UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Análise e fusão de imagens 2D e 3D com vistas para detecção e classificação de sinais de trânsito verticais em prol da segurança viária com veículos robóticos inteligentes

Diego Renan Bruno

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

ICMC USP
SÃO CARLOS

Diego Renan Bruno

Análise e fusão de imagens 2D e 3D com vistas para detecção e classificação de sinais de trânsito verticais em prol da segurança viária com veículos robóticos inteligentes

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Fernando Santos Osório

USP – São Carlos
Junho de 2020

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

Bruno, Diego Renan
Análise e fusão de imagens 2D e 3D com vistas para detecção
e classificação de sinais de trânsito verticais em prol da
segurança viária com veículos robóticos inteligentes /
Diego Renan Bruno; Fernando Santos Osório. -- São Carlos, 2020.
191 p.

Tese (Doutorado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2020

1. Computação - Matemática. 2. Redes neurais (Computação)
3. Veículos Autônomos. 4. Visão Computacional.
5. Aprendizado de Máquina. I. Osório, Fernando
II. Universidade de São Paulo.

Diego Renan Bruno

Analysis and merging of 2D and 3D images for detecting and classifying vertical traffic signs to the benefit of the road safety with intelligent robotic vehicles

Doctoral dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Fernando Santos Osório

USP – São Carlos
June 2020

Dedico meu conhecimento obtido neste trabalho ao grande Mestre Menino Jesus de Praga.

Este trabalho é dedicado a minha namorada Lara Alicia, por ter feito parte de tudo isso, ter me apoiado e feito dos dias de pesquisa mais felizes e tranquilos, sempre me apoiando e me incentivando.

Aos meus amados pais, Pedro e Mariluce e meu irmão, Danilo, por todo apoio e carinho de sempre. As minhas avós, Ires e Isabel. Também dedico a Dona Eliete e Teresa Catanho, dois anjos que encontrei nessa vida.

A todos cientistas que dedicam suas vidas pela ciência. Em especial, ao meu orientador de doutorado, Fernando Santos Osório, que nunca mede esforços em seu trabalho e que sempre foi um exemplo para mim, contribuindo de maneira grandiosa na pessoa que sou hoje. Também a todos pesquisadores do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo.

Dedico também este trabalho para meu orientador de mestrado, Norian Marranghello, que por alguns anos me ensinou muitas coisas da Robótica e da vida, conhecimento que levarei para sempre.

Não poderia deixar de lembrar dos parceiros e irmãos de vida, Marco Di Foggi e Denis Mosconi, meus grandes e velhos conselheiros. Pessoas que sempre estiveram ao meu lado sendo fundamentais em tudo que já passei.

AGRADECIMENTOS

Agradeço mais uma vez ao grande Mestre Menino Jesus de Praga por sempre me guiar nos caminhos dos estudos, trazendo-me paz, sabedoria e humildade, pois, sem Ele este trabalho não seria possível.

Gostaria de agradecer especialmente o Professor Fernando Santos Osório, que me orientou desde o meu último ano de mestrado na pesquisa para a robótica e sempre mostrou grande dedicação em sua orientação, sendo mais que um orientador, foi um grande pai na academia que me incentivou a buscar desafios que eu nunca imaginava conseguir. Sua orientação contribui de maneira grandiosa na minha formação acadêmica.

Agradeço a minha namorada Lara Alicia por ter dado apoio e acreditado em mim nos momentos de pesquisa, me apoiando em cada nova loucura acadêmica mesmo com poucas horas de sono. Ela acreditou em mim quando nem eu mesmo acreditava.

Agradeço aos meus amados pais, Pedro e Mariluce, que sempre me apoiaram em fazer dos estudos um potencial diferenciado em minha vida. Não deixo de agradecer também ao meu irmão Danilo, que sempre teve suas curiosidades sobre o meu trabalho e sempre esteve ao meu lado durante os meus estudos. Agradeço também às minhas avós Ires e Isabel, que sempre contribuíram para a pessoa que sou hoje.

Agradeço à Universidade de São Paulo, ao Instituto de Ciências Matemáticas e de Computação e a todos os seus docentes que estiveram neste caminho de estudos, seja como professor de algumas disciplinas ou como banca de avaliação de alguns trabalhos que apresentei durante este tempo que estive como aluno. Agradeço também aos amigos deste mesmo instituto, especialmente aos amigos do Laboratório de Robótica Móvel. Em especial também agradeço os meus amigos Venilton Falvo Júnior, Catherine Martins, Luis Rosero, Germain Garcia Zanabria, Fernando Pereira dos Santos, Rafael Berri e tantos outros que estiveram comigo.

Agradeço a dois grandes Mestres e amigos que me motivaram, apoiaram e inspiraram para a vida acadêmica, professores Tácio Barbeiro e Marcio Marques.

Não poderia me esquecer de duas grandes mães, a Dona Eliete Estavam Gomes e Teresa Catanho, que sempre investiram em minha carreira, sendo não somente amigas de trabalho, mas também pessoas cheias de luz e que iluminaram os meus caminhos.

A todos minha eterna gratidão, sem cada um de vocês não seria possível atinger este mais alto nível da área acâdemica, o que me habilita a ser um professor melhor e contribuir cada vez mais para a sociedade em que faço parte.

"As máquinas me surpreendem muito frequentemente."
(Alan Mathison Turing)

"A Matemática é a única linguagem que temos em comum com a natureza."
(Stephen Hawking)

"O lado negro não é mais poderoso, apenas mais rápido, mais fácil e mais sedutor."
(Mestre Yoda)

"O desenvolvimento humano depende fundamentalmente da invenção. Ela é o produto mais importante de seu cérebro criativo. Seu objetivo final é o completo domínio da mente sobre o mundo material e o aproveitamento das forças da natureza em favor das necessidades humanas."
(Nikola Tesla)

RESUMO

BRUNO, D. R. **Análise e fusão de imagens 2D e 3D com vistas para detecção e classificação de sinais de trânsito verticais em prol da segurança viária com veículos robóticos inteligentes**. 2020. 191 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Os Veículos Robóticos Inteligentes são aplicados principalmente em beneficio da redução de acidentes de trânsito, possibilitando então reduzir falhas e imprudências humanas com sistemas que utilizam Visão Computacional, Inteligência Artificial, Automação e outras tecnologias, para auxiliar o condutor em sua tarefa de dirigir. Aplicando a robotização, também é possível o aumento do nível da segurança viária por meio do desenvolvimento de veículos autônomos totalmente livres do controle humano e que são programados para navegarem dentro das leis de trânsito. Sendo que as falhas humanas são a causa de mais de 90% para acidentes fatais em todo o mundo. Esta pesquisa de doutorado teve como objetivo principal o estudo, proposta, desenvolvimento, adaptações e testes de um conjunto de técnicas e métodos de Visão Computacional e Inteligência Artificial, com vistas para um sistema de percepcão com fusão de imagens $2D$ e $3D$ mais robusto para detecção de sinais de trânsito verticais. Também foi desenvolvido um modelo de Atenção Visual *Fuzzy*, capaz de analisar a prioridade de cada informação detectada por meio do sistema de percepção, possibilitando então dar suporte para a tomada de decisão do veículo envolvendo situações de emergência (acidentes e obras na via), utilizando como base, os valores de prioridade de cada regra de trânsito.

O sistema de Visão Computacional Robótica deve ser capaz de detectar, classificar e analisar a prioridade dos sinais de trânsito verticais utilizados atualmente e, que são funcionais para o trânsito envolvendo motoristas humanos no mundo real, não exigindo adaptações da sinalização. O sistema de visão deve então auxiliar um veículo totalmente autônomo, ou semi-autônomo, a navegar dentro das regras de trânsito locais, assim, detectando informações de grande importância, como: velocidade máxima, parada obrigatória, cones de emergência e cores do semáforo. Em casos de navegação autônoma, apenas o sistema de percepção e análise de sinais de trânsito verticais deve ser utilizado. Já para a navegação semi-controlada, ou seja, com auxílio de um humano, o sistema de visão externo deve trabalhar em conjunto com a análise do condutor e dos dados de controle do veículo, ativando rotinas automáticas corretivas com base nos erros detectados na tarefa de dirigir, possibilitando evitar graves acidentes relacionados com o desrespeito as sinalizações de trânsito e que são gerados por falha humana e imprudência.

Palavras-chave: Veículos Robóticos Inteligentes, Visão Computacional, Atenção Visual, Deep Learning, Fusão de Sensores, ADAS.

ABSTRACT

BRUNO, D. R. **Analysis and merging of 2D and 3D images for detecting and classifying vertical traffic signs to the benefit of the road safety with intelligent robotic vehicles**. 2020. 191 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Intelligent Robotic Vehicles are mainly applied for the benefit of the traffic accidents reduction, thus enabling to reduce human faults and recklessness with systems that use Computer Vision, Artificial Intelligence, Automation and other technologies to assist the driver in his driving task. By applying robotization, it is also possible to increase the level of road safety by developing autonomous vehicles totally free of human control and which are programmed to navigate within traffic laws. Human failures in driving take responsability over 90% for fatal accidents worldwide. The main objective of this doctoral research was the study, proposal, development, adaptations and tests of a set of techniques and methods of Computer Vision and Artificial Intelligence, aiming at a $2D$ and $3D$ image fusion perception system, being more robust for detecting vertical road signs. A Fuzzy Visual Attention model was also developed, capable of analyzing the priority of each information detected through the perception system, thus enabling the decision making of the vehicle involving emergency situations (accidents and road works) to be supported. The Fuzzy Visual Attention model uses the priority values of each traffic sign as a basis.

The Robotic Computer Vision system shall be capable of detecting, classifying and analyzing the priority of currently used vertical traffic signs that are functional for traffic involving human drivers in the real world and do not require signaling adaptations. The vision system should then assist a fully autonomous or semi-autonomous vehicle to navigate within local traffic rules, thus detecting important information such as maximum speed, mandatory stop, emergency cones and traffic light colors. In cases of autonomous navigation, only the system of perception and analysis of vertical traffic signs should be used. For semi-controlled navigation, that is, with the help of a human, the external vision system should work in conjunction with driver analysis and vehicle control data, activating automatic corrective routines based on errors detected in the task execution. Thus making it possible to avoid serious accidents related to disregarding traffic signs, due mainly to human errors and recklessness.

Keywords: Intelligent Robotic Vehicles, Computer Vision, Visual Attention, Deep Learning, Sensor Fusion, ADAS.

LISTA DE ILUSTRAÇÕES

LISTA DE ALGORITMOS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AHP	*Analytic Hierarchy Process*
CAN	*Controller Area Network*
CaRINA	Carro Robótico Inteligente para Navegação Autônoma
CNN	*Convolutional Neural Networks*
CPU	*Central Processing Units*
CRob	Centro de Robótica da USP de São Carlos
DARPA	*Defense Advanced Research Projects Agency*
DCNN	*Deep Convolutional Neural Network / ConvNet*
FIS	*Fuzzy Inference System*
GPU	*Graphics Processing Unit*
HSV	*Hue, Saturation e Value*
ICMC	Instituto de Ciências Matemáticas e de Computação
InSAC	Instituto Nacional de Ciência e Tecnologia para Sistemas Autônomos Cooperativos
LIDAR	*Light Detection and Ranging*
LRM	Laboratório de Robótica Móvel
LSTM	*Long Short-Term Memory*
MADM	*Multiple Attribute Decision-Making*
MDL	*Minimum Description Length principle*
MLP	*Multi-Layer Perceptron*
PDI	Processamento Digital de Imagens
RefCN	*Reflectance ConvNet*
RGB-D	*Red, Green e Blue + Depth*
RNA	Rede Neural Artificial
SVM	*Support Vector Machine*
TOF	*Time-of-Flight*
TOPSIS	*Technique for Order of Preference by Similarity to Ideal Solution*
TSR	*Traffic Sign Recognition*
USP	Universidade de São Paulo
YOLO	*You Only Look Once*

SUMÁRIO

CAPÍTULO

1

INTRODUÇÃO

1.1 Contextualização

Os robôs móveis autônomos utilizam técnicas computacionais de grande complexidade para que seja possível navegar em variados tipos de ambientes dinâmicos, deste modo, evitando colisões com obstáculos e sempre buscando otimizar a melhor rota, possibilitando então o funcionamento de um sistema de controle seguro e preciso. Para que uma navegação neste nível seja possível de ser realizada, são utilizadas variadas técnicas de sensoriamento e de algoritmos de controle. Uma área de grande interesse na robótica móvel está muito ligada a navegação de veículos robotizados. Estes veículos tem a capacidade de navegar de maneira autônoma, sem a necessidade de controle humano e, também, em outros casos, podem auxiliar motoristas por meio de modelos de assistência a condução.

Estes mecanismos de auxílio para navegação de veículos inteligentes são aplicados em duas situações principais e que estão relacionadas com as maiores motivações deste trabalho de doutorado: (a) **Veículos autônomos**: onde a tarefa de dirigir é totalmente feita pelo controle automático, dispensando a tarefa de dirigir por parte de um humano. Este tipo de aplicação é voltada para pessoas que não tem a capacidade de dirigir um veículo por algum tipo de restrição e e desejam ter acessibilidade para uma locomoção rápida e segura em um ambiente urbano. Ainda para veículos autônomos, por meio de um piloto automático podem dar suporte para um motorista que precisa descansar em uma viagem de longa distância e, também, para (b) **Suporte a condução de veículos semi-autônomos**: aplicados em prol de um ambiente de trânsito mais seguro, assim, sendo possível evitar acidentes gerados principalmente pela falha humana e imprudência. Para este tipo de aplicação, um motorista humano ainda trabalha para dirigir o veículo, no entanto, por meio de assistência automatizada, uma falha pode ser corrigida ou alertada, podendo evitar então um grave acidente.

1.2 Motivação

Os acidentes de trânsito estão entre as principais causas de morte em todo o mundo (BUCKERIDGE, 2015) e sendo a falha humana responsável pela grande maioria dos casos, totalizando 95% das vitimas fatais (Amditis *et al.*, 2010). As principais falhas dos seres humanos na atividade de dirigir estão relacionadas com: embriaguez, sonolência e as distrações do motorista de maneira geral, principalmente pelo uso do celular (Murata *et al.*, 2011). Apesar de todas as leis que são aplicadas de maneira rígida em todo o mundo, a fiscalização é incapaz de abordar todo condutor dirigindo de maneira imprudente, consequentemente apenas 1% dos motoristas embriagados são detectados por meio da polícia (CDC, 2013) ou agindo de outra forma irregular. As mortes geradas por motoristas embriagados nas estradas americanas no ano de 2008 chegou ao valor de 30% do total de vítimas fatais neste mesmo período (WILSON FA, 2010). A fiscalização de trânsito brasileira também leva em consideração que a sonolência é um indicador de embriaguez (OGAMA, 2014). Já no Brasil, um estudo feito em 143 cidades brasileiras no ano de 2009, apontou que beber e dirigir é normal e aceitável para 35% da população (PECHANSKY; BONI R.; DIEMEN, 2009).

Outro problema é a sononolência, que para os sociólogos Steve Kroll-Smith e Valerie Gunter, esta situação é uma nova embriaguez, pois é um estado culpável, já que o condutor se permite realizar a tarefa de dirigir sob esta condição insegura, então sendo um caso bem parecido com a direção com efeitos alcoólicos (WILLIAMS, 2011).

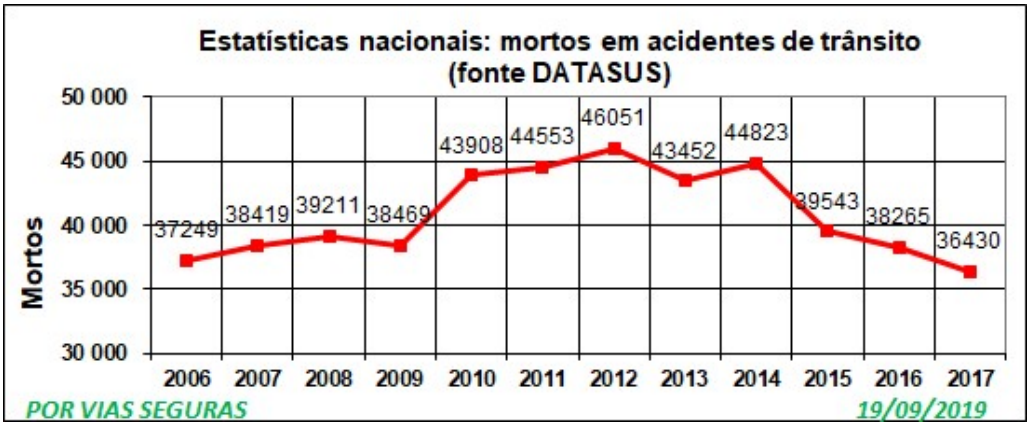
Um dos maiores problemas está relacionado com as distrações e desrespeito as regras de trânsito geradas pelo motorista (PEISSNER; DOEBLER; METZE, 2011), sendo que muitas vezes as informações passadas por meio dos sinais de trânsito verticais são desrespeitadas. O motorista que não percebe ou ignora os sinais de trânsito, estará desobedecendo as regras locais e poderá gerar um possível acidente de trânsito. Para quantificar este problema, cerca de 63.000 acidentes no ano de 2010 foram ocasionados pela desatenção do motorista (SALVADOR, 2011). Em dados mais recentes, cerca de 30% das mortes no trânsito são causadas por desrespeito as leis, gerando um total de 23 mil mortos no ano de 2017 e 2018. Para a situação de falta de atenção para as leis de trânsito, um total de 23%, gerando 15 mil mortos e 276 mil feridos (INFRAESTRUTURA, 2019). Sendo que as distrações no trânsito são ocasionadas em três diferentes situações: visual, mãos fora do volante e cognitiva (STRAYER *et al.*, 2013).

Na distração visual o motorista perde o foco da visão que era para estar na estrada e sinalizações de trânsito, com outra situação. As mãos fora do volante são retiradas para que seja feita outra tarefa (que irá desviar o foco da atenção do motorista), sendo esta uma falha de grande perigo. Para exemplificar esta situação, um motorista falando no celular enquanto dirige (STRAYER; DREWS; JOHNSTON, 2003), irá desviar o foco da visão e retirar pelo menos uma mão do volante. Já a distração cognitiva está muito ligada a distração do motorista com alguma situação ou outra tarefa executada ao mesmo tempo. Portanto, uma ligação via celular pode implicar nas três distrações ao mesmo tempo (STRAYER *et al.*, 2013), consequentemente

perdendo processamento das informações visuais e, também, tendo uma queda do nível de percepção (STRAYER *et al.*, 2013). Dirigir falando no celular é um problema alarmante para as distrações no trânsito, gerando um aumento de até 6 vezes no risco de colisão (REDELMEIER; TIBSHIRANI, 1997) (STRAYER; DREWS, 2004).

O Brasil é o país com um dos maiores números de mortes em acidentes de trânsito causados por falha humana. No gráfico da Figura 1 pode ser observado os dados estatísticos coletados entre os anos de 2004 e 2017 e que são referentes a quantidade de mortos em acidentes de trânsito em território nacional.

Figura 1 – Número de óbitos em acidentes de trânsito registrados pelo Ministério da Saúde.



Fonte: (SEGURAS, 2016).

Visando estes problemas no trânsito mundial e, que são gerados por diversos tipos de falhas e imprudência humana, é buscado o desenvolvimento de novas tecnologias para dar suporte ao aumento da segurança viária. Estas tecnologias são bastante ligadas aos veículos autônomos, onde um condutor é dispensado totalmente da tarefa de dirigir e, também, aos veículos semi-autônomos, onde modelos automáticos de assistência ao condutor são aplicados. Para exemplificar este cenário de pesquisas, pode ser destacado o Carro Robótico Inteligente para Navegação Autônoma (CaRINA), sendo um Fiat Palio *Adventure* automático (SALES *et al.*, 2012), o primeiro veículo autônomo desenvolvido e testado em vias públicas na America Latina. Desenvolvido junto ao Laboratório de Robótica Móvel (LRM) do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo (USP) de São Carlos. Outro exemplo de grande projeto é Google *Self-Driving Car* (FISHER, 2014). Estes dois grandes trabalhos são aplicados em prol dos veículos robóticos com capacidade de navegação autônoma. Então, o desenvolvimento de veículos robóticos autônomos e semi-autônomos junto ao LRM/ICMC/USP, está evoluindo e acompanhando o estado-da-arte nesta área de pesquisas, mostrando um grande esforço científico e de grande potencial aplicado nesta grande área da robótica móvel, possibilitando aumentar consideravelmente a segurança viária.

Para exemplificar e dar um melhor entendimento do problema de rotas de veículos robóticos em ambientes dinâmicos com regras de trânsito, pode-se citar os eventos da *Defense Advanced Research Projects Agency* (DARPA), sendo eles o *Grand Challenge* e o *Urban Challenge* (DARPA, 2013). Os desafios tiveram como objetivo uma corrida realizada entre robôs (carros autônomos não tripulados). O evento contou com duas edições, a primeira edição, no ano de 2004, teve como objetivo atravessar o deserto de Mojave, nos Estados Unidos, com 240km de extensão, no entanto, neste ambiente não existiam regras de trânsito. Já no próximo desafio, no ano de 2007, a complexidade do desafio foi aumentada, sendo considerado um ambiente dinâmico de trânsito urbano, típico de cidades do estado da Califórnia nos EUA (DARPA, 2013). O desafio foi nomeado como DARPA *Urban Challenge*, teve como maior complicação as regras de trânsito locais, para que fosse possível se locomover com segurança sobre este ambiente dinâmico. O desafio teve 55 robôs inscritos e apenas três deles cumpriram todas as etapas da prova (DARPA, 2013). Este último desafio do DARPA é muito ligado com esta pesquisa de doutorado, onde o reconhecimento das regras de trânsito locais é fundamental para que o veículo consiga navegar de maneira segura em um ambiente controlado por regras. É importante destacar também que o DARPA *Urban Challenge* foi realizado em uma pista de testes especialmente criada para este fim, sendo portanto um ambiente artificial que não representava situações reais de trânsito urbano, sendo apenas um simulação de um ambiente urbano real (BRUNO, 2016).

1.3 Pergunta Científica

É possível por meio da integração de um modelo de Visão Computacional Robótico baseado em técnicas de percepção para detecção e reconhecimento de sinais de trânsito com fusão de imagens 2D e 3D, associada a análise semântica do contexto da cena com múltiplas informações conflitantes, ter uma maior robustez para o tratamento de falsos positivos e falsos negativos e dar suporte para uma melhor tomada de decisão para veículos robóticos inteligentes com base em informações de sinalização de trânsito?

1.4 Hipótese

A fusão de imagens 2D e 3D permite a definição de um novo modelo do estado-da-arte de Visão Computacional Robótica mais robusto na detecção e reconhecimento de sinais de trânsito verticais, bem como a redução de falsos objetos e falsas situações, o que possibilita obter um sistema de Atenção Visual focado em analisar e priorizar sinais de trânsito de maior relevância em situações não mapeadas e com conflitos de informações sobre as regras de trânsito locais para veículos autônomos e semi-autônomos.

1.5 Desenvolvimento

Nesta pesquisa de doutorado foram realizadas as adaptações e desenvolvimento de técnicas e métodos de percepção baseados em Visão Computacional e Aprendizado de Máquina com vistas para a detecção, classificação e análise de sinais de trânsito verticais em ambientes não mapeados ou em constante modificação (desvios, acidentes, obras), utilizando fusão de imagens $2D$ e $3D$. As imagens $3D$ são capturadas e geradas por meio de sensores a lasers e de imagens estereoscópicas, os quais permitem obter dados em profundidade (sendo em coordenadas X, Y e Z), possibilitando então eliminar falsos objetos de trânsito em imagens puramente $2D$. Já as imagens $2D$, são geradas por câmeras monoculares e aplicadas para a tarefa de reconhecimento de objetos em dados *RGB* das superfícies dois sinais de trânsito, aproveitando o potencial das redes de *Deep Learning* para trabalhar com cores, texturas e formas.

Atualmente, grande parte das técnicas de Visão Computacional presentes na literatura, voltadas para o reconhecimento de sinais de trânsito verticais, utilizam apenas imagens $2D$. Porém, a maioria delas possui algumas deficiências que podem gerar problemas na identificação de elementos da cena, como por exemplo: falta de informação de profundidade e dificuldade em trabalhar em ambientes reais onde existe a variação de luminosidade. Por outro lado, sistemas de visão $2D$ e $3D$ podem trabalhar em conjunto, possibilitando uma análise mais robusta de imagens de sinais de trânsito verticais. O sistema de percepção com fusão de dados $2D$ e $3D$ que foi desenvolvido nesta tese, é capaz de, por meio da fusão de dados $2D$ e $3D$, detectar e caracterizar atributos presentes em variados tipos de sinais de trânsito verticais, já que nem sempre será possível obter padrões bem definidos para estas informações em ambientes dinâmicos.

Também foi desenvolvido nesta tese um modelo de Atenção Visual, capaz de analisar a prioridade de cada sinal de trânsito detectado, onde a presença de múltiplos sinais de trânsito pode acabar gerando conflitos quanto a prioridade e relevância de cada um destes. Este modelo é capaz, de por meio de regiões de interesse e algoritmos de tomada de decisão baseados em lógica *Fuzzy*, classificar o nível de relevância de cada informação de trânsito avaliada.

O modelo de Atenção Visual tem a capacidade de analisar um conjunto de sinais de trânsito detectados ao mesmo tempo pelo sistema de percepção, eliminando o problema de informações múltiplas em conflito. Esta análise deve ser gerada por uma camada inteligente entre o sistema de percepção (detecção e classificação), e o sistema de tomada de decisão do veículo, garantindo um tratamento em maior nível para as regras de trânsito, que em algumas situações, geram conflitos em vias não mapeadas: velocidade, parada obrigatória, semáforo e sentido da via.

1.6 Contribuições Sociais

Os modelos de percepção e Atenção Visual desenvolvidos nesta tese de doutorado, podem dar suporte para veículos inteligentes com base nos sinais de trânsito verticais e, que são provenientes da detecção, classificação e análise de imagens em situações de navegação autônoma ou semi-autônoma. Garantindo que um veículo autônomo possa navegar dentro das regras de trânsito que são informadas para motoristas humanos. Possibilitando também, em casos de navegação semicontrolada, alertar e penalizar o motorista sobre seu comportamento inadequado perante as regras desrespeitadas ou despercebidas, visto que o policiamento e radares de velocidade, não cobrem todos os problemas atuais. Os dados de percepção externa também devem auxiliar na geração de soluções previamente treinadas em uma base de conhecimento, onde o sistema automático deve corrigir falhas humanas (redução de velocidade e frenagem), ou em casos extremos, assumir o controle do veículo e possibilitando evitar um grave acidente.

A integração do modelo de percepção para sinais de trânsito, em conjunto com um Sistema Avançado de Auxílio ao Condutor (*Advanced Driver Assistance Systems* - ADAS), deve ser capaz de contribuir também com a educação no trânsito, já que pode auxiliar na detecção de desrespeito as regras de trânsito locais, permitindo avaliar como o veículo está sendo dirigido por um humano. Por meio deste tipo de diagnóstico deve ser possível que:

- As leis de trânsito sejam controladas automaticamente e sem auxílio de radares para controlar a velocidade e guardas de trânsito para outras situações, já que os dados de percepção dos sinais de trânsito, em conjunto com os dados do veículo, permitiriam avaliar e detectar as situações de imprudência;
- Possibilita também auxiliar no controle da carteira de habilitação, fazendo com que um motorista que desrespeita as leis de trânsito perca o direito de dirigir e causar risco para a sociedade;
- Uma seguradora possa penalizar ou bonificar seu cliente por seu mau ou bom comportamento no trânsito.

O sistema de percepção externo deve possibilitar então auxiliar dois tipos de aplicações distintas: (1) quando em uma navegação autônoma, sendo então possível a avaliação dos sinais de trânsito locais e, com isso, auxiliar o sistema autônomo de navegação do veículo robótico a navegar dentro das regras ou (2) em uma navegação semicontrolada, onde uma percepção externa do ambiente de navegação com sinais de trânsito e, também, dos dados internos do veículo que são gerados pelo comportamento do motorista, devem dar suporte para a detecção e correção de falhas humanas. Visando estes aspectos, foi possível definir estas aplicações que aproveitam o potencial desta pesquisa de doutorado, gerando uma possível contribuição com o aumento da segurança viária com veículos robóticos inteligentes.

1.7 Objetivos Alcançados: Contribuições Científicas

Com o presente trabalho de doutorado foi possível gerar contribuições científicas significativas e junto ao estado-da-arte para modelos de percepção e Atenção Visual de veículos inteligentes. Envolvendo técnicas de extração de características $3D$, Processamento de Imagens, Visão Computacional, Aprendizado de Máquina e Análise Semântica *Fuzzy*, por meio da fusão de imagens $2D$ e $3D$ aplicadas para detecção, reconhecimento e análise semântica de sinalizações de trânsito verticais. Possibilitando então um maior nível de automação inteligente para sistemas de visão robóticos, dando suporte para ADAS e veículos autônomos em prol do aumento do nível da segurança viária e que envolve regras de trânsito locais. A pesquisa e desenvolvimento foram voltadas especificamente para o trabalho nos seguintes tópicos:

- Fusão de informações $2D$ (imagens) e $3D$ (profundidade dos elementos da cena) de modo a melhor detectar e reconhecer os sinais de trânsito verticais com noção de cores, texturas, formas e profundidade dos objetos;
- Manipulação dos dados de sensores inteligentes $3D$ do estado-da-arte, adaptando e desenvolvendo novas técnicas para o processamento dos dados destes sensores;
- Adaptação e aplicação de novas técnicas de extração de features $3D$;
- Pesquisa, adaptação e desenvolvimento de métodos avançados de Aprendizado de Máquina baseados em Redes Neurais, *Deep Learning* e Lógica *Fuzzy*, visando melhorar o desempenho na detecção e reconhecimento de sinais de trânsito verticais;
- Percepção de sinais de trânsito verticais em relação a área navegável, dando suporte para a análise semântica entre os sinais de trânsito detectados e suas vias correspondetes;
- Desenvolvimento de um modelo de Atenção Visual baseado em Lógica *Fuzzy* e, que possibilita idealizar um sistema robotizado de visão com uma base de regras inspiradas no comportamento da visão humana e, também, no conhecimento especialista para análise semântica da cena. O modelo de Atenção Visual é aplicado para análise de prioridades dos sinais de trânsito detectados em conflito de regras, interpretando problemas de emergência na via em situações não mapeadas: desvios, acidentes, obras na pista e outras situações que envolvem novas sinalizações de trânsito temporárias;
- Suporte junto aos sistemas de tomada de decisão de veículos autônomos e semi-autônomos, por meio da integração da percepção e análise de prioridade dos sinais de trânsito;
- Análise integrada da sinalização de trânsito juntamente com os dados de condução do veículo, possibilitando o ajuste e correção de falhas no controle autônomo ou semi-autônomo.

1.8 Aplicações

A pesquisa desenvolvida nesta tese de doutorado é direcionada para aplicações de veículos autônomos. Suas maiores contribuições são focadas para a (a) redução de acidentes de trânsito gerados por falhas humanas na tarefa de dirigir, (b) um piloto automático para dar suporte em viagens de longas distâncias e (c) possibilitar que um deficiente visual ou com outra deficiência que o impeça de dirigir, tenha o auxílio deste tipo de veículo para sua mobilidade.

O trabalho que foi desenvolvido nesta tese, teve como objetivo, criar métodos e técnicas computacionais que permitam beneficiar modelos de Visão Computacional e Atenção Visual para que trabalhem em conjunto com fusão de imagens $2D$ e $3D$, podendo dar suporte tanto para ADAS, quanto para a navegação autônoma e livre do trabalho humano para a tarefa de dirigir.

Esta pesquisa de doutorado contribui então especificamente com projetos de carros autônomos desenvolvidos no Laboratório de Robótica Móvel (LRM[1]) do Instituto de Ciências Matemáticas e de Computação (ICMC), ligado ao Centro de Robótica da USP de São Carlos (CRob), e ao Instituto Nacional de Ciência e Tecnologia para Sistemas Autônomos Cooperativos (InSAC) [2]. No entanto, podendo também dar suporte para pesquisas de carros autônomos desenvolvidos por outros grupos de pesquisas do Brasil e do mundo.

As contribuições geradas nesta pesquisa tiveram suporte e, também, são direcionadas em prol do projeto CaRINA[3]. Em desenvolvimento desde 2010, o projeto conta atualmente com três veículos autônomos (Figura 2), aumentando então o nível de percepção de cada um destes veículos em relação ao sinais de trânsito verticais.

Figura 2 – Veículos robóticos inteligentes utilizados no projeto CaRINA.

Fonte: Adaptada de (CARINA, 2020).

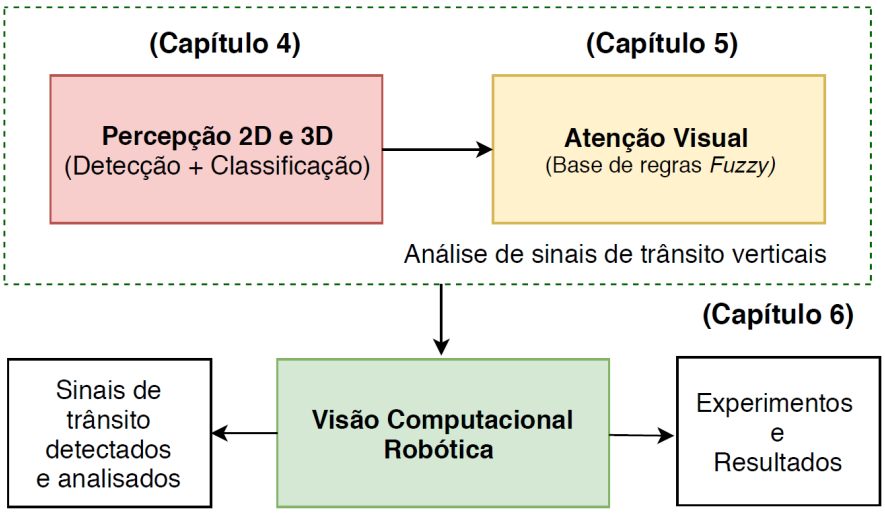
[1] <http://www.lrm.icmc.usp.br/>
[2] <http://www2.eesc.usp.br/insac/>
[3] <http://lrm.icmc.usp.br/carina/>

1.9 Estrutura da Tese de Doutorado

Nesta tese de doutorado foi desenvolvido um modelo de Visão Computacional Robótica para detecção e análise de sinais de trânsito verticais. A estrutura principal do trabalho pode ser observada na Figura 3. A tese está organizada nas seguintes seções:

- Capítulo 2: são apresentados os trabalhos relacionados para detecção, classificação e análise de sinais de trânsito verticais;
- Capítulo 3: é apresentado o referencial teórico, envolvendo as técnidas, métodos e aplicações para Processamento Digital de Imagens, Visão Computacional e Aprendizado de Máquina, junto ao estado-da-arte para sistemas de perceção que foram aplicados neste trabalho de doutorado;
- Capítulo 4: são apresentados os métodos e técnicas que foram desenvolvidos e adaptados para detecção e classificação de sinais de trânsito verticais com fusão de dados $2D$ e $3D$;
- Capítulo 5: é apresentado o modelo de Atenção Visual baseado em lógica *Fuzzy* que foi desenvolvido para analisar a prioridade dos sinais de trânsito detectados em conflito, definindo seus níveis de prioridade e dando suporte para o sistema de tomada de decisão do veículo;
- Capítulo 6: são apresentados os experimentos, resultados e discussões que foram gerados ao longo da pesquisa desenvolvida nesta tese de doutorado. As conclusões e trabalhos futuros são apresentados no Capítulo 7.

Figura 3 – Estrutura do trabalho desenvolvido nesta tese de doutorado.



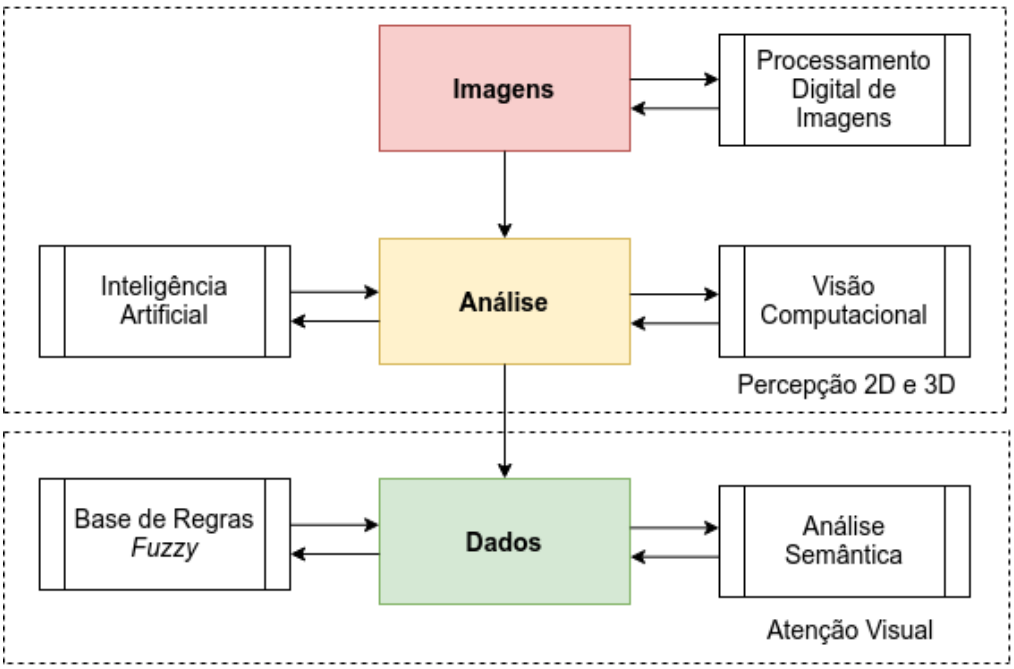
Fonte: Elaborada pelo autor.

1.10 Síntese do Trabalho Desenvolvido

Neste trabalho, existem vários processos para que uma imagem capturada chegue em sua etapa final, envolvendo detecção, classificação e análise de prioridades dos sinais de trânsito verticais. Para uma melhor leitura desta tese, estes processos são divididos em dois grandes grupos e com as seguintes funções:

- **Percepção 2D e 3D**: Nesta etapa, primeiramente é aplicado o Processamento Digital de Imagens sobre os dados provenientes do sensoriamento $2D$ e $3D$. Em seguida, métodos de Visão Computacional e Inteligência Artificial são aplicados para detectar e classificar os sinais de trânsito;
- **Atenção Visual**: Por meio da detecção e classificação dos sinais de trânsito, os dados de percepção são gerados e, sobre estas informações, uma análise semântica em conjunto com uma base de regras *Fuzzy* bem definida, são aplicadas para analisar a prioridade de cada sinal detectado.

Figura 4 – Síntese do sistema desenvolvido.



Fonte: Elaborada pelo autor.

CAPÍTULO

2

TRABALHOS RELACIONADOS

Nesta seção, os trabalhos relacionados com a pesquisa desenvolvida nesta tese de doutorado serão apresentados. As pesquisas destacadas estão junto ao estado-da-arte de sistemas de percepção e Atenção Visual para veículos autônomos, contribuindo significativamente com o avanço científico da área.

Por meio de uma revisão bibliográfica sistemática na área de percepção e Atenção Visual para a detecção, classificação e análise de sinais de trânsito verticais com fusão de imagens $2D$ e $3D$, foi constatado que o tema deste projeto de pesquisa de doutorado possui publicações relacionadas bem recentes (2014, 2015, 2016, 2017, 2018, 2019 e 2020).

Dentre os trabalhos mais recentes, poucos apresentam modelos de percepção com fusão de dados $2D$ e $3D$. Também não foi encontrado nenhum trabalho de Atenção Visual que faça análise de prioridades dos sinais de trânsito detectados e, também, em relação a área navegável. Isto indica que este projeto pode contribuir cientificamente com esta área, gerando grandes avanços em uma boa direção e junto ao estado-da-arte em termos de análise de sinais de trânsito verticais. Também foi feita uma revisão bibliográfica sobre Sistemas Avançados de Assistência ao Condutor (*Advanced Driver Assistance System* - ADAS), visando encontrar modelos de Atenção Visual para percepção de sinais de trânsito que podem auxíliar o motorista na tarefa de dirigir. Por último, são destacadas algumas das principais lacunas relacionadas com trabalhos publicados por outros autores desta mesma área de pesquisa.

2.1 Sistemas de Percepção para Detecção e Classificação de Sinais de Trânsito Verticais

Sistemas de percepção baseados em modelos de Visão Computacional, envolvem tecnologias de grande importância para dar suporte junto a navegação de veículos robóticos inteligentes. Estes sistemas são baseados em técnicas de Processamento Digial de Imagens (Timofte; Zimmer-

mann; Gool, 2009a); (SALTI *et al.*, 2015) ; (Zhu *et al.*, 2016), Visão Computacional (CHEN *et al.*, 2012) ; (Marinas *et al.*, 2012); (BALALI; Golparvar Fard, 2015) e Aprendizado de Máquina (HUVAL *et al.*, 2015); (Zeng *et al.*, 2017), possibilitando auxiliar o sistema de navegação autônomo, ou semi-autônomo, na tarefa de detectar, classificar e analisar semanticamente informações de trânsito informadas por meio de sinais de trânsito verticais.

Por meio da realização de uma revisão bibliográfica na área de detecção e classificação de sinais de trânsito verticais, foram destacados os principais autores com seus respectivos trabalhos levantados neste estudo. Nesta seção (2.1.1), estes trabalhos são apresentados de maneira que a evolução nessa área de pesquisa seja entendida de uma forma completa:

2.1.1 Percepção com fusão de imagens 2D e 3D

São apresentados nesta seção, alguns trabalhos que utilizam técnicas do estado-da-arte de percepção com dados provenientes da fusão de imagens $2D$ e $3D$.

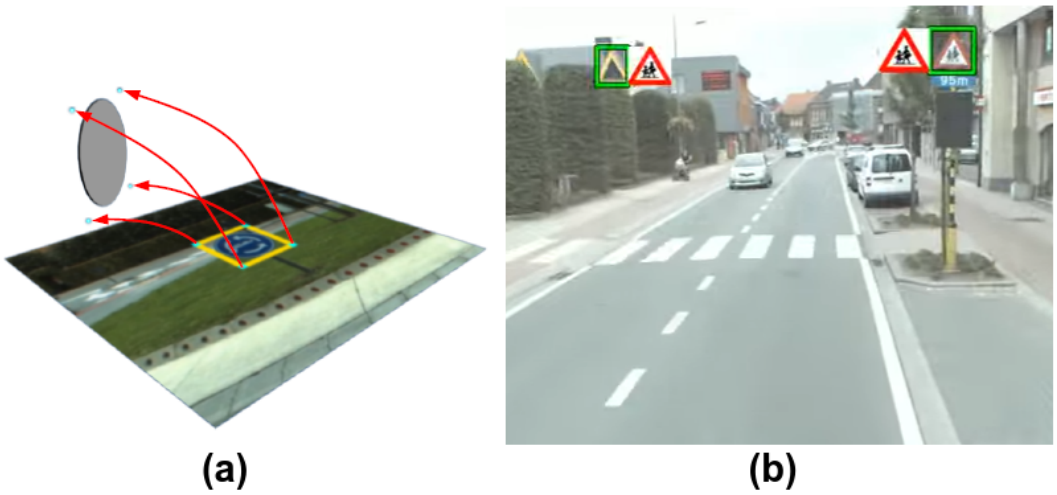
Em um trabalho feito por (Timofte; Zimmermann; Gool, 2009a) em 2014, foi apresentado um sistema que visa a detecção e reconhecimento de sinais de trânsito verticais em prol ao reconhecimento das regras de trânsito locais informadas por meio de placas. Neste trabalho, foi desenvolvido um método de reconhecimento de placas baseado em análise de imagens $3D$, utilizando uma técnica intitulada de *Minimum Description Length principle* (MDL). Este foi um dos primeiros trabalhos realizados utilizando imagens $3D$ para detecção de sinais de trânsito verticais (Timofte; Zimmermann; Gool, 2009a).

O sistema de Timofte e Zimmermann (Timofte; Zimmermann; Gool, 2009a), utiliza um mecanismo de detecção $2D$ baseado em caixas delimitadoras (*Slide Window*). Por meio dos sinais detectados, são feitas duas comparações puramente geométricas entre os dados $2D$ e $3D$ ($RGB + D$), possibilitando classificar em três classes, cada tipo de forma dos sinais: regulamentação, advertência ou de indicação auxiliar. Por meio do método de comparação geométrica, também é feito o processamento e análise de características para verificar se o objeto é um sinal de trânsito real com base no seu formato $3D$ (Timofte; Zimmermann; Gool, 2009a).

Para que a avaliação $3D$ dos sinais de trânsito verticais seja aplicada, são gerados pontos de referência na imagem capturada (*Bounding Box*), assim, criando a caixa delimitadora sobre a forma geométrica $3D$ equivalente da sinalização a ser analisada (Timofte; Zimmermann; Gool, 2009a). Possibilitando então analisar o objeto placa em sua geometria $3D$ com o auxílio de um algoritmo de recuperação planar aplicado na nuvem de pontos (Figura 5-(a)). Por fim, utilizando também dados $2D$ para a classificação do tipo do sinal de trânsito detectado (Limite de velocidade, Parada obrigatória, etc) (Mathias *et al.*, 2013).

Porém, quando uma análise semântica da cena não é feita, alguns erros são gerados. Na Figura 5-(b) pode ser observada uma situação que o telhado da casa (triangulo amarelo) foi detectado como um sinal de trânsito, declarando um sinal aleatório em sua classificação, podendo gerar grandes erros deste tipo para o sistema de percepção.

Figura 5 – Detecção de sinais de trânsito via análise de geometria (a) pontos correspondentes na detecção da região de interesse e (b) sinais detectados pelo sistema com fusão de dados $2D$ e $3D$.



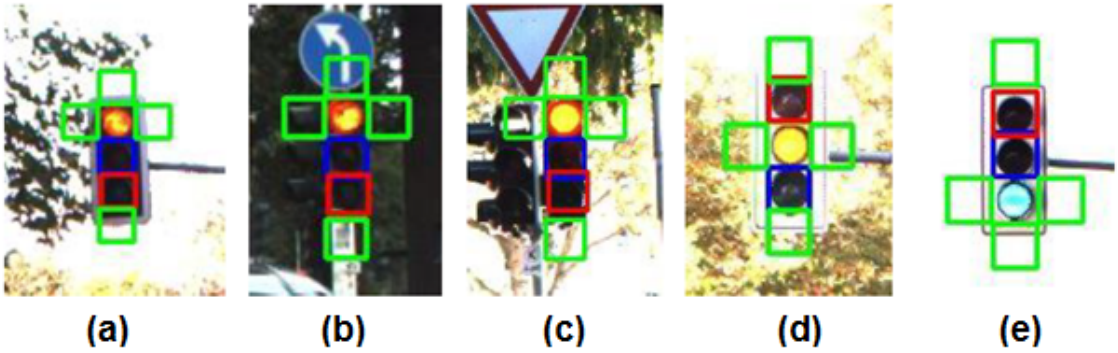
Fonte: Adaptada de (Mathias *et al.*, 2013).

Um outro ponto que foi tratado no trabalho desenvolvido nesta tese, está relacionado ao reconhecimento das informações contidas nos sinais de trânsito verticais por meio de redes de *Deep Learning*, sendo estas redes especialistas em classificação de imagens $2D$. Sendo que a classificação do tipo do sinal de trânsito é muito importante para navegar dentro das regras de trânsito locais. Ao contrário do objetivo do trabalho de (Mathias *et al.*, 2013), que focou apenas em uma boa detecção. Porém, Timofte e Zimmermann (Mathias *et al.*, 2013) tem contribuído fortemente para esta área de pesquisa nos últimos anos (Timofte; Zimmermann; Gool, 2009a); (TIMOFTE *et al.*, 2011)e (Mathias *et al.*, 2013).

Em um trabalho feito por (SALTI *et al.*, 2015), foi desenvolvido um sistema para a detecção dos semáforos de trânsito utilizando a tranformada de Hough. O objetivo principal deste trabalho, está ligado a um sistema de visão capaz de detectar os semáforos com base em suas formas em imagens $2D$. O sistema visa principalmente a identificação de qual cor está acionada em um semáforo, auxiliando o veículo em sua navegação.

No entanto, o foco deste trabalho está relacionado ao reconhecimento de informações de semáforos por meio da detecção de regiões de interesse em imagens $2D$ (SALTI *et al.*, 2015) e não se aplica em sinais de trânsito informados por meio de placas.

Figura 6 – Semáforos detectados pelo filtro proposto.



Fonte: Adaptada de (SALTI *et al.*, 2015).

O método de (SALTI *et al.*, 2015) utiliza a transformada de Hough em conjunto com um filtro *RGB* para cada uma das três cores que serão detectadas em uma semáforo de trânsito, assim, sendo possível identificar qual destas está acionada (Figura 6). Para que isso seja possível, é buscada a cor predominante, aplicando sobre cada uma das posições (vermelho, amarelo e verde) o filtro *RGB*. Já para as outras cores não ativas, são geradas duas caixas auxiliares (azul e vermelha), possibilitando então uma confirmação de que as três cores do semáforo foram detectadas e, possivelmente a cor acionada, foi processada corretamente. Nas Figuras 6-(a), (b) e (c) podem ser observadas três situações em que os semáforos com as cores vermelhas estão acionadas e foram detectadas corretamente. A caixa delimitadora com abas laterais está sobre a cor vermelha nestas três situações, declarando o estado atual do semáforo. Já nas Figuras 6-(a) e (b), podem ser observadas respectivamente as cores amarela e verde sendo declaradas como acionadas.

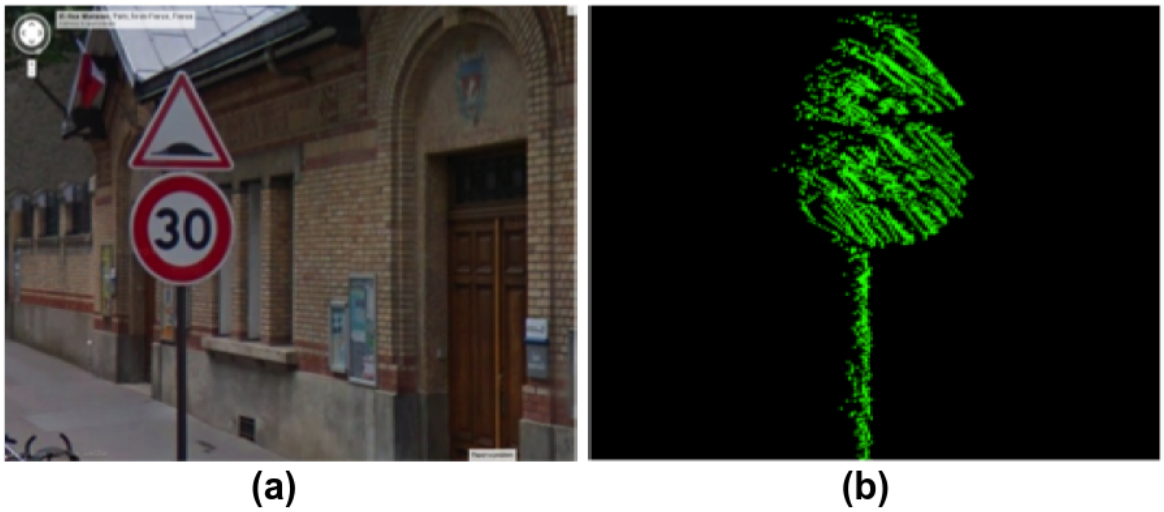
Em um trabalho feito por (Zhou; Deng, 2014), foi desenvolvido um sistema baseado em *Light Detection And Ranging* (LIDAR) e algoritmos aplicados em classificação de objetos em nuvem de pontos em beneficio da análise de imagens de sinalizações. A tecnologia do sensoriamento LIDAR é parecida com o sensoriamento ultrassônico, no entanto, utilizando feixes direcionados de luz (*laser*) como sinal ativo. O sistema visa processar imagens $3D$ com o objetivo de se conseguir uma maior robustez na detecção de sinais de trânsito verticais. Por meio de dados $3D$ da nuvem de pontos (aglutinação de pontos - *point-cloud*), são detectados todos os possíveis sinais. O algoritmo que analisa as características de cada sinal de trânsito vertical detectado é baseando em uma *Support Vector Machine* (SVM), sendo então aplicada como um classificador (Zhou; Deng, 2014).

Para que seja possível um melhor entendimento do modelo de visão via análise de objetos $3D$ feito por (Zhou; Deng, 2014), é apresentado um exemplo de sensoriamento em nuvem de pontos (*point-cloud*) na Figura 7-(a). Este método de segmentação utiliza a aglutinação de pontos

para gerar a imagem 3*D*, possibilitando analisar os sinais de trânsito com base em suas formas. Esta técnica de captura de imagens é utilizada também pelo sistema *Kinect* da Microsoft (HAN *et al.*, 2013), possibilitando capturar imagens 3*D* em nuvem de pontos e analisar o comportamento do jogador em um jogo de movimentos corporais.

Na Figura 7-(a), uma imagem 2*D* contendo dois sinais de trânsito (Velocidade máxima e Lombada) é capturada pelo sistema de sensoriamento em câmera monocular do Google *Street View*. Já na Figura 7-(b), uma nova imagem 3*D* é gerada por meio de uma nuvem de pontos (*point-cloud*) que detectou os sinais equivalentes (AIJAZI *et al.*, 2016) e as segmentou. Pode ser observada nesta situação (Figura 7-(b)), os sinais segmentados em dados de LIDAR. Isto possibilita detectar as regiões de interesse em profundidade, representando melhor as informações dos sinais de trânsito. Já o fundo da imagem (janelas, portas e paredes) é eliminado graças a noção de profundidade da cena, ficando todo preto.

Figura 7 – Detecção de sinais de trânsito (a) imagem via Google *Street View* com duas placas de sinalização e (b) imagem segmentada via nuvem de pontos.



Fonte: Adaptada de (AIJAZI *et al.*, 2016).

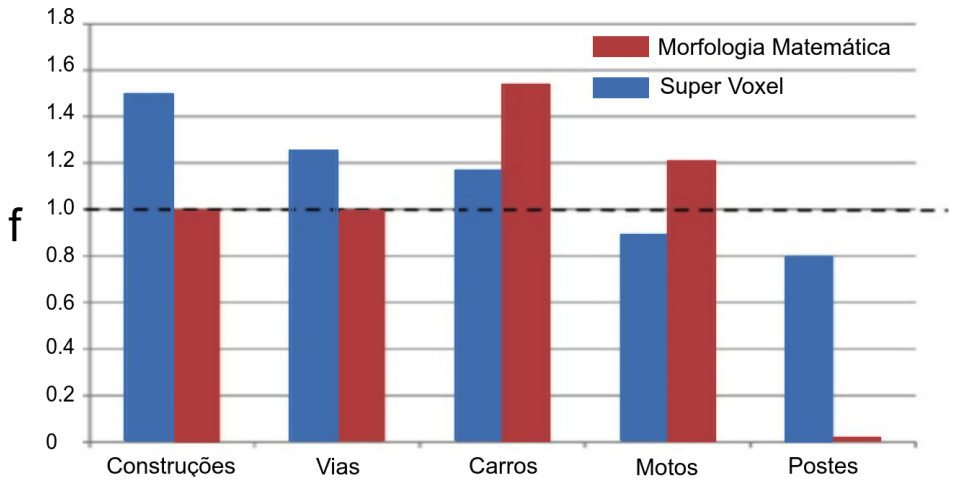
Ainda se tratando do trabalho desenvolvido por (AIJAZI *et al.*, 2016), foram implementadas e comparadas duas técnicas para segmentação de imagens em nuvem de pontos, sendo estas: (1) segmentação por *super-voxel*: segmentação por grupos de *voxel*, transformando em *super-voxel* e (2) segmentação morfológica: segmentação por *watershed* e classificação por SVM. Este trabalho focou no reconhecimento de objetos genéricos (árvores, postes, construções e placas) e não somente sinais de trânsito. No entanto, o estudo mostrou que para a deteção de sinais de trânsito em postes seria melhor a segmentação por *super-voxel*.

Um problema encontrado no método de segmentação por *super-voxel*, está ligado a detecção de objetos diferentes em uma mesma região de interesse. Isso acontece pelo motivo de que as regiões adjacentes entre estes mesmos objetos gera uma similaridade entre valores

de intensidade e refletância. Contudo, o método de visão utilizado por (AIJAZI *et al.*, 2016) e, que utiliza agrupamentos de *voxels* para segmentação e classificação utilizando descritores geométricos, pode ser inviável para detecção de sinais de trânsito verticais, já que normalmente mais de um sinal de trânsito deve ser detectado e analisado ao mesmo tempo.

Para exemplificar esta situação, no caso da Figura 7-(b), as placas agrupadas foram declaradas como sendo uma pequena árvore, em função deste tipo de problema, seria inviável aplicar este algoritmo para o reconhecimento de sinais de trânsito. Porém, este método foi bem aceito para o reconhecimento de construções, veículos e área navegável (Figura 7).

Figura 8 – Comparação dos métodos de segmentação (1) morfológica e (2) *super voxel*.



Fonte: Adaptada de (AIJAZI *et al.*, 2016).

No método de segmentação por morfologia matemática de (AIJAZI *et al.*, 2016), são utilizados apenas imagens do LIDAR $3D$. Seria viável uma fusão destas imagens com informações de cores, formas e texturas $2D$, gerando uma maior precisão na avaliação de cada objeto de trânsito. Uma comparação entre segmentação por morfologia matemática e *super-voxel* é também apresentada (Figura 8), destacando que a técnica de *super-voxel* é melhor para detecção de sinais de trânsito verticais em postes. Onde F é uma medida de qualidade de reconhecimento e quanto maior o valor de F, maior o acerto da segmentação.

O método faz com que cada item da cena seja *clusterizado* separadamente, o solo (ruas, estradas e outros caminhos) é segmentado primeiramente e depois os objetos verticais (árvores, postes e placas). Para que isso seja possível, é utilizada uma técnica de segmentação baseada em *watershed* para separar os objetos (AIJAZI *et al.*, 2016). Para que cada objeto da cena seja classificado, é utilizada uma SVM. O método de segmentação morfológica apresentou melhores resultados para os objetos grandes: carros, motocicletas e construções (AIJAZI *et al.*, 2016).

