

UNIVERSIDADE FEDERAL DO RIO GRANDE
Centro de Ciências Computacionais
Programa da Pós-Graduação em Computação
Curso de Mestrado em Engenharia de Computação

DISSERTAÇÃO DE MESTRADO

ALEXSANDER VASCONCELOS CANEZ

**DETECÇÃO DE COLISÃO COM VISÃO MONOCULAR PARA
AUXILIAR NA ORIENTAÇÃO E MOBILIDADE
DE PESSOAS COM DEFICIÊNCIA VISUAL**

RIO GRANDE, 2018.

ALEXSANDER VASCONCELOS CANEZ

**DETECÇÃO DE COLISÃO COM VISÃO MONOCULAR PARA
AUXILIAR NA ORIENTAÇÃO E MOBILIDADE
DE PESSOAS COM DEFICIÊNCIA VISUAL**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande, como requisito parcial para a obtenção do grau de Mestre em Engenharia de Computação.

Orientadora:
Prof^a. Dr^a. Regina Barwaldt

Coorientador:
Prof. Dr. Ricardo Nagel Rodrigues

RIO GRANDE, 2018.

Ficha catalográfica

C221d Canez, Alexander Vasconcelos.
Detecção de colisão com visão monocular para auxiliar na
orientação e mobilidade de pessoas com deficiência visual /
Alexander Vasconcelos Canez. – 2018.
93 f.

Dissertação (mestrado) – Universidade Federal do Rio Grande –
FURG, Programa de Pós-Graduação em Computação, Rio Grande/RS,
2018.
Orientadora: Dra. Regina Barwaldt.
Coorientador: Dr. Ricardo Nagel Rodrigues.

1. Deficiência Visual 2. Tecnologia Assistiva 3. Visão Monocular
4. Tempo de Colisão 5. Detecção de Colisões I. Barwaldt, Regina
II. Rodrigues, Ricardo Nagel III. Título.

CDU 004-056.262



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

DISSERTAÇÃO DE MESTRADO

**DETECÇÃO DE COLISÃO COM VISÃO MONOCULAR PARA
AUXILIAR NA ORIENTAÇÃO E MOBILIDADE
DE PESSOAS COM DEFICIÊNCIA VISUAL**

ALEXSANDER VASCONCELOS CANEZ

Banca examinadora:

Prof.ª. Dr.ª. Raquel de Miranda Barbosa

Prof. Dr. Vagner Santos da Rosa

Prof.ª. Dr.ª. Regina Barwaldt
Orientadora
Prof. Dr. Ricardo Nagel Rodrigues
Coorientador

AGRADECIMENTOS

A minha esposa Magda, por todo incentivo, amor e companheirismo. A minha mãe por todo apoio e amor incondicional e por sempre ter incentivado a minha educação para que eu pudesse realizar os meus objetivos. Ao meu pai (*in memoriam*) por sempre ter me apoiado com orgulho. Aos meus irmãos pelo apoio, carinho, confiança e incentivo. Aos colegas e amigos do Centro de Ciências Computacionais, por qualquer tipo de apoio durante o mestrado. Aos meus orientadores, Prof. Dr. Ricardo Nagel Rodrigues e Prof^a. Dr^a. Regina Barwaldt, pela confiança em mim depositada, pelos ensinamentos, pelo apoio, estímulo e ajuda no desenvolvimento deste trabalho. A banca examinadora dessa dissertação, por ter aceitado participar da defesa e pelas sugestões de melhorias que deram. Aos demais professores que contribuíram para minha formação.

