Uma contribuição ao desenvolvimento de sistemas
baseados em visão estéreo para o auxílio a
navegação de robôs móveis e veículos inteligentes

Leandro Carlos Fernandes

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura:

Uma contribuição ao desenvolvimento de sistemas
baseados em visão estéreo para o auxílio a navegação
de robôs móveis e veículos inteligentes

Leandro Carlos Fernandes

Orientador: Prof. Dr. Fernando Santos Osório

Co-orientador: Prof. Dr. Denis Fernando Wolf

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.
EXEMPLAR DE DEFESA.

USP – São Carlos

Agosto de 2014

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

F363c Fernandes, Leandro Carlos
 Uma contribuição ao desenvolvimento de sistemas
baseados em visão estéreo para o auxílio a navegação
de robôs móveis e veículos inteligentes / Leandro
Carlos Fernandes; orientador Fernando Santos Osório;
co-orientador Denis Fernando Wolf. -- São Carlos,
2014.
 78 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, 2014.

1. Calibração. 2. Visão estéreo. 3. Veículos
inteligentes. I. Osório, Fernando Santos, orient.
II. Wolf, Denis Fernando, co-orient. III. Título.

Resumo

ESTA tese visa apresentar uma contribuição ao desenvolvimento de sistemas computacionais, baseados principalmente em visão computacional, usados para o auxílio a navegação de robôs móveis e veículos inteligentes. Inicialmente, buscou-se apresentar uma proposição de uma arquitetura de um sistema computacional para veículos inteligente que permita a construção de sistemas que sirvam tanto para o apoio ao motorista, auxiliando-o em sua forma de condução, quanto para o controle autônomo, proporcionando maior segurança e autonomia do tráfego de veículos em meio urbano, em rodovias e inclusive no meio rural. Esta arquitetura vem sendo aperfeiçoada e validada junto as plataformas CaRINA I e CaRINA II (Carro Robótico Inteligente para Navegação Autônoma), que também foram alvo de desenvolvimentos e pesquisas junto a esta tese, permitindo também a experimentação prática dos conceitos propostos nesta tese. Neste contexto do desenvolvimento de veículos inteligentes e autônomos, o uso de sensores para a percepção 3D do ambiente possui um papel muito importante, permitindo o desvio de obstáculos e navegação autônoma, onde a adoção de sensores de menor custo tem sido buscada a fim de viabilizar aplicações comerciais. As câmeras estéreo são dispositivos que se enquadram nestes requisitos de custo e percepção 3D, destacando-se como sendo o foco da proposta de um novo método automático de calibração apresentado nesta tese. O método proposto permite estimar os parâmetros extrínsecos de um sistema de câmeras estéreo através de um processo evolutivo que considera apenas a coerência e a qualidade de alguns elementos do cenário quanto

ao mapa de profundidade. Esta proposta apresenta uma forma original de calibração que permite a um usuário, sem grandes conhecimentos sobre visão estéreo, ajustar o sistema de câmeras para novas configurações e necessidades. O sistema proposto foi testado com imagens reais, obtendo resultados bastante promissores, se comparado aos métodos tradicionais de calibração de câmeras estéreo que fazem uso de um processo iterativo de estimação dos parâmetros através da apresentação e uso de um padrão xadrez. Este método apresenta-se como uma abordagem promissora para realizar a fusão dos dados de câmeras e sensores, permitindo o ajuste das matrizes de transformação (parâmetros extrínsecos do sistema), a fim de obter uma referência única onde são representados e agrupados os dados vindos dos diferentes sensores.

Abstract

THIS thesis aims to provide a contribution to computer systems development based on computer vision used to aid the navigation of mobile robots and intelligent vehicles. Initially, we propose a computer system architecture for intelligent vehicles where intention is to support both the driver, helping him in driving way; and the autonomous control, providing greater security and autonomy of vehicular traffic in urban areas, on highways and even in rural areas. This architecture has been validated and improved with CaRINA I and CaRINA II platforms development, which were also subject of this thesis and allowed the practical experimentation of concepts proposed. In context of intelligent autonomous vehicles, the use of sensors that provides a 3D environment perception has a very important role to enable obstacle avoidance and autonomous navigation. Therefore the adoption of lower cost sensors have been sought in order to facilitate commercial applications. The stereo cameras are devices that fit these both requirements (cost and 3D perception), standing out as focus of the proposal for a new automatic calibration method presented in this thesis. The proposed method allows to estimate the extrinsic parameters of a stereo camera system through an evolutionary process that considers only the consistency and the quality of some elements of the scenario as to the depth map. This proposal presents a unique form of calibration that allows a user without much knowledge of stereo vision, adjust the camera system for new settings and needs. The system was tested with real images, obtaining very promising results as compared to traditional methods of calibration of stereo cameras that use an iterative process of parameter estimation through the presentation and use of a checkerboard pattern.

This method offers a promising approach to achieve the fusion of the data from cameras and sensors, allowing adjustment of transformation matrices (extrinsic system parameters) in order to obtain a single reference in which they are grouped together and represented the data from the different sensors.

Sumário

Resumo	i
Abstract	iii
Lista de Figuras	vii
Lista de Tabelas	ix
1 Introdução	1
1.1 Justificativa	4
1.2 Objetivos	5
1.3 Contribuições do trabalho	6
1.4 Organização sistematizada do texto	7
2 Veículos Autônomos e o Projeto CaRINA	9
2.1 Plataformas	15
2.2 Arquitetura	16
2.3 ROS - Robotic Operating System	18
2.4 Trabalhos em Percepção e Controle	20
2.4.1 Direção Assistida e Detecção de Obstáculos	20
2.4.2 Identificação de Zonas Navegáveis	23
2.4.3 CaRINA 1: Desenvolvimento da Plataforma e Arquitetura do Sistema	25
2.4.4 Sistema de Navegação Topológica	26
2.4.5 Projeto CaRINA: Arquitetura do Sistema e Aplicações	26
2.5 Considerações Finais	27

3	Visão Computacional: Geometria Projetiva, Calibração e Reconstrução 3D	31
3.1	Geometria projetiva	33
3.1.1	Transformações projetivas	34
3.1.2	Projeções no espaço 3D	37
3.2	Calibração de Câmeras	38
3.2.1	Parâmetros Intrínsecos	38
3.2.2	Parâmetros Extrínsecos	41
3.3	Imagens Estéreo	42
3.3.1	Geometria Epipolar	43
3.3.2	Matriz Fundamental	44
3.3.3	Retificação	46
3.3.4	Mapa de disparidade	46
3.4	Considerações Finais	49
4	Calibração Automática	51
4.1	Métodos tradicionais de Calibração	52
4.2	Visão estéreo em sistemas multi-baseline	52
4.3	Abordagem Evolutiva	55
4.4	Desenvolvimento	58
4.5	Experimentos e Resultados	60
5	Considerações finais	69
5.1	Publicações geradas	71
	Referências Bibliográficas	73

Lista de Figuras

2.1	Linha do tempo destacando projetos e plataformas experimentais na pesquisa de veículos inteligentes	10
2.2	Primeiras plataformas experimentais (semi-autônomas).	11
2.3	Veículos autônomos que competiram no Urban Challenge. <i>Os robôs Boss e Junior que venceram a competição, ficando em primeiro e segundo colocados. A direita a equipe que AnnieWAY cuja participação no evento deu início as pesquisas do Instituto de Tecnologia de Karlsruhe - KIT na área de veículos autônomos.</i>	13
2.4	Plataformas autônomas mais recentes	13
2.5	Plataformas robóticas desenvolvidas no âmbito do projeto CaRINA	16
2.6	Módulos conceituais e pontos de integração dos diversos elementos do sistema	17
2.7	Visão aérea do trajeto realizado durante os experimentos do sistema de direção assistida e detecção de obstáculos	21
2.8	Técnica usando fusão sensorial para detecção mais eficiente de obstáculos. .	22
2.9	Classificação de regiões navegáveis usando análise de atributos relativos a cor e textura em imagens	24
2.10	Sistema de percepção da plataforma CaRINA 1, representado pelos sensores dois laser frontais (paralelo ao solo e voltado para o solo), uma câmera monocular montada sobre o teto, e mais dois sensores lasers montados na traseira do veículo	25
2.11	CaRINA II - Plataforma com os Sensores e o Campo de Percepção.	27
3.1	Modelo de câmera <i>pinhole</i> e seus componentes geométricos.	32
3.2	Modelo geométrico de plano projetivo.	34
3.3	Distorções causadas por uma transformação <i>affine</i>	36
3.4	Modelo geométrico de uma câmera simplificada.	38

3.5	Tipos básicos de distorções causadas por lentes e consequências para a imagem	40
3.6	Relação entre os sistemas de coordenadas de mundo e da câmera no modelo geométrico projetivo.	42
3.7	Ilustração de um sistema de visão estéreo	43
3.8	Relação geométrica epipolar entre planos de imagem	44
3.9	Interpretação geométrica do processo de retificação de imagens	47
3.10	Exemplo de imagem de de profundidade obtida a partir de um par de imagens estéreo	48
4.1	Exemplo de regiões selecionadas pelo usuário com características de gradiente, homogeneidade e descontinuidade	57
4.3	Mapas de Disparidade utilizados pela função de fitness na avaliação da qualidade da transformação linear caracterizada por cada indivíduo.	63
4.4	Disparidades avaliadas pelo método de proposto após o ajustar do <i>baseline</i> para 36 cm.	64
4.5	Valores médios do fitness em populações com 10 e 30 indivíduos, mutação fixa em 1% e crossover em 80% e 90%	65
4.6	Valores médios do fitness em populações com 10 e 30 indivíduos, mutação fixa em 5% e crossover em 80% e 90%	66
4.7	Valores médios do fitness em populações com 10 e 30 indivíduos, mutação fixa em 8% e crossover em 80% e 90%	67

Lista de Tabelas

4.1	Parâmetros do sistema estéreo para uma câmera Bumblebee2 da Point Grey.	53
4.2	Comparação entre valores teóricos e experimentais na detecção de obstáculos	54
4.3	Configurações do sistema de visão estéreo utilizado nos experimentos . . .	60
4.4	Rodadas exploratória dos parâmetros do processo evolutivo. <i>Experimentos visavam analisar o comportamento do processo evolutivo assim como aspectos de variabilidade e convergência da população para diferentes combinações de parâmetros do AG. Em todas as rodas os testes consideraram as mesmas imagens, seleções de regiões e parâmetros intrínsecos e extrínsecos para o sistema de visão.</i>	61

Introdução

TALVEZ nenhuma outra tecnologia tenha transformado tão profundamente nosso cotidiano e estilo de vida neste último século quanto o automóvel. Servindo como um poderoso amplificador de nossas capacidades, esses veículos nos possibilitaram ir mais longe, transportar maiores volumes e fazer tudo isso de maneira mais rápida.

Desde a sua invenção, passaram de elementos de status social para substitutos legais dos antigos meios de transportes e agiram como poderosos agentes transformadores do ambiente em que vivemos. Ao longo deste tempo exerceram um significativo impacto no desenvolvimento do cenário urbano e se tornaram elementos fundamentais em nosso modelo econômico e de nosso estilo de vida.

Apesar das inúmeras contribuições que trouxeram para a sociedade, os automóveis também estão associados a custos humanos e materiais altos. Em dezembro de 2006 o IPEA¹ em conjunto com o DENATRAN² publicou um relatório analítico onde avaliou o impacto social e econômico resultante dos acidentes de trânsito nas rodovias brasileiras. Na ocasião foi reportado um custo total estimado em aproximadamente R\$ 6,5 bilhões computado apenas em relação aos acidentes ocorridos em rodovias federais (DENATRAN e IPEA, 2006).

¹Instituto de Pesquisa Econômica Aplicada

²Departamento Nacional de Trânsito

Esse montante totalizava custos ligados a atendimentos hospitalares, danos materiais dos veículos e ao local do acidente, além de despesas judiciais e de policiamento. Embora igualmente importantes, não considerou ainda os elementos de natureza não quantificável, tais como os custos com tratamentos relacionados a estresse pós-traumático dos envolvidos e danos ao meio ambiente.

De 2007 para 2008 o Brasil subiu da 9º para a 6º colocação no ranking dos países com mais automóveis e veículos comerciais leves, com um total de 2.671.316 veículos e que representava uma evolução de 14,98%, segundo o anuário publicado pela Fenabrave - Federação Nacional de Distribuição de Veículos Automotores (Fenabrave, 2008). Isso demonstra um crescimento significativo da frota de veículos em trânsito e um potencial agravamento dos problemas existentes.

Entendido como um caso de saúde pública há anos e reafirmada como uma área carente de medidas eficientes, o problema da segurança no transporte de pessoas e cargas é um tema importante e que demanda atenção principalmente pelos altos custos humanos envolvidos. Ações imediatas e de longo prazo tornam-se mandatórias diante do quadro que vivenciamos.

A indústria automobilística tem se empenhado para tornar seus veículos mais seguros, trabalhado em novas tecnologias e desenvolvendo sistemas de segurança como freios ABS, AirBags e Controle de tração, dentre outros. Infelizmente muitos destes sistemas não oferecem assistência pró-ativa e dependem fortemente das ações do motorista e da sua maneira de conduzir o veículo.

Atualmente os sistemas automobilísticos de segurança podem ser classificados em quatro diferentes categorias (Freyman, 2009):

- **Segurança Passiva (Passive Safety):** cujo objetivo é minimizar as consequências após um acidente ou impacto. Como exemplos podemos citar os cintos de segurança, os air-bags e dispositivos de pré-colisão.
- **Segurança Ativa (Active Safety):** responsáveis por uma assistência mais efetiva de condução tentam prevenir que o condutor cause acidentes ou bata em algum obstáculo. As principais aplicações são a detecção de pedestres e os alertas de colisão, que impedem o motorista de cometer erros por falta de atenção.
- **Assistência de Condução (Driver assistance):** que auxiliam o motorista em sua condução e sobre as situações do carro, emitindo alertas para saídas de pista, melhora da condução em condições de pouca visibilidade através de sistemas de

visão noturna, detecção de obstáculos em pontos cegos do condutor e assistência em manobras para estacionamento.

- **Condução Veicular (Vehicle Guidance):** que tenta evitar acidentes através do controle da direção, aceleração e frenagem do veículo, atuando de modo conjunto em paradas e retomadas em trechos urbanos, seguindo o veículo à frente com velocidade e distância apropriadas ou oferecendo controle automático dos movimentos longitudinais em pistas pouco sinuosas quando em rodovias.

Um dos motivos que fazem da robótica móvel uma área com grande potencial de contribuição nestas questões, é a ênfase de suas pesquisas na compreensão do ambiente, interpretação de contexto, deliberação e produção de ações de reposta.

Vale ressaltar que comportar-se de modo adequado em meio a um cenário urbano e de tráfego não implica somente em lidar com a aquisição de grandes volumes de dados, mas também na habilidade de reconhecer elementos importantes dentro deste contexto, como sinais, cruzamentos, veículos estacionados e em movimento; além de o fazê-lo de forma coerente e atendendo as restrições no tempo de resposta. A partir da exploração de técnicas de percepção e representação oriundas da robótica móvel é possível distinguir elementos estruturais bem como tratar e inferir a dinâmica de outros elementos participantes do tráfego. Assim a pesquisa e o aperfeiçoamento de técnicas para a percepção são de grande importância para a execução de tarefas ligadas ao desvio de obstáculos e a compreensão de elementos do ambiente, tendo caráter fundamental no desenvolvimento de sistemas de apoio a condução segura de veículos.

O trabalho desenvolvido nesta tese foi realizado buscando estudar, desenvolver e contribuir com a criação de sistemas que possibilitem o aperfeiçoamento de veículos inteligentes e autônomos provendo subsídios para uma condução mais segura.

Destaca-se aqui a importância da percepção do ambiente no qual está inserido o veículo e do uso destas informações no auxiliar da condução segura do veículo. No desenvolvimento desta tese foram realizados inicialmente trabalhos que permitiram o estudo, acompanhamento de trabalhos, desenvolvimento e implementação de diferentes métodos de percepção do ambiente em 2D e 3D (p.ex. sensores laser do tipo LIDAR (Fernandes *et al.*, 2010, 2012), visão monocular (Sales *et al.*, 2014, 2012; Shinzato *et al.*, 2010) e visão estéreo (Fernandes *et al.*, 2014; Klaser, 2014), além da atuação no projeto e desenvolvimento de veículos inteligentes, resultando em importantes contribuições no desenvolvimento do projeto CaRINA (abordado no capítulo 2).

Esta tese contribuiu para o amadurecimento da arquitetura adotada no veículo CaRINA I e II (Fernandes *et al.*, 2014), e mais em particular, através da proposta e validação

de um novo método para a calibração de câmeras estéreo, elemento este de grande importância para a percepção 3D do ambiente, dentro do contexto do desenvolvimento e evolução das pesquisas relativas as plataformas do projeto CaRINA.

1.1 Justificativa

Historicamente as principais aplicações de robôs móveis autônomos buscavam a exploração de ambientes inóspitos ou temas relativos ao desenvolvimento de sistemas para transporte automatizados. Uma realidade mais concreta e carente de atenção que também pode se beneficiar desta área de conhecimento é a questão do trânsito urbano e rodoviário. Veículos inteligentes e autônomos tem um enorme potencial de aplicação, seja em aplicações de transporte urbano e em rodovias, seja em aplicações agrícolas, industriais e de exploração. Estes veículos podem trazer diversas contribuições nestas áreas de aplicação: reduzindo riscos, aumentando a segurança, aumentando a produtividade e otimizando o uso de recursos.

Em nossa sociedade, cada vez mais dependemos dos veículos, sejam eles automóveis, ônibus ou até pulverizadores e colheitadeiras agrícolas. Por outro lado, as pessoas são muito suscetíveis a erros, distrações, imperícia ou incapacidade na operação de veículos (p.ex. conduzir sob efeito de medicamentos ou de bebida alcoólica) e mesmo devido a limitações de ordem pessoal (p.ex. problemas de visão, problemas de reflexos lentos devido a idade avançada, deficiências físicas e motoras). Uma condução inadequada de um veículo pode levar a acidentes graves, onde uma operação incorreta/indevida de tais veículos, usualmente acarreta em prejuízos e danos materiais e humanos.

Uma das maneiras de reduzir o número de acidentes é através de campanhas de conscientização e de educação. Uma outra forma, mais efetiva, é através da introdução de mecanismos e sistemas automatizados que simplifiquem e aumentem a segurança na condução de veículos, seja pela introdução de sistemas de apoio e alerta ao motorista, ou também, pela substituição do motorista por um sistema completamente automatizado. Esta é a principal justificativa para o desenvolvimento desta tese, a pesquisa e o desenvolvimento de sistemas e dispositivos que permitam aumentar a segurança na condução de veículos através do apoio ao motorista ou através da condução autônoma.