Em um trabalho feito por (BALALI; Golparvar Fard, 2015), também foi desenvolvido um método de detecção e classificação de sinais de trânsito verticais baseado na fusão de imagens

2D e 3D. O foco do trabalho foi aplicado no desenvolvimento de um método automático para detectar e classificar os sinais de trânsito. O método de detecção utiliza o algoritmo *Sift* para detectar os sinais em imagens 2D e aplica uma análise em 3D para filtrar as formas. Já o método de classificação proposto é baseado em uma SVM que trabalha com histogramas orientados a cores (BALALI; Golparvar Fard, 2015). Diferente do método de (AIJAZI *et al.*, 2016), este modelo é aplicado também no reconhecimento de sinais de trânsito e não somente na detecção.

Em um trabalho feito por (WANG *et al.*, 2014), foi desenvolvido um sistema de detecção $RGB+D$ (*Red, Green, Blue and Depth*), visando a detecção de faixas de pedestres e de trafego por meio da fusão de câmeras 3D e LIDAR. O maior objetivo deste trabalho foi desenvolver um sistema capaz de auxiliar pessoas com deficiência visual para quando estão se locomovendo em ambientes desconhecidos (WANG *et al.*, 2014). Este sistema não foi aplicado em veículos e, sim, em pedestres, no entanto pode sofrer adaptações para contribuir com aplicações de veículos robóticos envolvendo detecção de sinais de trânsito horizontais.

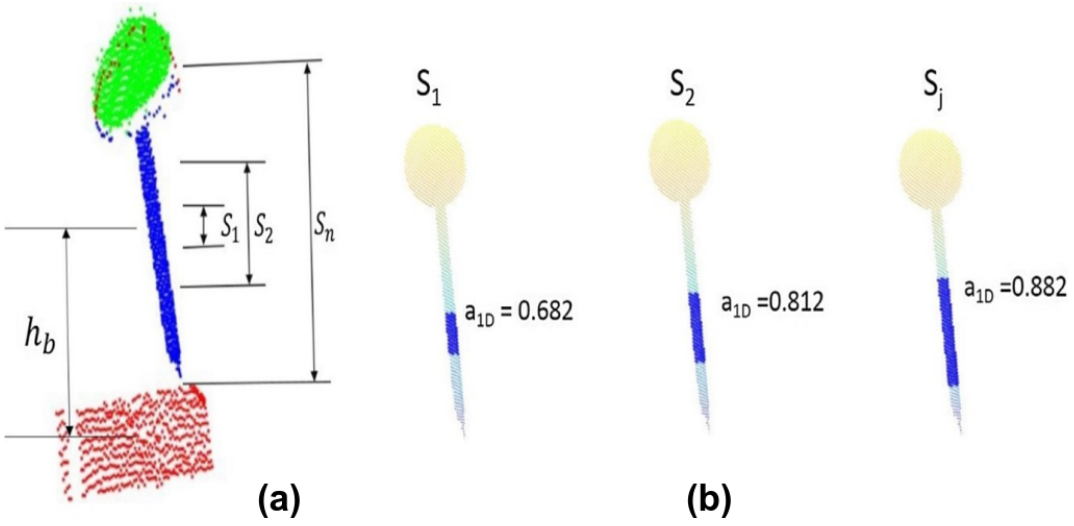
Em um trabalho feito por (Wu *et al.*, 2015), também foi desenvolvido um sistema baseado em LIDAR. Este sistema diferencia-se dos outros pelo fato de que toda análise da cena é feita por meio do sistema de percepção 3D e, para que isso seja possível, utiliza pontos de referência para auxiliar na detecção dos sinais de trânsito. O objetivo de se trabalhar com análise 3D "pura" é conseguir um maior nível de robustez com as variações de iluminação da cena (Wu *et al.*, 2015). O método utiliza a análise de pontos de referência geográficos. Estes pontos de referência tem ligações com um mapa de GPS fornecido *a priori*. O mapa auxilia o sistema de visão com os possíveis lugares que podem ser encontradas as sinalizações de trânsito verticais (Wu *et al.*, 2015) e que estão relacionados com cruzamentos, curvas e outras situações.

Em um trabalho desenvolvido por (Soilan *et al.*, 2015), também é utilizado processamento de imagens 3D via nuvem de pontos para detectar sinais de trânsito verticais. Cada sinal de trânsito é detectado automaticamente utilizando a técnica de sensoriamento de LIDAR em conjunto com algoritmos baseados em análise geométrica aplicados como detectores (Soilan *et al.*, 2015). A classificação 3D dos sinais de trânsito é feita via SVM aplicada nos dados extraídos do objeto. Na Figura 9, pode ser observada a reconstrução da superfície do sinal de trânsito feita por meio da nuvem de pontos. Dada esta região bem definida, é onde serão aplicados os algoritmos de extração de características 3D.

Um ponto de extrema importância neste trabalho, está ligado a detecção de sinais de trânsito verticais via análise de localização espacial (Soilan *et al.*, 2015). Para que isso seja possível, é esperado que os sinais de trânsito estejam em um poste próprio para placas, poste de luz, ou até mesmo, em uma árvore, possibilitando fazer uma estimativa de onde estas informações devem ser detectadas com maior frequência e baseando-se na sua distância em relação ao veículo e, também, da área navegável (Figura 9).

Para a análise de localização espacial, é utilizado um coeficiente (S) para que seja declarada uma possível localização do sinal de trânsito vertical em meio a cena. Para um fator S

Figura 9 – Análise de sinais de trânsito via características de altura (a) reconstrução via nuvem de pontos e (b) avaliação para cada segmento de *Si*.



Fonte: Adaptada de (Soilan *et al.*, 2015).

que foi obtido via análise do conjunto de seguimentos ($S1$, $S2...Sj$) maior que $0,5$, o objeto é declarado ser um possível sinal de trânsito estando dentro da região de interesse e será feita a extração de características em prol do classificador $3D$ (Soilan *et al.*, 2015).

Para que seja possível a análise geométrica espacial são utilizadas algumas variáveis que possibilitam verificar se este objeto é realmente um sinal de trânsito vertical:

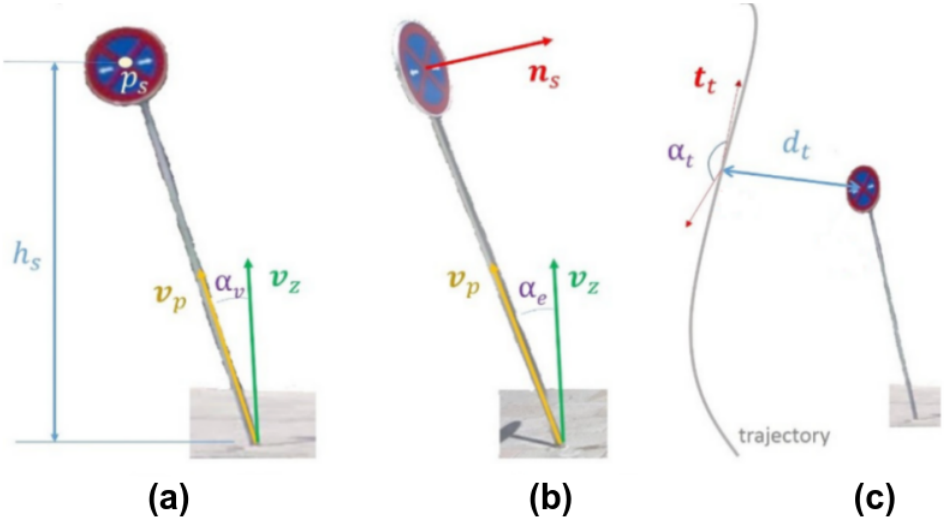
- **Posição da placa** *(Ps)*: sendo este o centróide para o processamento da imagem;
- **Altura da placa** *(hs)* : sendo esta a diferença de altura entre *Ps* e a base do sinal;
- **Distância entre a trajetória do veículo e o sinal de trânsito** (dt): é a distância Euclidiana mínima entre a posição do sinal e a rota do veículo;
- V_p **e** V_z: são os eixos para se calcular o ângulo da inclinação do sinal em sua base.

Por meio destas variáveis é feita a detecção do sinal, logo a seguir é feita uma análise geométrica sobre a detecção do objeto $3D$, facilitando no processo de eliminação de falsos positivos e falsos negativos (Soilan *et al.*, 2015). As características dos objetos $3D$ serão extraídas apenas quando o sistema de percepção baseado em localização declarar que o objeto é um possível sinal de trânsito.

Outro fator de grande importância a ser considerado, é que neste trabalho (Soilan *et al.*, 2015) é feita também uma estimativa da distância (dt) que um sinal de trânsito vertical pode ser detectado em relação ao veículo (Figura 10-(c)), eliminando informações de trânsito de longas

distâncias irrelevantes para o ponto atual de navegação do veículo, dispensáveis temporariamente para uma navegação segura. O trabalho de (Soilan *et al.*, 2015) mostrou ser de grande relevância para aplicações de detecção de sinais de trânsito verticais. Pois utilizou um método com grande potencial de análise geométrica e localização espacial com vistas para localização de sinais de trânsito, possibilitando então diferenciar sinais de trânsito verticais de outros objetos da cena, graças as relações de distâncias e que são muito úteis para o sistema de percepção eliminar falsos positivos e falsos negativos.

Figura 10 – Parâmetros geométricos calculados para sinais de trânsito verticais informados por meio de placas (a) vista frontal de uma placa (b) vista lateral de uma placa e (c) visão de uma trajetória em conjunto com o sinal de trânsito.



Fonte: Adaptada de (Soilan *et al.*, 2015).

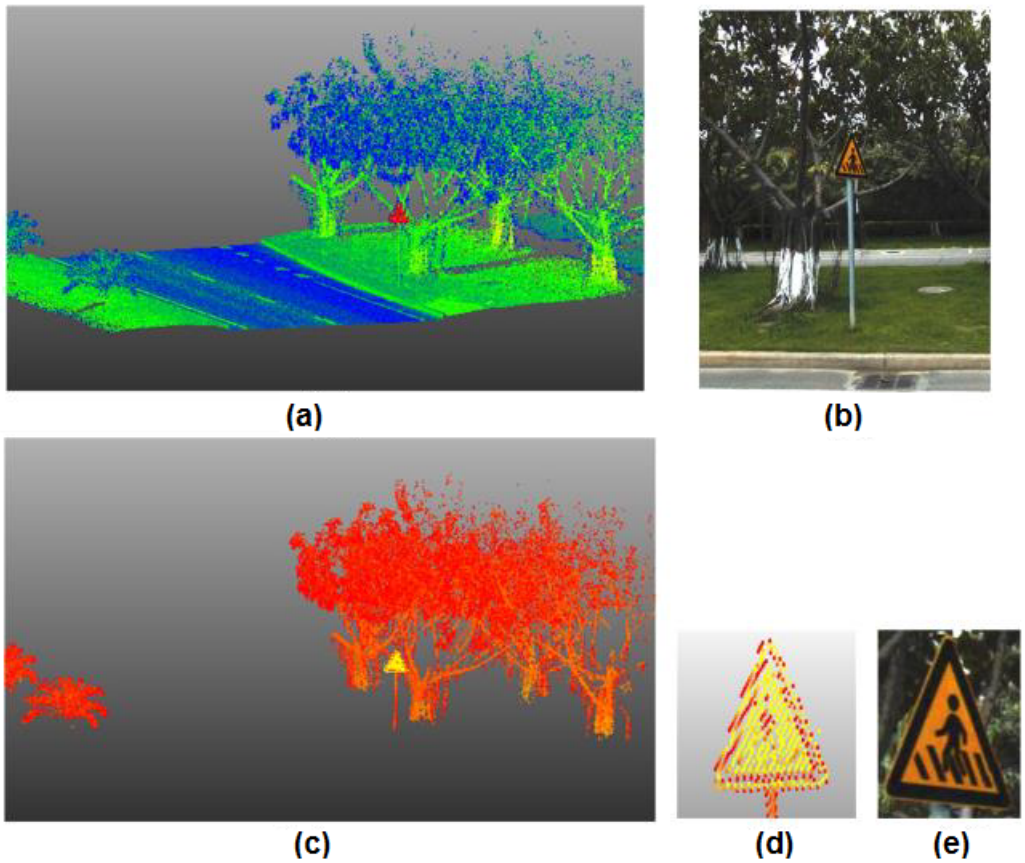
O trabalho de (Wu *et al.*, 2015) utilizou o mesmo princípio do trabalho desenvolvido por (Soilan *et al.*, 2015), assim, possibilitando localizar sinais de trânsito verticais com uma maior precisão em cada cena. Por meio da Figura 11, podem ser observadas as etapas para a detecção e análise dos sinais de trânsito.

Na Figura 11-(a), pode ser observada a cena completa capturada por meio do sensoriamento LIDAR gerando a nuvem de pontos $3D$. Também pode ser observado nesta imagem, que o sinal de trânsito (Figura 11-(b)) foi detectado e destacado em uma outra cor, pois este mesmo já foi declarado como um possível sinal de trânsito vertical com base em sua forma e posição na cena. Isso foi possível graças ao sistema de georreferenciamento que gera uma estimativa da localização dos sinais de trânsito em relação da área navegável do veículo (Wu *et al.*, 2015). Já na Figura 11-(c), pode ser visualizada a aplicação de um filtro que eliminou a superfície da estrada, garantindo então uma redução da carga de dados irrelevantes do LIDAR, facilitando no processo de extração de características de cada sinal de trânsito que está acima da via.

Ainda no trabalho de (Wu *et al.*, 2015), nas Figuras 11-(d) e (e), podem ser observados, respectivamente, o sinal de trânsito segmentado na imagem de nuvem de pontos $3D$ e o sinal equivalente na imagem $2D$. Pode ser notado que a região do sinal foi bem definida, porém, a informação contida nesta mesma ficou um tanto quanto precária na nuvem de pontos (Figura 11-(d)), tendo uma melhor representação na imagem $2D$ (Figura 11-(e)).

Uma boa classificação do tipo do sinal de trânsito deve ser obtida por meio da imagem $2D$ e que suporta dados de cores, texturas e formas. Já o método de segmentação utilizado para a nuvem de pontos $3D$, é baseado no cálculo das características de intensidade de luz e do contraste do histograma de cores, possibilitando detectar as bordas dos sinais informados por meio das placas que tem uma refletância diferenciada para os outros objetos da cena (árvores, carros e pessoas). Este método de segmentação foi aplicado nas imagens obtidas via LIDAR (Wu *et al.*, 2015) e os dados de imagens $2D$ foram obtidos por meio de uma câmera monocular.

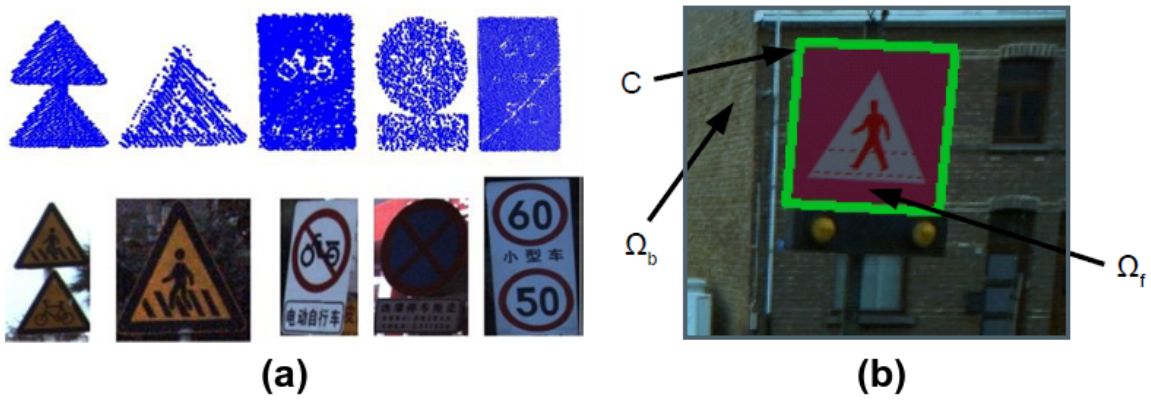
Figura 11 – Detecção de sinais de trânsito (a) sensoriamento LIDAR via nuvem de pontos, (b) imagem $2D$ real da placa, (c) superfície da estrada já eliminada, (d) sinal de trânsito segmentado via imagem $3D$ e (e) sinal $2D$ correspondente.



Fonte: Adaptada de (Wu *et al.*, 2015).

Por meio da Figura 12-(a), podem ser observadas as relações das superfícies $3D$ detectadas e que são referentes as informações dos sinais de trânsito, devendo ser classificadas posteriormente com base na região da caixa delimitadora C e sua imagem $2D$ equivalente (Figura 12-(b)). Por meio desta região de interesse (C), muitos trabalhos envolvendo percepção aplicam *Deep Learning* para classificação da imagem $2D$ (TIMOFTE *et al.*, 2011).

Figura 12 – Detecção de sinais de trânsito (a) superfícies $3D$ detectadas e (b) representação do sinal por meio da região C da placa.



Fonte: Adaptada de (TIMOFTE *et al.*, 2011).

2.1.2 Percepção com imagens 2D

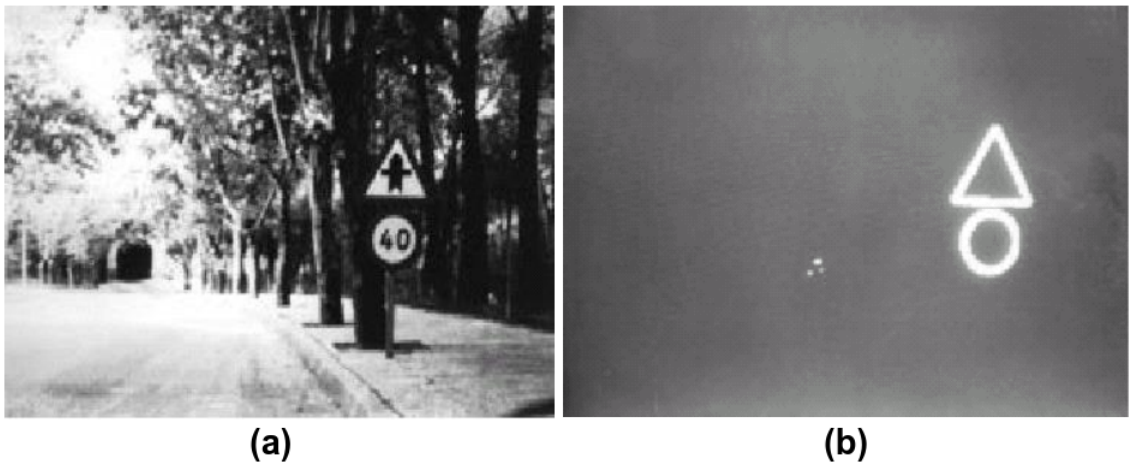
Nesta seção serão apresentados alguns trabalhos que focam suas técnicas para o estado-da-arte de percepção em dados provenientes de imagens puramente $2D$.

Em um trabalho feito por (Hossain; Hyder, 2015), foi desenvolvido um sistema capaz de detectar sinais de trânsito informados por meio de placas. Um dos principais objetivos deste trabalho foi aplicar um sistema automático $2D$ para detectar os sinais informados por meio de placas (Hossain; Hyder, 2015). Neste trabalho, o sistema de visão desenvolvido foi direcionado para detectar sinais de trânsito em ambientes com situações climáticas adversas e não controladas, focando principalmente em situações onde a iluminação é variável, sendo este, um dos grandes problemas para os sistemas de detecção e classificação de imagens em ambientes reais (Hossain; Hyder, 2015).

Para que fosse possível detectar os sinais de trânsito verticais em imagens $2D$, duas etapas de grande importância são aplicadas: (a) *Threshold* de cores (Figura 13-(a)) e (b) aplicação do filtro de bordas de Sobel (Figura 13-(b)). A mudança de intensidade de brilho e cor de um objeto em relação ao seu fundo é chamada de borda e, neste trabalho, o filtro de Sobel (Figura 13-(b)) foi aplicado para identificar as regiões de interesse por meio destas variações. Para facilitar este processo, a imagem real passa por um processo de limiarização (*threshold*), assim, fortificando a variação de intensidade nas bordas dos sinais de trânsito.

Depois de segmentada a região de interesse e, que possivelmente existem informações referentes a sinais de trânsito verticais, é aplicada a transformada de Hough para detectar as formas com maior destaque. O próximo passo é identificar quais informações foram detectadas. Porém, no trabalho de (Hossain; Hyder, 2015), não foi desenvolvido um método capaz de classificar as informações contidas em cada sinal de trânsito detectado. O método de detecção proposto apresentou simplicidade na tarefa de detecção, no entanto, gerando grande volume de falsos objetos detectados. Este problema é relacionado com a análise puramente $2D$ (sem dados de profundidade) e, também, pela detecção focada em formas geométricas, gerando detecções de outros objetos que não são sinais de trânsito.

Figura 13 – Conjunto de sinais de trânsito detectados (a) limiarização de cor e (b) sinais detectados com as bordas evidentes.



Fonte: (Hossain; Hyder, 2015).

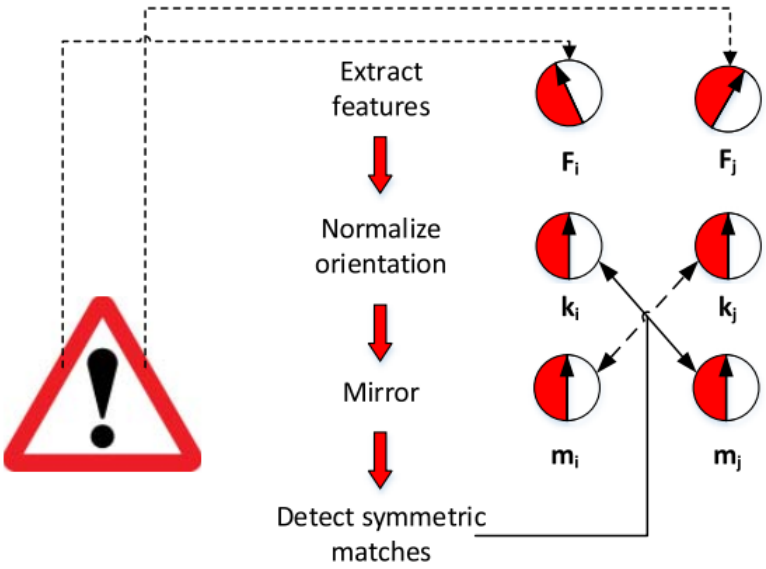
No trabalho de (Yuan; Xiong; Wang, 2017), foi desenvolvido um *Framework* para detecção e análise de sinais de trânsito verticais. Neste trabalho, é aplicado o filtro de Kalman para auxiliar na análise de regiões sobre cada *pixel* que representa um possível objeto. O filtro de Kalman, em conjunto com a transformada de Hough, trabalham em transformações de imagens $2D$ *pixel* a *pixel*, possibilitando então uma análise probabilística aplicada na vizinhança de *pixels* e suas variações de tons de cinza correspondentes. Por meio da análise probabilística, é possível identificar se um determinado *pixel* faz parte ou não de uma determinada aglutinação de *pixels* e que representa possivelmente o mesmo objeto.

Também é evidenciado neste trabalho, que a transformada genérica de Hough pode ser aplicada para detectar formas de interesse presentes nos sinais de trânsito, como exemplo: círculos, triângulos e retângulos (Yuan; Xiong; Wang, 2017). Também é destacado que esta transformada é mais robusta em relação a métodos de segmentação baseados em cores. Depois de feita a segmentação e a extração de características da região de interesse, é aplicado um par de classificadores para identificar as informações de cada sinal detectado. Os dois métodos

aplicados para o reconhecimento de sinais são formados por uma variação de SVM e Redes Neurais Artificiais (Yuan; Xiong; Wang, 2017).

Ainda no trabalho de (Yuan; Xiong; Wang, 2017), foi aplicado um sistema de aprendizado de máquina que funciona de maneira *on-line* analisando amostras detectadas no ambiente em tempo real. Para que o sistema seja capaz de reconhecer de maneira eficiente as amostras de sinais de trânsito detectadas, cadastrando apenas os sinais reais no banco de conhecimento, são consideradas as propriedades de simetria de cada placa. Isso é possível graças as placas de trânsito que são desenhadas com formas simétricas e regulares. Utilizando então a rotação invariante (Figura 14) para verificar a compatibilidade dos pares de simetria entre as características extraídas, possibilitando o cálculo de magnitude simétrica (Yuan; Xiong; Wang, 2017).

Figura 14 – Exemplo de cálculo de simetria de sinal de trânsito.



Fonte: (Yuan; Xiong; Wang, 2017).

Primeiramente é feita a extração de características dos sinais de trânsito informados por meio das placas, logo seguinte, é feita a normalização da orientação. Por meio da normalização é feito o "espelho" da imagem e, por meio deste, é feita uma avaliação das equivalências simétricas (Yuan; Xiong; Wang, 2017).

Resumidamente, para uma dada imagem de entrada, as características invariantes (Fi e Fj) são extraídas e normalizadas (Ki e Kj), facilitando o processo de classificação por meio dos seus "espelhos"(Mi e Mj). Este processo de orientação deve reduzir os erros gerados pelo sistema de classificação, verificando então a compatibilidade da simetria entre os pares de informações contidos em cada sinal de trânsito e que são informadas por meio de placas (Yuan; Xiong; Wang, 2017).

Este procedimento foi um grande diferencial do trabalho de (Yuan; Xiong; Wang, 2017) se comparado a outros trabalhos encontrados na literatura. Este diferencial deve contribuir no potencial das tarefas de detecção e classificação de sinais de trânsito, eliminando então falsos positivos e falsos negativos, consequentemente aumentando a acurácia. Este processo também contribui para os classificadores na tarefa de reconhecer os sinais com oclusão parcial, já que a simetria garante reconhecer a parte da informação oclusa por meio da parte evidente.

Em um trabalho de (Jung *et al.*, 2016), foi desenvolvido um sistema de detecção e reconhecimento de sinais de trânsito nomeado como *Traffic Sign Recognition* (TSR), sendo baseado em processamento de imagens em conjunto com um modelo de *Deep Learning*. O sistema também é capaz de reconhecer sinais de trânsito em imagens $2D$. Neste trabalho, o TSR é apresentado como sendo um ADAS, possibilitando reconhecer informações relevantes sobre um ambiente controlado por regras de trânsito locais, possivelmente contribuindo na redução de acidentes causados por distrações na tarefa de dirigir (Yuan; Xiong; Wang, 2017).

Neste trabalho, um conjunto com milhares de imagens com tipos diferentes de sinais de trânsito é armazenado em um *dataset* para o treinamento de um classificador que utiliza uma *Convolutional Neural Networks* (CNN) baseada no modelo *LeNet-5* (Yuan; Xiong; Wang, 2017). Sendo que este tipo de rede precisa de várias imagens para seu treinamento. Para um melhor entendimento deste modelo de rede pode ser consultado o trabalho de (LECUN; KAVUKCUOGLU; FARABET, 2010).

As redes de *Deep Learning* são bastante utilizadas na literatura para problemas de classificação de imagens digitais, tendo sua arquitetura fundamentada em uma rede *perceptron* de várias camadas. Neste trabalho, esta rede foi otimizada, reduzindo a quantidade de entradas de dados e, consequentemente, reduzindo também o número necessário de parâmetros para um bom aprendizado. Isso foi possível graças ao compartilhamento de pesos e recursos de entrada vindo de treinamentos anteriores para se criar os mapas de aprendizado (Yuan; Xiong; Wang, 2017). Essa técnica de aprendizado é bastante conhecida como *Transfer Learning*.

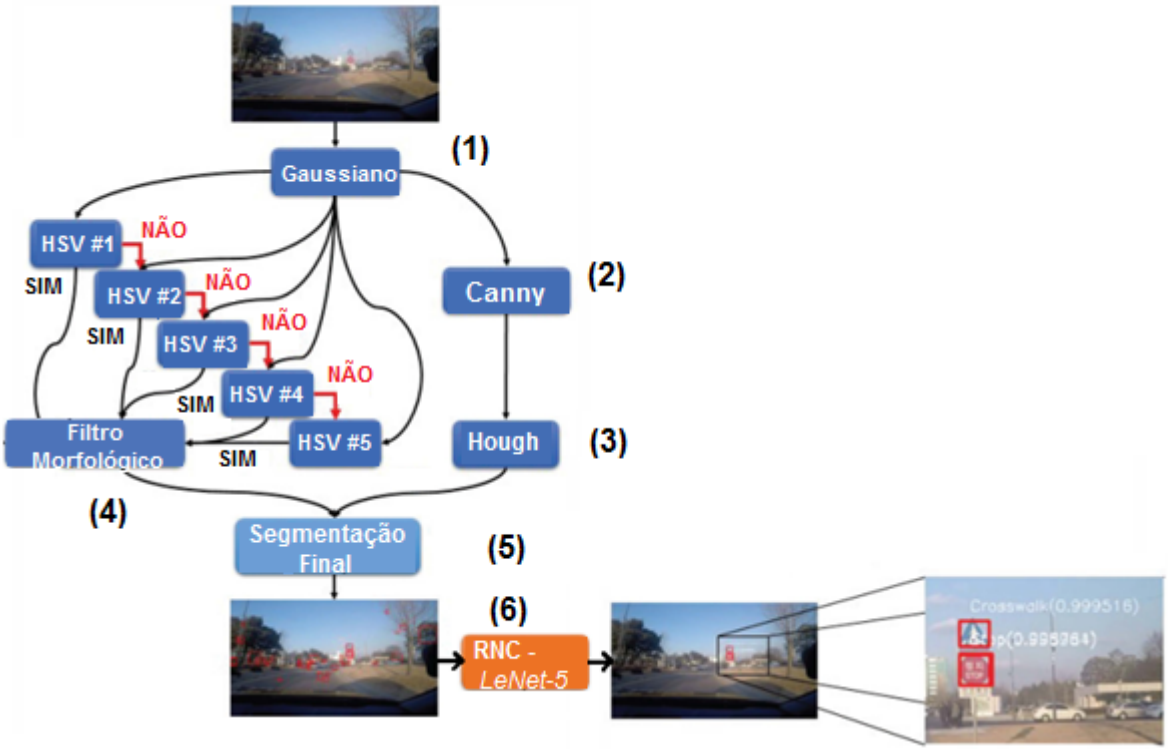
Para que seja possível tratar as imagens capturadas, inicialmente é utilizado um método de segmentação baseado em filtros de ruídos em conjunto com um filtro de bordas, eliminando interferências do sensoriamento e realçando as bordas dos elementos da cena para um detector de formas baseado na transformada de Hough, respectivamente. Esta transformada possibilita capturar as regiões de interesse por meio de suas formas (círculo, triângulo, quadrado e retângulo) e que representam as placas que contém as informações dos sinais de trânsito e também outros objetos de mesmo formato (Yuan; Xiong; Wang, 2017). Por meio da Figura 15, pode ser observado o processo feito pelo sistema de detecção e análise de sinais de trânsito verticais.

Como pode ser observado no fluxograma da Figura 15: (1) primeiramente a imagem passa por um filtro Gaussiano para uma redução do ruído do sensoriamento de imagens. (2) O próximo passo é aplicar na imagem filtrada um detector de bordas de Canny. Este detector de bordas deve salientar os contornos, contribuindo para a próxima etapa. (3) Nesta etapa, a transformada de

Hough é aplicada sobre a imagem pré-processada pelos dois filtros anteriores. Esta transformada deve reconhecer as formas dos sinais que estão sendo buscados. (4) Paralelamente a estes três passos, também é aplicado um filtro morfológico utilizando o padrão de cores de imagens *Hue, Saturation e Value* (HSV). Este quarto passo deve auxiliar na redução do ruído de iluminação e, também, no realce da imagem, trabalhando sobre as cores predominantes dos sinais informados por meio de placas. (5) Por final, é gerada uma segmentação com base em todos os outros filtros anteriores que salientaram as bordas com base em cores, eliminaram ruídos e detectaram as formas básicas. Possibilitando detectar as regiões de interesse que representam os sinais de trânsito verticais. Nestas regiões bem definidas e que já foram evidenciadas as suas características, é aplicada a (6) CNN para classificar os tipos das informações detectadas por meio dos sinais de trânsito verticais (Yuan; Xiong; Wang, 2017).

No trabalho (Yuan; Xiong; Wang, 2017), foi buscado um bom método de detecção de sinais utilizando como suporte os métodos de segmentação e análise de formas geométricas aplicados em dados $2D$. Por meio destas detecções e suas regiões de interesse, foi aplicada uma rede de *Deep Learning* para classificação do tipo da informação. No entanto, o trabalho gera os mesmos problemas apresentados anteriormente em pesquisas de outros autores, detectando formas geométricas equivalentes em outros objetos que não são sinais de trânsito verticais e, também, tendo problemas com variação de iluminação das imagens $2D$.

Figura 15 – Fluxograma do sistema TSR e seu resultado de reconhecimento de sinais de trânsito.



Fonte: Adaptada de (Yuan; Xiong; Wang, 2017).

Outros trabalhos relevantes também foram desenvolvidos em prol aos sistemas de percepção de sinais de trânsito verticais, sendo estes estudados no desenvolvimento desta revisão bibliográfica sobre o tema (CHEN *et al.*, 2012); (Marinas *et al.*, 2012); (Mogelmose; Trivedi; Moeslund, 2012) ; (Mogelmose; Trivedi; Moeslund, 2012) ; (MOUTARDE *et al.*, 2009); (Nunn; Kummert; Muller-Schneiders, 2008) ; (Timofte; Zimmermann; Gool, 2009b) e, também, outros trabalhos aplicados na área e que utilizaram Aprendizado de Máquina para o reconhecimento de sinalizações de trânsito (Chen *et al.*, 2015); (Cireşan *et al.*, 2011) ; (CIRESAN *et al.*, 2012) ; (Hossain; Hyder, 2015); (Houben *et al.*, 2013); (HUVAL *et al.*, 2015) ; (Li; Wen, 2016); (Li; Lv; Wang, 2016); (Sermanet; LeCun, 2011); (STALLKAMP *et al.*, 2012); (Zhu *et al.*, 2016); (Zeng *et al.*, 2017) e (WAN *et al.*, 2014). No entanto, os trabalhos detalhados anteriormente, foram aqueles considerados mais relevantes em relação a pesquisa que foi desenvolvida nesta tese.

O que pode ser notado com esta revisão bibliográfica, é que devido ao avanço dos métodos de *Deep Learning*, alguns trabalhos mais recentes continuaram a trabalhar e dar mais ênfase aos dados $2D$. Isso se deve ao fato de que com estas redes, é possível obter um alto poder de detecção e classificação de objetos neste tipo de informação. No entanto, tais abordagens baseadas em redes de *Deep Learning*, detectam um grande volume de informações falsas, principalmente por trabalharem com imagens puramente $2D$ sem noção de profundidade. Para exemplo, existem vários trabalhos envolvendo *Deep Learning* que utilizam um simples algoritmo de *slide window* para percorrer a cena toda e, para cada quadro, o método de Aprendizado de Máquina faz a classificação e retorna se encontrou um sinal de trânsito ou não e, também, qual o tipo da classe detectada. Sendo portanto este tipo de método, uma abordagem mais exaustiva de processamento das imagens e com menor grau de inteligência.

Nesta pesquisa de doutorado foram utilizadas redes de *Deep Learning* para detectar e classificar sinais de trânsito em imagens $2D$, no entanto, utilizando um filtro $3D$ para auxiliar na detecção e filtragem de falsas informações. Possibilitando então obter uma maior robustez e desempenho do sistema de percepção, eliminando estes problemas encontrados na detecção que utiliza apenas dados $2D$.

Para um melhor entendimento sobre o problema de detecção de falsos positivos, pode ser observado o exemplo do trabalho de (Mathias *et al.*, 2013), onde o sistema de detecção de sinais de trânsito verticais captura algumas formas geométricas de telhados triangulares (Figura 16-(a)) e janelas (Figura 16-(b)), e as confunde com as mesmas geometrias que são encontrados nos sinais de trânsito.

Por outro lado, um sistema utilizando imagens $3D$ pode ser capaz de ter um melhor nível de percepção, evitando estes falsos positivos com base em informações de profundidade e de dimensões reais dos sinais de trânsito em relação ao ambiente de navegação. Com estas propriedades aplicadas na relação dos sinais de trânsito em função do contexto da cena, é possível obter: (a) distância do sinal em relação a área navegável, (b) distância do sinal em relação ao veículo, (c) formato $3D$ do objeto e (d) análise de dados $2D$ e $3D$ equivalentes, aproveitando as

Figura 16 – Detecção de falsos positivos (a) detecção de telhado triangular e (b) detecção de vidraça da janela em formato de rombo.

Fonte: Adaptada de (Mathias *et al.*, 2013).

propriedades destes dois tipos de imagens.

Com este tipo de análise utilizando fusão de dados $2D$ e $3D$, é possível dar suporte para um modelo de visão robótica baseado em características de Atenção Visual, evitando então detectar situações falsas (Figura 16) e, também, possibilitando analisar a importância de cada informação detectada. Na próxima Seção (2.2), é apresentada a definição sobre modelos de Atenção Visual junto com alguns trabalhos relacionados.

2.2 Modelos de Atenção Visual

Por meio de uma revisão bibliográfica na área de Visão Computacional e junto ao tópico de Atenção Visual, foram encontrados alguns trabalhos de grande relevância para esta tese (FIGUEIRA A, 2019) ; (STALLKAMP *et al.*, 2012) ; (Schlosser; Montemerlo; Salisbury, 2010) e (Lee; Kim, 2018). Nesta seção (2.2), serão destacados alguns dos trabalhos de maior relevância.

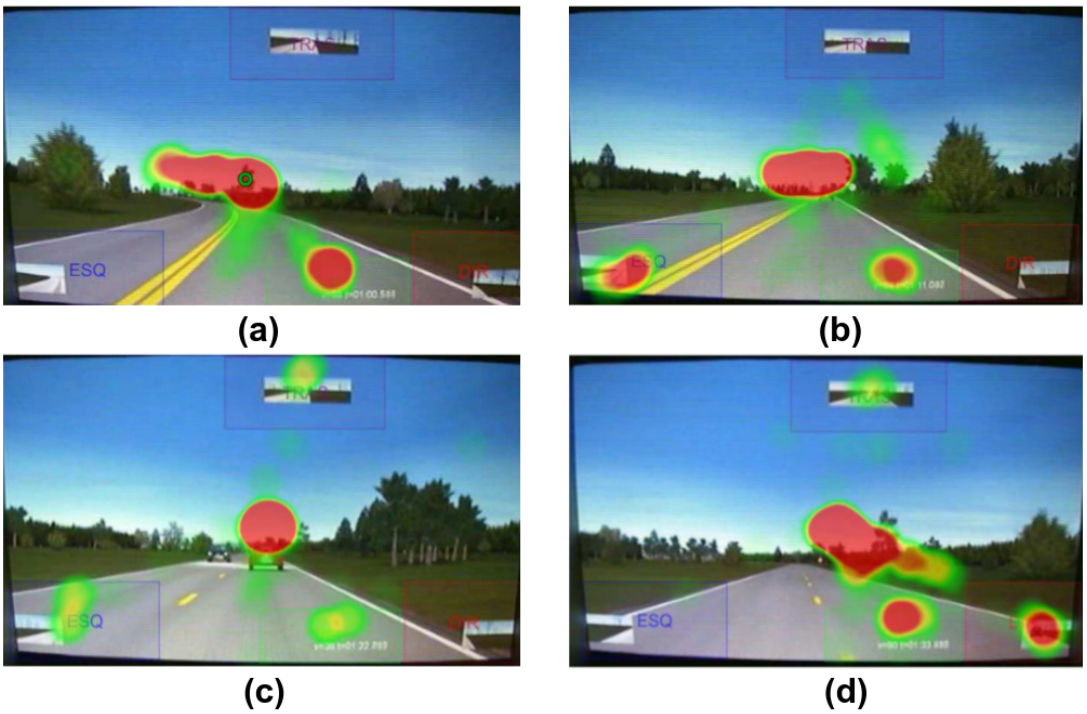
Para contextualizar melhor o problema de Atenção Visual, considera-se que essa tarefa está ligada com a capacidade que um sistema de visão (biológico ou artificial) em conjunto com seu sistema cognitivo, tem em focar seu campo de visão para situações de maior importância na cena observada. Essa capacidade é dada por meio da análise de relevância que cada região de observação tem para um dado instante de tempo. Para exemplificar este problema, o trabalho de (FIGUEIRA A, 2019) faz um estudo de um modelo de Atenção Visual biológico humano para análise de manobras de ultrapassagem.

A pesquisa de doutorado de (FIGUEIRA A, 2019) apresentou uma análise da Atenção Visual de um conjunto de motoristas em ambiente de simulação no momento da ultrapassagem

de veículos. O principal objetivo foi avaliar para onde o campo de visão do motorista está focado no momento deste tipo de manobra. Para que isso fosse possível, foi feito um sistema de rastreio do olhar por meio de técnicas de Processamento de Imagens Digitais, Visão Computacional e Aprendizado de Máquina, aplicadas em imagens geradas por uma câmera dentro do carro e posicionada de frente com o motorista.

Por meio da Figura 17, pode ser observada uma situação de ultrapassagem com o mapa de calor das cores relacionadas aos pontos de observação do motorista para a manobra. Na Figura 17-(a), pode ser observado o ponto inicial, onde o motorista começa a detectar a presença de um veículo lento a sua frente, sendo que todo seu campo de visão está focado na traseira do veículo que está impedindo a passagem e na faixa que possibilita a ultrapassagem. Já na Figura 17-(b), pode ser observado que o motorista olhou para os retrovisores direito e esquerdo (bolinhas nos cantos inferiores da tela) para verificar a presença de carros vindo por trás e impedindo a manobra de ultrapassagem.

Figura 17 – Atenção Visual para a situação de ultrapassagem em rodovias de faixa única: (a) momento em que um carro é detectado a frente, (b) início da verificação de possibilidade de manobra, (c) início da manobra de ultrapassagem e (d) final da manobra de ultrapassam.



Fonte: (FIGUEIRA A, 2019).

Na Figura 17-(c), foi verificado além dos retrovisores externos laterais, o retrovisor interno, assim, garantindo um maior campo de visão para a manobra. Por fim, na Figura 17-(d), o motorista faz a ultrapassagem mantendo seu foco de visão principalmente no velocímetro (ponto no centro da tela), na pista livre e levemente aos retrovisores, sendo este o momento de maior

nível de perigo e que exige uma maior atenção envolvendo múltiplos focos de visão (FIGUEIRA A, 2019).

Este exemplo destaca bem como funciona a Atenção Visual humana em seu sistema de percepção. Graças a essa inspiração da visão humana, o trabalho apresentado nessa tese de doutorado foi desenvolvido, buscando sempre uma boa aproximação a este modelo biológico por meio de modelos de sistemas artificiais, envolvendo: Aprendizado de máquina, Redes Neurais Artificiais, *Deep Learning* e Lógica *Fuzzy*. A seguir são apresentados alguns trabalhos que seguem algumas pequenas características de Atenção Visual, porém, ainda muito distantes de uma análise robusta da cena como mostrado no modelo de visão biológico estudado por (FIGUEIRA A, 2019). No entando, serviram de base para esta pesquisa de doutorado.

Alguns pesquisadores também aplicam características de Atenção Visual humana para detectar sinais de trânsito em diferentes situações. O trabalho de (STALLKAMP *et al.*, 2012) apresenta alguns dados relevantes para o estudo de modelos de Atenção Visual bioinspirados.

No trabalho desenvolvido por (STALLKAMP *et al.*, 2012), foi realizado um estudo comparativo entre a parcepção para sinais de trânsito por humano e computador. Neste trabalho, o foco principal foi comparar a capacidade cognitiva de algoritmos de Aprendizado de Máquina e humana (STALLKAMP *et al.*, 2012). Para que fosse possível uma potencial comparação, foram utilizadas variáveis ambientais que afetam um sistema artificial e que estão presentes em situações reais: iluminação, condições climáticas e oclusões parciais (STALLKAMP *et al.*, 2012). Sendo que estas variáveis geram problemas para um humano em sua tarefa de reconhecer sinais de trânsito.

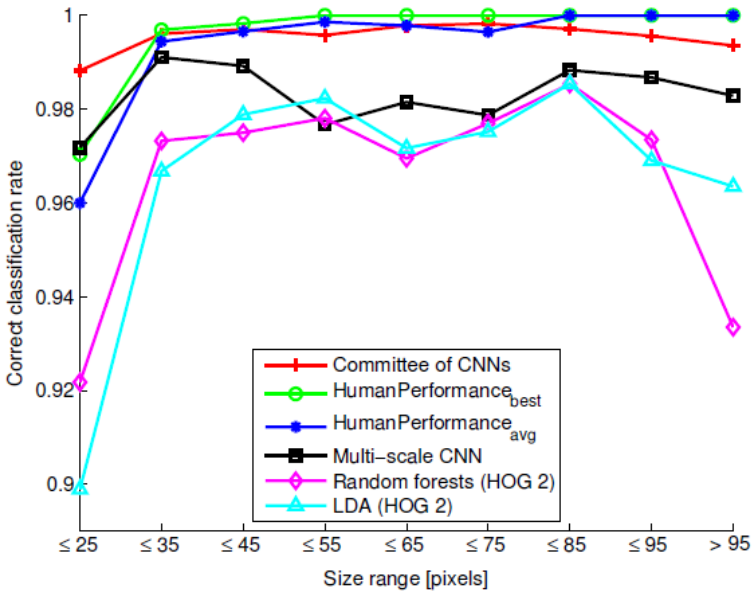
O trabalho de (STALLKAMP *et al.*, 2012) foi focado principalmente na variação de tamanho das imagens, simulando a variação de distância entre o carro e os sinais de trânsito. É dado que imagens menores fornecem pouca resolução e, consequentemente, gerando uma maior dificuldade para a visão humana que tem pouca capacidade para enxergar em longas distâncias.

Já para a Visão Computacional, as imagens muito grandes e que são detectadas com maior proximidade ao veículo, geralmente ficam muito desfocadas, apresentando também "fantasmas" (imagens duplicadas). Este problema se deve ao fato do movimento do veículo em relação ao sinal de trânsito (STALLKAMP *et al.*, 2012). Então existem duas dificuldades diferentes para a visão artificial e a visão humana.

Para determinar o desempenho do reconhecimento humano, foi utilizada uma interface que apresenta para cada pessoa avaliada três imagens da mesma placa, sendo: (a) imagem real, (b) imagem ampliada para facilitar o reconhecimento e uma (c) imagem desfocada. O humano deveria indicar em um painel qual a placa correspondente em um tempo controlado (STALLKAMP *et al.*, 2012), garantindo uma simulação mais próxima de uma situação real.

No gráfico da Figura 18, pode ser observada uma comparação que foi feita para o desempenho de reconhecimento humano com as variadas técnicas de classificadores de imagens artificiais. Também foi utilizado como fator de classificação a variação do tamanho das imagens detectadas (STALLKAMP *et al.*, 2012), já que este é um fator de grande importância para o reconhecimento de sinais de trânsito em veículos inteligentes.

Figura 18 – Desempenho do reconhecimento de sinais de trânsito relacionado ao tamanho da imagem.



Fonte: (STALLKAMP *et al.*, 2012).

Pode ser observado neste gráfico que o reconhecimento humano foi feito em duas análises distintas, vermelho (grupo de pessoas 1[*best*]) e azul marinho (grupo de pessoas 2 [*avg*]). Os resultados foram de alta capacidade para o reconhecimento de imagens por humanos (Figura 18). Pode também ser observado que os métodos de Aprendizado de Máquina ficaram bem próximos aos valores da capacidade humana. Sendo as RNAs profundas (*CNNs*) as que apresentaram melhores resultados se comparadas as outras técnicas aplicadas para classificação. Outras técnicas também apresentaram bons resultados no processo de classificação, sendo estas: (*Multi-scale CNN*, *Random Forests* e *Linear Discriminant Analysis*).

Pode ser notado também que as taxas de reconhecimento humana são mais baixas para as imagens menores, isto se deve ao fato da resolução ser menor (Figura 18). A baixa resolução com características de imagens borradas e desfocadas, trás problemas para a capacidade de visão humana. Isto pode gerar grandes problemas para o reconhecimento de sinais númericos ou outros sinais de trânsito com mais chances de serem confundidos (STALLKAMP *et al.*, 2012).

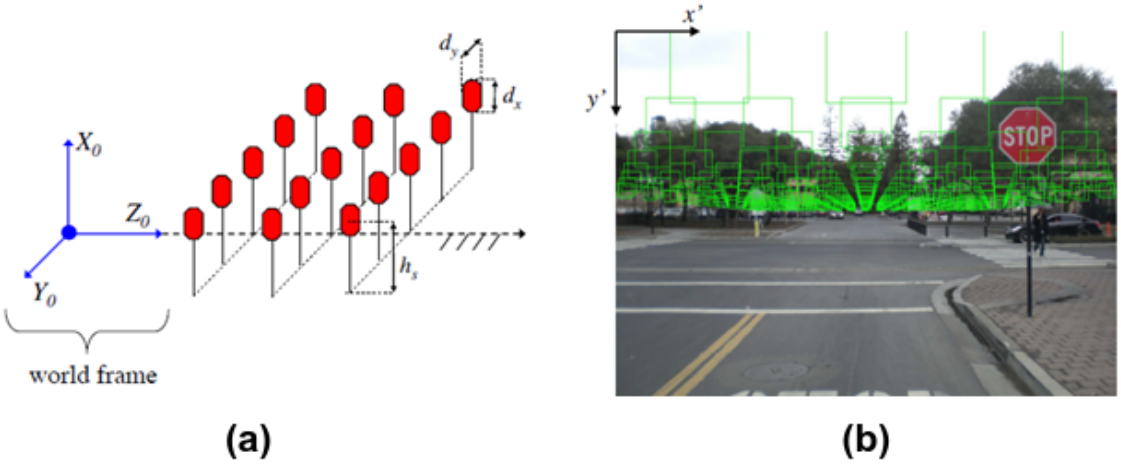
De maneira geral, é possível concluir que o reconhecimento humano é melhor para imagens de sinais de trânsito maiores e mais próximas, já que a visão humana é limitada a uma

pequena distância. Já o reconhecimento artificial, é melhor para imagens menores, pois para as imagens muito grandes e próximas ao veículo, existe o problema de imagens duplicadas devido ao desfoque gerado pelo deslocamento (STALLKAMP *et al.*, 2012).

Em um trabalho desenvolvido por (Schlosser; Montemerlo; Salisbury, 2010), é feita uma proposta para uma nova estrutura que possibilite a detecção dos sinais de trânsito verticais com uma maior rapidez utilizando regiões de interesse. O foco do trabalho é voltado para a redução do custo de processamento computacional, assim, aumentando o desempenho dos métodos de detecção baseados em *Deep Learning* já existentes. Para que isso seja possível, é feita uma busca no espaço da imagem com regiões de interrese pré-definidas e utilizando geometria $3D$ para avaliar cada sinal de trânsito (Figura 19).

O resultado final do trabalho de (Schlosser; Montemerlo; Salisbury, 2010), é um modelo de Atenção Visual capaz de trabalhar com informações relativas às dimensões relacionadas com as escalas da imagem e, utilizando as variáveis de altura em relação ao solo (Si) e distância do veículo (Sv). Essa análise deve possibilitar que um sinal de trânsito vertical seja encontrado em determinadas regiões com maior probabilidade, evitando uma busca exaustiva por toda a imagem (Figura 19-(a)).

Figura 19 – Sistema de análise de imagens por regiões de interesse em funcionamento (a) modelagem da visão em $3D$ representando possíveis locais de sinais de trânsito e (b) projeção de possíveis locais de sinais de trânsito.



Fonte: Adaptada de (Schlosser; Montemerlo; Salisbury, 2010).

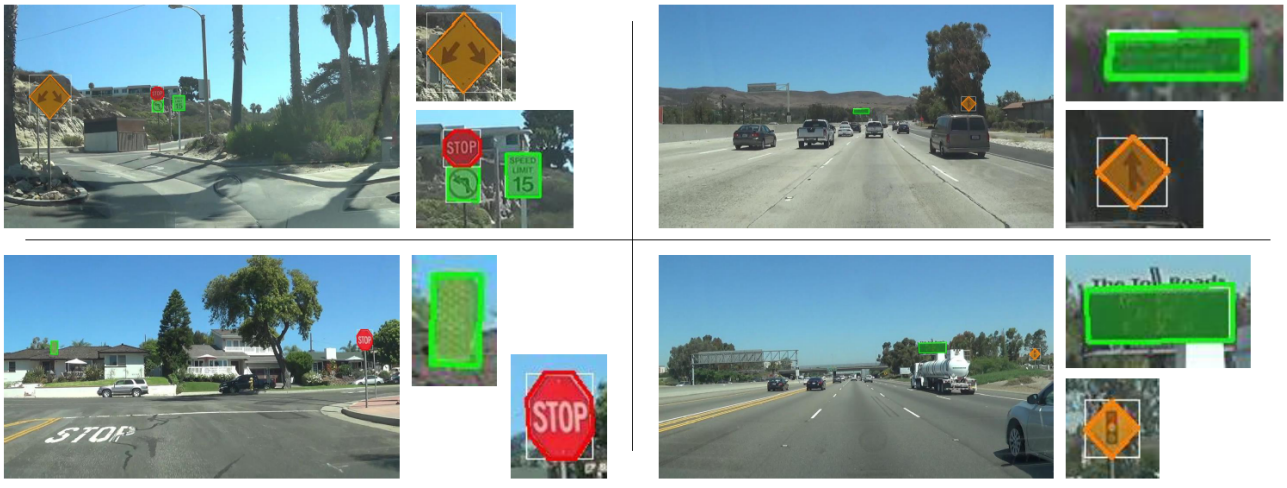
Para que o sinal de trânsito seja buscado nestas regiões, é utilizada uma máscara de tamanho definido (*Slide Window*) que percorre somente as regiões de interesse (Figura 19-(b)). Isso deve reduzir o custo de processamento, já que nas regiões em que não é esperado encontrar nenhum sinal de trânsito, o processamento para detecção não é aplicado.

O sistema resultou além de um menor tempo de computação, a redução de falso positivos, já que objetos fisicamente fora das regiões de interesse de busca são eliminados. Porém, o método de (Schlosser; Montemerlo; Salisbury, 2010) não trabalha com a identificação de área navegável, o que gera problemas em relação a definição das regiões de interesse onde espera-se encontrar os sinais de trânsito verticais. Para corrigir isso, é utilizado um parâmetro de incerteza relacionado a posição do veículo, no entanto, não sendo tão eficiente para o problema.

Neste trabalho os dados $3D$ são relacionados apenas com a profundidade (d) e altura (h) que se espera encontrar um sinal de trânsito, assim, dando suporte na definição das regiões de interesse (Schlosser; Montemerlo; Salisbury, 2010). No entanto, neste trabalho de doutorado um melhor aproveitamento dos dados $3D$ foi feito, possibilitando então detectar os sinais de trânsito verticais também por sua estrutura física (formas geométricas tridimensionais) e relacionando cada informação detectada com a área navegável e outros objetos de trânsito da cena (cones, placas, pessoas), evitando gerar falsos positivos e falsos negativos.

Em um outro trabalho de (Lee; Kim, 2018), foi desenvolvido um sistema de detecção de múltiplos sinais de trânsito no mesmo instante de tempo. O sistema é capaz de adaptar seu modelo de detecção para diferentes poses e formatos de sinais de trânsito verticais. No entanto, quando existem sinais em conflito (velocidade, sentido da via, parada obrigatória e semáforo verde) e que são detectados no mesmo tempo, o sistema não é capaz de analisar a prioridade de cada um (Figura 20). Nesta tese de doutorado, este problema foi atacado visando um modelo de Atenção Visual mais robusto para situações de conflito de informações de trânsito.

Figura 20 – Detecção de múltiplos sinais em dados $2D$.



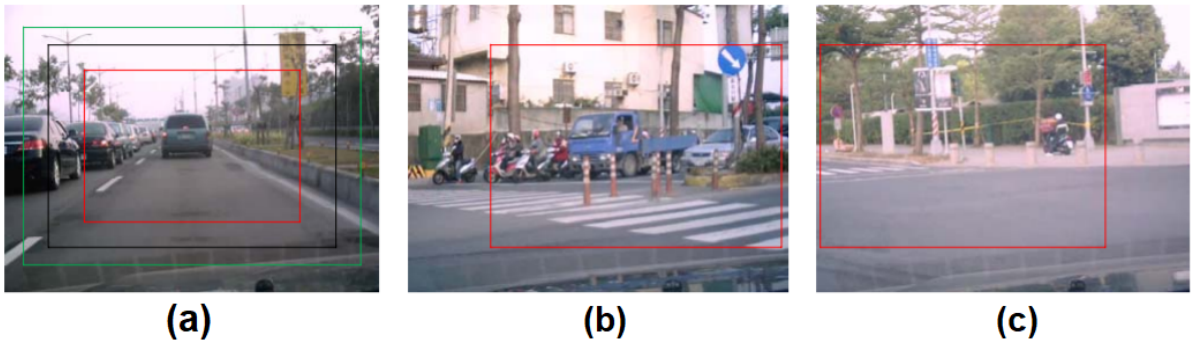
Fonte: (Lee; Kim, 2018).

Em um trabalho de Lin e Wang (LIN; WANG, 2012), um sistema de Atenção Visual *Fuzzy* foi desenvolvido relacionando o ângulo da direção do veículo à região equivalente do campo de visão do veículo (Figura 21). O sistema também pode ser ajustado de acordo com a velocidade

da navegação e detectar sinais de trânsito em diferentes distâncias, possibilitando dar prioridade maior para os sinais mais próximos e menor prioridade para sinais mais distantes, já que a tomada de decisão do veículo deve obedecer primeiro os sinais que são detectados primeiro e logo depois os que são detectados mais distantes (LIN; WANG, 2012).

O trabalho desenvolvido nesta tese de doutorado também utilizou regiões de interesse *Fuzzy*, no entanto, com uma base de regras e análise semântica melhor otimizada para definir a prioridade de cada sinal de trânsito detectado.

Figura 21 – Modelo de Atenção Visual para deteção de sinais de trânsito com base na velocidade e esterçamento: (a) Região de interesse adaptável com velocidades diferentes: $(a1)$ a região de interesse central para velocidade normal, $(a2)$ a região periférica 1 para alta velocidade e a $(a3)$ região periférica 2 para velocidade lenta, (b) exemplos de regiões de interesse adaptativas: virando à direita e (c) virar à esquerda.