RESUMO

Esta dissertação apresenta detalhes de um algoritmo que foi desenvolvido e implementado para estimar o Tempo de Colisão entre um usuário e um obstáculo. Consiste em um método para detectar colisões com visão monocular por meio da câmera de um *smartphone*. A detecção de colisão em um vídeo adquirido por câmera monocular é um problema desafiador com aplicações que vão desde robótica, navegação humana e automação. A aplicação alvo é um sistema de Tecnologia Assistiva para auxiliar pessoas com deficiência visual para evitar a colisão com obstáculos enquanto se locomovem em um ambiente, auxiliando em sua orientação e mobilidade, assim garantindo maior segurança e autonomia para essas pessoas, no entanto, a integração com o sistema será uma proposta futura. No entanto, o método poderá ser aplicado em diferentes áreas. O método apresentado é baseado em um novo algoritmo para estimar o Tempo de Colisão que segue as seguintes etapas: i) aquisição da sequência de imagens por uma câmera monocular, ii) pré-processamento onde as imagens são redimensionadas e é aplicado um filtro para reduzir ruídos, iii) fluxo óptico para estimar o movimento, iv) Tempo de colisão, a partir do modelo de câmera pinhole uma equação analítica é desenvolvida para estimar esse tempo para cada ponto no plano da imagem e v) técnicas de processamento de imagem para reduzir possíveis erros. O algoritmo foi implementado na linguagem C/C++, com a biblioteca de Visão Computacional OpenCV. Para avaliar o método foram realizados três experimentos em diferentes cenários, permitindo uma análise de seu desempenho. A avaliação da performance em relação à detecção de colisões, foi dada em termos de precisão, revocação e acurácia; e a validação do Tempo de Colisão, por um erro médio absoluto entre o tempo estimado e um *ground truth*.

Palavras-chaves: Deficiência visual, Tecnologia Assistiva, Visão Monocular, Tempo de Colisão, Detecção de colisões.

ABSTRACT

This dissertation presents details of an algorithm that was developed and implemented to estimate the Time To Collision between a user and an obstacle. It consists of a method to detect collisions with monocular vision through the camera of a smartphone. However, the method can be applied in different areas. Collision detection on a video acquired by a monocular camera is a challenging problem with applications ranging from robotics, human navigation and automation. The target application is an Assistive Technology system to assist visually impaired people to avoid collision with obstacles while moving around in an environment, helping their orientation and mobility, thus guaranteeing greater security and autonomy for these people, however, integration with the system will be a future proposal. The method presented is based on a new algorithm to estimate the Time To Collision that follows the following steps: i) acquisition of the sequence of images by a monocular camera, ii) preprocessing where the images are resized and a filter is applied to reduce noise, iii) optical flow to estimate motion, iv) Time To Collision, from the pinhole camera model an analytical equation is developed to estimate this time for each point and v) image processing techniques to reduce possible errors. The algorithm was implemented in the C / C ++ language, with the OpenCV Computer Vision library. To evaluate the method were performed three experiments in different scenarios, allowing an analysis of their performance. The evaluation of performance in relation to the detection of collisions, was given in terms of accuracy, recall and accuracy; and the Time To Collision validation, by an absolute mean error between the estimated time and a *ground truth*.

Keywords: Visual impairment, Assistive Technology, Monocular Vision, Time To Collision, Collision Detection

LISTA DE FIGURAS

Figura 1.1: Ilustração da futura aplicação do método desenvolvido.	12
Figura 2.1: Porcentagem da população com deficiência em 2010.	18
Figura 2.2: Causas de Cegueira e Deficiência Visual no Mundo (%).	20
Figura 2.3: Estrutura básica do olho humano.	21
Figura 2.4: Campo visual humano. (a) campo de visão monocular. (b) campo de visão binocular.	22
Figura 2.5: As cinco técnicas de Mobilidade de uma pessoa cega.	23
Figura 3.1: Analogia entre olho humano e uma câmera.	27
Figura 3.2: Par de câmeras posicionadas paralelamente, direcionadas para uma mesma cena.	28
Figura 3.3: Câmera monocular com seu campo visual em cinza.	29
Figura 4.1: Alguns exemplos de TA para cegos.	33
Figura 4.2: A bengala eletrônica alerta sobre obstáculos acima da linha da cintura.	35
Figura 4.3: Bengala eletrônica. (a) Esquema básico funcionamento. (b) Protótipo.	35
Figura 4.4: Ilustração de um deficiente visual em rota de colisão e o protótipo do óculos.	36
Figura 4.5: Partes constituintes do sistema.	36
Figura 4.6: Usuário com o protótipo do sistema Drishti em um ambiente externo.	37
Figura 4.7: A proposta do sistema baseado em RFID.	38
Figura 4.8: Sistema de navegação por luz visível.	39
Figura 4.9: Sistema com visão estéreo e sensores de ultrassom.	40
Figura 4.10: Prévia do protótipo.	41
Figura 4.11: Detalhes do experimento de Ereno.	41
Figura 4.12: Teste do protótipo do sistema para smartphone.	42
Figura 4.13: Fluxo de funcionamento do sistema.	42
Figura 5.1: Classificação dos algoritmos para detecção de obstáculos com visão monocular.	47
Figura 5.2: Esquema do vetor de fluxo óptico para detecção de obstáculos.	48
Figura 5.3: Abordagem de Lee - (a) robô em rota de colisão; (b) robô fora da rota de colisão.	49
Figura 5.4: Geometria óptica para o TTC.	50
Figura 6.1: Fluxograma básico do algoritmo de detecção de colisões.	60
Figura 6.2: Geometria do modelo de câmera pinhole.	61