Neste contexto, um dos elementos de maior importância são os sensores, capazes de perceber os obstáculos e elementos do ambiente no qual está atuando o veículo. Existem diferentes tipos de sensores que podem ser adotados de acordo com as necessidades de cada aplicação, desde sensores baseados em imagens (2D) até os sensores de percepção em

3D, como por exemplo, sensores a laser do tipo LIDAR e câmeras estéreo. Os sensores com percepção em 3D são de grande importância, pois permitem uma melhor estimativa de distâncias e de posicionamento dos objetos e elementos do ambiente, provendo usualmente um mapa de profundidade ou uma nuvem de pontos (*3D point cloud*), conforme será apresentado e discutido posteriormente no 3.

Por outro lado, o uso de sensores de uma forma individualizada não permite uma grande robustez em relação a detecção de obstáculos e interpretação de uma cena. Em função disto, usualmente são utilizados múltiplos sensores e técnicas de fusão de sensores para compor uma representação mais completa e fiel da cena tratada. Esta fusão de sensores necessita de uma calibração precisa para estabelecer uma coerência entre os dados dos sensores, onde é usual realizar a transformação dos dados para um sistema de referência único onde são representados e agrupados os dados vindos dos diferentes sensores. O processo de calibração dos sensores e de transformação dos dados a fim de integrá-los é um item muito importante no desenvolvimento de sistemas baseados em sensores inteligentes, como os adotados em veículos inteligentes.

Existem na literatura diversos métodos propostos para a calibração e integração de dados de múltiplos sensores. Porém, destaca-se aqui, a importância do desenvolvimento de métodos que sejam de fácil uso, a fim de possibilitar sua utilização em aplicações práticas, sem a necessidade da participação de especialistas ou de impor restrições e/ou condições especiais de uso. Esta é uma outra importante justificativa para o desenvolvimento desta tese, que foi o de buscar o desenvolvimento de sistemas automatizados de calibração que possam ser realizados com um mínimo de intervenção e conhecimentos por parte dos usuários.

1.2 Objetivos

O principal objetivo desta tese é apresentar uma contribuição ao desenvolvimento de sistemas computacionais, baseados principalmente em visão computacional, usados para o auxílio a navegação de robôs móveis e veículos inteligentes.

Neste contexto, destaca-se a proposta de um novo método automático de calibração dos parâmetros extrínsecos de um sistema de câmeras estéreo. Através de um processo evolutivo que considera apenas a coerência e a qualidade de alguns elementos do cenário quanto ao mapa de profundidade, propomos uma forma original de calibração que permite a um usuário, sem grandes conhecimentos dos fundamentos e da geometria da visão estéreo, ajustar o sistema de câmeras para novas configurações e necessidades.

Este método apresenta-se como uma abordagem promissora para realizar a fusão dos dados de câmeras e sensores, permitindo o ajuste das matrizes de transformação (parâmetros extrínsecos do sistema), a fim de obter uma referência única onde são representados e agrupados os dados vindos dos diferentes sensores.

Como objetivo secundário adicional, busca-se apresentar uma proposição de uma arquitetura inteligente – com uma divisão apropriada de tarefas, elementos funcionais e características – que permitam a construção de um sistema que sirva tanto como apoio ao motorista, auxiliando-o em sua forma de condução; quanto para o controle autônomo, proporcionando maior segurança e autonomia do tráfego de veículos em meio urbano. Esta arquitetura vem sendo aperfeiçoada e validada junto as plataformas CaRINA I e CaRINA II, que também foram alvo desta pesquisa, permitindo o desenvolvimento e experimentação prático dos conceitos propostos nesta tese.

1.3 Contribuições do trabalho

A principal contribuição desta tese é a proposta, implementação e validação de um método de fácil uso para a calibração e ajuste dos parâmetros extrínsecos de câmeras estéreo. O método desenvolvido não depende do uso de padrões quadriculados, os denominados *checkerboards* (tradicionalmente usados em métodos de calibração estéreo), cujo uso é amplamente difundido, porém de difícil implantação em aplicações práticas junto a usuários leigos.

A lista das principais contribuições desta tese segue abaixo:

- Método para a calibração dos parâmetros extrínsecos de câmeras estéreo baseado em otimização por Algoritmos Evolutivos;
- Método para ajuste dos parâmetros extrínsecos de câmeras de um sistema multi-câmeras com diferentes *baselines*, sujeito a alterações da configuração (posicionamento) das câmeras, ou que necessite de uma recalibração sendo realizada de tempos em tempos;
- Proposta de uma metodologia de calibração e ajuste de parâmetros que pode ser aplicada a outros tipos de sensores, deste de que se estabeleça uma forma adequada de avaliação do ajuste dentro do novo contexto (definição de uma função adequada de *fitness*);

- Desenvolvimento de uma arquitetura de referência para veículos inteligentes e autônomos, que foi adotada junto aos veículos CaRINA I e II, e que vem sendo constantemente aperfeiçoada;
- Contribuição no projeto e desenvolvimento da automação do veículo CaRINA I;
- Contribuição no desenvolvimento dos sistemas computacionais dos veículos CaRINA I e II;
- Contribuição na implementação e desenvolvimento de diversos experimentos que foram realizados com as plataformas CaRINA I e II, incluindo atividades relacionadas ao uso de sensores laser (LIDAR) do tipo Sick e Velodyne, e de sistemas de aquisição e processamento de imagens (monoculares e estéreo);
- Preparação, submissão e publicação de diversos artigos Nacionais e Internacionais, que contribuíram de forma importante para a difusão das pesquisas desenvolvidas junto a esta tese, mas também relacionados aos trabalhos de pesquisa desenvolvidos junto ao LRM e junto ao projeto CaRINA;
- Participação em atividades ligadas aos projetos de pesquisa do INCT-SEC, CROB/SC e mais recentemente junto ao projeto S-TRUCK (este último é um projeto em parceria com uma empresa privada e que envolve sigilo na divulgação das pesquisas e resultados alcançados).

1.4 Organização sistematizada do texto

No capítulo 2 proporciona-se uma visão geral sobre veículos inteligentes e autônomos assim como sobre o projeto CaRINA ao qual este trabalho está diretamente relacionado. Além de relatos sobre a trajetória associada a esta pesquisa, também são apresentados alguns dos trabalhos desenvolvidos que se relacionam com as demais pesquisas conduzidas pelo grupo.

Um referencial matemático dos princípios e modelos geométricos relacionados a formação de imagens e de visão computacional são apresentados no capítulo 3. Fundamentos sobre espaços projetivos e a abordagem analítica que permite relacionar objetos tri-dimensionais com suas projeções bi-dimensionais são apresentados com o propósito de oferecer o suporte necessário as discussões ligadas ao problema referente a calibração automática de câmeras estéreo.

O método proposto nesta tese é descrito no Capítulo 4, onde é detalhado o processo desenvolvido visando a calibração automática dos parâmetros extrínsecos de uma câmera estéreo, baseado no uso de um processo Evolutivo (Algoritmos Genéticos). Neste capítulo também são apresentadas as configurações adotadas, são descritos os experimentos realizados e os resultados obtidos nos testes realizados com dados reais.

Por fim, o último capítulo apresenta as considerações finais e as conclusões relacionadas ao desenvolvimento e resultados obtidos nesta tese. Neste capítulo também são apresentados os trabalhos futuros e perspectivas de aplicação dos conceitos e métodos proposto nesta tese.

Veículos Autônomos e o Projeto CaRINA

EMBORA a robótica autônoma possa ser considerada uma área de pesquisa jovem e recente, a iniciativa para o desenvolvimento de veículos inteligentes ou autônomos teve seus primeiros esforços há cerca de três décadas. Uma linha de tempo representada na figura 2.1, pontua de modo cronológico alguns dos projetos mais importantes na área veículos autônomos e inteligentes.

Aplicações de caráter militar sempre tiveram motivações contundentes, como por exemplo, os altos custos humanos tanto pela inospitabilidade do ambiente quanto pelos traumas físicos e psicológicos peculiares a estas campanhas. Assim como em outros casos, a tecnologia por trás do desenvolvimento de um veículo autônomo também interessava as Forças Armadas.

Dando início ao que podemos dizer ser o primeiro grande investimento em um projeto relacionado a este tema, o programa ALV - *Autonomous Land Vehicle* intencionava trazer para o Exército tecnologias emergentes que pudessem servir no campo de batalha. Utilizando um sistema de navegação baseado em dados de um radar laser e num mapa de terreno, um veículo de grande porte (Figura 2.2(a)) apresentando mais de três metros de altura e oito rodas obteve relativo sucesso em atravessar um terreno escarpado enquanto evitava encostas, barrancos e pedras (Leighy, 1986).

Figura 2.1: Projetos e plataformas relacionados ao desenvolvimento de veículos inteligentes. *Abaixo da linha do tempo encontram-se indicados os grandes projetos temáticos da área, com destaque para seus períodos de duração (representados por barras horizontais coloridas). Logo acima, algumas das plataformas experimentais que se destacaram nestes projetos e tornaram-se referência.*

Pouco depois o primeiro sistema de navegação desenvolvido pelo grupo NavLab da universidade Carnegie Mellon tentava identificar ruas através da análise de cor e textura das imagens, classificando os pixels em navegáveis e não-navegáveis. Numa fase posterior, as bordas da rua eram detectadas estabelecendo-se o ponto de fuga e os parâmetros de orientação (Thorpe *et al.*, 1987).

Ao longo dos anos várias plataformas robóticas foram desenvolvidas pelo grupo (num total de 11) e que serviram para testes de diferentes sistemas, tais como o SCARF (Crisman e Thorpe, 1993), UNSCARF (Crisman e Thorpe, 1991) e o ALVINN. Este último, talvez o mais famoso, operava o NavLab-5 (Figura 2.2(b)) e consistia num sistema de controle baseado em rede neural. Nele uma rede MLP representava uma retina bi-dimensional de 30x32 que recebia as imagens da câmera do veículo e produziam 30 valores de saída identificando a direção que o veículo deveria seguir (Pomerleau, 1989).

Em vigor pelo período de 1987 à 1995, o projeto Prometheus - *PROgramMme for a European Traffic of Highest Efficiency and Unprecedented Safety* foi a maior iniciativa em pesquisas para o desenvolvimento de um veículo autônomo da época, recebendo um investimento financeiro cerca de um bilhão de dólares em termos de valores atuais (Williams, 1992).

Seu ponto alto deu-se em outubro de 1994, quando dois veículos robôs – um Daimler-Benz (VITA-2) e um Mercedes 500 SEL (VaMP) – com sistemas baseados em visão dirigiram por uma autoestrada de três pistas próximo ao aeroporto Charles-de-Gaulle em Paris, alcançando velocidades de até 130 Km/h.

Nesta demonstração o sistema foi capaz de rastrear outros veículos e realizar ultrapassagem de veículos mais lentos através da faixa da esquerda, após a manobra ser aprovada pelo piloto. O grupo de Ernst Dickmanns apresentou uma abordagem bastante inovadora ao incluir um movimento sacádico¹ artificial em seu sistema de visão computacional, permitindo que o sistema focasse sua atenção nos detalhes mais relevantes das imagens obtidas (Dickmanns *et al.*, 1994).

(a) ALV

(b) NavLab-5

(c) VITA-2

Figura 2.2: Primeiras plataformas experimentais (semi-autônomas).

O desenvolvimento de veículos autônomos é um tópico de pesquisa amplo e complexo, possibilitando abordagens diferentes e que podem ser vistas em alguns surveys como (Dickmanns, 1998) (Bertozzi *et al.*, 2000) (DeSouza e Kak, 2002) e (Bonin-Font *et al.*, 2008).

Grandes projetos temáticos e chamadas públicas foram (e ainda são!) capazes de reunir em torno de objetivos comuns os esforços da academia, grupos colaborativos de pesquisa, indústria e entidades governamentais, permitindo que ocorra um crescimento acelerado no avanço do conhecimento e amadurecimento das técnicas envolvidas.

Aproximadamente 18 anos após o ALV, a Agência Americana de Pesquisas Avançadas em Defesa - DARPA realiza uma chamada pública intitulada *The Grand Challenge Race* desafiando os participantes para projetar robôs capazes de percorrer grandes distâncias (de 240 à 336 km) em terreno desconhecido e acidentado de forma totalmente autônoma.

A percepção, análise e decisão deveriam ser totalmente automatizadas, sendo vetado ao veículo receber qualquer comando durante toda a corrida, exceto o de parada. A rota permitia certa flexibilidade desde que a trajetória passasse obrigatoriamente pelo conjunto de coordenadas de GPS dadas, que chegavam a apresentar variações entre 45 e 300 metros de distância umas das outras. Tanto o desafio quanto o prêmio de um milhão de dólares atraíram bastante atenção, totalizando 106 equipes inscritas.

¹Nome dado ao movimento ocular rápido cujo objetivo é posicionar a imagem sobre a fóvia, isto é, a área de maior precisão visual.

Em 2004, das 106 equipes inscritas apenas 25 grupos foram selecionados para a competição e o resultado foi, no mínimo, surpreendente. Nenhum time conseguiu concluir a tarefa! A melhor equipe alcançou cerca de 60 km/h em alguns momentos e percorreu aproximadamente 12 km dentro da rota, isto é, apenas 5% se tomarmos como referência o menor percurso que era possível.

No ano seguinte o desafio foi novamente lançado e os resultados mostraram que o desafio de fato não era fácil de ser sobrepujado (DARPA, 2005). Apenas 5 dentre os 23 competidores selecionados concluíram totalmente o percurso em meio ao deserto. A estratégia utilizada pela equipe Stanford, vencedora do desafio, foi criar um mapa 3D em tempo real do terreno a frente do robô, desviando de possíveis obstáculos e depressões. Também foram utilizadas câmeras de vídeo para identificar mudanças no padrão do terreno e, além dos sensores, foram utilizados seis microcomputadores dentro do carro para o processamento das informações e tomada de decisão (Thrun *et al.*, 2006).

Considerando que os resultados obtidos até então eram satisfatórios, mas não suficientes, o DARPA propõe uma nova edição para 2007 e sugere uma mudança de contexto (DARPA, 2007). O novo desafio consistia em trafegar de maneira autônoma por 97 km em meio a ambientes urbanos e os robôs deveriam ser capazes de interagir com outros veículos, identificar ruas e respeitar as regras de trânsito local.

O vencedor do desafio foi o “Boss” (Figura 2.3(a)), um veículo autônomo que foi capaz de dirigir de modo seguro com velocidade de até 48 km/h e dotado de um sistema de planejamento em três camadas que combina missão, comportamento e planeja as trajetórias para a navegação em meio ao ambiente urbano. Equipado com uma grande quantidade de sensores, sua instrumentação buscava atender as necessidades de cobertura e redundância sensorial, proporcionando ao Boss a capacidade de combinar informações (Urmson *et al.*, 2008), através de técnicas como a de fusão de sensores.

Desenvolvido por pesquisadores de Stanford, o “Junior” (Figura 2.3(b)) obteve a segunda colocação cruzando a linha de chegada 19 minutos após o Boss (Montemerlo *et al.*, 2008). Equipado com uma unidade inercial integrada com GPS, lasers posicionados a frente e nas laterais, além de radares de longo alcance instalados na grade dianteira, a detecção de obstáculos e a movimentação de outros veículos era feita principalmente com base na informações de um Velodyne HDL-64E montado no teto combinado com os outros lasers. Faixas e sinalizações horizontais eram determinadas por diferentes índices de reflexão do feixe infra-vermelho dos lasers.

Ao longo desse tempo os objetivos iniciais das pesquisas sofreram uma mudança natural e gradativa em seu foco, agregando as inovações tecnológicas mais recentes e direcionando a busca no desenvolvimento de sistemas de controle mais robustos e autônomos,

Figura 2.3: Veículos autônomos que competiram no Urban Challenge. *Os robôs Boss e Junior que venceram a competição, ficando em primeiro e segundo colocados. A direita a equipe que AnnieWAY cuja participação no evento deu início as pesquisas do Instituto de Tecnologia de Karlsruhe - KIT na área de veículos autônomos.*

compreensão mais detalhada do ambiente e de contexto, além de novas formas de interação homem-máquina (Thrun, 2010).

Entre as principais capacidades assistivas que seriam oferecidas pelos projetos estariam a habilidade de monitoramento de pontos cegos, alertas para mudanças involuntárias de pista, controle de aceleração e frenagem por um piloto automático adaptativo, além de assistência ao passar em cruzamentos ou durante manobras de estacionamento.

Dentre os grupos e iniciativas mais recentes, podemos destacar (AutoNOMOS, 2006) que em 2011 realizou a “*Mission: Brandenburger Tor*” trafegando seus dois veículos - o “Spirit of Berlin” e o “MIG” (*Made In Germany*) - de forma totalmente autônoma pelas ruas e rodovias de Berlin. E também o (VisLab, 2009) que continuou suas pesquisas após o “Argo” e criou cerca de mais sete protótipos. Dentre as plataformas mais recentes estão o “BRAiVE” e o “Porter”, sendo que este último participou do “*VIAC - Vislab Intercontinental Autonomous Challenge*”, em que dois “Porter”s percorreram uma rota de 13 mil Km da Itália para China organizados como comboio. (vide Figura 2.4)

Figura 2.4: Veículos autônomos experimentais utilizados nas pesquisas mais recentes.

Talvez o veículo mais famoso atualmente e considerado estado da arte é o “Google Self-Driving Car” (Google, 2009) que tomou como ponto de partida toda a tecnologia desenvolvida no DARPA Urban Challenge e tendo Chris Urmson como diretor chefe.

No âmbito nacional, as pesquisas relacionadas a esse tema são bastante recentes e consequentemente não muito numerosas. O crescente suporte financeiro oferecido por agências de fomento e iniciativas governamentais a projetos desta natureza evidenciam que prosseguir neste caminho para o domínio destas tecnologias é de forte interesse para o país.

Sob coordenação da Escola de Engenharia de São Carlos (EESC) o projeto Sistema Embarcado de Navegação Autônoma - SENA (SENA, 2009) foi proposto em 2009 com o objetivo desenvolver tecnologias de auxílio ao motorista e de navegação autônoma, com trabalhos relacionados a instrumentação e percepção do ambiente através de sensores lasers.

O projeto “Driving4u” concebido pelo Grupo de Automação e Tecnologia da Informação da UNIFEI (Goncalves *et al.*, 2010) é mais uma plataforma nacional de pesquisa em veículos inteligentes, mas que teve seus trabalhos descontinuados após a saída de alguns membros do grupo.

