com distâncias elevadas.....	83
Figura 58: Identificação da estrada por nuvem de pontos. ....	84
Figura 59: Identificação de obstáculos por nuvem de pontos. ....	85
Figura 60: Identificação por fusão de sensores.....	87
Figura 61: Unificação dos resultados. ....	88

## LISTA DE TABELAS

Tabela 1: Configuração da ConvNet mostrando em colunas de A à E, aumentando o número de camadas, adaptado de (SIMONYAN e ZISSERMAN, 2014).....	67
Tabela 2: Número de parâmetros para cada configuração (em milhões), adaptado de (SIMONYAN e ZISSERMAN, 2014).....	67
Tabela 3: Performance de cada configuração da ConvNet, adaptado de (SIMONYAN e ZISSERMAN, 2014). ....	68
Tabela 4: Percentual de identificação de objetos pela CNN com imagens em diferentes níveis de visibilidade.....	80
Tabela 5: Percentual de identificação de objetos pela CNN com imagens realizadas durante as coletas. ....	81
Tabela 6: Percentual de identificação de estradas através do RANSAC. ....	84
Tabela 7: Percentual de identificação de objetos nas nuvens de pontos pelo método de DBSCAN....	86
Tabela 8: Percentual de obstáculos identificados através de fusão dos sensores. ....	86

## LISTA DE SIGLAS E ABREVIATURAS

6DoF	6 Graus de Liberdade
ACFR	Centro Australiano de Robótica de Campo
BIR	Instituto Brasileiro de Robótica
C	Linguagem de Programação
C++	Linguagem de Programação compilada multi-paradigma
CCD	Dispositivo de Carga Acoplado
DARPA	Agência de Projetos de Pesquisa Avançada de Defesa
DBSCAN	Agrupamento Espacial Baseado em Densidade e Ruído
EMBRAPII	Associação Brasileira de Pesquisa e Inovação Industrial
GPS	Sistema de Posicionamento Global
IMU	Unidade de Medição Inercial
ITV	Instituto Tecnológico Vale
LHD	Carregadeira Transportadora de Carga
LIDAR	Medição de propriedade de luz refletida por laser
MIT	Licença de softwares de computadores
MVC	Matriz de Variâncias e Covariâncias
PTU	Unidade para Controle de Movimento
QML	Linguagem de Programação para descrever objetos e interface
RGB	Sistema de Cores Vermelho, Verde e Azul
SENEAI	Serviço Nacional de Aprendizagem Industrial
SO(3)	Grupo ortogonal especial de dimensão 3

# SUMÁRIO

1.	INTRODUÇÃO .....	12
1.1	Projeto de Direção Assistida.....	14
1.2	Motivação .....	17
1.2.1	Aplicações no setor de Mineração.....	17
1.3	Objetivos.....	19
1.3.1	Objetivos Específicos.....	19
1.4	Organização da Dissertação.....	20
2	REFERENCIAL TEÓRICO E FUNDAMENTAÇÃO CIENTÍFICA .....	21
2.1	Sistemas de cores aditivas .....	25
2.2	Representação da imagem digital .....	27
2.2.1	Processamento da Imagem Digital.....	28
2.3	Sistema Laser Scanner .....	30
2.3.1	Nuvem de Pontos .....	34
2.4	Fusão de dados e calibração de sensores .....	37
2.4.1	Calibração da câmera RGB com Laser .....	37
2.4.1.1	Calibração grosseira: .....	38
2.4.1.2	Calibração fina.....	38
2.5	Aprendizado de máquina .....	39
2.5.1	Aprendizado supervisionado.....	40
2.5.2	<i>AdaBoost</i> .....	40
2.5.3	<i>Support Vector Machine</i> .....	41
2.5.4	Redes Neurais Artificiais.....	42
2.5.5	Aprendizado não supervisionado.....	49
2.5.6	RANSAC .....	49

2.5.7	DBSCAN .....	49
2.5.8	<i>Iterative Closest Point (ICP)</i> .....	50
3	MATERIAIS E MÉTODOS .....	51
3.1	Materiais .....	51
3.2	Métodos .....	53
3.2.1	Calibração da Câmera RGB com Laser .....	55
3.2.2	Coleta de dados em campo .....	61
3.2.3	Identificação da estrada por imagem.....	62
3.2.4	Identificação de obstáculos por imagem RGB.....	65
3.2.5	Identificação de estrada pela nuvem de pontos.....	69
3.2.6	Identificação de objetos por nuvem de pontos .....	71
3.2.7	Identificação por fusão dos sensores .....	73
4	RESULTADOS .....	77
4.1	Calibração dos sensores .....	77
4.2	Identificação da estrada por imagem RGB .....	79
4.3	Identificação de obstáculos por imagem RGB .....	80
4.4	Identificação da estrada por nuvem de pontos.....	83
4.5	Identificação de obstáculos por nuvem de pontos .....	84
4.6	Identificação por fusão de sensores .....	86
4.7	Comparação dos resultados .....	87
5	CONCLUSÕES .....	88
5.1	Contribuições.....	89
5.2	Trabalhos futuros .....	90
6	REFERÊNCIAS .....	91

## 1. INTRODUÇÃO

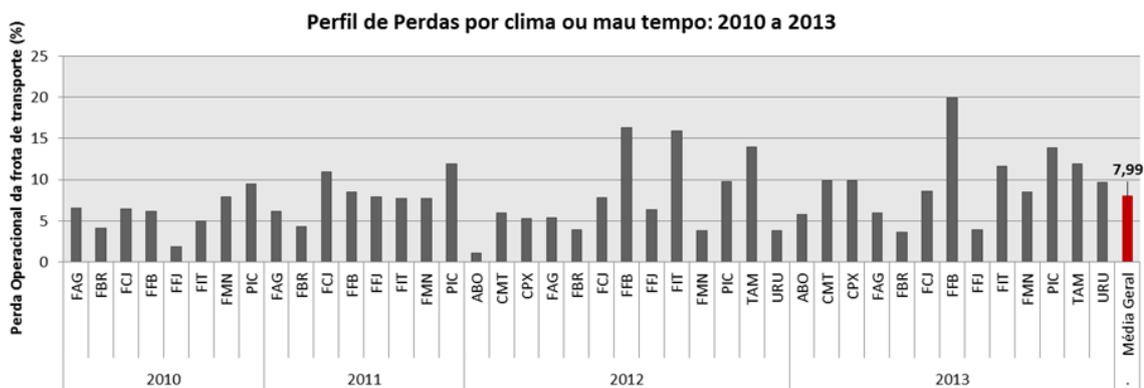
As minas da Vale no Brasil, em que predominam o clima tropical, sofrem o impacto das condições climáticas adversas, principalmente nos períodos chuvosos, com neblinas intensas.

Algumas pesquisas indicam que os acidentes em mineração estão relacionados a um de três fatores contribuintes: fator humano, fator veicular e fator viário-ambiental. Assim o investimento adequado em treinamentos para operadores e outros usuários da via, padrão de fabricação e manutenção dos equipamentos, projetos de normas e manutenção de vias, podem prevenir acidentes de tráfego. Além disso, as tecnologias embarcadas nos veículos podem bloquear o comportamento ou reduzir o tempo de resposta humano cada vez mais, prevenindo acidentes ou minimizando seus impactos (OECD, 2012).

Para os caminhões do setor de mineração, em que o risco de acidentes é causado principalmente por comportamento humano e condições adversas de visibilidade, um sistema de identificação de obstáculos poderia aumentar a segurança de operação. Trafegar em uma mina depende de inúmeras variáveis e uma delas é a condição climática e a luminosidade na mina, que pode prejudicar a visibilidade do operador. Atualmente para este caso a solução mais segura é interromper a produção.

O gráfico da Figura 1 resume as perdas operacionais da frota de transporte por mina do ano de 2010 a 2013 devido a condições climáticas adversas para os sistemas de produção de ferrosos. O perfil de perdas representa o percentual de horas paradas devido a visibilidade em relação ao total de horas ociosas. As perdas operacionais obtiveram uma média de 7,99% de perdas anuais, sendo que em períodos chuvosos as perdas podem chegar a 10,48% da produção. Em termos de redução de custos, a baixa produtividade das frotas de transporte, com redução da utilização dos equipamentos, impacta fortemente nos custos operacionais (US\$/ton movimentada). Na região amazônica, as perdas são muito significativas em Carajás, Salobo e Sossego, principalmente nas estações chuvosas, onde o clima quente e úmido

favorece a evaporação e a geração de neblina. As chuvas torrenciais, que reduzem significativamente a visibilidade, como mostra a Figura 2, também contribuem fortemente para a redução da produtividade nas minas a céu aberto.



**Figura 1: Perfil de perdas anuais nas minas de ferrosos da Vale por condições adversas de clima (Relatório de Acompanhamento Proj. Direção Assistida 2017).**

É de grande importância para o setor mineral viabilizar a operação dos equipamentos mesmo em condições de baixa visibilidade do operador. A redução das horas ociosas devido à falta de visibilidade permite aumentar a produtividade dos ativos de uma mina, e ao mesmo tempo mantém a operação segura.

O aumento da produtividade dos equipamentos impacta positivamente na produção da mina ou então, na redução de custos, possibilitando produzir a mesma quantidade de minério com uma frota mais enxuta.



**Figura 2: Condições de visibilidade em situações adversas (Treinamento de Direção de Mina - Vale, 2012).**

Até o presente momento não foi desenvolvida uma tecnologia que viabilize a operação dos caminhões nos períodos com chuva e neblina. Operar na mina a céu aberto com baixa visibilidade compromete a segurança dos operadores, bem como aumentam os riscos de perdas materiais significativas nos equipamentos em operação.

Este trabalho consiste em uma etapa preliminar do “Projeto de Direção Assistida”, que é descrito na próxima seção. Por motivos de simplificação, será realizado com dados obtidos em boas condições de visibilidade utilizando apenas dois sensores: laser 3D de múltiplos feixes e câmera de vídeo RGB. Também é importante ressaltar que todo o processamento será realizado através de fotos e nuvem de pontos estáticos em um processamento *off-line*. Validando as técnicas estudadas, é possível migrar para os cenários mais desafiadores de visibilidade e processamento através de vídeo em trabalhos futuros. Os resultados e conhecimento adquirido neste trabalho auxiliarão e integrarão o “Projeto de Direção Assistida”, que possui maior escopo e complexidade.

## **1.1 Projeto de Direção Assistida**

O SENAI/CIMATEC está desenvolvendo, em parceria com a empresa VALE S.A. e o Instituto Tecnológico Vale (ITV), o projeto “Sistema de Direção Assistida para Caminhões”, com apoio da Empresa Brasileira de Pesquisa e Inovação Industrial (EMBRAPII).

Como citado anteriormente, grande parte dos acidentes envolvendo direção de veículos ocorre por falha humana. Tais riscos aumentam consideravelmente em condições de baixa visualização provocada por chuva, neblina e poeira. Além disso, o problema de condução de veículos em condições adversas também corresponde a um entrave crítico no desenvolvimento e utilização de veículos autônomos.

O projeto “Sistema de Direção Assistida” propõe o desenvolvimento de uma interface de fácil acesso ao usuário, projetada em um painel ou para-brisa do veículo utilizado para transporte do minério nas dependências de uma mina. A solução

disponibilizará dados referentes à localização, rota, obstáculos e velocidade que serão captados por dispositivos instalados no veículo. Dessa forma, entende-se que o sistema estará conectado a um conjunto de sensores independentes ao sistema do veículo, coletando e realizando a fusão dos dados dos sensores e seu processamento computacional para que em seguida possa exibi-lo para o usuário (motorista do veículo).

O sistema será composto por um conjunto de sensores (câmeras, laser, radar, GPS) capazes de perceber o ambiente ao redor do equipamento e um display onde são apresentadas informações gráficas ao motorista para auxiliar a direção do equipamento conforme apresentado na Figura 3.

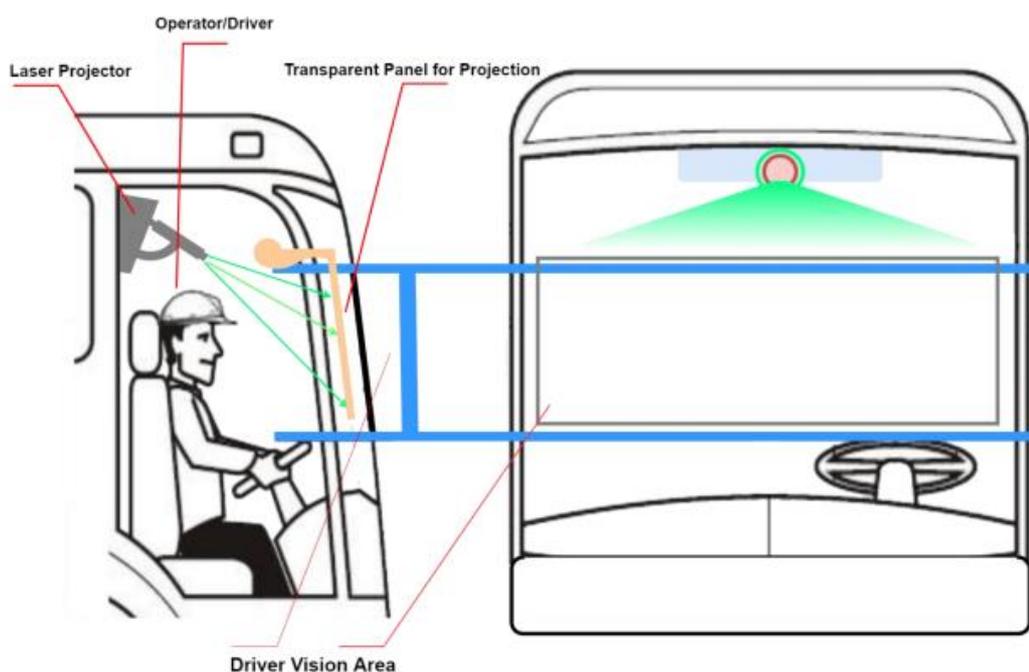


Figura 3: Visão conceitual: Disposição do hardware para exibição da interface gráfica (Relatório de Acompanhamento Proj. Direção Assistida 2017).

Os hardwares adicionados ao equipamento são descritos abaixo como:

- *Driver Vision Area* - Área de Visão do operador/motorista, espaço retangular que permite ao operador visualizar o trecho.
- *Transparent Panel for Projection* – Painel ou película transparente para projeção. É um tipo de película especial que permite a exibição das imagens provenientes dos projetores a laser.

- *Laser Projector* – Equipamento de projeção de imagem RGB na área de visão do operador.

Na Figura 4 é apresentado o conceito da interface a ser exibido para o usuário tendo como alguns requisitos obrigatórios:

- Informações visuais ao operador para a condução do veículo sem visibilidade;
- Representação de obstáculos visualmente para o operador;
- Exibição das distâncias dos obstáculos existentes na rota.

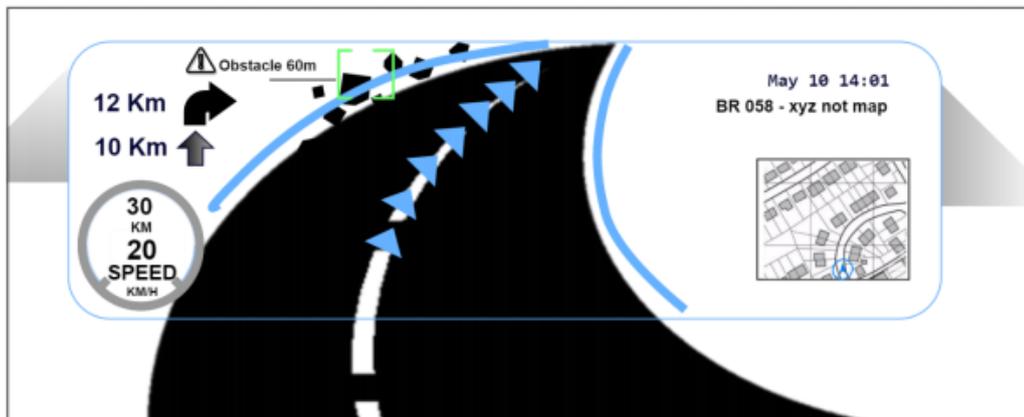


Figura 4: Interface para o painel do operador (Relatório de Acompanhamento Proj. Direção Assistida 2017).

A interface proposta apresentará um painel a fim de manter a atenção do operador na tarefa central que é a direção do veículo. O painel contém informações como velocidade, distância e velocidade média. Avisos serão exibidos em forma de *cards* na parte superior da tela, e também há informações ou avisos plotados na tela para informar sobre obstáculos próximos. Uma rota deve ser exibida para orientar a direção, e sendo desejável também que sejam plotados os limites do trecho/estrada, além do pequeno mapa na parte inferior da tela. A interface com o usuário preza por simplicidade e leveza, contendo poucos elementos visuais.

O *back-end* ou motor de processamento de dados deve ser construído utilizando as linguagens C, C++ e *Python*, lógica a ser seguida pela interface utilizando componentes em QML (*Qt Meta Language*). Futuramente poderão ser identificadas outras necessidades ou mesmo a inclusão de outras linguagens ou *frameworks* que possam auxiliar ou melhorar a experiência do usuário.

Como descrito anteriormente, o sistema será composto por um conjunto de sensores (câmeras, laser, radar, GPS) capazes de perceber o ambiente ao redor do equipamento. Porém, as informações coletadas pelos sensores devem ser processadas, filtradas e fundidas em uma única imagem para o operador.

Esta dissertação tem como foco testar somente dois tipos de sensores, câmera RGB e laser, para suportar o projeto “Sistema de Direção Assistida” e posteriormente somar com o processamento e informações dos demais sensores.

## **1.2 Motivação**

Nos últimos anos, os veículos autônomos estão cada vez mais inseridos no setor de mineração. O crescimento dessa tecnologia permite direcionar pesquisas com o objetivo de aumentar a segurança no trânsito, oferecer conforto e outras facilidades para o usuário. Essa tecnologia pode também melhorar a eficiência do transporte de cargas e passageiros e reduzir os impactos ambientais (HABERMANN, 2016).

Uma importante motivação para esse estudo foi buscar uma solução que não dependa de estruturas externas dos equipamentos. A proposta é que os sensores sejam instalados no próprio equipamento.

### **1.2.1 Aplicações no setor de Mineração**

Uma pesquisa realizada em (Ramos, 2008) descreve que as operações do setor de mineração no futuro incluirão:

- Os equipamentos de mina terão sensores usados para fornecer dados da face das bancadas de lavra e praças de operação;
- Os caminhões de mineração serão autônomos e portadores de sensores para processamento e atualização do avanço da mina;
- Ferramentas de estatística e probabilidades permitirão mapear o minério através dos sensores instalados na mina ou nos equipamentos móveis.

Equipamentos inteligentes vêm se tornando cada vez mais atrativos na cadeia de produção mineral, caminhões, tratores, perfuratrizes, carregadeiras operadas remotamente além de aplicações de logística automatizadas por uma central de operações. Na mineração vêm crescendo a utilização dessas soluções e equipamentos, visando melhorar o desempenho, condições de saúde e segurança, buscando afastar os trabalhadores das áreas de risco (FREITAS, 2015).

Uma pesquisa realizada em (BACKBLOM, 2010) com o objetivo de desenvolver uma projeção da mina no futuro para as multinacionais do setor mineral que investem na automação e pesquisa da área de tecnologia da mineração descreve diversas melhorias em relação à mineração nas próximas décadas:

- Não haver acidentes pessoais e materiais;
- Contribuir para sustentabilidade ambiental das mineradoras reduzindo a emissão de CO<sub>2</sub>, uso de energia elétrica sustentável e redução dos rejeitos;
- A competitividade das empresas será através de investimentos em tecnologia para melhorar os processos como:
  - **Central de operações:** recebe os dados processados da qualidade do minério, da equipe técnica e dos sensores, possibilitando os ajustes mais precisos na operação;
  - **Redução da presença de pessoas nas áreas operacionais:** as operações e coleta de dados, incluindo o mapeamento geológico, serão controlados a distância. Existirão equipamentos automatizados para executar as manutenções dos equipamentos em campo;
  - **Lavra contínua:** fluxo de lavra sem interrupção de troca de operações ou por baixa visibilidade operacional. O fluxo de lavra contínuo é de grande importância para o aperfeiçoamento do setor mineral.

Atualmente, devido às incertezas associadas à qualidade do minério, ao estado de conservação dos equipamentos e improdutividade operacional as mineradoras operam de maneira conservadora para cumprir a demanda dos clientes e garantir o lucro da operação (FREITAS, 2015).

### **1.3 Objetivos**

O objetivo principal desta dissertação é investigar técnicas de processamento de dados de dois sensores, câmera RGB e laser, bem como sua fusão para auxiliar a identificação de objetos e estradas durante a operação dos equipamentos de mineração.

Atualmente existem diversas técnicas de processamento de dados e faz parte dos objetivos compreender quais podem auxiliar no estudo. Um dos desafios principais é desenvolver um algoritmo computacional para realizar o processamento dos dados dos sensores lasers tipo LiDAR e câmera RGB trabalhando em conjunto. O algoritmo deverá identificar os principais objetos ou obstáculos à frente do equipamento e os limites de uma pista dentro de um ambiente de mineração.

Esse estudo consiste numa etapa preliminar de processamento de dados inserida no “Projeto de Direção Assistida” que possui um escopo maior. Nesta etapa, por motivos de simplificação, são utilizados cenários em boas condições de iluminação, devido à complexidade de coletar dados em diversas condições de visibilidade em ambiente de mina.

#### **1.3.1 Objetivos Específicos**

Os objetivos específicos deste estudo consistem em:

- Obter dados de cenas do ambiente de mineração a céu aberto, ainda sem condições adversas;
- Validar técnicas de processamento das imagens capturadas pela câmera RGB e a nuvem de pontos do laser como remoção de ruídos, filtrar os dados e evidenciar pontos de interesse;
- Elaborar algoritmo computacional para identificação de obstáculos e pista de rolagem através dos sensores individualmente;
- Aplicar técnicas de fusão de dados dos sensores para identificação das áreas ou pontos de interesse;

- Aplicar técnicas estruturadas de processamento de imagem e nuvem de pontos propostas por outros pesquisadores para o ambiente de mineração.

#### **1.4 Organização da Dissertação**

Esta dissertação é composta por cinco capítulos. Este capítulo introduz o assunto abordado durante a pesquisa, descrevendo o problema encontrado na operação de mina a céu aberto, os objetivos desejados e considerações sobre trabalhos na área.

O referencial teórico e fundamentação científica são apresentados no Capítulo 2. Nessa etapa serão apresentadas as considerações sobre operação de veículos em condições adversas, os sensores usados para auxiliar ou operar equipamentos e veículos, técnicas de detecção de obstáculos durante seu trajeto, uma descrição sobre câmera RGB e imagem digital junto com nuvem de pontos com mapeamento laser, que serão os dois sensores usados neste relatório.

O Capítulo 3 apresenta o modelo de câmera RGB para coleta das imagens digitais, e o sensor Velodyne para captura da nuvem de pontos. Neste mesmo capítulo serão apresentados os métodos de calibração dos dois sensores, coleta de dados e informações em ambiente urbano e de mineração e o processamento de dados da imagem digital com nuvem de pontos.

Os resultados serão apresentados no Capítulo 4, sendo mostrado um processamento da imagem digital em ambiente urbano e de mineração, resultados da calibração dos sensores e o processamento dos dados para detecção de obstáculos e pista de rolagem.

No Capítulo 5 será apresentada a conclusão da dissertação, assim como as sugestões para trabalhos futuros.

## 2 REFERENCIAL TEÓRICO E FUNDAMENTAÇÃO CIENTÍFICA

Os estudos sobre veículos inteligentes começam no final da década de 50 de forma muito precária e com mais abrangência dos anos 70 e 80 em diante, de modo mais abrangente. Porém, os avanços foram mais significativos com a iniciativa de países europeus em investir nessa tecnologia, dando início ao projeto PROMETHEUS (Williams, 1992). Após o projeto PROMETHEUS, existiram outras iniciativas que fomentaram os estudos para veículos inteligentes (HABERMANN, 2016).

Pesquisas para identificação de objetos como o DBSCAN (Clusterização Espacial Baseada em Densidade de Aplicação com Ruído) descrito em (ESTER, 1996), que identifica agrupamentos de pontos de forma arbitrária e de diferentes tamanhos, começaram a contribuir para o desenvolvimento de técnicas usadas para o estudo de veículos inteligentes. Sem qualquer informação preliminar sobre os dados o DBSCAN separa os ruídos e identifica os aglomerados de pontos. Atualmente diversas linhas de pesquisa utilizam o método descrito para identificação de obstáculos para veículos inteligentes (GAVA et al., 2013).