Figura 6.3: Ilustração da relação entre os pontos p_l e p_a 66

Figura 6.4: Ilustração dos pontos p_l e p_a considerando o offset. 67

Figura 7.1: Exemplos de imagens do experimento 1..... 73

Figura 7.2: Gráfico *TTC* estimado e *ground truth* de um dos vídeos, para todos os *frames*. 74

Figura 7.3: Erro médio para um dos vídeos do primeiro experimento (em segundos). 75

Figura 7.4: Gráfico do *TTC* real x *TTC* estimado para os instantes de 1, 2 e 3 segundos. 76

Figura 7.5: Exemplos de imagens do experimento 2..... 77

Figura 7.6: Gráfico do melhor caso para o Grupo 1..... 79

Figura 7.7: Gráfico do pior caso para o Grupo 1..... 79

Figura 7.8: Gráfico do melhor caso para o Grupo 2..... 80

Figura 7.9: Gráfico do pior caso para o Grupo 2..... 80

Figura 7.10: Exemplos de imagens do experimento 3. 81

LISTA DE TABELAS

Tabela 4.1: Quadro comparativo da classificação de TA	31
Tabela 4.2: Resumo das TA para deficientes visuais apresentadas	44
Tabela 7.1: Erros mínimo, máximo e médio o experimento	75
Tabela 7.2: Resultados para o Grupo 1	78
Tabela 7.3: Resultados para o Grupo 2	78
Tabela 7.4: Resultados para todos os vídeos	80
Tabela 7.5: Resultados para o experimento 3	82
Tabela 7.6: Performance e validação do algoritmo	83

LISTA DE ABREVIATURAS

2D	Duas dimensões
3D	Três dimensões
ACAS	Active Contour Affine Scale
ADA	American with Disabilities Act
CAT	Comitê de Ajudas Técnicas
DMRI	Degeneração Macular Relacionada à Idade
EUA	Estados Unidos da América
FN	False Negative
FoE	Focus of Expansion
FP	False Positive
FPS	Frames por Segundo
GIS	Sistema de Informação Geográfica
GPOD	Ground Plane Obstacle Detection
GPS	Sistema de Posicionamento Global
IBD	Image Brightness Derivatives
IBGE	Instituto Brasileiro de Geografia e Estatística
IMU	Inertial Measurement Unit
LBS	Location-Based System
LED	Light Emitting Diode
OMS	Organização Mundial da Saúde
OpenCV	Open Source Computer Vision
RFID	Radio Frequency Identification
SEDH/PR	Secretaria Especial dos Direitos Humanos da Presidência da República
SIRS	Scale Invariant Ridge Segments
SLAM	Simultaneous Localization and Mapping
TA	Tecnologia Assistiva
TN	True Negative
TP	True Positive
TTC	Time-To-Collision
TTS	Text-to-Speech
USB	Universal Serial Bus
VLC	Visible Light Communication
WHO	World Health Organization