Na Universidade Federal de Minas Gerais (UFMG), o grupo de Pesquisa e Desenvolvimento de Veículos Autônomos - PDVA equipou e instrumentou um veículo Chevrolet Astra Sedan 2.0 agora batizado de CADU (Carro Autônomo Desenvolvido na UFMG) utilizado em pesquisas relacionadas a técnicas de sensoriamento, planejamento e controle para veículos inteligentes (de Lima, 2010).

No Centro de Tecnologia da Informação Renato Archer (CTI) sediado em Campinas-SP, o projeto VERO (Veículo Robótico) desenvolveu um veículo elétrico com capacidade de carga de 200 Kg que pode servir tanto para aplicações agrícolas quanto para ambientes urbanos (Bueno *et al.*, 2009).

O projeto CaRINA (CaRINA, 2009), cujo nome é acrônimo de CARRO ROBÓTICO INTELIGENTE PARA NAVEGAÇÃO AUTÔNOMA, têm seu início em 2009 estabelecendo como objetivo a investigação e o desenvolvimento de sistema de controle para veículo autônomos em meio a ambientes urbanos. Estabelecendo-se como uma das linhas centrais de pesquisas do Laboratório de Robótica Móvel - LRM, os trabalhos do grupo tem sido direcionados a investigação dos temas relacionados a compreensão de contexto, assistência ao motorista e controle do veículo, além do desenvolvimento de plataformas autônomas experimentais para uso em ambientes urbanos e agrícolas.

Esta tese foi desenvolvida junto ao LRM e teve os veículos autônomos desenvolvidos pelo grupo como plataformas de testes para seus experimentos.

2.1 Plataformas

O Laboratório de Robótica Móvel - LRM possui atualmente três plataformas de pesquisa e testes em veículos autônomos e inteligentes, dentre as quais as duas primeiras resultam de investimentos do INCT-SEC² e são descritas a seguir. A terceira, em convênio com a indústria, consiste em um veículo de grande porte usado no transporte de cargas e encontra-se ainda em fase de desenvolvimento.

CaRINA 1: trata-se de um veículo elétrico de pequeno porte que foi modificado para servir de plataforma robótica. Construído a partir de um modelo utilitário é capaz de locomover-se em ambientes abertos diversos, tais como condomínios e ambientes fabris. É impulsionado por um motor elétrico de 48 Volts e 10 HP, capaz de desenvolver velocidade máxima de 30 Km/h e transportar até 360Kg de carga útil (Figura 2.5(a)). Além de apresentar as características cinemáticas do modelo *Ackermann*, possui automatização dos seus controles de aceleração e esterçamento. Permite a combinação de vários sensores como base para os sistemas de percepção, dentre eles, lasers, receptor GPS, unidade inercial, bússola digital, encoder de roda e câmera de vídeo. Uma descrição detalhada da plataforma e sua arquitetura são apresentados em (Fernandes *et al.*, 2012).

CaRINA 2: desenvolvida a partir de um veículo comercial comum, o modelo Palio Adventure Locker foi modificado para que pudesse ter seus sistemas de aceleração, frenagem e esterçamento controlados via computador. Instrumentado com um conjunto bastante amplo de sensores, permitindo operar com diferentes configurações e ainda intercambiá-los com outras plataformas. Pode ser equipado com um sensor laser rotativo de 32 feixes (Velodyne HDL32), câmeras estéreo e 360° (Bumblebee e Ladybug 2 da Point Grey), unidade inercial (Xsens MTi-G) e receptor GPS com correção RTK (Septentrio AsteRX). Esta tem sido a principal plataforma de testes e a primeira em âmbito nacional a circular de maneira totalmente autônoma em meio as vias de uma cidade (Figura 2.5(b)).

Ambos os veículos são utilizados como plataforma de testes em experimentos de campo e condições reais para os diversos trabalhos desenvolvidos pelo grupo.

²Instituto Nacional de Ciência e Tecnologia em Sistemas Embarcados Críticos (<http://www.inct-sec.org>)

(a) CaRINA 1

(b) CaRINA 2: Fiat Palio Adventure modificado

Figura 2.5: Plataformas robóticas desenvolvidas no âmbito do projeto CaRINA. *Com menos de 1 ano entre os testes operacionais iniciais da segunda plataforma, o CaRINA 1 serviu entre outros propósitos para consolidar as soluções de automação e ampliar a compreensão dos problemas que seriam enfrentados no próximo estágio. O desenvolvimento do CaRINA 2 permitiu o avanço das pesquisas em controle, modelagem cinético-dinâmica e percepção, permitindo submeter toda a estrutura de software a testes em condições reais de tráfego e ambiente.*

2.2 Arquitetura

Considerando o objetivo secundário desta tese, quanto a estruturação e organização dos componentes necessários a um sistema de controle para veículos autônomos, parte de nossos esforços foram dedicados a proposição de uma arquitetura que conciliasse funcionalidades relativas a localização, planejamento, navegação e mapeamento aplicadas a um meio de características estruturadas e dinâmicas, próprios de ambientes urbanos.

Adotamos para a arquitetura uma organização hierárquica e distribuída, dada a necessidade de manipulação e tratamento concomitante das informações por diferentes módulos funcionais, além do balanceamento e distribuição da carga computacional em diferentes unidades de processamento (Figura 2.6).

Nesse contexto torna-se relevante a existência de elementos capazes de prover telemetria e suporte a depuração, que se caracterizavam como quesito importante não apenas para o desenvolvimento, mas também na operação do sistema fornecendo meios de rastre-

Figura 2.6: Módulos conceituais e pontos de integração dos diversos elementos do sistema. *As informações captadas pelos sensores (em azul) são processadas por diferentes módulos de percepção (em vermelho), responsáveis por realizar tarefas específicas de interpretação e tratamento dos objetos de interesse. Módulos de controle reativos (laranja) e comportamentais (em verde) são encarregados do controle em diversas atividades que vão desde o planejamento da rota até uma manobra de desvio. As ações sobre o veículo são realizadas através dos módulos de atuação (em roxo), que fazem a interface entre o sistema e o veículo propriamente dito.*

amento e identificação das decisões assumidas pelo sistema na assistência e/ou controle de condução do veículo.

Contendo componentes básicos para interpretação de dados, na divisão de tarefas e modularização do sistema concebeu-se a necessidade de módulos específicos dedicados a detecção de elementos estruturais e a identificação e rastreamento de obstáculos, a partir dos quais é possível determinar as áreas navegáveis e prover adicionalmente, para um nível hierárquico superior, meios para a identificação situações de risco que determinassem ações preventivas que devem ser tomadas.

Esta arquitetura (Fernandes *et al.*, 2014), assim como a definição dos componentes comportamentais e dos módulos funcionais, está em constante evolução e desenvolvimento. É preciso destacar que as plataformas do projeto CaRINA são protótipos de pesquisa, e como tal, tem sido usadas para explorar diferentes aspectos relativos a autonomia e a inteligência dos veículos, aplicadas também a diferentes cenários e situações. Isto faz com que muitas vezes sejam explorados individualmente certos problemas relacionados a robótica móvel, adotando e selecionando apenas partes da arquitetura proposta, ou até mesmo realizando modificações nesta a fim de validar soluções alternativas. A arquitetura

proposta serve como um "modelo de referência" no projeto e desenvolvimento de novos experimentos e aplicações. Além disto, sua construção e evolução ao longo dos anos, permitiu um amadurecimento dos conhecimentos e de ideias que levaram a proposta do uso de sensores baseados em lasers e câmeras estéreo, tal como vem sendo discutido e adotado nesta tese.

A organização dos elementos funcionais responsáveis pela identificação de áreas navegáveis, detecção de obstáculos e rastreamento de elementos dinâmicos do cenário tem sua forma de integração diretamente influenciada pelo modelo representativo do ambiente, assim como pelos níveis de acoplamentos e fusão sensorial que caracterizam uma camada de sensoriamento.

Inicialmente foi adotada no CaRINA 1 uma percepção do ambiente fortemente baseada em câmeras monoculares e sensores a laser de feixe único (Fernandes *et al.*, 2012) (Shinzato *et al.*, 2010) (Fernandes *et al.*, 2010). Posteriormente, houve uma migração visando uma adoção mais relacionada a sensores que permitam uma melhor descrição tridimensional do ambiente e de seus obstáculos e elementos. Passou-se a utilizar câmeras estéreo e o Velodyne (laser com 32 feixes adotado no CaRINA 2) (Fernandes *et al.*, 2014) (Klaser, 2014), assim como foi dada uma atenção especial aos problemas de calibração e fusão de sensores.

Desta maneira, a interpretação da arquitetura vem sofrendo variações de acordo com a implementação de cada trabalho. Com a evolução das pesquisas do grupo concernentes a percepção, temos migrado para formas de representação do mundo mais próximas ao modo como os humanos percebem o ambiente, ou seja, através de uma percepção tridimensional e da identificação dos elementos do ambiente visando a criação de modelos semânticos (os quais dependem fortemente desta percepção espacial dos elementos do ambiente). Podemos citar como exemplo desta identificação e análise estrutural e semântica dos elementos do ambiente, os trabalhos desenvolvidos junto ao LRM (Habermann *et al.*, 2013) (Hata *et al.*, 2013), que realizam a segmentação e reconhecimento (classificação) de diferentes estruturas identificadas junto as informações 3D providas pelos sensores do veículo.

2.3 ROS - Robotic Operating System

O ROBOTIC OPERATING SYSTEM (ROS) consiste numa plataforma de software concebida para operar como um meta-sistema operacional no desenvolvimento de aplicações em robótica, isto é, tem por objetivo oferecer funcionalidades e serviços relacionados a abs-

tração de hardware, controles de dispositivos, gerenciamento de pacotes, integração com outros *frameworks* de robótica, recursos e ferramentas para a visualização, gravação e depuração das informações, além dos principais componentes de sistemas distribuídos, como infraestrutura de comunicação, troca de mensagens entre processos, chamadas remotas de procedimentos (RPC), elementos de sincronismo, recursos para a definição e interpretação de estruturas customizadas para a troca de dados através de uma linguagem independente.

Seguindo a filosofia *Open Source* e tendo sua primeira versão disponibilizada em 2010, esta plataforma tem sido adotada mundialmente por laboratórios de pesquisa, universidades e outros grupos relacionados ao desenvolvimento em robótica, tornando-se progressivamente referência em aplicações na área e com uma grande comunidade de usuários. Inclusive alguns produtos comerciais, tais como sensores e plataformas robóticas, oferecem drivers e bibliotecas em versões compatíveis com os padrões do ROS.

Neste *framework* as aplicações são estruturadas segundo o conceito de grafo de computação, isto é, como uma rede *peer-to-peer* de processos que realizam suas tarefas de forma colaborativa e que são organizados em elementos fundamentais chamados nodos, tópicos e serviços.

Os nodos são programas unitários responsáveis por desempenhar alguma tarefa específica e que podem interagir com outros nodos através da troca de mensagens. Uma mensagem é uma estrutura de dados que encapsula a informação a ser intercambiada entre os elementos de processamento que compõem o sistema.

Os tópicos funcionam como canais de comunicação para um tipo específico de mensagem. Nele qualquer nodo interessado em receber a informação que ali trafega pode se inscrever realizando uma operação de *subscribe* e passa a ser notificado a cada chegada de novos dados. Por sua vez, os nodos que produzem uma informação podem publicá-la através de um tópico após se inscrever com uma operação de *advertise*. Esse mecanismo permite que as informações sejam distribuídas simultaneamente e numa relação de muito-para-muitos, mas envolvendo apenas os nodos relacionados.

Um modelo secundário de comunicação é provido através do conceito de serviços, onde um fluxo bi-direcional de informação é estabelecido em uma interação do tipo requisição/-resposta. Este difere-se do mecanismo funcional dos tópicos em três aspectos: apresenta uma relação entre nodos de um-para-um; opera com pares de mensagens, sempre gerando um fluxo de resposta para cada requisição realizada; e a condição espera bloqueada durante a execução de uma transação.

Para que seja possível tornar conhecido tópicos e serviços oferecidos pelo sistema, assim como os próprios nodos que estão ativos, a plataforma utiliza um elemento de

gerenciamento chamado Master. Seu papel é registrar e manter numa tabela de nomes quais são os nodos e canais de comunicação que formam o grafo de computação, de modo que cada elemento do sistema possa descobrir quais são seus *peers* e estabelecer comunicação diretamente com eles.

2.4 Trabalhos em Percepção e Controle

Durante o desenvolvimento desta tese foram realizados diversos trabalhos, os quais tiveram por objetivo o desenvolvimento e aperfeiçoamento da arquitetura do sistema e das plataformas robóticas autônomas CaRINA 1 e 2, bem como, visavam um aprimoramento das técnicas de percepção que pudessem contribuir para o auxílio a navegação de robôs móveis e veículos inteligentes. Estes trabalhos resultaram em trabalhos publicados em conferências e periódicos, onde serão descritas brevemente estas publicações e as trajetórias de pesquisas e desenvolvimentos que as mesmas representam.

2.4.1 Direção Assistida e Detecção de Obstáculos

Um dos quesitos fundamentais para a navegação autônoma, o problema da detecção de obstáculos foi endereçado neste primeiro trabalho explorando a capacidade de detecção do sensor laser (Sick). Dadas as características do campo de varredura produzido pelo sensor, sua localização junto ao para-choque frontal tornava possível a detecção de objetos posicionados tanto a frente do veículo quanto em suas laterais.

Inicialmente, o veículo foi conduzido pelo caminho a ser percorrido, definindo uma trajetória em termos de coordenadas GPS (figura 2.7). Considerando que o veículo fosse percorrer novamente esta mesma trajetória, porém agora fazendo uso de um sensor laser para identificar os obstáculos que pudessem estar no seu caminho, e consequentemente representando um risco a sua navegação. O objetivo deste trabalho era gerar alertas ao condutor, informando sobre a presença de obstáculos a frente do veículo.

Considerando as características e dimensões do campo radial produzido pelo sensor laser Sick de um único feixe, uma região retangular correspondente a um recorte da área de percepção do dispositivo foi utilizada para definir uma zona de atenção. Esta zona de atenção se encontra definida dentro de uma seção da área frontal de percepção do sensor, representando uma parte da trajetória que deverá vir a ser percorrida pelo veículo. Desta maneira era possível uma diferenciação dentre a presença dos inúmeros elementos do ambiente, que nem sequer poderiam ser considerados como potenciais obstáculos, daqueles dos quais se diriam ser potencialmente perigosos.

Figura 2.7: Visão aérea do local onde foram realizados os experimentos do sistema de direção assistida e detecção de obstáculos nas dependências do Campus 1 da USP em São Carlos. *Destaque para a reconstrução da trajetória realizada a partir das coordenadas GPS registradas durante a coleta de dados. Representados em azul, o percurso de ida; em vermelho, o retorno em frente a portaria principal; em verde, o percurso de volta; e a manobra final em roxo.*

Entretanto, usando apenas a informação do sensor laser e da região retangular definindo uma zona de atenção, era gerado um número excessivo de falsos alertas, decorrentes da detecção de múltiplos elementos que estavam localizados dentro da área de recorte, mas posicionados fora da trajetória prevista a ser seguida pelo veículo. A falta da definição de uma zona de atenção mais apurada, por onde realmente o veículo iria passar, não era capaz de impedir a ocorrência de muitos falsos alertas (figura 2.8), onde em uma curva, os obstáculos a frente do veículo eram considerados como potencialmente perigosos, apesar da trajetória do veículo indicar que o mesmo iria virar na curva e acabar desviando dos obstáculos mais a sua frente.

(a) Imagem capturada pela câmera de vídeo durante uma coleta de dados. Nela é possível notar a presença de árvores em ambas laterais da rua; à esquerda, a parede de um dos prédios daquela região, e ao fundo, a área de estacionamento dos blocos e uma grande lixeira amarela de metal, distante cerca de 30m e bem a frente do veículo.

(b) Os *waypoints*, coletados anteriormente e indicados na figura pelos pontos em roxo, definem a trajetória que o veículo deve realizar. Em vermelho, temos algumas árvores e o prédio sendo classificados como obstáculos quando utilizamos o sistema com uma região fixa de atenção.

Figura 2.8: Técnica usando fusão sensorial para detecção mais eficiente de obstáculos.

Combinando os dados sobre o posicionamento do veículo fornecidos pelo GPS e de orientação pela bússola, juntamente com o conjunto de *waypoints* obtidos previamente, foi possível identificar qual a trajetória que deveria ser seguida pelo veículo e assim ajustar a zona de atenção para uma cobertura mais adequada. Essa região podia se adaptar diante das diferentes situações, como por exemplo em uma curva acentuada, onde os limites da zona de atenção eram ajustados de acordo com o *waypoint*, considerando o ponto mais distante da trajetória e que permanecesse contido dentro dos limites absolutos da área de percepção. Desta maneira tornava-se possível ainda um novo nível de alerta ao condutor, diferenciando entre obstáculos considerados apenas como potenciais elementos de atenção daqueles realmente perigosos.

Uma descrição mais completa dos experimentos e dos resultados obtidos neste trabalho são apresentados em (Fernandes *et al.*, 2010). Vale ressaltar ainda que neste trabalho já temos uma associação entre os dados de GPS, bússola e laser, demonstrando a importância e relevância da fusão de sensores, do tratamento e da análise de dados de percepção no contexto desta tese e das pesquisas envolvendo veículos inteligentes.

2.4.2 Identificação de Zonas Navegáveis

A identificação de áreas navegáveis é uma abordagem bastante adotada em aplicações de robótica móvel, diferenciando-se em relação aos tipos de sensores usados para identificar uma região navegável de outra não navegável. Uma abordagem que foi desenvolvida inicialmente por Shinzato (Shinzato e Wolf, 2011) (Shinzato *et al.*, 2010) realiza a classificação de regiões nas categorias navegável e não-navegável, através da análise de atributos associados a cor e textura em imagens captadas por uma câmera de vídeo monocular.

Nesta abordagem, as imagens adquiridas por uma câmera de vídeo colorida eram processadas para a extração de características de cor e textura, que então eram encaminhadas para uma rede neural para a classificação da região correspondente. Empregando um método de classificação por blocos, os quadros de 320×240 eram reticulados formando uma matriz bidimensional de blocos de 10×10 pixels. Cada um dos 768 conjuntos que representavam a imagem era processado para a extração de uma lista de atributos associados as características de cor e textura daquela coleção de pixels.