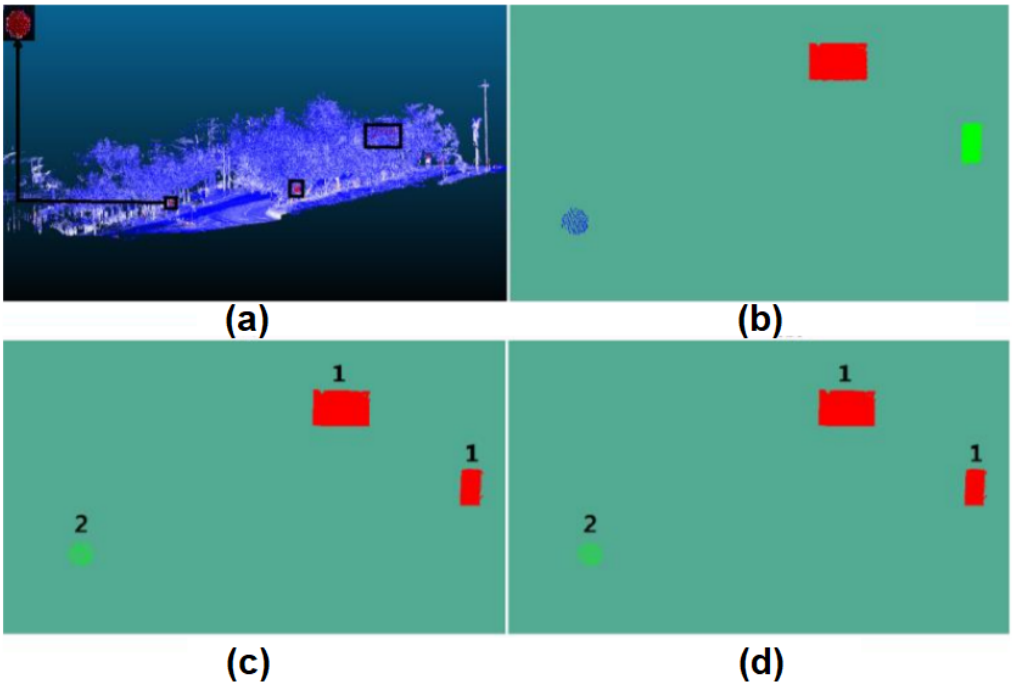


Fonte: (LIN; WANG, 2012).

Outra pesquisa que apresentou um trabalho com detecção de múltiplos sinais de trânsito, foi desenvolvida por (WENG *et al.*, 2016). O trabalho foca na detecção de sinais de trânsito em dados $3D$ por meio da nuvem de pontos do LIDAR (Figura 22-(a)). No entanto, falsos objetos com os mesmos formatos dos sinais de trânsito geram um grande volume de falsos positivos (Figura 22-(b)). O filtro desenvolvido para este erro trabalha com uma análise de refletância dos sinais de trânsito informados por meio de placas (Figura 22-(c) e (d)). No entanto, ainda assim detectando outros tipos de objetos que não são sinais de trânsito e gerando muitos falsos positivos. O sistema também não é capaz de classificar a classe do sinal de trânsito detectado, já que só trabalha com dados $3D$, impossibilitando classificar a superfície de cada objeto detectado.

Visando os trabalhos relacionados apresentados (Schlosser; Montemerlo; Salisbury, 2010) ; (Lee; Kim, 2018) e (WENG *et al.*, 2016), é notável que alguns pesquisadores já utilizam o conceito de Atenção Visual em seus trabalhos. No entanto, com poucos recursos para avaliar a cena de maneira semântica, ou até mesmo, avaliar situações de conflitos entre informações da cena. Nesta tese de doutorado é apresentado no Capítulo 5, um modelo de Atenção Visual capaz de tratar estes tipos de problemas. O modelo de Atenção Visual foi desenvolvido com base em lógica *Fuzzy*.

Figura 22 – Detecção de múltiplos sinais de trânsito em nuvem de pontos $3D$ (a) nuvem de pontos da cena, (b) detecção de possíveis sinais de trânsito e (c), (d) filtragem de falsos positivos por meio da segmentação da nuvem de pontos e análise de refletância.



Fonte: (WENG *et al.*, 2016).

2.3 Sistemas Avançados de Assistência ao Condutor

Nesta seção, serão apresentados alguns trabalhos que são aplicados para dar suporte para o motorista em sua tarefa de dirigir, em muitas vezes, possibilitando evitar acidentes de trânsito. Nesta tese de doutorado, os sistemas de percepção e de Atenção Visual desenvolvidos para detecção e análise de sinais de trânsito podem também contribuir com ADAS, garantindo um maior nível de percepção de regras de trânsito locais para o motorista.

As tecnologias aplicadas em ADAS beneficiaram em uma redução entre 30% a 40% de acidentes de trânsito causados em países europeus (LEFÈVRE; LAUGIER; IBAÑEZ-GUZMÁN, 2011), consequentemente, fazendo destes sistemas um grande diferencial em termos de segurança no trânsito.

Por meio da realização de uma revisão bibliográfica na área de ADAS para veículos robotizados, destacamos os principais estudos experimentais e que estão ligados a esta pesquisa de doutorado:

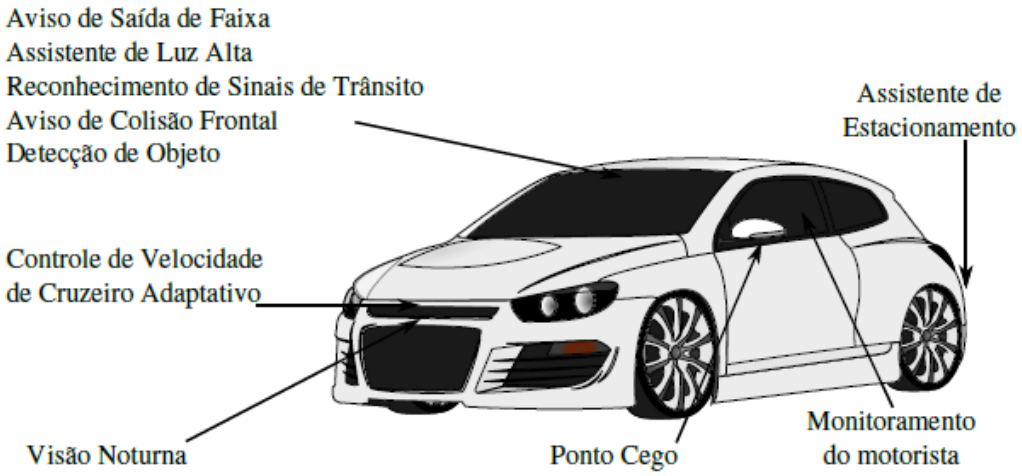
2.3.1 ADAS com percepção externa

Um tipo de ADAS bastante conhecido nos veículos é o controle lateral. Estes sistemas procuram auxiliar o motorista em saídas da pista por descuido ou mudança de faixa de maneira perigosa. Dois mecanismos que podem ser citados, são: a) aviso de saída de faixa (WILSON FA, 2010) e (b) monitoramento de ponto cego.

É notável que a maioria do sensoriamento frontal (Figura 23) é feito por sensores de imagem (câmeras), assim, permitindo detectar um obstáculo ou sinal de trânsito. Lembrando que o sensor pode ser o mesmo para diferentes situações, porém, o processamento dos dados deve ser diferenciado.

Neste cenário, o objetivo desta pesquisa de doutorado é aplicado aos sistemas de percepção longitudinal, voltando-se para a percepção de sinais de trânsito e que devem dar suporte para a detecção de falhas do motorista em diversas situações de desrespeito das regras de trânsito locais.

Figura 23 – Síntese de alguns Sistemas Avançados de Assistência ao Condutor (ADAS).



Fonte: Adaptada de (KISACANIN, 2011).

Os ADAS também devem ser auxiliados por métodos de percepção externos e que são capazes de detectar obstáculos e outros objetos, assim, possibilitando evitar acidentes gerados por colisões. No entanto, o obstáculo deve ser detectado pelo sistema de sensoriamento e solucionado pelo sistema de controle do veículo robótico, aplicando uma frenagem de emergência e/ou uma solução de desvio de rota. Os ADAS utilizados para dar suporte neste tipo de situação devem tomar medidas automáticas para evitar qualquer acidente gerado por uma colisão, possibilitando então eliminar ou mitigar uma falha humana.

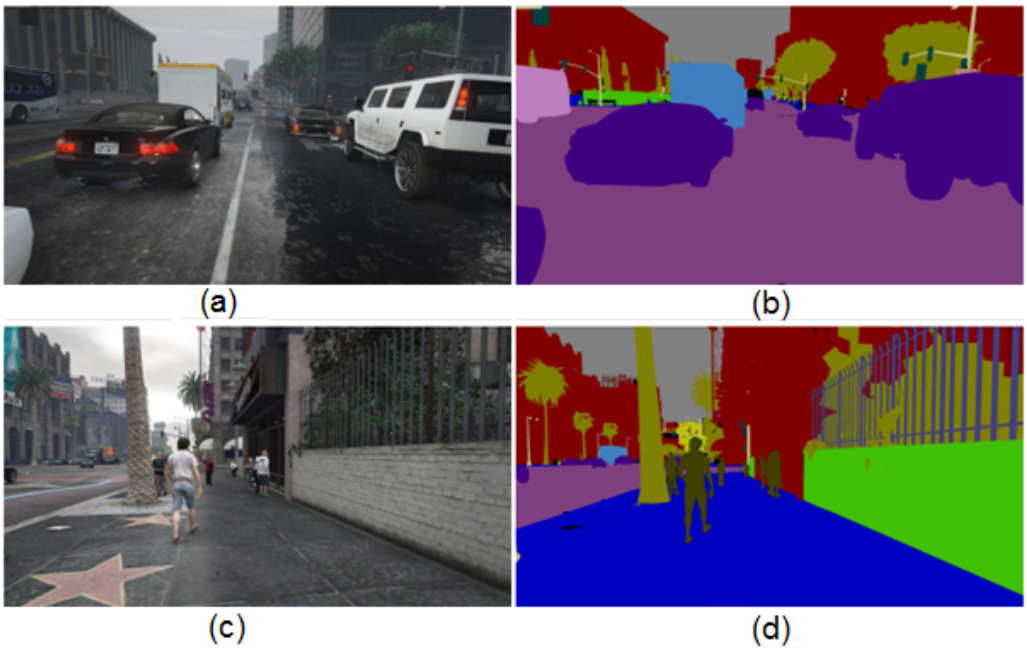
No trabalho feito por (RICHTER *et al.*, 2016), foi desenvolvido um sistema de percepção que aplica uma rede de *Deep Learning - Segnet* (*A Deep Convolutional Encoder-Decoder Archi-*

tecture for Image Segmentation), sendo capaz de detectar e classificar os elementos encontrados em um ambiente de trânsito (Figura 24). No sistema desenvolvido neste trabalho, foi utilizado para os experimentos um ambiente virtual 3*D* baseado em um jogo digital bastante conhecido no mundo dos jogos de veículos em ambientes de trânsito, principalmente por ter gráficos de alta qualidade, mundos extensos, movimentos dos veículos com grande realismo, texturas realistas e, também, uma grande variação de elementos dos "mundos" do jogo. O ambiente digital utilizado foi o *Grand Theft Auto V* (RICHTER *et al.*, 2016).

No trabalho de (RICHTER *et al.*, 2016), o principal objetivo foi desenvolver um sistema de percepção em conjunto com um classificador de imagens baseado em uma rede de *Deep Learning*, assim, possibilitando reconhecer os elementos da cena, como por exemplo: humanos, carros, ônibus, motos e outros objetos também presentes em ambientes de trânsito reais (Figura 25).

Um dos objetivos de maior potencial em se trabalhar com sistemas de visão em ambientes de simulação, está relacionado com a disponibilidade rápida de um grande volume de dados para treinamento e testes, permitindo o desenvolvimento de métodos de classificação de imagens não limitados a dados de treinamento reais, garantindo também a criação de novos cenários não encontrados ou de difícil acesso em situações reais (RICHTER *et al.*, 2016). Na Figura 24, pode ser observado o sistema de Visão Computacional desenvolvido por (RICHTER *et al.*, 2016) em funcionamento.

Figura 24 – Imagens capturadas do jogo e mapeamento de visão geradas pela abordagem de classificação semântica (*a* e *b*) imagens extraídas do jogo *jogo Grand Theft Auto V* e (*c* e *d*) mapeamento semântico para os objetos da cena.

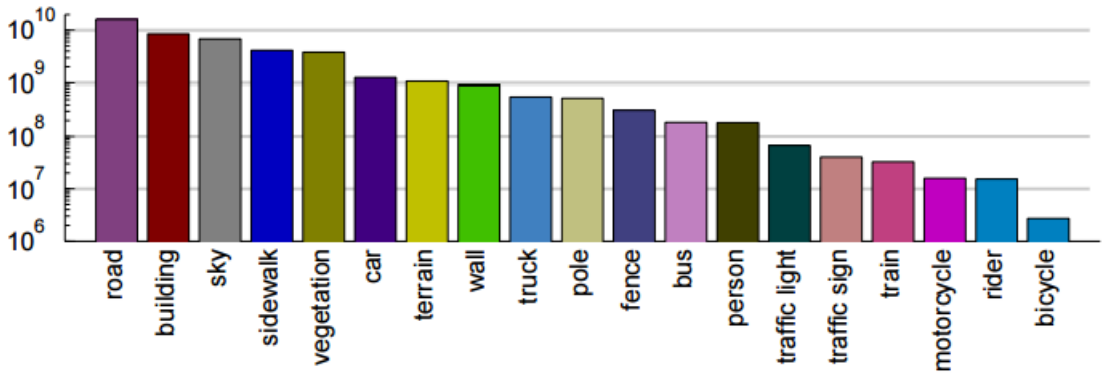


Fonte: Adaptada de (RICHTER *et al.*, 2016).

Por meio da Figura 24-(*a*, *c*), pode ser observado o ambiente de trânsito capturado em imagens 2*D* e que foram geradas pelo jogo *Grand Theft Auto V*. Já na Figura 24-(*b*, *d*), pode ser observado o mapeamento de rótulos semânticos e, por meio deles, é possível reconhecer diferentes elementos encontrados e segmentados no ambiente de navegação (RICHTER *et al.*, 2016). Este processo deve auxiliar o sistema de controle do veículo a tomar medidas compatíveis para evitar diferentes tipos de obstáculos, como por exemplo: um carro tem uma velocidade maior que a de um humano, então a tomada de decisão para este primeiro deve ser mais rápida do que para este segundo.

Por meio do gráfico da Figura 25, podem ser observados os conjuntos de dados referentes aos elementos do ambiente em questão e que foram detectados e classificados utilizando o sistema de visão baseado na rede de *Deep Learning* Segnet.

Figura 25 – Número de *pixels* capturados, segmentados e classificados no conjunto de dados do sistema de visão.



Fonte: (RICHTER *et al.*, 2016).

O trabalho de (RICHTER *et al.*, 2016) foi desenvolvido para um ambiente virtual onde as variações físicas ambientais de um cenário real, como por exemplo, as variações de iluminação, são totalmente controladas. Sendo assim, este sistema pode ser ineficiente se aplicado em um ambiente real. Porém, o ambiente disponibiliza imagens 2*D* e 3*D* e, adaptações podem ser feitas para que o sistema seja eficiente também em ambientes reais. Possibilitando então o aproveitamento de um sistema de grande potencial de percepção treinado e desenvolvido em ambiente de simulação para aplicações reais.

2.3.2 ADAS com percepção interna

Para que seja possível analisar o comportamento do motorista, são utilizados sistemas de monitoramento internamente ao veículo robótico. Estes sistemas concentram-se no estado fisiológico e comportamental do motorista, possivelmente, informando quando a tarefa de dirigir

não pode mais ser feita de maneira segura. Duas situações principais que devem ser analisadas por meio do sistema de percepção estão relacionadas com a (a) detecção de distrações e de (b) sonolência (KUTILA, 2006).

Em um trabalho feito por (DAI *et al.*, 2010), é apresentada uma solução que visa a detecção de uma direção perigosa feita por um motorista embriagado. O sistema utiliza técnicas dependentes de um programa instalado em um celular que deve possuir obrigatoriamente os sensores de acelerômetro e de orientação. Este programa tem como finalidade extrair características do modo de direção do motorista e compará-las com um modelo de motorista embriagado previamente cadastrado em um banco de conhecimento (DAI *et al.*, 2010). Essa comparação deve ser feita por meio de uma RNA (DAI *et al.*, 2010). O sistema deve alertar a polícia sobre a direção perigosa, evitando então um possível acidente de trânsito.

Em um trabalho feito por (CARROLL; BELLEHUMEUR; CARROLL, 2013), também foi desenvolvido um sistema capaz de detectar uma possível embriaguez na direção. O sistema analisa o teor alcoólico no sangue do motorista por meio de um sensor transdérmico. O sistema utiliza um modelo de aprendizado de máquina baseado em uma RNA previamente treinada, assim, sendo possível declarar a embriaguez do motorista (CARROLL; BELLEHUMEUR; CARROLL, 2013).

Assim como no trabalho de (DAI *et al.*, 2010) e (CARROLL; BELLEHUMEUR; CARROLL, 2013), existem outros autores que desenvolveram funções em prol da detecção de embriaguez do motorista utilizando técnicas precisas (EDMONDS; HOPTA, 2001) ; (HAILE, 1987) ; (HAILE, 1992) ; (MURATA *et al.*, 2011) ; (SHIRAZI; RAD, 2012) ; (BERRI *et al.*, 2013) e (BERRI R, 2019).

Já para os trabalhos direcionados para a detecção de sonolência do motorista em plena atividade, destaca-se o sistema desenvolvido por (AKROUT; MAHDI, 2013). O sistema utiliza uma câmera frontal ao motorista, visando localizar a face e em seguida os olhos (AKROUT; MAHDI, 2013). Isso é feito usando a técnica de Viola-Jones (VIOLA; JONES, 2004). A transformada de Hough (CAUCHIE; FIOLET; VILLERS, 2008) é aplicada nas regiões que representam os olhos, assim, sendo possível identificar a íris e as pálpebras.

A análise do estado dos olhos do condutor é obtida por meio da classificação de duas características, sendo destas uma não linear e outra estacionária. Por meio desta análise é feita a detecção da sonolência verificando a posição da visão.

Em um trabalho feito por (EBRAHIM *et al.*, 2014), foi desenvolvido um sistema capaz de detectar as características de amplitude ocular, a energia e as velocidades máximas e médias de abertura e fechamento dos olhos e, concomitantemente, as características de frequência de fechamento e o tempo de permanência dos olhos fechados (EBRAHIM *et al.*, 2014). Por meio desta detecção, é possível identificar a sonolência do motorista, sendo então possível tomar atitudes para se evitar um acidente de trânsito. Assim como nos trabalhos de (AKROUT; MAHDI,

2013) e (EBRAHIM *et al.*, 2014), existem outros que desenvolvem funções em prol da detecção de sonolência do motorista utilizando variados mecanismos (BERRI *et al.*, 2013); (DKHIL *et al.*, 2015); (KUMAR; SIMON, 2015) e (LENSKIY; LEE, 2012).

2.4 Suporte para Correção de Falhas Humanas

De maneira geral, o foco da presente pesquisa de doutorado está relacionado com o desenvolvimento de um novo sistema de análise e fusão de imagens $2D$ e $3D$ em conjunto com um modelo de Atenção Visual, visando então a detecção e classificação de sinais de trânsito e análise de suas respectivas prioridades, respectivamente. O modelo de percepção para sinais de trânsito pode ser integrado a um sistema de assistência ao motorista, visando o aumento da segurança viária, assim, comparando a visão externa (com regras de trânsito) com os dados de comportamento do motorista e que são provenientes da análise da tarefa de dirigir (CARROLL; BELLEHUMEUR; CARROLL, 2013) ; (DAI *et al.*, 2010) ; (EDMONDS; HOPTA, 2001) ; (HAILE, 1987) ; (HAILE, 1992) ; (MURATA *et al.*, 2011) ; (SHIRAZI; RAD, 2012) e, também, dos dados internos do veículo obtidas via rede *Controller Area Network* (CAN), como por exemplo, indicando a velocidade (aceleração e frenagem) e esterçamento do veículo. Possibilitando então detectar falhas e desrespeito as regras de trânsito locais.

Por meio da deteção de falhas humanas e de desrespeito as regras de trânsito locais disponibilizadas por meio do sistema integrado de percepção externa e interna ao veículo, um auxílio automatizado a condução do veículo pode ser ativado de acordo com uma base de regras equivalente para cada problema encontrado. Com isto, é possível corrigir falhas de maneira automática, ou em casos extremos, assumir o controle do veículo. O sistema de percepção também pode auxiliar o motorista simplesmente a detectar sinais de trânsito que não foram percebidos.

2.5 Considerações

Neste capítulo foram apresentados trabalhos relacionados com sistemas de percepção para sinais de trânsito verticais e modelos de Atenção Visual em prol da navegação de veículos robóticos inteligentes. Também foram apresentados alguns modelos de ADAS aplicados para a detecção e correção de falhas humanas na tarefa de dirigir. Estes trabalhos estão dentro do mesmo escopo e junto ao estado-da-arte da pesquisa realizada nesta tese.

Alguns dos trabalhos encontrados na revisão da literatura, principalmente dos autores (Timofte; Zimmermann; Gool, 2009a) e (Zhou; Deng, 2014), trabalham com percepção de sinais de trânsito em imagens com fusão de dados $2D$ e $3D$, no entanto, ainda existem pontos negativos a serem tratados e que são sugeridos como trabalhos futuros pelos próprios autores. Estes pontos negativos estão muito relacionados com a detecção de falsos sinais envolvendo

imagens 2*D*. Também são destacados problemas no sensoriamento 3*D*, principalmente com dados provenientes de LIDARs que trabalham com nuvens de pontos esparsas e, para o caso dos sinais de trânsito, geram grandes problemas por serem objetos pequenos e com uma aglutinação de pontos muito baixa. Também são destacados problemas relacionados com *sprays* em imagens 3*D* provenientes de erros na estimativa de profundidade e disparidade entre pontos equivalentes.

Estes problemas impossibilitam detectar de maneira completa as informações das superfícies dos sinais de trânsito, consequentemente, gerando problemas para um reconhecimento preciso. Para a correção destes problemas, entre outros relacionados ao longo desta revisão da literatura, foi desenvolvido um sistema de percepção baseado na fusão de sensoriamento 2*D* e 3*D*, possibilitando um melhor nível de informações e características dos objetos da cena.

Outro problema que não foi tratado em outras pesquisas e, que foi desenvolvido nesta tese, está ligado a um modelo de Atenção Visual *Fuzzy* capaz de classificar a prioridade de cada sinal de trânsito detectado em conflito de informações. O modelo de Atenção Visual foi definido como uma camada entre o sistema de percepção e de tomada de decisão do veículo, possibilitando então obter um maior nível de informações do ambiente de navegação para dar suporte a tomada de decisão baseada em regras de trânsito.

Algumas lacunas e pontos negativos encontrados por meio da revisão sistemática da literatura e que merecem ser destacados, sendo estes tratados de maneira completa ou parcial nesta pesquisa de doutorado, são:

- A maioria das pesquisas atuais na área de detecção de sinais de trânsito é focada no poder das redes de *Deep Learning* sobre imagens 2*D*. O que fez com que as pesquisas com imagens 3*D* ficassem em menor quantidade depois do ano de 2015. Sendo que estas redes trabalhando com dados puramente 2*D* ainda apresentam grandes problemas com detecção de objetos falsos;

- No trabalho de (Mathias *et al.*, 2013), foi desenvolvido um bom sistema para detecção de sinais de trânsito, no entanto, trabalhando com uma análise 3*D* muito voltada para a geometria da placa, gerando uma grande quantidade de falsos positivos relacionados com objetos do mesmo formato, como por exemplo, telhados de casas (triangulares) e vidraças de igreja (losango) (Mathias *et al.*, 2013). Porém, depois do avanço dos trabalhos de detecção e reconhecimento de sinais de trânsito verticais em imagens 2*D* por meio das redes *Deep Learning*, estes autores voltaram suas pesquisas para este tipo de imagem;

- No trabalho de (ZHAO; YUAN; DANG, 2014), também foi feita uma aposta na detecção de sinais de trânsito verticais em imagens 3*D*. Em todos seus trabalhos aplica-se como principal sensor o $3D-LIDAR$ (*Velodyne HDL32*). Estes autores apostam em uma percepção puramente 3*D* em conjunto com uma SVM para classificação do tipo da informação detectada, no entanto, não aproveitando o potencial de imagens 2*D* em conjunto com métodos de aprendizado de máquina baseados em *Deep Learning* para a tarefa de classificação;

- No trabalho de (Li; Lv; Wang, 2017), é aplicado um método de detecção de sinais de trânsito baseado em imagens 2*D* via sensoriamento por câmeras. Utilizando para a detecção a rede de *Deep Learning You Only Look Once* (YOLO), onde é aliada com a análise de refletância dos objetos da cena por meio de dados do LIDAR, gerando um método intitulado como *Reflectance ConvNet* (RefCN). Porém, neste método, ainda é necessário um ajuste no filtro de refletância para que se consiga detectar melhor os sinais de trânsito, eliminando falsos objetos com refletância parecida (Li; Lv; Wang, 2017). Outro problema é relacionado com a aglutinação de pontos do LIDAR, sendo muito esparsa e impossibilitando detectar sinais em longas distâncias (> 25 metros);

- No trabalho de (Soilan *et al.*, 2015), foi desenvolvido um sistema de percepção capaz de relacionar os sinais de trânsito com a via. No entanto, o sistema utiliza apenas a via como referência para detectar as posições dos sinais de trânsito verticais. Não é feita uma análise para verificar a qual via a informação pertence. Este trabalho não resolve o problema de sinais em bifurcações ou rotas auxiliares;

- Alguns trabalhos envolvendo Atenção Visual (FIGUEIRA A, 2019) ; (STALLKAMP *et al.*, 2012) ; (Schlosser; Montemerlo; Salisbury, 2010) ; (Lee; Kim, 2018), focam seus objetivos em desenvolver uma maior capacidade de detecção de múltiplos sinais de trânsito. No entanto, em nenhum trabalho foi desenvolvido um modelo capaz de tratar informações em conflito: velocidade, sentido da via, semáforos e parada obrigatória;

- No trabalho de (RICHTER *et al.*, 2016), é notável o potencial da rede de *Deep Learning* SegNet para clusterização de objetos em imagens puramente 2*D*, porém, os problemas neste tipo de imagem ainda continuam sendo uma desvantagem para detecção de objetos de interesse, gerando muitos falsos positivos e falsos negativos;

- Foi eliminada nesta tese a possibilidade de percepção de sinais de trânsito baseada apenas em dados 2*D* pelos tantos problemas relatados ao longo da pesquisa. Por mais que as redes de *Deep Learning* apresentem bons resultados em dados 2*D*, ainda são encontrados grandes problemas neste tipo de imagem em bons trabalhos (Hossain; Hyder, 2015) ; (Zeng *et al.*, 2017); (Zhe; Jingyi; Chaoqian, 2016) ; (Zuo *et al.*, 2017);

- Os métodos de ADAS para percepção externa estão voltados particularmente para detecção de faixas horizontais para o controle lateral (WILSON FA, 2010). Não são encontrados sistemas de percepção de sinais de trânsito aliados aos dados internos do veículo (velocidade, frenagem, esterçamento). Possibilitando detectar imprudências, distrações e falhas na tarefa de dirigir com desrespeito as regras de trânsito. Os atuais trabalhos são aplicados, na maioria do casos, na análise fisiológica do condutor (LEFÈVRE; LAUGIER; IBAÑEZ-GUZMÁN, 2011) ; (CARDARELLI, 2012) ; (KUTILA, 2006).

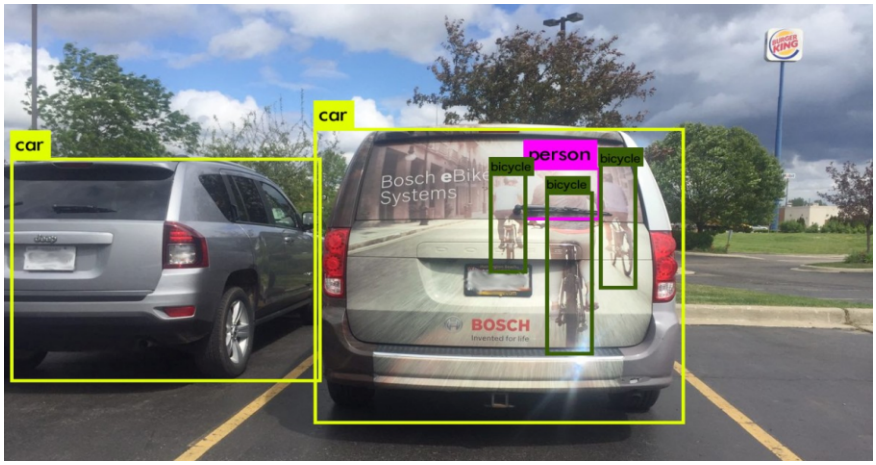
Por meio da Tabela 1, é apresentada uma síntese para efeito de comparação entre os trabalhos relacionados que são de maior importância para esta pesquisa de doutorado. Nesta tabela, pode ser notado que poucos trabalhos utilizam fusão de sensores $2D$ e $3D$, sendo que nenhum deles faz uma análise de prioridades dos sinais de trânsito que são detectados de maneira múltipla e com conflitos de informações. Também não foram encontrados trabalhos que fazem análise de bifurcações e seus sinais correspondentes em situações não mapeadas *a priori*.

Tabela 1 – Tabela de comparação entre os trabalhos relacionados com a tese.

Comparação entre os trabalhos relacionados																
Trabalhos relacionados	(Timofte; Zimmermann; Gool, 2009a)	(SALTI *et al.*, 2015)	(Zhou; Deng, 2014)	(AIJAZI *et al.*, 2016)	(BALALI; Golparvar Fard, 2015)	(WANG *et al.*, 2014)	(Wu *et al.*, 2015)	(Soilan *et al.*, 2015)	(Lee; Kim, 2018)	(Hossain; Hyder, 2015)	(Yuan; Xiong; Wang, 2017)	(Jung *et al.*, 2016)	(Schlosser; Montemerlo; Salisbury, 2010)	(Lee; Kim, 2018)	(LIN; WANG, 2012)	**Tese**
Tipos de dados																
Dados 2D	X	X		X	X	X			X	X	X	X		X	X	X
Dados 3D	X		X	X	X	X	X	X					X			X
Fusão de Sensores			X	X			X	X								X
Reconhecimento																
SVM				X	X	X					X					
RNA	X		X			X	X			X			X			X
DCNN	X	X				X		X	X			X		X	X	X
Atenção Visual																
Fuzzy															X	X
DCNN									X							X
Regiões de interesse													X	X	X	X
Funções																
Detecção	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
Reconhecimento	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X
Localização dos sinais								X								X
Eliminação de falsos sinais	X						X				X					X
Análise de prioridades dos sinais																X
Análise de sinais em função da via								X								X

Alguns autores focam seus trabalhos sendo direcionados apenas ao uso de imagens 2*D*, buscando obter bons resultados baseados no alto potencial das redes de *Deep Learning*. No entanto, utilizando apenas este tipo informação, pode ser muito complexo o tratamento de problemas relacionados a falsos positivos e falsos negativos. Por meio da Figura 26, é possível observar um exemplo para uma situação em que o sistema de detecção e classificação de obstáculos de um veículo autônomo identifica não somente dois carros, mas também, detecta incorretamente três ciclistas. Este problema foi causado por um adesivo que foi colado atrás de um dos carros (MIT, 2017). Além disso, nesta figura, existe um sinal de *fast food* que pode ser detectado como um sinal de trânsito, uma vez que em 2*D* não é possível obter as informações sobre o tamanho e a distância desse objeto.

Figura 26 – Número de *pixels* capturados e classificados no conjunto de dados do sistema de visão.



Fonte: (MIT, 2017).

Este problema dificilmente seria possível de ser tratado em imagens 2*D*, no entanto, se um sistema de detecção 3*D* fosse aplicado em conjunto, seria facilmente corrigido. Considerando a noção de profundidade e geometria 3*D*, seria simples identificar que são apenas dois carros nesta cena e não existe nenhum outro objeto na mesma posição em que foram detectados. Ou seja, as bicicletas, carros e pedestres não podem estar no mesmo plano (x, y e z) para uma situação real.

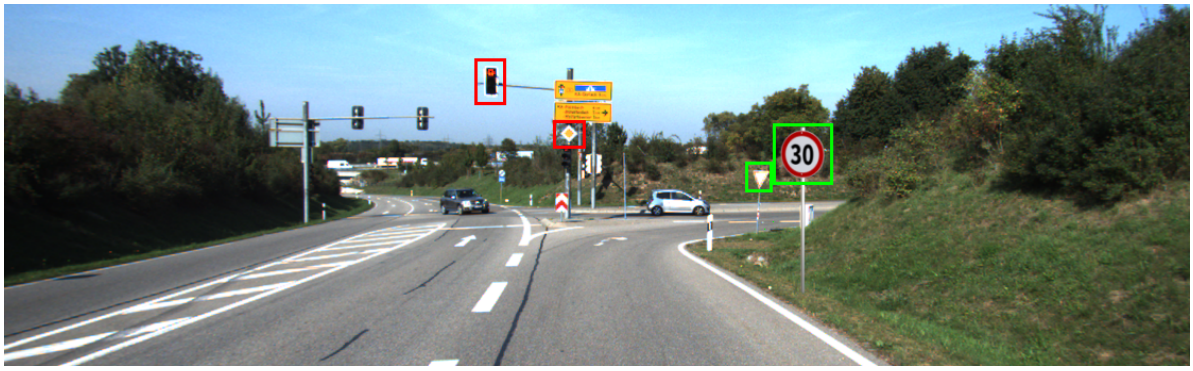
Outro problema bastante frequente está relacionado com os sinais de trânsito que são colados na traseira de veículos de tranporte escolar ou de carga (Figura 27). Estas informações são relacionadas com a velocidade máxima permitida ao veículo em particular e não a velocidade máxima da via para os outros veículos. Estes tipos de sinais geram então falsos positivos para a navegação e são intratáveis atualmente utilizando apenas dados 2*D*. Já que não seria possível identificar sua profundidade na cena e declarar que este sinal é apenas um adesivo na traseira de um terceiro veículo na via. Também não seria possível avaliar a posição x, y e z do sinal de trânsito e verificar que ele está em movimento, ou em uma posição da rodovia não habitual.

Figura 27 – Falsos sinais de trânsito: (*a*) estacionamento com vários ônibus com sinais de velocidade colados e (*b* e *c*) falsos sinais de trânsito detectados em 2*D* pela rede YOLO.

Fonte: Elaborada pelo autor.

Outro problema não tratado atualmente em sistemas de percepção, está relacionado com a análise de prioridades de múltiplos sinais de trânsito detectados no mesmo instante. Por meio da Figura 28, pode ser observada uma situação onde vários sinais são detectados ao mesmo tempo em uma bifurcação, no entanto, dependendo da escolha da rota, sinais diferentes devem ser obedecidos. Caso a via da direita seja escolhida, os sinais de: Velocidade máxima de 30Km/h e "Dê a preferência" são prioritários. Caso a escolha seja seguir em frente, a velocidade máxima deve ser mantida com base em informações anteriores e o sinal vermelho deve ser obedecido. Este tipo de análise semântica é fundamental para verificar a importância de cada regra de trânsito em um dado momento da navegação, garantindo também interpretar informações de emergência na pista em situações não mapeadas: obras, acidentes e desvios.

Figura 28 – Problema de detecção de múltiplos sinais de trânsito em bifurcações.



Fonte: Elaborada pelo autor.

Em resumo, estes tipos de falhas em dados provenientes de imagens $2D$ podem gerar grandes problemas relacionados com a detecção de sinais de trânsito. No entanto, com a fusão de dados $2D$ e $3D$, é possível corrigir estas falhas e gerar novos potenciais para os sistemas aplicados nesta tarefa, conseguindo então uma maior robustez e eficiência na detecção de sinais de trânsito com base em informações de profundidade. Já com os dados $2D$ é possível obter um melhor aproveitamento das cores, formas e texturas das superfícies dos sinais de trânsito, assim, possibilitando que os métodos de *Deep Learning* apresentem melhores resultados na tarefa de classificação. Visto que este tipo de rede trabalha muito bem com estes tipos de informações encontradas apenas em imagens $2D$, resultando então em classificadores com acurácia próxima de $98 - 99\%$.

CAPÍTULO

3

REFERENCIAL TEÓRICO

Por meio desta seção, serão apresentados conceitos importantes relacionados com a pesquisa de doutorado desenvolvida nesta tese, possibilitando então um melhor entendimento dos modelos, métodos, técnicas e ferramentas computacionais que foram utilizados.

3.1 Processamento de Imagens Digitais

O Processamento Digital de Imagens (PDI) consiste na aplicação de operações e técnicas matemáticas para que seja possível obter/extrair dados de interesse em imagens digitais. Estas técnicas são direcionadas para aplicações computacionais e que estão relacionadas com algum processamento inicial para modelos de Visão Computacional. Para que isso seja possível, é necessário transformar as situações em questão em modelagem computacional de problemas reais e, logo seguinte, gerar os programas de computador que são ajustados para resolver um determinado problema de maneira automática (SILVA; SPATTI; FLAUZINO, 2010) ; (GONZALEZ; WOODS, 2010).

Primeiramente, uma imagem digital é formada por uma matriz (*linha* x *coluna*) em que cada endereço é correspondente a um *pixel* que representa a imagem discretizada do mundo real e contínuo para o mundo digital e discreto.

O processo é iniciado com a aquisição da imagem por meio de um sensor em dados $2D$ (câmera monocular) ou $3D$ (câmera estéreo ou LIDAR). Esta imagem capturada possui informações relacionadas com suas características em forma de sinais elétricos contínuos. Já nos computadores, o trabalho é feito de maneira discreta, então é necessária uma conversão dos sinais (SILVA; SPATTI; FLAUZINO, 2010); (GONZALEZ; WOODS, 2010).

Depois que uma imagem é capturada pelo sistema de aquisição de imagens, é necessário então começar o tratamento desta para que seja melhor aproveitada para a aplicação desejada. Visando essa motivação, filtros são aplicados para remoção dos ruídos e melhoramento do contraste,

assim, facilitando obter bons resultados nas próximas etapas (SILVA; SPATTI; FLAUZINO, 2010); (GONZALEZ; WOODS, 2010).

Depois destas duas etapas, o processamento de imagens pode ser voltado para a segmentação, visando também uma posterior extração das características das imagens. Nesta etapa, é necessário que uma segmentação dos dados em grupos de *pixels* que formam os objetos e, também, o fundo, seja feita (SILVA; SPATTI; FLAUZINO, 2010); (GONZALEZ; WOODS, 2010). O objeto na maioria dos casos é o que mais interessa, possibilitando então tratá-lo, já o fundo, certamente será descartado. Como exemplo, na robótica móvel, em alguns casos, os objetos estão relacionados com os obstáculos na rota do robô.

3.2 Visão Computacional em Robótica

No tópico anterior, foi descrito basicamente como funciona o Processamento de Imagens Digitais. Já neste tópico, será explicado em mais detalhes o funcionamento de um sistema de Visão Computacional que também utiliza técnicas de processamento de imagens como pré-processamento. Possibilitando o desenvolvimento de sistemas robustos e capazes de realizar tarefas de grande complexidade na área de visão robótica.

Visando enfrentar a complexidade de se trabalhar com grandes volumes de dados em sistemas de Visão Computacional e, que são capturados por meio de sensores $2D$ e $3D$, são necessárias técnicas e métodos para processamento destes sinais com a capacidade de se trabalhar em tempo real. Esta questão deve ser considerada, uma vez que os sistemas de percepção robóticos necessitam de uma alta velocidade na resposta.

Para trabalhar com reconhecimento de padrões, existem hoje na computação, métodos robustos de extração de características (*features*) (KRIG, 2016). A extração de características, nada mais é do que uma abstração dos dados reais para um conjunto condensado de características que são de maior relevância para dada aplicação. Estas características de maior importância devem eliminar, principalmente, a redundância, assim, evitando armazenar e processar um grande volume de dados. Resumidamente, uma característica nada mais é que uma representação compacta e rica de um conjunto de dados bem definido (SALES, 2017).

Depois de aplicada a extração de características, é necessário então armazenar estas informações, utilizando então um formato de dados conhecido em um descritor. Este processo é feito para que seja possível em uma etapa futura, os dados serem comparados por alguma métrica. Os dois principais elementos de um descritor com dados $3D$ são (TOMBARI *et al.*, 2014):

- **Descritividade**: armazenar de maneira abstrata as informações predominantes de um conjunto de dados, sendo então capaz de fornecer dados para um classificador distinguir diferentes classes;

- **Robustez**: está muito ligada à invariância a um conjunto de transformações ou ruídos (TOMBARI *et al.*, 2014).

Os descritores podem ser divididos em locais ou globais. Sendo os locais para a representação de áreas bem definidas como junções, cantos e bordas. Para este tipo de descritor, as aplicações são robustas a variações por oclusão e ruído. Nos descritores globais, são fornecidas as descrições de um objeto em sua forma total, sendo então melhores aplicadas em tarefas de classificação (SALES, 2017) ; (BAY *et al.*, 2008).

Os descritores locais de (LOWE, 1999), tiverem um grande impacto na disseminação desta técnica, onde apresentaram grandes vantagens decorrentes dessa abordagem e que gerou grandes trabalhos de Visão Computacional, utilizando esta técnica para processar imagens digitais com grande confiabilidade e repetibilidade (KRIG, 2016). Na área de visão envolvendo processamento de imagens $2D$, esta técnica já é bastante comum e também de grande potencial. As técnicas *SIFT* (LOWE, 1999) e *SURF* (BAY *et al.*, 2008) são dois exemplos de aplicações de grande potencial para extração de características $2D$ (SALES, 2017).

Com a evolução dos sensores $3D$ (câmeras estereoscópicas e LIDARs $3D$) e também da aplicação de fusão de sensores (acelerômetros, câmeras $2D$ e unidades inerciais), o volume de informação que deve ser tratado é muito grande, o que exige maior custo computacional e de tempo. Dados estes problemas, é então necessária a aplicação de métodos de extração de características especializados para dados $3D$ (KRIG, 2016) ; (SALES, 2017) .

3.3 Análise de Objetos 3D

Como o foco deste trabalho é voltado para percepção de sinais de trânsito envolvendo fusão de sensores $2D$ e $3D$, foi dedicada esta seção para explicar um pouco melhor o funcionamento da extração de características em imagens provenientes destes tipos de sensores.

Em se tratando de descritores $3D$ locais, estes incorporam características da vizinhança, utilizando um ponto de referência específico para cada situação (SALES, 2017). Os descritores $3D$ podem ser divididos em três categorias (TOMBARI *et al.*, 2014): (a) descritores baseados em assinatura, (b) baseados em histogramas e (c) descritores locais híbridos (SALES, 2017).

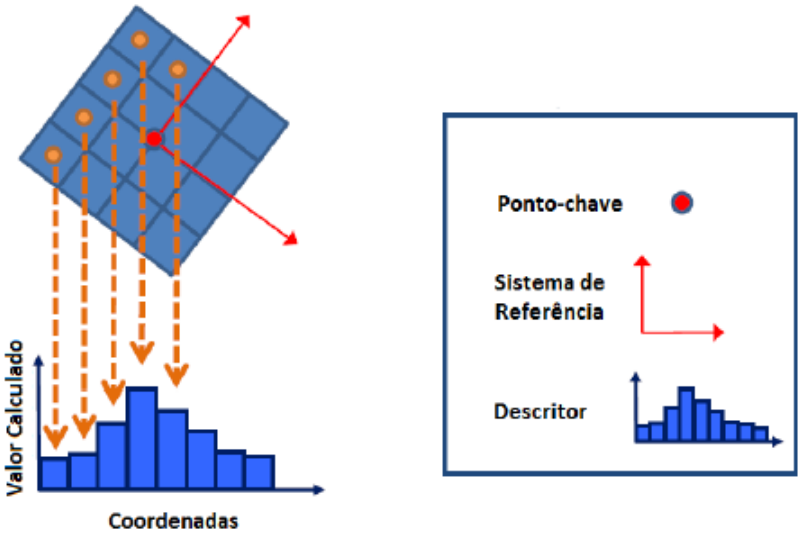
Os descritores globais encapsulam as informações relacionadas com cada objeto como um todo em uma mesma estrutura, sendo mais sensíveis a oclusão se comparado com as técnicas de descritores locais. Visando estas características, trabalham com maior invariância de escala e rotação, sendo então aplicados na tarefa de classificação (SALES, 2017). Os descritores $3D$ globais são ainda divididos em quatro categorias segundo (AKGUL *et al.*, 2009): descritores baseados em histogramas, baseados em transformações, baseados em visão $2D$ e descritores baseados em grafos.

3.3.1 Descritores locais baseados em assinatura

Descritores baseados em assinatura utilizam a computação das relações geométricas e que são calculadas individualmente em cada ponto da vizinhança, dado o ponto de referência. Para auxiliar neste processo, um sistema de referência é utilizado para que os pontos sejam processados de maneira correta e, também, para adicionar invariância na rotação e translação. O ponto fraco desta classe de descritores é sua sensibilidade a ruídos. No entanto, tendo um grande potencial em descritibilidade (SALES, 2017).

Existem alguns exemplos de trabalhos relacionados com descritores locais baseados em assinatura: *Point Signatures* (CHUA; JARVIS, 1997), KPQ (MIAN; BENNAMOUN; OWENS, 2010) e 3*D SURF* (KNOPP *et al.*, 2010). Por meio da Figura 29, é apresentada a composição de um descritor local baseado em assinatura.

Figura 29 – Representação dos dados em descritores locais baseados em assinatura.

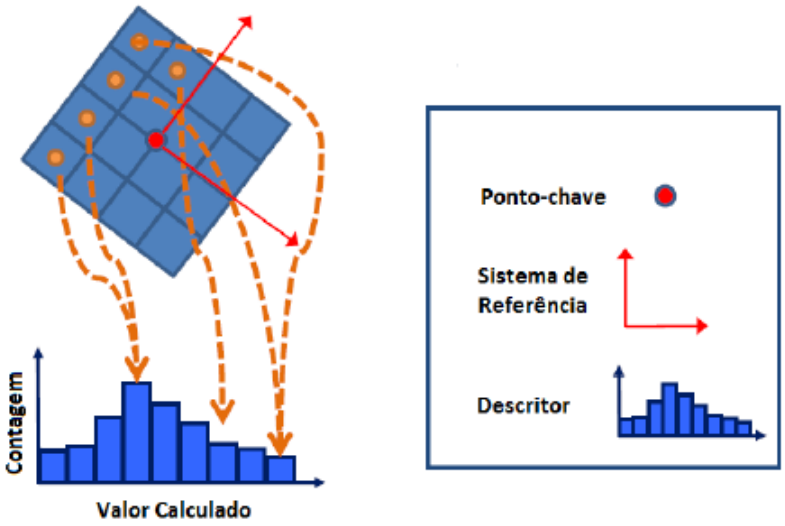


Fonte: (TOMBARI *et al.*, 2014).

3.3.2 Descritores locais baseados em histograma

Nos descritores locais baseados em histograma, dados da topologia local da imagem como vértices e arestas, são quantificados em alguma variável, como por exemplo: as coordenadas dos pontos, curvaturas e ângulos normais relacionadas as superfícies (SALES, 2017). Inversamente ao descritor apresentado anteriormente, tem uma maior robustez a ruídos e menor descritibilidade (SALES, 2017). Para uma visão melhor sobre esta classe de descritores, pode se consultada a seguinte literatura envolvendo alguns exemplos: 3*D*SC (FROME *et al.*, 2004), FPFH (RUSU; BLODOW; BEETZ, 2009) e HIGH (ZHAO; YUAN; DANG, 2014). Por meio da Figura 30, pode ser observada a representação em descritores locais baseados em histograma.

Figura 30 – Representação dos dados em descritores locais baseados em histograma.

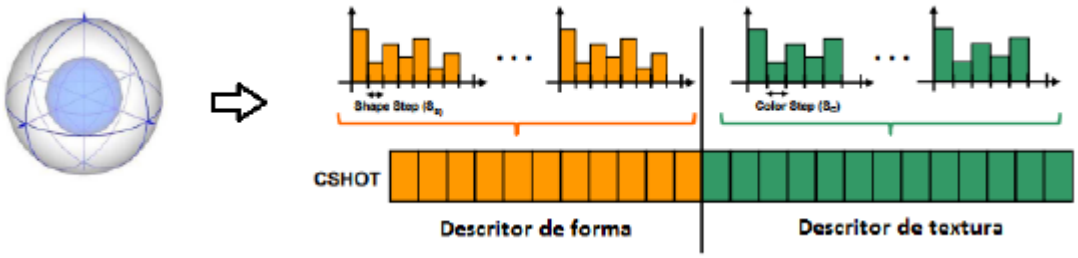


Fonte: (TOMBARI *et al.*, 2014).

3.3.3 Descritores locais híbridos

Os descritores locais híbridos combinam as duas técnicas descritas anteriormente. São destacados alguns descritores que são mais conhecidos nesta categoria (SALES, 2017): *MeshHoG* (ZAHARESCU *et al.*, 2009) e *SHOT* (TOMBARI; SALTI; STEFANO, 2010), podendo ser observados na Figura 31.

Figura 31 – Estrutura para assinatura e histograma do descritor *SHOT*.

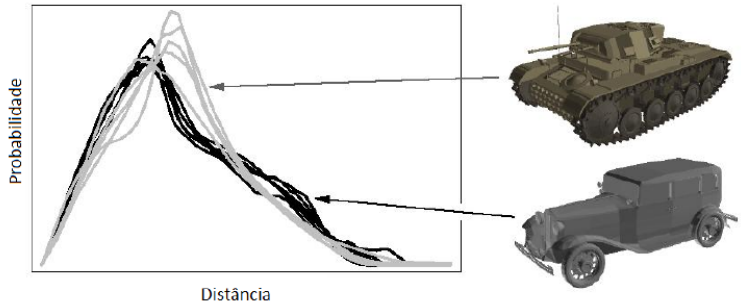


Fonte: (TOMBARI *et al.*, 2014).

3.3.4 Descritores globais baseados em histogramas

Os descritores globais baseados em histogramas detectam e armazenam uma contagem de características globais dos objetos em imagens 3*D* (SALES, 2017). É bastante robusto, porém, com menor descritividade. Como exemplo, podem ser destacados os seguintes descritores desta classe (SALES, 2017): *View point Feature Histogram* (RUSU *et al.*, 2010), *Clustered-VFH* (ALDOMA *et al.*, 2011), 3*D Shape Histograms* (ANKERST *et al.*, 1999), *Orientation Histograms* (HORN, 1984) e *Shape Distributions* (OSADA *et al.*, 2002). Este último pode ser observado por meio da Figura 32.

Figura 32 – Histograma gerado com o descritor *Shape Distributions*.

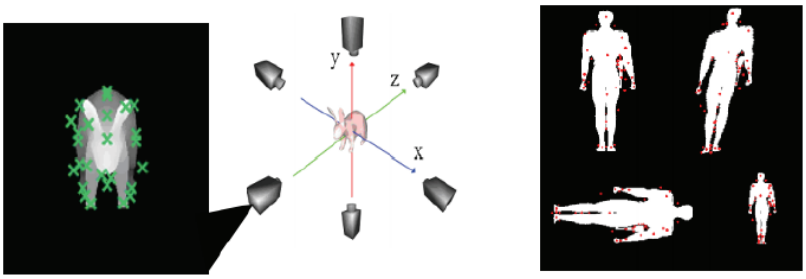


Fonte: (OSADA *et al.*, 2002).

3.3.5 Descritores globais baseados em visão 2D

Por meio deste descritor, é possível transformar a superfície de uma dada imagem 3*D* em um grupo de projeções em 2*D* de diferentes ângulos de vista. Com isso, é possível aplicar descritores 2*D* convencionais em cada uma destas imagens (SALES, 2017) resultando na possibilidade de aplicação de *Fourier descriptors* e *SIFT* (LOWE, 1999). Este último descritor é ilustrado na Figura 33.

Figura 33 – Múltiplas projeções 2*D* e uso do descritor *SIFT* nas imagens resultantes.

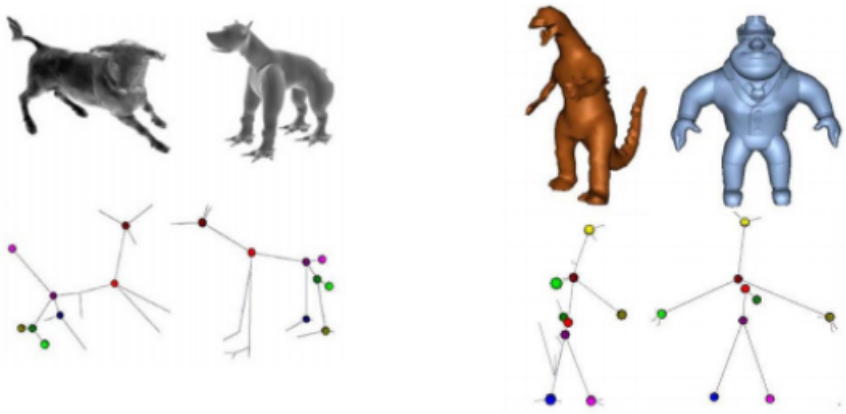


Fonte: (OHBUCHI *et al.*, 2008).

3.3.6 Descritores globais baseados em grafos

Já nos descritores baseados em grafos, a representação é feita em torno da topologia do objeto, posteriormente convertendo o mesmo em um vetor com relações numéricas entre cada nó que representa o objeto. As abordagens mais destacadas nesta classe de descritores são: *Reeb graph* (TUNG; SCHMITT, 2005), *Topology-based* (HILAGA *et al.*, 2001) e *Skeleton-based* (SUNDAR *et al.*, 2003). O último exemplo pode ser visualizado por meio da Figura 34.

Figura 34 – Dois tipos de animais representados com o descritor *Skeleton-Based*.



Fonte: (SUNDAR *et al.*, 2003).

Resumidamente, a extração de características 3*D* é uma aplicação necessária para diferentes áreas envolvendo grandes volumes de dados, principalmente onde é preciso o reconhecimento de objetos com variações de tamanho e rotação. Como no caso desta tese de doutorado, onde os sinais de trânsito devem ser reconhecidos por meio de placas verticais e que tem suas formas e cores bem definidas em imagens 2*D* e 3*D*, no entanto, com seus tamanhos e posições variáveis.

De maneira geral, um bom aproveitamento dos dados 3*D* gerados e que são capturados por sensores de última geração, devem facilitar no processo percepção de sinais de trânsito. Sendo que uma boa extração de características deve contribuir com os métodos de Aprendizado de Máquina para que trabalhem de forma adequada na solução de problemas envolvendo detecção e classificação automática e que foram envolvidos neste trabalho.

3.4 Aprendizado de Máquina

Por meio desta seção é apresentada uma visão detalhada sobre métodos de Aprendizado de Máquina voltados para reconhecimento de padrões e classificação de objetos. Os métodos apresentados são direcionados para sistemas de percepção robóticos.

Em resumo, o Aprendizado de Máquina é uma área de conhecimento muito ligada com a Inteligência Artificial e, que trabalha com algoritmos com o potencial de solucionar problemas

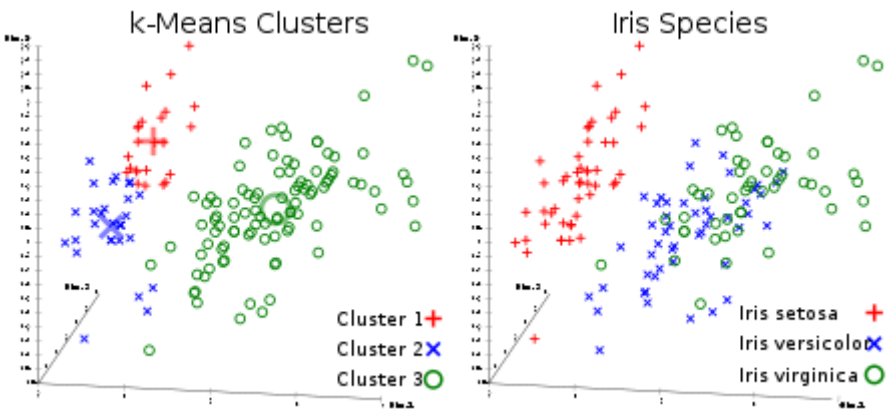
de maneira automática, utilizando como base um treinamento feito com um conjunto de dados bem definido (MITCHELL, 1997); (SALES, 2017). Os modelos de Aprendizado de Máquina são fortemente utilizados em problemas envolvendo Visão Computacional e que estão relacionados com a robótica móvel.

Uma outra visão sobre o Aprendizado de Máquina, está ligada ao reconhecimento de padrões, onde os métodos são voltados para detecção automática de regularidades, possibilitando então detectar algum padrão de interesse. Dando suporte para classificar os dados em classes bem definidas (BISHOP, 2006) ; (THEODORIDIS; KOUTROUMBAS, 2008). Os métodos de reconhecimento de padrões podem trabalhar em conjunto com métodos de Visão Computacional, assim, dando o potencial de classificar objetos de interesse para uma dada aplicação. O Aprendizado de Máquina é dividido em três categorias principais: (a) aprendizado supervisionado, (b) não-supervisionado e (c) aprendizado por reforço (MITCHELL, 1997).

3.4.1 Aprendizado não-supervisionado

No aprendizado não-supervisionado, não existe a necessidade de rotular as diferentes classes dos elementos do conjunto de dados. Para que isso seja possível, os dados são tratados por sua similaridade entre os mesmos e, para isso, é usada alguma métrica, possibilitando então separar os dados em classes bem definidas (SALES, 2017) ; (MITCHELL, 1997). O fato de não se precisar rotular um conjunto de dados é ideal para casos onde é necessário eliminar a tarefa de um especialista em dada aplicação, sendo então bastante interessante para tarefas onde esse recurso é custoso (SALES, 2017) ; (MITCHELL, 1997).

Figura 35 – *K-Means*: Agrupamento de amostras formando três diferentes classes.



Fonte: (http://www.wikiwand.com/es/K-means).

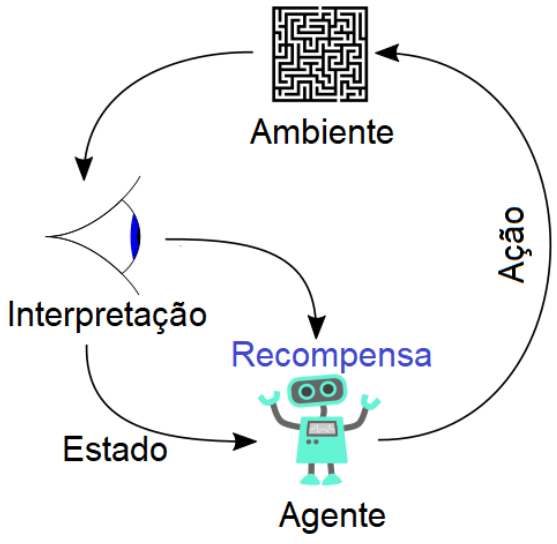
O processo é bastante simples, quando dois elementos são similares, são atribuídos a mesma classe. Caso um elemento não seja similar a nenhum outro, então é criada uma nova classe para este mesmo ser inserido (SALES, 2017) ; (KLASING; WOLLHERR; BUSS, 2008). Um exemplo para esta classe é o algoritmo *K-Means*.