A pesquisa de (PISANO et al., 2008) revela que a grande maioria dos acidentes envolvendo direção de veículos ocorre por falha humana e estes riscos aumentam consideravelmente em condições climáticas adversas, incluindo a presença de chuva, neblina e poeira. Desta forma, os estudos de sistemas de direção assistida, ou auxílio ao motorista, vêm recebendo grande atenção ao longo dos anos. Além disso, o problema de direção em condições climáticas adversas também corresponde a um entrave crítico no desenvolvimento e utilização de veículos autônomos.

Para (GAVA et al., 2013), os veículos inteligentes necessitam de mais de um tipo de sensor, devido a aplicabilidade de cada um deles em diversos ambientes diferentes. No caso da utilização de múltiplos sensores com diferentes resoluções e princípios de funcionamento, é necessário também o emprego de técnicas de fusão de dados.

Num estudo realizado em (PINHO et al, 2005) define a fusão de dados como conceitos e ferramentas para união de dados ou informações de origens distintas com

objetivo de aumentar a qualidade das informações. Qualquer interação de dados oriunda de diferentes origens capaz de melhorar a análise pode ser considerada como fusão.

Uma questão inicial consiste em avaliar quais sensores podem ser utilizados para o mapeamento de ambientes. Uma análise de sensores e técnicas de instrumentação a serem empregadas em veículos autônomos é apresentada em (PEYNOT et al., 2009), incluindo odometria com *encoders* (dispositivos eletromecânicos que através da rotação de seu eixo reproduzem ou contam sinais eletrônicos), unidades inerciais, compassos geomagnéticos ou sensores *Doppler* (dispositivo que observa ondas quando refletivas ou emitidas de um objeto em movimento) para medição de movimento e identificação de objetos através de sonar ou aplicação de técnicas de visão computacional.

Cada um dos instrumentos de medição previamente mencionados é mais adequado a uma determinada condição de operação, de forma que não existe um sensor ideal com desempenho ótimo para mapeamento de ambientes com presença de chuva, neblina e poeira. Uma solução mais robusta consiste então em combinar vários instrumentos de medição para mapear ambientes naturais (ROSETO, 2017).

Outras iniciativas isoladas mostraram interesses em desenvolver veículos inteligentes, mas elas ganharam visibilidade quando a agência americana DARPA (Agência de Projetos de Pesquisa Avançada de Defesa) começou a estimular pesquisas nessa área com o DARPA *Grand Challenge* realizado em 2005 com a participação de diversas equipes (Buehler, 2009).

Conforme (GROETELAARS, 2011), quando é necessário o uso de diversos instrumentos simultâneos se faz necessário converter os dados de sensores em modelos complexos para criação de superfícies e sólidos. Para conversão geralmente são utilizadas as etapas de planejamento, aquisição de dados, pré-processamento, processamento, análise de precisão do modelo e a exportação dos resultados encontrados.

Para (ZHAO et al., 2014), a fusão dos dados geométricos dos sensores com imagens digitais pode ser realizada de diversas formas diferentes. Uma delas é a lógica *fuzzy* (forma de lógica multivalorada onde os valores lógicos estão entre zero, sendo falso e um, sendo verdadeiro) para o processo de fusão de dados. A lógica *fuzzi* permite modelar funções não lineares de alta complexidade, tolera resultados imprecisos dos dois sensores e é uma estrutura de fusão de dados flexível, sendo que os resultados de mais sensores podem ser integrados ao sistema.

Como descrito em (PAVAN, 2015), o uso do *LASER scanning* para reconstrução tridimensional de ambientes internos e principalmente externos vem crescendo de forma exponencial. Esse equipamento utiliza laser com alta capacidade de escaneamento do ambiente e devido à quantidade de feixes de lasers é possível criar diversas nuvens de pontos do mesmo ambiente de forma rápida e com precisão representando um novo modelo geométrico tridimensional.

Num estudo realizado em (KELLY et al., 2016) é apresentado um sistema que utiliza potenciômetro, *encoder*, DGPS (*Differential Global Positioning System*) e radar para determinar a posição; sensor de toque, lasers, câmeras coloridas, térmicas e estéreo para mapear o ambiente e detectar obstáculos presentes no caminho do veículo. Uma solução diferente apresentada em (PEYNOT et al., 2009), utiliza um sistema com GPS para obter a posição do veículo, integrados com lasers, radar, câmeras coloridas e infravermelha para o mapeamento do ambiente.

A nuvem de pontos é confiável para detecção de obstáculos, mas não possui informações para realizar uma classificação confiável de categorias de objetos específicos (ZHAO et al., 2014). Por isso, muitos pesquisadores utilizam câmeras a cores quando é necessário classificação de objetos.

Para (HWANG et al., 2016), o sensor mais popular para veículos inteligentes é o sensor laser LiDAR (*Light Detection And Ranging*). O princípio de funcionamento desse sensor é baseado no chamado “tempo de vôo”, utilizado para medir a distância entre os objetos e o sensor.

Para (HABERMANN, 2016), o desenvolvimento de um veículo inteligente, ou até mesmo com interação com o motorista, exige conhecimentos em diversas áreas, como processamento de imagem digital, mapeamento de ambiente com uso de laser, sistemas de posicionamento global e outros sistemas similares. Os sistemas principais para um veículo inteligente podem ser observados na Figura 5.

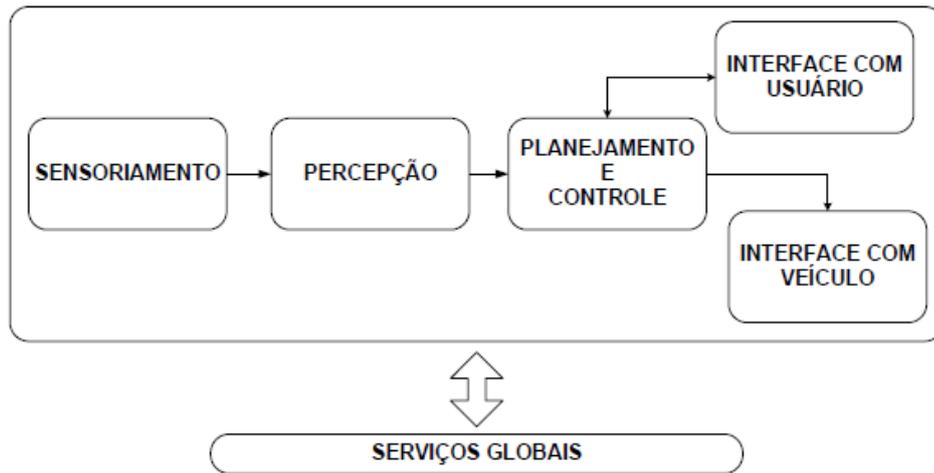


Figura 5: Sistemas principais para um veículo inteligente (HABERMANN, 2016).

Abaixo segue a descrição dos principais sistemas para um veículo inteligente:

- Sensoriamento: responsável por receber os dados dos sensores: radar, câmeras de imagem, laser e GPS;
- Percepção: é responsável por calcular a velocidade, o posicionamento, identifica obstáculos e estradas ao redor do veículo e coleta informações das sinalizações presentes nas vias;
- Planejamento e Controle: tem a responsabilidade de gerar os sinais eletrônicos para o controle da direção e atualiza a trajetória do veículo;
- Interface com o veículo: envia sinais mecânicos ou eletrônicos para sistemas de freios, sistemas de emergências e outros mais;
- Interface com o usuário: seleciona quais informações devem ser apresentadas ao usuário do veículo;
- Serviços globais: realiza a comunicações e atualizações de software e armazena informações do veículo.

## 2.1 Sistemas de cores aditivas

Quando se trata das cores, primeiramente devemos considerar a conformação do olho humano. Os seres humanos possuem uma visão tricrômica, ou seja, vemos todas as cores baseadas na combinação de intensidade das cores vermelha, o azul e verde (PONCE e FORSYTH, 2012).

Numa pesquisa realizada em (SCURI, 2002) revela que a luz é outro assunto importante para se entender as cores. A luz visível faz parte de um conjunto de vibrações eletromagnéticas, onde somente uma parte é percebida pelos olhos. A oscilação dos movimentos dos elétrons define o comprimento de onda, que pode ser em quilômetros ou centímetros (Figura 6).

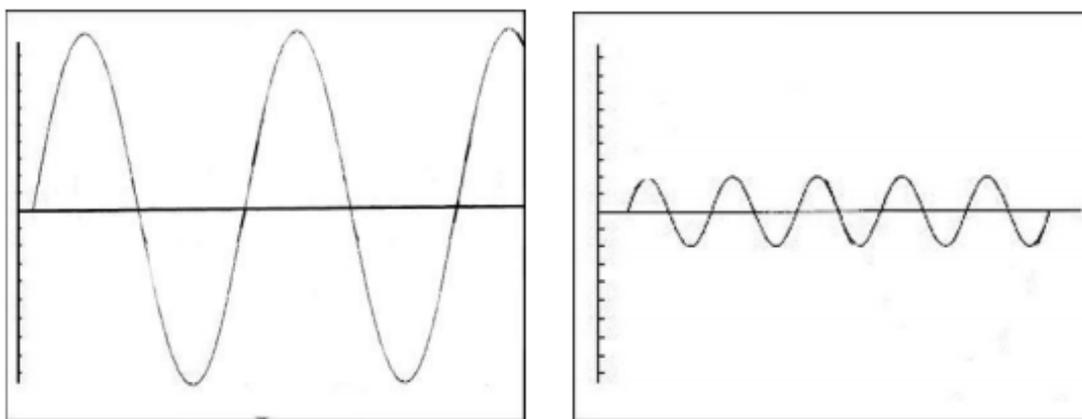


Figura 6: (a) Ondas curtas com alta frequência e (b) Ondas longas com baixa frequência. (SCURI, 2002).

Quando a oscilação aumenta, a amplitude da onda diminui, ou seja, o espaço percorrido é menor, ao se aproximar de nanômetros, o elétron libera um fóton e então temos a luz visível. Caso essa vibração continuasse teríamos os raios-x, raios-gama e assim por diante ilustrado na Figura 7 (SCURI, 2002).

A luz visível se encontra entre o tamanho de 370nm e a frequência de 790 THz e vai até 750nm e a frequência de 400 Thz. Esta porção é chamada de “espectro visível” (SCURI, 2002).

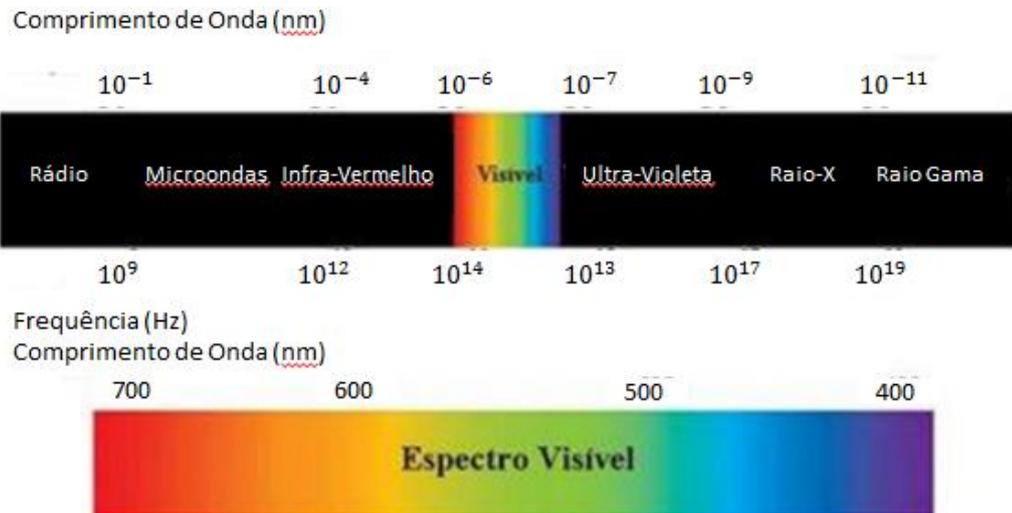


Figura 7: Comprimento de Onda e Espectro visível (SCURI, 2002).

Conforme (PENHARBEL et al., 2013), o sistema de imagens RGB, que é a abreviação do sistema de cores aditivas formada por vermelho (Red), verde (Green) e azul (Blue), representados na Figura 8(a), também tem o principal propósito de reproduzir cores em dispositivos eletrônicos como televisores, computadores, retroprojetores e demais dispositivos.

As cores aditivas RGB podem ser caracterizadas por um cubo, onde suas dimensões representam o lado R, o comprimento G e a altura B. A representação do formato RGB, na forma de cubo está ilustrada na Figura 8(b) onde é possível observar que as cores parecidas podem ser representadas por diversos valores de RGB podendo formar uma imagem digital (PENHARBEL et al., 2013).

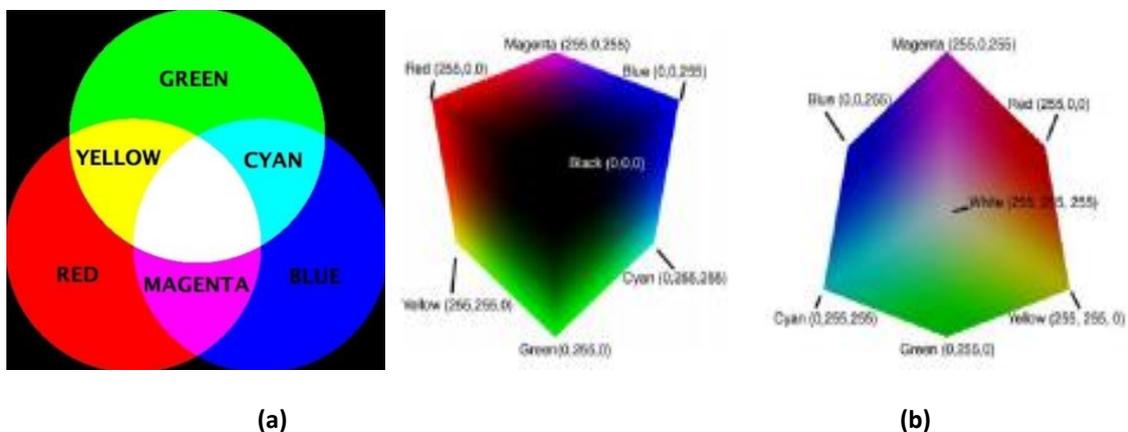


Figura 8: Sistema de cores RGB e Representação do cubo RGB.

## 2.2 Representação da imagem digital

A imagem digital pode ser capturada por câmeras fotográficas ou de vídeo. Estes sensores são capazes de coletar a energia proveniente do objeto, transformar o sinal para que possa ser registrado e apresentado de forma adequada à extração de informações. As câmeras são classificadas como sensores passivos, ou seja, não possuem uma fonte de radiação; elas medem a radiação solar refletida pelos objetos (SOUZA e CARDOZA, 2012).

Para (GONZALEZ et al., 2007), a imagem na Figura 9 pode ser representada por uma função bidimensional de  $f(X, Y)$ , onde  $X$  e  $Y$  indicam as coordenadas espaciais e  $f(X, Y)$  indica a intensidade do nível de cinza da imagem na dada coordenada  $(X, Y)$ .

Para o processamento computacional, a imagem digital pode ser representada como uma de matriz  $M \times N$ , onde cada posição  $A_{ij}$  indica um ponto da imagem e o valor contido nele representa a intensidade de brilho ou as cores RGB, como representado na Equação (1). Cada um dos pontos são denominados de *pixel* (“*Picture Elements*”) (GONZALEZ et al., 2007).



Figura 9: Convenção dos eixos para representação de imagens digitais (GONZALES et al., 2007).

$$f(X, Y) \cong \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0, N-1) \\ f(1,0) & f(1,1) & \dots & f(1, N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1, N-1) \end{bmatrix} \quad (1)$$

### 2.2.1 Processamento da Imagem Digital

Segundo (EASTON, 2010), as etapas de processamento de imagens digitais podem ser classificadas em duas abordagens principais, por escopo ou por resultado. Por escopo tem maior relação com as técnicas de processamento, enquanto por resultado está relacionada com a resposta de cada técnica.

A qualidade de imagem digital pode ser dividida em duas categorias: fidelidade e inteligibilidade. Processamentos que exigem fidelidade tem o objetivo de aproximar a imagem processada com a imagem original. Processamentos que se preocupam com a informação que se pretende extrair da imagem por algum processamento são classificados na categoria de inteligibilidade. Os programas para edição da imagem precisam ter cuidado com a fidelidade na forma de representar a cor, formas e texturas. Os programas de processamento, para reconhecimento e identificação de objetos, rostos e etc, são mais tolerantes com a baixa fidelidade das imagens (SCURI, 2002).

Para (SCURI, 2002), as etapas do processamento de imagens consistem em análises qualitativa e quantitativa. Cada análise possui outras subdivisões como ilustrado na Figura 10.

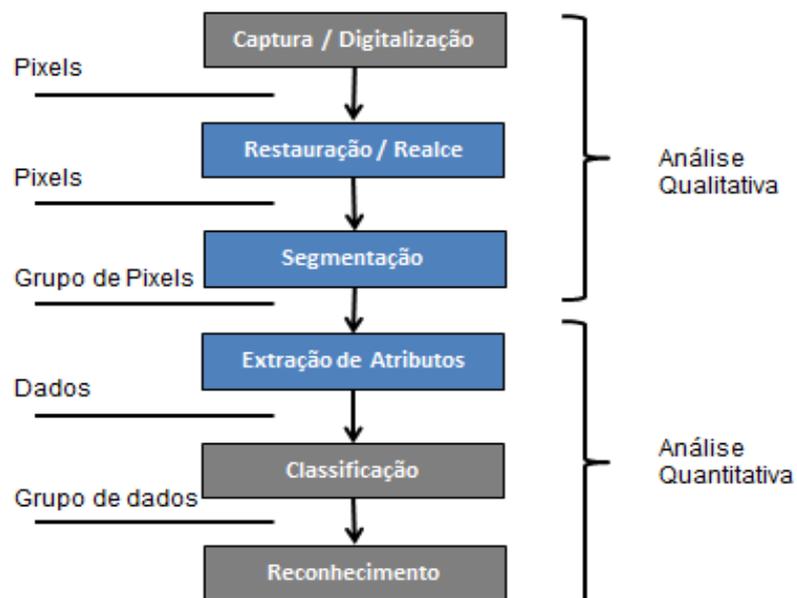


Figura 10: Sequência de etapas do Processamento de Imagens (SCURI, 2002).

As principais etapas do processamento de imagens consistem em:

- **Restauração/Realce:** busca compensar distorções específicas geradas na aquisição. Também é possível identificar a função da deformação e aplicá-la sobre a imagem deformada. São exemplos de reestruturação a correção de imagens sem foco e borradas por movimento;
- **Segmentação:** definição de região de interesse relacionadas com os objetivos encontrados. São necessárias etapas de Segmentação para isolar regiões e etapas de Extração de Atributos para encontrar diversos parâmetros. Operações mais comuns são as de limiarização por um limiar: pixels com intensidade acima de um determinado valor se tornam brancos, os demais se tornam pretos, obtendo assim uma imagem binária;
- **Extração de Atributos:** a partir das imagens binárias é possível extrair os dados apenas das regiões segmentadas como: número de objetos, propriedades geométricas, como área, perímetro, largura máxima e mínima, atributos relacionados a forma incluindo circularidade, concavidade, atributos relacionados a luminância como a média do nível de cinza de cada região, desvio padrão de cinza e também propriedades associadas a textura de cada região;
- **Classificação:** com as informações coletadas, é possível diferenciar objetos da imagem, separando classes distintas de acordo com sua semelhança. Esse processo é complexo e existem diversos níveis de automação;
- **Reconhecimento:** os processos mais sofisticados permitem reconhecer diferentes objetos através de técnicas de inteligência artificial que podem ser treinadas.

Para detecção de áreas de interesse ou objetos podem ser usados métodos baseados na detecção de bordas, tais como: *Sobel* e *Roberts*.

O filtro *Roberts* utiliza duas matrizes quadradas de ordem dois ilustradas na Figura 11, deslocadas em  $90^\circ$  para identificar as variações nas direções  $X$  e  $Y$ .

Semelhante ao filtro *Roberts*, o filtro *Sobel* também utiliza duas máscaras deslocadas em 90° para identificar variações nas direções *X* e *Y*. Essas máscaras identificadoras de bordas devem obter resposta de um valor máximo quando as mesmas estiverem orientadas em 45°(DIAS et al., 2011).

As máscaras são convoluídas com a imagem original e o resultado é obtido com a magnitude de cada direção. O filtro *Sobel* é menos sensível ao ruído devido as máscaras serem quadradas de ordem três ao invés de ordem dois (SANCHES et al., 2015). O filtro *Sobel* é usado neste estudo nas etapas de calibração dos sensores e identificação da estrada nas imagens RGB (GONZALEZ et al., 2007). As máscaras de cada filtro são ilustradas na Figura 11.

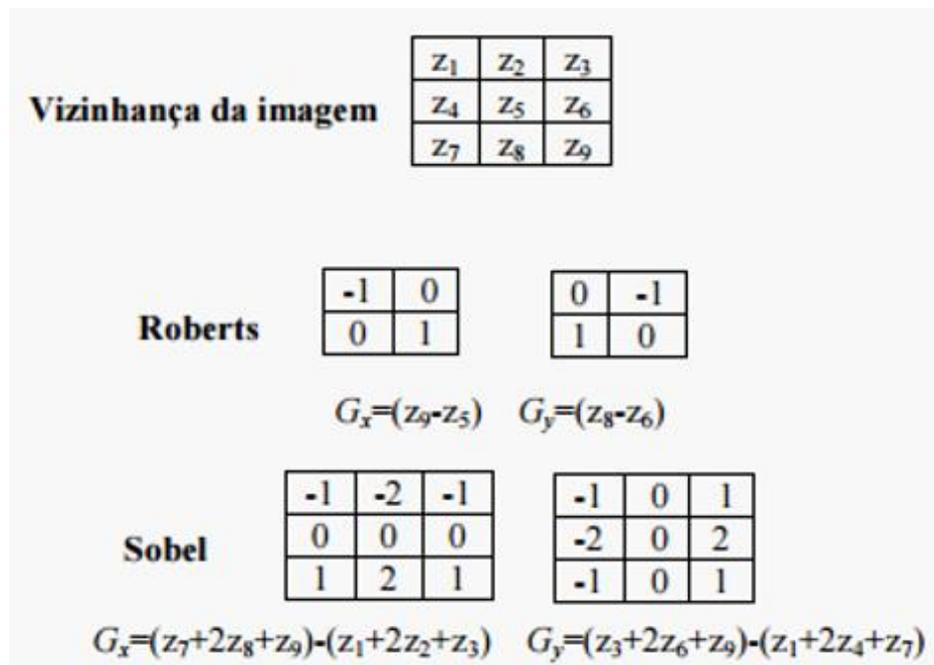


Figura 11: Máscara detectores de bordas e derivadas de primeira ordem (GONZALES et al., 2007).

### 2.3 Sistema Laser Scanner

O sistema *laser scanner* pode ser utilizado para gerar uma nuvem de pontos de duas ou três dimensões. O sistema gera pulsos de laser que são direcionados para o ambiente até ser interrompido por objetos ou pelo alcance do sensor. Tais objetos refletem os pulsos que retornam para o laser. Dessa forma, a distância entre o objeto

e o sensor pode ser calculada através do intervalo de tempo existente entre a emissão e reflexão de cada pulso emitido (DALMOLIN & SANTOS, 2004). Esse sistema de funcionamento é conhecido como “*time of flight*” (Figura 12) (BOEHLER et al., 2002).

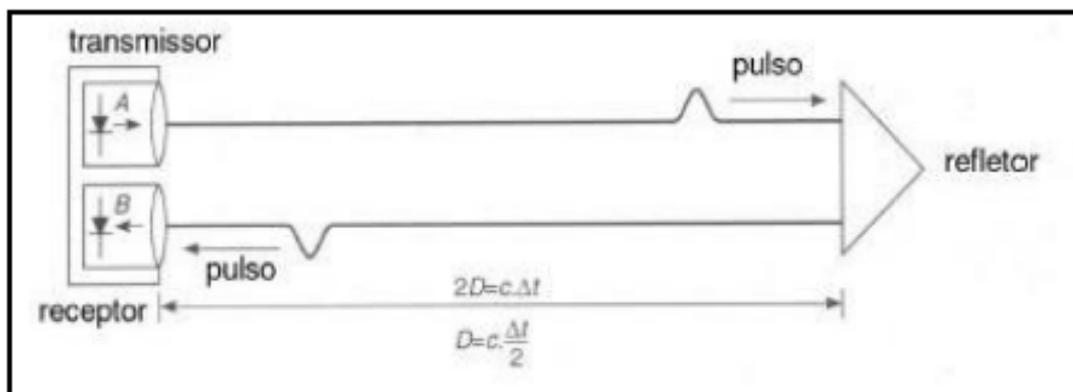


Figura 12: Princípio de *time of flight* (MAIA, 1999).