Bloco após bloco, seus atributos eram apresentados como entradas para uma rede neural do tipo MLP que os classificava em uma destas três categorias: navegável, não-navegável ou incerto. Cenários variados ou que incluíam capturas em diferentes condições de iluminação (Figuras 2.9(a), 2.9(d) e 2.9(g)) foram utilizados nos experimentos para a validação da metodologia proposta. Uma primeira versão deste trabalho foi testada utilizando um robô móvel Pioneer, onde foi publicado um trabalho em co-autoria junto com Shinzato (Shinzato *et al.*, 2010).

Inicialmente todos os blocos de um quadro, com exceção daqueles localizados as margens da imagem, eram utilizados no processo de treinamento e as respostas da rede eram avaliadas segundo o que era esperado para os demais quadros do vídeo. Nas figuras 2.9(b), 2.9(e) e 2.9(h) pode-se observar as respostas dada pela rede ao avaliar aquele cenário, enquanto as figuras 2.9(c), 2.9(f) e 2.9(i) evidenciam quais blocos foram incorretamente classificados, diferenciando-os entre falsos-positivos de falsos-negativos.

A necessidade de um treinamento supervisionado representava um fator complicador para o tipo de aplicação a que se pretendia. Através da escolha de regiões parciais da imagem, cuja classificação asseguradamente representariam exemplos das duas categorias, o processo de treinamento pôde ser automatizado. Ademais, para uma avaliação mais apurada da precisão das respostas da rede, uma análise desconsiderando os dados da região acima da linha do horizonte também foi feita.

De um modo geral, os mais altos índices de acerto resultaram da combinação dos atributos de média e entropia relativos aos canais RGB ou de uma composição destes valores

(a) Cenário 1: Embora apresentando uma aparente uniformidade, é possível notar a presença de folhas em meio ao caminho.

(b) Respostas bastante satisfatórias, produzindo uma área navegável bem delimitada.

(c) Poucos casos de falsos-positivos e que, por estarem localizados acima da linha do horizonte, não representam risco a navegação.

(d) Cenário 2: Calcamento construído com mosaico de tijolos conferindo padrões de textura à imagem.

(e) Boa distinção das áreas navegáveis, com destaque para a correta classificação do edifício ao fundo.

(f) Presença de falsos-negativos em regiões de fronteira. Formação de um pequeno cluster na porção inferior da imagem.

(g) Cenário 3: A mesma região utilizada no primeiro cenário, entretanto com condições bastante diferentes de iluminação.

(h) Mesmo com a presença de sombras projetadas no chão, uma boa diferenciação é obtida pela rede.

(i) Mantém os mesmos resultados do cenário 1, acrescentando na porção central as consequências dos fatores de iluminação.

Figura 2.9: A esquerda, os cenários utilizados para a detecção de regiões navegáveis através da análise de características de cor e textura. Ao centro, as regiões classificadas pela RNA como *navegáveis*, *não-navegáveis* e *indefinido*, representadas respectivamente em ciano, magenta e amarelo. Por fim, na coluna da direita, a avaliação dos resultados da rede, evidenciando os acertos (em verde), falsos-positivos (vermelho) e falsos-negativos (azul).

com elementos do espaço HSV. Uma discussão mais detalhada destes resultados podem ser encontrada em (Shinzato *et al.*, 2010). Este trabalho, apesar de relevante, demonstrou algumas das limitações em se trabalhar com imagens 2D, onde há a necessidade de um treinamento prévio e uma supervisão para a definição dos elementos que representam uma zona navegável (plana e sem obstáculos) e dos que representam uma zona não-navegável.

2.4.3 CaRINA 1: Desenvolvimento da Plataforma e Arquitetura do Sistema

Durante o desenvolvimento desta tese, uma parte importante do trabalho foi dedicada ao desenvolvimento da Plataforma CaRINA 1. Os trabalhos desenvolvidos tornaram possível a utilização prática e operacionalização do veículo autônomo CaRINA 1. Esta plataforma permitiu um importante aprendizado e a realização de diversos testes e implementações.

Figura 2.10: Sistema de percepção da plataforma CaRINA 1, representado pelos sensores dois laser frontais (paralelo ao solo e voltado para o solo), uma câmera monocular montada sobre o teto, e mais dois sensores lasers montados na traseira do veículo

Como resultado destes trabalhos, foi publicado um artigo (Fernandes *et al.*, 2012), juntamente com outros pesquisadores que participaram do desenvolvimento desta plataforma, onde é detalhada a Plataforma CaRINA 1, incluindo o sistema de percepção do veículo autônomo (figura 2.9(i)), os dispositivos de controle e atuação, e onde também é apresentada uma visão geral da arquitetura inicial que foi proposta no contexto do projeto CaRINA. Esta arquitetura foi a base para os trabalhos futuros e aperfeiçoamentos da mesma.

Neste trabalho também percebe-se uma preocupação para com a detecção de áreas navegáveis através de uma percepção baseada em câmera monocular e sensores laser de um único feixe. Estes dispositivos fornecem uma percepção restrita das características espaciais do ambiente, e já neste trabalho é discutido sobre o uso de câmeras estéreo para o auxílio a navegação e detecção de obstáculos.

Cabe destacar aqui algumas contribuições pontuais que fizeram parte deste trabalho, relacionadas a integração de *hardware e software* dos sistemas embarcados do veículo, como por exemplo, o desenvolvimento do sistema de controle de aceleração do CaRINA 1, o desenvolvimento das rotinas de interface com o controle do motor usado no esterçamento do veículo (RoboteQ), e inclusive *drivers* de dispositivo (p.ex. o *driver* inicialmente usado na aquisição de dados do Velodyne). Estes trabalhos, apesar de pontuais e de cunho mais técnico, tiveram uma importante contribuição para que fosse possível a operação do CaRINA 1 em modo autônomo e para o desenvolvimento de vários trabalhos de pesquisa.

2.4.4 Sistema de Navegação Topológica

A participação em trabalhos mais recentes, em parceria com D.Sales (Sales *et al.*, 2012) (Sales *et al.*, 2014), permitiu um aperfeiçoamento no desenvolvimento e uso da plataforma CaRINA 2. Estes trabalhos também permitiram evoluir a arquitetura proposta do sistema, bem como foram trabalhadas questões referentes à aquisição de imagens, navegação autônoma e percepção 2D e 3D.

Estes trabalhos visavam a navegação através do uso de Redes Neurais Artificiais e Autômatos Finitos, realizando o reconhecimento de “marcos topológicos”, ou seja, cruzamentos, curvas a esquerda e a direita, desvios, etc. O trabalho foi inicialmente desenvolvido com o uso de robôs móveis do tipo Pioneer P3AT (com o uso de sensor 3D do tipo Kinect, cujos dados são bastante similares aos obtidos, sob certas condições, por câmeras estéreo), sendo depois adaptado para o CaRINA 2, onde foram capturadas imagens a partir de uma câmera estéreo, porém fazendo uso de imagens 2D.

2.4.5 Projeto CaRINA: Arquitetura do Sistema e Aplicações

O artigo (Fernandes *et al.*, 2014) é um dos trabalhos mais abrangentes e completos que descrevem a arquitetura do CaRINA, as plataformas CaRINA 1 e 2, e diversos trabalhos que foram realizados ao longo dos 4 últimos anos (2010-2014) visando o desenvolvimento de sistemas inteligentes e de navegação autônoma para estes veículos.

Figura 2.11: CaRINA II - Plataforma com os Sensores e o Campo de Percepção.

Este artigo permite ter uma visão da evolução destas plataformas, onde pode-se constatar que recentemente houve um direcionamento dos trabalhos para o uso de sensores com percepção 3D (p.ex. Velodyne, câmera estéreo), onde a fusão de sensores também tem um papel fundamental. A figura 2.11 apresenta os sensores e o campo de percepção do CaRINA II, mostrando esta característica de percepção 3D e de sobreposição do campo de percepção entre os diferentes sensores, o que indica claramente a possibilidade de se realizar uma fusão de sensores para uma melhor aquisição de dados e representação do ambiente e de seus elementos.

Não cabe aqui reproduzir o conteúdo do referido artigo (Fernandes *et al.*, 2014), uma vez que o mesmo já foi publicado e está disponível para acesso, por outro lado, é recomendada a sua leitura de modo a complementar o presente documento. Sendo assim, este texto irá focar na apresentação de resultados de pesquisa mais recentes, e que não constam do referido artigo. Além disto, nos capítulos seguintes será dada uma ênfase as contribuições de pesquisa que possuem um caráter mais pessoal, sendo trabalhos desenvolvidos de forma individual, ao contrário de muitos dos demais trabalhos citados anteriormente que representam projetos e pesquisas realizadas em conjunto com outros colegas e pesquisadores do laboratório LRM do ICMC/USP.

2.5 Considerações Finais

Este capítulo apresentou as plataformas de veículos autônomos, desenvolvidas no exterior e no Brasil, e sua evolução. Foi dado um destaque ao projeto CaRINA que vem sendo desenvolvido junto ao Laboratório LRM do ICMC/USP, onde foram abordadas as plataformas CaRINA 1 e 2, que foram alvo desta pesquisa.

Os trabalhos de pesquisa e desenvolvimento relacionados a esta tese, resultaram na proposta e desenvolvimento de uma arquitetura do sistema computacional dos veículos autônomos CaRINA 1 e 2. Esta arquitetura de referência está em constante desenvolvimento, onde foram publicados trabalhos que descrevem tanto a arquitetura como aplicações e pesquisas realizadas com estas plataformas. Neste capítulo também foi apresentado um resumo dos trabalhos desenvolvidos e publicados durante o desenvolvimento da tese. Estes trabalhos e estudos levaram a orientar as nossas pesquisas em direção ao uso de sensores com percepção 3D, com um maior destaque para aqueles de menor custo, como as câmeras estéreo, e além disto, também constatou-se uma tendência em relação a fusão de sensores. No caso de sensores como o Velodyne, devido ao fato que este provê uma informação muito esparsa (composta por apenas 32 feixes que varrem o espaço), constata-se também a necessidade de “fundir” as informações deste sensor com informações provenientes de outros sensores.

Em função de uma forte interação com empresas em projetos de pesquisa realizados junto ao LRM-USP, também foi constatada uma demanda por projetos baseados em visão computacional, dadas suas características de baixo custo e a possibilidade de percepção espacial em sistemas de visão estéreo. Destaca-se neste contexto um projeto que foi desenvolvido em parceria com a empresa JACTO S.A. que produz implementos agrícolas. Esta empresa desenvolveu um pulverizador de defensivos agrícolas totalmente autônomo (JAV2 - *Jacto Autonomous Vehicle*), onde o desenvolvimento do sistema de percepção foi realizado em parceria com os pesquisadores do LRM-USP. O sistema de detecção de obstáculos do JAV2 da JACTO é baseado em um sistema de visão estéreo. Mais recentemente, outra importante empresa da área automobilística também estabeleceu um projeto de colaboração com a USP e o LRM, onde têm sido realizados estudos sobre o uso de câmeras estéreo.

Estes projetos tem abordado questões referentes a configuração e ajuste dos parâmetros que definem o par de câmeras estéreo, uma vez que dependendo da montagem das mesmas e das características dos equipamentos, podemos ter diferentes ajustes e resultados em relação a capacidade de percepção destes. Os experimentos realizados, por diferentes membros do nosso grupo, usando câmeras estéreo, permitiram identificar que um dos problemas referentes a este tipo de dispositivo está relacionado ao processo de calibração, seja devido a alterações da configuração do par de câmeras (p.ex. alteração da distância entre as duas câmeras, denominado de *baseline*), ou até mesmo devido a perda da calibração, indicando a necessidade da realização de uma nova calibração.

Ademais, o processo de calibração das câmeras estéreo tende a ser trabalhoso e depende de um certo conhecimento prévio para que o mesmo seja realizado adequadamente. Em

função disto, e da demanda de desenvolvimento de um método de calibração que fosse mais adequado para aplicações comerciais e usuários leigos, foi desenvolvida uma pesquisa que resultou na proposta de um novo método de calibração de câmeras estéreo, a ser detalhado nos capítulos seguintes.

Visão Computacional: Geometria Projetiva, Calibração e Reconstrução 3D

CÂMERAS digitais são em sua essência sensores que captam as características visuais dos objetos de um cenário tridimensional e as representam na forma de imagens bidimensionais.

Embora a maioria das câmeras se constituam como um sistema complexo, compreendendo partes interdependentes e elementos de influência direta na forma e qualidade da imagem obtida (tais como as lentes e propriedades do elemento de captura), o processo de formação de uma imagem pode ser satisfatoriamente descrito pelo modelo mais simplificado de câmera, chamado de estenopeica ou modelo *pinhole*.

Neste tipo de câmera, ilustrado na figura 3.1(a), os feixes que partem de fontes de luz e incidem nos objetos tridimensionais alcançam a câmera passando por um pequeno furo (*pinhole*) e atingem a parede oposta da caixa escura, onde a imagem é então formada. Desta forma, o modelo generalizado de uma câmera é fundamentado em termos de três elementos: o centro de projeção, pelo qual passam todos os raios de luz; o ponto focal,

que determina a proporção dos objetos e também o campo de visão da imagem; e o plano de projeção, onde a imagem é formada.

Assumimos o centro de projeção C sendo a origem do sistema de coordenadas em relação a câmera e o eixo óptico orientado de forma coincidente com o eixo Z_c . A distância entre o centro de projeção e o plano onde será formada a imagem é chamada de distância focal (f) e determina o ponto em que ocorre a secção do tronco de projeção. Deste corte temos a formação do plano da imagem Π , que é orientado de maneira perpendicular ao eixo óptico e é interceptado por ele no chamado ponto principal P (figura 3.1(b)).

Para que a imagem não se forme de maneira invertida, no modelo o plano de projeção é convenientemente posicionado entre o centro de projeção e os objetos da cena.

Figura 3.1: Modelo de câmera *pinhole* e seus componentes geométricos. O raio que parte do objeto $\mathbf{X}$ no espaço em direção ao centro de projeção intercepta o plano da imagem no ponto $\mathbf{x}$ de coordenadas x_c e y_c .

Através de um processo de calibração estabelece-se a correspondência entre os pontos conhecidos na cena (3D) e na imagem (2D), determinando a função que mapeia as relações projetivas entre esses elementos em termos do conjunto de características próprias da câmera, conhecidos como parâmetros intrínsecos e extrínsecos.

Os parâmetros intrínsecos correspondem as características construtivas e particulares de cada câmera, tais como a distância focal, o tamanho do pixel e as distorções de lente. Por equivalerem a propriedades internas e ligadas a natureza dos componentes, esses parâmetros são invariantes quanto ao ambiente observado e independem do posicionamento da câmera em relação a ele.

Por sua vez são denominados por parâmetros extrínsecos aqueles associados a orientação e o posicionamento da câmera em relação a um sistema de referência. A imagem

capturada dependerá diretamente dos valores assumidos uma vez que estes determinam a perspectiva dos objetos da cena.

3.1 Geometria projetiva

Na geometria plana define-se um ponto como um par ordenado (x, y) em $\mathbb{R}^2$ que é dado em relação a um sistema de referência. Se considerarmos o plano como um espaço vetorial euclidiano, as coordenadas do ponto passam a ser interpretadas como um único vetor dentro desse espaço que é dado por $\mathbf{x} = [x \ y]$.

De modo similar, uma reta que é expressa pela equação geral $ax + by + c = 0$ e cujos coeficientes a , b e c produzem diferentes retas, será representada na forma vetorial por $\mathbf{l} = [a \ b \ c]^T$. A relação entre retas e vetores não é única e independe de escala. Para qualquer constante $k \neq 0$ a reta $(ka)x + (kb)y + (kc) = 0$ é equivalente a $ax + by + c = 0$, embora representadas por vetores diferentes $[ka \ kb \ kc]^T$ e $[a \ b \ c]^T$.

Observe que essa propriedade é compatível com a definição de vetores equivalentes, onde um vetor particular $\mathbf{l}$ na verdade representa uma classe de vetores em $\mathbb{R}^3$ e que define um conjunto de vetores homogêneos. O conjunto de classes equivalentes de vetores em $\mathbb{R}^3$ com exceção de $[0 \ 0 \ 0]^T$ formam o *espaço projetivo* $\mathbb{P}^2$.

Uma reta $\mathbf{l} = [a \ b \ c]^T$ passa por um ponto $\mathbf{x} = [x \ y]^T$ se ambos satisfazem a equação $ax + by + c = 0$, o que equivale a escrevê-la como o produto interno dos vetores $[x \ y \ 1][a \ b \ c]^T = [x \ y \ 1]\mathbf{l} = 0$. Observe que o vetor tridimensional $\tilde{\mathbf{x}} = [x \ y \ 1]^T$ é uma maneira conveniente e não restritiva de representar o ponto (x, y) dado em $\mathbb{R}^2$, haja vista que para qualquer constante k não nula e linha $\mathbf{l}$ a equação $[kx \ ky \ k]\mathbf{l} = 0$ será verdadeira se e somente se $[x \ y \ 1]\mathbf{l} = 0$.

Intuitivamente torna-se natural considerar que o conjunto de vetores $[kx \ ky \ k]^T$, para $k \neq 0$, representam um ponto em particular em $\mathbb{R}^2$ expresso em termos de coordenadas homogêneas e cujas coordenadas originais são dadas por $(kx/k, ky/k)^T$. Assim os pontos dados na forma de vetores homogêneos também são elementos de $\mathbb{P}^2$.

Uma das conveniências desta representação está em permitir uma abordagem algébrica para se determinar o relacionamento entre pontos, retas e planos. Por exemplo, a intersecção de duas retas, $\mathbf{l}$ e $\mathbf{l}'$, corresponde a um ponto $\tilde{\mathbf{x}}$ que é determinado pelo produto cruzado ou vetorial, isto é $\tilde{\mathbf{x}} = \mathbf{l} \times \mathbf{l}'$. Seguindo o mesmo raciocínio, temos que a reta que une dois pontos é dada por $\mathbf{l} = \tilde{\mathbf{x}} \times \tilde{\mathbf{x}}'$.

Adicionalmente, o sistema de coordenadas homogêneas permite ainda que pontos e retas no infinito admitam uma representação finita em $\mathbb{R}^3$. Suponha duas retas paralelas

$ax + by + c = 0$ e $ax + by + c' = 0$ representadas respectivamente pelos vetores $\mathbf{l} = [a \ b \ c]^T$ e $\mathbf{l}' = [a \ b \ c']^T$ e de tal forma que diferem apenas pela última coordenada. Sua interseção é dada por $\mathbf{l} \times \mathbf{l}' = (c' - c)[b \ -a \ 0]^T$ e que se ignorarmos o fator de escala, corresponderá ao ponto $(b/0, -a/0)$ em termos de coordenadas não homogêneas. Embora pareça incoerente, este resultado sugere que o ponto tem coordenadas com valores infinitamente grandes e que as retas paralelas se encontram no infinito.