O algoritmo *K-Means* trabalha com a lógica de Aprendizado de Máquina não supervisionado, baseado em partições e bastante conhecido e utilizado na computação. O algoritmo inicialmente distribui centróides que representam o núcleo de cada grupo (Figura 35), assim, associando cada amostra com seus respectivos centróides, usando para isto uma medida de similaridade entre eles. A cada nova iteração do algoritmo, o centróide é realocado em direção à média de cada conjunto e as associações de cada amostra são recalculadas. A realocação dos centróides é feita até que cada elemento não mude mais suas associações com seus respectivos núcleos (SALES, 2017) ; (MURPHY, 2012).

3.4.2 Aprendizado por reforço

No aprendizado por reforço, o agente inteligente deve desenvolver suas funções em um ambiente dinâmico e por meio de tentativa e erro. Quando uma tentativa gera uma resposta positiva, o sistema é recompensado, já quando gera uma resposta negativa, é penalizado (MAXWELL, 2010). Então, dado um modelo de aprendizado por reforço, o agente inteligente inserido em um ambiente qualquer interage com o mesmo por meio de percepções e ações (Figura 36). Este processo deve ser capaz de reforçar ou penalizar o sistema de aprendizado dado o estado atual (MAXWELL, 2010).

Figura 36 – Modelo padrão de aprendizado por reforço.



Fonte:(http://www.each.usp.br/petsi/?p=8049.)

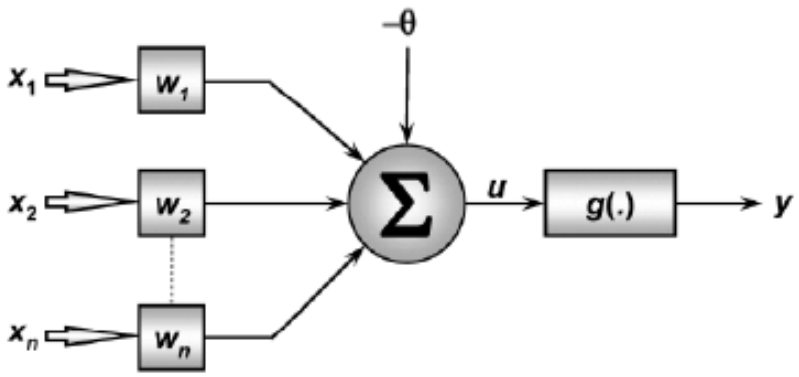
No entanto, para que seja possível que os métodos de aprendizado de máquina desenvolvam suas funções de maneira eficiente e possibilitando o aproveitando máximo do seu potencial, é necessário também que os dados sejam capturados de maneira precisa no ambiente em questão. Para que isso seja possível na área de visão robótica, são utilizados variados tipos de sensores $2D$ e $3D$ baseados em câmeras e LIDARs.

3.4.3 Aprendizado supervisionado

No aprendizado supervisionado, um treinamento é feito com base em um conjunto de exemplos rotulados, ou pré-classificados, para uma quantidade finita de classes conhecidas *a priori*. O treinamento deve ser capaz de fazer o sistema trabalhar generalizando seu conhecimento prévio para outras situações. Para um problema de valores discretos, existe um problema de classificação, já para um problema com valores contínuos, é gerado um problema de regressão.

O treinamento é feito por meio de pares de valores, sendo estes entradas e saídas. Por meio de um valor de entrada, é dada a saída esperada. Se a saída esperada não for alcançada, um novo ajuste é feito e uma reestruturação do sistema é feita (Figura 37). O ajuste do erro é feito até que a saída calculada esteja dentro de um valor esperado (MITCHELL, 1997) ; (SALES, 2017).

Figura 37 – Exemplo do modelo matemático do neurônio artificial.

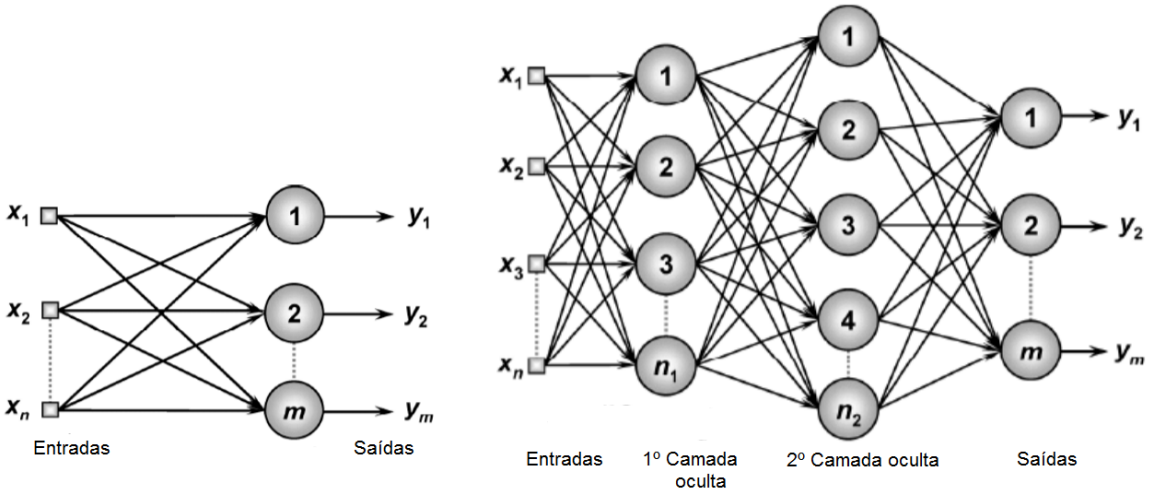


Fonte: (SILVA; SPATTI; FLAUZINO, 2010) e (LUDWIG; MONTGOMERY, 2007).

Uma RNA pode ser treinada por meio de um algoritmo e um conjunto de dados rotulados. O treinamento de uma RNA nada mais é do que o ajuste dos pesos das sinapses para que seja então dada a saída desejada em relação as entradas fornecidas. Este processo deve ser repetido até que um bom treinamento seja adquirido pela RNA ou um número de épocas predefinido seja alcançado (RUSSELL; NORVIG, 2003) ; (LUDWIG; MONTGOMERY, 2007).

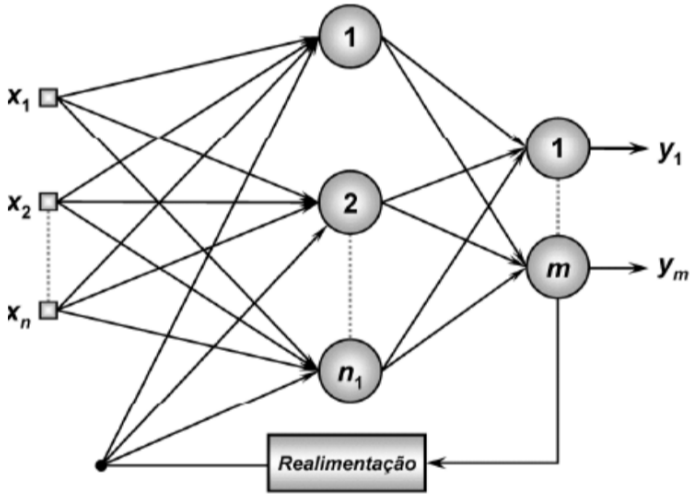
Existem diversificadas arquiteturas de RNAs e estas são definidas pela forma com que os neurônios se conectam entre si. Dependendo do problema, uma dada rede pode ser mais eficiente do que a outra. A disposição dos neurônios é então chamada de topologia, que tradicionalmente, são criadas em camadas. Uma arquitetura bastante conhecida é a baseada em *Perceptrons* (SILVA; SPATTI; FLAUZINO, 2010) ; (SALES, 2017), que pode ser organizada em diferentes camadas. Existem redes com conexões retro-alimentadas que são denominadas de Redes Recorrentes. Existem também as Redes Neurais com muitas camadas ocultas, e que têm sido denominadas de Redes de *Deep Learning* (Redes Profundas), onde destacam-se as redes CNN, conforme será discutido posteriormente (Seção 3.4.4).

Figura 38 – Arquiteturas de Redes Neurais Artificiais.



(a) Perceptron

(b) Multi-Layer Perceptron



(c) Recorrente

Fonte: (SILVA; SPATTI; FLAUZINO, 2010).

A arquitetura de uma RNA *Perceptron*, também conhecida como *single-layer feed-forward*, é o modelo de rede mais simples e básica da literatura, possuindo apenas uma camada de entrada e outra de saída, não tendo então nenhuma camada oculta (Figura 38-(a)). Sua aplicação é bastante limitada, conseguindo resolver apenas problemas lineares (SILVA; SPATTI; FLAUZINO, 2010) ; (SALES, 2017) ; (LUDWIG; MONTGOMERY, 2007).

Outra arquitetura bastante conhecida é a *Multi-layer Perceptron* (MLP) ou *multi-layer feedforward*. Esta arquitetura nada mais é do que uma rede *Perceptron* com adição de camadas ocultas (Figura 38-(b)), assim, aumentando sua capacidade para resolução de problemas lineares, possibilitando então, um aumento da precisão dos resultados e, também, estendendo suas aplicações para problemas não-lineares (HAYKIN, 1998).

Por último, existe a topologia recorrente (Figura 38-(c)), que utiliza um laço de *feedback*, realimentando a RNA em sua entrada ou em alguma camada anterior. Com isso, é possível trabalhar com uma memória de curto prazo em suas aplicações (HAYKIN, 1998).

Em resumo, as RNAs são capazes de trabalhar com robustez a ruídos e, também, com a imprecisão dos dados de entrada. No entanto, sua característica positiva de maior importância está ligada com sua grande capacidade de generalizar o conhecimento obtido no treinamento para diversificadas situações encontradas no futuro. Esse potencial de generalização é possível mesmo que estes novos casos gerados nas entradas, não tenham sido abordados explicitamente no momento do treino do sistema neural (HAYKIN, 1998) ; (SALES, 2017). Outro ponto importante, é que uma RNA pode ser treinada *offline*, já que esta etapa é a que mais consome tempo. Possibilitando então aplicar o algoritmo de execução e com os pesos já treinados na fase de teste e, na fase de processo, reduzindo o custo de tempo. Sendo então uma ótima escolha para casos onde se exige baixo tempo de execução (SILVA; SPATTI; FLAUZINO, 2010) ; (SALES, 2017).

Dado o referencial apresentado, neste trabalho de doutorado, serão utilizados dois tipos de RNAs em diversificadas tarefas: (a) detecção de objetos em imagens $2D$ e $3D$, tendo então uma saída binária dizendo se dado objeto é um sinal de trânsito vertical ou não e (b) classificação dos sinais de trânsito, informando para o veículo robotizado qual o tipo da informação que foi detectada. Sendo estas redes supervisionadas do tipo perceptron (basicamente sem recorrência) perceptron de múltiplas camadas ou profundas. Para este último caso, será aplicada uma rede neural profunda - *Deep Learning*, assim, aproveitando sua alta capacidade em reconhecimento de imagens $2D$.

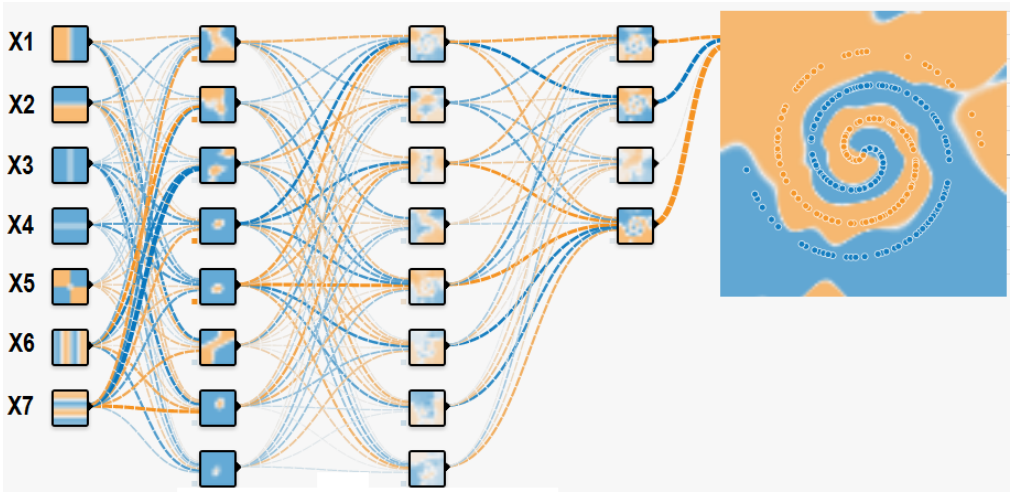
3.4.4 Deep Learning

Depois desta explicação mais detalhada sobre RNAs, existe então uma melhor base para o entendimento das redes de *Deep Learning* e que também foram utilizadas neste trabalho de doutorado para classificação e detecção em imagens $2D$. Estas redes são uma subdivisão das RNAs que se caracterizam por suas múltiplas camadas ocultas em sua estrutura (Zhou; Deng, 2014). O processo de aprendizado, com este tipo de estrutura de RNA, é chamado atualmente como aprendizado profundo.

Em aprendizado profundo, cada camada da rede deve receber em sua entrada a saída do neurônio anterior, possibilitando então uma maior robustez para a tarefa de classificação (Figura 38-(b)). Visando que este tipo de rede possibilita trabalhar com um maior volume de dados se comparada com as RNAs clássicas.

Os modelos mais conhecidos das redes de *Deep Learning*, são encontrados na literatura por *Convolutional Neural Networks* (CNN), sendo que também foram desenvolvidas e baseadas na computação bioinspirada do cérebro humano. Este tipo de rede foi desenvolvida principalmente para tratar problemas de classificação de imagens digitais, utilizando como base a fisiologia do córtex visual humano. As CNNs foram inicialmente propostas por (FUKUSHIMA, 1980) e (LECUN *et al.*, 1998), onde em seus trabalhos foram apresentados as vantagens deste tipo de arquitetura em relação a outras técnicas de classificação de imagens. Em meados de 2011, as técnicas baseadas em CNNs, começaram a ganhar a maioria das competições (como por exemplo, o *ImageNet*), disputando contra outras técnicas de Aprendizado de Máquina.

Figura 39 – Exemplo de uma arquitetura de rede neural usada pelo *Tensorflow*.



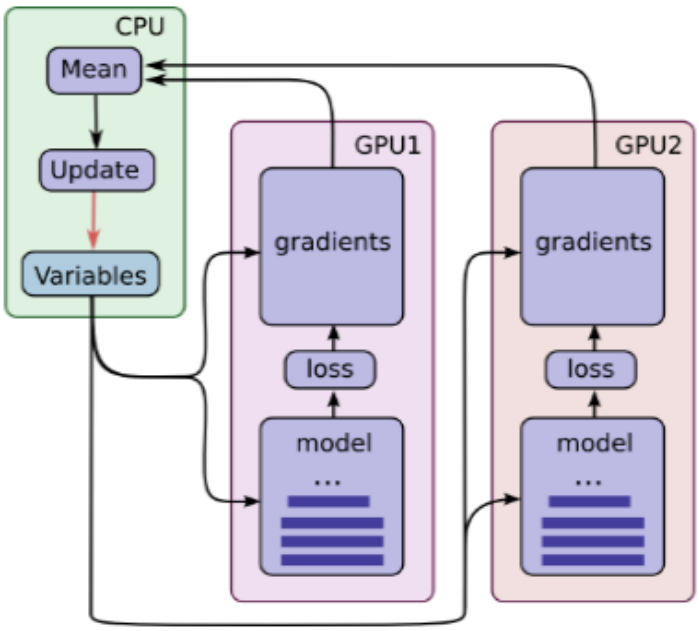
Fonte: (ABADI *et al.*, 2015).

Para esta pesquisa de doutorado, uma arquitetura de *Deep Learning* baseada na estrutura neural da Google no modelo *Inception-V3* foi aplicada. Os algoritmos neurais desenolvidos em *Tensorflow*, implementados na famosa *GoogLeNet*, utilizam muitas camadas ocultas, sendo então o que definem estes tipos de redes "profundas". O número de camadas também pode ser otimizado de acordo com o problema em questão (Bruno; Osorio, 2017).

Na Figura 39, pode ser observado um exemplo de uma rede otimizada no *playground* do *TensorFlow* (ABADI *et al.*, 2015) e, que foi construída com 1 camada de entrada e 3 camadas ocultas, contendo neurônios [7x8x8x4], respectivamente. Este modelo foi aplicado para uma classificação do problema da dupla espiral (um problema clássico na aprendizagem neural) (ABADI *et al.*, 2015).

O *TensorFlow* pode trabalhar com *Graphics Processing Unit* (GPU) em paralelo ou usando apenas *Central Processing Units* (CPU). Cada GPU pode calcular os gradientes para diferentes conjuntos de imagens (ABADI *et al.*, 2015), acelerando o treinamento de grandes conjuntos de dados (Figura 40). Isso possibilita o treinamento de uma arquitetura neural com melhor distribuição de carga e otimização do conjunto de dados para treinamento. A rede treinada final geralmente pode ser usada para classificar novos padrões em aplicações em tempo real.

Figura 40 – Configuração da arquitetura de *hardware* usada pelo *Tensorflow* no reconhecimento de imagens.



Fonte: (ABADI *et al.*, 2015).

A biblioteca de *software* de código aberto para Aprendizado de Máquina, *TensorFlow* (ABADI *et al.*, 2015), é amplamente utilizada hoje para aplicações de reconhecimento de imagens. Esta ferramenta foi aplicada na classificação de imagens encontradas na Internet (*ImageNet Competition*), e até mesmo para o reconhecimento de objetos em ambiente real, usado para auxiliar pessoas com deficiência visual. O *TensorFlow* foi desenvolvido para o Aprendizado de Máquina, focando principalmente em pesquisas para redes neurais profundas. No sistema desenvolvido neste trabalho de doutorado, as imagens são passadas para um modelo de aprendizado profundo, adotando uma rede *Deep Convolutional Neural Network / ConvNet* (DCNN) implementada usando o *framework TensorFlow*. Sendo que esta ferramenta foi utilizada em conjunto com a rede *Inception-V3* para classificar os sinais de trânsito previamente detectados pelo modelo de percepção desenvolvido.

O treinamento usado neste trabalho foi aplicado na camada superior de uma CNN, com base na rede *Inception-V3* (ABADI *et al.*, 2015), onde 70% do conjunto de dados foi utilizado para treinar a rede (adaptar os parâmetros de rede) (ABADI *et al.*, 2015). Esta arquitetura de *Deep Learning* possui em sua base de treino 1000 classes de objetos e, para o reconhecimento dos sinais de trânsito verticais, aplicou-se a técnica de *Transfer Learning*, assim, inserindo novas classes na base conhecimento. Essa técnica é abordada no Capítulo 4, onde é feita também uma explicação geral sobre o modelo de percepção desenvolvido.

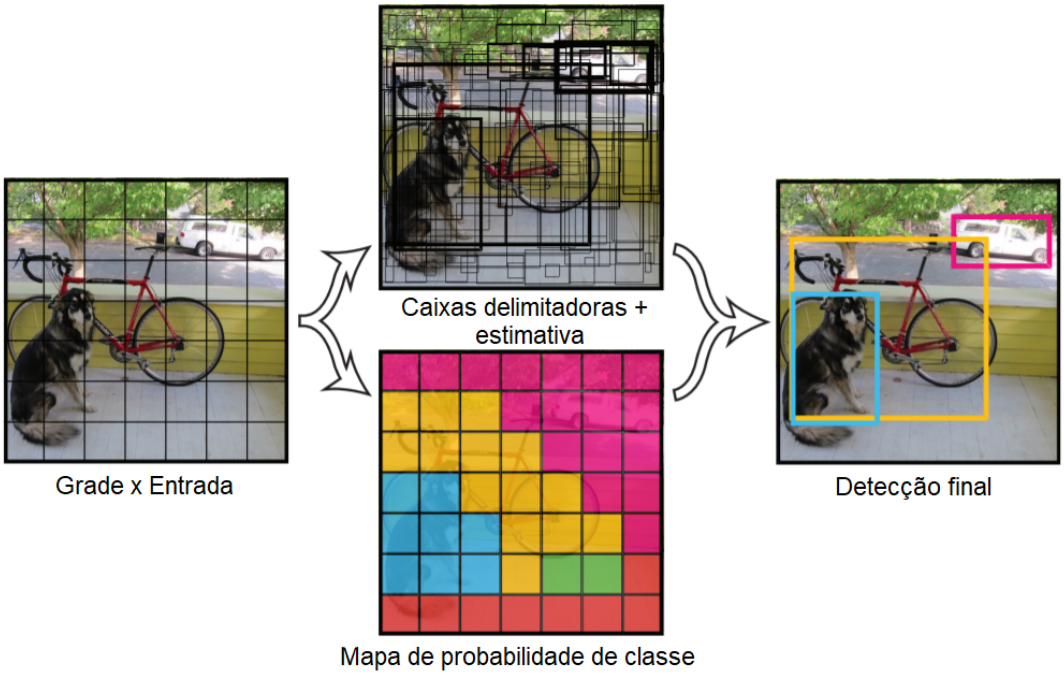
Outro exemplo de rede de *Deep Learning* é a YOLO (*You Only Look Once*) (REDMON *et al.*, 2016) ; (REDMON; FARHADI, 2018), apresentando uma nova abordagem para detecção e classificação de objetos em imagens $2D$ (Figura 41). A estruturação para detecção de objetos é definida como um problema de regressão para caixas delimitadoras separadas espacialmente com probabilidades de classes classificadas associadas. A rede YOLO é capaz de prever caixas delimitadoras juntamente com probabilidades de classes diretamente de imagens completas em uma única avaliação. Sendo que toda a detecção é feita em uma rede única, podendo ser otimizada diretamente no desempenho da detecção. Destacando também que esta rede é capaz de extrair características de maneira automática de objetos $2D$, sendo esta uma forte característica das redes de *Deep Learning* (REDMON *et al.*, 2016).

Este tipo de rede é capaz de trabalhar em 45 quadros por segundo. Uma versão menor da rede, *Fast* YOLO, processa surpreendentes 155 quadros por segundo. Se comparada aos sistemas de detecção atuais, a rede YOLO comete mais erros de localização na detecção, mas é menos provável prever falsos positivos, aprendendo muito bem as representações gerais de objetos.

A arquitetura da rede YOLO é inspirada no *GoogLeNet*, sendo este um modelo para classificação de imagens. Em seu total a rede possui 24 camadas convolucionais seguidas por 2 camadas totalmente conectadas (REDMON *et al.*, 2016).

Nesta tese de doutorado, foi aplicada a rede YOLO para detecção de sinais de trânsito em imagens $2D$.

Figura 41 – Sistema de detecção baseado na rede de *Deep Learning* YOLO.



Fonte: (REDMON *et al.*, 2016).

3.5 Atenção Visual

Comparado aos conceitos de visão humana, os modelos de Atenção Visual Artificiais também visam detectar regiões de interesse em imagens digitais. Esta área envolve diferentes ramos da pesquisa, como por exemplo: Psicólogos, neurobiologistas, engenheiros e cientistas da computação que trabalham atualmente com visão e cognição, focando em analisar e desenvolver modelos de Atenção Visual (FRINTROP; ROME; CHRISTENSEN, 2010).

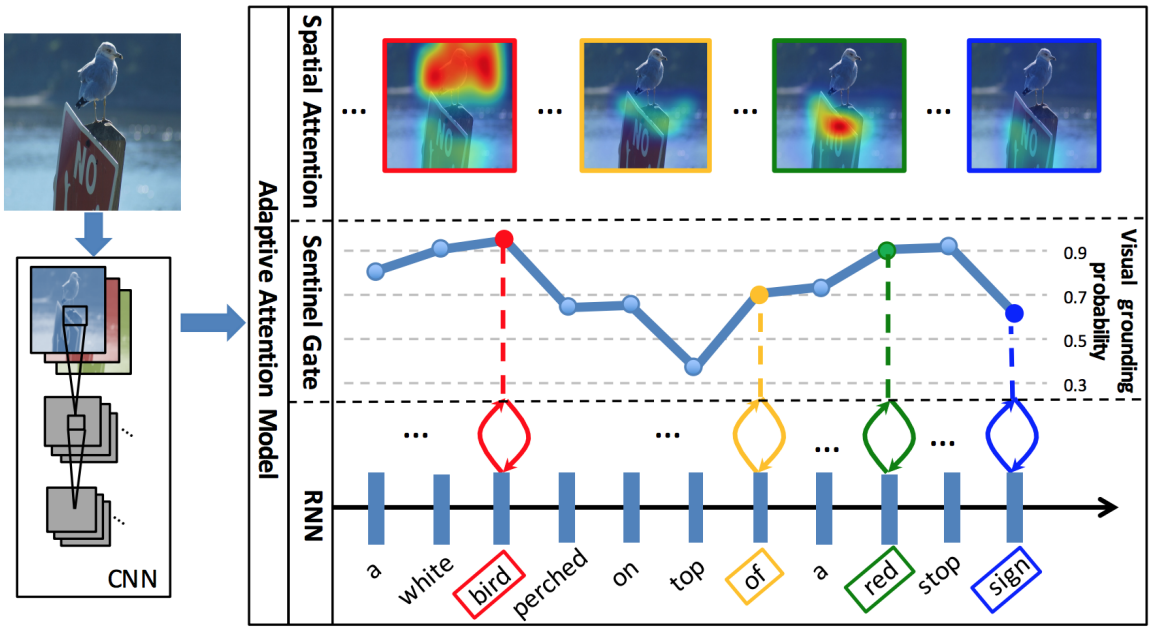
Por meio da Figura 42, pode ser observado um modelo de Atenção Visual baseado em *Deep Learning* que é capaz de fazer uma análise semântica na imagem e descrever a construção da cena com seus elementos (GRIGOREV *et al.*, 2018).

Estes modelos de Atenção Visual são direcionados para que seja possível uma interpretação visual da cena, indo muito além de detecções e classificações (GRIGOREV *et al.*, 2018). Na Figura 42, pode ser observada uma situação onde uma ave foi detectada em cima de um sinal de trânsito de parada obrigatória.

Para que a análise da cena seja possível, alguns modelos de Atenção Visual utilizam como suporte uma *Long Short-Term Memory* (LSTM), que é capaz de produzir as legendas das imagens condicionada ao vetor de contexto da cena. Por meio da LSTM, também é possível estimar o estado oculto atual da imagem com base em informações observadas anteriormente (XU *et al.*, 2015).

Por fim, é dado que a LSTM visa dar suporte no processo de ponderar os pesos de cada região da imagem, possibilitando então declarar quais objetos foram detectados e realizar a organização semântica da cena.

Figura 42 – Exemplo de Atenção Visual em imagens 2*D*.



Fonte: (GRIGOREV *et al.*, 2018).

No entanto, as abordagens que envolvem redes de *Deep Learning* para interpretar a cena, exigem um grande volume de dados para treinamento, necessitanto de imagens com muitas variações de contexto e, que em alguns casos, seria muito difícil de ser tratado. Como exemplo, pode ser destacado as situações de emergência na pista (obras, acidentes, desvios) e que atualmente não existem grandes *datasets* voltados para este tipo de problema.

3.6 Sensores

Os sensores são elementos essenciais na robótica móvel, já que por meio deles é possível fazer uma leitura do ambiente de navegação e, por meio destas leituras, gerar informações para o controlador realizar suas rotinas. Um sensor é um equipamento que recebe um tipo de sinal em sua entrada (imagem, temperatura, distância) e transforma em uma variável elétrica do tipo manipulável, enviada em sequência para o controle, sendo então mais fácil de ser tratada (MURPHY, 2012).

Basicamente existem dois tipos de sensores (HABERMANN *et al.*, 2013), os ativos e os passivos: (a) os sensores ativos emitem um sinal de energia e mensuram seu retorno, (b) já os sensores passivos, trabalham apenas com a captação da energia disponível no ambiente em questão (SIEGWART; NOURBAKHSH, 2004). A seguir serão descritos os principais sensores que estão ligados a este trabalho de doutorado.

3.6.1 Câmeras de vídeo 2D

Um sensor bastante eficiente na área de robótica são as câmeras de vídeo. Este tipo de sensor apresenta um grande volume de dados e que podem ser manipulados, assim, gerando um conjunto de informações de grande potencial para o sistema de percepção robótico (MATARIC, 2014) ; (ROMERO *et al.*, 2014).

Quando um robô utiliza uma câmera em conjunto com métodos de Visão Computacional, é possível então que ele consiga coletar informações do ambiente de grande importância. Os animais, na maioria das vezes, também utilizam a visão como seu principal sensor, possibilitando caminhar e reconhecer objetos encontrados em seu caminho. Bem próximo disso, ficam então os sistemas de Visão Computacional Robóticos. Porém, o tratamento de imagens é bastante custoso se comparado a outros sensores (ultrassom, laser e de contato), contudo, possibilitam um maior nível de informação da cena.

Os métodos de Processamento Digital de Imagens devem ser aplicados então para tratar as imagens capturadas pelas câmeras, já que sem um processamento destas informações, não é possível que o robô consiga algum conhecimento sobre as imagens "puras". Um exemplo para este problema, seria um robô que deve reconhecer alguns obstáculos em sua rota, para que isso seja possível, é necessário primeiramente extrair informações das imagens capturadas e, posteriormente, levar estes dados para um classificador que deve declarar qual o tipo de obstáculo detectado.

Um grande problema no uso de câmeras está muito ligado com a iluminação do ambiente, já que este tipo de sensor é bastante sensível a variação de luz, pois são passivos, ou seja, são dependentes da luz que existe no ambiente para que seja possível capturar a imagem. Algumas câmeras também são dotadas de sistemas ativos, emitindo um sinal sobre o ambiente e, que na maioria dos casos, é infravermelho e invisível para o olho humano.

Neste trabalho de doutorado foram utilizadas duas formas de imagens obtidas por meio de câmeras: (a) **Imagens 2D**, onde não é possível obter informações de profundidade, porém, informações de cores e texturas são bastante evidentes e (b) **Imagens 3D**, onde é possível trabalhar com a noção de profundidade graças a fusão das informações de duas câmeras (esquerda e direita), assim como funciona a visão humana, conseguindo então estimar com precisão a posição dos objetos detectados em relação ao mundo.

3.6.2 Sonares

Os sonares trabalham de maneira ativa, ou seja, geram uma emissão de um sinal sonoro de ultrassom, possibilitando medir o tempo que o som demora para sair do transmissor e voltar para o receptor depois que um obstáculo foi detectado. Com isso, é possível calcular a distância que o obstáculo foi detectado na rota de navegação do robô.

A precisão da medição é altamente dependente do ângulo em que as ondas atingem o obstáculo, já que quanto mais agudo este ângulo, atrapalham a reflexão deste sinal em relação ao receptor, dificultando então a leitura dos dados reais. Por outro lado, é um sensor de simples uso e de baixo custo, ideal para aplicações onde não se exige muita precisão. Os LIDARs (*Light Detection And Ranging*) apresentados na próxima seção (3.6.3), trabalham com a mesma lógica de funcionamento, trocando o sinal de portadora de som por luz.

3.6.3 Sensores 3D: LIDARs e câmeras de vídeo 3D

Em algumas aplicações, apenas as informações geradas por sistemas de sensoriamento $2D$ não são suficientes, principalmente pelo fato de que este formato de imagens gera dados que são bastante vulneráveis a erros relacionados com problemas de variação de iluminação e a falta de informações de profundidade.

Um dos problemas mais complexos na área de veículos robóticos, está relacionado com o reconhecimento de obstáculos, onde nem sempre um sistema de visão em $2D$ consegue resolver problemas semânticos sem as informações de profundidade. Gerando problemas de falsos positivos e falsos negativos, assim, detectando situações falsas e deixando de detectar objetos que deveriam ter sido detectados, respectivamente.

Por meio da Figura 43-(a), é apresentada uma situação de falso positivo, onde o sistema de visão em $2D$ detecta um semáforo vermelho, sendo este apenas uma propaganda de um *fast-food*. Já na Figura 43-(b), é apresentado um sinal de trânsito falso colado na traseira de um ônibus que serve apenas para o controle de velocidade do veículo em particular. Este tipo de problema é bastante complexo de ser solucionado em dados $2D$, pois não trabalha com informações de profundidade dos elementos da cena, consequentemente considerando todas as detecções no mesmo plano da imagem. Com o uso de dados $3D$ seria possível identificar os níveis de profundidade de cada objeto na cena, possibilitando eliminar estes problemas.

Hoje existe uma grande quantidade de sensores que tem a capacidade de trabalhar com dados 3*D*, sendo os principais tipos: a laser 3*D*-LIDAR (p.ex. *Velodyne HDL32*), câmera estéreo, sensores *Time-of-Flight* (TOF) e de Luz Estruturada.

Figura 43 – Problemas na detecção e classificação de sinais de trânsito (a) detecção de um semáforo em uma propaganda de *fast-food* e (b) um sinal de trânsito que indica a limitação de velocidade do ônibus e não da via.

Fonte: (MIT, 2018).

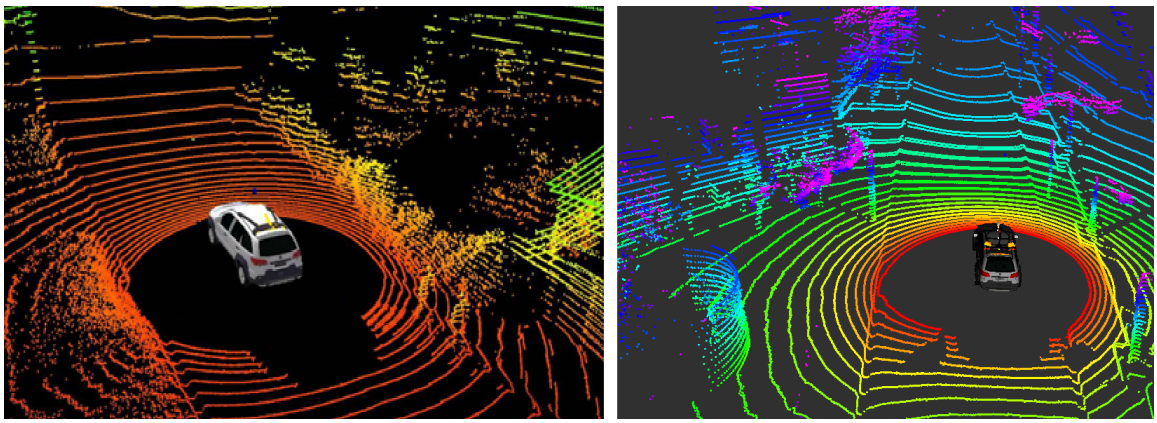
Particularmente os sensores a laser, como por exemplo, o 3*D*-LIDAR (*Velodyne HDL32*), estão trazendo resultados de grande potencial para a visão robótica, sendo então um dos sensores mais utilizados atualmente na área de veículos robóticos. Este sensor trabalha com múltiplas camadas de feixes de laser, permitindo criar mapas do ambiente de navegação com um volume considerável de informação (Figura 44). Também possibilitando detectar objetos de interesse em ambientes de trânsito que sofrem com problemas físicos de iluminação, fumaça, neblina e outros tipos.

O principal diferencial dos sensores 3*D*, se comparado com sensores 2*D*, está relacionado com a noção de profundidade. Possibilitando eliminar problemas relacionados com falsos positivos (Figura 43) e gerar dados de posicionamento 3*D* de cada objeto detectado na cena. Sendo que esta última informação é de grande importância para que seja possível interpretar os sinais de trânsito verticais em função da navegação do veículo.

Um outro tipo de sensor 3*D* é a câmera de luz estruturada, onde se utiliza uma técnica de visão estéreo ativa que projeta uma dada sequência conhecida de padrões nos elementos da cena, capturando então a deformação geométrica da forma destes padrões, possibilitando gerar a imagem equivalente 3*D* por meio de um par de câmeras combinadas. Para que isso seja possível, é utilizada uma câmera em posição diferenciada em relação ao meio de projeção (SARBOLANDI; LEFLOCH; KOLB, 2015). O mapa de disparidade é então gerado por meio da análise das distorções nos padrões da imagem.

Em resumo, imagens RGB (cores) + D (profundidade - (*Depth*)), podem ser adquiridas utilizando sensores 3*D* baseados em pares de câmeras estéreo ou LIDARs, possibilitando manipular informações com base nos dados de profundidade que representam os objetos encontrados no ambiente em questão. A noção de profundidade deve então contribuir fortemente para um sistema de percepção, que visa detectar alguns objetos em específico e relacioná-los com a navegação atual do veículo, como no caso desta pesquisa de doutorado, os sinais de trânsito verticais.

Figura 44 – Sistema de sensoriamento em nuvem de ponto do 3*D*-LIDAR (*Velodyne HDL32*).



Fonte: (lrm.icmc.usp.br/).

Por meio da (Figura 45-(b)) pode ser observada a imagem em profundidade gerada pelo Kinect da Microsoft. Esta imagem em 3*D* (Figura 45-(b)), é gerada por meio de um emissor de infravermelho. Já a imagem da Figura 45-(a), é gerada por um par de câmeras estéreo. Graças a fusão de imagens 2*D* (RGB) e, de profundidade (3*D-depth*), é possível então uma representação da cena em $RGB + D$, garantindo o aproveitamento de cores, texturas e, também, a noção de profundidade da cena.

Como conclusão parcial sobre sensores, a representação da cena com fusão de imagens 2*D* e 3*D* possui grandes vantagens sobre a representação em imagens puramente 2*D*, já que geram dados mais detalhados para o sistema de percepção envolvendo detecção e classificação. Alguns dispositivos, como por exemplo, o 3*D*-LIDAR (*Velodyne HDL32*), trabalham muito bem sob condições de pouca ou nenhuma iluminação, pois são ativos e independentes da iluminação natural ou artificial do ambiente, possibilitando uma melhor qualidade na detecção. Já as imagens 2*D* garantem uma melhor capacidade para a classificação de objetos com base nas informações de cores e texturas.

Figura 45 – Relação de profundidade das imagens do Kinect (a) imagem em RGB e (b) a mesma imagem em profundidade (o Kinect é mais adequado para ambientes internos).

Fonte: (SILBERMAN *et al.*, 2012).

3.7 Considerações

Por meio deste capítulo, foram apresentados modelos, métodos, técnicas, ferramentas computacionais e, também, alguns dos sensores $2D$ e $3D$ que foram utilizados junto ao estado-da-arte para detecção e classificação de sinais de trânsito verticais. De maneira geral, por meio deste levantamento de técnicas e métodos, foi possível destacar que as imagens $2D$ apresentam grandes limitações que estão muito ligadas com os problemas de variação de iluminação, variação do clima (neblina, neve e noite) e, também, de detecção de falsos positivos e falsos negativos, visto que câmeras monoculares fornecem apenas uma representação planar (x, y) do espaço de interesse observado.

Por meio das imagens $RGB + D$ (cores + profundidade) e, que são obtidas por meio de sensoriamento com fusão de dados $2D$ e $3D$, é possível analisar as cores, formas, texturas e profundidade dos objetos no ambiente em questão. Este potencial deve facilitar as tarefas de percepção envolvendo detecção e classificação de sinais de trânsito verticais. A representação do mundo com informações de profundidade tem uma maior robustez, já que possibilitam uma melhor descrição com um maior volume de informações do ambiente de navegação. Outra vantagem de imagens provenientes de sensores $3D$, é que seu funcionamento em LIDAR é satisfatório mesmo em condições de iluminação precária, já que são ativos na tarefa de percepção.

A fusão de dados $3D$ proveniente do sensoriamento em câmeras e LIDARs permite uma maior robustez na percepção, complementando pontos negativos em ambas as técnicas, visto que os dados destes sensores sofrem com problemas de variação de iluminação e baixa densidade na aglutinação de pontos que representam uma imagem, respectivamente.

Além do sensoriamento, também foram apresentados diversificados métodos de extração de características e Aprendizado de Máquina, destacando os principais que foram aplicados nesta pesquisa. Os métodos apresentados são direcionados para técnicas de extração de características $3D$, detecção de objetos com fusão de dados $2D$ e $3D$ e classificadores baseados em redes de *Deep Learning*. Possibilitando automatizar o processo de detecção, classificação e, também, de análise de sinais de trânsito verticais informados por meio de placas, cones e semáforos.

Destacando que as redes de *Deep Learning* trabalham muito bem como detectores e classificadores em imagens com dados de cores, formas e texturas provenientes de imagens $2D$. Já as imagens $3D$, possibilitam por meio dos seus dados de profundidade, validar os objetos detectados e classificados em imagens $2D$, filtrando então falsos positivos e falsos negativos. Sendo que a fusão destas imagens foi fundamental para o desenvolvimento desta tese de doutorado.

CAPÍTULO

4

DETECÇÃO E CLASSIFICAÇÃO DE SINAIS DE TRÂNSITO VERTICAIS COM FUSÃO DE DADOS 2D E 3D

Neste capítulo, serão apresentadas as técnicas e métodos que foram desevolvidos e adaptados para que seja possível a detecção e classificação de sinais de trânsito verticais em ambientes com regras de trânsito. O capítulo é dividido em duas seções principais: (a) Detecção de sinais de trânsito com fusão de dados $2D$ e $3D$ (Seção: 4.2) e (b) Classificação de sinais de trânsito em dados $2D$ (Seção: 4.4).

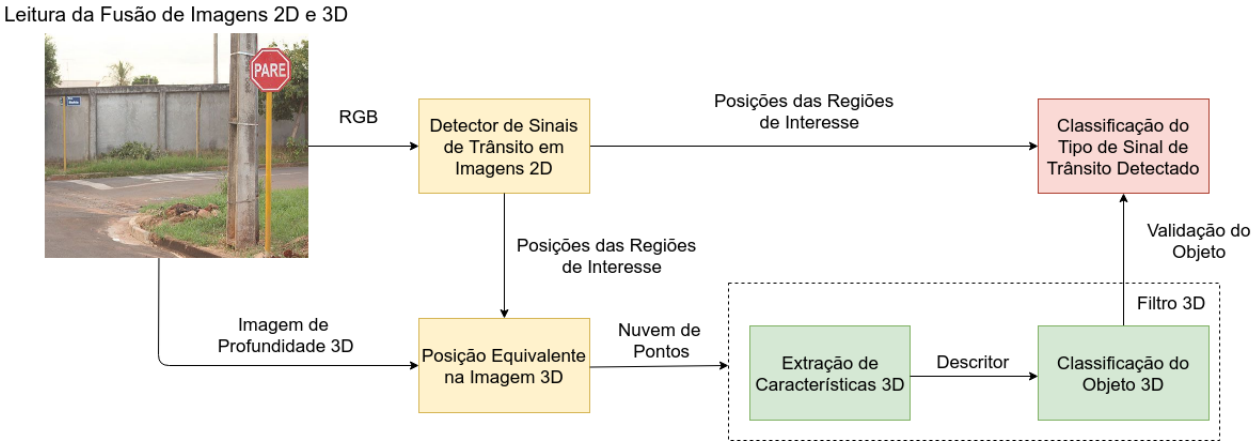
4.1 Configuração do Sistema de Visão Computacional Proposto

Por meio do fluxograma da Figura 46, é possível observar a comunicação de cada módulo do sistema de Visão Computacional desenvolvido. A entrada do sistema é gerada pela fusão de imagens $2D$ e $3D$. Visando a manipulação destas imagens, é possível detectar, filtrar e classificar o tipo de sinal de trânsito encontrado na cena (Bruno *et al.*, 2018a) ; (BRUNO *et al.*, 2019).

As etapas entre a detecção de um conjunto de imagens e a sua classificação correspondente, são apresentadas a seguir e, nas seções seguintes, uma descrição de cada método aplicado também é feita.

Já o modelo de Atenção Visual desenvolvido e, que utiliza os dados provenientes da percepção com fusão de imagens $2D$ e $3D$, é apresentado no Capítulo 5.

Figura 46 – *Pipeline* do sistema de detecção e classificação proposto.



Fonte: Elaborada pelo autor.

- **1) Detector de Sinais de Trânsito em Imagens** $2D$ **- (Seção 4.2-a)**

  Nesta etapa, a rede de *Deep Learning You Only Look Once* (YOLO) (REDMON; FARHADI, 2018) é aplicada para a detecção de sinais de trânsito em imagens $2D$. Por meio desta aplicação, é possível detectar os sinais de trânsito e suas regiões de interesse *Bounding Box* e, por meio destas regiões, dar suporte para filtrar os falsos positivos e falsos negativos com suas posiões (x, y) equivalentes nas imagens $3D$.

- **2) Posição Equivalente na Imagem** $3D$ **- (Seção 4.2-b)**

  Por meio das regiões de interesse geradas pela caixas delimitadoras na etapa de detecção, é possível obter as posições correspondentes da imagem $2D$ em sua imagem $3D$ equivalente. Trabalhando com esta equivalência de imagens, é possível avaliar o sinal de trânsito detectado em imagens $2D$ em sua imagem $3D$ correspondente, eliminando os objetos falsos positivos e falsos negativos.

- **3) Extração de Características 3D - (Seção 4.3)**

  Visando filtrar os objetos falsos em imagens $3D$, é necessário aplicar um método capaz de extrair as características de maior relevância, garantindo que seja possível verificar se o objeto detectado nas imagens $2D$ é um sinal de trânsito real ou um falso.

- **4) Classificação do Objeto** $3D$ **- (Seção 4.3.4)**

  Depois de extraídas as características, um método de Aprendizado de Máquina neural é aplicado para classificar o objeto $3D$ equivalente na região de interesse. A classificação é dividida em duas classes: (a) Sinal de trânsito ou (b) Um não sinal de trânsito. Caso seja um sinal de trânsito real, ele deve ser enviado para o próximo passo e ser classificado pelo seu tipo: velocidade, parada obrigatória, semáforo, sentido da via ou outras classes.

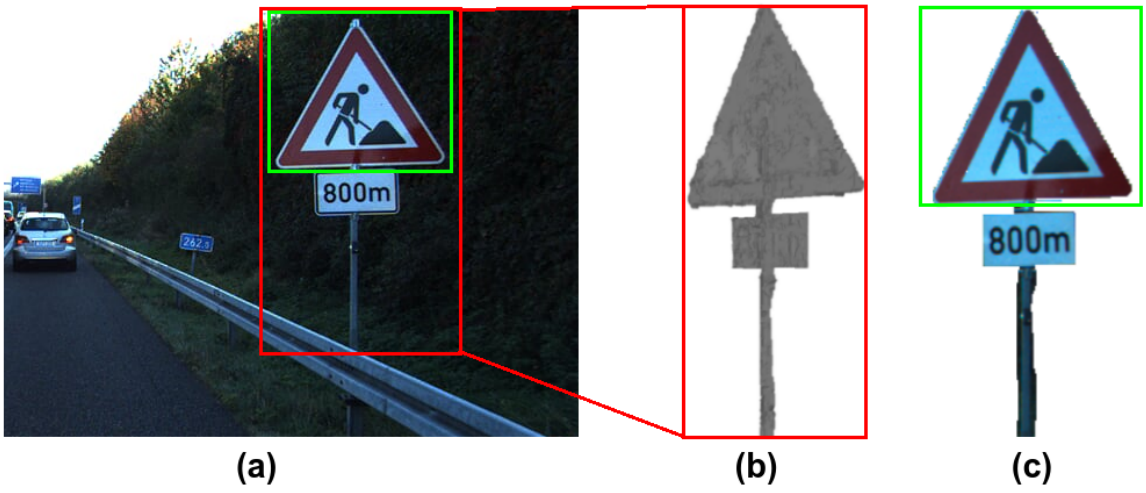
- **5) Classificação 2D do Tipo de Sinal de Trânsito Detectado - (Seção 4.4)**

  Depois de detectado em imagens 2*D* e filtrado em imagens 3*D*, o sinal de trânsito candidato a uma possível informação real, é então classificado em sua classe final. Então informando para o sistema de percepção qual a informação nele contida (velocidade, sentido da via, etc). Para que isso seja possível, foi aplicado uma rede de *Deep Learning - Inception-V3*, por sua especialidadade em trabalhar bem com cores, formas e texturas dos sinais de trânsito.

4.2 Detecção de Sinais de Trânsito em Dados 2D

Para a detecção de sinais de trânsito verticais, foi utilizada a fusão de imagens 2*D* e 3*D*, visando a detecção em longa distância e a eliminação de falsos positivos e falsos negativos, respectivamente. A detecção inicial é feita por meio da rede YOLO sobre imagens 2*D*, garantindo assim uma maior distância de percepção (aproximadamente 80 metros), já que as imagens 3*D* (câmeras e LIDARs) possibilitam apenas uma detecção em curta distância (inferior a 30 metros). O alcance na detecção está relacionado com as falhas na nuvem de pontos 3*D*, sendo estas geradas por *spray*, o que dificulta a reconstrução de um objeto e, consequentemente, prejudicando seu reconhecimento.

Figura 47 – Representação do sinal de trânsito em dados 2*D* e 3*D*. A caixa delimitadora vermelha representa a nuvem de pontos 3*D* da estrutura do sinal de trânsito detectado e a caixa delimitadora verde representa a intensidade RGB da placa de trânsito a ser classificada em 2*D*). (a) cena real (RGB + Profundidade), (b) imagem de profundidade e (c) RGB equivalente ao objeto 3*D*.



Fonte: Elaborada pelo autor.

Sendo que uma detecção 2D garante imagens melhores construídas em sua aglutinação de *pixels*, no entanto, gerando um grande volume de falsos positivos e falsos negativos por não ter a noção de profundidade, interpretando todos objetos no mesmo plano (x, y).

Visando este problema, uma vez que a detecção é feita em dados 2D (Figura 47-(a)), o próximo passo é aplicar o filtro (Seção 4.3) de falsos positivos e falsos negativos para avaliar a imagem equivalente em 3D (Figura 47-(b)). Permitindo então a filtragem dos sinais de trânsito falsos. São chamados sinais falsos aqueles que não pertencem às leis de trânsito locais, por exemplo: sinais em paredes de auto-escola, atrás de ônibus escolares, em *outdoors* e outros objetos de mesmo formato geométrico. Na Figura 47-(c), é possível observar o sinal de trânsito proveniente da fusão de dados 2D (cores e texturas) e 3D (objeto em profundidade).

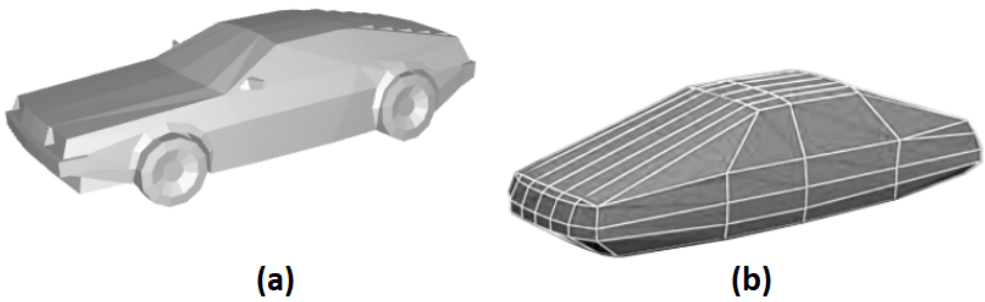
4.3 Análise e Filtragem de Sinais de Trânsito em Dados 3D

Para a tarefa de filtragem de falsos positivos, foi aplicado um filtro baseado no formato de cada sinal de trânsito. O filtro em questão utiliza extração de características 3D para que um método de Aprendizado de Máquina seja capaz de classificar o tipo de objeto encontrado nestas imagens.

4.3.1 Estimação da superfície em nuvem de pontos

O primeiro passo é estimar uma malha de superfície para a nuvem de pontos do objeto, convertendo a nuvem de pontos esparsa em um conjunto de polígonos (malha de superfície), obtendo a representação sólida do objeto de forma 3D. Para que isso seja possível, o algoritmo 3D *Convex Hull* é aplicado para essa tarefa, possibilitando que o contorno aproximado do objeto possa ser obtido.

Figura 48 – Faces geradas por: (a) triangulação e (b) *Convex Hull* 3D.



Fonte: (SALES, 2017).

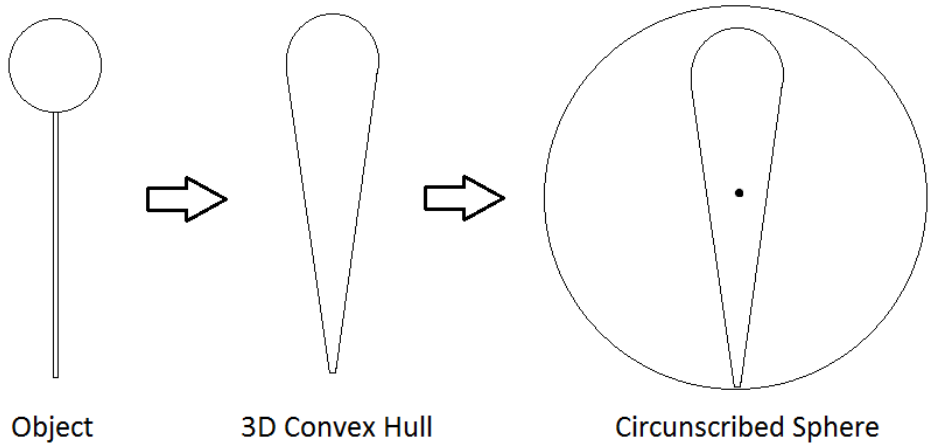
4.3.2 3D-CSD: Extração de características

O primeiro passo do 3*D-CSD* (*3D-Contour Sample Distances*) (SALES D, 2017) é trabalhar sobre uma malha de superfície para a nuvem de pontos, obtendo a representação sólida do objeto de forma 3*D*. O algoritmo 3*D Convex Hull* foi aplicado para esta tarefa, para que o contorno aproximado do objeto possa ser obtido.

O segundo passo consiste em considerar uma esfera circunscrevendo o objeto 3*D Convex Hull* (Figura 49). O centro da esfera é posicionado no centro de massa do objeto, e o raio é o ponto mais distante do objeto ao centro. Por meio da Figura 50, é apresentado um exemplo do processo realizado.

O terceiro passo consiste em selecionar os vários pontos-chave na esfera. Cada ponto chave corresponderá a uma única medida de distância e, portanto, uma posição no vetor de distâncias final (Figura 49).

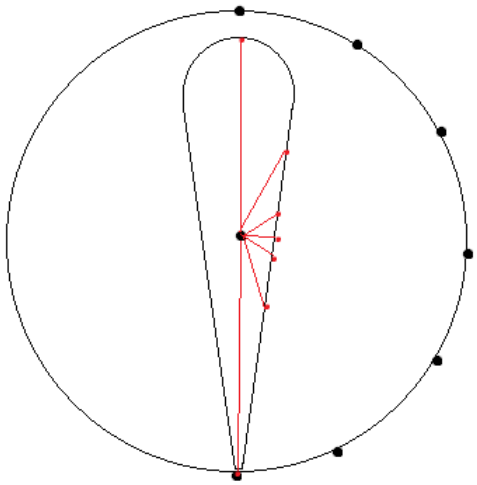
Figura 49 – Visão 2*D* do processo de geração de esfera.



Fonte: Elaborada pelo autor.

Depois de selecionar os pontos-chave, o descritor 3*D-CSD* pode ser gerado. Para cada ponto chave p_i na superfície da esfera, é considerada uma linha reta do centro de p_i para a esfera. Se esta linha reta cruzar a superfície do objeto em um ponto d_i, deve-se calcular a distância euclidiana do ponto d_i para o centro de massa do objeto. Caso contrário, devolve -1. Esse processo deve sempre ser feito sequencialmente, por exemplo, da esquerda para a direita e depois da sequência superior para a inferior. Por meio da Figura 50, é ilustrado o procedimento para medir distâncias.

Figura 50 – Exemplo de medição de distâncias do objeto.



Fonte: Elaborada pelo autor.

Por meio do Algoritmo 1, é descrita a sequência de etapas para iniciar o cálculo das distâncias para geração da estrutura de características do 3D-CSD (SALES D, 2017); (SALES; AMARO; OSóRIO, 2017).

Algoritmo 1 – Geração do conjunto das características de distâncias

1: **para todo** ponto de referência P_i na periferia da esfera **faça**
2: Considere uma reta do centro de massa do objeto C_{xyz} até P_i
3: **se** existe interseção da reta com alguma face do objeto em um ponto d_i **então**
4: **retorna** distância euclidiana no intervalo C_{xyz} e d_i
5: **senão**
6: **retorna** -1
7: **fim se**
8: **fim para**

Os pontos-chave devem ser igualmente distribuídos, ou prioritariamente, nas áreas mais representativas dos objetos. Para sinais de trânsito, uma estratégia genérica para distribuir pontos ao longo da superfície da esfera, pode ser realizada definindo um ângulo de azimute e altitude constante para a distribuição de pontos (Figura 49).

Depois que todas as medições de distâncias são feitas pelo 3*D-CSD*, os valores devem ser interpolados, isto é, normalizados, garantindo a entrada adequada para a RNA. Outro ponto importante é que esse procedimento de interpolação é a chave para fornecer invariância de escala, já que o aspecto da forma é estimado com base nas proporções de distância, em vez de medidas métricas. Todos os valores do descritor devem estar no intervalo [0..1], onde 1 é o raio da esfera.

4.3.3 Reconhecimento de Padrões 3D

Para poder reconhecer a assinatura 3*D* do objeto segmentado fornecida pelo método 3*D-CSD*, uma RNA *Multi-Layer Perceptron* (MLP) é aplicada devido às suas boas capacidades para capturar recursos subjacentes complexos e não-lineares com um alto grau de precisão (Algoritmo 2).

Por meio do Algoritmo 2, é descrito o ciclo completo de processamento para classificação de objetos 3*D* (SALES; AMARO; OSóRIO, 2017).

Algoritmo 2 – *Pipeline* de classificação de objetos

```
1: para todo conjunto de dados dado pelo sensoriamento 3D em um dado instante faça
2:     Segmentar elementos da cena 3D
3:     para todo elemento segmentado faça
4:         Compute o descritor 3D-CSD
5:         Aplica os dados do descritor na entrada da RNA
6:         retorna a classificação gerada na saída da RNA
7:     fim para
8: fim para
```

A camada de entrada da RNA deve conter o mesmo número de unidades que o tamanho do vetor das distâncias e contornos que são normalizados. A camada de saída da RNA é composta de uma única unidade, para uma saída binária, sendo: objeto de sinal de trânsito valendo 1 e qualquer outro objeto valendo 0. O tamanho da camada oculta da RNA foi definido empiricamente, possibilitando garantir o número mínimo de unidades necessárias para uma boa convergência e generalização da aprendizagem.