Para determinar as distâncias dos objetos é medida a variação de tempo ( $\Delta t$ ) que o sinal precisa para percorrer do sensor até o objeto, e então retornar ao sensor. Com base na velocidade da luz ( $c = 300 \times 10^3$  km/s) se calcula a distância ( $D$ ) entre o objeto e o sensor considerando somente metade do valor, pois o tempo medido é de envio e retorno do sinal. Assim chega-se na Equação (2) (LICHTI & HARVEY, 2002):

$$D = \frac{1}{2} \cdot c \cdot \Delta t \quad (2)$$

Os sensores laser captam distâncias relativas a si mesmo em coordenadas polares para sensores 2D sendo raio  $r$  e azimute  $\alpha$ . Para sensores 3D as coordenadas são esféricas sendo raio  $r$ , elevação  $\omega$  e azimute  $\alpha$ . É necessária a conversão dos dados esféricos do sensor para coordenadas cartesianas. A Figura 13 lista as fórmulas para converter as coordenadas esféricas  $(r, \omega, \alpha)$  em um sistema de coordenadas formado pelos vetores ortonormais  $X = [1 \ 0 \ 0]^T$ ,  $Y = [0 \ 1 \ 0]^T$  e  $Z = [0 \ 0 \ 1]^T$ .

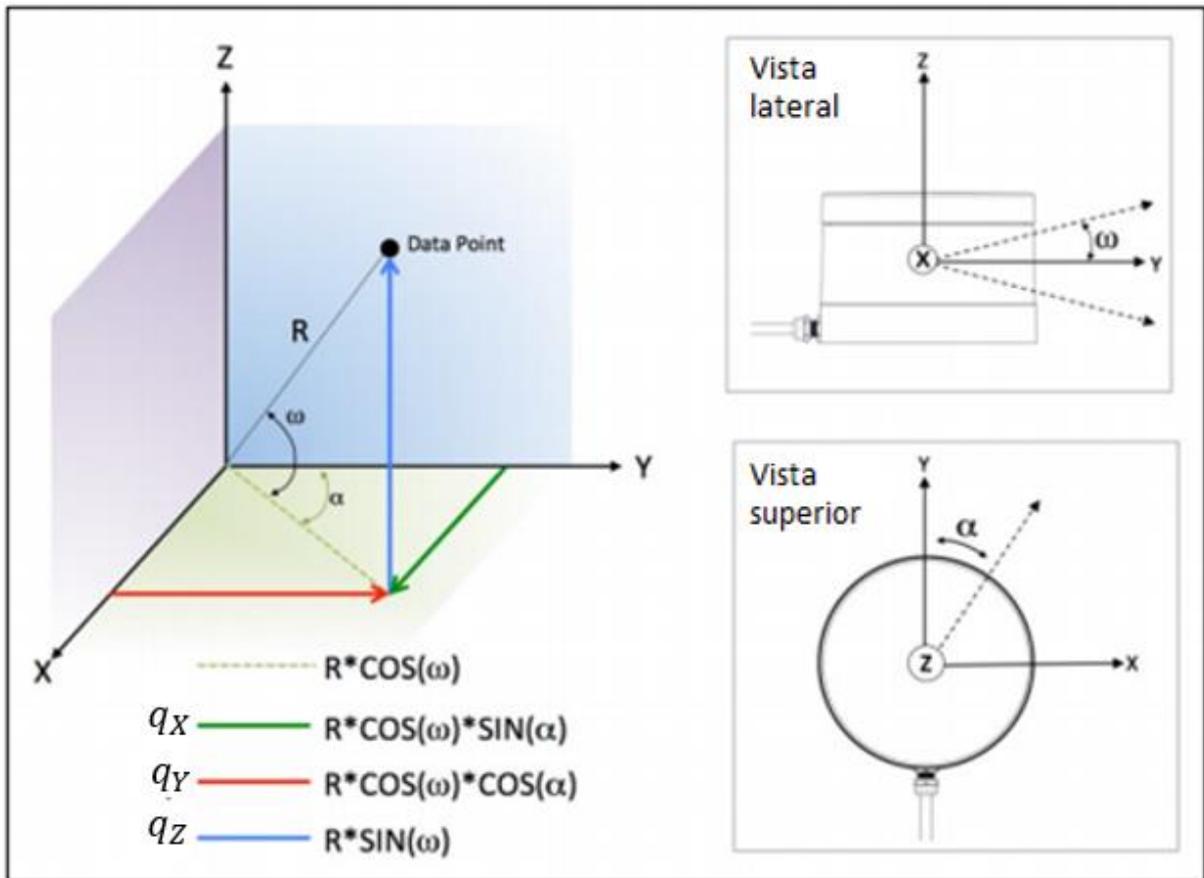


Figura 13: Sistema de Coordenadas do Sensor VLP-16, adaptado de velodyne.com manual do usuário.

Para (GROETELAARS, 2011), o *Laser Scanning* é uma tecnologia que executa uma varredura das superfícies dos objetos com diversos feixes de laser resultando em uma nuvem de pontos. Existem lasers com diversas características e dependendo da sua aplicação, das dimensões do cenário de coleta da nuvem de pontos, o laser pode ser usado na captura da geometria de inúmeros objetos, de tamanhos variados, de formas simples ou complexas com precisão milimétrica.

A representação obtida diretamente pelo *scanner* tridimensional é o modelo geométrico “nuvem de pontos”. Assim, cada ponto  $q_i$  obtido pelo laser possui 3 componentes com respeito ao sistema de coordenadas formado pelos vetores ortonormais  $(X, Y, Z)$ , tal que  $\vec{q}_i = [q_x \ q_y \ q_z]^T$  (HWANG et al., 2016).

Para (TOMMASELLI, 2003), os sistemas de varredura a laser possuem maior velocidade de coleta de dados e possibilidades de automação mais eficientes. Devido

à velocidade na coleta de dados, tempo de processamento e possibilidade de alta densidade de nuvem de pontos os lasers têm sido cada vez mais empregados em levantamento de superfícies. Seguem diversas outras características da varredura laser:

- É um método ativo, ou seja, o sensor emite feixes de laser e registra o sinal refletido, e são menos afetados por condições adversas de clima como neblina e poeira;
- O sensor pode ser operado remotamente, não havendo necessidade de intervenção humana durante seu funcionamento;
- Logo após a varredura, é possível ter a nuvem de pontos com identificação de superfícies, posicionamento de objetos em relação ao sensor e suas dimensões;
- É possível obter alta densidade da nuvem de pontos ocasionando na redundância de informações dos objetos;
- Realizando mais de uma varredura ou usando mais de um sensor, é possível combinar seus dados e cobrir grande parte das superfícies dos objetos;
- Diversos sistemas possuem software para identificação automática de objetos e realiza buscas de pontos isolados incluindo ou isolando o mesmo em aglomerados pertencentes a um objeto.

Existem dois sistemas de varredura laser: os sistemas estáticos que são posicionados em locais fixos e normalmente capturam dados do mesmo cenário, e os sistemas dinâmicos que a varredura pelo laser é realizada em plataformas móveis, como aviões ou veículos terrestres para medir a distância entre os objetos e o veículo que possui o sensor. Em ambos os sistemas de varredura os dados do laser podem ser coletados e posteriormente analisados ou processados em tempo real. Quando é necessário processamento e análise em tempo real se faz uso de processadores de alto desempenho embarcados (DALMOLIN & SANTOS, 2004).

Diversas áreas utilizam a varredura laser devido a precisão e rapidez da geração e aquisição dos dados e informações necessárias (ROCHA, 2002), são elas:

- Mineração e mapeamento geológico;
- Levantamento topográfico;
- Geotécnica, modelagem geofísica e de terrenos;
- Medições de estradas e ferrovias;
- Modelagem tridimensional de plantas industriais;
- Documentação de estruturas consideradas patrimônios históricos;
- *Design* de interiores;
- Fabricação de moldes de peças descontínuadas.

Para (ROCHA, 2002), o uso do laser scanner na mineração possui grandes vantagens em relação aos outros métodos tradicionais, o rendimento, a velocidade de atualização dos dados e a segurança da equipe de campo são fatores que destacam seu uso. Alguns exemplos de aplicações na mineração são:

- Medição de inclinação de talude;
- Cálculo de volume de pilhas pulmão e depósito de estéril;
- Monitoramento de erosões e trincas em taludes;
- Identificação de volume em caçambas dos caminhões de mina.

### **2.3.1 Nuvem de Pontos**

Durante o desenvolvimento de uma pesquisa em (MARCON e FERRAO, 2016), observou-se que para garantir a autonomia de um sistema para mapear ambientes e identificar objetos sem interferência humana são necessários sensores que medem profundidade. Portanto, métodos para detecção e extração de objetos em uma nuvem de pontos são de grande importância devido a sua presença cada vez mais frequente no sensoriamento de sistemas autônomos.

Num estudo realizado em (CÉSAR, 2016) mostra que a análise dos principais componentes e a classificação dos pontos na nuvem tridimensional pode ser realizada através de uma técnica para reduzir o número de variáveis usando Matriz de Variâncias e Covariâncias (MVC) da nuvem de pontos obtida. As variáveis que compõem o novo sistema de coordenadas, chamadas de Componentes Principais

(CP), podem ser encontradas da rotação do sistema original na mesma direção de variabilidade máxima dos dados, gerando uma descrição mais simples da estrutura de covariância (PONTOS e VIRTUAL, 2010).

Assim, em uma nuvem de pontos tridimensional, é possível encontrar a MVC de cada ponto considerando um subconjunto de  $i = 1, \dots, n$  pontos vizinhos localizados dentro de uma esfera de raio  $r$ , onde o centro é o próprio ponto de interesse  $q_i$  (CÉSAR, 2016), como ilustra a Figura 14.

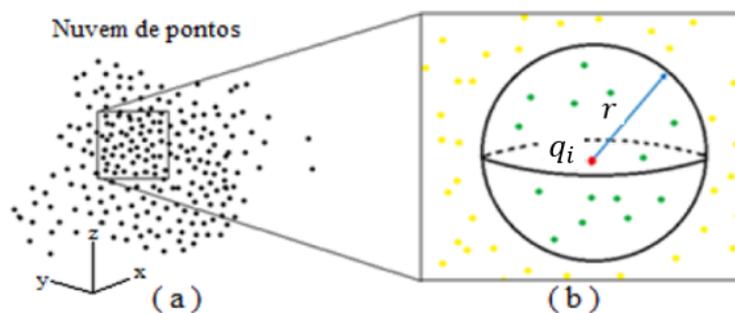


Figura 14: Nuvem de pontos tridimensional e seleção do conjunto de pontos situados a uma distância  $r$  do ponto central P (CESAR, 2016).

A técnica citada é normalmente usada para identificar aglomerados de pontos ou regiões planas na nuvem de pontos.

### 2.3.1.1 Extração de planos em nuvem de pontos

Para reconhecimento de planos em nuvem de pontos, uma abordagem amplamente usada é o método de consenso de amostragem aleatória, no inglês *Random Sample Consensus* (RANSAC), descrito por Fishler e Bolles em 1981 como um algoritmo de regressão robusta. Esse método tem sido aplicado com sucesso à extração de planos, tanto individuais (WAHL e colab., 2005) quanto integrado a diferentes *frameworks* de otimização (BAZARGANI et al., 2015).

A abordagem descrita por (TODA e KUBOTA, 2014) segue pela seguinte equação geral de um plano:

$$AX + BY + CZ + D = 0 \quad (3)$$

onde ao menos um dos parâmetros  $A$ ,  $B$ , e  $C$  devem ser não nulos. Supondo que um dado ponto  $\vec{q}_i = [q_X \ q_Y \ q_Z]^T$  no espaço euclidiano pertence ao plano definido pelo vetor de coeficientes da Equação (3),  $h = [A \ B \ C \ D]$ , tem-se, em notação vetorial, que:

$$[A \ B \ C \ D] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0. \quad (4)$$

$$h \cdot \begin{bmatrix} \vec{q}_i \\ 1 \end{bmatrix} = 0. \quad (5)$$

A distância euclidiana  $d_i$  entre qualquer ponto  $\vec{q}_i$  e um plano  $h$  é dada por:

$$d_i = \left| h \cdot \begin{bmatrix} \vec{q}_i \\ 1 \end{bmatrix} \right| \quad (6)$$

Se o ponto  $\vec{q}_i$  pertence ao plano  $h$  então  $d_i = 0$ , obedecendo à Equação (5).

Com a Equação (6), é possível determinar se um determinado ponto  $\vec{q}_i$  está dentro ou fora de um limite de distância  $t$  do plano  $h$ . Então, a quantidade de pontos da nuvem  $N = \{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n\}$  que estão dentro do limite de distância  $t$  do plano  $h$  pode ser computada usando a Equação (7):

$$f(h, N) = \sum_{i=1}^n (|h \cdot \begin{bmatrix} \vec{q}_i \\ 1 \end{bmatrix}| < t) \quad (7)$$

Esta é a função de aptidão para o RANSAC quando aplicado à extração de planos em nuvem de pontos. Esta função fornece a quantidade de pontos da nuvem  $N$  que se encontram a uma distância menor que  $t$  do plano candidato  $h$ . O limiar  $t$  é uma tolerância usada para compensar por irregularidades nos planos e ruídos nos dados (WAHL et al., 2005).

Os dados gerados pelo laser não são similares ao que é visualizado pelo olho humano, o que pode causar dificuldade à adaptação dos usuários como ilustrado na Figura 15. Porém, os dados do laser não sofrem grandes interferências como as câmeras RGB em condições climáticas ruins, o que possibilita identificar distâncias

com melhor precisão, desde que realizada por um sistema computadorizado (GROETELAARS, 2011).



Figura 15: Imagem de varredura de sensor de laser (velodynelidar.com).

## 2.4 Fusão de dados e calibração de sensores

Segundo (ROBERT A. SCHOWENGERDT, 2012), se divide os métodos de fusão de dados em três grupos: modelo de domínio espacial, domínio espectral e as que utilizam operações algébricas.

O método de domínio espacial separa as informações de alta frequência da imagem e a relaciona com a imagem multiespectral. A técnica *High-Pass Filter* e as funções que usam *Wavelets* fazem parte dessa categoria.

O modelo de domínio espectral possui processos para transformar a imagem multiespectral em um conjunto de bandas, sendo uma delas correlacionada com a imagem pancromática. São exemplos desse modelo as componentes *IHS*, *Gram-Schmidt*. As funções que operam pixel-a-pixel de forma aritmética com normalização de cores e de forma multiplicativa fazem parte dos modelos algébricos.

Segundo (VELAS et al., 2014), o grande desafio da tarefa de calibração é a capacidade de lidar com os sensores em posições diferentes, obter os requisitos mínimos para configurar uma cena preservando a precisão do resultado. Para cumprir esses objetivos, a calibração é realizada em duas etapas. Primeiro, uma calibração grosseira e depois um refinamento é aplicado para obter uma calibração mais precisa.

### 2.4.1 Calibração da câmera RGB com Laser

Informações sobre a calibração do sensor Velodyne com a câmera RGB podem ser encontradas em diversos trabalhos de classificação de imagens, percepção e

mapeamento de ambiente. Para este estudo será usado um método proposto em (VELAS et al., 2014), que descreve a calibração automática de imagens RGB com a nuvem de pontos do Velodyne.

De acordo com (VELAS et al., 2014), a solução consiste em duas etapas. Em primeiro lugar é realizada uma calibração grosseira, identificando um ponto que possa ser facilmente encontrado na câmera e nos dados do Velodyne, e então são encontrados os demais pontos correspondentes.

Após a calibração grossa usando um marcador 3D, é realizada uma estimativa mais precisa pela busca de todos os 6 graus de liberdade (translação e rotação). Esse método segue o processo de busca de grade regular proposto em (LEVINSON e THRUN, 2014), para sua detecção de erros de validação.

Os parâmetros finais de calibração são escolhidos como o ponto do espaço do parâmetro com a função máxima ou o ponto obtido pela média dos múltiplos pontos com o valor de aptidão acima de um determinado limiar. Mesmo no caso mais geral da calibração, todos os seis graus de liberdade que descrevem a translação e rotação devem ser considerados (VELAS et al., 2014). As etapas de calibração são detalhadas na seguinte sequência:

#### **2.4.1.1 Calibração grosseira:**

- a. Identificação de bordas na imagem RGB e na nuvem de pontos;
- b. Detecção de um marcador 3D (centros de círculos):
  - i. Na imagem da câmera RGB usando transformada de Hough;
  - ii. Na nuvem de pontos usando o algoritmo de detecção de planos baseado em RANSAC;
- c. Estimativa de tradução entre ambos os sensores no espaço tridimensional.

#### **2.4.1.2 Calibração fina**

- a. Pesquisa de um subespaço menor para calibração;
  - i. Projeção da nuvem de pontos com detecção de bordas para o plano da imagem;

- ii. Detecção de borda e a transformação de distância inversa da imagem da câmera;
- iii. Correlação cruzada dessas duas imagens.

A detecção de bordas na nuvem de pontos gerada pelo laser baseia-se no pressuposto de que as discontinuidades de profundidade nos dados representam bordas na cena real. Ou seja, é possível atribuir um determinado valor a cada ponto  $V_i$  medido que depende da sua medida de profundidade e da profundidade de seus pontos vizinhos conforme mostra a Equação (8) descrita em (VELAS et al., 2014).

$$V_i = \max(q_{i-1}^r - q_i^r, q_i^r - q_{i+1}^r, 0)^\gamma \quad (8)$$

onde  $q_i^r$  representa o alcance do i-ésimo ponto da nuvem de pontos e  $\gamma$  é a constante.

O menor  $\gamma$  é o valor mais alto atribuído aos pontos de baixa distorção relativamente aos pontos com alta distorção. Os valores de  $V_i$  são normalizados dentro de um range (0;1).

Para detecção e calibração através da câmera RGB, usa-se o operador *Sobel*, que produz saídas menos ruidosas (LEVINSON e THRUN, 2014). Os objetos então são detectados usando a transformação de *Hough* (YUENH.K., 1989).

## 2.5 Aprendizado de máquina

Como descrito anteriormente, os veículos inteligentes são sistemas complexos. É necessário que eles interpretem as variáveis do seu ambiente, identificando os obstáculos ao seu redor e a área de navegabilidade a fim de tornar a dirigibilidade correta e segura (HABERMANN, 2016).

Em um cenário de mineração é de grande importância que o sistema do veículo inteligente identifique outros equipamentos, obstáculos diversos e a pista de rolagem. Para essa identificação são empregadas técnicas de aprendizado de máquina para identificar os obstáculos e superfícies ao seu redor do equipamento de forma automática. As pesquisas nessa área de conhecimento vêm sendo aprimorada nas

últimas décadas para serem aplicadas em diversas situações como fraudes em operações bancárias até a operação de veículos inteligentes (MITCHELL, 1997).

Para (P. MURPHY, 2012), a definição de aprendizado de máquina é o conjunto de métodos computacionais que pode detectar de forma automática os padrões em dados e posteriormente usar esses padrões para fazer previsões sobre os dados futuros.

As técnicas de aprendizagem de máquina são divididas em três grupos: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço (P. MURPHY, 2012). A seguir são apresentados o aprendizado supervisionado e não supervisionado; o aprendizado por reforço não será abordado neste trabalho.

### **2.5.1 Aprendizado supervisionado**

O principal objetivo do aprendizado supervisionado é usar exemplos que estão pré-classificados para induzir o modelo de classificação, ou seja, exemplos que estão nomeados dentro de uma classe determinada, também conhecida por conjunto de treinamento (HABERMANN, 2016).

### **2.5.2 AdaBoost**

O AdaBoost é um método de aprendizado de máquinas criado por Freund e Shapire no final da década de noventa. Seu princípio de funcionamento se baseia na ideia de criar uma regra de predição de elevada precisão por meio da combinação de várias regras de classificação de baixa precisão, chamados de classificadores fracos. O AdaBoost é um método adaptável: as classificações subsequentes são ajustadas favorecendo as instâncias anteriores classificadas negativamente. Esse algoritmo utiliza classificadores simples que com uma reta separa os dados em dois grupos (HABERMANN, 2016).

### 2.5.3 Support Vector Machine

Segundo (SCHÖLKOPF, 2001), *Support Vector Machine* (SVM) é um método de aprendizado de máquina supervisionado utilizado para classificação. Essa técnica procura encontrar o melhor hiperplano capaz de separar o conjunto de dados de treinamento em duas classes. No hiperplano existem margens que são determinadas pela distância entre este e o elemento mais próximo a ele. Quanto maior a margem de ambas as classes melhor o hiperplano. Na Figura 16, existem três hiperplanos, H1, H2 e H3. O hiperplano H2 e H3 conseguem separar o conjunto de treinamento, porém o H1 não consegue. Comparando H2 e H3, percebe-se que a margem de H3 (identificado na figura por  $W_3$ ) é maior do que H2.

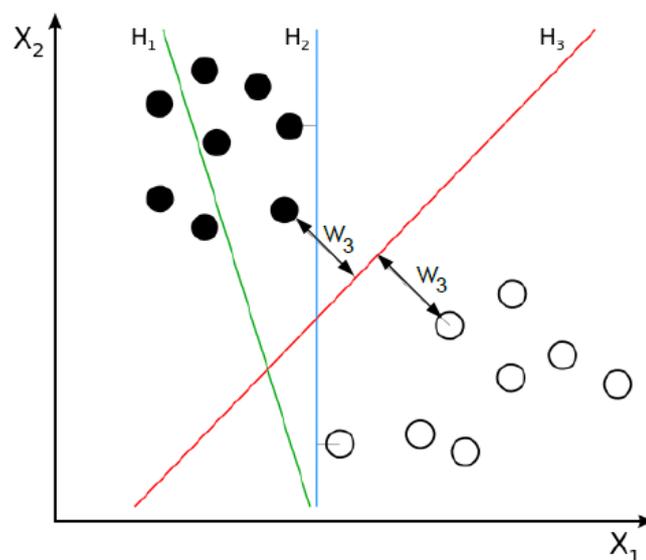


Figura 16: Hiperplanos separando pontos no treinamento. Hiperplano H3 possui maior margem (HABERMANN, 2016).

Na Figura 17, o hiperplano é representado pela linha vermelha e as setas entre as linhas tracejadas indicam a dimensão da margem. Na Figura 17(a), é apresentado o caso linear e a Figura 17(b) o caso não linear. Uma função *Kernel*, que é uma função não paramétrica para estimar a densidade de probabilidade relacionada a uma variável aleatória, deve ser usada no caso não linear para criar um espaço onde as classes são separáveis por um hiperplano. Uma vez feito isso, o método torna-se similar ao caso linear (SCHÖLKOPF, 2001).

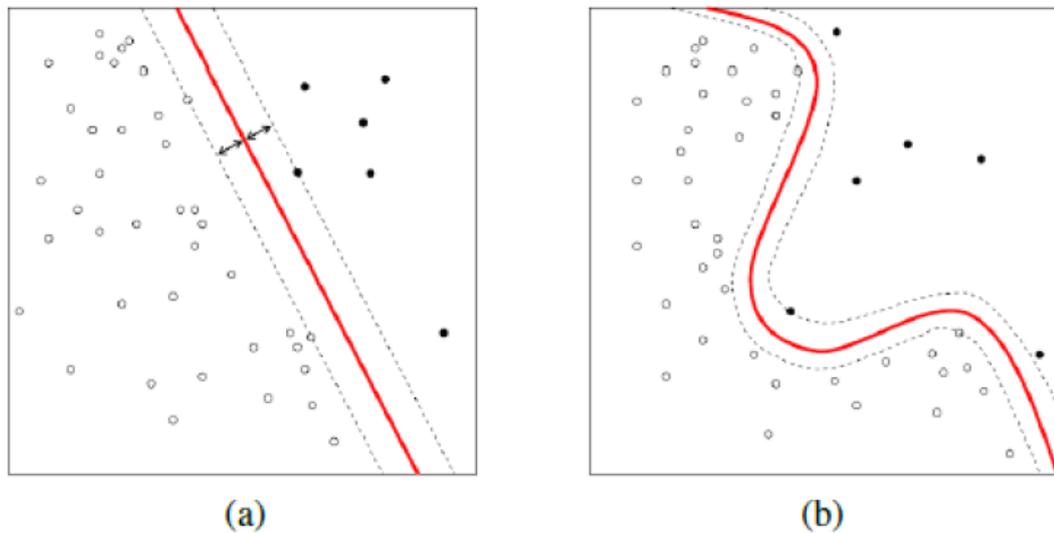


Figura 17: Separação de pontos por hiperplanos. Em (a) caso linear e (b) caso não linear (HABERMANN, 2016).

#### 2.5.4 Redes Neurais Artificiais

Para (MITCHELL, 1997), as Redes Neurais Artificiais (RNA) são modelos matemáticos inspirados na complexa rede de neurônios do cérebro humano. Essas RNAs podem ser aplicadas com sucesso para aprendizado de exemplos que possuem funções de valores reais ou discretos.