SUMÁRIO

CAPÍTULO 1 – INTRODUÇÃO	11
1.1. Introdução	11
1.2. Motivação	13
1.3. Objetivos.....	14
1.4. Metodologia.....	15
1.5. Estrutura da dissertação	16
CAPÍTULO 2 – DEFICIÊNCIA VISUAL.....	17
2.1. Introdução	17
2.2. Deficiência visual e principais causas	19
2.3. Visão humana	21
2.4. Orientação e Mobilidade	22
2.5. Considerações finais do capítulo	24
CAPÍTULO 3 – VISÃO COMPUTACIONAL.....	25
3.1. Introdução	25
3.2. Visão Computacional	26
3.2.2. Biblioteca OpenCV	26
3.3. Sistema de visão por computador	26
3.3.1. Visão Estéreo	27
3.3.2. Visão Monocular.....	28
3.4. Considerações finais do capítulo	29
CAPÍTULO 4 – TECNOLOGIA ASSISTIVA	30
4.1. Introdução	30
4.2. Tecnologia Assistiva para deficientes visuais	32
4.2.1. Métodos computacionais de Tecnologia Assistiva para deficientes visuais.....	34

4.2.1.1. Apoio a para deficientes visuais com informações não contextualizadas	34
4.2.1.2. Apoio a navegação para deficientes visuais com informações contextualizadas	39
4.3. Considerações finais do capítulo	43
CAPÍTULO 5 – TEMPO DE COLISÃO COM VISÃO MONOCULAR	46
5.1. Introdução	46
5.2. Algoritmos para detecção de obstáculos com visão monocular	46
5.2.1. Método baseado no fluxo óptico	47
5.3. Técnicas para estimar o TTC por Visão Monocular	48
5.4. Considerações finais do capítulo	58
CAPÍTULO 6 – O ALGORITMO PARA DETECÇÃO DE COLISÕES	59
6.1. Introdução	59
6.2. Etapas do algoritmo	60
6.3. Considerações finais do capítulo	71
CAPÍTULO 7 – EXPERIMENTOS E RESULTADOS	72
7.1. Introdução	72
7.2. Primeiro Experimento	72
7.3. Segundo Experimento	77
7.4. Terceiro experimento	81
7.5. Performance e validação	82
7.6. Considerações finais do capítulo	84
CAPÍTULO 8 – CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	86
REFERÊNCIAS	88

CAPÍTULO 1 – INTRODUÇÃO

1.1. Introdução

Para a maioria da população, a noção do espaço físico no qual cada um se situa é algo bastante simples de ser percebido, porém, muitas vezes pode não ser algo tão trivial, principalmente para uma parcela significativa da sociedade que possui dificuldades relacionadas à percepção desse espaço. Essa parcela é representada principalmente pelos deficientes visuais, entre eles os que possuem cegueira ou baixa visão.

Além disso, devido a essa dificuldade de percepção, são inúmeros os problemas e obstáculos que os deficientes visuais podem encontrar durante a locomoção em seu dia-a-dia, como postes, telefones públicos, placas de sinalização, caçambas de entulho, buracos, escadas, mesas, pedestres, veículos, entre outros, que podem estar em seu caminho podendo colocar em risco sua integridade devido a algum tipo de colisão.

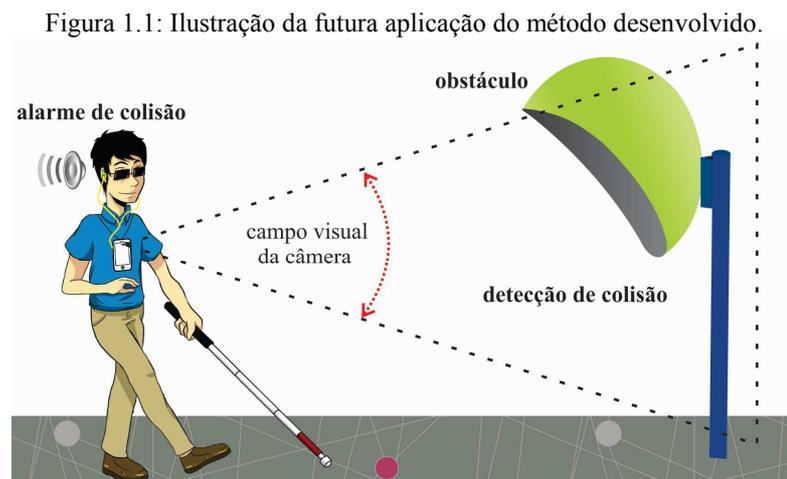
De acordo com dados da OMS - Organização Mundial da Saúde (*WHO - World Health Organization*), em pesquisa divulgada em agosto de 2012, o número de pessoas com deficiência visual no Mundo é de aproximadamente 285 milhões, sendo 246 milhões com perda parcial (visão subnormal) e 39 milhões com perda total da visão (cegos) (WHO, 2014). No Brasil, de acordo com o IBGE (Instituto Brasileiro de Geografia e Estatística) o Censo de 2010 apontou que a deficiência visual totalizou aproximadamente 35,8 milhões de pessoas, onde 6,6 milhões de pessoas com grave comprometimento visual (cegueira ou visão subnormal), dos quais 506,3 mil eram cegos (IBGE, 2012). Dessa forma, torna-se evidente que esta parcela significativa da população merece e necessita de mais atenção, respeito e cuidados.