O método de aprendizagem supervisionado da RNA, requer um conjunto de dados de treinamento que devem ser gerados *a priori*. É necessário montar um conjunto representativo de exemplos, compostos de um grande conjunto de objetos de diferentes cenas, aplicando o descritor 3*D-CSD* e rotulando cada exemplo como um objeto sinal/não-sinal de trânsito. Para poder reconhecer os objetos de diferentes pontos de observação (orientações diferentes), exemplos adicionais de objetos em diferentes rotações são incluídos no conjunto de dados de treinamento.

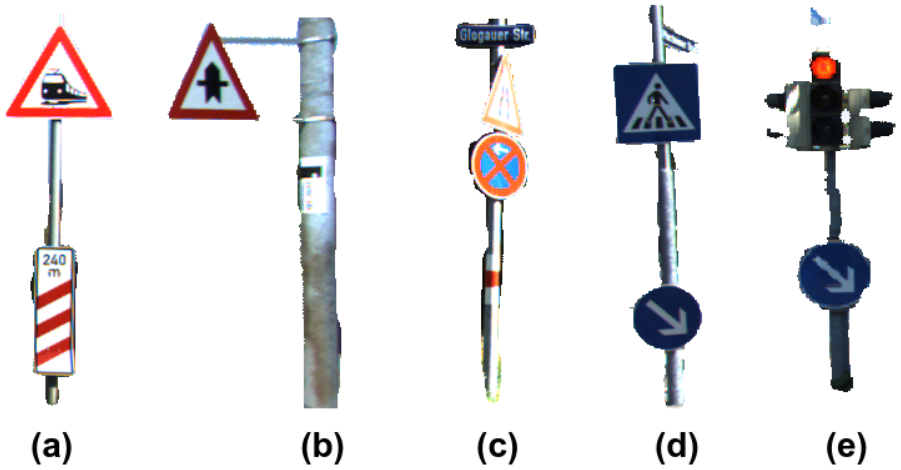
4.3.4 Filtro 3D para eliminação de falsos sinais de trânsito

O algoritmo para filtragem de falsos sinais de trânsito, utiliza em sua base de conhecimento para treinamento modelos de objetos 3*D* que representam os sinais de trânsito verticais e, que geralmente, podem ser encontradas no ambiente de navegação. Essa modelagem define um objeto único a ser detectado pelo sistema de Visão Computacional 3*D* (Figura 51), envolvendo suporte e placas. Sendo que no ambiente de trânsito urbano, nem sempre um sinal de trânsito é

colocado em um poste individual, em algumas situações, ele pode ser encontrado em um poste compartilhado com outros tipos de informações, por exemplo: placas de nome de rua, semáforos, ou até mesmo, em um poste com outros sinais de trânsito.

Uma Rede Neural Artificial (RNA) com saída binária foi treinada com estes vários casos de objetos 3*D* (Figura 51), onde sinais de trânsito informados por meio de placas, cones e semáforos podem ser encontrados. A RNA foi aplicada para resolver o problema de filtragem de sinalizações falsas utilizando imagens 3*D*. Para que isso seja possível, cada tipo de caso foi modelado (Figura 51) e baseado nos dados *Velodyne Light Detection and Ranging* (LIDAR) e, considerando também, um par de câmeras estéreo. Os dados provenientes destes sensores permitem que a RNA responda se o sinal de trânsito detectado em imagens 2*D* é real ou falso.

Figura 51 – Sinais de trânsito detectados em diferentes contextos (a) detectados em um fino poste metálico, (b) detectado em um poste de concreto, (c) e (d) duas placas de sinalização detectadas no mesmo poste e (e) sinal de trânsito detectado com um semáforo.

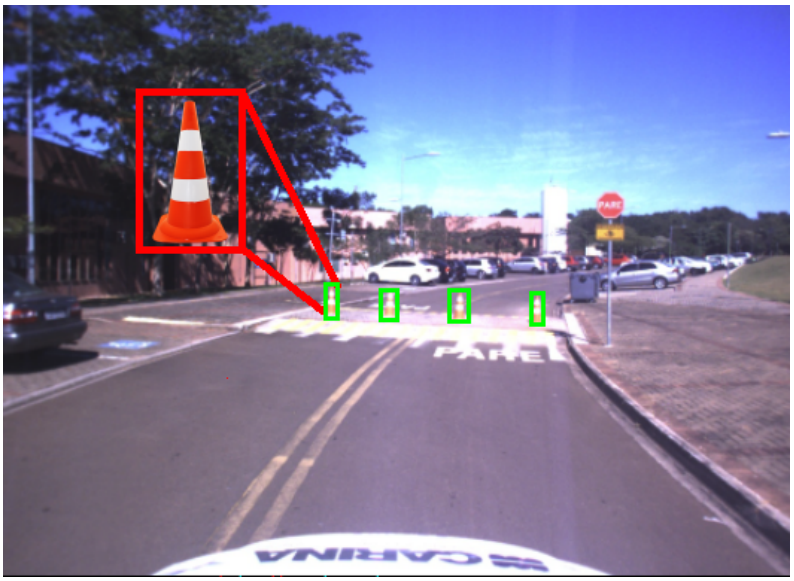


Fonte: Elaborada pelo autor.

Para o caso do algoritmo neural informar ao sistema de percepção que um sinal de trânsito foi detectado realmente na cena em imagens 3*D*, um segundo classificador baseado na rede de *Deep Learning Inception-V3* é ativado para classificar o tipo de sinal de trânsito que foi detectado na cena com fusão de dados 2*D* e 3*D* (*Red, Green e Blue* + *Depth* (RGB-D)): velocidade máxima, cones para desvio de rota, parada obrigatória, via preferencial, pedestres, ou também, outros tipos de informações.

Neste contexto, o sistema de percepção obtém a localização da detecção do sinal de trânsito em questão. Isso é possível graças as coordenadas (x, y e z) da imagem 3*D* (Figura 52). Sendo que estas informações são muito importantes para a classificação, mapeamento da informação e, também, para o sistema de Atenção Visual *Fuzzy* realizar a análise semântica da cena.

Figura 52 – Detecção de cones em uma cena usando dados 3*D* - cena real com dados fornecidos pelo sistema de câmera de visão estéreo.



Fonte: Elaborada pelo autor.

4.3.5 Metodologia de treinamento e teste para o filtro 3D

As etapas de treinamento e testes de filtragem de sinais de trânsito verticais foram aplicados no *Dataset* do KITTI (GEIGER; LENZ; URTASUN, 2012) com imagens 2*D* e 3*D*. Por meio dos resultados da filtragem de objetos 3*D*, foi possível eliminar os falsos objetos e declarar se cada detecção feita em 2*D* é um sinal de trânsito real ou não. Foi alcançada uma taxa de acerto de grande potencial para esta tarefa de detecção e filtragem 3*D*, possibilitando auxiliar na tarefa de percepção com um nível menor de falsas informações, sendo um diferencial desta pesquisa se comparado com outros trabalhos que usam apenas imagens 2*D* ou 3*D*, detectando vários casos de falsos positivos e falsos negativos.

Por meio da Figura 53, é possível observar alguns sinais de trânsito que foram detectados, filtrados e classificados com o sistema de percepção usando o conjunto de dados do KITTI (GEIGER; LENZ; URTASUN, 2012). Para que isso fosse possível, foram utilizadas as imagens estéreo disponíveis em conjunto com as imagens *RGB* equivalentes.

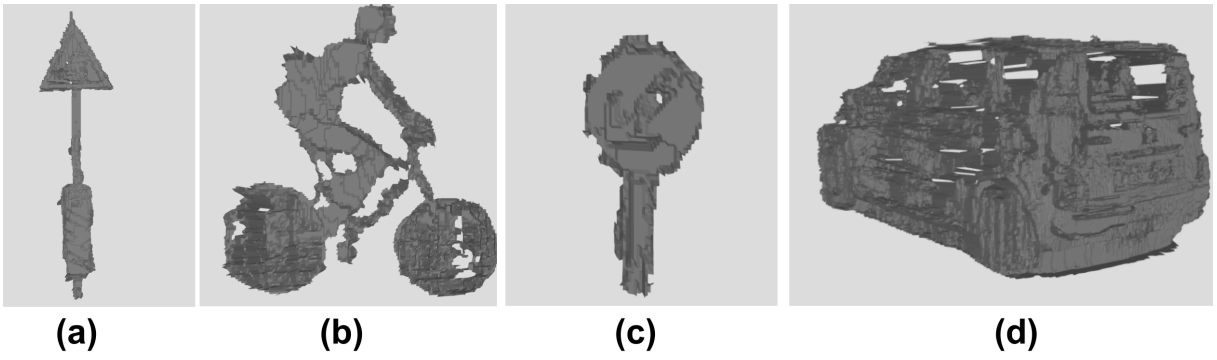
É importante destacar que o filtro para imagens 3*D* pode apresentar resultados de maneira incorreta para os variados formatos de sinais de trânsito, devido a imprecisão dos sensores 3*D*. Estes erros podem ser corrigidos pelo sistema de classificação de imagens 2*D*, uma vez que as formas (objetos em profundidade 3*D*) ou as imagens (cores, formas e texturas 2*D*) podem gerar erros no reconhecimento, mas geralmente não ocorrem ao mesmo tempo. Sendo então a fusão de imagens uma forma de unir caracterísiticas positivas nestes dois tipos de imagens.

Figura 53 – Exemplos de sinais de trânsito detectados no conjunto de dados KITTI.

Fonte: Elaborada pelo autor.

Para o treinamento da RNA foram utilizados outros tipos de objetos encontrados em ambientes de trânsito, como árvores, pessoas e carros (Figura 54), possibilitando então que o sistema de detecção e filtragem fosse capaz de diferenciar a classe sinas de trânsito da classe outros objetos.

Figura 54 – Objetos $3D$ equivalentes da detecção em $2D$: (a) e (c) sinais de trânsito, (b) ciclista e (d) carro.



Fonte: Elaborada pelo autor.

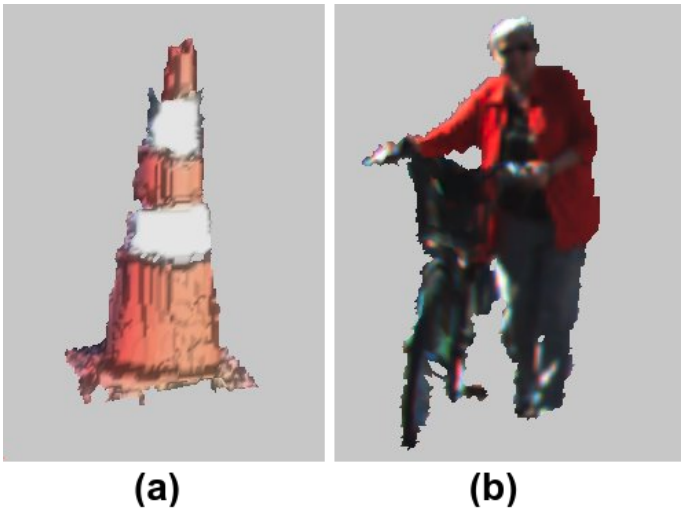
4.4 Classificação de Sinais de Trânsito Verticais

A classificação do tipo do sinal de trânsito que foi detectado pela etapa anterior, é feita por meio de uma rede *Deep Learning* que é aplicada em dados $2D$ (Imagem-*RGB*), aproveitando as informações de formas, texturas e cores disponíveis. Qualificando então que este tipo de rede é especialista para reconhecer os diferentes sinais de trânsito em suas diferentes classes.

Neste trabalho, foi aplicada a técnica de *Transfer Learning* em uma arquitetura de rede de *Deep Learning* baseada no modelo *DCNN inception-V3*, aproveitando a capacidade dos filtros e convoluções pré-treinados, gerando a capacidade deste modelo reconhecer e classificar imagens com grande acurácia. Incluindo imagens com oclusão, danificadas pelo tempo e, também, com precárias condições de iluminação. No entanto, sendo que esta tarefa pode ser muito difícil se considerados apenas os dados $3D$, já que este tipo de imagem apresenta uma nuvem de pontos muito esparsa e com precárias condições de superfície (Figura 54).

Na Figura 55, objetos detectados e filtrados pela visão estéro podem ser observados gerando a captura da nuvem de pontos $3D$ em conjunto com as cores, formas e texturas $2D$ que representam os elementos de uma cena de trânsito. Então, pode-se observar, que o sinal de trânsito para atenção na pista em formato de cone e também a ciclista, foram detectados respectivamente com noções de cores, texturas e profundidade ($RGB + Depth$) na fusão de imagens. No entanto, para a classificação da classe do sinal de trânsito, foi utilizada apenas a imagem $2D$ que representa a superfície da imagem (x, y) (Figura 53 e 55).

Figura 55 – Objeto $3D$ segmentado e sua imagem $2D$ - *RGB* equivalente: (a) Sinalização de trânsito em cone e (b) Ciclista



Fonte: Elaborada pelo autor.

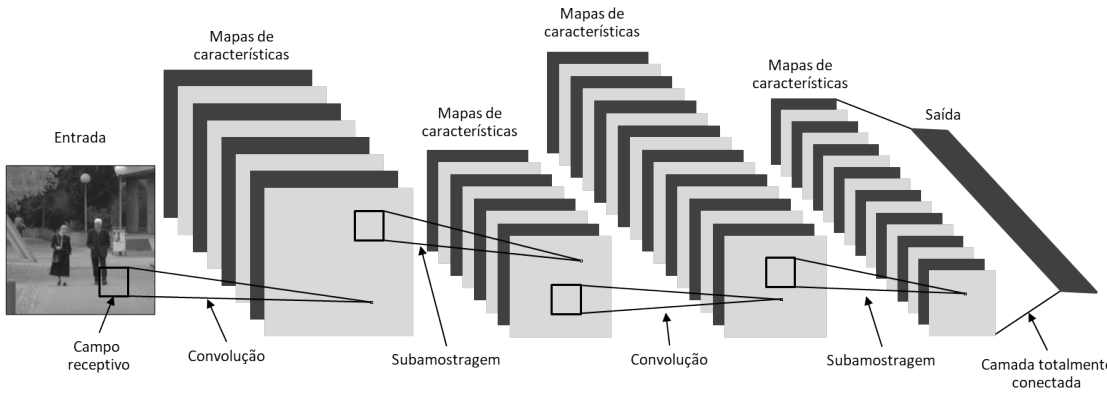
4.4.1 Transfer Learning

Transfer Learning (NILSBACK; ZISSERMAN, 2008) ; (Szegedy *et al.*, 2016) foi aplicado para facilitar a convergência do algoritmo de aprendizado de máquina baseado na rede de *Deep Learning Inception-V3*. Essa técnica gerou um aproveitamento dos pesos do modelo disponibilizado pelo *TensorFlow* e já treinado com o *Dataset* do *ImageNet* para inicializar o aprendizado de uma nova tarefa relacionada. Dada a transferência dos pesos para um novo terinamento da rede *Inception-V3*, foi possível obter a transferência de um conhecimento previamente

adquirido pelo treinamento com os dados do *ImageNet*, o que permite melhorar e agilizar o processo de aprendizado (GOODFELLOW; BENGIO; COURVILLE, 2017) e, também, aproveitar os filtros e concoluções do modelo adaptado.

Aplicando-se a técnica de *Transfer Learning* para classificação de objetos $2D$, pode-se estender a aplicação feita por meio da consideração de um domínio qualquer, $D = \{\mathscr{X}, P(X)\}$, construído por um conjuto geral de características $\mathscr{X}$ e uma distribuição de probabilidade uniforme $P(X)$, em que $X = \{x_1, ..., x_n\} \in \mathscr{X}$, gerando uma tarefa $\mathscr{T} = \{\mathscr{Y}, f(\cdot)\}$ (PAN; YANG, 2010), proveniente um espaço de classes $\mathscr{Y}$, juntamente a uma função de predição $f(\cdot)$. Por meio de um ponto de vista de probabilidades, a função $f(x)$ utilizada para predizer a classe correspondente de uma instância x deve ser formalmente descrita como $P(y|x)$. Visando estes aspectos fundamentais, dado um domínio de origem D_S correspondente a uma tarefa $\mathscr{T}_S$, um domínio final D_T e uma tarefa de classificação $\mathscr{T}_T$, o *Transfer Learning* gera suporte para o novo aprendizado com a função de predição $f_T(\cdot)$ utilizando o conhecimento de $\mathscr{T}_S$ e D_S, onde $D_S \neq D_T$, ou $\mathscr{T}_S \neq \mathscr{T}_T$ (PAN; YANG, 2010).

Figura 56 – Extração de características em dois estágios feita por uma rede de *Deep Learning*.



Fonte: Editada pelo autor de (LECUN; KAVUKCUOGLU; FARABET, 2010).

Em resumo, a técnica de *Transfer Learning* permitiu transferir as primeiras n camadas com seus filtros e convoluções e, que contém características genéricas para a classificação de objetos baseadas no treinamento proveniente dos dados do *ImageNet*, para as referentes n primeiras camadas da rede *Inception-V3*. Sendo que para esta aplicação de classificação, as camadas posteriores foram inicializadas aleatoriamente e treinadas do estado inicial para o novo modelo. Sendo que as camadas copiadas realizaram a retropropagação dos erros para ajustá-las ao problema de classifcação de sinais de trânsito verticais. A escolha desse procedimento de treinamento depende do tamanho do conjunto de dados da nova tarefa de classificação e do número de parâmetros nas primeiras n camadas (YOSINSKI *et al.*, 2014).

Destacando então que a principal abordagem do processo de *Transfer Learning* é aproveitar os filtros e convoluções de uma rede de *Deep Learning* pré-treinada, possibilitando uma extração de características mais robusta sobre objetos de novas classes de imagens 2*D* (Figura 56).

4.4.2 Metodologia de treinamento e teste para o classificador

Para o treinamento do classificador baseado na rede de *Deep Learning Inception-V3*, foi utilizado o conjunto de dados do *German Traffic Sign Benchmarks* (INI) (STALLKAMP *et al.*, 2011). No entanto, para os testes do sistema de percepção, foi aplicado um segundo banco de dados. Foram aplicadas as cenas do *Vision Benchmark Suite* (KITTI) (GEIGER *et al.*, 2013). Nestas cenas (GEIGER *et al.*, 2013) o sistema de detecção de sinais de trânsito com fusão de dados 2*D* e 3*D* foi aplicado e, em seguida, as imagens candidatas 2*D* provenientes da representação dos sinais de trânsito são enviadas para o classificador baseado na rede de *Deep Learning Inception-V3*. O uso de duas bases de dados possibilitou a validação do sistema, uma vez que foi testado com um conjunto de dados completamente diferente do usado no treinamento.

O treinamento baseado em *Trasnfer Learning*, e que foi aplicado junto ao modelo de rede *Inception-V3*, modificou apenas a camada final da rede (SZEGEDY *et al.*, 2016), garantindo a preservação de todos os filtros e convoluções que são genéricos para o reconhecimento de diferentes tipos de imagens baseadas em texturas, cores, formas e outros tipos de padrões (BRUNO; OSORIO, 2017). Para poder adaptar a rede ao novo modelo, 70% do conjunto de dados foi aplicado para realizar o treinamento (adaptando os parâmetros da rede - ajustando os pesos). Os 30% restantes foram aplicados apenas para avaliar o desempenho da tarefa de classificação nas etapas de teste e validação (DONAHUE *et al.*, 2014). Além disso, o uso de duas bases de dados para treino e testes possibilitou evitar *overtraining/overfit*. Garantindo um melhor aprendizado.

4.5 *Dataset*

4.5.1 Dataset para classificação

Por meio da Figura 57, são apresentadas as classes dos sinais de trânsito alemães e que fazem parte do *Dataset* (INI) (STALLKAMP *et al.*, 2011), sendo utilizados para o treinamento do classificador neural. No total, 21 classes contendo 21.000 imagens foram utilizadas para essa tarefa. O principal objetivo do uso deste conjunto de imagens está relacionado ao reconhecimento dos sinais de trânsito em seus diferentes tipos.

O conjunto de dados disponível contém imagens entre 15x15 e resolução de 250x250 *pixels* e estão no formato *Portable PixMap* (PPM). As imagens foram convertidas para o formato JPG (*Joint Photographic Experts Group*), configurado sem perda em dados de imagem e qualidade.

Figura 57 – Classes de sinais de trânsito utilizadas para *Transfer Learning*.

Fonte: (STALLKAMP *et al.*, 2011).

4.5.2 Dataset para detecção e filtragem 3D

O *Dataset* do KITTI (GEIGER; LENZ; URTASUN, 2012) disponibiliza um importante conjunto de dados para teste e validação de algoritmos para percepção, sendo bastante conhecido por pesquisadores da área de veículos inteligentes. Neste trabalho, foi utilizado o *Dataset* do KITTI por sua disponibilidade de cenas de trânsito com fusão de dados 2*D* e 3*D* (Figura 58).

Figura 58 – *Dataset* do KITTI: Cenas com sinais de trânsito com fusão de dados 2*D* e 3*D*.

Fonte: (GEIGER; LENZ; URTASUN, 2012).

O *Dataset* também tem suas imagens na forma de *tracking*, o que facilita o teste e validação de algoritmos de percepção para detecção de sinais de trânsito verticais em diferentes distâncias. Neste trabalho, foram utilizados os dados de 19.000 imagens: *RGB* e *RGB+D* em dados de *tracking*, sendo que foram rotulados todos os sinais de trânsito para os processos de treinamento, teste e validação para a tarefa de detecção e filtragem 3*D*.

4.5.3 Dataset modificado para a filtragem de falsos sinais

Para os testes do filtro 3*D* em sinais de trânsito falsos, foi necessário adaptar o *Dataset* do KITTI para esta tarefa, visto que não existe nenhum conjunto de imagens disponível atualmente que atenda essa necessidade. Para as adaptações necessárias, sinais de trânsito falsos foram colocados na traseira de veículos em imagens 2*D*. Na Figura 59 pode ser observada uma situação onde existem três sinais de trânsito, sendo dois reais (placas de "PARE") e um sinal de trânsito falso (velocidade máxima) que foi adaptado na traseira do veículo para os testes.

Figura 59 – Imagens modificadas do *Dataset* do KITTI para testes da detecção e filtragem 3*D*.

Fonte: Adaptada de: (GEIGER; LENZ; URTASUN, 2012).

4.6 Considerações

Neste capítulo, foi apresentado um novo sistema de percepção que é capaz de detectar e classificar sinais de trânsito por meio da fusão de imagens 2*D* e 3*D*, sendo mais robusto para falhas de detecção geradas por falsos positivos e falsos negativos. Isso foi possível graças a aplicação de um filtro que é capaz de eliminar falsos sinais de trânsito utilizando dados de profundidade.

Analisando os dados 2*D* e 3*D* gerados pela fusão de sensores do estado-da-arte de Visão Computacional, foi possível detectar os sinais de trânsito verticais: cones, sinais de desvio e placas convencionais, em duas divisões de trabalho:

- (a) **- Dados** 2*D***:** Permitiram uma detecção e classificação eficiente dos tipos de sinais de trânsito verticais, graças à capacidade das redes de *Deep Learning* sobre as imagens com cores, texturas e formas. Sendo que este tipo de informação não é encontrado atualmente com qualidade em imagens 3*D*.

- (b) **- Dados** 3*D***:** Possibilitaram identificar falsos positivos e falsos negativos: sinais de trânsito colados na traseira de veículos, estampados em muros e objetos do mesmo formato geométrico das placas, fossem analisados e, também filtrados. Para essa tarefa fosse possível, foi utilizado um bom descritor 3*D* aplicado na imagem de profundidade, permitindo extrair características de formato. Dando suporte para que uma RNA-MLP *(Multilayer Perceptron)* classifique em duas classes o tipo de objeto detectado na cena (1) sinais de trânsito e (2) não sinais de trânsito. Obtendo então uma maior robustez para o sistema de detecção que ficou mais preciso para eliminar informações falsas e que geram problemas para a navegação do veículo.

Destacando também que o sistema de detecção e classificação de sinais de trânsito apresentado neste capítulo, gera suporte para o modelo de Atenção Visual apresentado no Capítulo 5. O modelo de Atenção Visual é capaz de analisar a cena com múltiplos sinais de trânsito detectados em conflito. Possibilitando então declarar a prioridade de cada um destes sinais e suas respectivas informações para um dado instante da navegação do veículo. Sendo que para a Atenção Visual, é indispensável os dados de posicionamento (x, y e z) e profundidade 3*D* dos elementos da cena, já que necessita das informações de distâncias dos sinais de trânsito em relação ao veículo e, também, em relação a via e outros elementos da cena para a análise semântica desenvolvida.

CAPÍTULO

5

ATENÇÃO VISUAL FUZZY

Neste capítulo, é apresentado um modelo de Atenção Visual *Fuzzy* para trabalhar em conjunto com o sistema de percepção de sinais de trânsito verticais. O modelo de Atenção Visual é formado por uma base de regras bem definida em conjunto com regiões de interesse *Fuzzy*, sendo então capaz de classificar a prioridade de sinais de trânsito detectados em conflito, criando-se então uma camada entre o sistema de percepção (detecção + classificação) e o sistema de tomada de decisão do veículo (Bruno; Osório, 2019).

Esta tarefa de análise de sinais de trânsito é possível graças ao uso de Múltiplos Atributos de Decisão - *Multiple Attribute Decision-Making* (MADM) (HWANG C; YOON, 2011) e, também, aplicando conjuntos *Fuzzy* para que seja possível classificar o nível de prioridade dos sinais de trânsito detectados. O Processo de Hierarquia Analítica - *Analytic Hierarchy Process* (AHP) (SAATY, 1990), foi aplicado para calcular os pesos dos atributos e, a técnica para a Ordem de Preferência por Similaridade com a Solução Ideal - *Technique for Order of Preference by Similarity to Ideal Solution* (TOPSIS) (HSU-SHIHSHIH; SHYUR H; ZAVADSKAS E, 2007), aplicada para classificar os sinais de trânsito em seus níveis de importância.

O principal objetivo é contribuir com um novo sistema de percepção inteligente e, por meio de um conjunto de base de regras *Fuzzy* proveniente de regiões bem definidas, poder relacionar semanticamente a cena e definir qual sinal de trânsito é mais importante em um determinado momento da navegação. Gerando então o suporte para o sistema de tomada de decisão do veículo encontrar uma solução com base nas regras de trânsito locais.

5.1 Dados Gerados para o Sistema de Atenção Visual

O sistema de percepção $2D$ e $3D$ deve fornecer os dados *a priori* de detecção e classificação de sinais de trânsito para o modelo de Atenção Visual *Fuzzy*. O modelo de Atenção Visual é capaz de analisar o uso de sinais auxiliares (cones e sinais de emergência) e relacioná-los com os

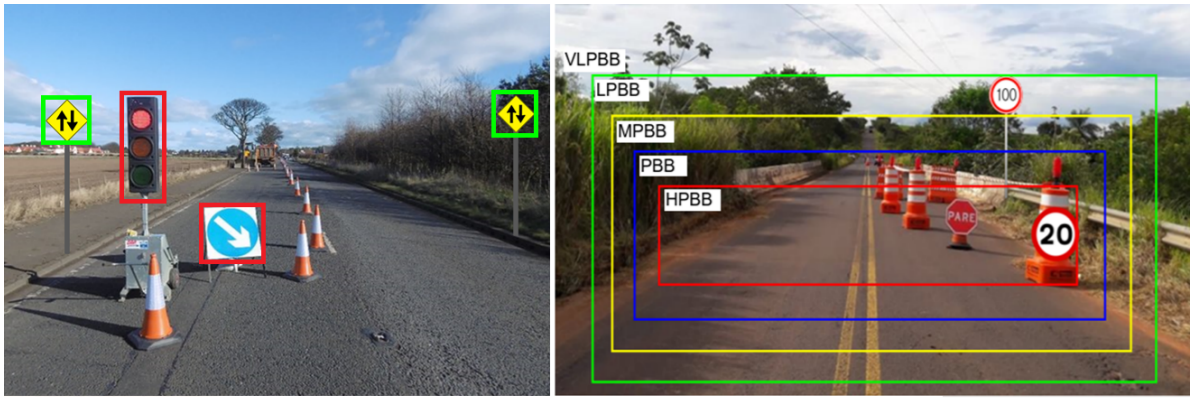
sinais de trânsito habituais.

Também foi possível relacionar os dados de detecção e classificação de sinais de trânsito com a detecção da área navegável, dando suporte para os casos de bloqueios de vias e bifurcações (obras na rodovia, com acidente de trânsito, etc) e, que geram novas situações com alterações na sinalização de trânsito habitual.

Destacando então que este modelo de Atenção Visual é altamente dependente dos resultados do sistema de percepção com fusão de dados $2D$ e $3D$, pois os resultados de detecção e classificação geram as entradas provenientes para a análise de prioridade dos sinais de trânsito. As informações que são mais importantes para o sistema de Atenção Visual, são: (a) detecção e classificação do sinal de trânsito e a sua (b) posição espacial.

A análise feita pela Atenção Visual deve então priorizar os sinais de emergência não habituais e que são mais importantes para a tomada de decisão do veículo em um trecho não mapeado (Figura 60), já que as regras de trânsito nestes tipos de situações são alteradas.

Figura 60 – Problema de detecção com múltiplos sinais (a) Exemplo 1: Conflitos de sinais de trânsito de direção e (b) Exemplo 2: regiões *Fuzzy* de interesse. Legenda: Caixa Delimitadora de Prioridade Máxima (*CDPM - HPBB*), Caixa Delimitadora de Prioridade (*CDP - PBB*), Caixa Delimitadora de Prioridade Média (*CDPM - MPBB*), Caixa Delimitadora de Prioridade Baixa (*CDPB - LPBB*) e Caixa Delimitadora de Prioridade Muito Baixa (*CDPMB - VLPBB*).



Fonte: Elaborada pelo autor.

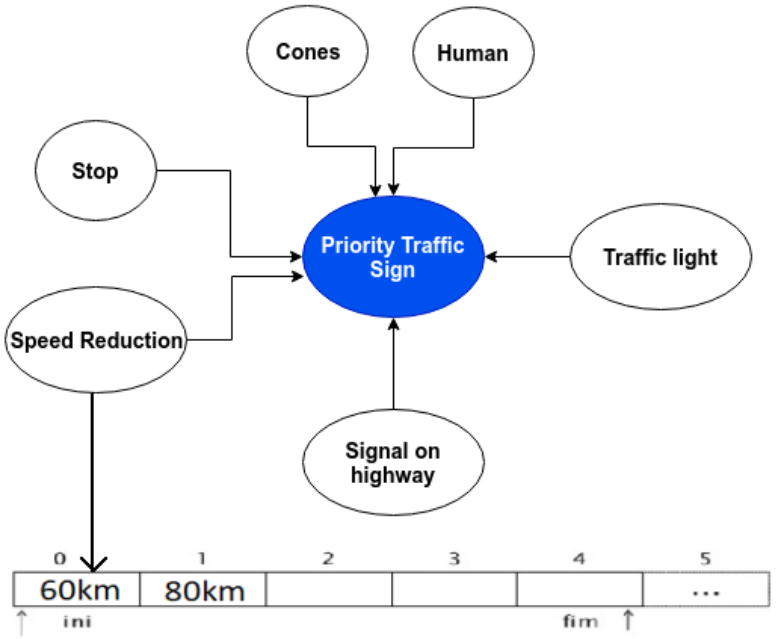
5.2 Análise Semântica

5.2.1 Influências: Rede semântica

Por meio da Figura 61, é possível observar a modelagem da rede semântica para as relações de regras de trânsito com outros elementos da cena. A construção das relações, modela todas as influências que um sinal de trânsito pode ter de acordo com a modelagem do problema.

Para que isso seja possível, foram aplicados alguns elementos que são de grande importância para uma análise semântica: (a) cones, (b) pessoas, (c) parada obrigatória, (d) redução de velocidade e (e) a relação do sinal de trânsito com a via (Figura 61). Na próxima seção (Seção 5.2.2), é apresentada uma descrição de cada relação feita entre os sinais de trânsito e que foram modeladas com auxílio da rede semântica.

Figura 61 – Rede Semântica: Influências da base de regras *Fuzzy* gerada para cada sinal de trânsito na cena.



Fonte: Elaborada pelo autor.

5.2.2 Conflitos de informações: Problemas de detecção com múltiplos sinais de trânsito

O sistema de percepção desenvolvido nesta tese, possui uma camada de processamento para análise semântica, permitindo então definir quais sinais de trânsito são mais importantes em um determinado momento (geralmente quando um conjunto de sinais de trânsito é detectado ao mesmo tempo) para a navegação do veículo (Figura 60 - (a e b). Essa camada de análise de sinais é capaz de trabalhar com uma característica muito importante da visão humana, sendo esta a Atenção Visual. A Atenção Visual deve ser capaz de definir quais sinais de trânsito são mais importantes por meio do relacionamento entre os elementos da cena, enviando uma lista de prioridades com estes sinais ao sistema de tomada de decisão do veículo.

Para uma melhor compreensão do problema de conflitos de sinais de trânsito na mesma cena, é possível observar na Figura 60-(a e b), onde duas informações de direção (Figura 60-(a) e dois sinais com limite de velocidade são detectados em conflito (Figura 60-(b)).

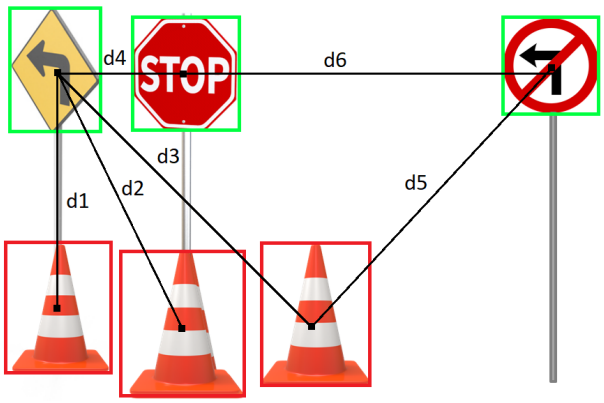
Nestas situações, há grandes problemas, porque apenas o sistema de percepção não deve ser capaz de interpretar a cena em alto nível e definir a direção real (Figura 60-(a)) ou o limite de velocidade desta rota (sinal de $20km/h$ ou $100km/h$ (Figura 60-(b)). No entanto, aplicando a camada de Atenção Visual e fazendo a análise semântica em cada elemento de trânsito detectado, é possível interpretar o contexto da cena.

Também é possível observar na Figura 60-(b) as regiões de interesse *Fuzzy* que são aplicadas para definir quais sinais de trânsito têm maior prioridade. Um sinal de trânsito no centro da via (Caixa Delimitadora de Prioridade Máxima - *CDPM*) tem uma prioridade mais alta do que um sinal de trânsito na periferia (Caixa Delimitadora de Prioridade Baixa - *CDPB*). Essas regiões de interesse auxiliam no processo de detecção do sinal de maior prioridade.

As áreas dos retângulos (Figura 60-(a)) representam não apenas as conexões entre os objetos, mas também, as distâncias dos sinais de trânsito em relação ao veículo. Um sinal de trânsito central (*CDPM*) está mais próximo do veículo e deve ser detectado e reconhecido antes de um sinal periférico mais distante (*CDPMB*). Pois, um sinal de trânsito mais distante, oferece um maior tempo para a tomada de decisão do veículo.

- *Problema 1*: Novamente usando os exemplos da Figura 60, em casos que dois sinais de velocidade são detectados ao mesmo tempo. Para resolver esse problema, foi aplicada uma análise das relações entre os elementos da cena com base nas variáveis de (a) distâncias euclidianas ($d1$, $d2...dn$) entre os elementos (Figura 62), (b) a relação com a área navegável e o (c) peso das conexões ($p1$, $p2...pn$) dadas pela base de regras em função das regiões de interesse.

Figura 62 – Cálculo das distâncias euclidianas entre os sinais de trânsito.



Fonte: Elaborada pelo autor.

- *Problema 2*: Também existe um segundo problema em que um semáforo é detectado junto com uma parada obrigatória. Se o semáforo estiver verde e um sinal de parada obrigatória for detectado, é gerado um grande problema para a tomada de decisão do veículo;
- *Problema 3*: Um fator importante que deve ser central na análise dos sinais de trânsito, é verificar se o sinal está dentro ou fora da via. Se estiver dentro, certamente deve ter uma prioridade mais alta. Possivelmente sendo um sinal móvel que foi colocado temporariamente para sinalizar uma situação de emergência. Para que esta última análise seja possível, é necessária a detecção da área navegável;
- *Problema 4*: Detecção de sinais de trânsito de sentido da via com direções em conflito;
- *Problema 5*: Por último, é possível observar na Figura 63, um sinal de trânsito móvel e que define uma nova regra para a via. Sendo esta informação de alta prioridade, pois é um sinal de emergência temporário (trechos de obras ou área escolar). Para este tipo de situação, a análise semântica deve dar um grande peso para um sinal de trânsito relacionado a uma pessoa.

Figura 63 – Detecção de cones, pessoa e sinal de emergência em uma cena usando dados 2*D*.

Fonte: Elaborada pelo autor.

5.3 Detecção de Área Navegável

Para poder relacionar os sinais de trânsito que estão dentro e/ou fora da via, e também, estimar para qual situação uma informação pertence, primeiro é necessária uma segmentação da área navegável. Para que isso fosse possível, foi utilizada uma rede de *Deep Learning* intitulada como *KittiSeg* para esta tarefa (FRITSCH; KUEHNL; GEIGER, 2013). Essa rede é capaz de segmentar a área navegável de *pixel* a *pixel* em imagens 2*D*, permitindo que o sistema de Atenção

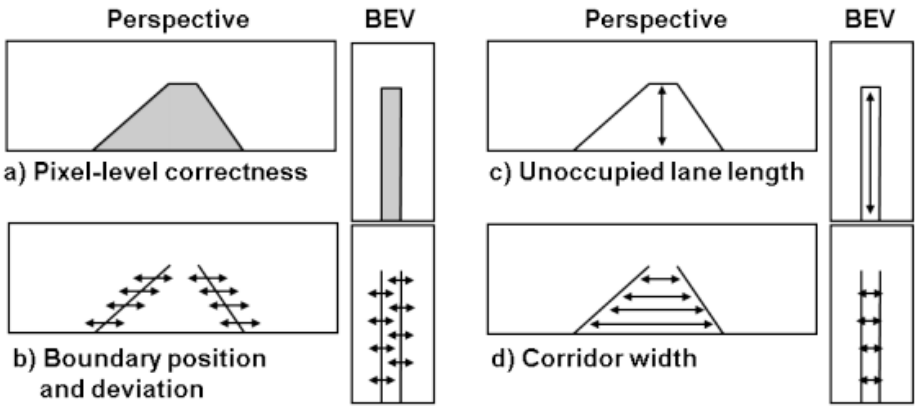
Visual interprete melhor a cena. Especialmente em situações em que uma rota auxiliar é criada (obras, acidentes, desvios) e existem sinais de trânsito fora da situação habitual e, também, sinais dentro da via. Sendo estes casos de maior prioridade, por geralmente informarem uma situação de emergência no trânsito (Figura 60).

5.3.1 Detecção de área navegável 2D

Os métodos utilizados pela rede de *Deep Learning KittiSeg*, aplicada para segmentação de área navegável, fornecem resultados de limites da via (Figura 64-(a)), sendo avaliados dadas as distâncias da faixa estimada de marcação central (Figura 64-(c)) até as bordas da área navegável (Figura 64-(b)). Para definir uma margem flexível para a contagem ideal de *pixels* candidatos, as taxas de CVP (Clássico Verdadeiro Positivo) e FP (Falso Positivo) devem ser obtidas (FRITSCH; KUEHNL; GEIGER, 2013). Sendo estes os *pixels* que representam respectivamente a área navegável e a área não navegável.

Além dos limites da área navegável em largura (Figura 64-(d)), o comprimento da via desocupada (Figura 64-(c)) é importante para o sistema de Atenção Visual, devendo ser combinada com as informações da área navegável lateral (FRITSCH; KUEHNL; GEIGER, 2013).

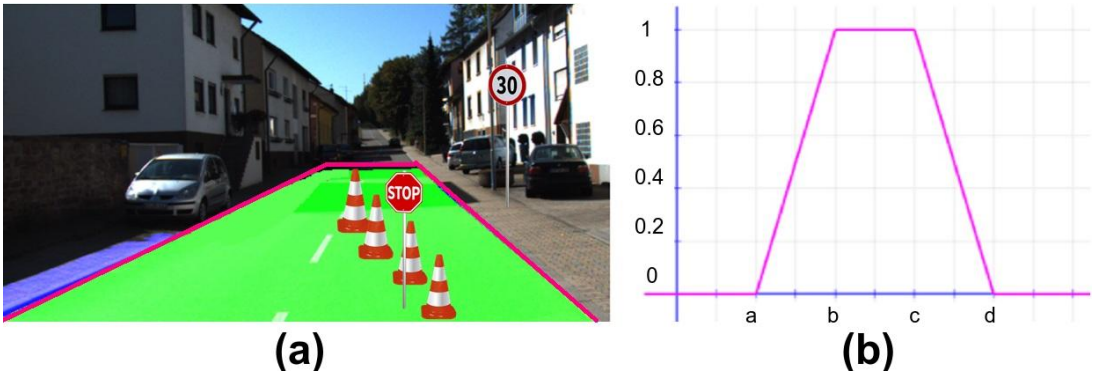
Figura 64 – Visualização de métricas de avaliação.



Fonte: (FRITSCH; KUEHNL; GEIGER, 2013).

Por meio da Figura 65, é possível observar uma via detectada. Nesta imagem, é possível avaliar se o sinal de trânsito é convencional (fora da via) ou emergencial (dentro da via). Lembrando que essa análise é apenas uma probabilidade condicional, que por meio da função trapezoidal, possibilita dar um valor de prioridade (0 a 1) a um determinado sinal de trânsito. Também existem outras variáveis que devem contribuir para essa análise. Outras condicionais serão apresentados nas seções posteriores.

Figura 65 – Região trapezoidal de interesse difuso (a) Detecção de vias e (b) Função trapezoidal correspondente.



Fonte: Elaborada pelo autor.

5.3.2 Detecção de área navegável 3D

Por meio das imagens $2D$ é possível detectar a área navegável, no entanto, não é possível obter as distâncias dos sinais de trânsito em relação a via e aos outros elementos da cena. Por este motivo, também foram utilizados os dados $3D$ para relacionar os sinais detectados com suas respectivas posições (x, y, e z). Essa análise possibilita informar ao modelo de Atenção Visual *Fuzzy* as distâncias dos sinais de trânsito e que também são entradas de dados para a análise de prioridade de cada informação.

Para esta análise, foi utilizado um mapa de grade baseado em dados estéreo, fundido com odometria visual. Usando dados estéreo e o método de localização ORB-SLAM (*Oriented FAST and Rotated BRIEF - Simultaneous Localization and Mapping*) (MUR-ARTAL; TARDóS, 2017), foi possível concatenar cada mapa de nuvem de pontos estéreo, agregando informações em cada nova região mapeada.

5.3.3 Localização visual

Como foram utilizadas imagens estéreo para gerar os pontos $3D$ do ambiente, possibilitando recriar a trajetória em uma grade de nuvem de pontos, é preciso uma localização confiável e fazer uma estimativa do veículo em cada quadro da câmera (*tracking*). Alguns sensores, como GPS e IMU, fornecem uma posição global precisa do veículo. O problema neste caso é a sincronização dos dados de vários sensores. Isso poderia agregar um ruído incremental que comprometeria a concatenação dos pontos e o mapa final.

A técnica de ORB-SLAM é baseada no extrator de recursos ORB (*Oriented FAST e Rotated BRIEF*). Esse detector de características é uma fusão do detector de ponto-chave FAST (ROSTEN; DRUMMOND, 2006) e do descritor BRIEF (CALONDER *et al.*, 2012). O sistema propõe um método de odometria visual baseado na extração de recursos de imagem

Figura 66 – (a) O sistema ORB-SLAM extraindo os recursos em verde, (b) Os recursos do ORB-SLAM mapeados em pontos vermelhos, em azul os quadros principais representativos e em verde a trajetória estimada.

Fonte: Elaborada pelo autor.

robusta às transformações, translação e rotação.

Durante a trajetória, os quadros-chave são selecionados para mapear o ambiente, os recursos são extraídos e mapeados nas imagens de par estéreo com sua posição $3D$ para que novas poses de quadros possam ser estimadas usando os quadros-chave mapeados (Figura 66). O primeiro quadro, é considerado a origem do sistema e, os novos quadros, têm a pose calculada como uma transformação da origem para sua posição estimada. Os recursos de correspondência nas imagens esquerda e direita são mapeados para apoiar a estimativa de pose sempre que esse item for encontrado. Com base nas informações detalhadas, os recursos são divididos em duas classes, *perto* e *longe*. Pontos mais próximos da câmera têm informações mais confiáveis sobre translação e rotações.

Esse conhecimento sobre a profundidade dos elementos da cena é importante, pois possibilita ter uma estimativa mais precisa da pose do veículo em relação aos sinais de trânsito e, também da área navegável. À medida que o veículo navega pelo ambiente, novas características são extraídas e salvas (Figura 66).

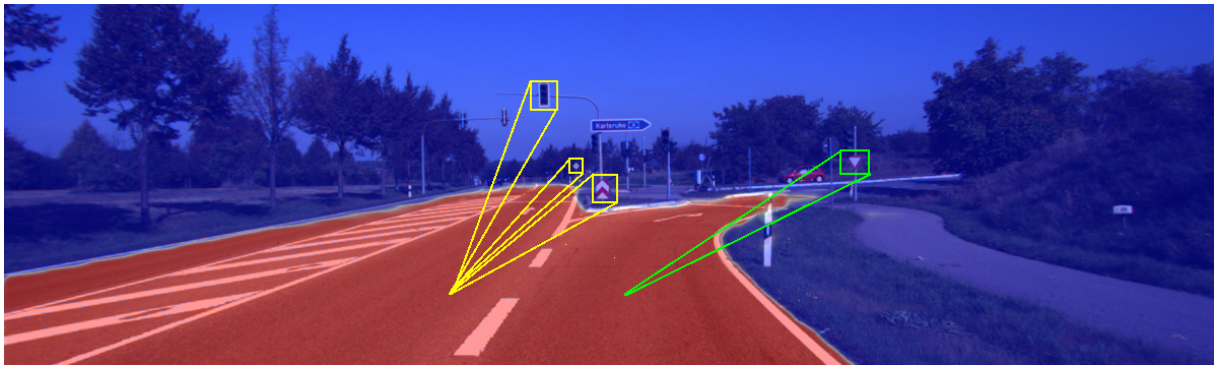
5.4 Análise de Bifurcações

Por meio dos dados provenientes da segmentação da área navegável, é possível detectar as bifurcações e relacionar os sinais de trânsito correspondentes. Essa análise é bastante importante para evitar sinais de trânsito não pertencentes a rota atual em situações não mapeadas *a priori*.

Como exemplo, pode ser observada a cena da Figura 67, onde na bifurcação existem sinais de trânsito pertencentes ao trecho do retorno a direita e seguindo em frente.

A análise dos sinais de trânsito em conjunto com a área navegável, permite avaliar qual sinal de trânsito deve ser obedecido dada a direção tomada pelo veículo na bifurcação. Sem a detecção da área navegável e suas bifurcações correspondentes, não seria possível que o sistema de Atenção Visual realizasse esta tarefa em situações não mapeadas *a priori*.

Figura 67 – Detecção de bifurcações e seus sinais de trânsito correspondentes.



Fonte: Elaborada pelo autor.

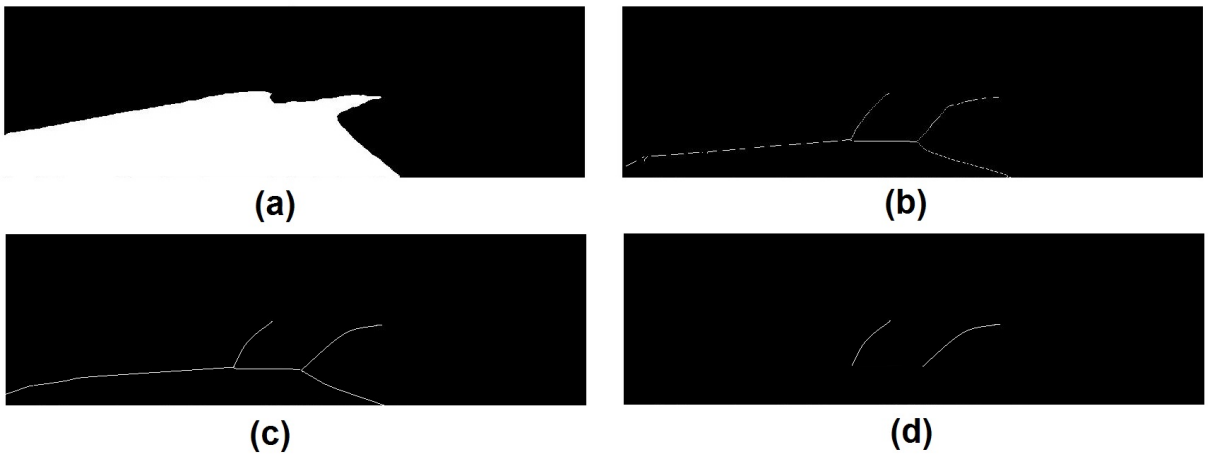
- **Segmentação:** Para que seja possível detectar as bifurcações, é necessário primeiramente ter a área navegável segmentada. A segmentação das regiões navegáveis foram feitas por meio da rede de *Deep Learning* KittiSeg em imagens $2D$ (Figura 68-(a)).

- **Esqueletização:** O segundo passo é detectar o centro de cada região navegável que foi segmentada *a priori*. Para que essa tarefa seja realizada, primeiramente é aplicada uma esqueletização na imagem, possibilitando definir uma linha de referência para cada via (Figura 68-(b)).

- **Morfologia Matemática:** No terceiro passo, são aplicados filtros morfológicos para melhor segmentar os esqueletos e suas linhas equivalentes. Primeiramente é aplicado o filtro de (a) dilatação e depois o filtro de (b) erosão (Figura 68-(c)). Depois de aplicados os filtros, é feito o preenchimento do trapezoide que corresponde a área navegável fechada e, logo seguinte, essa região é eliminada, assim, sobrando apenas as linhas que representam as bifurcações (Figura 68-(d)).

- **Contagem de elementos desconectados:** Por fim é aplicada uma contagem de elementos desconectados. Se dois elementos forem detectados de maneira desconectada (Figura 68-(d)), então uma bifurcação foi detectada. Para este caso, o modelo de Atenção Visual é capaz de relacionar os sinais detectados com as direções correspondentes e com base nas linhas desconectadas, possibilitando definir para qual sentido da via a regra de trânsito pertence.

Figura 68 – Análise de bifurcações: (a) Imagem da via segmentada, (b) Esqueletização, (c) Aplicação de de morfologia matemática e (d) Contagem de elementos desconectados.



Fonte: Elaborada pelo autor.

5.5 Atenção Visual Fuzzy: Classificação de Prioridade dos Sinais de Trânsito

O processo de tomada de decisão deve ser capaz de avaliar um conjunto de informações finitas, decidindo a ação mais apropriada de acordo com um conjunto de valores *Fuzzy*. Tornando possível classificar a pertinência do sinal de trânsito detectado.

Para que isso seja possível, é aplicada uma abordagem baseada na técnica MADM para tomada de decisão, e que é voltada para classes de acordos arbitrários e que possibilitam a decisão, garantindo uma avaliação de um conjunto de critérios diferentes. De acordo com Chen e Hwang (1992) (Chen *et al.*, 2014), o MADM se aplica muito bem ao problema de Atenção Visual para sinais de trânsito, porque é especializado em problemas com conjuntos finitos de alternativas e permite a avaliação nas etapas de decisão com base de regras *Fuzzy* (Pachêco Gomes *et al.*, 2018).

5.5.1 Análise hierárquica

Por meio da técnica *Analytic Hierarchical Process* (AHP), é criada uma estrutura hierárquica, possibilitando relacionar os componentes do problema de detecção de sinais de trânsito em conflito para uma análise de decisão. Com esse recurso de decomposição hierárquica, o elemento de decisão pode fazer uma comparação entre os elementos e classificá-los em seu nível de prioridade (Chen *et al.*, 2014) ; (Pachêco Gomes *et al.*, 2018).

5.5.2 Regiões Fuzzy de interesse: Múltiplos atributos

Um conjunto de valores *Fuzzy* é usado com os métodos *Multiple Attribute Decision Making* (MADM) para modelar incerteza e subjetividade na análise de decisão. Chen e Hwang (Chen *et al.*, 2014) descreveram algumas abordagens para as etapas do MADM. Neste trabalho, foi aplicado conjuntos *Fuzzy* para representar a importância de cada sinal de trânsito detectado em suas regiões de interesse, possibilitando então definir a relevância para cada sinal detectado (Pachêco Gomes *et al.*, 2018).

Para o sistema de Atenção Visual proposto, foi trabalhado com o modelo linguístico *Fuzzy*. O método foi aplicado pelas duas etapas descritas abaixo:

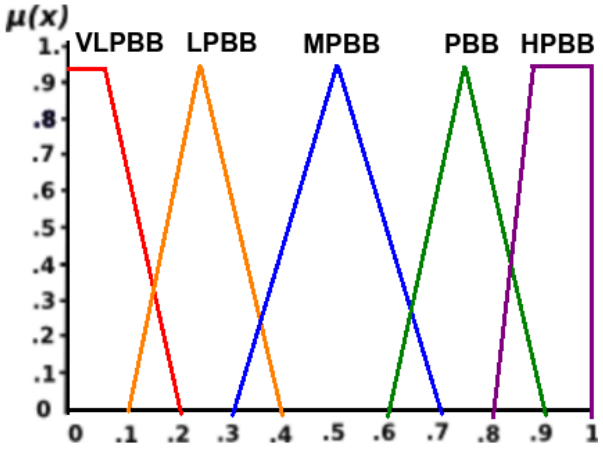
- *Etapa 1: Converta termos linguísticos em números difusos*

  A abordagem aplicada neste sistema utiliza um conjunto finito de termos linguísticos que podem ser ajustados para melhor descrever a natureza dos atributos, onde, por exemplo:

$$Fuzzy - U = \{muito\ alto,\ alto\ a\ muito\ alto,\ alto,\ razoavelmente\ alto,\ médio,\ razoavelmente\ baixo,\ baixo,\ baixo\ a\ muito\ baixo.\}.$$

  A primeira etapa da lógica *Fuzzy*, identifica uma escala de conversão (de 0 a 1) que representa o subconjunto usado para caracterizar o atributo. Por meio da Figura 69, é possível observar um exemplo de como os números difusos nessas escalas podem substituir os termos linguísticos (Chen *et al.*, 2014).

Figura 69 – Regiões *Fuzzy* de interesse: Caixa Delimitadora de Prioridade Maxíma (*CDPM - HPBB*), Caixa Delimitadora de Prioridade (*DCP - PBB*), Caixa Delimitadora de Prioridade Média (*CDPM - MPBB*), Caixa Delimitadora de Prioridade Baixa (*CDPB - LPBB*) e Caixa Delimitadora de Prioridade Muito Baixa (*CDPMB - VLPBB*).



Fonte: Elaborada pelo autor.

- *Etapa 2: Converta números Fuzzy em pontuação nítida*

  Depois de identificar a escala correspondente de cada valor e substituir os termos linguísticos por números *Fuzzy*, é preciso aplicar um método de pontuação para converter esses números em dados precisos. Para esta tarefa, foi proposto um método de pontuação *Fuzzy* para estimativas μ_T, chamado: Pontuação total do número *Fuzzy* M. Usando a esquerda (μ_E) e a direita (μ_D), dos conjuntos de maximização e minimização (μ_{max} e μ_{min}, respectivamente, e a função de associação de $M(\mu_M)$, onde:

$$\mu_E(M) = \sup_x[\mu_M(x) \wedge \mu_{max}(x)]$$

$$\mu_D(M) = \sup_x[\mu_M(x) \wedge \mu_{min}(x)]$$

$$\mu_T(M) = \frac{[\mu_R(M) + 1 - \mu_L(M)]}{2}$$

  Finalmente, substituindo cada número *Fuzzy* pela pontuação nítida correspondente, o método resulta em uma matriz com apenas dados precisos, permitindo assim que os métodos clássicos do MADM classifiquem as alternativas (Pachêco Gomes *et al.*, 2018) ; (Chen *et al.*, 2014).

5.5.3 Tomada de decisão com múltiplos atributos

Por meio do MADM, é possível resolver o problema de decisão aplicando uma matriz M, onde as linhas representam as alternativas (A), e as colunas são os atributos (T), que avaliam as alternativas. A matriz M pode ser expressa como (Chen *et al.*, 2014):

$$M_{m,n} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{pmatrix}$$

Onde temos que: $A = \{a_1, a_2, \cdots, a_m\}$, $T = \{t_1, t_2, \cdots, t_n\}$, e $x_{i,j}$, com $i = 1, 2, \cdots, m$ e $j = 1, 2, \cdots, n$, são os valores de cada atributo.

5.5.4 Técnica para Preferência de Ordem por Similaridade à Solução Ideal - TOPSIS

Para classificar a solução ideal, foi aplicado o algoritmo baseado em TOPSIS, partindo do conceito central de que essa abordagem é capaz de encontrar a melhor alternativa por meio da distância euclidiana mais próxima da solução ideal. As principais etapas do TOPSIS são mostradas abaixo no algoritmo 3 (Chen *et al.*, 2014); (Pachêco Gomes *et al.*, 2018):

Algoritmo 3 – Algoritmo TOPSIS

***Passo 1**: Calcular a matriz de decisão normalizada*

$$r_{i,j} = \frac{x_{i,j}}{\sqrt{\sum_{k=1}^{n} x_{k,j}^2}}, j = 1, \cdots, m$$

***Passo 2**: Calcular a matriz de decisão normalizada ponderada*

$$v_{i,j} = w_j r_{i,j},$$

with $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n$.

***Passo 3**: Determinar as soluções ideais positivas e negativas*:

$$A^+ = \left\{ \left(\max_i v_{i,j} \mid j \in J \right), \left(\min_i v_{i,j} \mid j \in J' \right) \right\} \tag{5.1}$$

Onde $= 1, 2, \cdots, m$ e J são os atributos de benefício e J' são os atributos de custo. Para o problema de decisão, foi feito uma otimização para maximizar J e minimizar J'. O mesmo se aplica ao ideal negativo.

***Passo 4**: Calcular as medidas de separação*

Nesta etapa, a distância de cada alternativa da solução positiva e negativa ideal deve ser calculada.

***Passo 5**: Calcular a proximidade relativa à solução ideal*

A proximidade relativa é definida como:

$$c_{i^+} = \frac{s_{i^-}}{s_{i^+} + s_{i^-}}, 0 < c_{i^+} < 1, i = 1, 2, \cdots, m$$

***Passo 6**: Classifique a ordem de preferência*

Por meio da proximidade relativa c_{i^+} de cada alternativa é possível classificá-las.