Para (HABERMANN, 2016), as características que mais se destacam nas RNA são a capacidade de aprender através de exemplos (padrões, amostras, etc), a generalização do conhecimento recebido, a atuação como mapeadores de entrada e saída (aprendizado supervisionado), além de ser tolerante a falhas.

O *Multi-Layer Perceptron* (MLP) é umas das RNAs mais conhecidas. Um estudo realizado por (HABERMANN, 2016) descreve as RNAs como um conjunto de neurônios localizados em camadas organizadas em sequência. Geralmente uma rede neural MLP dispõe de 3 camadas, como ilustrado na Figura 18. A camada de entrada, que recebe os dados a serem inseridos na rede, a camada escondida, a qual realiza o cálculo dos pesos (conexões) com os dados de entrada, e a camada de saída, que fornece os resultados (HABERMANN, 2016).

Cada neurônio de uma camada é conectado com todos os neurônios das camadas posteriores e anteriores, com um determinado peso. Para que os dados de entrada sejam corretamente relacionados às saídas esperadas, é preciso estimar corretamente os pesos que conectam os neurônios da rede. A estimativa é realizada na etapa de aprendizado, que consiste em usar exemplos já classificados e um algoritmo próprio de treinamento (HABERMANN, 2016).

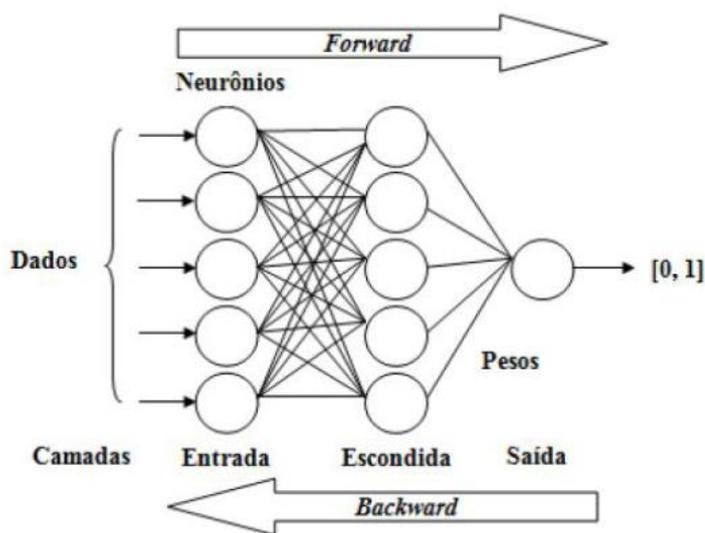


Figura 18: Ilustração de uma rede neural artificial (Haykin, 1998).

O método de treinamento das RNA MLP comumente usado é o aprendizado supervisionado *back-propagation*. Esse algoritmo opera em forma de pares de entrada e saída por meio da correção de erros, ajustando os pesos da RNA. O treinamento acontece em duas etapas: *forward* e *backward*. A etapa *forward* define a saída da rede para um determinado padrão de entrada. Já a etapa *backward* utiliza a saída fornecida pela RNA e a saída desejada para determinar o erro da resposta e atualizar os pesos de suas conexões (HABERMANN, 2016).

Existem também as Redes Neurais Convolucionais (CNNs) que são baseadas nas RNAs, mas utilizam a convolução em uma ou mais camadas de sua estrutura. As CNNs se mostraram ser eficientes em tarefas de reconhecimento de vídeo e imagens, porém necessitam de grande quantidade de dados rotulados para o treinamento (CARNEIRO e SILVA, 2017).

As CNNs são compostas por uma sequência de camadas e cada uma destas possui uma função específica na propagação do sinal de entrada. A Figura 19 mostra

a arquitetura de uma *Convnet* e suas três principais camadas: convolucionais, de *pooling* e totalmente conectadas. Os atributos dos volumes de entrada são extraídos pelas camadas convolucionais. As camadas de *pooling* tem a finalidade de reduzir a dimensão do volume resultante posteriormente das camadas convolucionais e ajudam a tornar a representação invariante e pequenas translações de entrada. As camadas totalmente conectadas têm a responsabilidade de propagar o sinal através da multiplicação ponto a ponto e também o uso de uma função de ativação. O resultado de saída de uma CNN é basicamente a probabilidade da imagem de entrada pertencer a uma das classes para qual a rede foi treinada (REN et al., 2017).

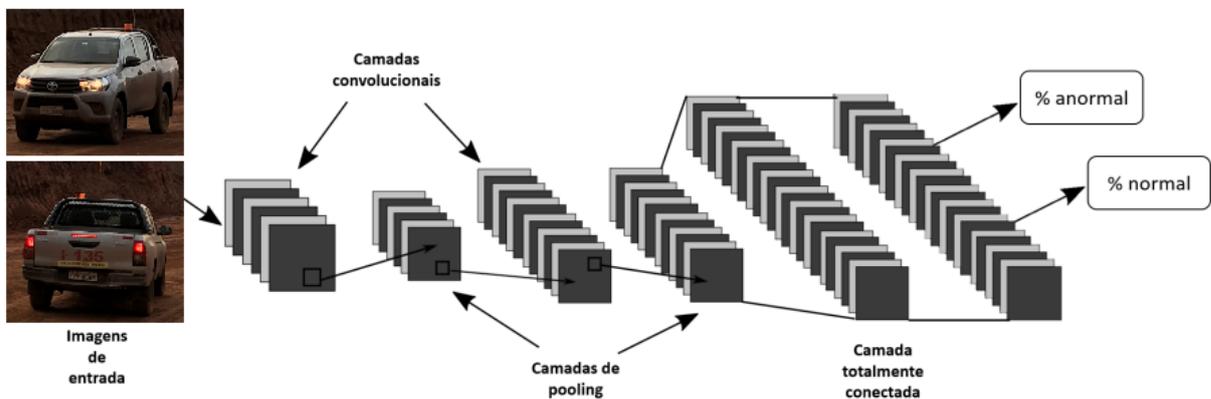


Figura 19: Configuração da rede neural *Convnet* classificadora de imagens e suas principais camadas. Adaptado de (Ren et al., 2017).

### 2.5.4.1 Camada convolucional

Para (SALIMANS et al., 2017), as camadas convolucionais são compostas por um conjunto de convoluções que recebem um volume ou arranjo 3D na entrada. As convoluções se estendem por toda profundidade do volume de entrada e possui dimensões reduzidas. Por exemplo, quando a imagem é colorida ela possui três canais e o filtro pertencente à primeira camada convolucional será 5x5x3, sendo 5 pixels de altura, 5 de largura e 3 de profundidade. Durante o processo de treinamento da rede, esses filtros são ajustados para que sejam ativados somente quando houver presença de características importantes identificadas no volume de entrada, como orientação de bordas ou manchas de cores.

Cada filtro dá origem a uma estrutura conectada localmente que percorre toda extensão do volume de entrada. A operação de convolução é a somatória do produto ponto a ponto entre os valores do filtro e a cada posição do volume de entrada, como ilustrada na Figura 20. Os valores resultantes após a operação de convolução são destinados a uma função de ativação, a função *ReLU* (*Rectified Linear Units*) é a mais comum e pode ser calculada pela Equação (9), onde  $f(x)$  é o valor da saída e o  $x$  e a quantidade de entradas (SALIMANS et al., 2017).

$$f(x) = \max(0, x) \quad (9)$$

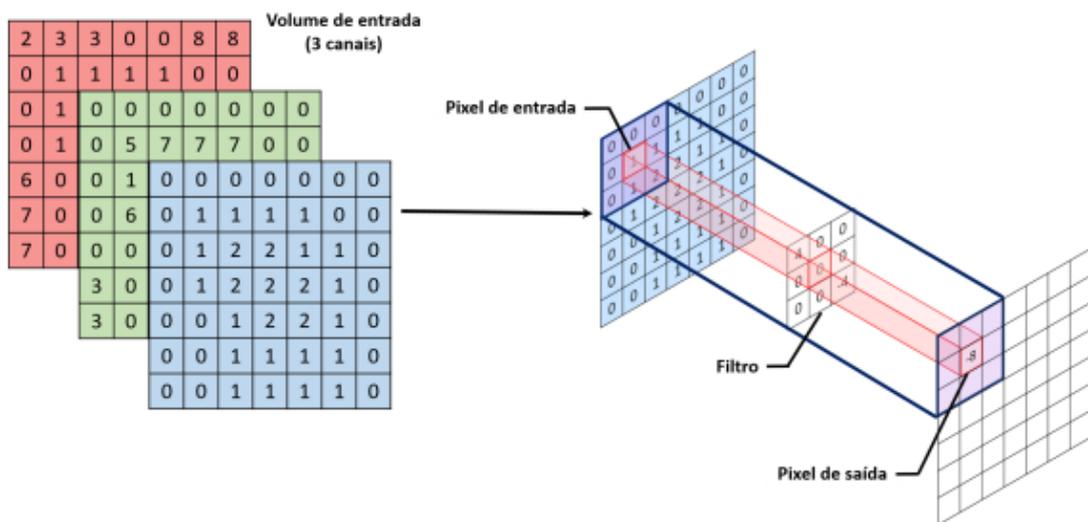


Figura 20: Representação de uma convolução de filtro de ordem 3 e o volume de entrada (CARNEIRO e SILVA, 2017).

As dimensões do volume resultante após a operação de convolução são controladas por três parâmetros: passo (*stride*), profundidade (*depth*) e *zero-padding* (Ujjwal 2016). Pode-se observar que a profundidade (*depth*) do volume resultante é igual ao número de filtros utilizados. Cada um dos filtros é responsável por extrair diferentes características no volume de entrada. Então, quanto maior o número de filtros, maior o número de características se pode extrair, porém a complexidade computacional também será maior (SALIMANS et al., 2017).

Os parâmetros do passo e *zero-padding* determinam a altura e largura do volume resultante. O passo especifica o tamanho do salto na operação de convolução, como mostra na Figura 21. Quando o passo é 1, o filtro salta uma posição por vez. Quando o passo é 2, o filtro salta duas posições por vez. Quanto maior o valor do passo, a

largura e altura do volume resultante serão menores. Porém, características importantes podem ser perdidas. Por esse motivo, não é comum utilizar um valor de salto maior que 2.

A operação de *zero-padding* tem a função de preencher a borda do volume de entrada com zeros. Essa operação permite controlar a largura e altura do volume resultante, fazendo com que eles tenham as mesmas dimensões do volume de entrada.

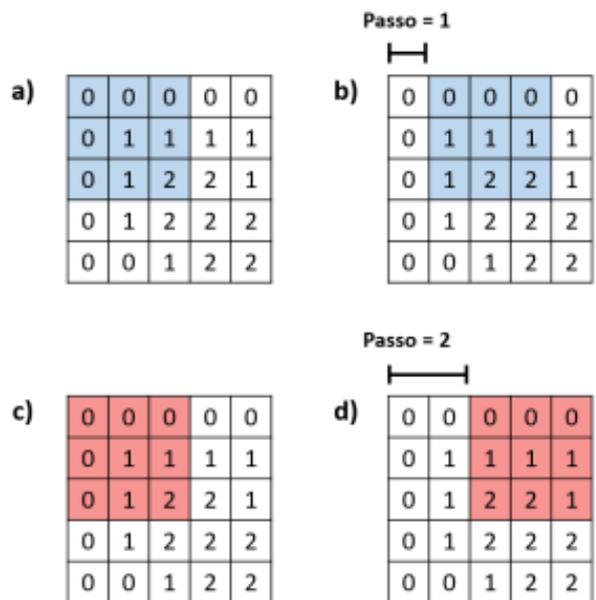


Figura 21: Representação da atuação do passo no deslocamento do filtro de ordem 3 em duas etapas seguidas de convolução. Em (a) e (b) com passo 1 e em (c) e (d) com passo 2 (CARNEIRO e SILVA, 2017).

Portanto, é possível calcular a largura ( $LC$ ) e a altura ( $AC$ ) do volume resultante da camada convolucional através das Equações (10) e (11), respectivamente.

$$LC = \frac{L - F + 2P}{S} + 1 \quad (10)$$

$$AC = \frac{A - F + 2P}{S} + 1 \quad (11)$$

onde  $L$  e  $A$  correspondem a largura e altura do volume de entrada,  $S$  é o valor do passo,  $F$  é o tamanho dos filtros utilizados e  $P$  corresponde ao valor do *zero-padding* (CARNEIRO e SILVA, 2017).

### 2.5.4.2 Camada de *Pooling*

Normalmente existe uma camada de *Pooling* após uma camada convolucional com objetivo de reduzir gradualmente as dimensões espaciais do volume de entrada. Conseqüentemente, a redução diminui o custo computacional da rede (SALIMANS et al., 2017). Nesta operação os valores gerados pelas camadas convolucionais, que pertencem a uma determinada região, são substituídos por alguma métrica dessa mesma região. A forma mais comum de *pooling* consiste em substituir os valores de uma região pelo valor máximo (GOODFELLOW et al., 2016), como ilustra a Figura 22.

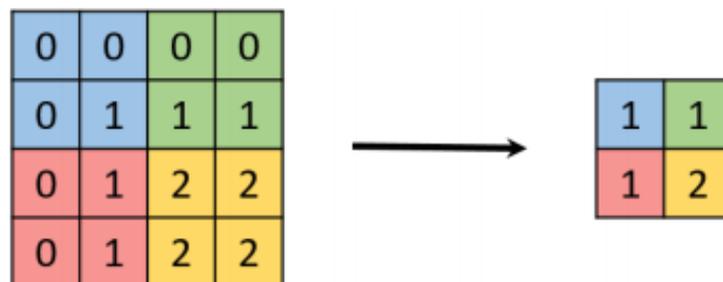


Figura 22: Representação de uma operação *max pooling* em uma imagem de ordem 4 utilizando filtro de ordem 2. Reduz o tamanho da imagem e o processamento para camadas seguintes (FERREIRA, Santos 2017).

Essa operação é chamada de *max pooling*, sendo útil para reduzir as dimensões da representação dos dados eliminando os valores desprezíveis. Além disso, essa operação consegue acelerar a computação necessária para as etapas seguintes (FERREIRA, Santos, 2017).

A largura ( $LP$ ) e a altura ( $AP$ ) do volume resultante após realizada a operação *max pooling* podem ser calculados pelas seguintes Equações (12) e (13), respectivamente:

$$LP = \frac{L - F}{S} + 1 \quad (12)$$

$$AP = \frac{A - F}{S} + 1 \quad (13)$$

onde as variáveis  $L$  e  $A$  correspondem respectivamente à largura e altura do volume de entrada, sendo  $S$  o valor do passo e  $F$  o tamanho da janela utilizada. Lembrando que a profundidade do volume de entrada não se altera pela operação *max pooling*.

### 2.5.4.3 Camada totalmente conectada

Os dados de saída das camadas convolucionais e da operação *max pooling* representam as informações extraídas da imagem de entrada. O objetivo das camadas totalmente conectadas é utilizar essas características para classificar a imagem em uma classe pré-determinada, como mostra na Figura 23. As camadas totalmente conectadas são exatamente iguais a uma rede neural artificial convencional (*Multi Layer Perceptron* ou MLP) (SALIMANS et al., 2017), que utiliza funções de ativação *softmax* na última camada de saída.

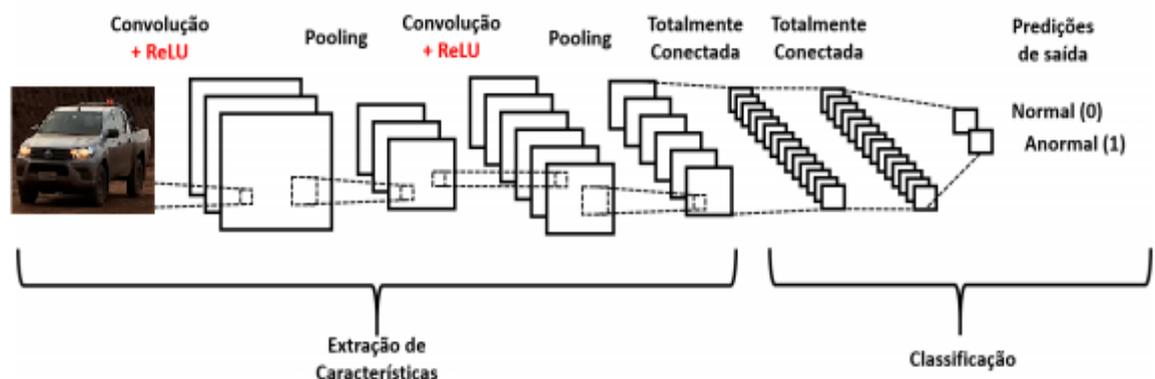


Figura 23: Representação de uma extração de padrões da imagem de entrada por uma rede neural *Convnet* e sua classificação. Adaptado de (SALIMANS et al., 2017).

Essas camadas são constituídas por unidades de processamento, também conhecida por neurônio. O termo “totalmente conectado” se refere aos neurônios da camada anterior que estão conectados com todos os neurônios da camada posterior (GOODFELLOW et al., 2016).

### 2.5.4.4 Camada *Softmax*

A camada *softmax* é a última da rede neural utilizada. Uma propriedade dessa camada é que a soma das saídas é 1 e todas as saídas são positivas. Dessa forma pode-se interpretar a saída da rede como uma distribuição de probabilidade discreta da entrada pertencer a cada uma das classes de interesse (ROCHA, 2015).

As CNNs são usadas neste estudo para identificação de obstáculos nas imagens RGB.

### **2.5.5 Aprendizado não supervisionado**

O aprendizado de máquina não supervisionado não possui classes já pré-determinadas. O principal ponto é encontrar como os dados estão organizados, ou seja, selecionar os dados em diferentes classes ou grupos, baseado na semelhança dos padrões sem fornecer previamente uma nomenclatura associada a cada padrão e a uma determinada classe (HABERMANN, 2016).

O objetivo é identificar os padrões nos dados de entrada e distingui-los por diferentes categorias. Se dois elementos possuem padrões semelhantes, ambos deverão ser associados a mesma classe dentro da aprendizagem não supervisionada. Caso seja encontrado algum valor de entrada que não se assemelha ao de nenhuma classe já estabelecida, então é criada uma nova classe para abrigar esse valor de entrada (HABERMANN, 2016).

Para (FERREIRA, 2012), existem três principais métodos para agrupamento de dados, baseados em partições, em densidade e em distribuição. Para segmentação de dados gerados por um sensor laser o método baseado em partições é o mais indicado quando existe grande quantidade de dados (KLASING et al., 2008). Mas ainda assim a otimização ou segmentação de nuvem de pontos tridimensional pode ser realizada por três técnicas descritas a seguir:

### **2.5.6 RANSAC**

O RANSAC, já descrito anteriormente é usualmente empregado para detectar os planos em uma nuvem de pontos. A principal diferença desse método é que se inicia com uma quantidade menor de dados e se expande quando possível (FISCHLER e BOLLES, 1981).

Esse método é usado neste estudo nas etapas de calibração dos sensores e identificação de planos da pista de rolagem.

### **2.5.7 DBSCAN**

Esse método não supervisionado efetua uma varredura na nuvem de pontos permitindo gerar aglomerados. São dois parâmetros base neste método, o número

mínimo de pontos necessários para criar um aglomerado, e a distância base que delimita a vizinhança de um ponto. Uma vizinhança é formada quando encontrado uma quantidade mínima de pontos dentro de um determinado raio. Então é iniciado um aglomerado, cada ponto deste segue o mesmo padrão acrescentando pontos no conglomerado ou classificando-o como ruído (ESTER, 1996).

O DBSCAN é usado neste estudo para identificação de aglomerados de pontos pertencente a um objeto que possa ser um obstáculo.

### **2.5.8 Iterative Closest Point (ICP)**

De acordo com (SEGAL et al., 2009) o *Iterative Closest Point* (ICP) é um algoritmo usado para alinhar as nuvens de pontos, ou seja, reduzir a diferença de posição entre elas. Porém possui custo operacional elevado.

O ICP pode ser descrito em dois passos (SEGAL et al., 2009):

1. Calcular a correspondência entre dois escaneamentos;
2. Calcular a transformação que minimiza a distância entre os pontos correspondentes.

Neste capítulo foi apresentado o princípio de funcionamento dos sensores câmera RGB e laser. Também foi descrito algumas formas de processamento dos dados gerados pelos sensores, tanto em aprendizado supervisionado como não supervisionado. No próximo capítulo é descrito os materiais e métodos usados neste estudo.

### 3 MATERIAIS E MÉTODOS

Este capítulo descreve os materiais e os métodos para classificação de obstáculos e pista de rolagem em uma mina a céu aberto. Os dados coletados em campo foram utilizados para realizar o processamento das imagens e nuvem de pontos no MATLAB.

Para identificação e prova de conceito de processamento foi utilizada uma base de imagens RGB e laser em um ambiente urbano e outro em ambiente de mineração, ambos contemplando estrada e obstáculos. Os materiais utilizados para coleta de dados são dispositivos disponíveis no laboratório de robótica do ITV. Como descrito anteriormente, a coleta de dados foi realizada em cenários com boa visibilidade dos obstáculos e estrada.

#### 3.1 Materiais

O sensor usado para captura de imagens RGB foi uma câmera do fabricante AXIS, modelo P1224-E que é composta em duas partes: uma IP66 – unidade de sensor nominal (lente e sensor de imagem), que pode ser instalada dentro ou fora e em espaços limitados, e uma unidade principal, que pode ser colocada até oito metros de distância em uma área interna (Figura 24). Esse modelo de câmera possui um campo de visão horizontal de 145°.



Figura 24: Câmera AXIS P1224-E ([axis.com/br](http://axis.com/br)).

Para o mapeamento da nuvem de pontos, o sensor usado foi o VLP-16 do fabricante Velodyne LiDAR (Figura 25). O VLP-16 envia sinal para 360° na horizontal

e 30° na vertical usando dezesseis feixes de laser montados em uma caixa circular compacta. O laser dispara milhares de vezes por segundo, fornecendo uma nuvem de pontos 3D em tempo real. Suas características incluem:

- 16 feixes de laser que registram ~300.000 pontos por segundo;
- Campo de visão horizontal de 360°;
- Campo de visão vertical de 30°;
- Velocidade de rotação de 5-20 rotações por segundo (ajustável);
- Retornos de até 100 metros (o intervalo útil depende de sua aplicação).



**Figura 25: Sensor Velodyne LiDAR VLP-16 (velodyne.com).**

Para aumentar o número de pontos devido ao campo de visão vertical de 30° do VLP-16, o laser pode ser acoplado uma unidade de controle de movimentos (*Pan Tilt unit* - PTU). Para esse estudo foi usado o modelo PTU-D47 do fabricante FLIR, ilustrado na Figura 26.



**Figura 26: Unidade para controle de movimentos PTU-D47.**

Para alimentação dos dispositivos eletrônicos e sensores, é necessário inserir uma bateria portátil recarregável no conjunto. O modelo BT-70791 ilustrado na Figura 27 atende as necessidades para efetuar os testes.



<b>Dimensions</b>	Length 112 mm Width 61 mm Height 127 mm
<b>Weight</b>	1.4 kg
<b>Nominal Voltage</b>	20.0 V, Two (2) x 14.4 V / section
<b>Maximum Voltage</b>	33.0 V, Two (2) x 16.5 V / section * 33.6 V, Two (2) x 16.8 V / section (with the BTA 70903 adaptor)
<b>Capacity</b>	9.9 Ah in 24 V Mode 19.8 Ah in 12 V Mode
<b>Discharge</b>	10 A maximum continuous / section
<b>Pulse Discharge</b>	18 A (5 sec ON/25 sec OFF) / section
<b>Operating Temperature</b>	- 20°C to + 60°C
<b>Recommended Storage Temperature</b>	- 40°C to + 40°C
<b>Connector</b>	Floating type per US Army DWG # SC-C-179495
<b>State of Charge Indicator</b>	2 separate 5 segment LCDs with constant display
<b>Disposal</b>	Check local regulations (Contains 0% Mercury or Cadmium)

Figura 27: Bateria portátil recarregável BT-70791.

Para alimentação dos sensores com a bateria descrita é necessário o uso de conversores de tensão, pois a bateria fornece 28,8 volts quando totalmente carregada e ambos os sensores necessitam de 12 volts de alimentação. Portanto, foi usado o conversor DC/DC da VICOR V48C12C200BL ilustrado na Figura 28.