Assim a adoção de uma representação por coordenadas homogêneas amplia o espaço euclidiano transformando-o num novo objeto geométrico, que é o conjunto de todos os vetores homogêneos tridimensionais e chamado de espaço projetivo $\mathbb{P}^2$. Em $\mathbb{P}^2$ todos os pontos no infinito possuem a forma $[x_1 \ x_2 \ 0]^T$ e são conhecidos como pontos ideais, definidos apenas pela proporção $x_1 : x_2$. Denotados por $\tilde{x}_\infty$, representam uma direção de reta e correspondem a pontos no plano x_1x_2 em $\mathbb{R}^3$.

Podemos entender $\mathbb{P}^2$ como sendo o conjunto de todos os raios que passam pela origem do sistema. Cada ponto $\tilde{\mathbf{x}}$ em $\mathbb{R}^3$ projeta sua imagem $\mathbf{x}$ no plano Π , e por sua vez, a linha que intercepta dois pontos em Π representa um plano em $\mathbb{R}^3$ que também atravessa a origem (figura 3.2).

Figura 3.2: Modelo de plano projetivo. Os pontos e linhas em $\mathbb{P}^2$ são representados em $\mathbb{R}^3$, respectivamente, por raios e planos que passam pela origem.

3.1.1 Transformações projetivas

Uma transformação linear no espaço euclidiano $\mathbb{R}^n$ é representada por uma matriz que deve multiplicar as coordenadas de cada ponto. Da mesma maneira, uma transformação

no espaço projetivo $\mathbb{P}^n$ é um mapeamento das coordenadas homogêneas que representam um ponto (expresso por um vetor $n+1$ -dimensional), cujas coordenadas são multiplicadas por uma matriz não-singular.

Formalmente uma transformação projetiva é uma transformação linear que mapeia $\mathbb{P}^n \mapsto \mathbb{P}^n$ dentro do espaço $\mathbb{R}^{n+1}$ e que é dada na forma

$$\mathbf{x}' = \mathbf{H}\mathbf{x}$$

onde $\mathbf{H}$ é uma matriz quadrada de ordem $n+1$ e representa um conjunto de transformações não singulares que mapeiam um plano de projeção em outro.

Dentre as operações que $\mathbf{H}$ representa, existem quatro subgrupos de transformações que diferenciam-se pelas características que se mostram invariantes após sua aplicação. Cada uma destas classes de transformações é explanada a seguir.

- **Euclidiana ou Isométrica:** tem por objetivo modelar o movimento de corpos rígidos no plano $\mathbb{R}^2$, sendo uma operação que preserva a distância euclidiana. Composta de uma rotação seguida de translação, é dada na forma:

$$\mathbf{x}' = \mathbf{H}_E \mathbf{x} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} \epsilon \cos \theta & -\sin \theta & t_x \\ \epsilon \sin \theta & \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} \quad (3.1)$$

onde $\epsilon \pm 1$. Se ϵ assume valor positivo, então a isometria preserva a orientação; e, no caso contrário, a inverte. A rotação é determinada pela matriz $\mathbf{R}$ enquanto a translação é dada pelo vetor $\mathbf{t}$. Neste tipo de transformação as distâncias entre dois pontos, os ângulos entre linhas e áreas são invariantes.

- **Similaridade:** é uma transformação equivalente a operação de isometria com escala isotrópica. Considerando um fator de escala na matriz de rotação, essa operação permite a dilatação ou contração dos objetos no espaço e é definida por:

$$\mathbf{x}' = \mathbf{H}_S \mathbf{x} = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} s \cos \theta & -s \sin \theta & t_x \\ s \sin \theta & s \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} \quad (3.2)$$

onde s equivale a escala isotrópica. Esta transformação preserva a “forma” uma vez que os ângulos permanecem inalterados enquanto as distâncias entre pontos e a área, que embora não sejam preservados, têm suas proporções conservadas.

- **Affine:** corresponde a uma transformação linear não-singular seguida de uma translação, dada através de duas rotações e duas escalas não isotrópicas e que são expressas por:

$$\mathbf{x}' = H_A \mathbf{x} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} \quad (3.3)$$

A matriz $\mathbf{A}$ pode ser expressa em termos de uma decomposição SVD (*Singular value decomposition*) como:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T = (\mathbf{U}\mathbf{V}^T)(\mathbf{V}\mathbf{D}\mathbf{V}^T) = \mathbf{R}(\theta)(\mathbf{R}(-\phi)\mathbf{D}\mathbf{R}(\phi))$$

uma vez que $\mathbf{U}$ e $\mathbf{V}$ são matrizes ortogonais. Deste maneira a matriz $\mathbf{A}$ pode ser entendida como uma rotação (dada por ϕ); uma escala por λ_1 e λ_2 nas direções de x e y (expressas pela matriz diagonal $\mathbf{D}$ de ordem 2); o retorno da rotação inicial (dada por $-\phi$) e por fim uma operação de rotação (dada por θ). A consequência destas componentes pode ser observada na figura 3.3.

Por conta das escalas não isotrópicas as similaridades entre as proporções de comprimento e os ângulos entre as linhas não são preservadas. Entretanto, linhas paralelas, proporção entre segmentos de linha e áreas são conservados.

Figura 3.3: Distorções causadas por uma transformação *affine*.

- **Projetiva:** pode ser vista como uma transformação linear não singular mais geral que a *affine*, diferindo-se por apresentar um vetor não-nulo responsável por reproduzir os efeitos não-lineares decorrentes de uma perspectiva de visualização. Essa

transformação é definida na forma:

$$\mathbf{x}' = H_P \mathbf{x} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^T & v \end{bmatrix} \mathbf{x} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \mathbf{x} \quad (3.4)$$

Dentre os nove elementos da matriz, como apenas a razão entre eles é significativa, a transformação é especificada por oito parâmetros. Nota-se que não se pode afirmar que sempre será possível aplicar escala a operação, uma vez que o elemento v pode ser zero. Essa transformação preserva fundamentalmente duas características projetivas em ambos os planos: a colinearidade e a proporção cruzada de distâncias.

Uma transformação projetiva pode ser decomposta como uma combinação de três transformações encadeadas, onde cada matriz na sequência representa uma operação hierarquicamente superior a anterior.

$$H = H_S H_A H_P = \begin{bmatrix} sR & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} K & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ \mathbf{v}^T & v \end{bmatrix} = \begin{bmatrix} A & \mathbf{t} \\ \mathbf{v}^T & v \end{bmatrix} \quad (3.5)$$

Assim a matriz A resulta da combinação linear $A = sRK + \mathbf{t}\mathbf{v}^T$, sendo K uma matriz triangular superior normalizada com determinante igual a 1. Esta decomposição é válida para $v \neq 0$ e será única caso $s > 0$.

3.1.2 Projeções no espaço 3D

Nos parágrafos anteriores tratamos dos aspectos geométricos de perspectiva relacionados ao plano, espaço projetivo $\mathbb{P}^2$. Entretanto as propriedades de um ambiente tridimensional podem ser encaradas como generalizações destes fundamentos.

No espaço projetivo $\mathbb{P}^3$ os pontos são descritos por vetores de quatro coordenadas, $\tilde{\mathbf{x}} = [x_1 \ x_2 \ x_3 \ x_4]^T$ dados em coordenadas homogêneas. De modo análogo ao que ocorre em $\mathbb{P}^2$, as coordenadas do ponto em $\mathbb{R}^3$ são equivalentes a $(x_1/x_4, x_2/x_4, x_3/x_4)$ e, quando $x_4 = 0$, o vetor é a representação de pontos localizados no infinito em $\mathbb{R}^3$. Por conseguinte, as transformações projetivas no espaço $\mathbb{P}^3$ correspondem a transformações lineares de vetores quadri-dimensionais homogêneos, que são representados por matrizes não-singulares 4×4 .

3.2 Calibração de Câmeras

A calibração de câmeras é uma etapa vital quando se deseja obter informações métricas e uma representação tridimensional do ambiente a partir de um par de imagens. Dá-se o nome de calibração ao processo que determina o conjunto de parâmetros internos e externos ligados a formação de imagens. A estimação desses parâmetros é realizada relacionando pontos conhecidos no ambiente com suas respectivas projeções dentro do plano de imagem.

3.2.1 Parâmetros Intrínsecos

Os parâmetros internos da câmera, ou PARÂMETROS INTRÍNSECOS, compreendem o conjunto de características fundamentais de uma câmera e definem principalmente os aspectos geométricos de formação da imagem, sendo dados pela distância focal, ponto central de projeção, dimensões pixel, distorções induzidas pelo sistema de lentes e a inclinação entre os eixos de projeções.

Pressuposto que as câmeras podem ser generalizadas por sua versão mais simplificada, o modelo *pinhole* (figura 3.4), estabelece-se o modelo geométrico assumindo que o centro de projeção C é dado como origem do sistema de coordenadas e o eixo ótico é colinear ao eixo Z_c .

Figura 3.4: Modelo geométrico de uma câmera simplificada. A reta definida pelo ponto X no espaço e o centro ótico C intercepta o plano da imagem Π no ponto projetivo x , de coordenadas (x_c, y_c)

Sendo a distância focal f a medida entre o centro de projeções e o plano da imagem, um ponto $\mathbf{X}_c$ no espaço dado pelas coordenadas (X_c, Y_c, Z_c) é mapeado para o ponto $\mathbf{x} = (x_c, y_c)$ no plano da imagem por meio da seguinte equação:

$$[x \ y]^T = [fX/Z \ fY/Z]^T \quad (3.6)$$

A equação 3.6 é concebida tendo o ponto principal $\mathbf{P}$ como origem do sistema de coordenadas e sendo coincidente com o centro do plano da imagem. Entretanto, dada a dificuldade de uma montagem perfeitamente ortogonal, deve-se considerar um possível deslocamento do centro do plano da imagem em relação ao ponto principal nas direções de x e y que são dados respectivamente por c_x e c_y .

A proporção entre a largura e a altura de um pixel é denominada razão de aspecto (*aspect ratio*) e, quando não são perfeitamente quadrados, definem fatores de escala diferentes no sentido de cada eixos. Essa variação é modelada considerando valores focais independentes para cada direção e relacionada ao número de pixels por unidade de distância nas direções x e y (valores η_x e η_y , respectivamente).

Na geometria projetiva tanto os pontos no espaço quanto os pontos no plano são dados em termos de coordenadas homogêneas, assim podemos reescrever a equação 3.6 na forma de uma combinação linear que incorpora cada um destes elementos:

$$s \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} = \begin{bmatrix} f\eta_x & 0 & c_x & 0 \\ 0 & f\eta_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (3.7)$$

$$\tilde{\mathbf{x}} = [\mathbf{K} | 0^T] \tilde{\mathbf{X}}_c \quad (3.8)$$

$$\tilde{\mathbf{x}} = \mathbf{P} \tilde{\mathbf{X}}_c \quad (3.9)$$

onde $\mathbf{P}_{4 \times 3}$ é chamada de matriz de projeção de câmera e define a transformação homogênea que relaciona os pontos no mundo real com suas projeções no plano de imagem em função de $\mathbf{K}_{3 \times 3}$, conhecida como MATRIZ FUNDAMENTAL ou de PARÂMETROS INTRÍNSECOS da câmera.

Com o objetivo de focar os raios e ajustar a imagem que é projetada sobre o sensor, a maioria das câmeras faz uso de um conjunto ótico composto por uma ou mais lentes. Mesmo quando confeccionadas com precisão, dada sua natureza esférica as lentes inevitavelmente provocam deformações na imagem projetada. Esses efeitos são conhecidos

como distorções radiais e produzem aspectos visuais em forma de “barril” (*barrel*) e/ou “almofada” (*pincushion*), como ilustrado nas figuras 3.5(b) e 3.5(c).

Figura 3.5: Tipos básicos de distorções causadas por lentes e consequências para a imagem. *Em sequência temos uma imagem dada sem distorção (a) e os efeitos das distorções do tipo barril (b) e formato almofada (c). Em (d) temos uma imagem das ruínas de Petra (Jordânia) e a correspondente (e) após correções de distorção.*

As distorções são deslocamentos dos pontos da imagem em relação ao esperados pela perspectiva projetiva perfeita. Entre os efeitos causados estão a sinuosidades em linhas retas e a divergências em elementos paralelos, como se observa nas imagens 3.5(d) e 3.5(e).

A modelagem precisa das lentes é uma tarefa complexa. Segundo (Brown, 1971) a distorção radial é a componente mais importante a ser considerada, podendo ser modelada na direção de cada eixo por uma série de potências, cujo valor é dado em função da distância (raio) em relação ao centro:

$$x' = x \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + 2p_1 xy + p_2(r^2 + 2x^2) \quad (3.10)$$

$$y' = x \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + p_1(r^2 + 2y^2) + 2p_2 xy \quad (3.11)$$

onde $r^2 = x^2 + y^2$ e os coeficientes radiais e tangenciais correspondem respectivamente a k_n e p_n . Elementos de mais alta ordem exercem pouca influencia e assim costumam ser desprezados na maioria das aplicações.

Os coeficientes de distorção também são considerados parte do conjunto de parâmetros intrínsecos da câmera.

3.2.2 Parâmetros Extrínsecos

A equação 3.7 permite determinar a projeção que um ponto $\mathbf{X}$ do mundo real terá, desde que suas coordenadas sejam dadas em relação ao sistema de referência da câmera. Entretanto, é bastante comum que os elementos de interesse e inclusive a própria câmera, sejam expressos em relação a um sistema global de referência (sistema de coordenadas do mundo). A figura 3.6 ilustra a relação existente entre os dois sistemas de coordenadas.

Os pontos no espaço possuem duas representações equivalentes, sendo $\mathbf{X}_c$ expressa em coordenadas da câmera e $\mathbf{X}_w$ em coordenadas do mundo. Supondo o ponto $\mathbf{O}$ como origem do sistema de coordenadas do mundo, há uma transformação linear dada por uma rotação e translação que relaciona os dois sistemas de referência e por conseguinte os pontos no plano da imagem com seus correspondentes no mundo.

Se reescrevemos a equação 3.7 de maneira a incorporar essa transformação e tomar o sistema de coordenadas do mundo como referência, temos:

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_1 \\ R_{21} & R_{22} & R_{23} & t_2 \\ R_{31} & R_{32} & R_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3.12)$$

ou na forma contraída por:

$$\tilde{\mathbf{x}} = \mathbf{K}[\mathbf{R}|\mathbf{t}]\tilde{\mathbf{X}}_w \quad (3.13)$$

$$\tilde{\mathbf{x}} = \mathbf{P}\tilde{\mathbf{X}}_w \quad (3.14)$$

Figura 3.6: Relação entre os sistemas de coordenadas de mundo e da câmera no modelo geométrico projetivo. *Uma transformação de corpo rígido descreve a relação entre os sistemas de coordenadas e permite expressar a correspondência entre um ponto no espaço dado em coordenadas de mundo e sua projeção dada em termos do sistema de referência da câmera.*

em que os elementos dados por $[R|t]$ são conhecidos como PARÂMETROS EXTRÍNSECOS.

A matriz $P = K[R|t]$ é responsável pelo mapeamento entre pontos no mundo e seus correlatos no plano de imagem, modelando o sistema projetivo em função dos parâmetros intrínsecos e extrínsecos da câmera.

3.3 Imagens Estéreo

Um sistema composto por duas ou mais câmeras permite a captura simultânea de uma cena sob diferentes perspectivas. Em função da distância que os elementos estão das câmeras as projeções numa e noutra imagem se formam de maneira diferenciada.

Por exemplo, num arranjo disposto de maneira horizontal com câmeras a esquerda e a direita, como ilustrado na figura 3.7, os objetos posicionados mais próximos das câmeras terão um deslocamento horizontal maior entre as imagens enquanto os mais distantes apresentam pouca variação entre as projeções.

Uma vez conhecida a posição entre as câmeras em relação a um sistema de coordenadas global e calculada a diferença entre pontos correspondentes nas áreas de sobreposição das

Figura 3.7: Ilustração de um sistema de visão estéreo. *As imagens capturadas por cada uma das câmeras apresentam diferentes perspectivas da mesma cena. De acordo com a posição dos objetos no espaço, suas projeções aparecem localizadas em lugares diferentes em cada imagem.*

imagens é possível estimar a profundidade dos pontos. De acordo com o deslocamento em pixels e das relações geométricas do arranjo, a posição espacial dos elementos de cena pode ser determinada pelo processo inverso ao da projeção abordada na seção 3.1. Na literatura essa inversão é conhecida como *back-projection*.

O processo de calibração de um sistema estéreo consiste então em determinar a transformação que descreve a relação entre os sistemas de coordenadas das câmeras a partir de elementos previamente conhecidos e suas projeções nas duas imagens. Usualmente a pose de uma das câmeras é dada em relação a outra, que por sua vez tem sua posição em relação ao sistema de referência global conhecida.

3.3.1 Geometria Epipolar

A geometria epipolar é usada em visão estéreo para limitar o espaço de busca na correspondência de pontos nas imagens. Equivale a modelagem geométrica entre duas perspectivas de um ambiente e depende apenas dos parâmetros intrínsecos e extrínsecos das câmeras, não do cenário.

A modelagem fundamenta-se na triangulação entre os pontos no espaço, os centros óticos e as projeções destes pontos nos planos de imagem. Através da Figura 3.8 nota-se

que os pontos $\mathbf{X}$, $\mathbf{O}_1$ e $\mathbf{O}_2$ formam um plano no espaço 3D que intercepta os planos de imagem. Este plano é chamado de plano epipolar e as linhas decorrentes de sua intersecção com os planos das imagens, $\mathbf{l}_1$ e $\mathbf{l}_2$, são conhecidas como linhas epipolares.

Figura 3.8: Relação geométrica epipolar entre dois planos de imagem. *O plano epipolar é determinado pela relação entre os pontos do espaço, suas projeções a os centros óticos das câmeras tidos como coplanares. Os pontos e_1 e e_2 são chamados de epípolos e correspondem a projeção do centro de ótico de uma câmera em relação ao plano de imagem da outra.*

Observe que as projeções do ponto $\mathbf{X}$ nos planos das imagens dadas pelos pontos $\mathbf{x}_1$ e $\mathbf{x}_2$ também pertencem ao plano epipolar e, por consequência, pertencem as linhas epipolares em cada um dos planos da imagem. Essa propriedade se caracteriza como uma importante restrição para o processo que procura estabelecer a correspondência (*matching*) entre pontos nas imagens, permitindo que a busca seja realizada localmente (ao longo da linha epipolar) e não em toda a imagem.