5.6 Base de Conhecimento e Regras

5.6.1 Base de conhecimento Fuzzy: Atributos do sistema

As variáveis de entrada do sistema de Atenção Visual e, que são importantes para a modelagem da base de conhecimento, são mapeadas para os seguintes conjuntos *Fuzzy*:

5.6.1.1 Regiões de Interesse Fuzzy

No gráfico da Figura 69, é possível observar as regiões de interesse *Fuzzy* em seus termos linguísticos e que são geradas para representar a Atenção Visual em seus níveis de importância. Sendo estas regiões definidas como: Caixa Delimitadora de Prioridade Máxima (*CDPM*), Caixa Delimitadora de Prioridade (*DCP*), Caixa Delimitadora de Prioridade Média (*CDPM*), Caixa Delimitadora de Prioridade Baixa (*CDPB*) e Caixa Delimitadora de Prioridade Muito Baixa (*CDPMB*).

Por meio da Figura 65-(a), é possível observar o trapézio gerado para definir a área de navegação (Equação 5.2). Todo sinal de trânsito detectado no trapézio tem prioridade mais alta do que os que estão fora. Assim, quanto mais próximo o sinal de trânsito estiver da borda da via, ou fora dela, a função trapezoidal deve gerar um valor menor (Figura 65-(b)). As informações das regiões prioritárias *Fuzzy* (Figura 60) e, da área navegável trapézoidal (Figura 65), são combinadas para definirem as regiões de maior interesse para a Atenção Visual.

$$trap(x;a,b,c,d) = max(min(\frac{x-a}{b-a},1,\frac{d-x}{d-c}),0) \tag{5.2}$$

5.6.1.2 Distâncias Fuzzy

O recurso de distâncias *Fuzzy* está vinculado às distâncias euclidianas analisadas do sinal de trânsito em questão, em relação ao conjunto de outros objetos de emergência na cena (cones, sinais de trânsito verticais, semáforos) (Figura 62). As distâncias euclidianas dos objetos, são aplicadas a uma análise relacionada com as suas conexões, tornando então possível avaliar semanticamente a importância do sinal de trânsito detectado em relação ao contexto dos outros elementos de sinalizações e estruturação da via (Figura 62).

5.6.1.3 Fator de Conectividade

O fator de conectividade está ligado ao grau da conexão entre os sinais detectados em conflito de informação e os outros elementos da cena (Figura 70) (cones, área navegável, sinais de trânsito dentro da via, sinais de parada obrigatória e pessoas). O fator de conectividade é dado por duas variáveis:

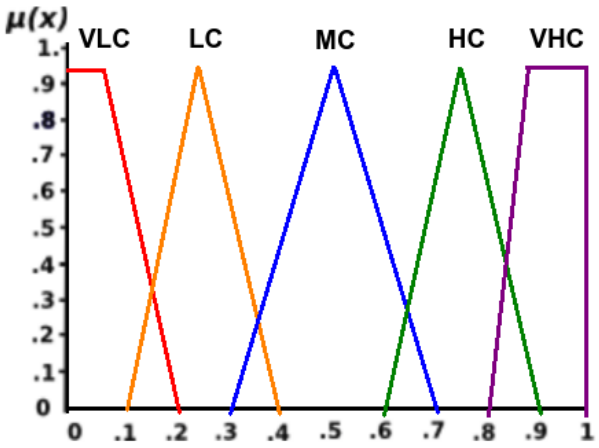
Tabela 2 – Base de regras *Fuzzy* - peso das conexões

Base de regras *Fuzzy*		
Aumento de velocidade	<	Redução de velocidade
Cone + Parada obrigatória	>	Semáforo verde
Vire a direita + Humano	>	Vire a esquerda

A) Essa análise é feita por meio da relação entre as distâncias das regiões de interesse (cada distância tem um custo de: $HPBB = 0,2$; $PBB = 0,2$; $MPBB = 0,2$; $LPBB = 0,2$; $VLPBB = 0,2$) (Figura 60-(b)).

B) E por último, o valor do peso de cada conexão que é dado pelo nível de importância de cada elemento na cena em relação ao sinal de trânsito detectado (por exemplo: cone = 1, pessoas = 0,9, STOP = 0,8, redução de velocidade = 0,7 e sinal dentro da rodovia = 0,9) (Tabela 2).

Figura 70 – Fator de conectividade dos sinais de trânsito - Legenda: Conectividade Muito Baixa (*CMB*), Conectividade Baixa (*CB*), Conectividade Média (*CM*), Conectividade Alta (*CA*), Conectividade Muito Alta.



Fonte: Elaborada pelo autor.

5.6.2 Base de regras Fuzzy

Nesta subseção, é apresentada a base de regras para que seja possível analisar cada sinal de trânsito com auxílio da base de conhecimento gerada.

Os atributos foram selecionados por meio das capacidades de detecção e classificação de sinais de trânsito, provenientes do sistema de percepção com fusão de imagens $2D$ e $3D$. Também foi aplicado o conhecimento especialista e experiência do autor sobre problemas de Atenção Visual Robótica e humana para ambientes de trânsito. Os problemas mais comuns foram listados para fazer parte dos atributos ($t1$, $t12$..., tn) junto ao algoritmo TOPSIS, permitindo

realizar a classificação da solução ideal com base nas alternativas bem definidas ($a1$, $a2$..., an). Os atributos e alternativas são listados na Tabela 3.

Tabela 3 – Atributos e alternativas para a análise e diagnóstico.

Alternativas	**Definição**
a_1	Seleciona Sinal de Trânsito A
a_2	Seleciona Sinal de Trânsito B
a_3	Seleciona Sinal de Trânsito C
Atributos	**Definição**
t_1	Sinal de emergência detectado
t_2	Sinal de trânsito dentro da via
t_3	Detecção de pessoa perto do sinal de trânsito
t_4	Sinal de trânsito conectado com elementos de emergência
t_5	Redução de velocidade
t_6	Distância para o agrupamento de sinais de emergência

Os atributos (t), são baseados nas características importantes de um sinal de trânsito a ser avaliado em relação a semântica da cena. As alternativas (a), definem a escolha do sinal de trânsito mais relevante para a navegação do veículo em um determinado trecho da via (Tabela 3), possibilitando o tratamento de informações em conflito. O modelo baseado em regras bem definidas, permitiu obter um Sistema de Inferência Inteligente - *Fuzzy Inference System* (FIS), capaz de classificar a prioridade de cada sinal de trânsito e, para esta tarefa, utilizando dados provenientes da análise de regiões de interesse e da base de regras *Fuzzy* em conjunto com o algoritmo TOPSIS (Algoritmo 3).

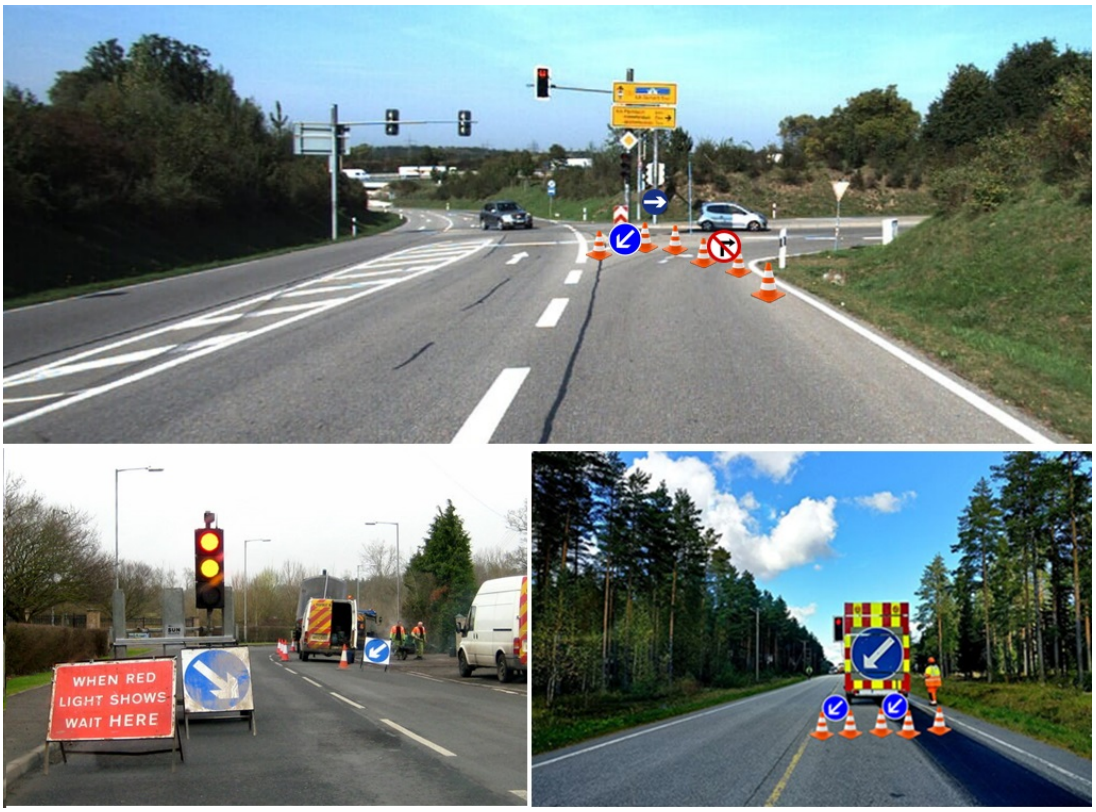
Tabela 4 – Termos linguísticos para os valores de atributos de diagnóstico.

Alternativas	**Definição**			
		a_1	a_2	a_3
t_1	SIM	muito bom	bom	muito ruim
	NÃO	muito ruim	bom	muito bom
t_2	SIM	muito bom	bom	muito ruim
	NÂO	muito ruim	bom	muito bom
t_3	SIM	muito ruim	ruim	muito bom
	NÂO	muito bom	bom	muito ruim
t_4	SIM	muito bom	bom	muito ruim
	NÃO	ruim	bom	muito bom

5.7 Dataset

Para a realização dos testes do sistema de Atenção Visual, foi criado um conjunto de dados para imagens de trechos de vias, com: acidentes de trânsito, manutenção e mudança de rota com desvios, intitulado como: **Cenas com Sinais de Trânsito de Emergência** - (CSTE). Esse tipo de imagem não é facilmente encontrado em conjuntos de dados padrão para veículos autônomos e, por conta deste problema, a base de dados foi criada com imagens diversas e que incluem: Imagens do Google, imagens do KITTI, imagens do Waymo, imagens capturadas nas ruas brasileiras com o CaRINA 2 e imagens editadas.

Figura 71 – Cenas de testes para o sistema de Atenção Visual *Fuzzy*.



Fonte: Elaborada pelo autor.

5.8 Considerações

Neste capítulo, foi apresentado um modelo de Atenção Visual capaz de analisar e classificar o nível de prioridade de sinais de trânsito detectados em conflitos de regras de trânsito. Por meio de uma camada auxiliar baseada nos Sistemas de Inferência *Fuzzy*, e que está ativa entre o sistema de percepção de sinais de trânsito e o sistema de tomada de decisão do veículo, foi possível analisar cada sinal de trânsito detectado. O modelo completo e, que é capaz de

processar cada sinal de trânsito em função de sua prioridade, é nomeado nesta tese como Atenção Visual. Para que este modelo funcione de maneira efetiva, os dados de percepção gerados pela detecção e classificação dos sinais de trânsito e, que são provenientes da fusão de imagens $2D$ e $3D$, são fornecidos *a priori*.

Para que a classificação da prioridade de cada sinal de trânsito detectado fosse possível, foi aplicado um modelo de decisão baseado em Múltiplos Atributos de Decisão (*Multiple Attribute Decision-Making* - MADM) e conjuntos *Fuzzy*. O Processo de Hierarquia Analítica (*Analytic Hierarchy Process* - AHP), foi aplicado para calcular os pesos dos atributos e, a técnica para a Ordem de Preferência por Similaridade com a Solução Ideal (*Technique for Order of Preference by Similarity to Ideal Solution* - TOPSIS), aplicada para classificar os sinais de trânsito em seus níveis de prioridade.

O sistema de Atenção Visual foi capaz de lidar com as situações de problemas propostos e, também, dar suporte para o sistema de tomada de decisão do veículo com base nas regras de trânsito locais avaliadas.

Outra situação que foi tratada por meio do modelo de Atenção Visual desenvolvido, está relacionada com as bifurcações e seus sinais de trânsito correspondentes. Sem este tipo de análise conjunta entre os sinais de trânsito e a área navegável, não seria possível definir quais sinalizações devem serem obedecidas em situações de bifurcações que não consideram um mapa *a priori*.

Destacando que outros trabalhos relacionados com esta tese e, que também trabalham com percepção de sinais de trânsito verticais, não fazem estes tipos de análises apresentadas: (a) **Classificação da prioridade dos sinais detectados em conflito** e de (b) **Bifurcações e trechos auxiliares**, o que dificulta a tomada de decisão do veículo com base nas regras de trânsito locais em situações não mapeadas.

CAPÍTULO

6

EXPERIMENTOS, RESULTADOS E DISCUSSÕES

Neste capítulo, são apresentados os experimentos, discussões e resultados do sistema desenvolvido para detecção e classificação de sinais de trânsito utilizando fusão de dados $2D$ e $3D$ e, também, do modelo de Atenção Visual *Fuzzy*. Os experimentos são apresentados em etapas, e foram realizados em prol da avaliação e validação de cada técnica.

Este capítulo é dividido em três tópicos de experimentos e resultados: (a) Detecção de sinais de trânsito em imagens $2D$, (b) Filtragem de falsos positivos e falsos negativos em imagens $3D$ e (c) Atenção Visual *Fuzzy*. Por último, é apresentada uma discussão sobre os resultados, gerando comentários sobre os pontos fortes e dificuldades encontradas no desenvolvimento desta pesquisa de doutorado.

Figura 72 – *Intersection over Union* - (IoU) para detecção de sinais de trânsito verticais.

(a) Ruim **(b) Bom** **(c) Excelente**

Fonte: Elaborada pelo autor.

6.1 Detecção 2D

6.1.1 Experimentos para detecção de sinais de trânsito verticais

Nesta seção, são apresentados os experimentos e resultados gerados por meio da detecção de sinais de trânsito verticais com imagens $2D$. A principal métrica para avaliação do sistema de detecção em imagens $2D$, é baseada nos resultados da Intersecção pela União (*Intersection over Union* - IoU), onde a detecção é comparada diretamente ao *ground truth* (Figura 72).

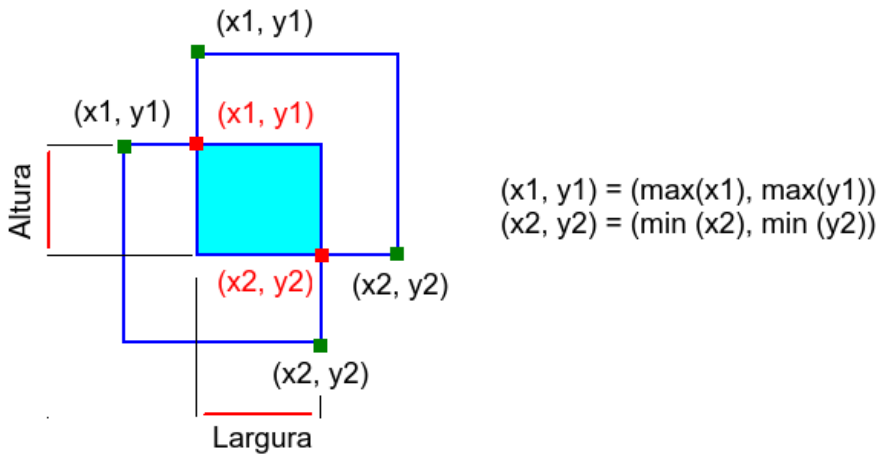
Por meio da Figura 72, é possível observar três situações na qualidade de detecção utilizando a rede de *Deep Learning* YOLO (*You Only Look Once*) : (a) Ruim, (b) Bom e (c) Excelente. Para gerar cada valor do IoU, foi aplicado um algoritmo para a comparação entre o *ground truth* e o objeto detectado, representando o sinal de trânsito. O algoritmo de avaliação é baseado na equação de quantificação (Equação 6.1):

$$IoU = \frac{A \bigcap B}{A \bigcup B} \tag{6.1}$$

Onde é dado que cada $A \cap B$, é gerada pela intersecção do *bounding box* do *ground truth* e da predição de detecção. Já a $A \cup B$, é gerada pela união do *bounding box* do *ground truth* e da predição de detecção.

Para exemplificar melhor como é feito o cálculo destes valores, é possível ser observado por meio da Figura 73, a sobreposição e união de maneira gráfica.

Figura 73 – *Técnica Intersection over Union* - (IoU).



Fonte: Adaptada de (RESTREPO, 2018).

Para o cálculo, é necessário levar em consideração a largura e a altura da intersecção. Se os valores de largura e altura forem positivos, então o valor da sobreposição (SP) é dado por largura vezes altura (Equação 6.2):

$$SP = (x2 - x1) * (y2 - y1) \tag{6.2}$$

Caso contrário, a sobreposição recebe o valor zero (RESTREPO, 2018). O último passo, é calcular a região combinada e, para isso, o seguinte cálculo é aplicado (Equação 6.3):

$$RC = area(GT) + area(P) - SP \tag{6.3}$$

Onde a variável GT representa o valor do *ground truth*, P representa o valor da predição da detecção e SP representa o valor da sobreposição calculada.

Essa métrica de avaliação foi aplicada para cada imagem que foi feita a detecção de sinais de trânsito em $2D$. Por meio dessa avaliação, foi possível quantificar os valores de detecção que são apresentados na próxima seção.

6.1.2 Resultados para a detecção de sinais de trânsito verticais

A média de resultados para o IoU apresentou valores em torno de 78% (Tabela 5), sendo que uma boa taxa de acerto para essa tarefa é considerada igual ou maior que 50%. Destacando que para o sistema de percepção desenvolvido, o detector $2D$ deve apenas informar a localização (x, y) do sinal de trânsito que foi detectado na cena, dando suporte para as etapas seguintes, de filtragem e classificação.

Figura 74 – IoU para a detecção de sinais de trânsito em imagens $2D$ no *Dataset* de *Tracking* do KITTI.

Fonte: Elaborada pelo autor.

Por meio da Tabela 5, podem ser observados os resultados de treinamento da rede de *Deep Learning* YOLO em função do número de épocas e, levando também em consideração,

Tabela 5 – Comparação do treinamento em função do número de iterações.

Comparação para o número de épocas de treinamento			
Iterações	**Avg Loss**	**mAP (%)**	**IoU (%)**
1000	0.2933	36.26	25.31
2000	0.2722	44.35	31.74
3000	0.2601	55.96	38.33
4000	0.2057	65.81	45.56
5000	0.1549	80.57	53.91
10000	0.1045	86.43	62.41
15000	0.0497	89.01	72.55
20000	**0.0421**	**91.85**	**78.21**
25000	0.0519	90.39	77.61

os seguintes parâmetros: *Avg Loss*, mAP [1] e IoU [2].

- **Treinamento:** Uma característica importante observada nos testes da rede YOLO, é que os sinais de trânsito rotulados para treinamento devem estar presentes e relacionados junto ao restante de sua cena completa. Sendo que imagens recortadas e rotuladas somente com os sinais e suas informações equivalentes, geram resultados ruins para a tarefa de detecção. Essa característica é relacionada a dependência das redes de detecção com as informações do contexto da cena e, sem essas informações, o modelo não consegue obter o aprendizado necessário para um bom funcionamento;

- **Classificação:** Para a classificação das 21 classes de sinais de trânsito verticais, a rede VGG, habilitada originalmente para esta tarefa junto a rede YOLO, não apresentou bons resultados, gerando uma taxa bastante baixa de acerto, em média, de 70%. Visando este problema, a rede YOLO foi treinada para classificar todos os sinais de trânsito detectados como uma única classe ("sinal de trânsito"), deixando a tarefa de classificação para a última etapa, utilizando a rede *Inception-V3*;

- **Detecção:** Para a tarefa de detecção, a rede YOLO também apresentou melhores resultados considerando o treinamento com apenas uma classe representando todos os sinais de trânsito. A detecção com todas as 21 classes separadas, teve uma taxa de acerto em média (mAP) de 78%, contra 92% para o treinamento com uma classe única. Sendo que o erro gerado (8%), está relacionado com objetos falsos positivos e falsos negativos.

[1] maP: Valor médio da precisão em todos os valores de *recall*.
[2] IoU: Métrica para avaliação de detectores de objetos baseados em *Deep Learning* capaz de medir a taxa de sobreposição entre o *ground truth* e a detecção.

6.2 Filtragem 3D

Nesta seção, são apresentados os experimentos e resultados gerados pelo sistema de filtragem de sinais de trânsito com dados 3*D*. Os dados de profundidade gerados por este tipo de imagem, possibilitam que o filtro consiga eliminar falsos positivos e falsos negativos com uma maior precisão. Sem estes dados, não seria possível tratar imagens de sinais de trânsito falsas e, também, não seria possível estimar sua posição na cena e dar suporte para o sistema de Atenção Visual.

6.2.1 Experimentos para a filtragem de sinais de trânsito verticais

O algoritmo para filtragem de sinais de trânsito em dados 3*D*, utiliza um conjunto de possíveis formatos modelados na nuvem de pontos (Figura 75 - (b)) e que representam os sinais de trânsito verticais em situações que geralmente podem ser encontradas no ambiente de navegação (Figura 75 - (a)). Sendo que no ambiente urbano, nem sempre um sinal de trânsito é colocado em um poste individual, em algumas situações, ele pode ser encontrado em um poste compartilhado com outros tipos de informações, como por exemplo: placas de identificação de ruas, semáforos, ou até mesmo, junto com outros sinais de trânsito (Figura 75 - (a)).

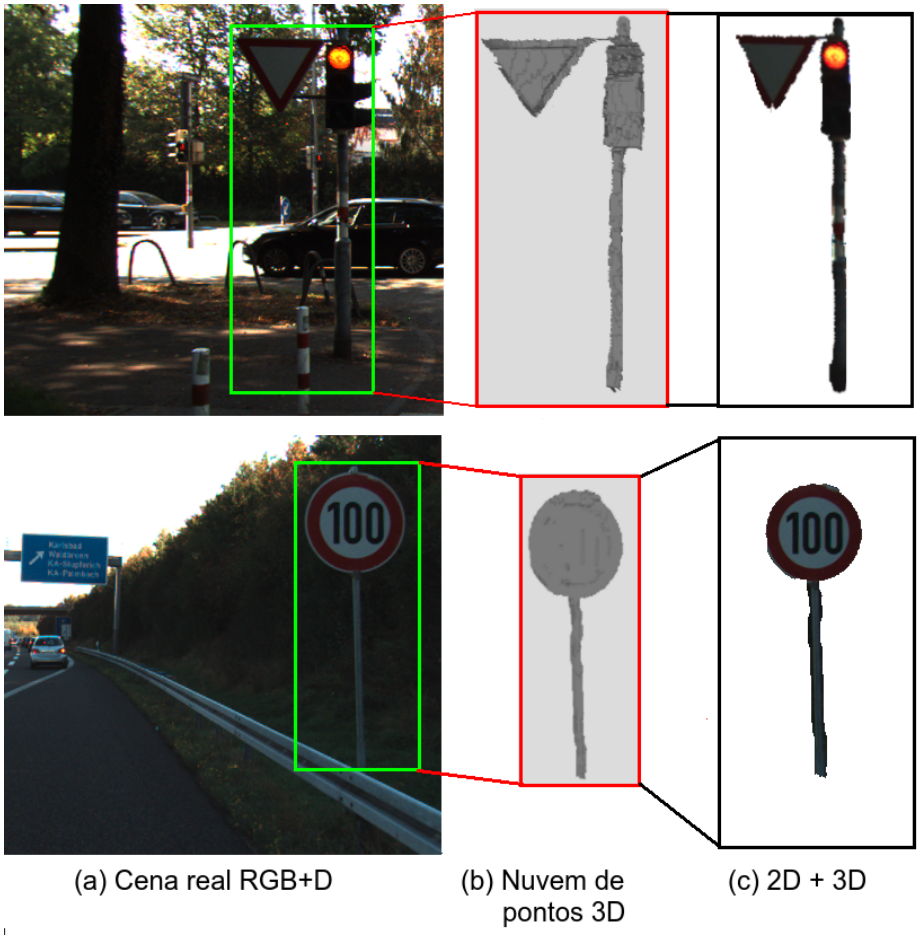
Os objetos no formato 3*D*, passam por uma extração de características por meio do método $3D - CSD$ (SALES D, 2017). A extração destes dados, permite que o método de classificação neural classifique os objetos detectados em duas classes: (a) sinais de trânsito e (b) não sinais de trânsito.

Para a tarefa de classificação, uma RNA-MLP com saída binária foi treinada com estes diferentes modelos de sinais de trânsito (Figura 75 - (b)). Para que isso seja possível, cada tipo de formato de sinal de trânsito foi modelado com base nos dados 3*D* do *Velodyne* LIDAR (*Light Detection and Ranging*) e, também, considerando os dados de um par de câmeras estéreo. Permitindo então que a RNA possa responder se um sinal de trânsito real foi detectado ou não, eliminando falsos objetos. A arquitetura da RNA utilizada para classificação de objetos 3D, utiliza uma *Multilayer Perceptron* de 30 entradas, 3 câmadas ocultas e 1 saída. A função de ativação utilizada foi a *sigmóide*, garantindo um melhor comportamento da rede.

6.2.2 Resultados para a filtragem de sinais de trânsito verticais

Os resultados dos testes do filtro 3*D* apresentaram uma acurácia de 88% (Figura 76). O erro gerado no filtro está relacionado a falsos positivos (8%) e falsos negativos (4%) provenientes de uma análise de objetos aplicada em imagens 3*D*. A filtragem de falsos positivos e falsos negativos, considera também em seu treinamento, objetos semelhantes em relação a suas formas, garantindo uma melhor capacidade de generalização: árvores, pessoas e carros.

Figura 75 – Filtragem de sinais de trânsito com fusão de dados 2*D* e 3*D*.



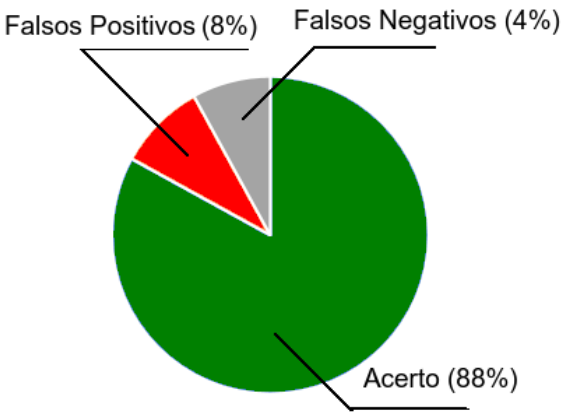
Fonte: Elaborada pelo autor.

Por meio da análise das matrizes de confusão (Tabela 6), pode-se concluir que 70% dos erros de classificação para falsos positivos e falsos negativos e, que estão relacionados com os sinais de trânsito verticais, acontecem nas classes “árvore” e “pessoas”. Destacando que os problemas gerados para classificação 3*D* e, que consideram as classes treinadas para estes tipos de objetos, estão muito relacionados com os *sprays*, gerando imperfeições na aglutinação da nuvem de pontos. Outro problema é que os sinais de trânsito, as pessoas e as árvores tem formatos muito parecidos.

Tabela 6 – Matrizes de confusão resultantes para 4 testes do filtro 3*D* - Relação da matriz de confusão: Sinais de trânsito, pessoas, árvores e carros.

Teste 1				**Teste 2**				**Teste 3**				**Teste 4**			
88	2	1	3	87	2	1	2	92	2	2	3	88	0	1	2
2	15	0	0	3	13	1	0	0	14	0	0	1	17	0	0
6	1	12	1	5	1	13	0	2	0	13	1	5	0	13	2
4	0	2	16	5	2	0	18	6	2	0	16	6	1	1	14

Figura 76 – Acurácia para o filtro de objetos 3D.



Fonte: Elaborada pelo autor.

Para que fosse possível testar o filtro 3*D* em sinais de trânsito falsos pintados e/ou colados na traseira de veículos (Figura 77), foi adaptado o *Dataset* de *tracking* do KITTI (GEIGER; LENZ; URTASUN, 2012) para este problema. Essa abordagem foi a mais próxima possível de um teste em imagens reais, visto que não existe um *Dataset* com problemas parecidos aos que são tratados nesta pesquisa. As adaptações feitas não impediram a qualidade dos testes do filtro em imagens 3*D*, já que as edições foram feitas apenas nas imagens *RGB* em 2*D*.

Para este tipo de problema de falsos sinais de trânsito, o filtro 3*D* apresentou uma acurácia de 100%, garantindo que todo sinal falso e, não relacionado com as regras de trânsito locais, fosse eliminado. Essa alta taxa de acerto foi possível graças a informação de profundidade, já que os sinais de trânsito colocados na traseira de veículos são todos na mesma profundidade dos mesmos, sendo então uma tarefa trivial para o filtro 3*D*.

Figura 77 – Teste do filtro 3*D* para sinais de trânsito falsos.

Fonte: Elaborada pelo autor.

6.3 Classificação 2D

Nesta seção, são apresentados os experimentos e resultados gerados pelo sistema de classificação de sinais de trânsito com imagens 2*D*. Os dados gerados por estas imagens, são fundamentais para esta tarefa, visto que suas informações de cores, texturas e formas são de grande potencial para classificação de imagens com redes de *Deep Learning*.

6.3.1 Experimentos para a classificação de sinais de trânsito verticais

O treinamento aplicado na rede de *Deep Learning Inception-V3* modificou apenas a camada final, gerando uma nova camada de classificação. Com isso foi possível preservar todos os filtros e convoluções genéricos para o reconhecimento de diferentes tipos de imagens com base em texturas, cores, formas e outros tipos de padrões.

Foram realizados 10 treinamentos para o sistema de classificação. Para que fosse possível realizar estes testes, foi utilizado o *Dataset* do INI (STALLKAMP *et al.*, 2011). A base de treino, teste e validação foram alteradas em todos os 10 treinamentos do classificador neural, sempre mantendo 70% para treinamento e 30% para teste e validação. Para isso, as imagens foram escolhidas aleatoriamente para gerar esses dois conjuntos (treinamento / teste). A precisão do sistema foi avaliada por meio das imagens de teste (Conjunto de teste de 30%), consequentemente, esse conjunto de dados não está contido nos dados de treinamento.

Todos os testes realizados tiveram uma taxa de reconhecimento superior a 96%, com uma média final de 98.1%. Este resultado indica que a rede de *Deep Learning Inception-V3* é muito eficiente, principalmente porque está trabalhando com imagens ricas em informações 2*D* e, também, com um grande conjunto de dados (muitos exemplos de sinais), incluindo imagens de variadas resoluções e várias situações ambientais (baixa luz, sombras, desfoque e sol contra) no momento da captura da imagem. Os resultados dos teste podem ser visualizados na Tabela 7.

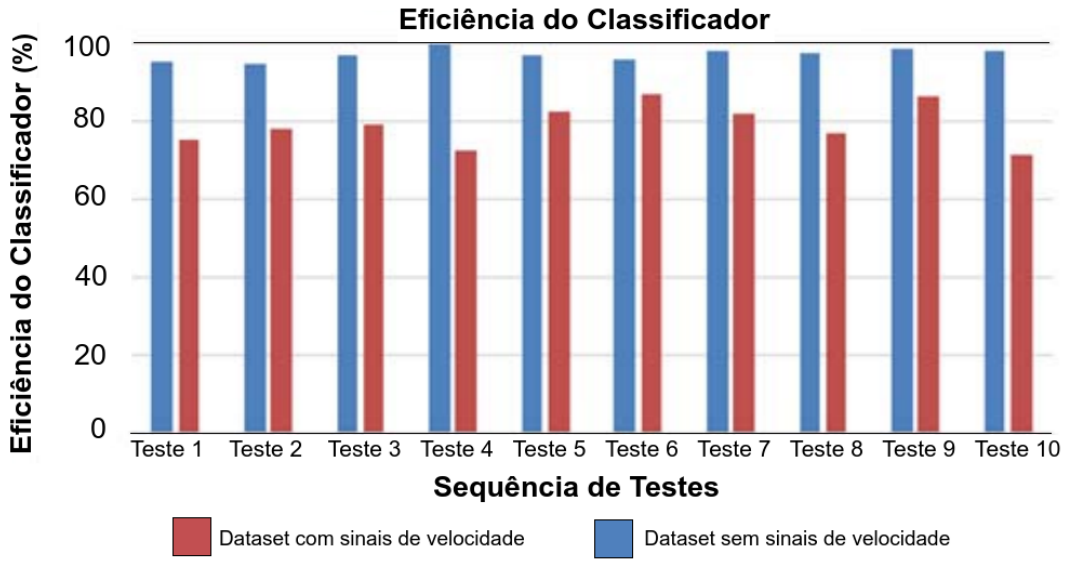
Tabela 7 – Conjunto de testes para treino e teste do classificador.

Conjunto de testes realizados	**Eficiência de reconhecimento (%)**
Treino 1	97.6
Treino 2	98.4
Treino 3	96.5
Treino 4	97.0
Treino 5	98.9
Treino 6	97.9
Treino 7	98.5
Treino 8	98.2
Treino 9	98.1
Treino 10	97.9
Média Geral	**98.1**

6.3.2 Resultados para a classificação de sinais de trânsito verticais

Por meio do gráfico da Figura 78, é possível observar o problema que os sinais de velocidade máxima geram para os resultados da classificação. A razão para a menor eficiência na classificação de sinais numéricos, está relacionada ao problema de oclusão (Figura 79). Sabendo então que nesses casos de limite de velocidade, é muito difícil reconhecer valores quando existe oclusão de dígitos por completo.

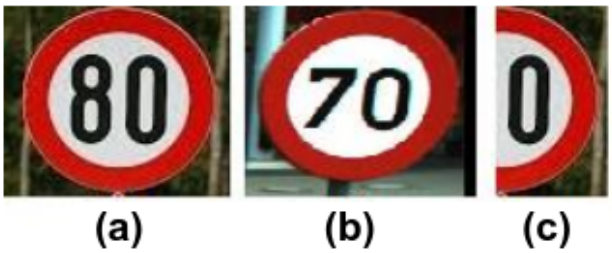
Figura 78 – Resultados do classificador neural para sinais de trânsito.



Fonte: Elaborada pelo autor.

Por meio da Figura 79, é possível observar uma situação em que dois sinais de velocidade têm zero em seu final (Figura 79 - (a) e (b)). Se apenas o final zero for detectado (Figura 79 - (c)), seria impossível classificar corretamente o sinal de velocidade, se levado em conta a finita quantidade de sinais numéricos com valor final zero. Este problema também aconteceria com um humano, visto que não seria possível estimar o primeiro digito.

Figura 79 – Problema grave de oclusão de sinais de trânsito (a) 80km (b) 70km e (c) problema de oclusão.



Fonte: Elaborada pelo autor.

Para que fosse possível testar o algoritmo neural de classificação, foram utilizadas algumas imagens com oclusão. Alguns exemplos para testes podem ser observados na Tabela da Figura 80.

Figura 80 – Classificação de sinais de trânsito com oclusão.

Sinal de Trânsito	Classe	Acurácia (%)
	PARE	99.1
	PARE	99.6
	PARE	98.3
	PARE	96.2
	PREFERÊNCIA	96.2
	PREFERÊNCIA	97.6
	PREFERÊNCIA	88.5
	PREFERÊNCIA	89.2
	SIGA EM FRENTE OU A DIREITA	94.1
	SIGA EM FRENTE OU A DIREITA	88.2

Fonte: Elaborada pelo autor.

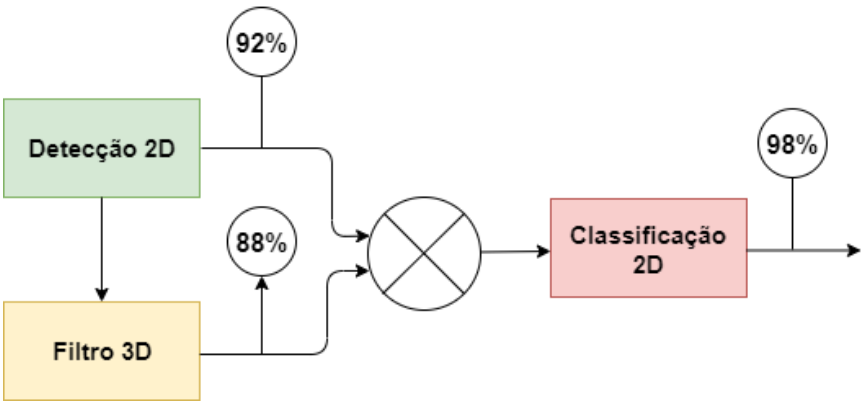
Em alguns casos, o algoritmo para detecção de sinais de trânsito em imagens $2D$ pode errar e detectar placas de maneira parcial. No entanto, esse problema também pode ocorrer em algumas situações naturais devido a uma real oclusão do objeto. Por exemplo, árvores em frente ao sinal de trânsito, vegetação da rodovia, deterioração das placas, ou até mesmo, outros tipos de objetos na cena (carro, caminhão, motocicleta e pessoas) gerando oclusão para o sistema de visão. Essas situações podem criar problemas para detectar o sinal de trânsito de maneira completa.

No entanto, a rede de *Deep Learning Inception-V3* funcionou muito bem em situações em que algumas informações não estão disponíveis para classificar uma imagem, ou seja, para sinais de trânsito com oclusão parcial em até 70% da área total da placa. O sistema de classificação utilizado exibe sempre os 5 primeiros sinais de trânsito que podem representar a imagem detectada. Os resultados aparecem de maneira decrescente, sempre dando a maior porcentagem à melhor estimativa. Para considerar uma boa classificação, foi definido que o primeiro lugar entre os cinco da classificação deve ter uma pontuação pelo menos 50% maior que o segundo colocado.

6.4 Análise dos Resultados de Percepção

Nesta seção, é apresentada a análise dos resultados do sistema de percepção para detecção e classificação de sinais de trânsito verticais. Por meio da Figura 81, é possível visualizar a comunicação dos resultados obtidos para detecção, filtragem e classificação.

Figura 81 – Comunicação entre os resultados do sistema de percepção com fusão de dados 2*D* e 3*D*.



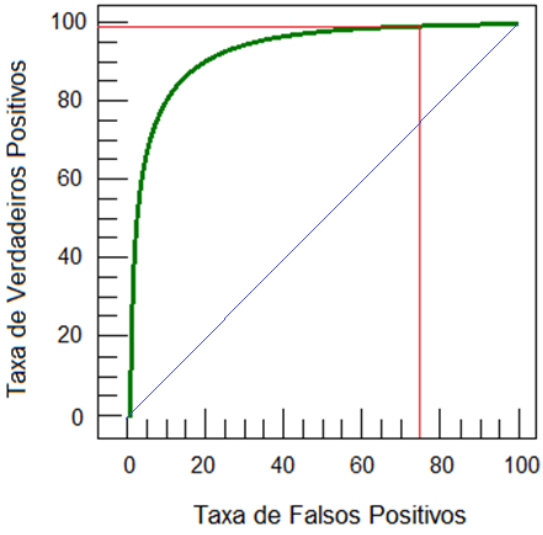
Fonte: Elaborada pelo autor.

- **Detecção 2D:** A detecção 2*D* utilizando a rede de *Deep Learning* YOLO, gerou uma taxa de acerto de 92%, possibilitando detectar sinais de trânsito com uma boa precisão (Figura 81);

- **Filtragem 3D:** O filtro 3*D* apresentou uma taxa de acerto de 88%, garantindo que sinais de trânsito falsos positivos e falsos negativos sejam eliminados por meio da filtragem proveniente da análise de informações de profundidade. Sendo que o filtro 3*D*, retorna apenas uma etiqueta com a taxa de certeza para cada sinal de trânsito detectado na etapa anterior. Declarando se é uma informação real ou falsa com valores de certeza entre 0% e 100% em uma função sigmóide na saída do classificador neural (Figura 81);

- **Taxa de certeza:** Um valor de corte para os valores da taxa de certeza e, que são provenientes da etapa de filtragem, pode ser definido, possibilitando eliminar os sinais de trânsito falsos que foram detectados. O valor de corte que foi aplicado para a taxa de certeza é de 75%, podendo ser observado por meio da curva de ROC (*Receiver Operating Characteristic Curve*) na Figura 82. Sendo que este valor baixo permite que um maior número de sinais falsos negativos sejam candidatos para a última etapa, a de classificação do tipo do sinal de trânsito. Garantindo então um maior número de informações de trânsito detectadas e classificadas pelo sistema de percepção;

Figura 82 – Curva de ROC: Taxa de TFP (Taxa de Falsos Positivos) e TVP (Taxa de Verdadeiros Positivos) ajustada para o modelo de percepção.



Fonte: Elaborada pelo autor.

- **Classificação 2D:** A classificação do tipo do sinal de trânsito detectado teve bons resultados, apresentando uma taxa de acerto de 98.1%. Para esta tarefa, foi aplicada a rede de *Deep Learning Inception-V3* sobre todas imagens candidatas para sinais de trânsito verticais e que foram selecionadas nas duas etapadas anteriores (detecção e filtragem).

6.5 Resultados para a Detecção de Rotas Auxiliares e Bifurcações

Com base na estimativa de pose do sistema ORB-SLAM (*Oriented FAST and Rotated BRIEF - Simultaneous Localization and Mapping*), é possível concatenar corretamente cada nuvem de pontos. O principal problema consiste em manter a distribuição dos pontos originais

no novo mapa de grade. Os intervalos das nuvens de pontos de entrada usados nos experimentos são mostrados na Tabela 8, aplicando $b = 0,1$ para o parâmetro de tamanho das células do mapa.

Tabela 8 – Intervalos de nuvem de pontos de entrada.

	Largura	**Altura**	**Profundidade**
Min	0.6	0.2	0.4
Max	7.0	2.0	10.0

Para avaliar essa distribuição, foi calculado a média e o desvio padrão da nuvem de pontos concatenada da maneira original e da nuvem de pontos da grade gerada. Na Tabela 9, pode ser observado que os valores médios do método de grade permanecem próximos da distribuição original.

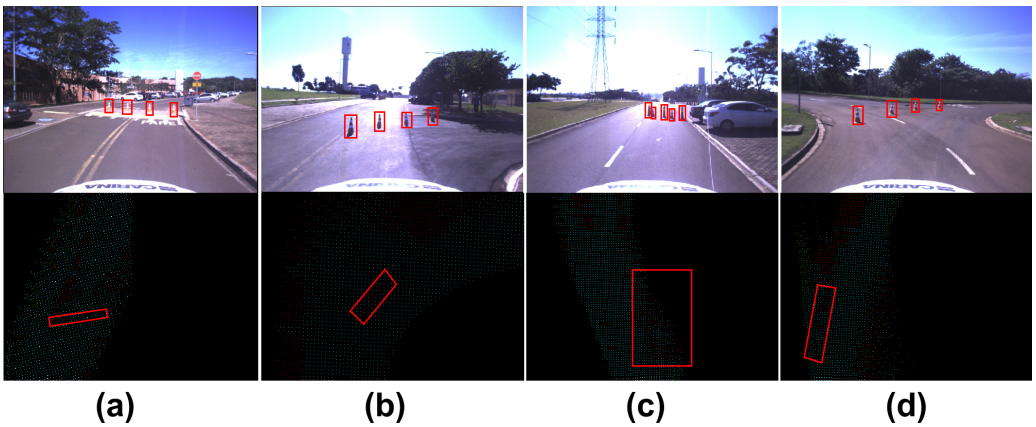
Tabela 9 – Média, desvio padrão e número de pontos da nuvem de pontos concatenados original e da nuvem de pontos da grade gerada.

	Nuvem de pontos completa			**Grade de nuvem de pontos**		
	Média	Variação	Pontos	Média	Variação	Pontos
X	1.9261	11.0822	424817	3.2194	21.0107	14923
Y	0.2347	0.0033	424817	0.2243	0.0049	14923
Z	-23.1317	97.4283	424817	-24.7042	110.3002	14923

A diferença na variação de x e z se deve ao fato de que, na nuvem de pontos original, ocorre um acúmulo de informações no centro da trajetória, enquanto na grade, todos os pontos são igualmente distribuídos no x e intervalos de z. Além disso, na variação y, existe a distribuição de altura que fica bem próxima da nuvem de pontos original.

Por meio da Figura 83, é possível observar os cones que são detectados com os dados de percepção provenientes do método de fusão de dados $2D$ e $3D$.

Figura 83 – Detecção de rotas auxíliares por meio de sinais de atenção visual em forma de cones.



Fonte: Elaborada pelo autor.

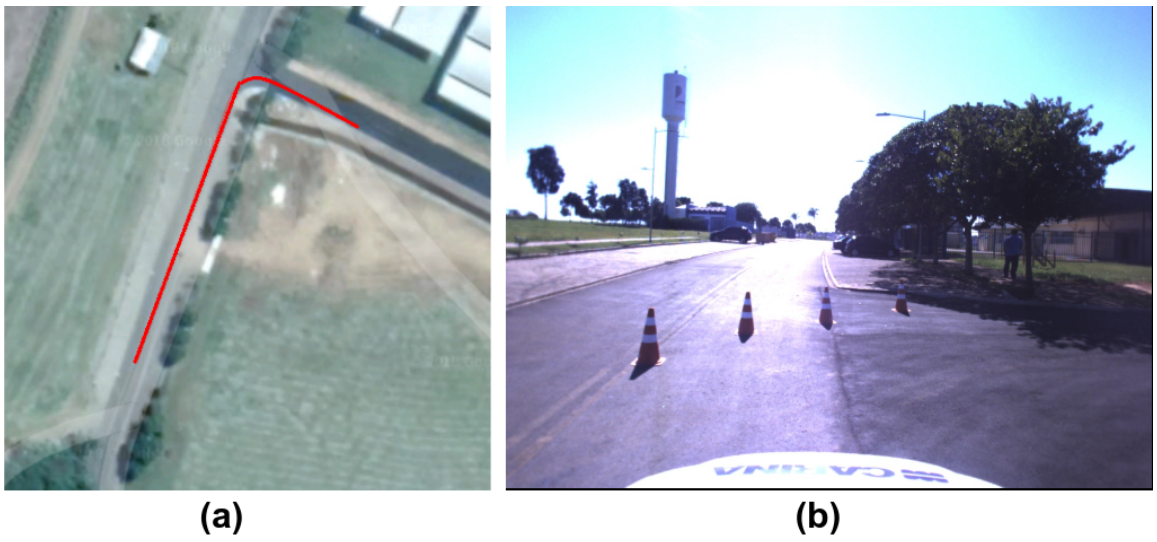
Nos dados de imagem $3D$, também é gerado o mapa de grade com as trajetórias. Os cones são identificados e mapeados como obstáculos positivos, conforme marcado nas caixas vermelhas, e as rotas não mapeadas geram rotas auxiliares.

Com a concatenação da nuvem de pontos, os resultados mostram que existem pontos alocados igualmente nas faixas de largura e profundidade e mantendo a distribuição da altura. Além disso, pode-se reduzir o volume de pontos aproximadamente em 96%, armazenando menos dados e mantendo a distribuição original.

Os resultados mostram que por meio da detecção da via, o sistema de percepção e mapeamento é capaz de identificar o bloqueio e encontrar uma possível rota auxiliar, podendo ser usada para o replanejamento de rotas e, também, dando suporte para o modelo de Atenção Visual interpretar novos sinais não mapeados.

À medida que o veículo navega pelo ambiente, cada vez mais características são extraídas e salvas. Os recursos já mapeados foram usados para estimar a posição real no mapa de quadro. Na Figura 84, é possível observar uma comparação entre a localização ORB-SLAM com base nas imagens estéreo e a visualização de satélite do Google *maps* da trajetória executada no Campus 2 da USP de São Carlos. Também é possível observar a reta que foi impedida pelos cones e a rota auxiliar detectada a direita para dar suporte ao replanejamento de rota.

Figura 84 – (a) - Pontos de trajetória ORB-SLAM em vermelho sobrepostos na visualização da trajetória do Google *Maps* e (b) Curva detectada pelo sistema de visão computacional embarcado no carro.



Fonte: Elaborada pelo autor.

6.6 Atenção Visual Fuzzy

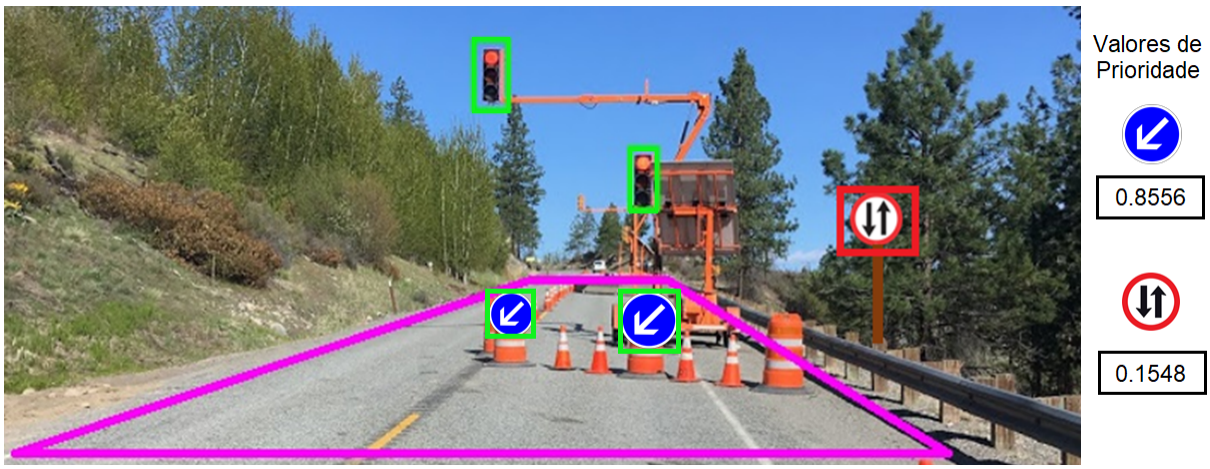
Nesta seção, são apresentados os experimentos e resultados gerados pelo modelo de Atenção Visual proposto para analisar a prioridade dos sinais de trânsito, focando em situações de conflito de informações. Os testes para situações de emergência na via (obras, acidentes e trechos não mapeados), foram realizados em cenas de trânsito reais e adaptadas, já que este tipo de situação é de difícil captura, visando o motivo de serem temporárias. Outro problema é que não existe nenhum *Dataset* voltado para este tipo de problema, sendo assim, as imagens utilizadas para o modelo de Atenção Visual são provenientes de várias fontes: Google imagens, *KITTI - Dataset* (GEIGER; LENZ; URTASUN, 2012), *INI - Dataset* (STALLKAMP *et al.*, 2011), *Waymo - Dataset* (WAYMO, 2020) e imagens capturadas pelo CaRINA 2 (CARINA, 2020).

6.6.1 Experimentos para o modelo de Atenção Visual Fuzzy

Os experimentos aplicados para o modelo de Atenção Visual são voltados para situações onde conflitos de sinais de trânsito são gerados. Para que um conflito de informações seja analisado efetivamente, é avaliado se mais de uma informação foi detectada por meio do sistema de percepção para a mesma regra: velocidade máxima, sentido da via, parada obrigatória e semáforo verde.

Por meio da Figura 85, pode ser observada uma situação onde dois sinais de trânsito estão em conflito: (a) Mão dupla e (b) Mantenha-se à esquerda, onde este último anula a navegação na faixa da direita, gerando conflito entre as duas informações. E ainda existe o semáforo temporário que vai regular o trânsito. Sendo que o sinal de trânsito de "Mão dupla" é o convencional para as regras locais, não sendo usual para a situação de emergência em questão.

Figura 85 – Situação com sinais de trânsito de sentido da via em conflito.



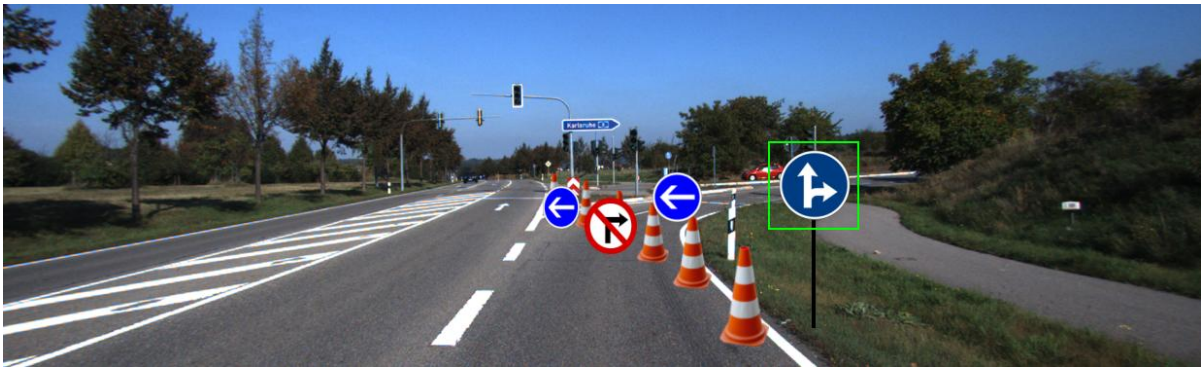
Fonte: Elaborada pelo autor.

Para este tipo de situação, em algumas rodovias, os sinais convencionais são tampados, no entanto, nem sempre esse tipo de cuidado é tomado, gerando confusão até mesmo para a navegação por um motorista humano.

6.6.2 Resultados para o modelo de Atenção Visual Fuzzy

As redes de *Deep Learning* para detecção, focam seus algoritmos de percepção para priorizar objetos que são representados por imagens com maior tamanho e nitidez, processando estes dados em função do volume de *pixels* relacionados a cada elemento da cena. Essa característica é capaz de fazer com que os objetos que mais se destacam na cena, sejam detectados com uma maior taxa de certeza, consequentemente, atribuindo uma maior relevância para este tipo de informação. No entanto, para a Atenção Visual, esse potencial presente em redes de detecção, pode gerar problemas semânticos relacionados com a tarefa de analisar a prioridade real de cada sinal de trânsito (Figura 86).

Figura 86 – Sinais de trânsito alterados em uma situação de emergência.



Fonte: Elaborada pelo autor.

Por meio da Tabela 10, podem ser observados os valores das taxas de certeza para a detecção $2D$ em comparação com a análise de prioridades *Fuzzy* relacionados com cada sinal de trânsito presente na cena da Figura 86.

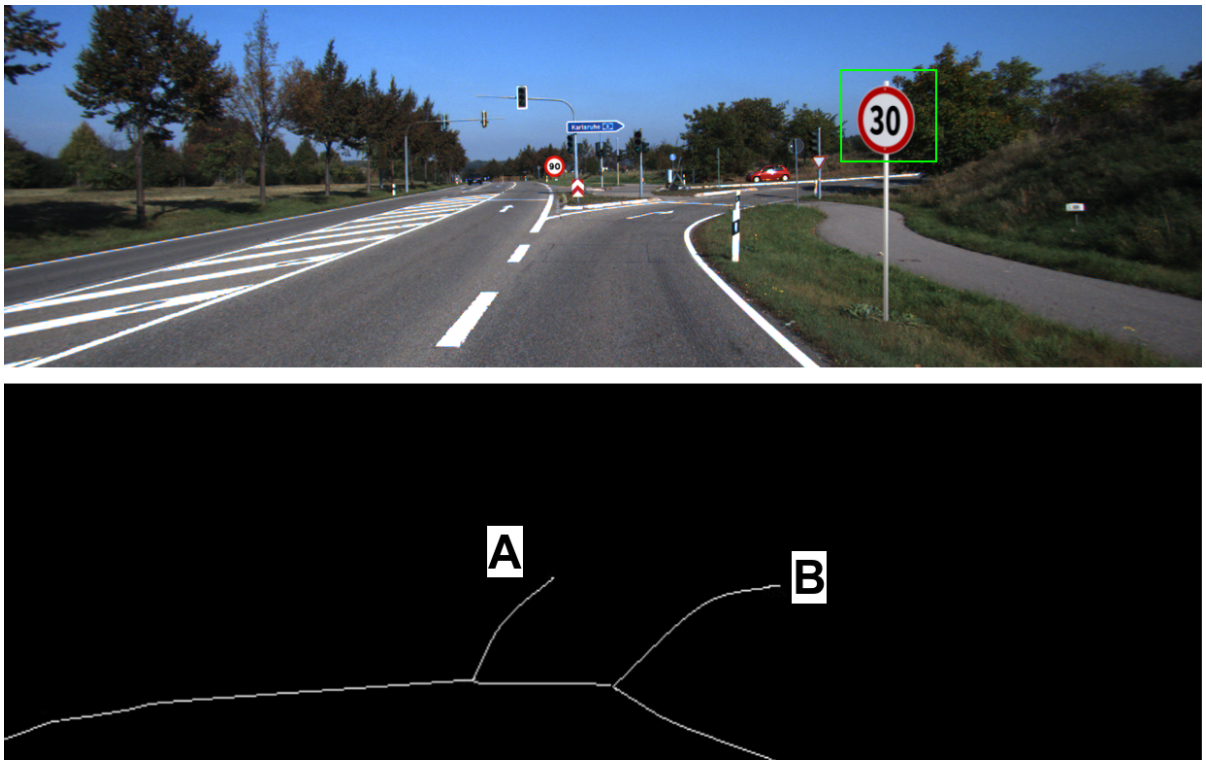
Tabela 10 – Relação dos resultados da análise de prioridades entre detecção e Atenção Visual - Cena 1.

	Comparação para taxas de certeza: Detecção e Atenção Visual	
	Detecção com YOLO (%)	**Atenção Visual (%)**
Siga em Frente ou Vire à Direita	94	14
Proibido Virar à Direita	80	86
Mantenha-se à Esquerda 1	74	86
Mantenha-se à Esquerda 2	69	86

Os dados gerados pelos testes de percepção, mostram que o sinal habitual de "Vire a direita ou siga em frente" foi detectado pela rede de *Deep Learning* YOLO com uma maior taxa de certeza (94%) se comparado aos outros sinais de sentido da via, respectivamente (80%, 74% e 69%). Para esta situação, a rede de detecção gerou uma taxa de certeza maior para uma informação que não é mais válida. No entanto, o modelo de Atenção Visual classificou a prioridade corretamente (Tabela 10).

Por meio da Figura 87, pode ser observada outra situação, onde por meio da Atenção Visual deve ser possível avaliar para qual lado da bifucação (*A* ou *B*) o sinal de trânsito detectado pertence. Se em modo autônomo, sabe se vai seguir o caminho *A* ou *B* (planejamento de rota). Se humano, tem que esperar para ver (esterçamento da direção e seta).