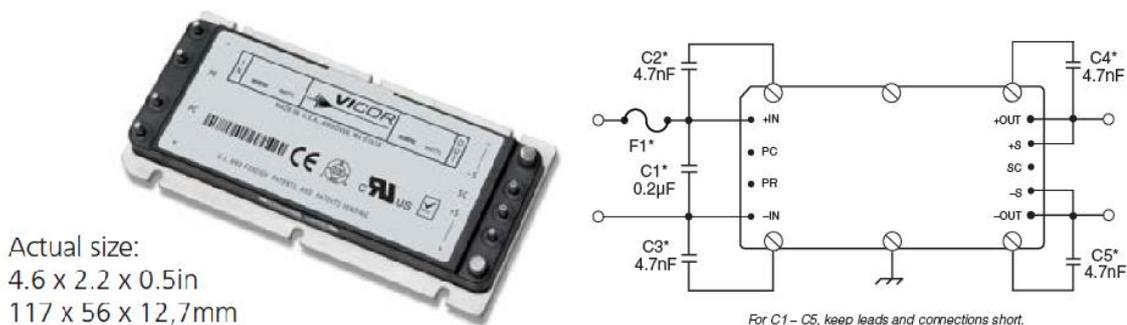


Figura 28: Conversor de tensão DC/DC do fabricante VICOR.

O processamento dos dados foi realizado em um Notebook com processador Core 2 Duo, com 4 Gigabytes de memória RAM e o software usado é o MATLAB 15.

### 3.2 Métodos

Antes de iniciar a aquisição dos dados, é necessário analisar se o cenário corresponde com o objetivo do estudo, se possui objetos e estradas bem definidas, as distâncias dos mesmos e o detalhamento necessário. É recomendado o registro fotográfico do ambiente para facilitar a identificação dos objetos a serem estudados, e de problemas que podem vir a acontecer, como locais com sombras ou com baixa luminosidade e áreas de difícil acesso.

Durante o procedimento de varredura a laser para capturar os dados é necessário fotografar os cenários com objetivo de registrar todos os obstáculos do ambiente, estradas, veículos, taludes, placas, entre outros.

Os dados da nuvem de pontos são gerados em 360° horizontalmente e 30° verticalmente e a câmera RGB possui campo de visão horizontal de 145°. Portanto, os pontos da nuvem não correlacionados com a imagem RGB devem ser descartados para este trabalho.

Conforme observado em (VELAS et al., 2014), nesta etapa é de grande importância que os sensores estejam instalados o mais próximo possível um do outro. Isso facilita a calibração dos sensores para obtenção do mesmo cenário.

Os processos de identificação por imagem e por nuvem de pontos são realizados separadamente, ou seja, sem interação entre eles. Posteriormente, são usados os dados de ambos os sensores para identificação dos obstáculos por fusão dos dados. Na Figura 29 é ilustrado o fluxograma para identificação de estradas e obstáculos e nas sessões a seguir será apresentado cada método.

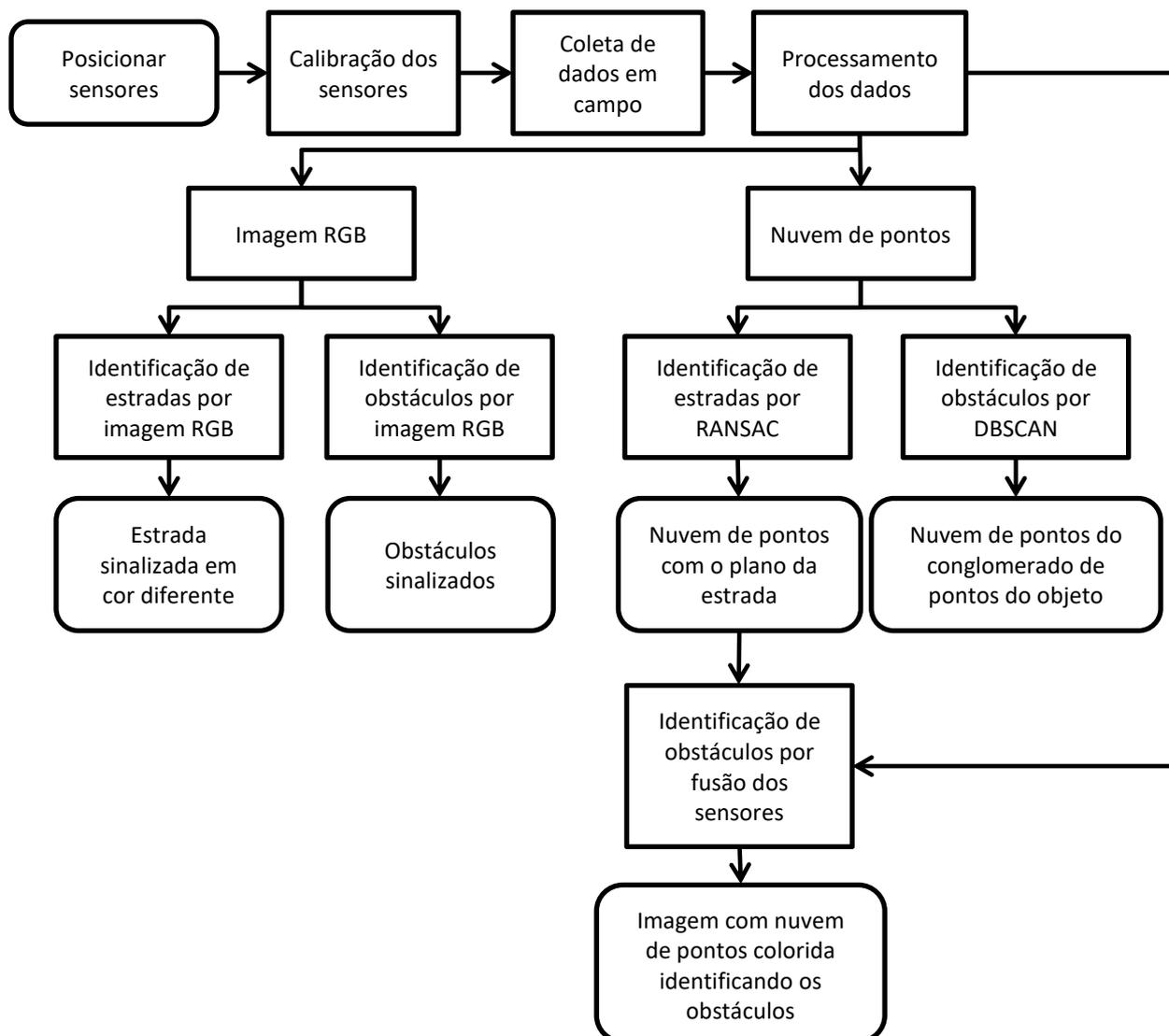


Figura 29: Fluxograma do processo de detecção de estradas e obstáculos.

### 3.2.1 Calibração da Câmera RGB com Laser

Os sensores descritos anteriormente foram fixados na parte superior de uma caixa de madeira e os acessórios na parte interior da mesma para melhor proteção contra intempéries como ilustrado na Figura 30.



Figura 30: Suporte para os sensores e acessórios.

Após a fixação no suporte da câmera e do laser é necessário calibrar os sensores. Essa etapa foi realizada no laboratório do ITV utilizando método de calibração descrito em (VELAS et al., 2014). Na Figura 31 é apresentado um fluxograma do das etapas da calibração.

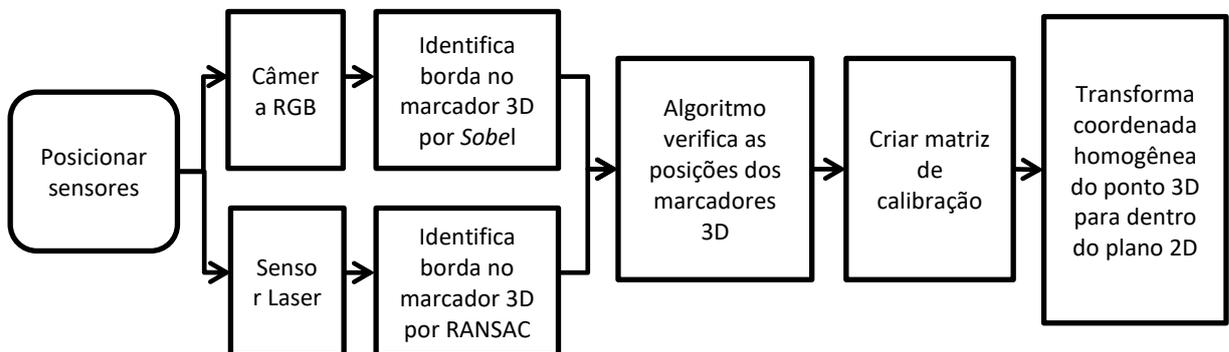


Figura 31: Fluxograma do processo de calibração.

Este método de calibração basicamente consiste em encontrar os parâmetros de posição e orientação do laser relacionado à câmera RGB, ou seja, os vetores  $t$  e  $R$  como ilustrado na Figura 32. Esse problema é chamado de busca por 6 graus de liberdade (6DoF, em inglês *Degree Of Freedom*) de rotação e translação dos três eixos do espaço tridimensional.

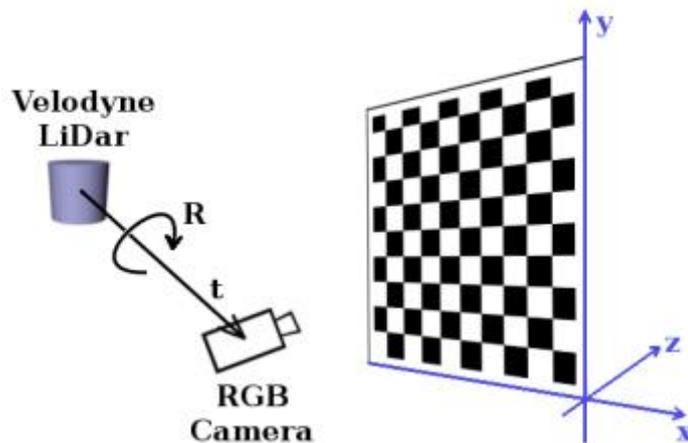


Figura 32: Calibração do laser com câmera RGB através dos vetores R e t. (VELAS et al., 2014).

Inicialmente são estimados os parâmetros de calibração usando um marcador 3D. Esta estimativa permite uma aproximação das posições dos sensores. Em seguida, é usado um marcador que deve ser facilmente identificado pela imagem RGB e nos dados do laser.

O pré-processamento da câmera e do laser incluem uma detecção de bordas. Na Figura 33 pode observar como as bordas são capturadas usando o laser.

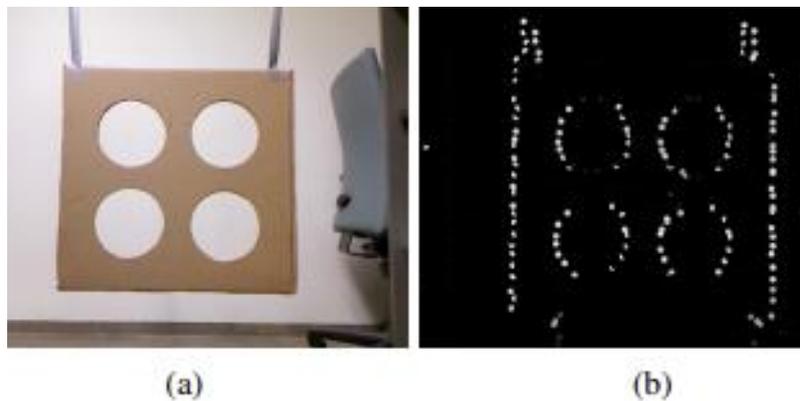


Figura 33: Marcador plano com furos circulares em (a) e arestas detectadas pelo laser (b). Adaptado de (VELAS et al., 2014).

A identificação pelo laser é baseada na descontinuidade das profundidades em relação os pontos vizinhos diretos, conforme Equação (8). Os valores de  $q_i^r$  representam o alcance do  $i$ -ésimo ponto da nuvem de pontos e  $\gamma$  é a constante. O menor  $\gamma$  é o valor mais alto atribuído aos pontos de baixa distorção relativamente aos pontos com alta distorção. Posteriormente, os valores de  $V_i$  são normalizados dentro do range (0;1).

Para eficiência, os pontos com valor atribuído sob um determinado limite são descartados. Isso leva a redução do número de pontos (aproximadamente 90% dos pontos são removidos), preservando a precisão da detecção.

O plano dos círculos é localizado pelo algoritmo RANSAC e os *outliers* são descartados. Para detecção e através da câmera RGB, usa-se o filtro *Sobel*, que produz saídas menos ruidosas (LEVINSON e THRUN, 2014). Para a detecção, assume-se que o tamanho e as posições dos círculos são conhecidos e são considerados como objetos planos. Então o algoritmo verifica as posições dos círculos encontrados pela imagem RGB e pela nuvem de pontos como na Figura 34.

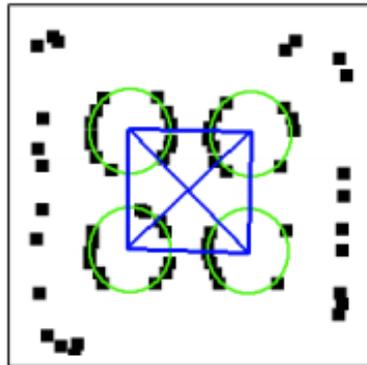


Figura 34: Detecção dos quatro círculos e verificação através das distâncias dos centros. Adaptado de (VELAS e colab., 2014)

Para o processo de calibração dos sensores é necessário transformar as coordenadas homogêneas do ponto 3D para o plano 2D. Para isso é definido o sistema formado pelos vetores ortonormais  $(X, Y, Z)$ . Quando a coordenada  $q_z$  for diferente de zero, podemos dividir  $(q_x, q_y, q_z)$  por  $q_z$ . Dessa forma obtemos o ponto  $(q_{X(2D)}, q_{Y(2D)}, 1)$ . Os valores  $[\frac{q_{X(2D)}}{q_z}, \frac{q_{Y(2D)}}{q_z}]$  são conhecidos por Coordenadas Cartesianas do ponto homogêneo e podem ser encontrados através da Equação (14):

$$\begin{bmatrix} q_{X(2D)} \\ q_{Y(2D)} \\ q_z \end{bmatrix} = P \cdot K \cdot \begin{bmatrix} q_x \\ q_y \\ q_z \\ 1 \end{bmatrix} \quad (14)$$

onde  $P$  é a matriz de projeção da câmera já conhecida e  $K$  é a matriz de calibração que descreve a transformação homogênea entre os sensores. Os parâmetros são apresentados nas seguintes Equações (15) e (16):

$$P = \begin{bmatrix} f & 0 & O_X & 0 \\ 0 & f & O_Y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (15)$$

$$K = \begin{bmatrix} 1 & 0 & 0 & t_X \\ 0 & 1 & 0 & t_Y \\ 0 & 0 & 1 & t_Z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (16)$$

onde  $f$  representa a distância focal da câmera e  $(O_X, O_Y)$  são as coordenadas do ponto principal em pixels, ou seja, os parâmetros da câmera intrínsecos conhecidos.

As únicas variáveis desconhecidas que são necessárias serem calculadas são os componentes do vetor de translação  $[t_X, t_Y, t_Z]^T$ .

O primeiro componente desconhecido de translação, que deve ser encontrado, converte na diferença de profundidade entre os sensores chamada de  $t_Z$ . É possível calcular o  $t_Z$  usando as correspondentes dos raios identificados pela nuvem de pontos ( $r_{3D}$ ) e pela imagem ( $r_{2D}$ ) dos círculos detectados e posteriormente as demais componentes (VELAS et al., 2014):

$$t_Z = \frac{r_{3D} \cdot f}{r_{2D} - q_Z} \quad (17)$$

$$t_X = \frac{(q_{X(2D)} - O_X) \cdot (q_Z + t_Z)}{f} - q_X \quad (18)$$

$$t_Y = \frac{(q_{Y(2D)} - O_Y) \cdot (q_Z + t_Z)}{f} - q_Y \quad (19)$$

Após a primeira etapa da calibração usando um marcador 3D, é realizada calibração fina com uma estimativa mais precisa pela busca de todos os 6 graus de liberdade (translação e rotação) em um subespaço de pequenos parâmetros. Essa procura é realizada usando uma amostragem densa da vizinhança do vetor de calibração aproximado de  $[t_X, t_Y, t_Z]^t$  onde a rotação foi omitida.

Na busca pelos parâmetros de calibração, a capacidade de avaliar cada conjunto de parâmetros é essencial. Isso significa que precisamos de uma função  $f$  Equação (20) que atribui um valor proporcional à qualidade de calibração para cada conjunto de parâmetros 6DoF entre os sensores:

$$f: \mathbb{R}^6 \rightarrow \mathbb{R} \quad (20)$$

A métrica de erro usada em (LEVINSON e THRUN, 2014), se baseia no pressuposto de que as bordas podem ser encontradas de forma robusta na imagem da câmera e combinadas com as descontinuidades do intervalo na nuvem de pontos do LiDAR.

Primeiramente, as bordas são encontradas na nuvem de pontos. Durante esse processo, o valor de  $V_i$  descrito na Equação (8) é atribuído a cada ponto e é considerado o valor de intensidade desse pixel. Então a nuvem do ponto é transformada usando os parâmetros que se pretende avaliar. Finalmente, esses pontos são projetados no plano da imagem usando a matriz de projeção (Equação (14)) e podem ser comparados com a imagem da câmera pré-processada no espaço 2D usando alguns critérios de similaridade tradicionais.

O pré-processamento da imagem da câmera consiste em detecção de borda, em caso 2D usando o filtro *Sobel*, criando a imagem de bordas. Para tornar a função (16) mais suave, aplica-se a transformada de distância inversa (IDT) (LEVINSON e THRUN, 2014) (Equação 20). IDT atribui cada pixel nas coordenadas  $[i, j]$  o valor  $D_{i,j}$  proporcional à distância e a força da borda  $E_{X,Y}$  no vizinho do pixel nas coordenadas  $(X, Y)$ :

$$D_{i,j} = \alpha \cdot E_{i,j} + (1 - \alpha) \cdot \max_{X,Y} \{E_{X,Y} \cdot \beta^{\max(|q_{X(2D)} - i|, |q_{Y(2D)} - j|)}\} \quad (21)$$

O fator  $\alpha$  aumenta a intensidade das bordas dando impacto nos contornos e o fator  $\beta$  amplia as arestas de uma área efetivamente impactada.

E então, os critérios de similaridade de correlação cruzada (Equação (21)) são aplicados à nuvem de pontos projetada  $I_v$  e a  $I_c$  do quadro de câmera processada para encontrar o valor da função.

$$S_E = \sum_X \sum_Y I_c(q_{X(2D)}, q_{Y(2D)}) \cdot I_v(q_{X(2D)}, q_{Y(2D)}) \quad (22)$$





Figura 36: Sensores fixados no veículo de apoio para coleta de dados.

O MATLAB foi usado para as etapas de processamento, segmentação, reconhecimentos de padrões e objetos a partir de imagens e nuvem de pontos coletadas em campo.

### 3.2.3 Identificação da estrada por imagem

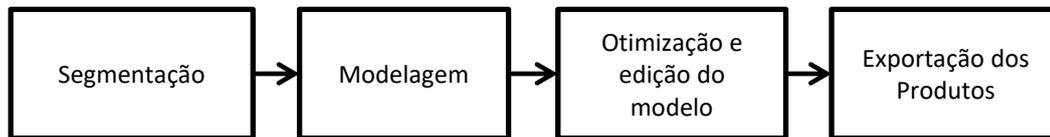
Existem diversos experimentos utilizando filtros para identificação de objetos e planos em uma imagem, porem quando se trata de imagens a céu aberto com diversos veículos, vegetações, diferentes texturas e cores variadas, eliminar as informações indesejadas ou ruídos se torna complexo.

Os dados de imagem coletados são transformados em escala de cinza para então, utilizar a técnica de limiarização e *Sobel* para identificar regiões de variações de intensidade dos pixels. Dessa forma a imagem original (Figura 37(a)) transformada em binária, como na Figura 37(b).



Figura 37: (a) Imagem original e (b) com pré-processamento por classificação de pixel.

Conforme descrito em (GROETELAARS, 2011), a fase de processamento contempla as operações ilustrada na Figura 38.



**Figura 38: Operações da fase de processamento de imagens.**

Para (GROETELAARS, 2011), na etapa de segmentação é realizado o agrupamento de regiões ou objetos semelhantes, para este caso estradas, equipamentos e pessoas. Esse processo ocorre a partir de critérios de agregação, como identificação de limites dos dados por variações de cores, planos, texturas e por meios interativos de seleção de trechos das imagens.

A modelagem é realizada para identificação de formatos ou modelos geométricos através de processos automáticos ou semiautomáticos.

Concluindo a fase de modelagem, é possível realizar a otimização dos dados preliminares. Em (GROETELAARS, 2011), são citados dois exemplos de otimização do modelo:

- Minimização dos ruídos e dados duplicados;
- Modificação da malha triangular – quando é necessário reduzir o tamanho do arquivo de dados preservando os detalhes mínimos requeridos.

Para manter somente a área de interesse, nesse caso a rua, é possível remover as regiões com menor quantidade de pixels, menor que área mínima, usando a função *bwareopen* do MATLAB. É uma operação morfológica para abertura de área de interesse cujo resultado pode ser observado na Figura 39.



Figura 39: (a) Imagem com pré-processamento e (b) com eliminação dos maiores ruídos no Matlab.

Na Figura 40 foi usada uma operação para preenchimento dos vazios e suavização das curvas, tomando os dados uniformes (GROETELAARS, 2011).

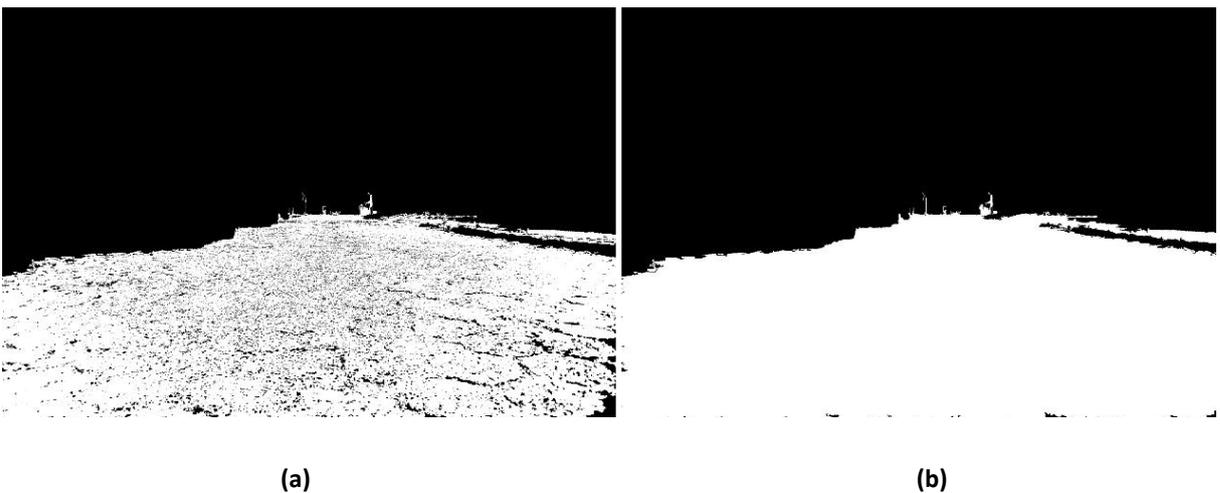


Figura 40: (a) Imagem com eliminação dos maiores ruídos e (b) com preenchimento de espaços no Matlab.

Com a imagem processada e definida a região de interesse, pode-se fundir com a imagem original da câmera usando a função *imfuse* do MATLAB para ilustrar a região da estrada na imagem original como mostra na Figura 41.

Existem diversas estratégias complementares para identificação de estradas, que foram desenvolvidas com base em suposições sobre as características da estrada. Por exemplo, os métodos baseados em bordas identificam faixas ou limites das estradas. Entretanto os algoritmos funcionam melhor em estradas bem projetadas,

pavimentadas e pintadas, resultando em uma riqueza de contornos e de contrastes adequados.



(a)

(b)

Figura 41: (a) Imagem original (b) detecção da estrada no Matlab.

Outra técnica que pode ser implementada é o filtro de Gabor descrita por (RASMUSSEN, 2016), que identifica a estrada por textura. Mas para o sucesso dessa técnica, é necessário que não tenha obstáculos ao longo da estrada. Para o estudo proposto, diversos cenários possuem obstáculos, dificultando a implementação da técnica do filtro de Gabor.

As estradas de uma mina não possuem bordas bem definidas, diferenciação das cores RGB com os limites das estradas, dificultando os métodos tradicionais de acompanhamento de estrada. Portanto, para esse estudo a identificação da estrada será realizada somente através do sensor laser apresentado mais à frente.

### 3.2.4 Identificação de obstáculos por imagem RGB

Conforme proposto em (CARNEIRO e SILVA, 2017), a identificação de objetos nas imagens RGB pode ser feita através das redes neurais convolucionais (CNNs). A identificação através das CNNs se torna mais eficiente quando existem quantidades significativas de amostras rotuladas representando diversas classes. Algumas das vantagens da CNN descritas em (CARNEIRO e SILVA, 2017) são:

- Capacidade de identificação e extração de padrões importantes devido ao aprendizado de transformações (*kernels*);

- Menor dependência de parâmetros de ajustes em relação às redes neurais totalmente conectadas com a mesma quantidade de camadas ocultas.