Conforme discutido da seção 3.1.1, uma transformação homográfica H permite relacionar pontos entre diferentes planos de imagem. Aplicando essa transformação a um ponto $\mathbf{p}'$ que não coincida com o ponto $\mathbf{X}$ fará com sua projeção $\mathbf{x}_2'$ no plano de imagem da outra câmera apareça deslocado ao longo da linha.

3.3.2 Matriz Fundamental

A representação algébrica da geometria epipolar é dada a partir da MATRIZ FUNDAMENTAL (F), que equivale ao mapeamento projetivo dos pontos de uma imagem para as linhas epipolares de outra.

Segundo (Hartley e Zisserman, 2003) a matriz F apresenta propriedades, a saber:

Matriz Fundamental: relaciona-se de uma matriz de ordem 3×3 com 7 graus de liberdade e rank 2.

Correspondência entre pontos: Se $\mathbf{x}_1$ e $\mathbf{x}_2$ são pontos correspondentes nos dois planos de imagem, então é válida a relação $\mathbf{x}_1^T F \mathbf{x}_2 = 0$ para todos os pares de pontos correspondentes.

Transposição: se F é uma matriz fundamental que relaciona o par de câmeras (P, P') , então F^T é a matriz para o par dado na ordem oposta (P', P) .

Linhas epipolares: para todo ponto $\mathbf{x}$ da primeira imagem, a linha epipolar da segunda imagem é definida por $\mathbf{l}_2 = F \mathbf{x}_1$. Similarmente, $\mathbf{l}_1 = F^T \mathbf{x}_2$ define a correspondência entre um ponto $\mathbf{x}_2$ da segunda imagem com a linha epipolar $\mathbf{l}_1$ da primeira imagem.

Epipolos: para qualquer ponto $\mathbf{x}_1$ da primeira imagem (exceto e_1), a linha epipolar $\mathbf{l}_2 = F \mathbf{x}_1$ contém e_2 . Assim e_2 satisfaz a expressão $e_2^T (F \mathbf{x}_1) = (e_2^T F) \mathbf{x}_1 = 0$ para qualquer ponto $\mathbf{x}_1$. Isto implica que $e_2^T F = 0$, permitindo que e_2 seja calculado pelo espaço nulo à esquerda de F . De modo análogo, $F e_1 = 0$ e e_1 é computado pelo espaço nulo à direita de F .

Assim, supondo um par de pontos correspondentes nas duas imagens, dados por $\mathbf{x}'$ e $\mathbf{x}$, a matriz fundamental F é definida por:

$$\mathbf{x}'^T F \mathbf{x} = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \quad (3.15)$$

Calculando o produto das matrizes da 3.15 obtemos:

$$x'x f_{11} + x'y f_{12} + x' f_{13} + y'x f_{21} + y'y f_{22} + y' f_{23} + x f_{31} + y f_{32} + f_{33} = 0 \quad (3.16)$$

Para computarmos os elementos da matriz F é necessário alguns pares de pontos que possibilitem a composição de um sistema de equações em que a relação $\mathbf{x}'^T F \mathbf{x} = 0$ seja válida.

Considerando um conjunto de n pares de pontos correspondentes (dados por $\mathbf{x}'_i$ e $\mathbf{x}_i$) e dispondo os elementos de F na forma de um vetor, podemos reescrever a equação 3.16 como o seguinte produto:

$$\mathbf{A}\mathbf{f} = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ x'_2x_2 & x'_2y_2 & x'_2 & y'_2x_2 & y'_2y_2 & y'_2 & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ \vdots \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0 \quad (3.17)$$

onde $\mathbf{A}$ é uma matriz de ordem $n \times 9$ dada pelo conjunto de pares de pontos correspondentes.

Nesta equação, o elemento f_{33} pode ser definido como 1 em função da matriz fundamental ser singular. Por apresentar ainda oito incógnitas é necessário ao menos oito pares de pontos para seja possível determinar os elementos da matriz (Zhang, 2013).

3.3.3 Retificação

O processo de retificação em imagens estéreo consiste em determinar transformações 2D para que os planos das imagens sejam posicionados de modo em que as linhas epipolares se tornem conjugadas e paralelas a um dos eixos da imagem. A grande vantagem associada a retificação é oferecer uma maneira conveniente para a correspondência estéreo. Este conceito é ilustrado na Figura 3.9.

As imagens retificadas podem ser consideradas novas imagens, como se tivessem sido adquiridas por um novo conjunto de câmeras estéreo paralelas. Na prática são obtidas a partir da rotação das câmeras originais em torno de seus centros óticos até que as imagens se tornem coplanares.

Após a retificação um par de retas epipolares conjugadas se tornam colineares e paralelas a um dos eixos da imagem, tendo a mesma coordenada para o segundo eixo das imagens. Adicionalmente, os epipolos estarão localizados no infinito pois os planos-imagem serão paralelos linha de base.

3.3.4 Mapa de disparidade

Uma vez que as imagens tenham sido retificadas e as linhas epipolares estejam alinhadas, é possível extrair informações de profundidade através da identificação de pontos correspondentes e de seus deslocamentos entre as imagens.

Figura 3.9: Interpretação geométrica do processo de retificação de imagens. *Aplica uma transformação as imagens para que as linhas epipolares apresentem-se de maneira paralela ao eixo X e conjugadas na mesma altura em Y*

Os algoritmos de visão estéreo buscam pela correspondências entre pontos nas duas imagens de entrada, computando a diferença de localização para cada pixel e armazenado esses valores em uma matriz bidimensional conhecida como mapa de disparidades.

As técnicas de correspondência estéreo utilizam padrões definidos em ambas as imagens para medir sua correlação, sendo as mais comuns: Soma das diferenças Absolutas - SAD (*Sum of Absolute Difference*), Soma das diferenças Absolutas - SSD (*Sum of Squared Difference*) e Correlação Cruzada Normalizada - NCC (*Normalized Cross Correlation*).

Utilizando um processo de triangulação que considera as informações métricas do sistema estéreo, as disparidades são convertidas em diferentes níveis de profundidade que representam a estrutura 3D da cena de acordo com a equação:

$$Z = \frac{Bf}{d} \quad (3.18)$$

onde B é a distância entre os centros óticos das câmeras (*baseline*), f a distância focal do arranjo e d a medida de disparidade ($x_d - x_e$).

Na Figura 3.10(c) temos uma imagem resultante dos diferentes níveis de profundidade estimada a partir das figuras 3.10(a) e 3.10(a). As áreas mais claras do mapa de profundidade representam as superfícies dos objetos que estão mais próximos ao plano da imagem, enquanto as áreas escuras, as superfícies mais distantes.

Figura 3.10: Exemplo de imagem de de profundidade obtida a partir de um par de imagens estéreo. *Considerando as disparidades calculadas para cada pixel é possível criar um mapa de profundidade da cena, em que os elementos mais claros estão posicionados mais próximos e os mais escuros mais longe das câmeras.*

Conforme pode ser constatado através da figura 3.10(c), usualmente o mapa de disparidade resulta em uma imagem onde para cada pixel (x, y) existe associada uma informação de profundidade “*depth/disparity*” constituindo assim o mapa de profundidade da cena (*depth map*). É importante destacar que em certas partes da imagem podemos ter pixels cuja disparidade não pode ser calculada, usualmente por que não foi encontrada uma correspondência de um mesmo ponto nas duas imagens capturadas pela câmera estéreo.

A profundidade pode ser representada por uma escala discreta de valores, que representam e discriminam os elementos mais próximos e mais afastados presentes na cena. Portanto, um mapa de disparidade, ou mapa de profundidade, pode ou não ser metricamente definido, ou seja, os valores de disparidade podem não estar diretamente relacionados com valores de distância representados em uma escala métrica (p.ex. centímetros ou metros). Através de um ajuste é possível obter a partir de um mapa de disparidade uma nuvem de pontos 3D (*point cloud*), onde cada ponto desta nuvem representa uma coordenada no espaço tridimensional (x, y, z) obtida a partir de um ponto da imagem $(x, y, depth)$, inclusive onde é desejável que os pontos possuam uma relação métrica de distâncias entre si e respeitando da melhor forma possível as distâncias metricamente definidas entre os elementos da cena. Para se obter uma nuvem de pontos 3D a partir do mapa de disparidades é necessário realizar ajustes (transformações usando parâmetros previamente estimados), pois os valores crescentes/decrescentes de disparidade podem não representar uma relação linear em termos de distâncias, sendo necessário a aplicação de uma função de ajuste.

3.4 Considerações Finais

Neste capítulo foram apresentados os conceitos relacionados a visão computacional em sistemas baseados na aquisição de imagens por câmeras estéreo. Foi apresentada uma revisão de conceitos sobre câmeras, geometria projetiva, transformações projetivas e sobre calibração de câmeras. Estes conceitos formam a base de conhecimentos necessários para descrever a proposta desenvolvida nesta tese, relacionada a calibração automática de câmeras estéreo.

O processo de calibração de câmeras estéreo, visando a obtenção de mapas de disparidade/profundidade, requer a definição de parâmetros de calibração, denominados de parâmetros intrínsecos e parâmetros extrínsecos das câmeras, sendo representados por matrizes de transformação, onde tais parâmetros também foram descritos e detalhados neste capítulo. Este processo de calibração é usualmente realizado através do uso de padrões de referência, que são apresentados de forma interativa ao sistema, e assim permitem estimar os parâmetros intrínsecos e extrínsecos do sistema estéreo.

Os parâmetros intrínsecos (matriz fundamental) são relacionados as propriedades internas da própria câmera e podem ser determinados pelo fabricante através de um processo de ajuste e calibração dos mesmos. Este processo pode ser realizado uma única vez. Por outro lado, os parâmetros extrínsecos dependem da configuração (posicionamento) de uma câmera em relação a outra, sendo que podem variar (p.ex. alteração da distância entre as duas câmeras, que define o chamado *baseline*, ou mesmo podem sofrer perturbações (desajustes na montagem das câmeras), podendo assim levar a necessidade da realização de uma nova calibração nestes casos. A determinação dos parâmetros extrínsecos é portanto de grande importância no processo de obtenção do mapa de disparidade, sendo que por sua vez os parâmetros intrínsecos podem ser previamente definidos e fixados previamente.

O processo de calibração dos parâmetros extrínsecos é usualmente demorado e exige um certo conhecimento por parte do usuário na sua execução. Este processo é tipicamente realizado através da apresentação de um padrão quadriculado (xadrez), precisamente definido, que é apresentado repetidas vezes a fim de obter pontos de referência usados na estimação da matriz que definem os parâmetros extrínsecos. A realização deste processo de calibração por um usuário leigo, e junto a aplicações comerciais (fora de um laboratório de pesquisa) pode apresentar diversas dificuldades e problemas para a sua realização, que vão desde a falta da disponibilidade de padrões de calibração até a dificuldade em obter um número adequado (e variado) de pontos de referência para o processo de ajuste dos parâmetros. Pode ser citado como exemplo disto, o padrão *checkerboard* adotado

pelo LRM para realizar experimentos com veículos autônomos em ambientes *outdoor*, que possui uma dimensão de 70cm x 100cm (uma placa de quase 1 metro quadrado).

Considerando a importância do processo de calibração para que seja possível realizar o cálculo do mapa de disparidade a partir dos parâmetros estimados neste processo, conforme descrito neste capítulo, e considerando as dificuldades da aplicação do processo tradicional de calibração através da apresentação interativa de um padrão quadriculado de referência, foi constatada a necessidade do desenvolvimento de um método de calibração para o ajuste dos parâmetros extrínsecos que pudesse ser realizado de forma mais automática, com o mínimo de interação por parte de um usuário. Um processo de calibração automática para câmeras estéreo, que atende aos requisitos acima, e que está sendo proposto nesta tese, será apresentado e descrito no próximo capítulo.

Calibração Automática

Nesta tese é proposta uma metodologia para a calibração automatizada entre sensores que é capaz de definir as relações entre sistemas baseando-se apenas na avaliação de características esperadas para certos objetos do ambiente.

Esta metodologia foi aplicada na calibração dos parâmetros extrínsecos de um sistema de visão estéreo, utilizando-se de Algoritmos Genéticos para realizar o processo de calibração (otimização) dos parâmetros. O processo desenvolvido apresenta um relevante diferencial em relação às técnicas tradicionais, simplificando o processo e a interação com o sistema a fim de permitir a calibração do par de câmeras.

O processo descrito nesta tese permite determinar, através da avaliação de elementos do cenário selecionados pelo usuário e de algumas de suas propriedades básicas, as relações de transformação entre sistemas de coordenadas locais de modo que seja possível correlacioná-los em termos de um sistema referencial comum.

Os métodos de calibração tem por objetivo determinar os valores das matrizes de transformação que expressam as relações geométricas entre os espaços vetoriais quer euclidianos ou projetivos. Na seção a seguir descrevemos os processos tradicionais de calibração para sistemas de visão estéreo, que foi o sistema escolhido para a experimentação da metodologia proposta.

Na seção 4.3 o método proposto é descrito em detalhes, enquanto os aspectos relativos à implementação são tratados na seção 4.4.

4.1 Métodos tradicionais de Calibração

De acordo com as discussões na subseção 3.3.2, a tarefa de calibrar um sistema estéreo consiste em determinar os elementos da matriz fundamental a partir da associação de 8 correspondentes nos dois planos das imagens. Em função disto, este algoritmo é conhecido algoritmo dos 8 pontos.

De igual modo, vale lembrar que a matriz fundamental F é singular, tem rank 2 e os epipolos e_1 e e_2 podem ser calculados pelos espaços nulos à esquerda e à direita de F . Dado que as linhas epipolares são coincidentes com os epipolos, a matriz fundamental computada pela Equação 3.17 nem sempre é singular, fazendo com que as linhas epipolares não sejam coincidentes.

Segundo (Hartley e Zisserman, 2003) esse problema pode ser resolvido impondo uma restrição de singularidade, onde F é substituída pela matriz F' , de tal modo que o determinante de F' é 0 e minimiza $\|F - F'\|$. A matriz F' pode ser obtida por um processo de *Decomposição em Valores Singulares* - (SVD).

Conhecido como “Algoritmo dos 8 Pontos”, este método considera um conjunto de $n \geq 8$ pares de pontos correspondentes que são transformados em termos de coordenadas normalizadas e emprega a restrição de singularidade dada no parágrafo anterior. Sua execução é descrita pelas seguintes etapas:

1. As coordenadas dos pontos $\mathbf{x}'_i$ e $\mathbf{x}_i$ são normalizadas através de uma transformação T composta por uma rotação e escala, tal que $\hat{\mathbf{x}}'_i = T\mathbf{x}'_i$ e $\hat{\mathbf{x}}_i = T\mathbf{x}_i$
 2. os elementos da matriz fundamental $\hat{F}$ são calculados utilizando os pontos normalizados no passo anterior de acordo com a Equação 3.17.
 3. calcula-se a matriz $\hat{F}'$ a partir de $\hat{F}$, sendo que $\hat{F}'$ respeita a restrição de singularidade imposta, isto é, o $\det(\hat{F}') = 0$ e $\hat{F}$ minimiza $\|\hat{F} - \hat{F}'\|$
- (iv) a matriz fundamental F para os pontos é então obtida a partir do processo inverso da normalização $F = T'^T \hat{F}' T$

4.2 Visão estéreo em sistemas multi-baseline

Os problemas associados a erros de calibração podem ser minimizados quando adota-se o uso de uma câmera estéreo ao invés de um conjunto composto por duas câmeras. A montagem cuidadosa dos componentes durante a fabricação e o processo controlado de calibração

permite que se alcance bastante precisão. Por sua vez, um conjunto estéreo formado por duas câmeras monoculares oferece mais flexibilidade permitindo a combinação de câmeras e lentes que sejam mais adequadas a aplicação.

A imprecisão de um sistema estéreo, dada por A , pode ser calculada segundo a expressão:

$$A = \frac{Z^2 \varepsilon}{Bf} \quad (4.1)$$

onde ε o erro de correlação ou calibração, enquanto B e f correspondem respectivamente a medida do *baseline* e a distância focal. Usualmente $\varepsilon = 0.2$ é considerada uma estimativa razoável, assumindo uma superfície texturizada.

A partir da equação 4.1 observa-se que imprecisão do sistema é proporcional a distância, aumentando em ordem quadrática em relação a Z . É possível concluir ainda que tanto o aumento da distância focal quanto do *baseline* promovem a diminuição da imprecisão para maiores distâncias.

Como consequência do aumento na resolução ou no *baseline* está a elevação do custo computacional associado ao cálculo das disparidades, que segundo a equação 3.18, é definida como:

$$d = \frac{Bf}{Z} \quad (4.2)$$

Assim a escolha do *baseline* e das características focais do conjunto estão diretamente ligadas a precisão desejada em termos do intervalo das distâncias de interesse da aplicação.

Estudos realizados no LRM procuraram determinar os aspectos de imprecisão relacionados a diferentes configurações de sistemas estéreo para serem empregados em processos de detecção de obstáculos. Considerando uma câmera Point Grey¹ Bumblebee2, cujos parâmetros informados pelo fabricante são dados na tabela 4.1, foi possível calcular a imprecisão do sistema para diferentes distâncias e consequentemente o tamanho mínimo de um obstáculo detectável.

Propriedade	Valor	Distância focal	Valor Normalizado
Resolução horizontal	640 pixels	2.5 mm	0.44
Distância focal	3.8 milímetros	3.8 mm	0.768
<i>Baseline</i>	0.12 metros	6 mm	1.27

Tabela 4.1: Parâmetros do sistema estéreo para uma câmera Bumblebee2 da Point Grey.

Sabendo que a distância em pixels é dada por $f = nr$, onde n é a distância focal normalizada e r corresponde a resolução horizontal, para a configuração dada na tabela

¹<http://www.ptgrey.com/products/stereo.asp>

4.1, temos que $f = 0.768 \times 640 = 491.51$. Supondo o erro de correlação $\varepsilon = 0.2$, a imprecisão do sistema para uma distância de 10 metros será:

$$A = \frac{Z^2 \varepsilon}{Bf} = \frac{10^2 \times 0.2}{0.12 \times 491.51} = 0.3391 \quad (4.3)$$

ou seja, a 10 metros de distância a imprecisão do sistema é de aproximadamente 33 centímetros.

Diferentes configurações para o sistema foram testados utilizando um método de detecção de obstáculos configurado para registrar quaisquer obstáculos de dimensões maiores que 33 cm de altura, a partir da superfície da rua. Na tabela 4.2 estão sumarizados os resultados dos experimentos.