Existem algumas maneiras para facilitar a vida dessas pessoas, por exemplo, com a inserção ou aumento de políticas públicas de acessibilidade e de inclusão social ou com Tecnologia Assistiva (TA). A TA são recursos e serviços que contribuem para proporcionar ou ampliar habilidades funcionais de pessoas com deficiência, e visam proporcionar maior independência, qualidade de vida e inclusão social, através da ampliação de sua comunicação, mobilidade, controle de seu ambiente, habilidades de seu aprendizado, trabalho e integração com a família, amigos e sociedade (BERSCH, 2013).



Nesse contexto, o presente trabalho faz parte do desenvolvimento de um sistema computacional de TA que será utilizado para auxiliar essas pessoas em algumas atividades cotidianas, como por exemplo, o desvio de obstáculos. Desse modo, esse projeto visa unir as áreas de robótica e visão computacional com a área de TA.

Entretanto, o objetivo desta dissertação não consiste em desenvolver e/ou detalhar o projeto como um todo. O escopo deste estudo aborda apenas uma das funcionalidades do projeto de TA para deficientes visuais, a detecção de colisões. A detecção é dada a partir da estimação do Tempo de Colisão (*Time-To-Collision - TTC*) entre o deficiente visual (câmera) e algum obstáculo em seu caminho. Então, o algoritmo desenvolvido poderá ser integrado ao sistema TA. Cabe ressaltar que para este projeto, o algoritmo desenvolvido não foi implementado no sistema, ficando essa tarefa como trabalho futuro. A Figura 1.1 ilustra um esboço da futura aplicação do método desenvolvido.



Fonte: o autor.

Ainda sobre o *TTC*, ele é o tempo que um observador saindo de um ponto 'A' leva para atingir um obstáculo localizado em um ponto 'B'. Alguns autores também o chamam de Tempo de Contato ou Tempo de Impacto, como por exemplo: Alenyà et al. (2009), Camus (1995), Müller et al. (2009), Sagrebin et al. (2008), Watanabe et al. (2015). Conhecer esse tempo é uma tarefa importante para desviar de obstáculos, pois a partir dele, pode-se tomar as decisões para impedir colisões. Um dos primeiros estudos sobre o assunto foi proposto por Lee (1976), onde definiu o *TTC* como sendo a distância entre um observador e um obstáculo (também conhecida como profundidade - Z) dividida pela velocidade constante (V) de deslocamento do observador, conforme a Equação (1.1). Porém, com a evolução das técnicas computacionais e a necessidade em obter resultados mais precisos, o método foi aprimorado e novos algoritmos têm sido desenvolvidos.

$$TTC = \frac{Z}{V} \quad (1.1)$$

Na literatura diferentes pesquisas vêm sendo realizadas para estimar o *TTC*. Muitos, por exemplo, utilizam sensores laser ou ultrassom para adquirir subsídios relevantes e obter uma estimativa desse tempo; outros, trazem técnicas de visão computacional que se baseiam no tratamento de imagens adquiridas por sistemas de visão por câmera.