Atualmente, existem diversas bibliotecas que providenciam implementações das principais operações utilizadas pelas CNNs, tais como: *Caffe* (JIA et al., 2014), *MatConvNet* (VEDALDI e LENC, 2015), *Theano* (BOCK et al., 2017), *Torch* (COLLOBERT et al., 2011), *TensorFlow* (ABADI et al., 2016). Porém, se familiarizar com essas bibliotecas e compreender como as CNNs efetivamente trabalham com as imagens não são tarefas simples.

Para este trabalho foi usado a *MatConvNet*, que basicamente é uma caixa de ferramentas MATLAB que implementa redes neurais convolucionais (CNNs) para aplicações de visão computacional (VEDALDI e LENC, 2014).

No início, a *MatConvnet* foi usada para identificação de caracteres, como dígitos numéricos e códigos postais. Diversas melhorias foram propostas nos últimos anos, porém as versões melhoradas da CNN compartilham dos mesmos conceitos fundamentais (REN et al., 2017).

#### **3.2.4.1 Configuração**

A rede neural *MatConvnet* foi executada em seis diferentes configurações propostas em (SIMONYAN e ZISSERMAN, 2014). A Tabela 1 mostra as configurações dos seis modelos testados. Configuração A e A-LRN com onze camadas, B com treze camadas, C e D com dezesseis camadas e E com dezenove camadas. A Tabela 2 possui o número de parâmetros para cada configuração.

Num estudo realizado em (SIMONYAN e ZISSERMAN, 2014) foi identificado que o aumento de camadas é benéfico, pois reduz o erro da resposta. Porém, conforme a Tabela 3, a configuração E com 19 camadas apresentou erro próximo, da configuração D com 16 camadas.

Portanto, para o trabalho proposto foi usada a rede neural *Convnet* com a configuração D, que possui dezesseis camadas e mostrou maior confiabilidade nos resultados.

Tabela 1: Configuração da ConvNet mostrando em colunas de A à E, aumentando o número de camadas, adaptado de (SIMONYAN e ZISSERMAN, 2014).

Configuração ConvNet					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 x 224 RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
softmax					

Tabela 2: Número de parâmetros para cada configuração (em milhões), adaptado de (SIMONYAN e ZISSERMAN, 2014).

Network	A, A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

Tabela 3: Performance de cada configuração da ConvNet, adaptado de (SIMONYAN e ZISSERMAN, 2014).

ConvNet config (Tabela 1)	top - 1 Valor erro(%)	top - 5 Valor erro(%)
A	29,6	10,4
A-LRN	29,7	10,5
B	28,2	9,6
C	27,7	9,2
<b>D</b>	<b>26,6</b>	<b>8,6</b>
E	26,9	8,6

Conforme (SIMONYAN e ZISSERMAN, 2014), essa rede neural foi treinada com mais de quatorze milhões de imagens e aproximadamente mil classes diferentes. A configuração D da *Convnet* possui treze camadas convolucionais, cinco camadas de *pooling*, três camadas totalmente conectadas, por último uma camada *softmax* e são no total 53.682.197 neurônios. A Figura 42 ilustra a arquitetura completa da rede neural.

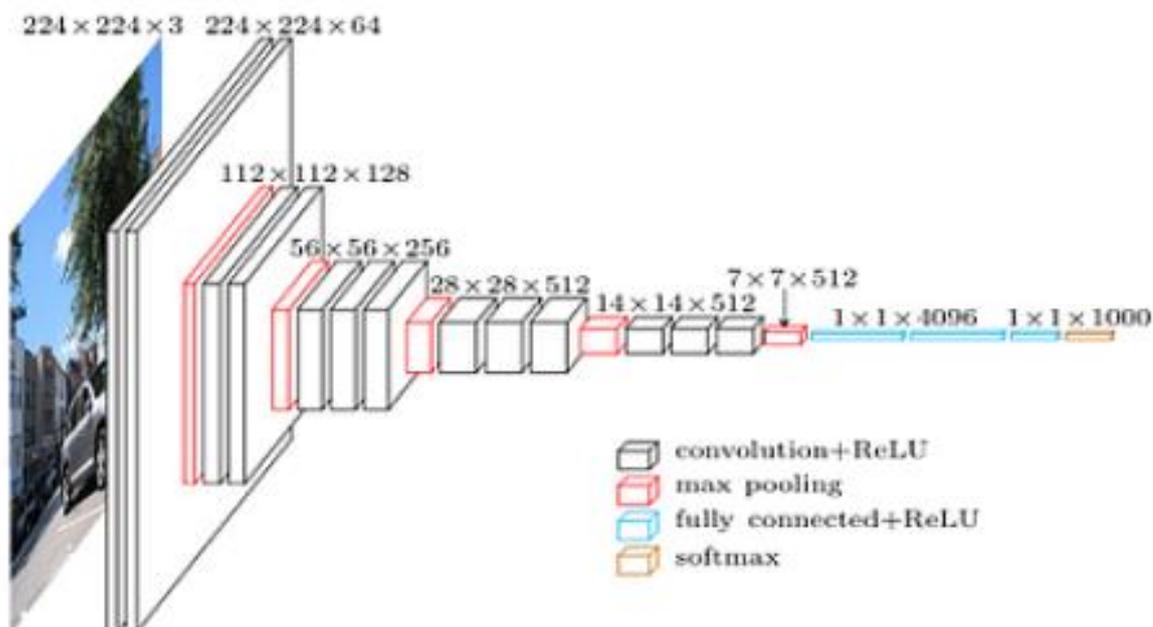


Figura 42: Arquitetura geral da rede neural de 16 camadas usada. Adaptado de (NOH et al., 2015).

Dentre as classes treinadas nessa rede neural, estão as classes dos veículos, que são os principais obstáculos dos cenários coletados em campo para esse estudo. Na Figura 43 observa-se a identificação de dois carros e um animal pela rede neural

treinada *Convnet* em um cenário urbano. Nos resultados deste trabalho serão mostrados os demais cenários com obstáculos.



Figura 43: Identificação de objetos por rede neural.

### 3.2.5 Identificação de estrada pela nuvem de pontos

Para a identificação de estrada através da nuvem de pontos, usou-se o método *Random Sample Consensus* – RANSAC, descrito anteriormente, para extração de planos dentro da nuvem de pontos.

Com a Equação (6), é possível determinar se um dado ponto  $\vec{q}_i$  pertence ou não a uma região com um limiar de distância  $t$  do plano  $h$ . Então, a quantidade de pontos da nuvem  $NP = \{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n\}$  que pertencem ao limiar distância  $t$  do plano  $h$  pode ser computada usando as Equações (6) e (7).

Quando as nuvens de pontos são densas, onde os espaços vagos entre os pontos gerados pelos feixes de laser são pequenos, o valor de  $t$  pode ser pequeno. Dessa forma, os planos terão espessura menor e mais fidedigna com a realidade. Para a nuvem de pontos gerada em ambientes externos, onde pode haver regiões de interesse sem pontos, o valor de  $t$  deve ser maior. A Figura 44 ilustra a variação da distância dos feixes de laser conforme varia o ângulo de  $\beta$ . Para se determinar o plano da pista de rolagem, o valor de  $t$  deve ser maior que a distância  $D_n$ . O último feixe de laser com distância segura para identificação do condutor do veículo. Em

contrapartida, o plano se torna mais espesso ocultando algumas irregularidades da pista, que para ambientes de mineração é aceitável.

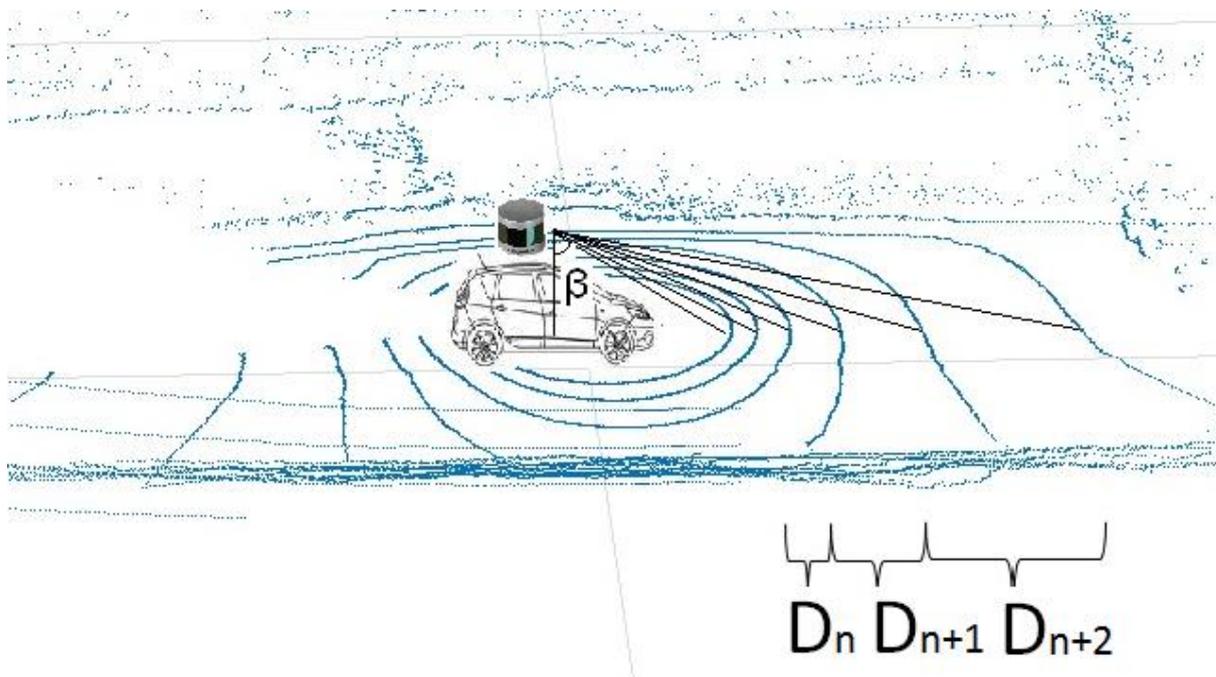


Figura 44: Distância entre feixes de laser para determinar valor de  $t$ .

Um dos parâmetros para configuração inicial do algoritmo RANSAC é a variável  $g$  que pode obter o valor de zero a um. Quando o objetivo é encontrar o maior plano da cena, o valor de  $g$  é zero. Quando o objetivo é extrair os principais planos da cena, o valor de  $g$  se aproxima de um.

A Figura 45(a) ilustra um cenário em ambiente de mineração onde apresentam dois principais planos, a estrada e o talude. Esse mesmo cenário é apresentado na Figura 45(b) em forma de nuvem de pontos e a identificação desses dos dois planos são mostrados na Figura 45(c), em branco o plano da estrada e em vermelho o plano do talude.

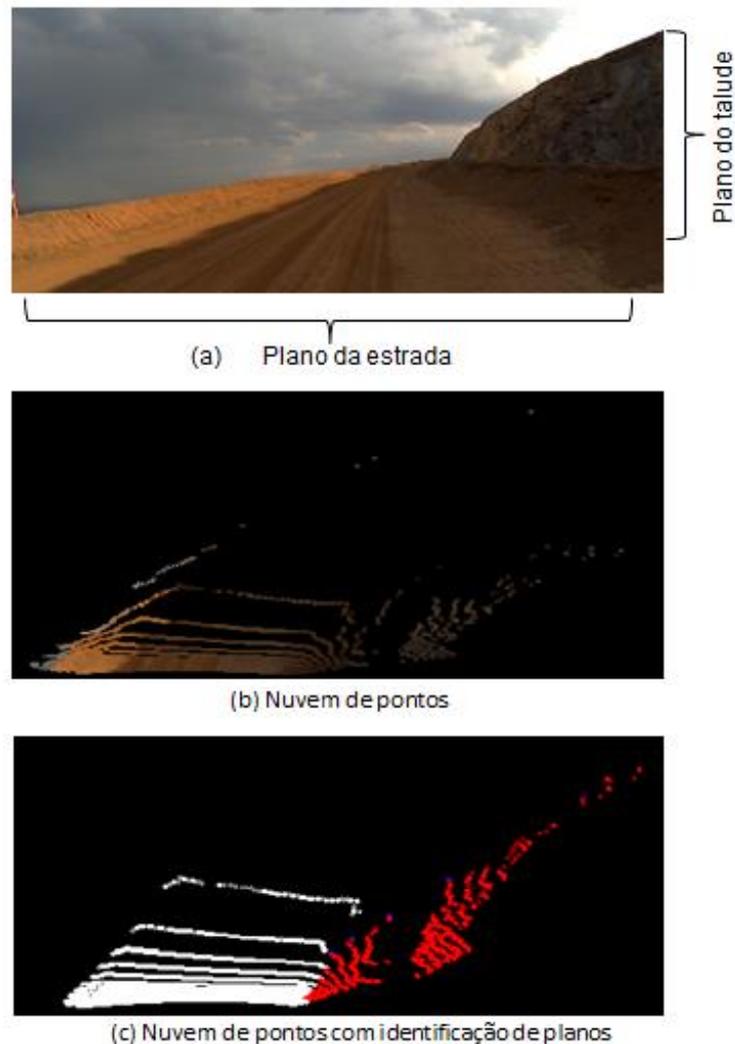


Figura 45: (a) Cenário de coleta da nuvem de pontos, (b) nuvem de pontos original, (c) nuvem de pontos com identificação de planos.

### 3.2.6 Identificação de objetos por nuvem de pontos

Para a identificação de objetos (aglomerados de pontos), o método de agrupamento de pontos DBSCAN descrito em (ESTER, 1996) se mostra eficaz.

Esse método se diferencia por não necessitar da quantidade de agrupamentos, do contrário dos métodos comuns de agrupamentos. O DBSCAN é capaz de identificar pontos isolados que podem ser ruídos do aparelho ou ignorá-los se necessário.

O DBSCAN efetua uma varredura generalizada da nuvem de pontos, identificando distância entre os pontos, permitindo gerar aglomerados de qualquer formato. Esse



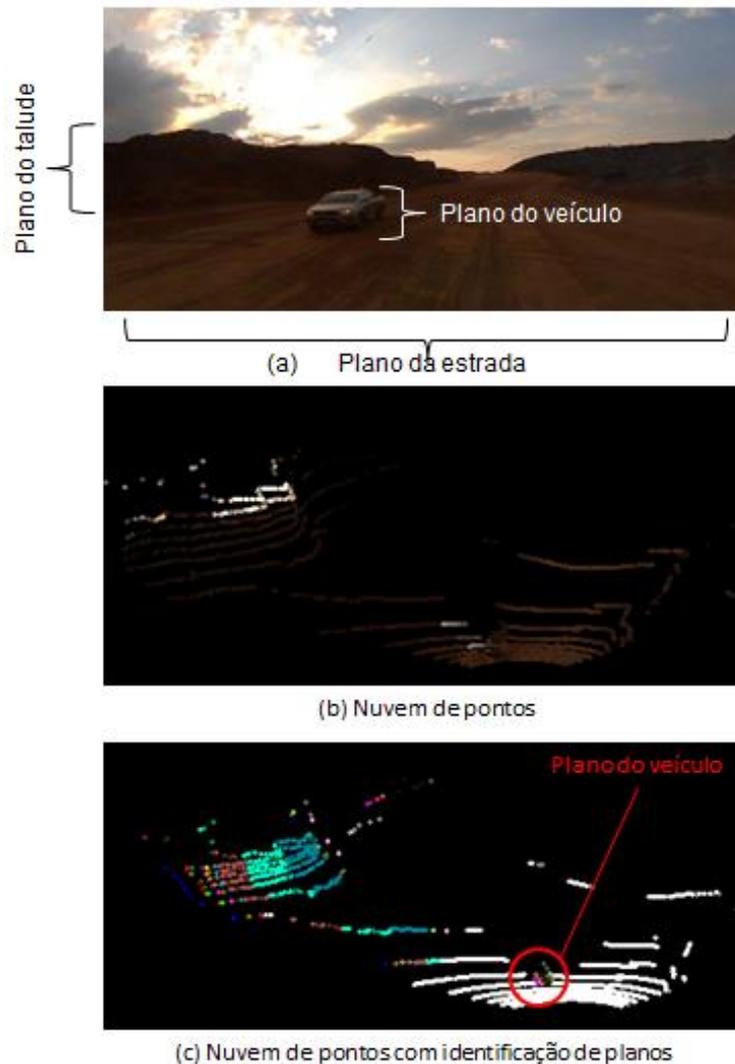


Figura 47: (a) Cenário de coleta da nuvem de pontos, (b) nuvem de pontos original, (c) nuvem de pontos com identificação de plano e aglomerado de pontos do obstáculo.

### 3.2.7 Identificação por fusão dos sensores

A identificação da estrada e obstáculos por fusão dos sensores também é conhecida por sistemas multiparâmetros (REGIS e colab., 2013). Nas tecnologias que abordam o processamento de dados de modo embarcado, ou seja, com instrumentos no próprio equipamento, destaca-se a fusão de sensores, que é o processo de combinação dos dados providos de dois ou mais sensores com o objetivo de reduzir as falhas nos processos decisórios.

Só é possível realizar essa etapa do estudo após a calibração dos sensores. O principal objetivo é identificar a variação de cores RGB entre os pontos da nuvem tridimensional onde o algoritmo de RANSAC identifica uma mudança no plano da estrada, e então, confirmar tanto por nuvem de pontos quanto por alteração de cores RGB que existem um obstáculo na estrada.

O sistema de coordenadas  $(X, Y, Z)$  é usado como referência, onde  $Y$  representa a largura da via,  $X$  representa o comprimento da via e  $Z$  a altura, conforme Figura 48.

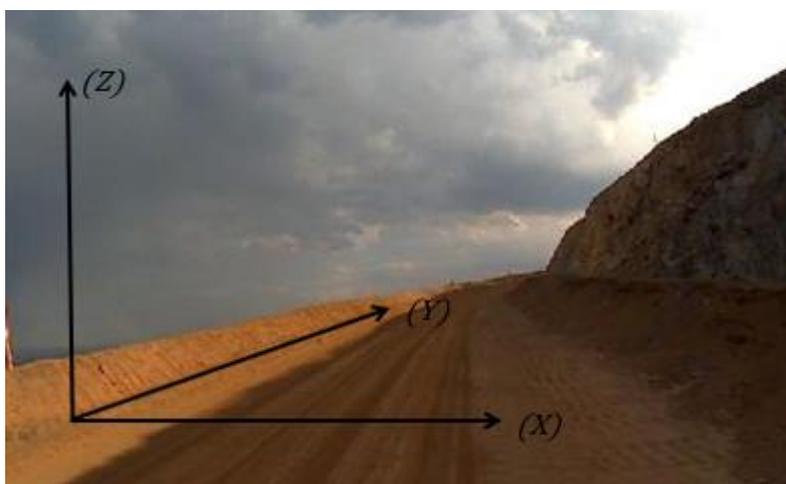


Figura 48: Plano tridimensional considerado para referência.

O algoritmo proposto apresenta melhor resultado para os dados coletados durante este estudo, pois os cenários no ambiente de mineração possuem características diferentes de outras pesquisas semelhantes. O método de identificação do algoritmo segue os seguintes passos para análise:

1. É identificado o padrão de cores RGB da estrada onde o algoritmo de RANSAC identificou como estrada;
2. Se os três parâmetros das cores RGB tiveram variações com o padrão de cores da estrada ou aquele ponto no mapa tridimensional tiver variação acima de um  $\Delta Y$  em relação ao plano da estrada, entende-se que naquele ponto há grandes chances de existir um obstáculo e o ponto é identificado na cor azul;

3. Se pelo menos dois parâmetros das cores RGB tiveram variações com o padrão de cores da estrada, entende-se que naquele ponto há grandes chances de existir um obstáculo e o ponto é identificado na cor verde;
4. Se pelo menos um parâmetro das cores RGB tiver variação com o padrão de cores da estrada, entende-se que naquele ponto há menores chances de existir um obstáculo e o ponto é identificado na cor amarela;
5. Se não houver variação nas cores RGB e variação da coordenada no eixo Z em relação ao plano da estrada, entende-se que não há obstáculo e o ponto permanece identificado como branco.

O algoritmo executa o passo dois ao cinco para todos os pontos da nuvem de pontos. Na Figura 49 é apresentado o modelo do algoritmo para essa etapa.

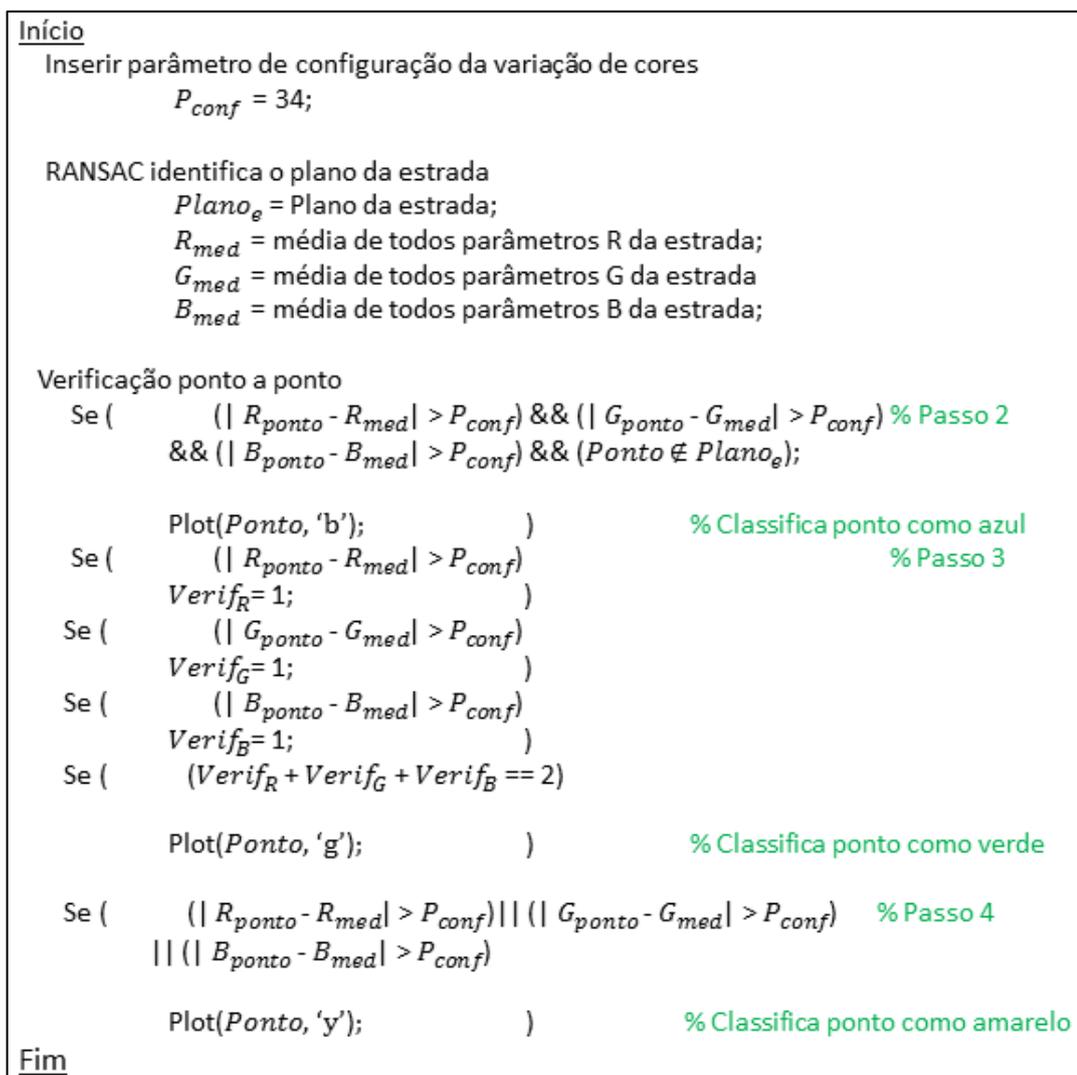
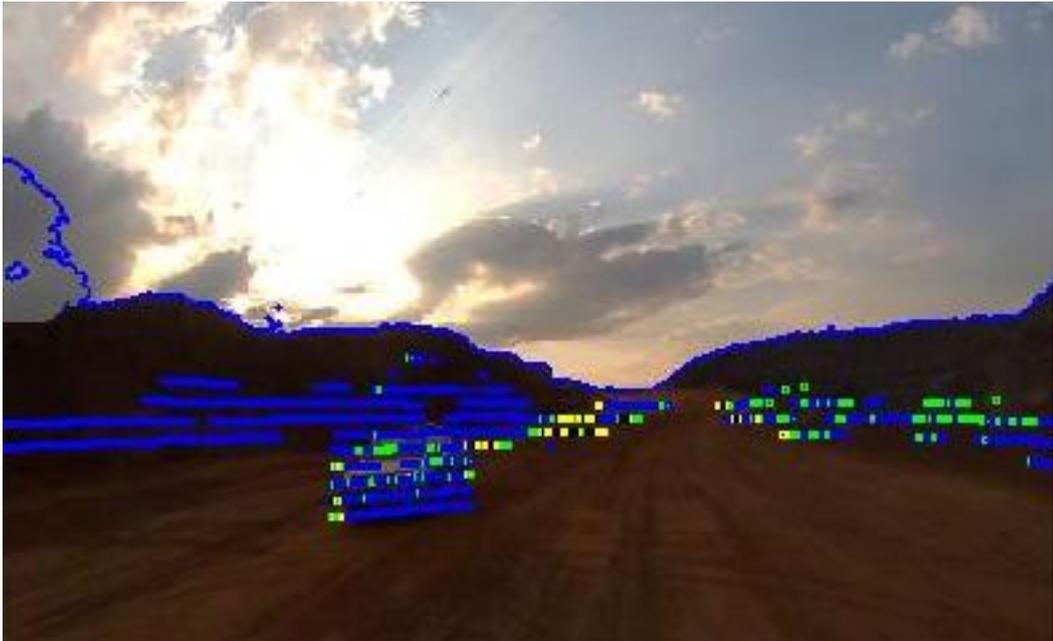


Figura 49: Modelo simplificado do algoritmo.