Resolução	<i>Baseline</i>	Dist. focal	Imprecisão	Valor teórico	Valor experimental
640	0.12	3.8	33 cm	10	7.5
640	0.12	3.8	50 cm	12.1	10.2
640	0.12	6	33 cm	12.7	12.3
640	0.12	6	50 cm	15.6	15
1280	0.12	3.8	33 cm	13.5	11.6
1280	0.12	3.8	50 cm	16.5	15.6
1280	0.24	3.8	33 cm	19.7	20
1280	0.24	3.8	50 cm	24.2	25.7

Tabela 4.2: Comparação entre valores teóricos e experimentais na detecção de obstáculos. *Valores obtidos a partir de três câmeras estéreo: Bumblebee2 de 3.8 mm (65-deg HFOVe), Bumblebee2 de 6 mm (43-deg HFOV) e BumblebeeXB3 de 3.8 mm (66-deg HFOV) fabricadas pela Point Grey.*

Neste experimento pode-se observar que a resposta do sistema de detecção era mais precisa de acordo com as características de configuração do sistema estéreo, de tal modo que o processo de método de detecção se tornaria mais robusto se um sistema flexível fosse empregado.

Em (Gallup *et al.*, 2008), os autores propõe uma abordagem que combina multi-resoluções e diferentes *baselines* com objetivo de que o erro nas medidas de profundidade cresçam de modo linear ao invés de quadrático. Contando com pares de câmeras de diferentes configurações, o método seleciona o melhor conjunto estéreo que permite amostrar as imagens oferecendo a precisão geométrica desejada.

A concepção de um arranjo estéreo bifocal montado a partir da combinação de dois conjuntos estéreo é dado em (Rovira-Más *et al.*, 2009). O primeiro par, com distância

focal de 4mm e *baseline* de 11cm, tinha por propósito prover a percepção de curto alcance; enquanto o segundo par, com distância focal de 16mm e *baseline* 22cm, servia para médio alcance. Essa combinação ofereceu diferentes faixas úteis de percepção, inclusive com certa área de sobreposição.

Os testes práticos demonstraram uma boa dispersão da nuvem de pontos resultante da reconstrução 3D dos sistemas de curto e médio alcance, permitindo que obstáculos posicionados a 9.1m (30 pés) e 23.5m (77 pés) de distância fossem adequadamente detectados. A fusão das áreas detectáveis pelas câmeras permitiu a cobertura de elementos posicionados entre 5 e 25 metros a frente do veículo, sendo coerente com a área de 30m esperado para esta configuração.

Em (Milella *et al.*, 2013) o sistema multi-baseline foi concebido para ser utilizado em um veículo autônomo que opere em ambientes naturais e agrícolas, executando tarefas de reconstrução 3D e segmentação. O sistema é composto por uma câmera trinocular Bumblebee XB3 de 3.8mm que opera com dois possíveis *baselines*, uma combinação de câmeras com 12 e outra com 24cm. O segundo conjunto é customizado sendo formado por três câmeras Flea3 idênticas, com distância focal de 12mm e montadas com *baselines* de 40 e 80cm. Este sistema permite a cobertura de uma área de aproximadamente 30 metros para a Bumblebee e de aproximadamente 60 metros a frente do veículo para o conjunto de Flea3.

Os parâmetros intrínsecos e extrínsecos para cada par de câmeras são determinados através de um processo convencional de calibração, empregando padrões xadrez e as funções da biblioteca OpenCV. Após esse passo, etapas adicionais de calibração são realizadas com o objetivo de alinhar todos os sistemas de coordenadas em relação ao sistema de referência do veículo. As reconstruções 3D para os diferentes *baselines* são integradas separadamente, resultando em duas nuvens de pontos que são então combinadas através do processo de transformação para o sistema de referência do veículo. Usando técnicas de filtragem estatísticas e voxelização a quantidade de pontos é reduzida de maneira que a nuvem possa ser segmentada e classificadas em regiões de chão e não-chão.

4.3 Abordagem Evolutiva

Como discutido nas subseções 3.2.1 e 3.2.2, os parâmetros intrínsecos de uma câmera são dados pelas propriedades de seus elementos e características de construção, enquanto os parâmetros extrínsecos são dados pela pose da câmera em relação ao sistema de referência. Deste modo os parâmetros intrínsecos são usualmente obtidos junto ao fabricante

enquanto os extrínsecos precisarão obrigatoriamente ser estabelecidos por algum processo de calibração após cada montagem.

Em linhas gerais, o nosso método tem início com a aquisição de um par de imagens capturadas simultaneamente por ambas as câmeras. Um usuário então seleciona algumas regiões da imagem que contenham características horizontais e/ou verticais de interesse, tais como áreas pertencentes ao chão, calçadas, postes, placas ou mesmo pessoas que estejam paradas (em pé) na região de sobreposição das imagens.

A partir destas imagens, tendo os parâmetros intrínsecos e o *baseline* pré-estabelecidos, um processo evolutivo conduzido por um algoritmo genético (AG) publica transformações geométricas equivalentes a matriz de rotação entre as câmeras. Para cada matriz gerada pelo AG, o mapa de disparidades é computado utilizando-se de algum método de correspondência estéreo. Em seguida as áreas do mapa equivalentes as regiões selecionadas pelo usuário na imagem são então avaliadas pela função de *fitness* segundo a natureza da região, conferindo uma nota de aptidão para aquela transformação.

O processo continua iterativamente gerando uma outra transformação, que novamente será aplicada e cujo mapa de disparidades terá as regiões de interesse avaliadas a luz das características esperadas para objetos daquele tipo. Uma estratégia elitista preserva a melhor solução obtida em cada geração, enquanto os operadores genéticos de cruzamento (*crossover*) e mutação atuam de modo a recombinar um conjunto de prováveis soluções produzindo outras novas.

A determinação dos valores adequados de cada um dos elementos da matriz de rotação se dá de maneira natural através do processo evolutivo que preserva as melhores soluções de uma etapa (geração) para outra e ainda permite que suas características sejam “espalhadas” para as demais hipóteses de forma proporcional a “nota” que receberam pela função de *fitness*.

Utilizamos uma função de avaliação (*fitness*) dada por três componentes diferentes, cada qual agregando aspectos específicos relativos aos objetos de interesse. Assim a função é dada por:

$$Fitness(n) = \sum_{R_g=1}^k Grad(R_g) + \sum_{R_v=1}^m Homo(R_v) + \sum_{R_d=1}^n Desc(R_d) \quad (4.4)$$

onde *Grad*, *Homo* e *Desc* são, respectivamente, subfunções que avaliam aspectos relacionados a GRADIENTES, HOMOGENEIDADE e DESCONTINUIDADES (limites) verticais.

A função de gradiente, ou *Grad*, potencializa elementos que tenham a tendência de se estender em profundidade, valorizando objetos cuja projeção siga em direção do ponto de

fuga. Uma análise das disparidades médias horizontais é realizada através de regressão linear, buscando analisar a progressividade e estabilidade dos valores ao longo do eixo Z

A avaliação dos componentes homogêneos pela função *Homo* tem por objetivo análise a ocorrência de discrepâncias e variabilidade nos níveis de disparidade do mapa para as regiões de interesse, usualmente associadas a falhas no processo de correlação ou oriundas de incoerências nas relações geométricas de elementos orientados de modo próxima da paralela ao sistema de câmeras.

Por fim, *Desc* é a componente responsável por conferir melhores “encaixes” verticais, a componente de descontinuidades analisa secções no mapa de disparidades que equivalham ao ponto de início da mudança de orientação entre elementos verticais e horizontais, como a que ocorre por exemplo entre o chão e a base de um poste.

Nestes termos, a função de *fitness* avalia o mapa de disparidade segundo as características esperadas para cada uma das k , m e n regiões r_x assinaladas na imagem, atribuindo uma nota aquela transformação que representa a configuração do sistema estéreo.

Figura 4.1: Exemplo de regiões selecionadas pelo usuário para a composição do fitness. As regiões em vermelho representam áreas de gradientes; dadas em azul, há duas seleções de componentes com homogeneidade esperadas dadas pela verticalidade do postes marcados; e por fim, as áreas de descontinuidade aparecem em verde.

Diante do exposto, o ganho (ou contribuição) de uma região em particular para o cálculo fitness pode ser computadas a partir das seguintes equações:

- Função de Gradiente, $Grad(R_g)$:

$$\begin{aligned}\xi &= \kappa \times \beta \times \Delta \\ \xi &= c \frac{A_r}{A_{img}} \times \frac{n \sum y - \sum x \sum xy}{n \sum x^2 - (\sum x)^2} \times \frac{d_{valid}}{A_{roi}} \quad \forall \beta \in \mathbb{R}\end{aligned} \quad (4.5)$$

o ganho ξ oferecido pela região é dado pelo produto entre a significância da região em relação a imagem (κ), a linearidade do gradiente β e o percentual de ocupação (Δ), calculado a partir das disparidades válidas d_{valid} em relação a área da região de interesse (A_{roi}).

- Função de Homogeneidade, $Homo(R_h)$:

$$\begin{aligned}\xi &= \kappa \times H \times \Delta \\ \xi &= c \frac{A_r}{A_{img}} \times \left(1 - \frac{\sqrt{\sum (d - \bar{d})^2}}{\frac{1}{n} \sum_{i=1}^n d_i} \right) \times \frac{d_{valid}}{A_{roi}} \quad \forall \mu, \sigma \neq 0\end{aligned} \quad (4.6)$$

onde o ganho ξ é dado por pelo produto entre a significância da região em relação a imagem (κ), o fator de homogeneidade H e o percentual de ocupação (Δ), calculado a partir das disparidades válidas d_{valid} em relação a área da região de interesse (A_r).

- Função de Descontinuidades, $Desc(R_d)$:

$$\begin{aligned}\xi &= \kappa \times H \times \Delta \\ \xi &= c \frac{A_r}{A_{img}} \times \left(1 - \frac{h_d}{h_r} \right) \times \frac{d_{valid}}{A_{roi}} \quad \exists h_d\end{aligned} \quad (4.7)$$

4.4 Desenvolvimento

Com o objetivo de manter-se compatível com os demais sistemas desenvolvidos optamos por adotar ferramentas compatíveis com os *frameworks* usados pelo grupo e no escopo do projeto CaRINA. Assim o desenvolvimento do sistema baseou-se no ROS, começando com a versão "F"(Fuerte), mas sendo portado ao longo das versões subsequentes até "I"(Indigo Igloo) a mais atual.

Por meio de um poderoso recurso de propagação de transformações entre diferentes sistemas de referência denominado TF (*Transform*), o ROS oferece um mecanismo que

permite descrever e armazenar todos os sistemas de coordenadas de um modelo, bem como, é responsável por manter constantemente atualizadas as transformações necessárias para a passagem de um sistema de referência para outro.

Se nenhuma informação que descreve a pose da câmera em termos do sistema de coordenadas do mundo é publicada, apenas o sistema de referência dado pelo próprio centro ótico das câmeras está disponível. Entretanto se uma mensagem TF que descrevendo a transformação rígida entre os pares é publicada, torna-se possível visualizar os dados em termos de uma nuvem de pontos 3D.

Assim, independentemente da pose que ambas as câmeras ocupam no sistema de coordenadas global, tanto para efeitos de frame de referencia quanto de correspondência geométrica, o processo de calibração deve determinar a posição da segunda câmera em relação ao sistema de coordenadas da primeira (tal qual descrito em 3.2).

Neste *framework* os drivers de câmeras são encapsulados em nodos responsáveis por publicam as imagens capturadas através de tópicos do tipo `sensor_msgs::Image`, juntamente com as informações sobre os seus parâmetros intrínsecos e extrínsecos em um tópico do tipo `sensor_msgs::CameraInfo` do sistema de câmeras.

Um nodo foi especialmente escrito para permitir o controle das mensagens relativa aos parâmetros do sistema de visão nomeado de `Artificial CamInfo`. É através deste elemento do que as matrizes de transformação produzidas pelo (node `Transform_2D_3D`). Os diferentes nodes e tópicos do sistema que desenvolvemos são ilustrados 4.2.

O algoritmo adotado para gerar os mapas de disparidade é o SGBM² disponível na biblioteca OpenCV, que consiste numa ligeira adaptação do método proposto por Hirschmüller (Hirschmüller, 2008).

As principais diferenças do SGBM em relação a versão original do método original são: realiza por padrão passagens simples e considera 5 direções ao invés de 8 com passagens duplas; buscar a correspondência entre blocos de pixels, ao invés de considerá-los individualmente; não implementar a função de custo com informação mútua, mas implementa a técnica de sub-pixel proposta por Birchfield-Tomasi; além de empregar etapas de pré e pós-processamento para a melhoria do *block matching*, tais como interpolação quadrática, verificação de unicidade e filtragem de *speckles*.

Para codificação do processo evolutivo que otimiza a matriz de rotação das câmeras utilizamos a biblioteca GALib (Wall, 1996), que provê um conjunto de classes em C++ que implementa os principais conceitos e estratégias relacionados ao Algoritmos Genéticos.

²*Semi-global Block Matching*

Figura 4.2: .

4.5 Experimentos e Resultados

Os experimentos foram realizados utilizando um dos veículos do LRM em condições reais de uso, simulando a necessidade de executar uma recalibragem do sistema estéreo *in-loco* decorrente de um novo ajuste no *baseline* ou da perda de calibração.

Os dados coletados foram registrados em vídeo em formato compatível com o ROS, propiciando a realização de testes variando condições de iluminação, objetos presentes no cenário, assim como áreas amplas com algumas centenas de metros de visada.

A montagem do sistema de visão empregou pares de câmeras idênticas em cada configuração, procurando combinar diferentes arranjos e *baselines*. A tabela 4.3

Câmera	Dist. focal	Resolução	<i>Baseline</i>	Frame rate	FoV
Flea3	12,5 mm	1920 x 1080	18 cm	15 FPS	21o h, 16o v
Flea3	6 mm	2.8 MPix	18 cm	15 FPS	56o h, 43o vert
Flea3	6 mm	2.8 MPix	36 cm	15 FPS	56o h, 43o vert

Tabela 4.3: Configurações do sistema de visão estéreo utilizado nos experimentos

Os experimentos iniciais procuraram explorar diferentes conjunto de parâmetros para o algoritmo genético relacionado ao tamanho da população, taxas de cruzamento e de mutação, analisando as características de convergência da população e variabilidade oferecidas pelo método. Foram executados 60 testes (*rounds*) no total explorando diferentes combinações dos parâmetros do AG acordo com a tabela 4.4.

Round	Pop	Mut	Cross	Round	Pop	Mut	Cross	Round	Pop	Mut	Cross	Round	Pop	Mut	Cross
#1	10	1%	70%	#16	30	1%	70%	#31	50	1%	70%	#46	100	1%	70%
#2	10	1%	80%	#17	30	1%	80%	#32	50	1%	80%	#47	100	1%	80%
#3	10	1%	90%	#18	30	1%	90%	#33	50	1%	90%	#48	100	1%	90%
#4	10	5%	70%	#19	30	5%	70%	#34	10	5%	70%	#49	100	5%	70%
#5	10	5%	80%	#20	30	5%	80%	#35	10	5%	80%	#50	100	5%	70%
#6	10	5%	90%	#21	30	5%	90%	#36	10	5%	90%	#51	100	5%	70%
#7	10	8%	70%	#22	30	8%	70%	#37	10	8%	70%	#52	100	8%	70%
#8	10	8%	80%	#23	30	8%	80%	#38	10	8%	80%	#53	100	8%	80%
#9	10	8%	90%	#24	30	8%	90%	#39	10	8%	90%	#54	100	8%	90%
#10	10	15%	70%	#25	30	15%	70%	#40	10	15%	70%	#55	100	15%	70%
#11	10	15%	80%	#26	30	15%	80%	#41	10	15%	80%	#56	100	15%	80%
#12	10	15%	90%	#27	30	15%	90%	#42	10	15%	90%	#57	100	15%	90%
#13	10	30%	70%	#28	30	30%	70%	#43	10	30%	70%	#58	100	30%	70%
#14	10	30%	80%	#29	30	30%	80%	#44	10	30%	80%	#58	100	30%	80%
#15	10	30%	90%	#30	30	30%	90%	#45	10	30%	90%	#58	100	30%	90%

Tabela 4.4: Rodadas exploratória dos parâmetros do processo evolutivo. *Experimentos visavam analisar o comportamento do processo evolutivo assim como aspectos de variabilidade e convergência da população para diferentes combinações de parâmetros do AG. Em todas as rodas os testes consideraram as mesmas imagens, seleções de regiões e parâmetros intrínsecos e extrínsecos para o sistema de visão.*

As imagens correspondentes aos melhores indivíduos de cada geração podem ser observados nas figuras 4.3 e 4.4, assim como os gráficos comparativos incluindo múltiplas rodadas para a mesma configuração de parâmetros.

Figura 4.3: Mapas de Disparidade utilizados pela função de fitness na avaliação da qualidade da transformação linear caracterizada por cada indivíduo. *Pertencentes a mesma execução, nota-se o aumento na definição dos elementos do cenário conforme o processo evolui, assim como na melhor caracterização da profundidade em direção ao ponto de fuga. Destaque para a variação gradativa e suave do chão a frente da câmera. As duas últimas imagens são, respectivamente, as disparidades para a melhor calibração e o cenário real capturado pela câmera da esquerda.*

(u) Mapa de disparidade obtido após rotação correspondente ao melhor indivíduo (788,34).

(v) Visão da câmera esquerda, com destaque para as três regiões informadas para a calibração

Figura 4.4: Sequência de Mapas de Disparidades referentes aos melhores indivíduos em uma das execuções do método de proposto, logo após o reajuste do *baseline* para 36 cm. *O valor atribuído pela função de fitness é dado abaixo de cada imagem.*

Figura 4.5: Comparação do fitness médio em cinco rodadas considerando populações com 10 e 30 indivíduos, taxa de mutação fixa de 1% e variando a taxa de crossover em 80% e 90%. *Observa-se que ambas as taxas de crossover oferecem uma evolução global da população, incorporando as boas características dos melhores indivíduos e elevando inclusive o desempenho dos piores.*

Figura 4.6: Comparação do fitness médio em cinco rodadas considerando populações com 10 e 30 indivíduos, taxa de mutação fixa de 5% e variando a taxa de crossover em 80% e 90%. *O aumento na taxa de mutação passa a favorecer melhoras mais frequentes na população, enquanto o grande percentual de crossover permite que tais características sejam incorporadas pela população em poucas gerações.*

Figura 4.7: Comparação do fitness médio em cinco rodadas, considerando populações com 10 e 30 indivíduos, taxa de mutação fixa de 8% e variando a taxa de crossover em 80% e 90%. *Uma taxa de mutação um pouco mais elevada favoreceu o aumento do fitness médio nas populações com menos indivíduos.*

Figura 4.8: Cenário 2.