Figura 87 – Sinais de trânsito relacionados com bifurcações correspondentes.



Fonte: Elaborada pelo autor.

Neste instante da cena (Figura 87), apenas o sinal de 30Km/h foi detectado pelo sistema de percepção. Porém, esta informação gera conflito nas rotas possíveis, pois só deve ser obedecida se o planejamento de rota for escolher a via à direita (Caminho *B*). Caso seja escolhido seguir em frente (Caminho *A*), o sinal de 30Km/h deve ser desconsiderado e a velocidade deve ser mantida considerando informações anteriores, visto que o próximo sinal de velocidade (90/Km/h) no caminho *A* está muito longe (+/- 200 metros) e ainda não pode ser detectado por meio do sistema de percepção. Para este tipo de caso, foi necessária a análise da área navegável, possibilitando detectar a bifurcação e relacionar os sinais de trânsito equivalentes para cada rota.

Para esta cena, a rede YOLO foi capaz de detectar com uma boa taxa de certeza o sinal de 30Km/h (93%), no entanto, não tendo a capacidade de analisar para qual sentido da via a informação deve ser obedecida, destacando o problema de que nem sempre o sinal detectado mais próximo deve ser mais importante para a tomada de decisão do veículo. O modelo de Atenção Visual declarou 100% de prioridade para o limite de velocidade em questão, caso a rota escolhida seja a *B*, visto que foi o único sinal detectado neste instante (Tabela 11). No entanto, considerando que a rota escolhida fosse seguir em frente (Caminho *A*), essa informação deve ser desconsiderada, recebendo a taxa de prioridade de 0%.

Tabela 11 – Relação dos resultados da análise de prioridades entre detecção e Atenção Visual - Cena 2.

	Comparação para taxas de certeza: Detecção e Atenção Visual	
	Detecção com YOLO (%)	**Atenção Visual (%)**
Limite de Velocidade: 30km/h	93	100
Limite de Velocidade: 90km/h	0	0
Semáforo Vermelho	0	0

Por meio da Figura 88, pode ser observada uma situação de obras na pista, onde um sinal móvel de "Parada obrigatória" é colocado para revesar a travessia pelo único lado navegável que ficou disponível. Para este caso, foram detectados com uma boa precisão dois sinais de trânsito em conflito: o sinal de "Parada Obrigatória" (92%) e o semáforo verde mais próximo (72%).

Figura 88 – Situação de obras na pista - Sinais de trânsito móveis.

Fonte: Elaborada pelo autor.

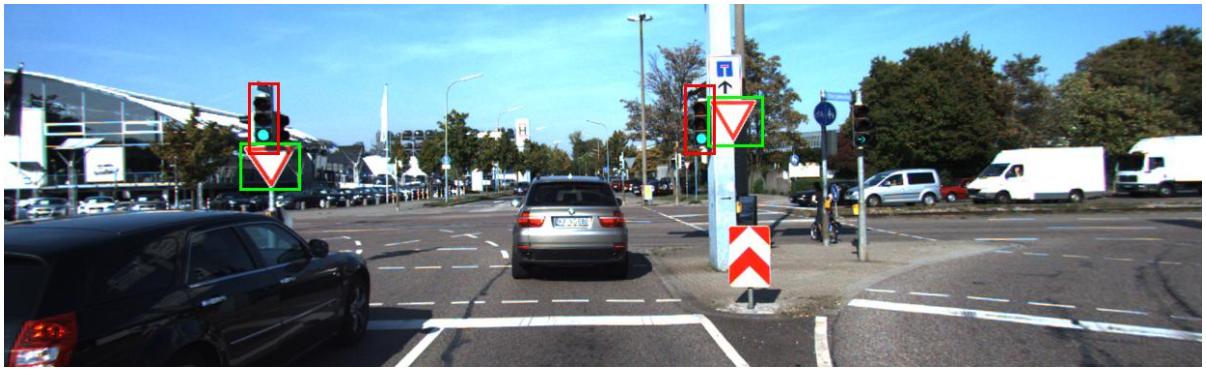
Nesta situação, o sinal de "Parada obrigatória" recebeu uma maior taxa de prioridade pela Atenção Visual por estar dentro da via e mais próximo a uma pessoa, cones e ao veículo, assumindo um valor de 90% (Tabela 12). Para esta cena, a detecção 2*D* e a análise de prioridades geraram valores parecidos, no entanto, é considerado sempre o valor da Atenção Visual, visando os problemas apresentados anteriormente e que estão relacionados com as redes de *Deep Learning*.

Tabela 12 – Relação dos resultados da análise de prioridades entre detecção e Atenção Visual - Cena 3.

	Comparação para taxas de certeza: Detecção e Atenção Visual	
	Detecção com YOLO (%)	**Atenção Visual (%)**
Parada Obrigatória	92	90
Semáforo Verde 1	72	10
Semáforo Verde 2	0	0

Por meio da Figura 89, pode ser observada uma última situação onde caso o semáforo esteja verde, o sinal "Dê a preferência" deve ser ignorado. Caso o semáforo esteja desligado, o sinal "Dê a preferência" deve ser obedecido. Para este problema, o modelo de Atenção Visual deve avaliar estas duas situações com base no conhecimento *a priori* de que os dois sinais de trânsito são dependentes para a mesma regra que controla o cruzamento.

Figura 89 – Situação de conflito de sinais de trânsito.



Fonte: Elaborada pelo autor.

Por meio da Tabela 13, podem ser observados os valores das taxas de certeza e de prioridade para uma análise aplicada na cena em questão (Figura 89). Todos os sinais foram detectados com uma boa precisão, no entanto, os valores de prioridade para os semáforos verdes são máximos, já que esta regra tem prioridade sobre o sinal "Dê a preferência" em uma situação concomitante.

Tabela 13 – Relação dos resultados da análise de prioridades entre detecção e Atenção Visual - Cena 4.

	Comparação para taxas de certeza: Detecção e Atenção Visual	
	Detecção com YOLO (%)	**Atenção Visual (%)**
Dê a preferência 1	95	0
Dê a preferência 2	92	0
Semáforo Verde 1	89	100
Semáforo Verde 2	86	100

Os resultados gerados para estas quatro cenas de exemplo (Figura 86, 87, 88 e 89) e que representam os problemas tratados nesta tese, mostram que por mais que as redes de *Deep Learning* tenham um grande potencial para detecção e classificação em imagens 2*D*, ainda apresentam muitos problemas relacionados com a interpretação dos dados gerados e que representam as informações extraídas da cena por meio da percepção.

Estes tipos de problemas são muito difíceis de serem tratados se considerado apenas os valores provenientes da detecção e classificação, desconsiderando informações semânticas com base em dados de profundidade (LIDAR + Câmera 3*D*) e, também, da relação entre os objetos presentes em ambientes de trânsito. A posição de cada elemento na cena é de extrema importância para que o modelo de Atenção Visual desenvolvido consiga realizar suas funções, assim, como também são importantes as informações de detecção e classificação, provenientes da percepção com fusão de dados 2*D* e 3*D*. Outro ponto importante é que muitas sinalizações podem sofrer alterações, devido a obras e modificações, temporárias ou permanentes, desta sinalização. Isto faz com que a navegação baseada apenas em mapas precise de um suporte de um sistema que possa considerar situações como as apresentadas aqui.

Outra característica muito importante do modelo de Atenção Visual desenvolvido, está ligada com a análise dos sinais de trânsito detectados em função da área navegável, possibilitando então relacionar as regras de trânsito de acordo com a sua rota pertencente. Sem esse tipo de análise seria bastante complicado tratar situações não mapeadas *a priori* e que são abordadas nesta tese.

Os testes para Atenção Visual foram aplicados em 36 diferentes cenas envolvendo conflitos de sinais de trânsito verticais e, também, de bifurcações, solucionando problemas parecidos com os que foram apresentados nas Figuras 86, 87, 88 e 89. No entanto, para novos casos que não foram modelados na base de regras *Fuzzy*, o modelo de Atenção Visual poderá não funcionar bem, sendo então necessários alguns ajustes para inserir os novos problemas e novas regras em questão na base de conhecimento do modelo.

6.7 Considerações

Neste capítulo, foram apresentados os experimentos, resultados e, também, as discussões relacionados com o modelo de Visão Computacional Robótico desenvolvido nesta tese e, que envolve, um sistema de percepção de sinais de trânsito verticais (ver Capítulo 4) em conjunto com um modelo de Atenção Visual *Fuzzy* (ver Capítulo 5).

O sistema de detecção de sinais de trânsito verticais em imagens 2*D* se comportou bem quando em conjunto com o filtro 3*D*, possibilitando desta forma eliminar informações falsas detectadas e que geram grandes problemas para uma navegação autônoma ou semicontrolada. Sendo que parte da contribuição científica desta tese está direcionada para a fusão de dados 2*D* e 3*D*, visando um sistema de percepção mais robusto contra falhas.

Inicialmente foi trabalhada a detecção puramente em dados 3*D*, no entanto, a nuvem de pontos da câmera estéreo e, também, do LIDAR, não oferece uma boa estrutura da aglutinação de pontos em distâncias maiores que 25 metros, gerando uma grande quantidade de *sprays* e imperfeições em sua forma. Devido a este problema, foi aplicada a detecção em imagens 2*D*, garantindo distâncias de até 100 metros para o modelo de percepção e, aplicando a filtragem 3*D* em sequência, para quando o sinal de trânsito está mais próximo do veículo, em torno de 25 metros. Uma evolução junto aos sensores 3*D* deve contribuir muito com futuros trabalhos relacionados ao estado-da-arte desta tese.

Os dados disponibilizados pelo *Dataset* do KITTI (GEIGER; LENZ; URTASUN, 2012) e, também, pelo *Dataset* do INI *German Traffic Sign Benchmarks* (STALLKAMP *et al.*, 2011), possibilitaram que o modelo de percepção e de Atenção Visual propostos nesta tese, fossem realizados com fusão de dados 2*D* e 3*D* de maneira efetiva, exigindo apenas algumas modificações para os testes de filtragem 3*D* e, também, do modelo de Atenção Visual.

O modelo de extração de características 3*D-CSD* se comportou bem em conjunto com um algoritmo neural, possibilitando uma boa filtragem de sinais de trânsito por meio da classificação de objetos 3*D*. No entanto, o custo computacional e de tempo deste método não possibilitou o seu funcionamento em tempo real. Contudo, o processamento em tempo real não foi o foco desta tese.

O modelo de Atenção Visual desenvolvido foi capaz de analisar a prioridade dos sinais de trânsito detectados em conflito. A análise é feita por meio dos dados provenientes do sistema de percepção: (a) Detecção e classificação de sinais de trânsito, (b) Detecção da área navegável e (c) Detecção de bifurcações. O modelo se comportou bem para situações de emergência na via, garantindo a classificação da prioridade de cada sinal de trânsito detectado. Destacando que para o bom funcionamento deste modelo, são necessários os dados provenientes do sistema de percepção com fusão de imagens 2*D* e 3*D* e da modelagem de cada problema.

O principal objetivo do modelo de Atenção Visual é dar suporte para o sistema de tomada de decisão do veículo, possibilitando gerar soluções de controle em função das regras de trânsito detectadas e analisadas. A base de regras para o modelo desenvolvido, pode ser alterada para o funcionamento adequado em outros países com outras regras de trânsito, outros tipos de sinais verticais e outros tipos de situações de emergência. Bastando que uma nova modelagem dos problemas encontrados seja feita e cadastrada na base de conhecimento *Fuzzy*.

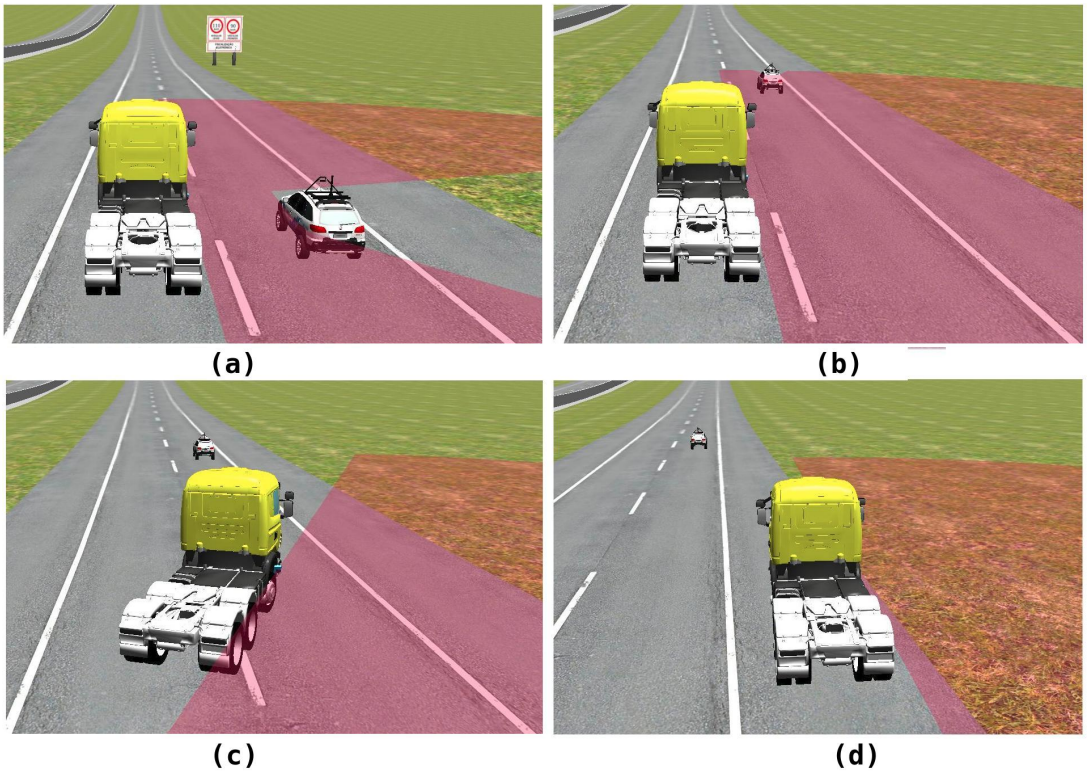
O modelo de Visão Computacional Robótica para detecção e análise de sinais de trânsito verticais e, que é formado por um sistema de percepção com fusão de dados 2*D* e 3*D* e, também, por um modelo de Atenção Visual, é altamente dependente de todas etapas em seu funcionamento, desde a captura das imagens por meio dos sensores, até o ponto em que é feita a análise das prioridades dos sinais detectados. Sendo assim, é muito importante para a evolução de trabalhos futuros, uma melhoria na qualidade da captura de imagens 3*D* (câmeras, LIDARs e outros), possibilitando uma melhor manipulação dos dados provenientes destes sensores, gerando

uma melhor capacidade de detecção e classificação. Consequentemente, gerando um melhor funcionamento do sistema de visão em sua forma completa.

Um ADAS foi desenvolvido para os testes de rotinas automáticas voltadas para o suporte à tomada de decisão. O sistema é baseado nos dados de percepção e Atenção Visual para detecção e reconhecimento de sinais de trânsito verticais. Este tipo de aplicação pode auxiliar o veículo robótico em situações de emergência envolvendo problemas com o desrespeito das leis de trânsito, consequentemente, possibilitando corrigir situações de perigo para a segurança viária (Bruno *et al.*, 2018b) ; (Bruno *et al.*, 2019).

Por meio da Figura 90, pode ser observada uma situação onde o veículo está sendo controlado de maneira insegura, portanto, uma rotina automática é ativada para estacioná-lo no acostamento, evitando um possível acidente.

Figura 90 – ADAS para suporte a navegação: (a) Obstáculo que impede a manobra, (b) livre de obstáculos, (c) início da manobra e (d) manobra efetuada.



Fonte: Elaborada pelo autor.

Este sistema foi desenvolvido em ambiente de simulação por ser de grande complexidade de testes em ambientes reais, no entanto, as implementações funcionam nos veículos do LRM/ICMC.

CAPÍTULO

7

CONCLUSÃO: CONSIDERAÇÕES FINAIS

Neste capítulo, serão apresentadas as conclusões finais, contribuições, resposta da pergunta científica e limitações e trabalhos futuros relacionados a esta tese.

Por meio dos estudos e pesquisas realizados durante este trabalho de doutorado, foi possível realizar o desenvolvimento de um sistema de Visão Computacional Robótico mais robusto e preciso para detecção de sinais de trânsito utilizando a fusão de dados $2D$ e $3D$. Também foi possível desenvolver um modelo de Atenção Visual capaz de classificar a prioridade dos sinais de trânsito detectados e, que geralmente, geram conflitos de informações para a navegação autônoma do veículo em situações não mapeadas *a priori*.

Nesta pesquisa foi possível contribuir junto ao estado-da-arte de modelos de percepção com fusão de imagens $2D$ e $3D$, possibilitando filtrar falsos positivos e falsos negativos e, também, gerando suporte para o modelo de Atenção Visual, que é capaz de avaliar a relevância de cada sinal de trânsito detectado. O modelo de Atenção Visual é inovador, capaz de interpretar a cena de maneira semântica e dar suporte para a tomada de decisão em uma navegação dentro das regras de trânsito locais, sendo que atualmente, não existem trabalhos relacionados que realizam as mesmas funções que foram apresentadas nesta tese (Capítulos 4 e 5).

Além de realizar a análise semântica da cena e avaliar cada sinal de trânsito detectado e relacioná-lo com seu valor de prioridade, o modelo de Atenção Visual desenvolvido é capaz de avaliar para qual rota o sinal em questão pertence. Possibilitando que um veículo que não utiliza um mapa *a priori*, consiga relacionar o sinal de trânsito detectado com a rota atual. Este problema é bastante comum em situações de bifurcações e trechos em obras.

7.1 Resposta da Pergunta Científica

O modelo de Visão Computacional baseado em Atenção Visual *Fuzzy* para análise semântica de prioridades de regras de trânsito conflitantes e, que utiliza técnicas de percepção com fusão de imagens $2D$ e $3D$ para detecção e classificação de sinais de trânsito verticais de maneira mais robusta, reduzindo a quantidade de falsos positivos e falsos negativos, corrobora respondendo positivamente a questão científica formulada inicialmente nesta tese de doutorado. Avançando junto ao estado-da-arte na área de Visão Computacional Robótica.

7.2 Declaração de autoria

Eu, MSc. Diego Renan Bruno, sob orientação do professor Dr. Fernando Santos Osório, confirmo que esta tese de doutorado não foi apresentada em apoio a uma solicitação de outro diploma nesta ou em qualquer outra instituição de ensino ou pesquisa. A pesquisa apresentada é o resultado do meu próprio trabalho e o uso de todo o material de outras fontes foi reconhecido de maneira adequada e plena. Pesquisas feitas em colaboração também são claramente indicadas. Trechos desta tese foram publicados ou submetidos à apreciação de conselhos editoriais de revistas, conferências e workshops, de acordo com a lista de publicações apresentadas na seção 7.4. Minhas contribuições para cada publicação estão todas relacionadas com os sistemas de percepção e Atenção Visual para veículos robóticos inteligentes.

7.3 Limitações e Trabalhos Futuros

As limitações deste trabalho estão direcionadas junto ao estado-da-arte de sensores de percepção e, que por mais que atualmente estejam evoluídos, ainda existem grandes problemas relacionados com a densidade e aglutinação de pontos em imagens $3D$, gerando deformações no objeto avaliado.

Outro problema que não foi tratado nesta pesquisa está relacionado ao tempo real. Os modelos desenvolvidos de percepção e Atenção Visual não foram otimizados para um funcionamento em baixo custo computacional e de tempo.

Outros trabalhos futuros estão relacionados com (a) Aplicações em veículos reais do LRM/ICMC ; (b) Considerar o tempo (sequência no tempo), para uma melhor tomada de decisão e (c) Integrar GPS + mapas $3D$ + o modelo de percepção e Atenção Visual desenvolvido nesta tese.

7.4 Publicações

Nesta seção serão apresentados os trabalhos publicados em eventos nacionais e internacionais e que estão relacionados com esta tese de doutorado. Também serão apresentados os trabalhos desenvolvidos como professor na Universidade de São Paulo (USP) e Faculdade de Tecnologia de Catanduva (FATEC).

7.4.1 Eventos

Os eventos em que os artigos foram apresentados estão ligados principalmente as áreas de: Robótica, Visão Computacional, Inteligência Artificial e Veículos Inteligentes.

- 2017 ***Latin American Robotics Symposium (LARS)*** *and 2017 Brazilian Symposium on Robotics (SBR) - Curitiba, Brazil*;
- 2018 ***International Joint Conference on Neural Networks (IJCNN)*** *IEEE World Congress on Computational Intelligence (IEEE WCCI) - Rio de Janeio, Brasil*;
- 2018 ***Latin American Robotics Symposium (LARS)*** *and 2018 Brazilian Symposium on Robotics (SBR) - João Pessoa, Brasil*;
- 2019 ***52nd Hawaii International Conference on System Sciences (HICSS)*** *- Maui, Hawaii - EUA*;
- 2019 ***IEEE Intelligent Vehicles Symposium (IV'19)*** *- Paris, França.*
- 2019 ***XV Workshop on Computational Vision (WVC'19)*** *- São Bernardo do Campo, Brasil*;
- 2019 ***IEEE RAS International Summer School on Deep Learning for Robot Vision*** *- Santiago, Chile*.

Os artigos apresentados, em sua maioria, estão relacionados com a tese de doutorado apresentada. Outros trabalhos estão relacionados com as pesquisas envolvendo os Veículos Robóticos do Laboratório de Robótica Móvel (LRM). Também existem trabalhos que estão relacionados com orientações realizadas na Faculdade de Tecnologia de Catanduva na área de Robótica Móvel e Visão Computacional.

7.4.2 Artigos

- BRUNO, DIEGO RENAN; OSORIO, FERNANDO SANTOS . **Image classification system based on deep learning applied to the recognition of traffic signs for intelligent robotic vehicle navigation purposes.** In: 2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR), 2017, Curitiba. 2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR), 2017. p. 1.

- BRUNO, DIEGO RENAN; SALES, DANIEL OLIVA ; AMARO, JEAN ; OSORIO, FERNANDO SANTOS . **Analysis and fusion of 2D and 3D images applied for detection and recognition of traffic signs using a new method of features extraction in conjunction with Deep Learning.** In: 2018 International Joint Conference on Neural Networks (IJCNN), 2018.

- BRUNO, DIEGO RENAN; Marranghello, Norian ; OSORIO, FERNANDO SANTOS ; PEREIRA, ALEDIR SILVEIRA . **Neurogenetic algorithm applied to Route Planning for Autonomous Mobile Robots.** In: 2018 International Joint Conference on Neural Networks (IJCNN), 2018, Rio de Janeiro. 2018 International Joint Conference on Neural Networks (IJCNN), 2018. p. 1.

- BRUNO, DIEGO RENAN; SANTOS OSORIO, FERNANDO . **A Comparison of Traffic Signs Detection Methods in 2D and 3D Images for the Benefit of the Navigation of Autonomous Vehicles.** In: 2018 Latin American Robotic Symposium (LARS), 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), 2018, João Pessoa.

- BRUNO, DIEGO RENAN; SANTOS, TIAGO C. ; SILVA, JUNIOR A.R. ; WOLF, DENIS F. ; OSORIO, FERNANDO S. . **Advanced Driver Assistance System Based on Automated Routines for the Benefit of Human Faults Correction in Robotics Vehicles.** In: 2018 Latin American Robotic Symposium (LARS), 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), 2018, João Pessoa. **(Best Paper)**.

- PACHECO GOMES, IAGO ; RENAN BRUNO, DIEGO ; SANTOS OSORIO, FERNANDO ; FERNANDO WOLF, DENIS . **Diagnostic Analysis for an Autonomous Truck Using Multiple Attribute Decision Making.** In: 2018 Latin American Robotic Symposium (LARS), 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), 2018, João Pessoa.

- BRUNO, D. R.; MATIAS, L. P. N. ; AMARO, J ; WOLF, DENIS F. ; OSÓRIO, Fernando S . **Computer Vision System with 2D and 3D Data Fusion for Detection of Possible Auxiliaries Routes in Stretches of Interdicted Roads.** In: Hawaii International Conference on System Sciences (HICSS) - 52nd, 2019, Maui, Hawaii. Smart (City) Application Development: Challenges and Experiences. Honolulu, Hawaii: Scholar Space, 2019. v. V1. p. 7372-7381.

- SANTOS, TIAGO C. ; BRUNO, D. R. ; OSÓRIO, Fernando S ; WOLF, DENIS F. . **Evaluation of lane-merging approaches for connected vehicles.** In: 2019 IEEE Intelligent Vehicles Symposium (IV-19) - IEEE Intelligent Transportation Systems Society (ITSS), 2019, Paris. IEEE Intelligent Transportation Systems Society (ITSS). EUA: IEEE Xplore, 2019. v. 1.

- BRUNO, D. R.; Assis, Marcelo ; OSÓRIO, Fernando S . **Development of a robotic sensing system to assist visually impaired people in their task of locomotion in urban environments.** In: Workshop de Visão Computacional (WVC-2019), 2019, São Bernardo do Campo. Workshop de Visão Computacional (WVC-2019) - FEI, 2019. v. XV.

- BRUNO, D. R.; PACHECO GOMES, IAGO ; OSÓRIO, Fernando S ; WOLF, DENIS F. **Advanced Driver Assistance System based on NeuroFSM applied in the detection of autonomous human faults and support to semi-autonomous control for robotic vehicles.** In: Latin American Robotics Symposium - (LARS 2019), 2019, Rio Grande do Sul. Latin American Robotics Symposium - (LARS 2019), 2019.

- BRUNO, D. R.; Assis, Marcelo ; OSÓRIO, Fernando S . **Development of a Mobile Robot: Robotic guide dog for Aid of Visual Disabilities in urban environments.** In: Latin American Robotics Symposium - (LARS 2019), 2019, Rio Grande do Sul. Latin American Robotics Symposium - (LARS 2019), 2019.

- BRUNO, D. R.; OSÓRIO, Fernando S . **Visual attention system based on Fuzzy Classifier to define priority of traffic signs for intelligent robotic vehicle navigation purposes.** In: International Conference on Advanced Robotics - (ICAR), 2019, Belo Horizonte. International Conference on Advanced Robotics, 2019 (**Selected as one of the Best Papers of the Conference**).

- BRUNO, D. R.; OSÓRIO, Fernando S. **Robotic Computer Vision System with 2D and 3D Data Fusion for TrafficSign Detection.** - IEEE RAS International Summer School on Deep Learning for Robot Vision. Santiago, Chile (2020).

A seguir, os resumos de cada trabalho serão apresentados em sua evolução até o presente momento, garantindo um melhor conhecimento sobre o desenvolvimento desta pesquisa de doutorado.

Image classification system based on Deep Learning applied to the recognition of traffic signs for intelligent robotic vehicle navigation purposes

Diego Renan Bruno., and Fernando Santos Osório., Member, IEEE - *University of São Paulo*

Abstract - **This paper presents a system for classifying images based on Deep Learning and applied in the recognition of traffic signals aiming to increase road safety increased road safety using autonomous and semi-autonomous intelligent robotic vehicles. This Advanced Driver Assistance System (ADAS) is a system created to automate vehicles, but also to help the human drivers to increase safety and the respect of traffic rules while driving the car. The system must be able to classify several different traffic signs (e.g. maximum speed allowed, stop, slow down, turn ahead, pedestrian), thus helping to make navigation within the local traffic rules. The obtained results are promising and very satisfactory, where we get 97.24% of test accuracy in a well known traffic sign benchmark dataset (INI - German Traffic Sign Benchmark).**

Keywords - Deep Learning; Tensorflow; Traffic Sign; Robotics Vehicles; ADAS.

I. INTRODUCTION

The structured road scenarios and urban street environments needs general and proper traffic rules, including varied information for drivers, so that a vehicle can safely navigate on it, whether it is in an autonomous or semi-autonomous mode, or even being conducted by a human. The most used signs are: traffic signs (speed control, stop and direction), traffic lights (danger, attention, proceed, traffic semaphore) and traffic lanes (regulation of flows and space). Traffic signs are used to assist the driver in the task of driving, however, also informing about the local traffic rules [1].

These signs are of great importance not only for the driver of a semi-autonomous/intelligent vehicle but also for a vehicle that travels autonomously and shares the same rules with other cars and humans. In this work, a Deep Learning based image classification method was applied for the recognition of traffic signs in benefit of road safety applied to intelligent robotic vehicles.

The research area of detection and classification of traffic signs has grown a lot in the last decade, mainly with the works that use machine learning techniques. This increase is closely linked to the evolution of robotized vehicles, where Advanced Driver Assistance Systems (ADAS) are applied to assist in the task of driving safely (conducted by a human driver) and also for autonomous vehicles. Safe navigation in traffic environments also depends on this information provided by means of traffic sign plates, traffic lights and traffic lanes. In this work it was proposed the use of Deep Learning, adopting the TensorFlow tool [2], to classify the traffic signals.

II. RELATED WORKS

Traffic sign recognition have been studied for a long time, but the in the 80's and 90's, most of the works suffered from problems with sign detection, illumination, noise and partial occlusion, and most of them obtained relatively poor performances of 70% to 80% of accuracy considering small datasets. The improvement in the development of machine learning techniques and the available datasets allowed to obtain better results more recently.

In a paper by Yaong Yu et. Al. [1], a system was developed based on an algorithm capable of classifying traffic signals for the benefit of navigation of robotic vehicles. The system uses a mobile laser scanning (MLS) sensor.

In a work done by Timofte and Zimmermann [3], a system was developed that aims at the detection of signaling signs in favor of the recognition of traffic rules. In this work, a method of plate recognition based on 3D image analysis was developed using a technique called Minimum Description Length principle (MDL). This was one of the first works using 3D images in favor of the analysis of traffic signs [3].

In a paper by Zhou and Deng [4], a system based on LIDAR (Light Detection And Ranging) and recognition algorithms for the analysis of images of signaling plates was developed. The LIDAR sensing technology is similar to ultrasonic sensing, however, using a light beam as a signal. The system aims to analyze 3D images in order to achieve greater robustness in the detection of plates. Through 3D point-cloud data (colors and point agglutination) the signaling of the board is analyzed. The algorithm that identifies the characteristics of each detected board is based on a Support Vector Machine (MVS), however, being applied as a classifier [4].

In a work developed by Soilán [5], each traffic signal is automatically recognized using the Light Detection And Ranging (LIDAR) technique in conjunction with algorithms

2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018.
IEEE 2018, ISBN 978-1-5090-6014-6

Analysis and fusion of 2D and 3D images applied for detection and recognition of traffic signs using a new method of features extraction in conjunction with Deep Learning

Diego Renan Bruno, Daniel Oliva Sales, Jean Amaro and Fernando Santos Osório., Member, IEEE
LRM Laboratory - ICMC - *University of São Paulo / USP*

***Abstract* - This paper presents a system for detection and recognition of traffic signs using 2D images and 3D scene data. The detection and recognition of the 3D structures (pole and signs) and the classification of the traffic signs is based on a new 3D features extraction method and the use of a Deep Learning method. The traffic sign recognition aims to increase road safety considering autonomous and semi-autonomous intelligent robotic vehicles in structured environments with traffic rules (urban streets or roads/highways). The proposed system can be used as an Advanced Driver Assistance System (ADAS) to help human drivers to increase safety and to respect the traffic rules while driving the car. It can also be adopted in fully autonomous vehicles, in the task of detecting traffic signs, making it possible to adapt the vehicle navigation control according to the local traffic rules. The system must be able to detect traffic signs, using 3D data in point clouds, and classify several different traffic signs, using 2D data considering colors, textures and shapes information (e.g. maximum speed allowed, stop, slow down, turn ahead, pedestrian crossing). The results are promising and very satisfactory, we obtained an accuracy of 97.64% in the 2D classification task and 76% accuracy in the single frame 3D detection task. These results were obtained using for testing a well-known dataset of street scenes with traffic signs, The KITTI Vision Benchmark Suite, and also another traffic sign benchmark dataset, the INI - German Traffic Sign Benchmark.**

Keywords - Traffic Sign Recognition; Deep Learning; Robotics Vehicles; Intelligent Vehicles, ADAS.

I. INTRODUCTION

The traffic scenarios, in highways and/or urban streets should display the local rules represented by traffic signs. This enables a person involved in the driving task to be able to travel safely and in an organized way, together with other drivers. Thus, making it possible to avoid accidents related to speed, obligatory stop, and also, due to other various possible hazards, including presenting warnings to the driver about the type of environment/road that is expected to be found in his/her path[1].

Traffic signs recognition are of great importance in two main situations: (1) for intelligent and semi-autonomous vehicles, where it still exists at a human driver assisted by an ADAS system; (2) for a fully autonomous vehicle, sharing the same transit environment and the same rules of traffic with humans. Both situations require systems that are able to automatically “detect and read” traffic signs.

In this paper we present a new system for detecting traffic signs using an innovative features extraction method. We proposed and adopted a new 3D descriptor, invariant to translation, rotation, and scale, named 3D-CSD (3D-Contour Sample Distances) [5]. Then, we applied it to represent and detect the surfaces of the scene objects, including the "traffic sign" present in the navigation environment. Once the traffic sign is detected, segmented and identified as a real traffic sign object, in order to determine which traffic sign it is, we use a supervised learning method – Artificial Neural Network (ANN) - for pattern recognition applied to the 2D data obtained from the 3D point cloud and stereo vision object data (RGB + Depth data).

In order to classify the specific type of each traffic sign (e.g. maximum speed, stop, direction required, among other signs), we use 2D data, thus taking advantage of the potential of the textures and visual clues found in the surfaces of the traffic sign plates. For a good classification to be possible, we used an ANN network based on Deep Learning, applied to 2D input data (images).In the developed system, the images are passed to a Deep Learning model, adopting a DCNN Net (Deep Convolutional Neural Network / ConvNet) implemented using TensorFlow[2],to classify the signs.

The open source software library for machine learning TensorFlow [2] is widely used today for image recognition and Deep Learning applications. The training used in this work was applied in the final layer of a CNN (Transfer Learning technique), based on the Inception's Network [2] [3] [4].

The remainder of the text of this paper is organized as follows: In section II the context and related works are presented. In section III we discuss about algorithms for traffic sign detection, and in section IV a theoretical background is presented and the TensorFlow/Deep Learning methods are introduced. Section V presents the experiments and results of tests performed in the recognition system. Finally, in section VI the conclusions and future works are presented.

2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018.
IEEE 2018, ISBN 978-1-5090-6014-6

Neurogenetic algorithm applied to Route Planning for Autonomous Mobile Robots

Diego Renan Bruno*, Norian Marranghello**, Fernando Santos Osório* and Aledir Silveira Pereira**

University of Sao Paulo (USP)*
Sao Paulo State University (UNESP)**

Abstract - We developed a bioinspired algorithm to assist in the navigation of an autonomous mobile robot in dynamic environments. The robotic controller uses both an Artificial Neural Network (ANN) and a Genetic Algorithm (GA), aided by a low computational cost vision system. The controller uses the vision system and the ANN to detect and recognize obstacles found in the robot's path. If the object is in the controller's knowledge bank a previously registered deviation solution is applied. Otherwise, the GA must optimize a new route alternative. We modeled and simulated the controller using robot's simulator V-REP, and the Computer Vision System using Scilab software. The contribution of this paper is the development of a hybrid neuro-genetic algorithm to control the navigation of autonomous mobile robots in dynamic environments.

Keywords- Mobile robot, Neural network, Genetic algorithm, Global path planning

I. INTRODUCTION

Much of the research in mobile robotics is intended for autonomous navigation, where route optimization techniques are constantly developed to control a robot in a dynamic environment. The controller must be able to optimize a collision-free route presenting the smallest path between the target points, as the smaller the path, the lower the energy, time and mechanical cost.

We developed a mobile robot neurogenetic controller for use in Automated Material Transportation Systems (AMTS), preventing damages to workers' health in potentially risky environments.

The bioinspired Hybrid controller we developed consists of a Genetic Algorithm (GA), and a Perceptron Artificial Neural Network (ANN). The ANN assesses whether the obstacle found is recognized or not. When recognized, a knowledge bank automatically provides a deviation solution to avoid the obstacle. When not recognized, the GA is applied to produce new optimal routes. Some related works using GA and ANN for mobile robots navigation are presented in the following paragraphs.

Santhosh developed a GA applied for the optimization of routes of autonomous mobile robots applied in war environments to find victims, avoiding the risk of a human being work in such an environment to save lives [1].

Panda and Choudhury developed a "pure" GA for mobile robot navigation control, for which they developed a method of representation of chromosomes, using a binary matrix to define the structure of each individual of the population. They also used sequential recombination, in which the first two individuals from a ranked list are selected and recombined [2].

Abinaya and Kumar developed a mobile robot hybrid controller, which uses a GA to define the deviation points and a GJL (Gilber-Johnson-Keerth) algorithm to calculate the distances of each solution. Based on these distances the best possible routes can be optimized [3].

Abbaspour and Alpour used a Particle Swarm Optimization algorithm (PSO) and GA to control a set of three mobile robots in order to work together, to load large and heavy objects. The controller uses GA and PSO to optimize the relative positioning among the three robots [4].

Gomez-Ortega et al used a GA in conjunction with a Kalman filter for routes planning. The GA is applied to optimize auxiliary routes, and the Kalman filter is applied to predict the future positions of the mobile obstacles [14].

Luo and Yang developed an ANN to declare, based on a knowledge bank, whether the obstacle is too close, or too far, or at middle ground. Thus avoiding collisions and generating deviation routes without wasting time and energy [7].

Heinen and Osorio developed the work most similar to ours, in which the ANN effectively controls the movements of the robot, while the GA is used for ANN training, assisting in the evolutionary adjustment of the synaptic weights. As the GA does not need local information to correct network errors, its use does not requires a real-time database [8].

All the papers presented in this section have used ANNs and GAs for the navigation of mobile robots, even in a hybrid way, but none of them have used the same methods and techniques for the same functions as proposed in this work. Heinen and Osorio [8], e.g., also used a neurogenetic algorithm, however, the functions, the mapping, ANN training and route

2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)

Diagnostic Analysis for an Autonomous Truck Using Multiple Attribute Decision Making

Iago Pachêco Gomes, Diego Renan Bruno, Fernando Santos Osório and Denis Fernando Wolf
Institute of Mathematical and Computer Sciences
University of São Paulo, São Carlos, São Paulo
Email: {iagogomes, diego.bruno, fosorio}@usp.br, denis@icmc.usp.br

***Abstract*—Autonomous vehicles are robots capable of navigate in urban or confined spaces with no human intervention. For that, the vehicle is equipped with specific hardware and software components. To accomplish the task of navigate in urban scenarios, some aspects must be analyzed, including the system integrity. In addition, the robot must make decisions based on sensor information. This paper proposes the use of Multiple Attributes Decision Making and Fuzzy sets, for an automatic analyzer of diagnostic information of an Autonomous Truck, and its perception data. The decision-maker task was to choose a safe maneuver to avoid or overcome a dangerous situation. Thus, Analytic Hierarchical Process (AHP) was used to compute the weights of the attributes, Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) to rank the alternatives, and Fuzzy to model the decision matrix. In the end, the system was tested in a real scenario where different events were created to validate the performance. The results show consistency in choices, which indicates the suitability of the approach for the task.**

***Keywords*-diagnotic; multiple attribute decision making; safety; autonomous truck;**

I. Introduction

Autonomous vehicles (AVs) are robotic and intelligent vehicles capable of traveling without the need of human intervention. They use its sensors to create a model of the environment, perform route and path planning to move from one place to another, make decisions about events, respect traffic rules, avoid collisions with obstacles, and do many other tasks to deal with all traffic conditions.

The lower the need of human intervention, more will be the autonomy, which can be classified according to NHTSA (National Highway Traffic Safety Administration, USA) as: *Level 0* - No-Automation; *Level 1* - Function-specific Automation; *Level 2* - Combined Function Automation; *Level 3* - Limited Self-Driving Automation; *Level 4* - Full Self-Driving Automation.

At the *Level 0*, the driver is responsible for all decision and control (brake, steering, and throttle) that have to be made. The majorities of the vehicles can be categorized in this level. For the *Level 1*, the vehicle can control a specific task, and in the *Level 2* more than one is combined, such as adaptive cruise control with lane keeping. In *Level 3* the vehicle is able to take all control of the driver under certain conditions, but when some situation occurs and the system is not able to deal with, the control is given back to the driver. At the last level all driver's tasks needed in any traffic conditions are carried out [3].

The higher the autonomy level the critical is the decision that must be done by the vehicle system. A situation when the decision is needed is, for example, which maneuver is better to perform in a specific traffic situation. Another kind of decision task is to evaluate the condition of the system and if it is safe or not continue traveling, that is, the decision maker could analyze all diagnostic information produced by the vehicle system and decide if everything is right or if some recovery protocol has to be initiated.

Diagnostic analysis is a critical piece of the software architecture for all autonomous vehicle [4]. In the level 3 of autonomy, if something wrong occurs, the recovery protocol gives the control back to the driver, but at the level 4, the system has all responsibility to deal with any kind of situation, and it includes system fault or instability, such as, poor data from essential sensors (GPS, Obstacle detector, etc.), higher lateral or angular error from local path planning, and inability to communicate with any software or hardware component.

This paper proposes an automatic analysis of diagnostic information produced by an Autonomous Truck software system, using Multiple Attribute Decision Making (MADM) method and Fuzzy sets, which model the subjectivity inherent in the data. The main contribution is combine diagnostic data with perception (obstacle detection) to evaluated a set of safe maneuvers alternatives to prevent or overcome dangers situation caused by some system fault or unreliability of essential sensors, such as GPS.

The rest of this paper is organized as follow: Section II presents some related works for maneuver choice and diagnostic analysis; Section III the architecture of the Autonomous Truck was described; Section IV explains the Multiple Attribute Decision Making (MADM) method; Section V discusses about the results of the proposed approach; and, Section VI concludes the paper.

II. Related Works

Since DARPA Urban Grand Challenge, 2007, the research about decision-making applied to an autonomous vehicle

978-1-5386-7761-2/18/$31.00 ©2018 IEEE
DOI 10.1109/LARS/SBR/WRE.2018.00059

2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)

Advanced Driver Assistance System based on automated routines for the benefit of human faults correction in robotics vehicles

Diego R. Bruno, Tiago C. Santos, Júnior A. R. Silva, Denis F. Wolf and Fernando S. Osório
Institute of Mathematics and Computer Science
University of São Paulo
São Carlos, Brazil
Emails: diego.bruno@usp.br, tiagocs@icmc.usp.br, junior.anderson@usp.br, fosorio@icmc.usp.br, denis@icmc.usp.br

***Abstract*—In this paper it is developed an Advanced Driver Assistance System with automatic routines to correct human faults and, also, for autonomous vehicles controllers. The system is able to generate an alternative solution for different problems encountered in the task of controlling a stand-alone vehicle in environments with traffic rules. Based on the premise of the detected problem, the automatic routines are activated (in the worst cases taking full control of the vehicle) to generate an alternative solution and avoid a possible accident. A precision of 90% of situations detected and solved the driver support routines is obtained, which can be considered a good efficiency. To validation the proposed system, traffic environment is modeled in MORSE (Modular Open Robots Simulation Engine) and ROS (Robot Operating System), allowing a considerable simulation quality. The proposed automated vehicle driving assistance showed good ability to correct human failures and also to take full control of the vehicle, where the vehicle should be stopped and parked at the roadside.**

***Keywords*-robotic vehicles; MORSE; ROS; ADAS.**

I. Introduction

The current traffic environment involves strict laws that are not respected by drivers in most cases. This type of situation generates a large number of fatal traffic accidents around the world. Despite all laws that are enforced rigidly around the world, the police forces surveillance is unable to address all forms of illegal driving, consequently only 1% of drivers are detected through the police [1] or otherwise acting irregular.

Deaths from drunk drivers on US roads in 2008 amount to 30% of total fatalities over the same period [22]. Brazilian traffic surveillance also takes into account that drowsiness is an indicator of intoxication [2]. In Brazil, a study of 143 cities in Brazil in 2009 indicated that drinking and driving is normal and acceptable for 35% of the population [3].

To address this problem, Advanced Driver Support Systems can be applied to manage and analyze the respect of the traffic rules the traffic rules detected by a vision system in conjunction with the mode the vehicle is being controlled (stand-alone or semi-autonomous). The most used signs are: traffic signs (speed control, stop and direction), traffic lights (danger, attention, follow-up, traffic lights) and traffic lanes (flow and space regulation). Traffic signs are used to assist the driver in the task of driving, however, also informing about the local traffic rules [4].

Traffic signs are of great importance not only for semi-autonomous/intelligent vehicles drivers, but also for autonomous vehicles, which follow the same rules as human drivers. However, traffic laws are not always respected and policing is insufficient to punish all drivers who break the law. As a result, only 1% of these drivers are approached by police [1]. Thereby, Advanced Driver Assistance Systems (ADAS) should contribute to increase road safety.

The ADAS research field has significantly increased in the last decade, mainly due to machine learning techniques. This increase is closely related to the evolution of robotized vehicles, which make use of ADAS to improve safety.

In this work the detection of a set of human faults is assumed, such as not obeying traffic laws, drowsiness and drunkenness. To overcome the fault, a routine is generated once the fault is known. In simple cases, only an alert or speed adjustment is able to correct a fault. In cases where the driver is completely out of control, an automatic routine that parks the vehicle on the road is executed, depending on the kind of each problem encountered. These problems are divided into three classes: (a) human failure or recklessness (b) failure of visual attention - no detection of traffic signals and (c) driver unable to perform the task of driving.

II. Related works

Most ADAS related works focus on driver analysis, but few of them shows methods to support the driver in an automated manner. In this section, works with similar subjects to those presented in this paper are surveyed.

A well-known ADAS feature is lateral control, which seeks to assist the driver in carriageway exits when he is acting carelessly or trying to change lane in a dangerous manner. Two mechanisms that can be mentioned are: track exit warning [5] and blind spot monitoring [6].

It is notable that most of the frontal sensing is done by an image sensor (camera) thus allowing to detect: an obstacle or road sign. Remembering that the sensor can be the same

978-1-5386-7761-2/18/$31.00 ©2018 IEEE
DOI 10.1109/LARS/SBR/WRE.2018.00029

2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)

A comparison of traffic signs detection methods in 2D and 3D images for the benefit of the navigation of autonomous vehicles

Diego Renan Bruno
University of Sao Paulo
LRM - Mobile Robotics Laboratory
Sao Carlos, Brazil
diego.bruno@usp.br

Fernando Santos Osorio
University of Sao Paulo
LRM - Mobile Robotics Laboratory
Sao Carlos, Brazil
fosorio@icmc.usp.br

***Abstract*—This paper presents a comparison between our computer vision system with 2D and 3D data fusion with vision systems that use purely 2D data. We introduce a system of obstacles recognition inspired by human visual attention and that uses the notion of depth to eliminate false positives and false negatives. In order for the system to have a greater robustness in data analysis, we apply a new method of extraction of 3D characteristics titled as 3D-Contour Sample Distances, being invariant to scale, translation and rotation.**

The system must be able to classify several different traffic signs (e.g. maximum speed allowed, stop, slow down, turn ahead, pedestrian), thus helping to make navigation within the local traffic rules. The obtained results are promising and very satisfactory, where we get 98.3% of test accuracy in a well known traffic sign benchmark dataset (INI - German Traffic Sign Benchmark).

Our results in the detection and recognition of the traffic signs with 2D and 3D data fusion showed better results and greater robustness compared to traffic signs detection systems working only with 2D data. Our system make it possible to reduce or eliminate false positives and false negatives which are a big problem for the autonomous vehicle vision systems.

***Keywords*-Deep Learning; Tensorflow; Traffic Sign Detection; Robotics Vehicles ; Computer Vision 3D;**

I. Introduction

The traffic scenarios, in highways and/or urban streets should display the local rules represented by traffic signs. This enables a people involved in the driving task to be able to travel safely and in an organized way, together with other drivers. Thus, making it possible to avoid accidents related to speed, obligatory stop, and also, due to other various possible hazards, including presenting warnings to the driver about the type of environment/road that is expected to be found in his/her path [1].

Traffic signs recognition are of great importance in two main situations: (1) for intelligent and semi-autonomous vehicles, where it still exists at a human driver assisted by an ADAS system; (2) for a fully autonomous vehicle, sharing the same transit environment and the same rules of traffic with humans. Both situations require systems that are able to automatically detect and read traffic signs.

Figure 1. Real problem reported in 2D image object detection and recognition.

As we can see in (Fig. 1), the process of detecting and recognizing a truck on an image, can also incorrectly, but also incorrectly detects two vertical traffic signs. This problem has occurred because an adhesive is glued behind the truck. This problem can't be treated or solved using 2D images, however treated in 2D images, however, if a 3D detection system were applied, this error would be easily corrected.

Considering the notion of depth (and 3D data), it would be easy to identify that there are only one truck in this scene, and no other objects in the same position that it was found the rear of the vehicle. The two verticals traffic signs can not be flat and sticked in the rear of the truck.

This problem would not be possible to be treated in 2D images, however, if a 3D detection system were applied, this error would be easily corrected. Considering the notion of depth (and 3D format), it would be easy to identify that there is only one truck in this scene, and no other object in the same position as the rear of the vehicle was found. The two road signs can not be flat and glued to the rear of the

978-1-5386-7761-2/18/$31.00 ©2018 IEEE
DOI 10.1109/LARS/SBR/WRE.2018.00015

Proceedings of the 52nd Hawaii International Conference on System Sciences | 2019

Computer Vision System with 2D and 3D Data Fusion for Detection of Possible Auxiliaries Routes in Stretches of Interdicted Roads

Diego Renan Bruno
University of São Paulo
diego.bruno@usp.br

Lucas P. Nunes Matias
University of São Paulo
lucas.matias@usp.br

Jean Amaro
University of São Paulo
jean.amaro@usp.br

Fernando Santos Osório
University of São Paulo
f.osorio@icmc.usp.br

Denis Wolf
University of São Paulo
denis@icmc.usp.br

Abstract

In this paper we present an intelligent system to help autonomous vehicles in real cities and with local traffic rules. A 2D and 3D visual attention system is proposed, capable of detecting the use of signs and aids in cases of major roadblock (road under work, with a traffic accident, etc.). For this to be possible, we analyze the cones and traffic signs that usually alert a driver about this type of problem. The main objective is to provide support for autonomous vehicles to be able to find an auxiliary route that is not previously mapped. For this we use a Grid Point Cloud Map. Using the ORB-SLAM visual odometry system we can correctly fit each stereo frame point cloud in the pose where the images were collected. With the concatenation of point clouds generated by the stereo camera, every grid block can draw the main characteristics of its region and an auxiliary route can be mapped. In this type of situation the vision system must work in real time. The results are promising and very satisfactory, we obtained an accuracy of 98.4% in the 2D classification task and 83% accuracy in the single frame 3D detection task.

1. Introduction

Autonomous car have been intensely studied since the last decade, more recently intelligent cities have been idealized and proposed methods, trying to spread the processing and decision making between intelligent vehicles and the smart city. The communication between an intelligent city and an autonomous car is based on the information exchange about the environment where the vehicle is inserted. The vehicle pass to the city its new collected information about the surroundings while the city send to the car the previous knowledge about the environment. With this mutual data swap, the vehicle can have a previous knowledge of the road and planned trajectory, and the city can update its stored information as new data is received from the intelligent vehicle.

In an ideal world, the city and the car have the map of the roads and safe routes can be easily calculated for the vehicle navigation. Yet, in some specific cases (urgent lane repair, broken car on the road and other eventual situations) the previously mapped road have to be temporarily obstructed. In these situations, the vehicle have an outdated map that can not be used for trajectory planning. The environment changes in an obstructed road with traffic signalization and cones have to be identified. Once this environment modification is noted, a vision system have to be activated and, if possible, a traversable auxiliary route have to be mapped, then, the vehicle can maintain its trajectory to the destination and notify the smart city data center about the temporarily road condition.

In this paper we propose a method of traffic signs and cone detection, and a grid point cloud mapping based on stereo 3D vision. Using the extracted object from the stereo point cloud, then we use a Multi-Layer Perceptron to recognize it, based on its 3D features. The cone features can be extracted from the road obstruction scene. Besides the cone detection we use Deep Learning to train a Convolutional Neural Network (CNN) for the case where a lane obstruction be informed by a traffic sign. Also, we present a Stereo Vision based Grid Mapping to identify possible auxiliary routes to recreate an alternative route to the planned destination.

In the next sections we gather related works, then we present our algorithm of traffic sign detection using 3D data, following we have the theoretical background and proposed 3D feature extraction and classification, thereafter the Deep Learning method to classify the type of traffic sign, then we present our 3D point cloud grid mapping, lastly we show our experimental results, conclusions and future works.

2. Related work

A great number of mapping methods propose a grid mapping based on 3D LiDAR sensor [1, 2]. A 2D grid is created and in each grid block is stored more detailed

URI: https://hdl.handle.net/10125/60174
ISBN: 978-0-9981331-2-6

2019 IEEE Intelligent Vehicles Symposium (IV'19)

Evaluation of lane-merging approaches for connected vehicles

Tiago C. dos Santos*, Diego R. Bruno*, Fernando S. Osório*, Denis F. Wolf *
*Institute of Mathmatics and Computer Science,
University of Sao Paulo, Avenida Trabalhador Sao-carlense, 400, Sao Carlos, Brazil
{tiagocs, fosorio, denis}@icmc.usp.br, diego.bruno@usp.br

***Abstract*— Interaction among vehicles based on wireless communication technologies is an efficient way to improve traffic conditions and safety. In this context, The Grand Cooperative Driving Challenge (GCDC) is an event that aims at developing cooperative systems for intelligent vehicles. Although the two editions of the event (2011 and 2016) have provided conditions for a series of experiments such as platoon stability and lane-merging, the traffic flow in lane-merging scenarios didn't receive so much attention. This paper presents a preliminary study on traffic flow using GCDC 2016 interaction protocol for lane-merging by varying the driver imperfection model and the number of vehicles. We also propose an extension of the standard protocol. We evaluated three lane-merging interaction protocols using the Simulation of Urban Mobility (SUMO). The first protocol is very similar to the GCDC 2016, the second is the default lane change model commonly used in SUMO and the third is a variation of the GCDC interaction protocol. The experiments were conducted using similar conditions and number of vehicles of the event. We analyzed the lane-merging total time for all vehicles to complete the maneuver, the maximum string platoon length and the average platoon speed. Therefore, we could observe that the interaction protocol extension proposed decreased the duration time to complete the lane-merging maneuver of all vehicles without drastically compromising maximum string length and average platoon speed. The flexibility of this extension is closer to real traffic scenario, in which vehicles can perform the merge without having a pre-defined order.**

I. Introduction

In recent years, there have been increasing improvements in wireless communication technologies enabling the development of complex connected systems. The communication among vehicles, infrastructure, and even people plays an important role in Intelligent Transportation Systems (ITS) which are directly benefited by these technologies. This allows to develop and build cooperative transportation called Cooperative Intelligent Transportation Systems C-ITS [3].

Previous research has indicated that the traffic flow can be optimized up to 273% using communication among vehicles which would increase the road capacity by reducing the horizontal and lateral spacing between the vehicles [7]. In [10] the authors conducted a study on the impact of Cooperative Adaptive Cruise Control (CACC) in traffic flow which revealed an increase in highway capacity near a lane drop, but this study also showed that CACC can not strongly increase roadway capacity as expected and it depends heavily on the CACC-penetration rate.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

More recently, in 2011, the Netherlands held a connected vehicle competition that became known as Grand Cooperative Driving Challenge (GCDC) [11]. The competition brought together efforts from academia and companies, which made it possible to exchange experiences from different groups with different technologies, as well as to analyze the interaction of different technologies that would have to work cooperatively. The main experiment conducted during the event was to join two platoons in a lane, in which the competitors had to smooth the approximation and accommodation of the platoons using communication between vehicles and CACC to control the longitudinal speed.

In 2016, a new edition of the event took place, again in the Netherlands, and was called Interoperable GCDC AutoMation Experience (i-GAME). They established three different challenges, the first was a lane-merging situation, in which two platoons had to accommodate in a single lane due to the roadworks. The second challenge was an automated T-junction and the last was a priority request from a special vehicle demanding passage. It is worth to mention that the competitions were conducted in a real scenario and not in a simulated environment.

GCDC conducted several experiments with different technologies together in a real test scenario. However, some questions have been raised about the traffic flow. The main goal of this paper is to evaluate the lane-merging duration time by varying the driver imperfection that can be considered as an imperfection of the system. We also evaluated the maximum string length and the average platoon speed.

II. Lane-merging Interaction Protocol

The organizers and developers of the event made a special network protocol for this experiment and also an interaction protocol to organize the lane-merging.

The Figure 1 shows the initial conditions of spatial positioning of vehicles on the lanes. The LA and LB vehicles are controlled by the organizers of the event. The goal of this experimental scenario is to merge the vehicles from platoon A (except the LA) with B to form a single platoon. For example, vehicle A1 should be positioned between LB and B1, A2 between B1 and B2, and so on, until the setting shown in Figure 2 is reached.

A. GCDC

The standard interaction protocol used in GCDC 2016 is described and discussed in [2], [4] and [9], summarizing it

2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)

Development of a Mobile Robot: Robotic guide dog for Aid of Visual Disabilities in urban environments

Diego Renan Bruno[1], Marcelo Henrique de Assis[2], and Fernando Santos Osório[3]
Institute of Mathematics and Computer Science
University of São Paulo[(1,3)]
São Paulo State Faculty of Technology[(1,2)]
Emails: diego.bruno@usp.br, marcelohassis@hotmail.com, fosorio@icmc.usp.br

***Abstract*— The general objective of this work is to develop a mobile robotic platform that is able to avoid obstacles for the benefit of visually impaired people. The idea is that this platform is a robot "guide dog" for visually impaired persons, and animal guide dogs like these have very high costs (on average 40 thousand dollars - Our current robotic prototype has a cost of 400 dollars), also need a long time for training (more than 1 year), being therefore inaccessible to many visually impaired persons. Our work then turns to social robotics, where we have a big problem with disabled people and who do not get freedom to move around in urban spaces without the help of a caregiver. By collecting data from ultrasound sensors, managing a possible path to the target point in an autonomous manner, using a computer vision system to recognize and treat the obstacles encountered, representing these in the form of audio to the user. The prototype was developed to serve as a platform for research in the development of navigational algorithms and computer vision systems of the automation laboratory of the Universidade de São Paulo (USP) and Faculdade de Tecnologia de Catanduva. The robot presented good results in tests with navigation algorithms and computer vision that were applied to validate this platform in tests with visually impaired people in IDVC - (Catanduva Institute for the Visually Impaired).**
Keywords: (Social Robotics, Computer Vision, Deep Learning, Path Planing);

I. INTRODUCTION

In Brazil there are a significant number of social bodies and projects for the inclusion of people with some degree of visual impairment in society, such as the Benjamin Constant Institute (IBC), which is the national reference center for the visually impaired, the Brazilian National Blind Organization (ONCB), the Dog Guide Brazil project, among others. All with the same objective of improving the quality of life, vocational training and independence of the visually impaired. This work fits into the development of a prototype to provide this public independence in their day to day. Allowing them to be guided during their locomotion, giving them greater autonomy, also allowing communication with the user through audio and informing which obstacle was detected in the route of the robot. With this type of information being "spoken", the user can take some kind of care and attention to a certain situation.