Após a verificação de todos os pontos da nuvem é formada uma nuvem de pontos colorida, classificando os pontos conforme identificação de ambos os sensores como ilustrado na Figura 50. A inclusão dos pontos brancos pertencentes à estrada será apresentada no próximo capítulo, junto com a unificação dos resultados de cada técnica.



**Figura 50: Identificação por fusão de sensores.**

## 4 RESULTADOS

Neste trabalho foram realizadas duas coletas de dados em campo, a primeira coleta foi em ambiente urbano, com vias asfaltadas, diferenciação de cores nítidas das vias e dos obstáculos, porém não havia sido realizado a calibração dos sensores. Na segunda coleta realizada em ambiente de mineração os limites das vias não são nítidos iguais à coleta anterior, mas a calibração já havia sido realizada em laboratório.

Para lidar com sensores de diferentes princípios de operação e também com posições diferentes foi necessário realizar a calibração apresentada na Seção 4.1.

Na Seção 4.2 é apresentada a identificação da estrada por imagem RGB nos dados das duas coletas através do filtro *Sobel* e classificação de pixel-a-pixel pertencente à estrada ou não. Na Seção 4.3 são descritos os resultados obtidos com a identificação de obstáculos por imagem através da rede neural *Convnet*.

Na Seção 4.4 é apresentada a identificação de estradas através da nuvem de pontos. Nos dois cenários de coleta foi usado o método de RANSAC para identificação de planos. Os obstáculos à frente do veículo foram localizados pelo método DBSCAN descrito na Seção 4.5, que identifica um aglomerado de pontos e o interpreta como objeto

A identificação de obstáculos por fusão dos sensores é apresentada na Seção 4.6. Essa etapa utiliza a variação de cores na imagem RGB e a variação de altura do ponto dentro da nuvem de pontos. Essa etapa só foi possível realizar com os dados da coleta em ambiente de mina devido a calibração dos sensores realizada anteriormente. Os parâmetros de configuração usados em cada etapa e seus resultados são descritos a seguir.

### 4.1 Calibração dos sensores

A calibração dos sensores foi realizada no laboratório do ITV seguindo a técnica proposta em (VELAS et al., 2014), descrita no Capítulo 3. Os componentes da matriz de calibração relativos a translação encontrados são:

$$t_z = -0,0946661 \quad t_x = 0,0722266 \quad t_y = -0,0609373$$

E os componentes relativos a rotação encontrados são:

$$r_z = 6,17115 \cdot 10^{-6} \quad r_x = -5,21805 \cdot 10^{-4} \quad r_y = 1,82257 \cdot 10^{-3}$$

A matriz de transformação homogênea  $G$  (Equação 16) de ordem 4, formada por quadro submatrizes, descreve a matriz de calibração  $MC$ .

$$G = \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0 & 1 \end{bmatrix} \quad (16)$$

onde a submatriz  $R_{3 \times 3}$  representa a matriz de rotação, pertencente ao grupo ortogonal especial de dimensão 3 ( $SO(3)$ ) (CORRADETTI, 2016), e  $T_{3 \times 1}$  o vetor de translação. Dessa forma, a matriz de calibração dos sensores assume a seguinte forma:

$$MC = \begin{bmatrix} r_z & r_x & 0 & t_x \\ r_y & r_z & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (17)$$

Os pontos da nuvem de pontos obtidos pelo laser são projetados na imagem RGB. A projeção é realizada de acordo com a Equação (14) usando projeção de matriz  $P$  (parâmetros da câmera intrínseca) e a matriz de calibração  $MC$  ao invés da  $K$ .

A Figura 51 ilustra a nuvem de pontos inserida na imagem. Para o trabalho proposto essa técnica de calibração se mostrou eficiente. Porém, para detalhamentos milimétricos, a calibração fina pode ser mais explorada.

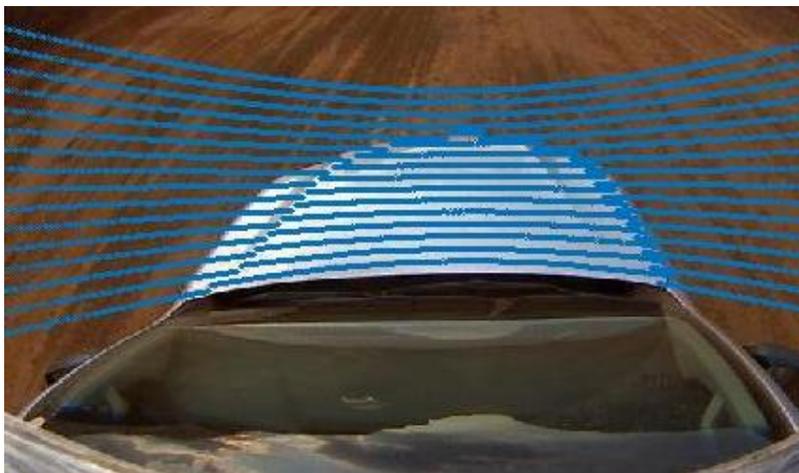


Figura 51: Resultado da calibração dos sensores de imagem e laser.

## 4.2 Identificação da estrada por imagem RGB

Para familiarização das metodologias de processamento de imagens, foi usado filtro *Sobel* com a configuração *Default* do Matlab para identificação da estrada. Suas matrizes para efetuar a convolução já são predefinidas.

Após a identificação de bordas, é realizada a limiarização como pertencente à estrada ou não. Essa técnica mostrou resultados melhores em ambientes urbanos, porém com falhas em regiões de sombras, como mostrado na Figura 52.

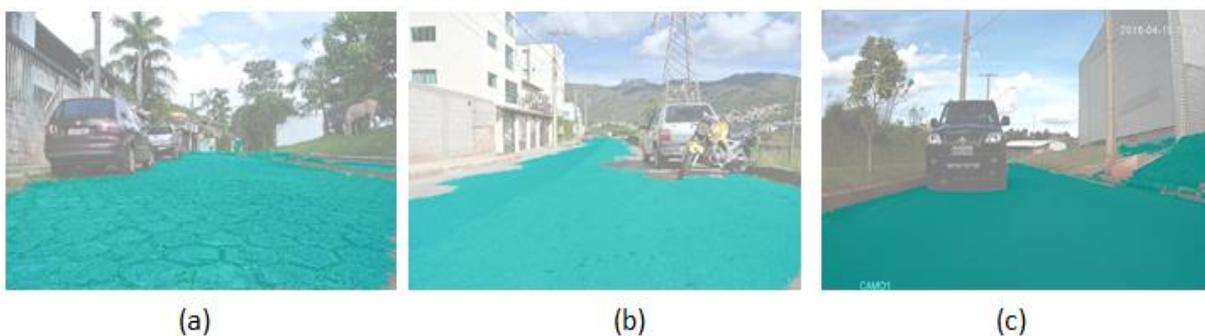


Figura 52: Identificação da estrada por imagem por detecção de bordas em ambiente urbano.

Para ambiente de mineração a identificação de bordas se torna mais complexa, pois os limites da via são do mesmo material e cor da via. A classificação por cores encontrou a mesma dificuldade. Na Figura 53 pode-se observar a mesma técnica, porém em ambiente de mina. As leiras que limitam a pista de rolagem foram identificadas como estrada na Figura 53(b) e (c), na Figura 53(a) parte da pista com maior claridade não foi identificada como estrada.

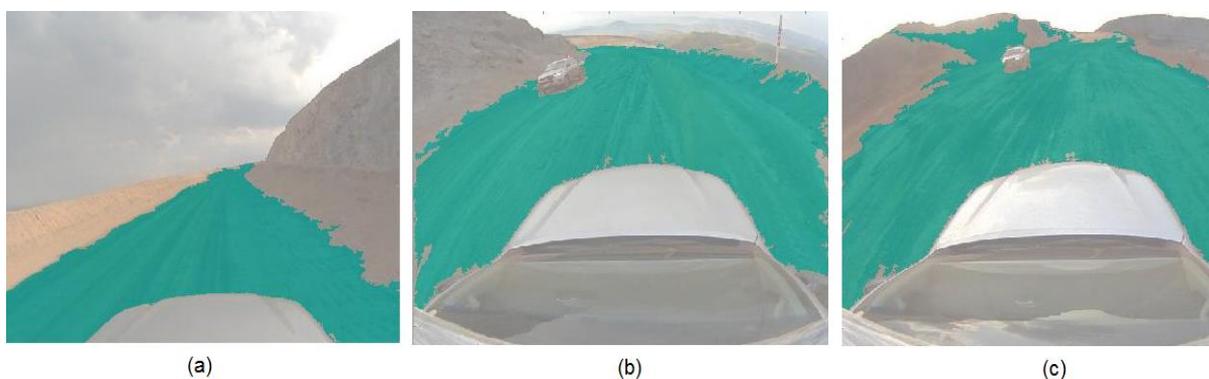


Figura 53: Identificação da estrada por imagem por detecção de bordas em ambiente de mina.

### 4.3 Identificação de obstáculos por imagem RGB

A identificação dos obstáculos na imagem RGB é realizada através da rede neural *Convnet*, que se caracteriza como um método supervisionado. A configuração com 16 camadas, descrita anteriormente, foi usada nessa etapa.

As dimensões do volume resultante da camada convolucional é controlada por três parâmetros: profundidade (*depth*) com o valor 3 devido as três cores RGB, passo (*stride*) com o valor de 1 e *zero-padding* também com o valor de 1 para preservar as características principais da imagem.

Antes de usar a CNN com as imagens dos dados coletados em campo, foram utilizadas imagens da internet que continham veículos, pedestres e animais. Essa etapa tem como finalidade identificar os obstáculos tanto em imagens obtidas em diferentes condições de visibilidade quanto durante o dia em boas condições de luminosidade, em condições de neblina e durante o período noturno, onde a visibilidade é menor. Foram usadas sessenta imagens, cada uma delas com um ou mais obstáculos a serem identificados. Algumas imagens são ilustradas na Figura 54.

O percentual de identificação de objetos durante o dia é de 91% de acerto, durante a noite de 84% e com neblina de 76%, ou seja, quanto menor a iluminação e nitidez da imagem maior o erro durante a identificação. Portanto é de grande importância utilizar diversos sensores de diferentes princípios de funcionamento para aumentar a confiabilidade dos resultados. A Tabela 4 mostra o percentual de identificação dos objetos.

Tabela 4: Percentual de identificação de objetos pela CNN com imagens em diferentes níveis de visibilidade.

Condições de visibilidade	Quantidade de objetos nas imagens	Quantidade de objetos identificados	Porcentagem de objetos identificados
Dia	108	98	91%
Noite	58	49	84%
Neblina	51	39	76%



Figura 54: identificação de objetos por imagem: (a) durante o dia, (b) durante a noite e (c) durante neblina (<https://br.123rf.com>).

As identificações realizadas nas imagens coletadas nesse estudo são ilustradas na Figura 55 e 56. Aplicada no cenário de mineração, o percentual de identificação de objetos através da CNN é de 90% de acerto como apresentado na Tabela 5.

Tabela 5: Percentual de identificação de objetos pela CNN com imagens realizadas durante as coletas.

Condições de visibilidade	Quantidade de objetos na imagem	Quantidade de objetos identificados	Porcentagem de objetos identificados
Dia	20	18	90%

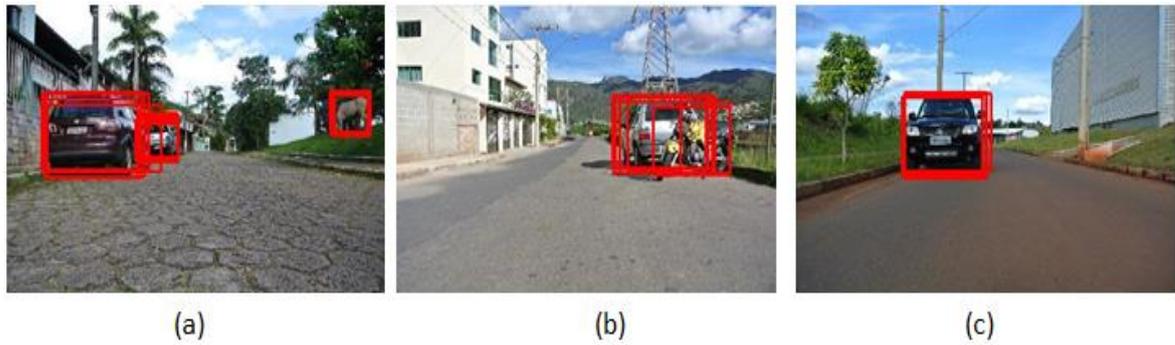


Figura 55: Identificação de obstáculos em ambiente urbano.

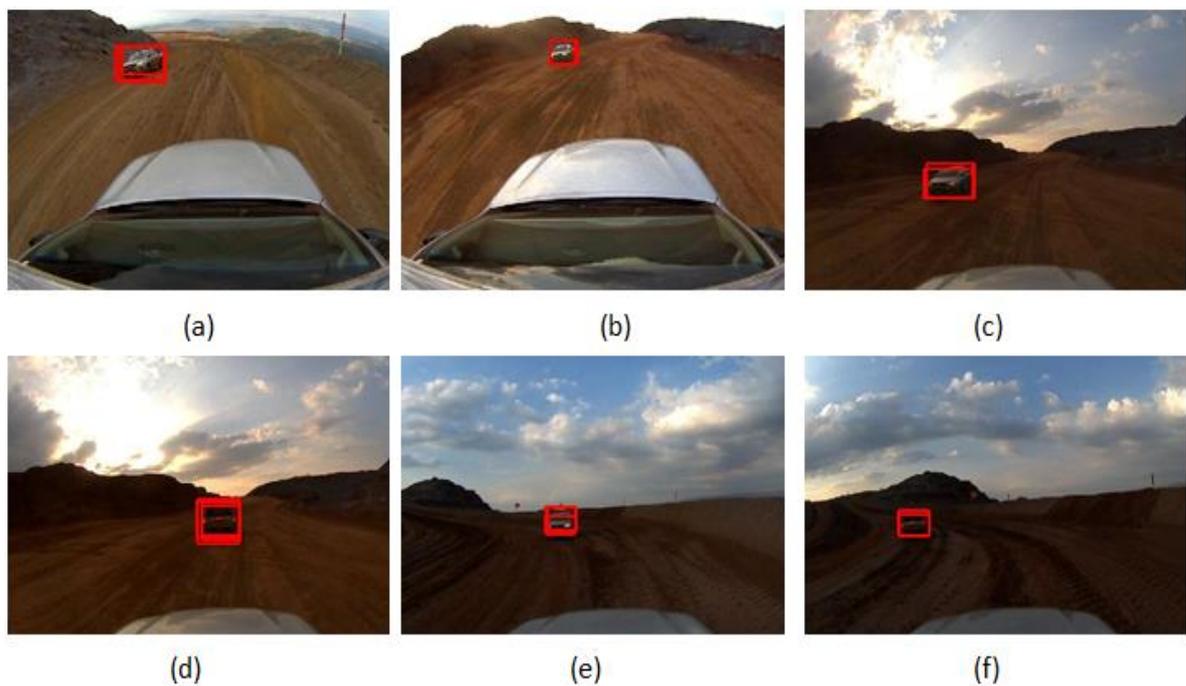


Figura 56: Identificação de obstáculos em ambiente de mina.

Para objetos com distâncias elevadas ou com baixa luminosidade a identificação de objetos com a rede neural *Convnet* apresentou erros. A quantidade de pixels pertencentes ao objeto não foi satisfatória para que o objeto fosse identificado e sinalizado.

A Figura 57 ilustra um veículo a uma distância de aproximadamente 71 metros não identificado. A Figura 56 apresenta um veículo identificado a aproximadamente 53 metros.



Figura 57: Erro na identificação de objetos com distâncias elevadas.

#### 4.4 Identificação da estrada por nuvem de pontos

Para identificação da estrada por nuvem de pontos, o método de RANSAC, descrito anteriormente, isola o plano com os pontos pertencentes a estrada e elimina os demais planos.

Os parâmetros de configuração para o método de RANSAC dependem de duas variáveis,  $t$  que é o limiar de distância dos pontos do mesmo plano, e  $g$  que define qual a proporção de pontos que deve ser extraído. Para o trabalho proposto foi usado um valor de 17 centímetros para  $r$  devido ao espaçamento da nuvem de pontos mostrado na Figura 44. O valor de 0,4 para  $g$  para que sejam encontrados somente os planos principais da nuvem de pontos, descartando a necessidade de incluir todos os pontos em algum plano. Com esses parâmetros pode-se identificar o plano da estrada mesmo com irregularidades normalmente encontradas em estradas de mina, e também identificar os principais planos, que serão estradas e taludes.

Os parâmetros de configuração do RANSAC podem ser alterados dependendo do modelo do sensor laser, da linearidade da estrada e das condições de operação.

Nas nuvens de pontos das coletas realizadas foram identificados 15 planos de estradas. Utilizando o método de RANSAC todas as estradas foram identificadas (Tabela 6). Os planos das estradas possuem grandes dimensões e grande quantidade de pontos, tornando o método eficaz.

Tabela 6: Percentual de identificação de estradas através do RANSAC.

Condições de visibilidade	Quantidade de estradas na nuvem de pontos	Quantidade de estradas identificadas	Porcentagem de estradas identificadas
Dia	15	15	100%

Na Figura 58 observa-se em diferentes cenários a identificação do plano da estrada e nos casos da Figura 58 (b), (d), (e) e (f) a eliminação dos pontos pertencentes ao veículo, por se tratar de outro plano.

As regiões da estrada logo à frente do veículo não foram identificadas devido a falta de pontos medidos nessa região.

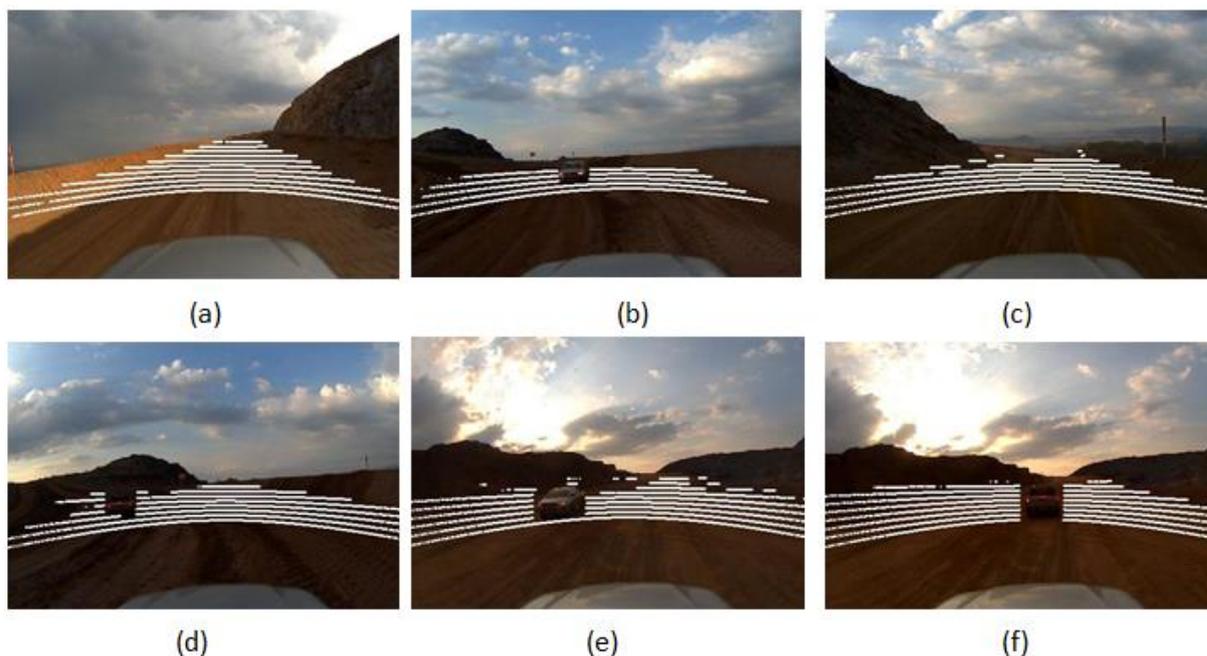


Figura 58: Identificação da estrada por nuvem de pontos.

#### 4.5 Identificação de obstáculos por nuvem de pontos

A identificação de obstáculos por nuvem de pontos utilizou o algoritmo DBSCAN para identificar os aglomerados de pontos e o classificar como um objeto.

Os parâmetros base deste método é o número mínimo de pontos necessários para criar um aglomerado, chamado de *MinPts*, e a distância radial que delimita a vizinhança de um ponto, denominado de *Eps*. Para este trabalho os parâmetros

usados foram de 2 pontos para *MinPts* e uma distância radial de 24 centímetros para *Eps*. O número mínimo de pontos foi determinado de acordo com a confiabilidade do laser. Caso sejam encontrados mais pontos que o número mínimo adotado, conclui-se a existência de um obstáculo neste local. A distância radial de 24 centímetros foi usada devido à variação no eixo *Z* ( $\Delta Z$ ) entre os feixes de laser do obstáculo mais distante dos cenários coletados. Dessa forma é possível identificar o conglomerado de pontos do obstáculo.

Nas Figuras 59 (a), (c) e (e) é possível visualizar a nuvem de pontos com os pontos pertencentes à estrada na cor branca, como visto no tópico anterior, e o aglomerado de pontos pertencentes ao veículo de outras cores. As Figuras 59 (b), (d) e (f) ilustram os mesmos cenários da nuvem de pontos para facilitar a visualização, e também os pontos isolados e caracterizados de vermelho pertencentes ao aglomerado de pontos do obstáculo.

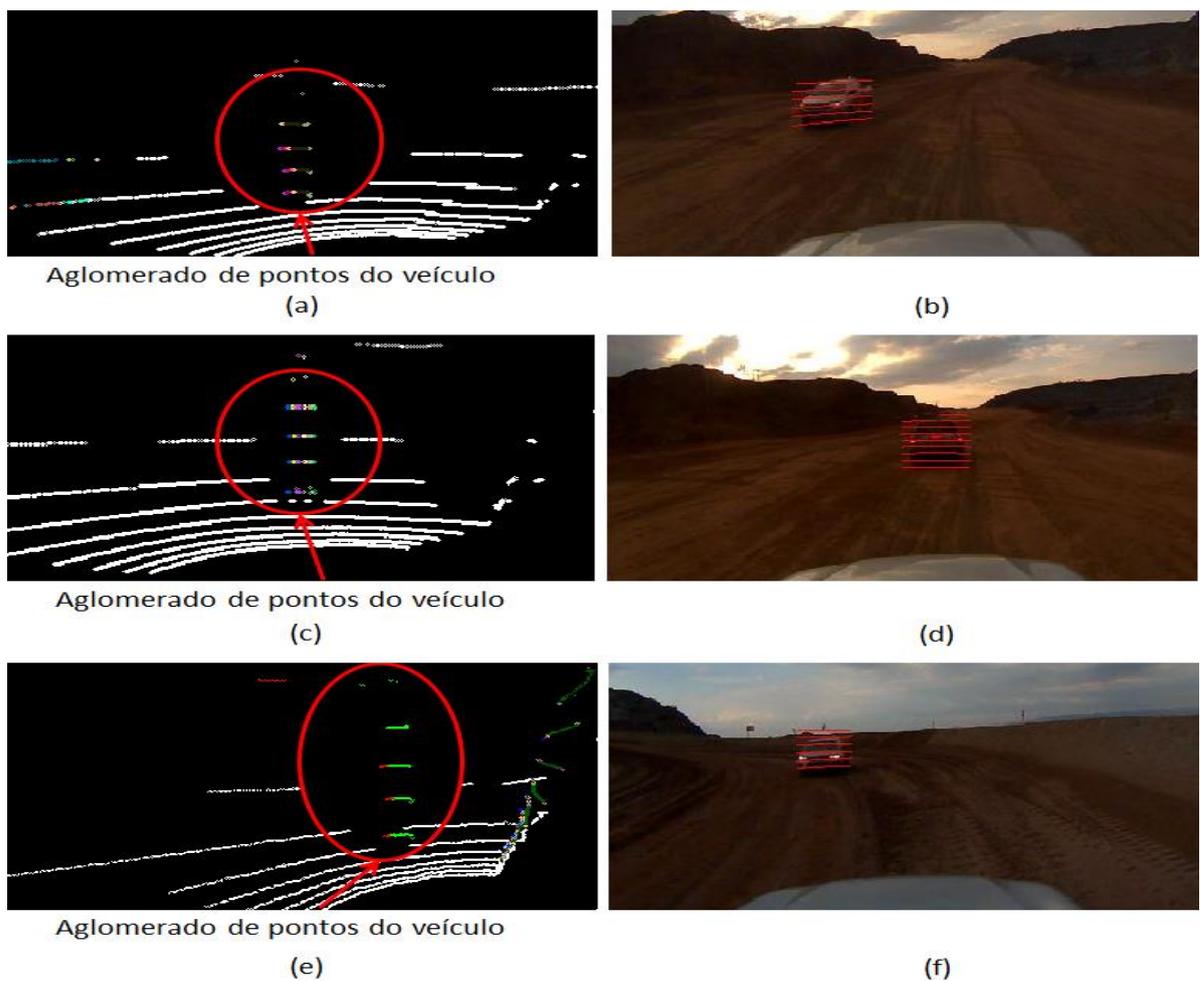


Figura 59: Identificação de obstáculos por nuvem de pontos.