Considerações finais

A principal contribuição desta tese foi a pesquisa, desenvolvimento, testes e validação de um método para a calibração e ajuste dos parâmetros extrínsecos de câmeras estéreo que não depende do uso de padrões de calibração, o conhecido padrão xadrez para calibração de sistemas de visão. O método proposto é uma abordagem original e baseada em Algoritmos Evolutivos, que obteve um resultado muito bom, se comparado com as técnicas tradicionais. Inclusive a metodologia proposta abre a perspectiva de ser adaptada para outras tarefas de ajuste de parâmetros para a fusão de sensores.

Por outro lado, podemos listar várias outras contribuições originárias desta tese, como por exemplo: o desenvolvimento de uma arquitetura computacional para veículos inteligentes e autônomos (adotada junto aos veículos CaRINA I e II), e que vem sendo constantemente aperfeiçoada; as diversas contribuições no projeto e desenvolvimento da automação e dos sistemas computacionais dos veículos CaRINA I e II; a contribuição na implementação e desenvolvimento de diversos experimentos, muitos em parceria com os demais membros do LRM (Laboratório de Robótica Móvel do ICMC - USP São Carlos), que foram realizados utilizando as plataformas CaRINA I e II; a publicação de diversos artigos científicos em periódicos e conferências nacionais e internacionais de reconhecida qualidade; e por fim, a participação e desenvolvimento de atividades junto ao INCT - SEC (Instituto Nacional de Ciência e Tecnologia em Sistemas Embarcados Críticos) e ao CROB/SC (Centro de Robótica da USP de São Carlos).

Considerando o acima exposto, pode-se afirmar que os principais objetivos desta tese foram alcançados e que dela foram geradas diversas contribuições, quer seja para o grupo de pesquisa do LRM, ou, para o INCT-SEC, para o CROB/SC e para a comunidade científica da área de robótica em geral.

Os resultados das pesquisas e desenvolvimentos desta tese alcançaram seus principais objetivos, entretanto, podemos listar alguns trabalhos futuros a serem desenvolvidos e que podem dar continuidade as pesquisas que foram apresentadas:

- Aperfeiçoamento da arquitetura computacional proposta na tese, visando a incorporação de aspectos referentes a segurança veicular: inclusão de um dispositivo tipo “caixa preta”, inclusão de módulos de segurança na comunicação (autenticação e criptografia na comunicação);
- Desenvolvimento de testes complementares relativos a obtenção dos parâmetros extrínsecos para configurações diferentes de posicionamento de câmeras, visando analisar a robustez do método proposto, por exemplo, considerando câmeras desalinhas;
- Analisar e propor variações da metodologia desenvolvida para a fusão de sensores distintos, como por exemplo, Velodyne e câmera estéreo. O ajuste dos parâmetros de transformação seriam obtidos através do processo evolutivo, porém teria que ser investigada a proposta de novas funções de *fitness* que permitissem um adequado funcionamento do método proposto nesta nova configuração;
- Realizar testes com dados simulados a fim de realizar medidas quantitativas e precisas em relação aos resultados obtidos no mapa de disparidade. O uso de imagens obtidas a partir de simuladores permite um controle total sobre os parâmetros de câmera (posicionamento), bem como um total conhecimento sobre os elementos dispostos no ambiente e suas coordenadas.

Agradecimentos:

Gostaria de deixar aqui registrados os meus sinceros agradecimentos aos órgãos de fomento pelo apoio dado a esta pesquisa, e em particular ao suporte recebido em termos de equipamentos e material de apoio a pesquisa disponibilizados junto ao LRM - ICMC/USP e tendo sido provido pelo MCT, CNPq e FAPESP através principalmente do INCT - SEC (Processos 573963/2008-9 and 08/57870-9).

5.1 Publicações geradas

Essa seção apresenta a lista de publicações geradas até o momento, tanto ligadas diretamente a este trabalho quanto decorrentes dos desdobramentos que tiveram cooperação com as demais pesquisas do grupo.

- FERNANDES, L. C.; DIAS, M. A.; OSÓRIO, F.; WOLF, D. F. A driving assistance system for navigation in urban environments. In: Kuri-Morales, A.; Simari, G., eds. *Ibero-American Conference on Artificial Intelligence (IBERAMIA)*, Bahia Blanca, Argentina: Springer-Verlag Berlin Heidelberg, 2010, p. 542–551 (*Advances in Artificial Intelligence - Lecture Notes in Computer Science (LNCS)*, v.6433).
- SHINZATO, P. Y.; FERNANDES, L. C.; OSÓRIO, F. S.; WOLF, D. F. Path recognition for outdoor navigation using artificial neural networks: Case study. In: *International Conference on Industrial Technology (ICIT)*, Viña del Mar, Chile: IEEE, 2010.
- FERNANDES, L. C.; SOUZA, J. R.; SHINZATO, P. Y.; PESSIN, G.; MENDES, C. C. T.; OSÓRIO, F. S.; WOLF, D. F. Intelligent robotic car for autonomous navigation: Platform and system architecture. In: *II Brazilian Conference on Critical Embedded Systems - CBSEC*, 2012.
- SALES, D. O.; FERNANDES, L. C.; OSÓRIO, F. S.; WOLF, D. F. FSM-based visual navigation for autonomous vehicles. In: *ViCoMoR - IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Algarve, Portugal, 2012.
- FERNANDES, L. C.; SOUZA, J. R.; PESSIN, G.; SHINZATO, P. Y.; SALES, D.; MENDES, C.; PRADO, M.; KLASER, R.; MAGALHÃES, A. C.; HATA, A.; PIGATTO, D.; BRANCO, K. C.; JR., V. G.; OSÓRIO, F. S.; WOLF, D. F. CaRINA Intelligent Robotic Car: Architectural design and applications. *Journal of Systems Architecture*, v. 60, n. 4, p. 372 – 392, 2014.
- SALES, D. O.; CORREA, D. O.; FERNANDES, L. C.; WOLF, D. F.; OSÓRIO, F. S. Adaptive finite state machine based visual autonomous navigation system. *Engineering Applications of Artificial Intelligence*, v. 29, p. 152 – 162, 2014.

Referências Bibliográficas

- AUTONOMOS Autonomos labs. <http://www.autonomos-labs.de/>, último acesso em Agosto de 2014, 2006.
- BERTOZZI, M.; BROGGI, A.; FASCIOLI, A. Vision-based intelligent vehicles: State of the art and perspectives. *Robotics and Autonomous Systems*, v. 32, n. 1, p. 1 – 16, 2000.
- BONIN-FONT, F.; ORTIZ, A.; OLIVER, G. Visual navigation for mobile robots: A survey. *Journal of Intelligent and Robotic Systems*, v. 53, n. 3, p. 263 – 296, 2008.
- BROWN, D. C. Close-range camera calibration. *Photogrammetric Engineering*, v. 37, n. 8, p. 855–866, 1971.
- BUENO, S. S.; AZEVEDO, H.; MIRISOLA, L. G. B.; PAIVA, E. C.; RAMOS, J. J. G. Uma plataforma para pesquisa e desenvolvimento em robótica terrestre de exterior. In: *IX Simpósio Brasileiro de Automação Inteligente*, Brasília - DF, 2009.
- CARINA Projeto carina – carro robótico inteligente para navegação autônoma. <http://www.lrm.icmc.usp.br/>, último acesso em Agosto de 2014, 2009.
- CRISMAN, J.; THORPE, C. Unscarf, a color vision system for the detection of unstructured roads. In: *Proceedings of IEEE International Conference on Robotics and Automation*, 1991, p. 2496–2501.

- CRISMAN, J.; THORPE, C. Scarf: a color vision system that tracks roads and intersections. *Robotics and Automation, IEEE Transactions on*, v. 9, n. 1, p. 49–58, 1993.
- DARPA Grand challenge '05. <http://archive.darpa.mil/grandchallenge05/index.html>, último acesso em Agosto de 2014, 2005.
- DARPA Urban challenge. <http://archive.darpa.mil/grandchallenge/index.html>, último acesso em Agosto de 2014, 2007.
- DENATRAN; IPEA *Impactos sociais e econômicos dos acidentes de trânsito nas rodovias brasileiras*. Relatório Técnico, Instituto de Pesquisa Econômica Aplicada (IPEA) e Departamento Nacional de Trânsito (DEANTRAN), 2006.
- DESOUZA, G. N.; KAK, A. C. Vision for mobile robot navigation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 2, p. 237–267, 2002.
- DICKMANNS, E.; BEHRINGER, R.; DICKMANNS, D.; HILDEBRANDT, T.; MAURER, M.; THOMANEK, F.; SCHIEHLEN, J. The seeing passenger car 'vampers-p'. In: *Intelligent Vehicles '94 Symposium, Proceedings of the*, 1994, p. 68–73.
- DICKMANNS, E. D. Vehicles capable of dynamic vision: a new breed of technical beings? *Artificial Intelligence*, v. 103, n. 1–2, p. 49–76, 1998.
- FENABRAVE *Anuário*. Relatório Técnico, Federação Nacional de Distribuição de Veículos Automotores, 2008.
- FERNANDES, L. C.; DIAS, M. A.; OSÓRIO, F.; WOLF, D. F. A driving assistance system for navigation in urban environments. In: KURI-MORALES, A.; SIMARI, G., eds. *Ibero-American Conference on Artificial Intelligence (IBERAMIA)*, Bahia Blanca, Argentina: Springer-Verlag Berlin Heidelberg, 2010, p. 542–551 (*Advances in Artificial Intelligence - Lecture Notes in Computer Science (LNCS)*, v.6433).
- FERNANDES, L. C.; SOUZA, J. R.; PESSIN, G.; SHINZATO, P. Y.; SALES, D.; MENDES, C.; PRADO, M.; KLASER, R.; MAGALHÃES, A. C.; HATA, A.; PIGATTO, D.; BRANCO, K. C.; JR., V. G.; OSÓRIO, F. S.; WOLF, D. F. Carina intelligent robotic car: Architectural design and applications. *Journal of Systems Architecture*, v. 60, n. 4, p. 372–392, 2014.

- FERNANDES, L. C.; SOUZA, J. R.; SHINZATO, P. Y.; PESSIN, G.; MENDES, C. C. T.; OSÓRIO, F. S.; WOLF, D. F. Intelligent robotic car for autonomous navigation: Platform and system architecture. In: *II Brazilian Conference on Critical Embedded Systems - CBSEC*, 2012.
- FREYMAN, R. Driver assistance technology to enhance traffic safety. In: ULBRICH, H.; GINZINGER, L., eds. *Motion and Vibration Control*, Springer Netherlands, p. 71–81, 2009.
- GALLUP, D.; FRAHM, J.-M.; MORDOHA, P.; POLLEFEYS, M. Variable baseline/resolution stereo. In: *IEEE Conference on Computer Vision and Pattern Recognition - CVPR*, 2008, p. 1 – 8.
- GONCALVES, L. M.; DE ALMEIDA, R. M. A.; DE MELLO HONORIO, L. Modelagem de busca de rotas para a navegação local de um veículo autônomo. In: *XVIII Congresso Brasileiro de Automática*, Bonito-MS, Brazil, 2010, p. 2073–2079.
- GOOGLE Google self-driving cars project. <https://plus.google.com/+GoogleSelfDrivingCars/about>, último acesso em Agosto de 2014, 2009.
- HABERMANN, D.; HATA, A.; WOLF, D. F.; OSORIO, F. S. Artificial neural nets object recognition for 3d point clouds. In: *Intelligent Systems (BRACIS), 2013 Brazilian Conference on*, Fortaleza, Brasil: IEEE, 2013, p. 101 – 106.
- HARTLEY, R.; ZISSERMAN, A. *Multiple view geometry in computer vision*. Second ed. Cambridge University Press, ISBN: 0521540518, 2003.
- HATA, A.; HABERMANN, D.; WOLF, D. F.; OSORIO, F. S. Crossroad detection using artificial neural networks. In: ILIADIS, L.; PAPADOPOULOS, H.; JAYNE, C., eds. *Engineering Applications of Neural Networks (EANN 2013)*, Halkidiki, Greece: Springer Berlin Heidelberg, 2013, p. 112 – 121.
- HIRSCHMÜLLER, H. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 30, n. 2, p. 328–341, 2008.
- KLASER, R. L. *Navegação de veículos autônomos em ambientes externos não estruturados baseada em visão computacional*. Dissertação de Mestrado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (USP), São Carlos, São Paulo, 2014.

- LEIGHTY, R. D. *Darpa alv (autonomous land vehicle) summary*. Relatório Técnico, Defense Advanced Research Projects Agency, 1986.
- DE LIMA, D. A. *Navegação segura de um carro autônomo utilizando campos vetoriais e o método da janela dinâmica*. Dissertação de Mestrado, Universidade Federal de Minas Gerais, 2010.
- MILELLA, A.; REINA, G.; FOGLIA, M. M. A multi-baseline stereo system for scene segmentation in natural environments. In: *Technologies for Practical Robot Applications (TePRA)*, IEEE Computer Society, 2013.
- MONTEMERLO, M.; BECKER, J.; BHAT, S.; DAHLKAMP, H.; DOLGOV, D.; ETTINGER, S.; HAEHNEL, D.; HILDEN, T.; HOFFMANN, G.; HUHNKE, B.; JOHNSTON, D.; KLUMPP, S.; LANGER, D.; LEVANDOWSKI, A.; LEVINSON, J.; MARCIL, J.; ORENSTEIN, D.; PAEFGEN, J.; PENNY, I.; PETROVSKAYA, A.; PFLUEGER, M.; STANEK, G.; STAVENS, D.; VOGT, A.; THRUN, S. Junior: The stanford entry in the urban challenge. *Journal of Field Robotics*, v. 25, p. 569–597, 2008.
- POMERLEAU, D. A. *Alvinn: An autonomous land vehicle in a neural network*. Relatório Técnico, Carnegie-Mellon University, 1989.
Disponível em <http://handle.dtic.mil/100.2/ADA218975>
- ROVIRA-MÁS, F.; WANG, Q.; ZHANG, Q. Bifocal stereoscopic vision for intelligent vehicles. *International Journal of Vehicular Technology*, v. 2009, p. 1 – 9, 2009.
- SALES, D. O.; CORREA, D. O.; FERNANDES, L. C.; WOLF, D. F.; OSORIO, F. S. Adaptive finite state machine based visual autonomous navigation system. *Engineering Applications of Artificial Intelligence*, v. 29, p. 152 – 162, 2014.
- SALES, D. O.; FERNANDES, L. C.; OSORIO, F. S.; WOLF, D. F. Fsm-based visual navigation for autonomous vehicles. In: *ViCoMoR - IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Algarve, Portugal, 2012.
- SENA Sistema embarcado de navegação autônoma.
<http://www3.eesc.usp.br/sena/url/pt/index.php>, último acesso em Agosto de 2014, 2009.
- SHINZATO, P. Y.; FERNANDES, L. C.; OSÓRIO, F. S.; WOLF, D. F. Path recognition for outdoor navigation using artificial neural networks: Case study. In: *International Conference on Industrial Technology (ICIT)*, Viña del Mar, Chile: IEEE, 2010.

- SHINZATO, P. Y.; WOLF, D. F. A road following approach using artificial neural networks combinations. *Journal of Intelligent & Robotic Systems*, v. 62, p. 527–546, 10.1007/s10846-010-9463-2, 2011.
Disponível em <http://dx.doi.org/10.1007/s10846-010-9463-2>
- THORPE, C.; HEBERT, M.; KANADE, T.; SHAFER, S. Vision and navigation for the carnegie-mellon navlab. *Annual Review of Computer Science*, v. 2, n. 1, p. 521–556, 1987.
- THRUN, S. Toward robotic cars. *Communication of the ACM*, v. 53, n. 4, p. 99–106, 2010.
- THRUN, S.; MONTEMERLO, M.; DAHLKAMP, H.; STAVENS, D.; ARON, A.; DIEBEL, J.; FONG, P.; GALE, J.; HALPENNY, M.; HOFFMANN, G.; LAU, K.; CELIAOAKLEY; PALATUCCI, M.; PRATT, V.; STANG, P.; STROHBAND, S.; DUPONT, C.; JENDROSSEK, L.-E.; KOELEN, C.; MARKEY, C.; RUMMEL, C.; VAN NIEKERK, J.; JENSEN, E.; ALESSANDRINI, P.; BRADSKI, G.; DAVIES, B.; ETTINGER, S.; KAEHLER, A.; NEFIAN, A.; MAHONEY, P. Stanley: The robot that won the darpa grand challenge. *Journal of Field Robotics*, v. 23, p. 661–69, 2006.
- URMSON, C.; ANHALT, J.; BAGNELT, D.; BAKER, C.; BITTNER, R.; CLARK, M. N.; DOLAN, J.; DUGGINS, D.; GALATALI, T.; GEYER, C.; GITTLEMAN, M.; HARBAUGH, S.; HEBERT, M.; HOWARD, T. M.; KOLSKI, S.; KELLY, A.; LIKHACHEV, M.; MCNAUGHTON, M.; MILLER, N.; PETERSON, K.; PILNICK, B.; RAJKUMAR, R.; RYBSKI, P.; SALESKY, B.; SEO, Y.-W.; SINGH, S.; SNIDER, J.; STENTZ, A.; WITTAKER, W. R.; WOLKIWICHI, Z.; ZIGLAR, J.; BAE, H.; BROWN, T.; DEMITHISH, D.; LITKOUHI, B.; NICKOLAOU, J.; SADEKAR, V.; ZHANG, W.; STRUBLE, J.; TAYLOR, M.; DARMS, M.; FERGUSON, D. Autonomous driving in urban environments: Boss and the urban challenge. *Journal of Field Robotics*, v. 25, p. 425–466, 2008.
- VISLAB Artificial vision and intelligent systems laboratory of parma university. <http://vislab.it/>, último acesso em Agosto de 2014, 2009.
- WALL, M. *Galib: A c++ library of genetic algorithm components*. Massachusetts Institute of Technology, versão 2.4, 1996.
Disponível em <http://lancet.mit.edu/ga/>
- WILLIAMS, M. The prometheus programme. In: *Towards Safer Road Transport - Engineering Solutions, IEE Colloquium on*, 1992, p. 4/1–4/3.

ZHANG, S., ed. *Handbook of 3d machine vision: Optical metrology and imaging*. Series in Optics and Optoelectronics, 1st ed. Bristol, PA, USA: Taylor & Francis, Inc., 2013.