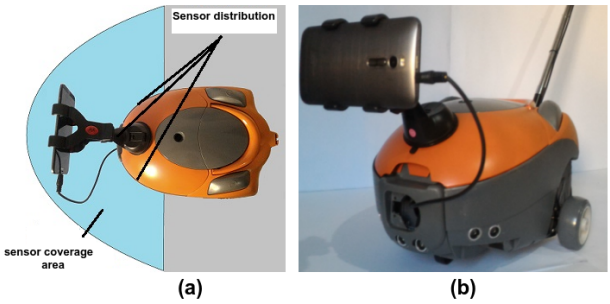


Fig. 1. Robot Bart: A platform to aid the visually impaired.

II. RELATED WORKS

A. Robotics System

We can divide the areas of Robotics into two parts, we have robot manipulators (robotic arms) and mobile robots [1]. The manipulator robots are focused on performing repetitive tasks where it is required precision in the assembly lines [1], also, we points out that manipulator robots act within a range limit (working volume) and with stationary (fixed) base. Focused on a mass production, the manipulator robots are dedicated to a process, once programmed, it must operate for months with little maintenance, thus guaranteeing a plan for profit.

Mobile robots have a locomotion system and are strongly connected with the areas of sensing and reasoning [1]. Robot soccer, humanoids, robots used in inspection operations (dangerous zones) [1] can be cited as examples. Mobile robots can also be used to give some support for human activities and also to interact with humans. Next, some mobile robots research related to this work will be better detailed.

In a paper by Cuturi et al., [2] a robot was also developed to assist visually impaired people in their task of locomotion. However, this robot has been applied to internal environments (homes, hospitals, schools) and differs from our application that is geared to urban outskirts.

In another work by Matsumura and Marranghello [3], a platform was also developed for the study of route planning algorithms. This platform has a hardware very similar to our work, however, focused on supporting the teaching of mobile robotics and not to aid the visually impaired.

2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)

Advanced Driver Assistance System based on NeuroFSM applied in the detection of autonomous human faults and support to semi-autonomous control for robotic vehicles

Diego Renan Bruno, Iago Pacheco Gomes, Fernando Santos Osório and Denis Fernando Wolf
Institute of Mathematics and Computer Science
University of São Paulo
São Carlos, Brazil
Emails: diego.bruno@usp.br, iagogomesl@usp.br, denis@icmc.usp.br, fosorio@icmc.usp.br

***Abstract*—This paper presents an ADAS (Advanced Driver Assistance System) applied in the detection of human faults and to support the semi-autonomous control of robotic vehicles in environments subject to traffic rules. The system must be able to detect and classify several different human faults which are related to a non-compliance with the local traffic rules (e.g. maximum speed allowed, stop signal, slow down, turn right/left, prohibited direction, pedestrian crossing zone), thus helping to make navigation according to the local traffic rules. We also use a new approach termed as Neuro-FSM (Neural Finite State Machine), to assess the state of the vehicle. Our ADAS system for detecting human faults, based in the Neuro-FSM, achieved an accuracy of 92.1% in the detection and classification of human actions (correct/incorrect behavior), having a great potential for the reduction of traffic accidents.**

The results are promising and very satisfactory, where we also obtained 98.3% of accuracy in the sign classification task in a traffic signal benchmark dataset (INI - German Traffic Sign Benchmark) and 83% of accuracy in the task of detecting traffic signs using 3D images in a dataset from KITTI (KITTI Vision Benchmark Suite). Through the traffic sign detection and recognition system, it was possible to compare the behavior of the driver and the vehicle state (via vehicle captured data - speed, steering, braking and acceleration), with the expected car navigation behavior according to the traffic rules present in the environment. Thus, allowing the detection of human car conduction failures, caused by imprudence or lack of attention to the visual signs (traffic rules).

I. Introduction

The structured road scenarios and urban street environments need proper traffic rules, including varied information for drivers, so that a vehicle can safely navigate on it, whether it is in autonomous or semi-autonomous mode, or even being conducted by a human. Traffic signs are used to assist the driver in the task of driving, and also inform about the local rules [1].

These signs are of great importance not only for the driver of a semi-autonomous/intelligent vehicle but also for a vehicle which travels autonomously and shares the same rules with other cars and humans. The problem is that traffic laws are not always respected, and policing is insufficient to identify all drivers who are disrespecting the local rules [2]. In view of this problem, Advanced Driver Assistance Systems (ADAS) contribute to increasing road safety.

The research area of traffic signs detection and classification has grown a lot in the last decade, mainly because the use of machine learning techniques. This growth is closely linked to the evolution of robotized vehicles, where ADAS are applied to assist in the task of driving safely (conducted by a human driver), and also, for autonomous vehicles. Safe navigation also depends on this information provided by means of traffic sign plates, traffic lights and traffic lanes.

In this paper it was proposed an integrated analysis of traffic signs detection together with the analysis of the driver/car behavior, aiming to detect faults in the task of driving, which are related to non-compliance with local traffic rules.

The proposed ADAS must be capable of alert when faults are detected, and also suggest maneuvers for overcome the situation. The problems detected in the human task of driving are divided into two classes: (a) human failure or recklessness; and (b) failure of visual attention - no detection of traffic signs. This paper main focus on the second problem, evaluating the vehicle behaviors in face of the current traffic signs detected.

II. Related Works

The technologies applied in ADAS have benefited in a reduction between 30% and 40% of traffic accidents caused in European countries [3], consequently, making these systems a great differential in terms of safety.

By conducting a bibliographic review in the area of ADAS in robotic vehicles, we highlight here some important experimental studies. In order to be able to analyze the behavior of the driver, monitoring systems are used internally in the robotic vehicle. These systems focus on the physiological and behavioral state of the driver, possibly informing when the task of driving can no longer be done safely. Two main situations that must be analyzed through the perception system are related to the detection of distractions and the identification of fatigue/sleepiness [4]. According to Young and Salmon (2012)

ICAR 2019 : International Conference on Advanced Robotics - Brazil

Visual attention system based on Fuzzy Classifier to define priority of traffic signs for intelligent robotic vehicle navigation purposes

Diego Renan Bruno and Fernando Santos Osório
Institute of Mathematics and Computer Science
University of São Paulo
São Carlos, Brazil
Emails: diego.bruno@usp.br and fosorio@icmc.usp.br

***Abstract*—In this paper we propose the use of Multiple Decision Attributes and Fuzzy Sets so that it is possible to classify the importance and priority of the detected traffic signs. The Analytic Hierarchy Process (AHP) was applied to calculate attribute weights, the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) to classify the traffic signs into their importance levels. The main objective is to contribute with a new system of perception and, through a knowledge base rules set, to be able to semantically relate the scene and to define which traffic sign is more important in a certain moment of navigation of the autonomous vehicle. The system of vision with 2D and 3D images must provide the a priori data of detection and classification of traffic signs for the fuzzy visual attention system, being able to detect the use of auxiliary signs (cones and emergency signs) and relates. Then, relate then to the detection of the navigable area in cases of road blocking (road at work, with a traffic accident, etc.) and give priority to the most important signs for the decision making of the vehicle. The results are promising and very satisfactory, we obtained an accuracy of 98.9% in the 2D classification task and 88% accuracy in the single frame 3D detection task.**

***Keywords*-robotic vehicles; Deep Learning, ADAS, Computer Vision, Traffic Sign Detection, Visual Attention System.**

I. Introduction

Autonomous cars have been extensively studied since the last decade, the most recent robotic systems have been devised and proposed methods, trying to disseminate processing and decision making using computer vision and artificial intelligence in most works. The union between the computer vision system and the artificial intelligence of this type of vehicle makes it possible to reach visual attention models capable of focusing the detection and classification of the images to situations of higher priority on the highway, as well as human drivers are capable of doing it. With this layer of visual attention, The vision system shall detect and classify the usual and emergency traffic signs (temporary), while the artificial intelligence, based on fuzzy rules, should select the information of higher priority and facilitate the decision making of the vehicle. With this mutual exchange of data, the vehicle may have prior knowledge of the planned trajectory, and update its stored information as new traffic signs are received by the perception system.

In an ideal model, the city and the car have the map of roads and routes defined a priori, thus being easily processed for the navigation of the vehicle. However, in some specific cases (urgent track repair, broken car on the road and other occasional situations) the previously mapped road has to be temporarily obstructed. In these situations, the vehicle has an outdated map that can not be used for trajectory planning.

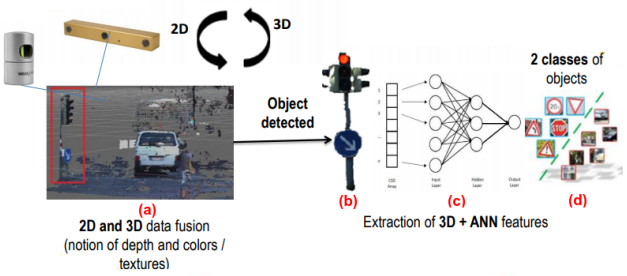


Figure 1: 3D vision system.

In this paper, we propose an innovative method of visual attention to detect and classify vertical traffic signs and emergency signs (cones and priority traffic signs), thus making it possible to interpret and react to dynamic emergency situations on the highway. This is possible using fuzzy regions to define which traffic sign has higher priority and should be passed to the vehicle's decision-making system. For the detection and classification task we use a 2D and 3D data fusion enabling greater robustness. In addition, we have developed a semantic-fuzzy knowledge bank capable of assisting the computer vision system in conjunction with fuzzy artificial intelligence in interpreting situations where the usual road signs of the highway conflict with temporary emergency signs.

In the next sections, we gather related works, then we present our algorithm of detection of traffic signs using 2D and 3D data fusion. We also present our main objective of this work, which is linked with a system of intelligent visual attention and able to focus detection of priority situations.

IEEE RAS International Summer School on Deep Learning for Robot Vision
Santiago, Chile (2019)

Robotic Computer Vision System with 2D and 3D Data Fusion for Traffic Sign Detection

Diego Renan Bruno
University of São Paulo
diego.bruno@usp.br

Fernando Santos Osorio
University of São Paulo
fosorio@icmc.usp.br

Abstract

In this paper we present an intelligent system to help autonomous vehicles in real cities and with local traffic rules. A 2D and 3D visual attention system is proposed, capable of detecting the use of signs and aids in cases of major roadblock (road under work, with a traffic accident, etc.). For this to be possible, we analyze the cones and traffic signs that usually alert a driver about this type of problem. The main objective is to provide support for autonomous vehicles to be able to find an auxiliary route that is not previously mapped. In this type of situation the vision system must work in real time. The results are promising and very satisfactory, we obtained an accuracy of 98.4% in the 2D classification task and 83% accuracy in the single frame 3D detection task.

1. Introduction

Autonomous car have been intensely studied since the last decade, more recently intelligent cities have been idealized and proposed methods, trying to spread the processing and decision making between intelligent vehicles and the smart city. The communication between an intelligent city and an autonomous car is based on the information exchange about the environment where the vehicle is inserted. The vehicle pass to the city its new collected information about the surroundings while the city send to the car the previous knowledge about the environment. With this mutual data swap, the vehicle can have a previous knowledge of the road and planned trajectory, and the city can update its stored information as new data is received from the intelligent vehicle.

In this paper, we propose an innovative method of visual attention to detect and classify vertical traffic signs and emergency signs (cones and priority traffic signs), thus making it possible to interpret and react to dynamic emergency situations on the highway.

2. Related work

In the work of Zhou and Deng [1], a system based on LIDAR (Light Detection and Ranging) and classification algorithms for the analysis of signaling plates images was used, using 3D data to improve the robustness of the task of detecting traffic signs. Through 3D point cloud data (color and spot clustering), the signboard in question was analyzed using Support Vector Machine (MVS) for classification of the traffic signs[1].

Another work, that also uses 3D data, was developed by Soiln et al. [2], each traffic sign was detected using the LIDAR sensing with classifiers based on semantic algorithms.

In the work of Wu et al. [3], a system that also uses LIDAR has been applied. This system does all the analysis through the 3D perception system. To make this possible, it uses landmarks to aid in the detection of signaling plates. [3].

3. Algorithm for detection of traffic signs with 3D data

Our algorithm for traffic signs detection uses a region of possible locations that plates usually can be found in the environment (Figure 1). In the urban transit environment, not always a traffic sign is placed on an individual pole, in some situations it can be found on a pole shared with other types of information (e.g. street name plates, light signals).

An Artificial Neural Network (ANN) with binary output has been trained with these various cases where boards (sign plates) and other elements can be found. The ANN was applied to solve this problem of classification and sign plate detection. For this to be possible, each type of case was modeled (Figure 1) based on the Velodyne LIDAR (Light Detection and Ranging) data and considering also a pair of stereo cameras, thus enabling the ANN network to respond if it is a board or an object that is not a sign plate.

REFERÊNCIAS

ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANé, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCKE, V.; VASUDEVAN, V.; VIéGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. **TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**. 2015. Software available from tensorflow.org. Disponível em: <https://www.tensorflow.org/>. Citado nas páginas 89, 90 e 91.

AIJAZI, A. K.; SERNA, A.; MARCOTEGUI, B.; CHECCHIN, P.; TRASSOUDAINE, L. Segmentation and classification of 3d urban point clouds: Comparison and combination of two approaches. In: ____. **Field and Service Robotics: Results of the 10th International Conference**. Cham: Springer International Publishing, 2016. p. 201–216. ISBN 978-3-319-27702-8. Disponível em: <https://doi.org/10.1007/978-3-319-27702-8_14>. Citado nas páginas 45, 46, 47 e 72.

AKGUL, C. B.; SANKUR, B.; YEMEZ, Y.; SCHMITT, F. 3d model retrieval using probability density-based shape descriptors. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 31, n. 6, p. 1117–1133, 2009. Citado na página 79.

AKROUT, B.; MAHDI, W. A visual based approach for drowsiness detection. In: **Intelligent Vehicles Symposium (IV), 2013 IEEE**. [S.l.: s.n.], 2013. p. 1324–1329. ISSN 1931-0587. Citado nas páginas 68 e 69.

ALDOMA, A.; BLODOW, N.; GOSSOW, D.; GEDIKLI, S.; RUSU, R.; VINCZE, M.; BRADSKI, G. Cad-model recognition and 6 dof pose estimation using 3d cues. In: **ICCV 2011, 3D Representation and Recognition (3dRR11)**. [S.l.: s.n.], 2011. Citado na página 82.

Amditis, A.; Bimpas, M.; Thomaidis, G.; Tsogas, M.; Netto, M.; Mammar, S.; Beutner, A.; Mohler, N.; Wirthgen, T.; Zipser, S.; Etemad, A.; Da Lio, M.; Cicilloni, R. A situation-adaptive lane-keeping support system: Overview of the safelane approach. **IEEE Transactions on Intelligent Transportation Systems**, v. 11, n. 3, p. 617–629, Sep. 2010. ISSN 1524-9050. Citado na página 32.

ANKERST, M.; KASTENMULLER, G.; KRIEGEL, H.-P.; SEIDL, T. 3d shape histograms for similarity search and classification in spatial databases. In: **Proc. Sixth Int'l Symp. Advances in Spatial Databases**. [S.l.: s.n.], 1999. p. 207–226. Citado na página 82.

BALALI, V.; Golparvar Fard, M. Recognition and 3d localization of traffic signs via image-based point cloud models. In: . [S.l.: s.n.], 2015. p. 206–214. 2015 ASCE International Workshop on Computing in Civil Engineering, IWCCE 2015 ; Conference date: 21-06-2015 Through 23-06-2015. Citado nas páginas 42, 46, 47 e 72.

BAY, H.; ESS, A.; TUYTELAARS, T.; GOOL, L. V. Surf: Speeded up robust features. **Computer Vision and Image Understanding (CVIU)**, 2008. Citado na página 79.

BERRI R, A. **Sistema ADAS para identificação de distrações e perturbações do motorista na condução de veículos.** 2019. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2019. Citado na página 68.

BERRI, R. A.; SILVA, A. G.; ARTHUR, R.; GIRARDI, E. Detecção automática de sonolência em condutores de veıculos utilizando imagens amplas e de baixa resolução. **Computer on the Beach**, p. 21–30, 2013. Citado nas páginas 68 e 69.

BISHOP, C. M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN 0387310738. Citado na página 84.

BRUNO. **"Algoritmo neurogenético com vistas para o planejamento de rota de robôs móveis autônomos**. 2016. <https://repositorio.unesp.br/handle/11449/138768>. [Online]. Citado na página 34.

Bruno, D. R.; Gomes, I. P.; Osório, F. S.; Wolf, D. F. Advanced driver assistance system based on neurofsm applied in the detection of autonomous human faults and support to semi-autonomous control for robotic vehicles. In: **2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE)**. [S.l.: s.n.], 2019. p. 92–97. ISSN 2639-1775. Citado na página 156.

BRUNO, D. R.; MATIAS, L. P. N.; AMARO, J.; OSóRIO, F. S.; WOLF, D. F. Computer vision system with 2d and 3d data fusion for detection of possible auxiliaries routes in stretches of interdicted roads. In: **Hawaii International Conference on System Sciences - HICSS**. [S.l.]: University of Hawai'i, 2019. Citado na página 101.

Bruno, D. R.; Osorio, F. S. Image classification system based on deep learning applied to the recognition of traffic signs for intelligent robotic vehicle navigation purposes. In: **2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR)**. [S.l.: s.n.], 2017. p. 1–6. Citado na página 90.

BRUNO, D. R.; OSORIO, F. S. Image classification system based on deep learning applied to the recognition of traffic signs for intelligent robotic vehicle navigation purposes. In: **2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR)**. [S.l.: s.n.], 2017. p. 1–6. Citado na página 113.

Bruno, D. R.; Osório, F. S. Visual attention system based on fuzzy classifier to define priority of traffic signs for intelligent robotic vehicle navigation purposes. In: **2019 19th International Conference on Advanced Robotics (ICAR)**. [S.l.: s.n.], 2019. p. 434–440. ISSN null. Citado na página 117.

Bruno, D. R.; Sales, D. O.; Amaro, J.; Osório, F. S. Analysis and fusion of 2d and 3d images applied for detection and recognition of traffic signs using a new method of features extraction in conjunction with deep learning. In: **2018 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2018. p. 1–8. ISSN 2161-4407. Citado na página 101.

Bruno, D. R.; Santos, T. C.; Silva, J. A. R.; Wolf, D. F.; Osório, F. S. Advanced driver assistance system based on automated routines for the benefit of human faults correction in robotics vehicles. In: **2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)**. [S.l.: s.n.], 2018. p. 112–117. ISSN null. Citado na página 156.

BUCKERIDGE, R. **"With autonomous, self-driving cars likely to be commonplace by around 2025, these vehicles will change our roads, our relationship with our cars and society at large. buckle up, a revolution is coming!"**. 2015. <http://factor-tech.com/feature/autonomous-cars-and-mans-future-the-road-ahead/>>. [Online]. Citado na página 32.

CALONDER, M.; LEPETIT, V.; OZUYSAL, M.; TRZCINSKI, T.; STRECHA, C.; FUA, P. Brief: Computing a local binary descriptor very fast. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 34, n. 7, p. 1281–1298, July 2012. ISSN 0162-8828. Citado na página 123.

CARDARELLI, E. Vision-based blind spot monitoring. In: ESKANDARIAN, A. (Ed.). **Handbook of Intelligent Vehicles**. [S.l.]: Springer London, 2012. p. 1071–1087. ISBN 978-0-85729-084-7. Citado na página 71.

CARINA. **"CaRINA 2 LRM-ICMC"**. 2020. <https://lrm.icmc.usp.br>. [Online]. Citado nas páginas 38 e 149.

CARROLL, J.; BELLEHUMEUR, D.; CARROLL, C. **System and method for detecting and measuring ethyl alcohol in the blood of a motorized vehicle driver transdermally and non-invasively in the presence of interferents**. [S.l.]: Google Patents, 2013. US Patent App. 2013/0027209. Citado nas páginas 68 e 69.

CAUCHIE, J.; FIOLET, V.; VILLERS, D. Optimization of an hough transform algorithm for the search of a center. **Pattern Recognition**, v. 41, n. 2, p. 567–574, 2008. ISSN 0031-3203. Citado na página 68.

CDC. **"Mobile vehicle safety-impaired driving"**. 2013. <http://www.cdc.gov/MotorVehicleSafety/Impaired_Driving/impaireddrv_factsheet.hml>. [Online]. Citado na página 32.

Chen, C.; Seff, A.; Kornhauser, A.; Xiao, J. Deepdriving: Learning affordance for direct perception in autonomous driving. In: **2015 IEEE International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2015. p. 2722–2730. ISSN 2380-7504. Citado na página 56.

Chen, J.; Zhao, P.; Liang, H.; Mei, T. A multiple attribute-based decision making model for autonomous vehicle in urban environment. In: **2014 IEEE Intelligent Vehicles Symposium Proceedings**. [S.l.: s.n.], 2014. p. 480–485. ISSN 1931-0587. Citado nas páginas 126, 127, 128 e 129.

CHEN, L.; LI, Q.; LI, M.; ZHANG, L.; MAO, Q. Design of a multi-sensor cooperation travel environment perception system for autonomous vehicle. **Sensors**, v. 12, n. 9, p. 12386–12404, 2012. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/12/9/12386>. Citado nas páginas 42 e 56.

CHUA, C. S.; JARVIS, R. Point signatures: a new representation for 3d object recognition. **International Journal of Computer Vision**, v. 25, n. 1, p. 63–85, 1997. Citado na página 80.

CIRESAN, D. C.; MEIER, U.; MASCI, J.; SCHMIDHUBER, J. Multi-column deep neural network for traffic sign classification. **Neural Networks**, v. 32, p. 333–338, 2012. Disponível em: <http://dblp.uni-trier.de/db/journals/nn/nn32.html#CiresanMMS12>. Citado na página 56.

Cireşan, D.; Meier, U.; Masci, J.; Schmidhuber, J. A committee of neural networks for traffic sign classification. In: **The 2011 International Joint Conference on Neural Networks**. [S.l.: s.n.], 2011. p. 1918–1921. ISSN 2161-4407. Citado na página 56.

DAI, J.; TENG, J.; BAI, X.; SHEN, Z.; XUAN, D. Mobile phone based drunk driving detection. In: IEEE. **Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS**. [S.l.], 2010. p. 1–8. Citado nas páginas 68 e 69.

DARPA. **"DARPA, - (Defense Advanced Research Projects Agency) Grand Challenge."**. 2013. <www.darpa.mil>. [Online]. Citado na página 34.

DKHIL, M. B.; NEJI, M.; WALI, A.; ALIMI, A. M. A new approach for a safe car assistance system. In: **Advanced Logistics and Transport (ICALT), 2015 4th International Conference on**. [S.l.: s.n.], 2015. p. 217–222. Citado na página 69.

DONAHUE, J.; JIA, Y.; VINYALS, O.; HOFFMAN, J.; ZHANG, N.; TZENG, E.; DARRELL, T. Decaf: A deep convolutional activation feature for generic visual recognition. In: XING, E. P.; JEBARA, T. (Ed.). **Proceedings of the 31st International Conference on Machine Learning**. Bejing, China: PMLR, 2014. (Proceedings of Machine Learning Research, 1), p. 647–655. Disponível em: <http://proceedings.mlr.press/v32/donahue14.html>. Citado na página 113.

EBRAHIM, P.; ABDELLAOUI, A.; STOLZMANN, W.; YANG, B. Eyelid-based driver state classification under simulated and real driving conditions. In: **Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on**. [S.l.: s.n.], 2014. p. 3190–3196. Citado nas páginas 68 e 69.

EDMONDS, D. S.; HOPTA, J. W. **Driver alcohol ignition interlock**. [S.l.]: Google Patents, 2001. US Patent 6,229,908. Citado nas páginas 68 e 69.

FIGUEIRA A, C. Manobra de ultrapassagem em pista simples bidirecional com simulador de direÇÃo. In: ESCOLA DE ENGENHARIA DE SãO CARLOS. **Tese de doutorado: USP São Carlos - EESC**. [S.l.], 2019. p. 583–588. Citado nas páginas 57, 58, 59 e 71.

FISHER, A. Inside google's quest to popularize self-driving cars. **Popular Science. Retrieved from http://www. popsci. com/cars/article/2013-09/google-self-driving-car**, 2014. Acessado em: 11.06.2015. Citado na página 33.

FRINTROP, S.; ROME, E.; CHRISTENSEN, H. I. Computational visual attention systems and their cognitive foundations: A survey. **ACM Trans. Appl. Percept.**, ACM, New York, NY, USA, v. 7, n. 1, p. 6:1–6:39, jan. 2010. ISSN 1544-3558. Disponível em: <http://doi.acm.org/10.1145/1658349.1658355>. Citado na página 92.

FRITSCH, J.; KUEHNL, T.; GEIGER, A. A new performance measure and evaluation benchmark for road detection algorithms. In: **International Conference on Intelligent Transportation Systems (ITSC)**. [S.l.: s.n.], 2013. Citado nas páginas 121 e 122.

FROME, A.; HUBER, D.; KOLLURI, R.; BULOW, T.; MALIK, J. Recognizing objects in range data using regional point descriptors. In: **Proceedings of European Conference on Computer Vision (ECCV)**. [S.l.: s.n.], 2004. v. 3, p. 224–237. Citado na página 80.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological Cybernetics**, v. 36, n. 4, p. 193–202, Apr 1980. ISSN 1432-0770. Disponível em: <https://doi.org/10.1007/BF00344251>. Citado na página 89.

GEIGER, A.; LENZ, P.; STILLER, C.; URTASUN, R. Vision meets robotics: The kitti dataset. **International Journal of Robotics Research (IJRR)**, 2013. Citado na página 113.

GEIGER, A.; LENZ, P.; URTASUN, R. Are we ready for autonomous driving? the kitti vision benchmark suite. In: **Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2012. Citado nas páginas 109, 114, 115, 141, 149 e 155.

GONZALEZ, R.; WOODS, R. **Processamento Digital De Imagens**. [S.l.]: Artliber, 2010. Citado nas páginas 77 e 78.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.: s.n.], 2017. ISSN 15487105. ISBN 9780521835688. Citado na página 112.

GRIGOREV, A.; SHANMUGAMANI, R.; BOSCHETTI, A.; MASSARON, L.; THAKUR, A. **TensorFlow Deep Learning Projects**. [S.l.]: Packt, 2018. Citado nas páginas 92 e 93.

HABERMANN, D.; HATA, A.; WOLF, D.; OSóRIO, F. 3d point clouds segmentation for autonomous ground vehicle. In: **Simpósio Brasileiro de Engenharia de Sistemas Computacionais (SBESC)**. [S.l.: s.n.], 2013. Citado na página 94.

HAILE, E. **Drunk driver detection system**. [S.l.]: Google Patents, 1987. US Patent 4,716,413. Citado nas páginas 68 e 69.

____. **Drunk driver detection system**. [S.l.]: Google Patents, 1992. US Patent 5,096,329. Citado nas páginas 68 e 69.

HAN, J.; SHAO, L.; XU, D.; SHOTTON. Enhanced computer vision with microsoft kinect sensor: A review. **IEEE Transactions on Cybernetics**, v. 43, n. 5, 2013. Citado na página 45.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. 2nd. ed. [S.l.]: Prentice Hall, 1998. Citado na página 88.

HILAGA, M.; SHINAGAWA, Y.; KOHMURA, T.; KUNII, T. L. Topology matching for fully automatic similarity estimation of 3d shapes. In: **Proceedings of. ACM SIGGRAPH**. [S.l.: s.n.], 2001. p. 203–212. Citado na página 83.

HORN, B. K. P. Extended gaussian images. In: **Proc. IEEE**. [S.l.: s.n.], 1984. v. 72, p. 1671–1686. Citado na página 82.

Hossain, S.; Hyder, Z. Traffic road sign detection and recognition for automotive vehicles. In: **International Journal of Computer Applications**. [S.l.: s.n.], 2015. p. 565–568. Citado nas páginas 51, 52, 56, 71 e 72.

Houben, S.; Stallkamp, J.; Salmen, J.; Schlipsing, M.; Igel, C. Detection of traffic signs in real-world images: The german traffic sign detection benchmark. In: **The 2013 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2013. p. 1–8. ISSN 2161-4407. Citado na página 56.

HSU-SHIHSHIH; SHYUR H, J.; ZAVADSKAS E, K. **An extension of TOPSIS for group decision making**. [S.l.]: Springer, 2007. Citado na página 117.

HUVAL, B.; WANG, T.; TANDON, S.; KISKE, J.; SONG, W.; PAZHAYAMPALLIL, J.; ANDRILUKA, M.; RAJPURKAR, P.; MIGIMATSU, T.; CHENG-YUE, R.; MUJICA, F. A.; COATES, A.; NG, A. Y. An empirical evaluation of deep learning on highway driving. **CoRR**, abs/1504.01716, 2015. Disponível em: <http://arxiv.org/abs/1504.01716>. Citado nas páginas 42 e 56.

HWANG C, L.; YOON, K. **Multiple Attribute Decision Making - Methods and Applications**. [S.l.]: Springer, 2011. Citado na página 117.

INFRAESTRUTURA. **"Associação brasileira de prevenção dos acidentes de trânsito"**. 2019. <http://transportes.gov.br/ultimas-noticias/7999-estudo-aponta-que-mais-de-50-dos-acidentes-de-tr>. Citado na página 32.

Jung, S.; Lee, U.; Jung, J.; Shim, D. H. Real-time traffic sign recognition system with deep convolutional neural network. p. 31–34, Aug 2016. Citado nas páginas 54 e 72.

KISACANIN, B. Automotive vision for advanced driver assistance systems. In: IEEE. **VLSI Design, Automation and Test (VLSI-DAT), 2011 International Symposium on**. [S.l.], 2011. p. 1–2. Citado na página 65.

KLASING, K.; WOLLHERR, D.; BUSS, M. A clustering method for efficient segmentation of 3d laser data. In: **IEEE International Conference on Robotics and Automation (ICRA)**. [S.l.: s.n.], 2008. p. 4043–4048. Citado na página 85.

KNOPP, J.; PRASAD, M.; WILLEMS, G.; TIMOFTE, R.; GOOL, L. V. Hough transform and 3d surf for robust three dimensional classication. In: **Proceedings of European Conference on Computer Vision (ECCV)**. [S.l.: s.n.], 2010. Citado na página 80.

KRIG, S. **Computer Vision Metrics - Survey, Taxonomy and Analysis of Computer Vision, Visual Neuroscience, and Deep Learning**. [S.l.]: Springer, 2016. Citado nas páginas 78 e 79.

KUMAR, A. M.; SIMON, P. Review of lane detection and tracking algorithms in advanced driver assistance system. **Int. J. Comput. Sci. Inf. Technol**, v. 7, n. 4, p. 65–78, 2015. Citado na página 69.

KUTILA, M. **Methods for Machine Vision Based Driver Monitoring Applications: Dissertation**. Tese (Doutorado) — Tampere University of Technology (TUT), Finland, 2006. Project code: 612. Citado nas páginas 68 e 71.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. In: **Proceedings of the IEEE**. [S.l.: s.n.], 1998. p. 2278–2324. Citado na página 89.

LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: **Proceedings of 2010 IEEE International Symposium on Circuits and Systems**. [S.l.: s.n.], 2010. p. 253–256. ISSN 0271-4302. Citado nas páginas 54 e 112.

Lee, H. S.; Kim, K. Simultaneous traffic sign detection and boundary estimation using convolutional neural network. **IEEE Transactions on Intelligent Transportation Systems**, v. 19, n. 5, p. 1652–1663, May 2018. Citado nas páginas 57, 62, 63, 71 e 72.

LEFÈVRE, S.; LAUGIER, C.; IBAÑEZ-GUZMÁN, J. Exploiting map information for driver intention estimation at road intersections. In: IEEE. **Intelligent Vehicles Symposium (IV), 2011 IEEE**. [S.l.], 2011. p. 583–588. Citado nas páginas 64 e 71.

LENSKIY, A.; LEE, J. Driver's eye blinking detection using novel color and texture segmentation algorithms. **International Journal of Control, Automation and Systems**, Springer, v. 10, n. 2, p. 317–327, 2012. Citado na página 69.

Li, L.; Lv, Y.; Wang, F. Traffic signal timing via deep reinforcement learning. **IEEE/CAA Journal of Automatica Sinica**, v. 3, n. 3, p. 247–254, July 2016. ISSN 2329-9266. Citado na página 56.

____. Deep convnet-based vehicle detection using 3d-lidar reflection intensity data. **Third Iberian Robotics Conference**, v. 3, n. 3, p. 247–254, July 2017. Citado na página 71.

Li, L.; Wen, D. Parallel systems for traffic control: A rethinking. **IEEE Transactions on Intelligent Transportation Systems**, v. 17, n. 4, p. 1179–1182, April 2016. ISSN 1524-9050. Citado na página 56.

LIN, C.-C.; WANG, M.-S. Road sign recognition with fuzzy adaptive pre-processing models. **Sensors**, v. 12, n. 5, p. 6415–6433, 2012. ISSN 1424-8220. Disponível em: <http://www.mdpi.com/1424-8220/12/5/6415>. Citado nas páginas 62, 63 e 72.

LOWE, D. G. Object recognition from local scale-invariant features. In: **Proceedings of the International Conference on Computer Vision**. [S.l.: s.n.], 1999. p. 1150–1157. Citado nas páginas 79 e 82.

LUDWIG, J.; MONTGOMERY. **Redes Neurais: fundamentos e aplicações com programas em C**. [S.l.]: Ciência Moderna, 2007. Citado nas páginas 86 e 88.

Marinas, J.; Salgado, L.; Arróspide, J.; Camplani, M. Traffic sign detection and tracking using robust 3d analysis. In: **2012 Third International Conference on Emerging Security Technologies**. [S.l.: s.n.], 2012. p. 78–81. Citado nas páginas 42 e 56.

MATARIC. **introdução a robótica**. Blucher/UNESP Editora,: [s.n.], 2014. Citado na página 94.

Mathias, M.; Timofte, R.; Benenson, R.; Van Gool, L. Traffic sign recognition — how far are we from the solution? In: **The 2013 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2013. p. 1–8. ISSN 2161-4407. Citado nas páginas 42, 43, 56, 57 e 70.

MAXWELL. Aprendizado por reforço. In: **Proceedings of the IEEE**. [S.l.: s.n.], 2010. Citado na página 85.

MIAN, A. S.; BENNAMOUN, M.; OWENS, R. A. On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. **International Journal of Computer Vision**, v. 89, n. 2-3, p. 348–361, 2010. Citado na página 80.

MIT. False positive detection in 2d data. In: . [S.l.: s.n.], 2017. Citado na página 73.

MITCHELL, T. **Machine Learning**. [S.l.]: McGraw-Hill, 1997. Citado nas páginas 84 e 86.

Mogelmose, A.; Trivedi, M. M.; Moeslund, T. B. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. **IEEE Transactions on Intelligent Transportation Systems**, v. 13, n. 4, p. 1484–1497, Dec 2012. ISSN 1524-9050. Citado na página 56.

MOUTARDE, F.; BARGETON, A.; HERBIN, A.; CHANUSSOT, L. Modular traffic sign recognition applied to on-vehicle real-time visual detection of american and european speed limit signs. **CoRR**, abs/0910.1295, 2009. Disponível em: <http://arxiv.org/abs/0910.1295>. Citado na página 56.

MUR-ARTAL, R.; TARDóS, J. D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. **IEEE Transactions on Robotics**, v. 33, n. 5, p. 1255–1262, Oct 2017. ISSN 1552-3098. Citado na página 123.

Murata, K.; Fujita, E.; Kojima, S.; Maeda, S.; Ogura, Y.; Kamei, T.; Tsuji, T.; Kaneko, S.; Yoshizumi, M.; Suzuki, N. Noninvasive biological sensor system for detection of drunk driving. **IEEE Transactions on Information Technology in Biomedicine**, v. 15, n. 1, p. 19–25, Jan 2011. ISSN 1089-7771. Citado na página 32.

MURATA, K.; FUJITA, E.; KOJIMA, S.; MAEDA, S.; OGURA, Y.; KAMEI, T.; TSUJI, T.; KANEKO, S.; YOSHIZUMI, M.; SUZUKI, N. Noninvasive biological sensor system for detection of drunk driving. **Information Technology in Biomedicine, IEEE Transactions on**, v. 15, n. 1, p. 19–25, Jan 2011. ISSN 1089-7771. Citado nas páginas 68 e 69.

MURPHY, K. **Machine Learning – A probabilistic Perspective**. [S.l.]: MIT Press, 2012. Citado nas páginas 85 e 93.

NILSBACK, M.-E.; ZISSERMAN, A. Automated flower classification over a large number of classes. In: **Indian Conference on Computer Vision, Graphics and Image Processing**. [S.l.: s.n.], 2008. Citado na página 111.

Nunn, C.; Kummert, A.; Muller-Schneiders, S. A novel region of interest selection approach for traffic sign recognition based on 3d modelling. In: **2008 IEEE Intelligent Vehicles Symposium**. [S.l.: s.n.], 2008. p. 654–659. ISSN 1931-0587. Citado na página 56.

OGAMA. **"O. Art. 306 do código de trânsito brasileiro: do texto original às mudanças surgidas com o advento da lei n. 12.760/12"**. 2014. Http://revista.pgsskroton.com.br/index.php/juridicas/article/view/302/283. [Online]. Citado na página 32.

OHBUCHI, R.; OSADA, K.; FURUYA, T.; BANNO, T. Salient local visual features for shape-based 3d model retrieval. In: **Proceedings of IEEE Shape Modeling International (SMI)**. [S.l.: s.n.], 2008. Citado na página 82.

OSADA, R.; FUNKHOUSER, T.; CHAZELLE, B.; DOBKIN, D. Shape distributions. **ACM Trans. Graphics**, v. 21, n. 4, p. 807–832, 2002. Citado na página 82.

Pachêco Gomes, I.; Renan Bruno, D.; Santos Osório, F.; Fernando Wolf, D. Diagnostic analysis for an autonomous truck using multiple attribute decision making. In: **2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)**. [S.l.: s.n.], 2018. p. 283–290. Citado nas páginas 126, 127, 128 e 129.

PAN, S. J.; YANG, Q. **A survey on transfer learning**. 2010. 1345–1359 p. Citado na página 112.

PECHANSKY; BONI R.; DIEMEN, L. V. B. D. P. I. Z. M. C. R. L. R. D. **"Highly reported prevalence of drinking and driving in brazil: data from the first representative household study"**. 2009. [Online]. Citado na página 32.

PEISSNER, M.; DOEBLER, V.; METZE, F. Can voice interaction help reducing the level of distraction and prevent accidents? 2011. Citado na página 32.

REDELMEIER, D. A.; TIBSHIRANI, R. J. Association between cellular-telephone calls and motor vehicle collisions. **The New England Journal of Medicine**, v. 336, n. 7, p. 453–458, 1997. Citado na página 33.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: Unified, real-time object detection. In: **The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. Citado nas páginas 91 e 92.

REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. **CoRR**, abs/1804.02767, 2018. Disponível em: <http://arxiv.org/abs/1804.02767>. Citado nas páginas 91 e 102.

RESTREPO. **"Intersect over Union (IoU)"**. 2018. <http:/http://ronny.rest/tutorials/module/localization_001/iou/>. [Online]. Citado nas páginas 136 e 137.

RICHTER, S. R.; VINEET, V.; ROTH, S.; KOLTUN, V. Playing for data: Ground truth from computer games. In: SPRINGER. **European Conference on Computer Vision**. [S.l.], 2016. p. 102–118. Citado nas páginas 65, 66, 67 e 71.

ROMERO, R. A.; PRESTES, E.; OSóRIO, F. S.; WOLF, D. F. **Robótica Móvel**. [S.l.]: LTC, 2014. ISBN 8521623038. Citado na página 94.

ROSTEN, E.; DRUMMOND, T. Machine learning for high-speed corner detection. In: LEONARDIS, A.; BISCHOF, H.; PINZ, A. (Ed.). **Computer Vision – ECCV 2006**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 430–443. ISBN 978-3-540-33833-8. Citado na página 123.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence: A modern approach**. [S.l.]: Pearson Education, 2003. Citado na página 86.

RUSU, R.; BLODOW, N.; BEETZ, M. Fast point feature histograms (fpfh) for 3d registration. In: **Proceedings of the International Conference on Robotics and Automation (ICRA)**. [S.l.: s.n.], 2009. Citado na página 80.

RUSU, R. B.; BRADSKI, G.; THIBAUX, R.; HSU, J. Fast 3d recognition and pose using the viewpoint feature histogram. In: **Proceedings of the 23rd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**. [S.l.: s.n.], 2010. Citado na página 82.

SAATY, T. L. **How to make a decision: The analytic hierarchy process**. [S.l.]: Springer, 1990. Citado na página 117.

SALES D, O. **Extração de features 3D para o reconhecimento de objetos em nuvem de pontos.** 2017. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2017. Citado nas páginas 105, 106 e 139.

SALES, D. O. Neurofsm: aprendizado de autômatos finitos através do uso de redes neurais artificiais aplicadas à robôs móveis e veículos autônomos. In: **Institudo de Ciências Matemáticas e de Computação - Universidade de São Paulo (ICMC-USP)**. [S.l.: s.n.], 2017. Citado nas páginas 78, 79, 80, 81, 82, 84, 85, 86, 87, 88 e 104.

SALES, D. O.; AMARO, J.; OSóRIO, F. S. 3d feature extraction for objects recognition. In: **IEEE Latin American Robotics Symposium (LARS)**. [S.l.: s.n.], 2017. Citado nas páginas 106 e 107.

SALES, D. O.; FERNANDES, L. C.; OSÓRIO, F. S.; WOLF, D. F. Fsm-based visual navigation for autonomous vehicles. In: **Workshop on Visual Control of Mobile Robots-IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Algarve, Portugall**. [S.l.: s.n.], 2012. Citado na página 33.

SALTI, S.; PETRELLI, A.; TOMBARI, F.; FIORAIO, N.; STEFANO, L. D. Traffic sign detection via interest region extraction. **Pattern Recogn.**, Elsevier Science Inc., New York, NY, USA, v. 48, n. 4, p. 1039–1049, abr. 2015. ISSN 0031-3203. Disponível em: <http://dx.doi.org/10.1016/j.patcog.2014.05.017>. Citado nas páginas 42, 43, 44 e 72.

SALVADOR, A. A culpa foi do celular. **Revista Veja, 21/12/2011**, 2011. Citado na página 32.

SARBOLANDI, H.; LEFLOCH, D.; KOLB, A. Kinect range sensing: Structured-light versus time-of-flight kinect. **CoRR**, abs/1505.05459, 2015. Disponível em: <http://arxiv.org/abs/1505.05459>. Citado na página 96.

Schlosser, J.; Montemerlo, M.; Salisbury, K. Intelligent road sign detection using 3d scene geometry. In: **2010 IEEE/RSJ International Conference on Intelligent Robots and Systems**. [S.l.: s.n.], 2010. p. 740–745. ISSN 2153-0866. Citado nas páginas 57, 61, 62, 63, 71 e 72.

SEGURAS, V. **"Associação brasileira de prevenção dos acidentes de trânsito"**. 2016. <http://www.viasseguras.com/layout/set/print/os_acidentes/estatisticas/estatisticas_nacionais>. [Online]. Citado na página 33.

Sermanet, P.; LeCun, Y. Traffic sign recognition with multi-scale convolutional networks. In: **The 2011 International Joint Conference on Neural Networks**. [S.l.: s.n.], 2011. p. 2809–2813. ISSN 2161-4407. Citado na página 56.

SHIRAZI, M. M.; RAD, A. B. Modeling the steering behavior of intoxicated drivers. In: **Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on**. [S.l.: s.n.], 2012. p. 648–653. ISSN 2153-0009. Citado nas páginas 68 e 69.

SIEGWART, R.; NOURBAKHSH, I. R. **Introduction to Autonomous Mobile Robots**. Scituate, MA, USA: Bradford Company, 2004. ISBN 026219502X. Citado na página 94.

SILBERMAN, N.; HOIEM, D.; KOHLI, P.; FERGUS, R. Indoor segmentation and support inference from rgbd images. In: **Proceedings of the 12th European Conference on Computer Vision - Volume Part V**. Berlin, Heidelberg: Springer-Verlag, 2012. (ECCV'12), p. 746–760. ISBN 978-3-642-33714-7. Disponível em: <http://dx.doi.org/10.1007/978-3-642-33715-4_54>. Citado na página 98.

SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais para Engenharia e Ciências Aplicadas**. [S.l.]: Artliber, 2010. Citado nas páginas 77, 78, 86, 87 e 88.

Soilan, M.; Riveiro, B.; Matinez-Sánchez, J.; Arias, P.; Wang, C.; Li, J. Traffic sign detection in mls acquired point clouds for geometric and image-based semantic inventory. In: **ISPRS Journal of Photogrammetry and Remote Sensing**. [S.l.: s.n.], 2015. p. 565–568. Citado nas páginas 47, 48, 49, 71 e 72.

STALLKAMP, J.; SCHLIPSING, M.; SALMEN, J.; IGEL, C. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In: **IEEE International Joint Conference on Neural Networks**. [S.l.: s.n.], 2011. p. 1453–1460. Citado nas páginas 113, 114, 142, 149 e 155.

____. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. **Neural Networks**, v. 32, p. 323–332, 2012. Disponível em: <http://dblp.uni-trier.de/db/journals/nn/nn32.html#StallkampSSI12>. Citado nas páginas 56, 57, 59, 60, 61 e 71.

STRAYER, D. L.; COOPER, J. M.; TURRILL, J.; COLEMAN, J.; MEDEIROS-WARD, N.; BIONDI, F. Measuring cognitive distraction in the automobile. **AAA Foundation for Traffic Safety - June 2013**, p. 1–34, 2013. Citado nas páginas 32 e 33.

STRAYER, D. L.; DREWS, F. A. Profiles in driver distraction: effects of cell phone conversations on younger and older drivers. **The Journal of the Human Factors and Ergonomics**, v. 46, n. 4, p. 640–648, 2004. Citado na página 33.

STRAYER, D. L.; DREWS, F. A.; JOHNSTON, W. A. Cell phone-induced failures of visual attention during simulated driving. **Journal of experimental psychology: Applied**, American Psychological Association, v. 9, n. 1, p. 23, 2003. Citado na página 32.

SUNDAR, H.; SILVER, D.; GAGVANI, N.; DICKINSON, S. Skeleton based shape matching and retrieval. In: **Proc. Shape Modeling Int'l**. [S.l.: s.n.], 2003. Citado na página 83.

Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 2818–2826. ISSN 1063-6919. Citado na página 111.

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. In: **CVPR**. [S.l.]: IEEE Computer Society, 2016. p. 2818–2826. Citado na página 113.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition, Fourth Edition**. 4th. ed. Orlando, FL, USA: Academic Press, Inc., 2008. ISBN 1597492728, 9781597492720. Citado na página 84.

TIMOFTE, R.; PRISACARIU, V. A.; Van Gool, L. J.; REID, I. Combining traffic sign detection with 3d tracking towards better driver assistance. In: CHEN, C. H. (Ed.). **Emerging Topics in Computer Vision and its Applications**. [S.l.]: World Scientific Publishing, 2011. Citado nas páginas 43 e 51.

Timofte, R.; Zimmermann, K.; Gool, L. V. Multi-view traffic sign detection, recognition, and 3d localisation. In: **2009 Workshop on Applications of Computer Vision (WACV)**. [S.l.: s.n.], 2009. p. 1–8. ISSN 1550-5790. Citado nas páginas 42, 43, 69 e 72.

____. Multi-view traffic sign detection, recognition, and 3d localisation. In: **2009 Workshop on Applications of Computer Vision (WACV)**. [S.l.: s.n.], 2009. p. 1–8. ISSN 1550-5790. Citado na página 56.

TOMBARI, F.; FIORAIO, N.; CAVALLARI, T.; SALTI, S.; PETRELLI, A.; STEFANO, L. D. Automatic detection of pole-like structures in 3d urban environments. In: **Proceedings of 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**. [S.l.: s.n.], 2014. Citado nas páginas 78, 79, 80 e 81.

TOMBARI, F.; SALTI, S.; STEFANO, L. D. Unique signatures of histograms for local surface description. In: **Proceedings of European Conference on Computer Vision (ECCV)**. [S.l.: s.n.], 2010. Citado na página 81.

TUNG, T.; SCHMITT, F. The augmented multiresolution reeb graph approach for content-based retrieval of 3d shapes. **International Journal on Shape Modeling**, v. 11, n. 1, 2005. Citado na página 83.

VIOLA, P.; JONES, M. Robust real-time face detection. **International journal of computer vision**, Springer, v. 57, n. 2, p. 137–154, 2004. Citado na página 68.

WAN, J.; WANG, D.; HOI, S. C. H.; WU, P.; ZHU, J.; ZHANG, Y.; LI, J. Deep learning for content-based image retrieval: A comprehensive study. In: **Proceedings of the 22Nd ACM International Conference on Multimedia**. New York, NY, USA: ACM, 2014. (MM '14), p. 157–166. ISBN 978-1-4503-3063-3. Disponível em: <http://doi.acm.org/10.1145/2647868.2654948>. Citado na página 56.

WANG, S.; PAN, H.; ZHANG, C.; TIAN, Y. Rgb-d image-based detection of stairs, pedestrian crosswalks and traffic signs. **J. Vis. Comun. Image Represent.**, Academic Press, Inc., Orlando, FL, USA, v. 25, n. 2, p. 263–272, fev. 2014. ISSN 1047-3203. Disponível em: <http://dx.doi.org/10.1016/j.jvcir.2013.11.005>. Citado nas páginas 47 e 72.

WAYMO. **"Open Dataset - Waymo"**. 2020. <https://waymo.com/open/>. [Online]. Citado na página 149.

WENG, S.; LI, J.; CHEN, Y.; WANG, C. Road traffic sign detection and classification from mobile lidar point clouds. In: . [S.l.: s.n.], 2016. p. 99010A. Citado nas páginas 63 e 64.

WILLIAMS, S. J. Our hard days' nights. **Contexts**, SAGE Publications, v. 10, n. 1, p. 26–31, 2011. Citado na página 32.

WILSON FA, S. J. Trends in fatalities from distracted driving in the united states. **US national library of medicine national institutes of health**, v. 100, n. 11, p. 2213–2219, Jan 2010. Citado nas páginas 32, 65 e 71.

Wu, S.; Wen, C.; Luo, H.; Chen, Y.; Wang, C.; Li, J. Using mobile lidar point clouds for traffic sign detection and sign visibility estimation. In: **2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)**. [S.l.: s.n.], 2015. p. 565–568. ISSN 2153-6996. Citado nas páginas 47, 49, 50 e 72.

XU, K.; BA, J.; KIROS, R.; CHO, K.; COURVILLE, A.; SALAKHUDINOV, R.; ZEMEL, R.; BENGIO, Y. Show, attend and tell: Neural image caption generation with visual attention. In: BACH, F.; BLEI, D. (Ed.). **Proceedings of the 32nd International Conference on Machine Learning**. Lille, France: PMLR, 2015. (Proceedings of Machine Learning Research, v. 37), p. 2048–2057. Disponível em: <http://proceedings.mlr.press/v37/xuc15.html>. Citado na página 92.

YOSINSKI, J.; CLUNE, J.; BENGIO, Y.; LIPSON, H. How transferable are features in deep neural networks? In: **Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2**. Cambridge, MA, USA: MIT Press, 2014. (NIPS'14), p. 3320–3328. Disponível em: <http://dl.acm.org/citation.cfm?id=2969033.2969197>. Citado na página 112.

Yuan, Y.; Xiong, Z.; Wang, Q. An incremental framework for video-based traffic sign detection, tracking, and recognition. **IEEE Transactions on Intelligent Transportation Systems**, v. 18, n. 7, p. 1918–1929, July 2017. ISSN 1524-9050. Citado nas páginas 52, 53, 54, 55 e 72.

ZAHARESCU, A.; BOYER, E.; VARANASI, K.; HORAUD, R. Surface feature detection and description with applications to mesh matching. In: **Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2009. Citado na página 81.

Zeng, Y.; Xu, X.; Shen, D.; Fang, Y.; Xiao, Z. Traffic sign recognition using kernel extreme learning machines with deep perceptual features. **IEEE Transactions on Intelligent Transportation Systems**, v. 18, n. 6, p. 1647–1653, June 2017. ISSN 1524-9050. Citado nas páginas 42, 56 e 71.

ZHAO, G.; YUAN, J.; DANG, K. Height gradient histogram (high) for 3d scene labeling. In: **Proceedings of 2014 Second International Conference on 3D Vision (3DV)**. [S.l.: s.n.], 2014. Citado nas páginas 70 e 80.

Zhe, X.; Jingyi, R.; Chaoqian, B. A traffic signs' detection method of contour approximation based on concave removal. In: **2016 Chinese Control and Decision Conference (CCDC)**. [S.l.: s.n.], 2016. p. 5199–5204. ISSN 1948-9447. Citado na página 71.

Zhou, L.; Deng, Z. Lidar and vision-based real-time traffic sign detection and recognition algorithm for intelligent vehicle. In: **17th International IEEE Conference on Intelligent Transportation Systems (ITSC)**. [S.l.: s.n.], 2014. p. 578–583. ISSN 2153-0009. Citado nas páginas 44, 69, 72 e 89.

Zhu, Z.; Liang, D.; Zhang, S.; Huang, X.; Li, B.; Hu, S. Traffic-sign detection and classification in the wild. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 2110–2118. ISSN 1063-6919. Citado nas páginas 42 e 56.

Zuo, Z.; Yu, K.; Zhou, Q.; Wang, X.; Li, T. Traffic signs detection based on faster r-cnn. In: **2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)**. [S.l.: s.n.], 2017. p. 286–288. ISSN 2332-5666. Citado na página 71.

APÊNDICE

A

TRABALHOS TÉCNICOS E ACADÊMICOS RELACIONADOS COM A TESE DE DOUTORADO: USP E FATEC

A.1 Trabalhos desenvolvidos como professor

A.1.1 Programa de Aperfeiçoamento de Ensino

Além das pesquisas relacionados ao trabalho de doutorado apresentado, também foi possível, nestes anos de doutorado, trabalhar no Programa de Aperfeiçoamento de Ensino (PAE) junto com o professor Dr. Fernando Santos Osório, nas disciplinas: (*SSC*5888) Robótica Móvel Autônoma, (*SSC*0715) Sensores Inteligentes e (*SSC*0714) Programação de Robôs Móveis (2016, 2017, 2018 e 2019). O programa PAE garantiu um enriquecimento de conhecimento e experiência para as atividades de docência, tanto na parte de ensino quanto de pesquisa. Foi possível também participar de bancas de Trabalhos de Conclusão de Graduação (TCC) e estágios na área de pesquisa desta tese.

A.1.2 Docência

Também foram desenvolvidos trabalhos como professor da Faculdade de Tecnologia de Catanduva (FATEC), ministrando aulas nas disciplinas de Laboratório de Automação e Redes Industriais. Garantindo uma maior experiência acadêmica no ensino em nível de graduação. Também foi possível realizar nesta faculdade, orientações para alunos em Trabalhos de Conclusão de Graduação (TCC) e, também, gerando publicações científicas. Um projeto também foi desenvolvido junto ao Laboratório Hacker da Red Bull *Basement*. Este projeto junto a Red Bull gerou um protótipo de Cão-guia Robótico para auxiliar deficientes visuais. Sendo selecionado entre os 5 melhores projetos *Maker hacker* do Brasil no ano de 2019 (Figura 91).

A.1.3 Extensão

Com o conhecimento adquirido no doutorado, foi possível também realizar palestras, minicursos na área de robótica e competições de robótica em vários eventos pelo Brasil:(a) *The Developers Conference - (TDC)* ;(b) Semana da Computação - (SECOMP Unicamp) ; (c) Semana da Computação - (SECOMP Ufscar) ; (d) Semana da Engenharia de Computação - (SEnC - USP) ; (e) Feira do Livro - USP Riberão Preto ; (f) Jornada Científica do IFMT, (g) Semana de Tecnologia da Fatec Catanduva, (h) Semana da Engenharia - Unifafibe, (i) Semana da Engenharia Elétrica - Unilago, (j) Olimpíada Brasileira de Robótica - (OBR) e organização do (k) Primeiro Campeonato de Batalha de Robôs da Fatec Catanduva.

Estes trabalhos são voltados para uma contribuição com a sociedade e possibilitaram um retorno do investimento de recursos humanos investido nesta pesquisa.

Figura 91 – Projetos de extensão: Cão-guia Robótico - *Red Bull Basement*.

Fonte: Elaborada pelo autor.

A.2 Trabalhos Técnicos

A.2.1 Revisor de periódicos

Neste tópico foram realizadas revisões e pareceres de periódicos na área de Visão Computacional e Inteligência Artificial para as seguintes revistas, visando uma contribuição para trabalhos de outros pesquisadores:

- **(2016 - Atual) Periódico:** Atual Periódico: Revista IEEE América Latina;
- **(2017 - Atual) Periódico:** JOURNAL OF INTELLIGENT e ROBOTIC SYSTEMS;
- **(2018 - Atual) Periódico:** Sinergia - Revista Científica do Instituto Federal São Paulo;
- **(2017 - Atual) Periódico:** RA-L - IEEE Robotics and Automation Society;
- **(2018 - Atual) Periódico:** ISTA - Innovative Solutions for Arthroplasty;
- **(2016 - Atual) Periódico:** - Atual Periódico: ITS - Intelligent Transport System.

A.2.2 Revisor de trabalhos completos para eventos na área desta tese

- Latin American Robotics Symposium (LARS). 2016, 2017, 2018 e 2019;
- IEEE Intelligent Vehicles Symposium (IV-19). 2019;
- Hawaii International Conference on System Sciences (HICSS). 2018 e 2019;
- BRACIS 2019 | Brazilian Conference on Intelligent Systems (BRACIS). 2019;
- International Joint Conference on Artificial Intelligence (IJCAI). 2019;
- CoDIT 2019 | IEEE 2019 International Conference on Control, Decision and Information Technologies (CoDIT). 2019;
- WCCI 2018 : World Congress on Computational Intelligence - International Joint Conference on Neural Networks (IJCNN). 2018;
- International Conference on Robotics and Automation (ICRA). 2019 e 2018.
- IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2017 e 2018;
- The International Conference on Advanced Robotics (ICAR). 2019.


ICMC
USP
SÃO CARLOS