O método de DBSCAN identificou 91% dos objetos presentes nas nuvens de pontos das duas coletas de dados, conforme a Tabela 7.

Tabela 7: Percentual de identificação de objetos nas nuvens de pontos pelo método de DBSCAN.

Condições de visibilidade	Quantidade de objetos nas nuvens de pontos	Quantidade de objetos identificados	Porcentagem de objetos identificados
Dia	11	10	91%

#### 4.6 Identificação por fusão de sensores

A identificação de obstáculos por fusão dos sensores, onde as variações de cores RGB e variação da altura do ponto dentro da nuvem de pontos em relação a estrada, sinalizam se existe ou não um obstáculo naquele local. Esse método se caracteriza como um sistema não supervisionado e mostrou bons resultados na identificação do veículo e dos limites da estrada das cenas testadas.

Foram considerados obstáculos os veículos e as leiras que limitam as estradas. O método proposto para fusão de sensores identificou 19 dos 20 obstáculos (Tabela 8).

Tabela 8: Percentual de obstáculos identificados através de fusão dos sensores.

Condições de visibilidade	Quantidade de obstáculos nos dados	Quantidade de obstáculos identificados	Porcentagem de obstáculos identificados
Dia	20	19	95%

Nos cenários apresentados na Figura 60 (a) e (b), o algoritmo identifica em azul os limites da estrada, onde predominou pontos com grande variação de cores e altura dos pontos. Nos cenários ilustrados na Figura 60 (c), (d), (e) e (f) foi identificado o veículo como obstáculo e também os limites da estrada. Os pontos pertencentes ao veículo, predomina-se a cor azul devido a variação de cor RGB em relação a pista de rolagem. Os limites da pista de rolagem possuem pontos verdes devido a variação de cores RGB ser menor, pois a leira que limita a estrada é formada pelo mesmo material da pista de rolagem.

A Tabela 5 mostra que 90% dos obstáculos foi identificado usando processamento somente dos dados de imagem RGB. Como observado na Tabela 7 a identificação dos obstáculos através da nuvem de pontos atingiu os mesmos 91% de acerto. Porém, utilizando o método de fusão dos dados dos sensores proposto nesse estudo, a identificação dos obstáculos foi de 95% de acerto. Esse aumento do percentual de identificação dos obstáculos gera maior confiabilidade para os resultados.

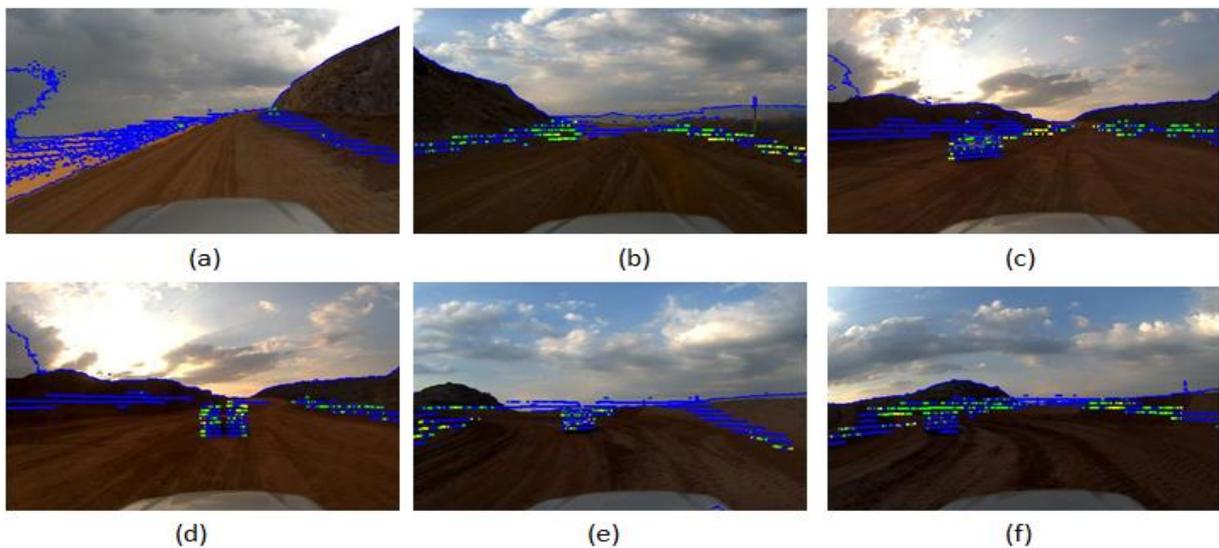


Figura 60: Identificação por fusão de sensores.

#### 4.7 Comparação dos resultados

Comparando as técnicas de identificação de obstáculos e estradas, é possível avaliar a diferença entre os resultados obtidos através das técnicas isoladas e com fusão dos dados.

Na Figura 61, todos os cenários apresentam pontos brancos para identificação do plano pertencente a estrada. Nos cenários (c), (d), (e) e (f) o veículo é identificado por um retângulo vermelho através da rede neural citada anteriormente. A nuvem de pontos colorida representa o resultado da fusão dos sensores.

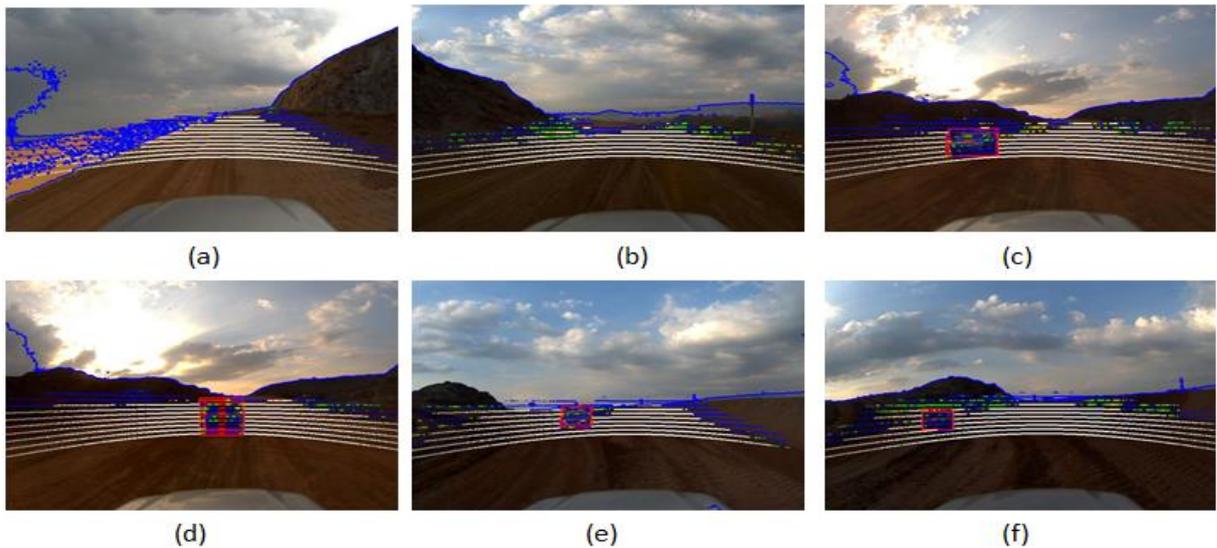


Figura 61: Unificação dos resultados.

## 5 CONCLUSÕES

O setor de mineral, principalmente no Brasil, sofre com as condições climáticas adversas, principalmente nos períodos de chuva e neblina. Trafegar em uma mina em segurança depende de inúmeros fatores e um fator de grande importância é a visibilidade dos operados, que depende diretamente da luminosidade na mina. Quando as condições operacionais não são favoráveis, a equipe de segurança interrompe a operação da mina. A interrupção por condições de visibilidade baixa pode chegar a mais de 10% de perda de produção anualmente, dependendo da mina.

O “Projeto de Direção Assistida” propõe o desenvolvimento de uma interface projetada no painel do equipamento. A solução disponibilizará dados referentes à localização, rota, obstáculos e velocidade que serão captados por dispositivos instalados no equipamento. Neste estudo foi testado somente dois tipos de sensores, câmera RGB e Laser, para identificação de obstáculos e o posicionamento da pista de rolagem dentro de uma mina a céu aberto.

Considerando os resultados apresentados neste trabalho, obtidos através da realização de experimentos em ambientes urbanos e de mineração, conclui-se que é de grande relevância a fusão dos dados oriundos de diferentes sensores, uma vez que um único sensor pode apresentar algum tipo de limitação e menor confiabilidade.

As câmeras, como sensor de imagem RGB, apresentam pontos positivos por ilustrar de forma simples de visualização, por existirem diversas técnicas de classificação de objetos, incluindo redes neurais treinadas com milhões de imagens e diversas classes e que permite acrescentarem mais informações. Porém, por se tratar de um sensor passivo, está sujeito a variações de iluminação.

Os pontos positivos do uso do laser, considerando um sensor ativo, incluem a precisão para identificar à distância dos objetos, o funcionamento mesmo com baixa luminosidade, e a probabilidade de geração de nuvem de pontos em 360° do sensor. Os pontos negativos desse sensor incluem a necessidade de grande quantidade de feixes de laser para ambientes amplos, quanto maior a distância dos objetos a quantidade de pontos pertencentes ao mesmo reduz, dificultando a sua identificação. Existe também a dificuldade de identificação visual pelo usuário e complexidade associada às técnicas de identificação de objetos e planos.

A fusão dos dados, onde os dois sensores identificam os obstáculos em conjunto, mostrou a importância da calibração dos sensores. As técnicas de calibração aumentaram a complexidade para a execução do trabalho. Em contrapartida, o resultado por fusão de sensores mostrou-se eficiente, identificando os objetos ou estrada por variação de imagem e de pontos da nuvem de pontos, trazendo confiabilidade para a detecção.

Neste trabalho foram aplicadas técnicas já desenvolvidas para identificação de obstáculos e estradas por outros pesquisadores, porém foi testado em ambiente de mineração. A técnica para integração dos sensores foi desenvolvida neste trabalho para caráter de prova de conceito, portanto, foi realizada no MATLAB. Notou-se também que o processamento dos dados demanda maior capacidade computacional, o que pode dificultar o processamento para sistemas em tempo real.

## **5.1 Contribuições**

O diferencial deste trabalho é a capacidade de identificação dos obstáculos e pista de rolagem a frente do veículo usando a fusão de dados de dois sensores de modelos diferentes.

As principais contribuições são as validações de técnicas de identificação de obstáculos por imagem em ambiente de mineração. A identificação de planos e conglomerados dentro de uma nuvem de pontos. Todos esses métodos podem ser usados em outras linhas de pesquisa com diferentes finalidades para o crescimento da tecnologia no setor mineral.

## **5.2 Trabalhos futuros**

O primeiro trabalho futuro consiste em utilizar vários sensores devidamente calibrados em torno do equipamento para uma percepção em 360 graus do ambiente ao redor, proporcionando maior segurança ao condutor durante manobras realizadas em operações de mina.

Para abrangência e continuação do estudo, é de grande importância realizar coletas de dados em ambiente de mineração em condições climáticas adversas. Dessa forma as técnicas usadas neste trabalho podem ser aperfeiçoadas ou modificadas para melhor percepção do ambiente, realizando medidas mais quantitativas de acertos e erros.

É de grande importância realizar a fusão dos resultados com outros sensores de diferentes princípios de funcionamento como câmera térmica, GPS, entre outros, como proposto no “Projeto de Direção Assistida”.

Para a melhor aplicabilidade do sistema em ambiente operacional, é preciso reduzir o tempo de processamento, de forma que a identificação dos objetos e estradas seja realizada no menor tempo possível.

Um dos últimos passos seria embarcar o sistema no equipamento de mina, considerando as dificuldades de software e hardware relacionados ao processamento online dos algoritmos aqui propostos.

## 6 REFERÊNCIAS

ABADI, Martín e colab. **TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems**. 2016. Disponível em: <<http://arxiv.org/abs/1603.04467>>.

BOCK, Julian e colab. **Self-learning Trajectory Prediction with Recurrent Neural Networks at Intelligent Intersections**. Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems, n. Vehits, p. 346–351, 2017. Disponível em: <<http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0006374003460351>>.

BOEHLER, Wolfgang e colab. **The potential of non-contact close range laser scanners for cultural heritage recording**. Cipa2001, n. 1, p. 430–436, 2002.

CARNEIRO, Allan C e SILVA, Romuere R V. **Redes Neurais Convolucionais com Tensorflow: Teoria e Prática**. p. 382–406, 2017.

CÉSAR, Renato. **Classification of LASER points cloud using the concept of principal components analysis and the non ambiguity factor**. p. 196–216, 2016.

COLLOBERT, Ronan e colab. **Natural Language Processing (Almost) from Scratch**. Journal of Machine Learning Research, v. 12, p. 2493–2537, 2011.

CORRADETTI, Daniele. **Grupos de Lie**. n. September, 2016.

DIAS, Fabian Gualda e colab. **Anais XV Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Curitiba, PR, Brasil, 30 de abril a 05 de maio de 2011, INPE p.7086**. p. 7086–7092, 2011.

EASTON, Roger L. **Digital Image Processing I**. n. September, 2010.

ESTER, Martin. **Density-Based Clustering Methods**. Comprehensive Chemometrics, v. 2, p. 635–654, 1996.

FERREIRA, Leonardo Nascimento. **Técnica de agrupamento de dados baseada em redes complexas para o posicionamento de cluster heads em rede de sensores sem fio**. p. 102, 2012.

FERREIRA, Santos. **Redes Neurais Convolucionais Profundas na Detecção de Plantas Daninhas em Lavoura de Soja**. p. 70, 2017.

FISCHLER, Martin A e BOLLES, Robert C. **Paradigm for Model**. v. 24, n. 6, 1981.

FREITAS, Sandro Bernard Moreira. **Planejamento estocástico de lavra**:

**metodologias de simulação, otimização e gestão de risco para a mina do futuro.** p. 181, 2015.

GAVA, Éverton Marangoni e colab. **O Algoritmo Density-Based Spatial Clustering of Applications With Noise ( DBSCAN ) na Clusterização dos Indicadores de Dados Ambientais.** v. d, 2013.

GONZALEZ, Rafael C e colab. **Digital Image Processing.** [S.l: s.n.], 2007.

GOODFELLOW, Ian e BENGIO, Yoshua e COURVILLE, Aaron. **Deep Learning.** ICML2013 Tutorial, p. 1–800, 2016. Disponível em: <[www.cs.nyu.edu/~yann/talks/lecun-ranzato-icml2013.pdf](http://www.cs.nyu.edu/~yann/talks/lecun-ranzato-icml2013.pdf)>.

GROETELAARS, Natalie Johanna. **Tecnologia 3D Laser Scanning : características , processos e ferramentas para manipulação de nuvens de pontos.** p. 1–5, 2011.

HABERMANN, Danilo. **Localização topológica e identificação de obstáculos por meio de sensor laser 3D (LIDAR) para aplicação em navegação de veículos autônomos terrestres.** 2016.

HWANG, Soonmin e colab. **Fast Multiple Objects Detection and Tracking Fusing Color Camera and 3D LIDAR for Intelligent Vehicles.** p. 234–239, 2016.

JIA, Yangqing e colab. **Caffe: Convolutional Architecture for Fast Feature Embedding.** 2014. Disponível em: <<http://arxiv.org/abs/1408.5093>>.

KELLY, Alonzo e colab. **Toward Reliable Off Road Autonomous Vehicles Operating in Challenging Environments.** [S.l: s.n.], 2016. v. 25.

KLASING, Klaas e WOLLHERR, Dirk e BUSS, Martin. **A clustering method for efficient segmentation of 3D laser data.** Icara, p. 4043–4048, 2008. Disponível em: <[http://velodynelidar.com/lidar/hdlpressroom/pdf/Articles/A\\_Clustering\\_Method\\_for\\_Efficient\\_Segmentation\\_of\\_3D\\_Laser\\_Data.pdf](http://velodynelidar.com/lidar/hdlpressroom/pdf/Articles/A_Clustering_Method_for_Efficient_Segmentation_of_3D_Laser_Data.pdf)>.

LEVINSON, Jesse e THRUN, Sebastian. **Unsupervised Calibration for Multi-beam Lasers.** 2014.

MITCHELL, Tom M. **Machine learning in ecosystem informatics and sustainability.** [S.l: s.n.], 1997.

NOH, Hyeonwoo e HONG, Seunghoon e HAN, Bohyung. **Learning deconvolution network for semantic segmentation.** Proceedings of the IEEE International Conference on Computer Vision, v. 2015 Inter, p. 1520–1528, 2015.

OECD. **Education at a Glance 2012.** [S.l: s.n.], 2012.

P. MURPHY, Kevin. **Machine Learning: A Probabilistic Perspective.** [S.l: s.n.], 2012. Disponível em: <[http://link.springer.com/chapter/10.1007/978-94-011-3532-0\\_2](http://link.springer.com/chapter/10.1007/978-94-011-3532-0_2)>.

PAVAN, Nadisson Luis. **UM MÉTODO AUTOMÁTICO PARA REGISTRO DE DADOS LASER An automatic method for registration of terrestrial laser scanning data using planar surfaces**. p. 572–589, 2015.

PENHARBEL, Eder a e colab. **Filtro de Imagem Baseado em Matriz RGB de Cores-Padrão para Futebol de Robôs**. 2013.

PEYNOT, Thierry e UNDERWOOD, James e SCHEDING, Steven. **Towards reliable perception for unmanned ground vehicles in challenging conditions**. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, p. 1170–1176, 2009.

PINHO, Carolina Moutinho Duque e RENNÓ, Camilo Daleles e KUX, Hermann Johann Heinrich. **Avaliação de técnicas de fusão aplicadas à imagens Quickbird**. XII Simpósio Brasileiro de Sensoriamento Remoto, n. 1999, p. 4225–4232, 2005. Disponível em: <<http://marte.sid.inpe.br/col/ltid.inpe.br/sbsr/2004/11.22.20.14/doc/4225.pdf>>.

PISANO, Paul A. e GOODWIN, Lynette C. e ROSSETTI, Michael A. **U.S. highway crashes in adverse road weather conditions**. 24th Conference on Institutional Information Processing System, p. 1–15, 2008. Disponível em: <<http://ams.confex.com/ams/pdfpapers/133554.pdf>>.

PONCE, J e FORSYTH, D. **Computer vision: a modern approach**. [S.l: s.n.], 2012. Disponível em: <<http://www.inria.fr/centre/paris-rocquencourt/actualites/computer-vision-a-modern-approach>>.

PONTOS, Nuvem De e VIRTUAL, Realidade. **Otimização de nuvens de pontos densas para softwares de visualização em ambiente de realidade virtual**. 2010.

RASMUSSEN, Christopher. **Texture-Based Vanishing Point Voting for Road Shape Estimation**. v. 1, 2016.

REGIS, Adriano e colab. **Estudo da tecnologia de fusão de sensores em lógica programável aplicado a sistema de monitoramento ambiental (1)** . n. 1, p. 4–7, 2013.

REN, Shaoqing e colab. **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks**. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 39, n. 6, p. 1137–1149, 2017.

ROBERT A. SCHOWENGERDT. **Remote Sensing: Models and Methods for Image Processing**. [S.l: s.n.], 2012. v. 95. Disponível em: <<http://www.cabdirect.org/abstracts/19790652875.html%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/22115514>>.

ROCHA, Antônio Carlos Pereira. **Aplicações do Scanner a Laser I-SiTE para levantamentos topográficos**. Rem: Revista Escola de Minas, v. 55, n. 4, p. 307–311,

2002. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0370-44672002000400013&lng=pt&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0370-44672002000400013&lng=pt&nrm=iso&tlng=pt)>.

ROCHA, Rafael Henrique Santos. **Reconhecimento de Objetos por Redes Neurais Convolutivas**. p. 49, 2015.

ROSERO, Luis Alberto. **Detecção de obstáculos usando fusão de dados de percepção 3D e radar em veículos automotivos**. 2017.

SALIMANS, Tim e colab. **Pixel Cnn ++ : Improving the Pixel Cnn With Discretized Logistic Mixture Likelihood and**. p. 1–10, 2017.

SANCHES, Carlos H e colab. **Técnicas de Suavização de Imagens e Eliminação de Ruídos**. p. 21–30, 2015.

SCHÖLKOPF, B. **Learning with kernels**. Journal of the Electrochemical Society, v. 129, n. November, p. 2865, 2001. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.167.5140&rep=rep1&type=pdf>>.

SCURI, Antonio Escaño. **Fundamentos da Imagem Digital**. p. 95, 2002.

SEGAL, Aleksandr V e HAEHNEL, Dirk e THRUN, Sebastian. **Generalized-ICP**. Proc. of Robotics: Science and Systems, v. 2, p. 4, 2009.

SIMONYAN, Karen e ZISSERMAN, Andrew. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. p. 1–14, 2014. Disponível em: <<http://arxiv.org/abs/1409.1556>>.

SOUZA, J. S. D. e CARDOZA, J. A. S. **Sensores de Imagem Digitais CCD E CMOS**. VII Congresso de Pesquisa e Inovação da Rede Norte e Nordeste de Educação Tecnológica, v. 2009, p. 1–8, 2012.

TOMMASELLI, Antonio Maria Garcia. **Um estudo sobre as técnicas de varredura a laser e fotogrametria para levantamentos 3D a curta distância**. GEODÉSIA Online-Revista da Comissão Brasileira de Geodésia, v. 4, n. 1415–1111, 2003.

VEDALDI, Andrea e LENC, Karel. **MatConvNet**. Proceedings of the 23rd ACM international conference on Multimedia - MM '15, p. 689–692, 2015. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2733373.2807412>>.

VEDALDI, Andrea e LENC, Karel. **MatConvNet - Convolutional Neural Networks for MATLAB**. 2014. Disponível em: <<http://arxiv.org/abs/1412.4564>>.

VELAS, Martin e colab. **Calibration of RGB Camera With Velodyne LiDAR**. BIOSCI. BIOTHECH BIOCHEM, p. 5, 2014. Disponível em: <[http://www.fit.vutbr.cz/research/view\\_pub.php?id=10578%5Cnhttp://arxiv.org/abs/1505.0729%5Cnhttp://mi.eng.cam.ac.uk/projects/segnet/](http://www.fit.vutbr.cz/research/view_pub.php?id=10578%5Cnhttp://arxiv.org/abs/1505.0729%5Cnhttp://mi.eng.cam.ac.uk/projects/segnet/)>.

WAHL, R e GUTHE, M e KLEIN, R. **Identifying Planes in Point-Clouds for Efficient Hybrid Rendering**. The 13th Pacific Conference on Computer Graphics and Applications, 2005.

YUENH.K., Princen J. **A COMPARATIVE STUDY OF HOUGH TRANSFORM METHODS FOR CIRCLE FINDING** H . K . Yuen , J . Princen , J . Illingworth and J . Kittler **Department of Electronics and Electrical Engineering**. p. 169–174, 1989.

ZHAO, Gangqiang e colab. **Fusion of 3D-LIDAR and camera data for scene parsing**. Journal of Visual Communication and Image Representation, v. 25, n. 1, p. 165–183, 2014.