Dentre as técnicas de visão computacional cita-se o fluxo óptico que possui diversas aplicações e pode auxiliar em várias tarefas importantes, tais como segmentação de regiões ou bordas, navegação de robôs, interpretação de cena, rastreamento de objetos, reconstrução 3D, codificação de vídeo, detecção de colisões.

Neste contexto, o estudo exposto na presente dissertação visa apresentar pontos considerados relevantes de um método que foi desenvolvido para detectar obstáculos. Foi implementado um algoritmo para estimar o *TTC* unicamente por visão monocular, utilizando técnicas de visão computacional para identificar obstáculos com potencial de risco ao usuário.

Para avaliar e validar o algoritmo foram realizados três experimentos, em condições e ambientes diferentes. A partir desses experimentos foi possível determinar a performance do algoritmo em termos de precisão, revocação e acurácia, possibilitando avaliar o método em relação à detecção de colisões. Também pode-se determinar um erro em segundos entre o tempo real para a colisão e o *TTC* calculado, possibilitando validar em relação ao *TTC*.

O algoritmo foi desenvolvido no ambiente QT Creator, em linguagem C/C++ e com a biblioteca de visão computacional OpenCV (*Open Source Computer Vision*) que possibilita obter respostas em tempo real. Vale destacar que processamento em tempo real é algo imprescindível quando trata-se de tarefas que necessitem de respostas rápidas, como é o caso do cálculo do *TTC* e conseqüentemente a detecção de obstáculos.

1.2. Motivação

De acordo com os dados do Censo de 2010, o Brasil tem uma população residente de 190.755.799 de pessoas, sendo que 23,9% deste total tem, pelo menos, algum dos tipos de deficiência (física, auditiva, visual ou mental). Isso significa que 45.623.910 brasileiros possuem alguma deficiência. O número de pessoas com deficiência visual totalizou

aproximadamente 35,8 milhões, onde 6,6 milhões de pessoas com grave comprometimento visual (cegueira ou visão subnormal), dos quais 506,3 mil eram cegos (IBGE, 2012).

Além disso, o dia a dia de uma pessoa com deficiência não é uma tarefa que pode ser considerada fácil. Como é o caso dos deficientes visuais que, entre outras dificuldades, enfrentam problemas durante sua locomoção, pois os riscos de colisão com obstáculos são constantes. Dessa forma, existem algumas técnicas para tentar evitar colisões, essas são apresentadas por Felipe (2001) que comenta como um deficiente visual pode se locomover: (i) Com ajuda de outra pessoa - guia vidente; (ii) Usando seu próprio corpo - autoproteções; (iii) Usando uma bengala - bengala longa; (iv) Usando um animal - cão-guia; (v) Usando a tecnologia - algum aparato eletrônico.

Devido ao grande número de pessoas com deficiência visual aliado a dificuldades de locomoção encontradas por elas, a principal motivação deste trabalho consiste em desenvolver um método que pode ser utilizado para minimizar os riscos de colisão com obstáculos; que, quando implementado no sistema TA, irá proporcionar maior autonomia, independência e confiança para essas pessoas, possibilitando sua locomoção com maior segurança.

1.3. Objetivos

- **Objetivo geral**

Este trabalho tem como objetivo geral desenvolver e implementar um algoritmo para detectar obstáculos a partir do *TTC*.

- **Objetivos específicos**

Os seguintes objetivos específicos foram traçados:

- ✓ Estudar algoritmos da biblioteca OpenCV;
- ✓ Detectar obstáculos com visão monocular;
- ✓ Desenvolver e implementar um algoritmo para estimar o *TTC*;
- ✓ Desenvolver um algoritmo com resposta em tempo real;
- ✓ Avaliar e validar o modelo desenvolvido.

1.4. Metodologia

A metodologia dessa pesquisa se resume nas seguintes tarefas:

- **Revisão bibliográfica sobre alguns assuntos pertinentes ao tema proposto:** foi feita uma revisão na literatura sobre alguns aspectos considerados importantes ao tema desta dissertação. Estes assuntos podem ser divididos em deficiência visual (visão geral sobre deficiência visual), TA (conceitos e tipos de TA, dando ênfase às que utilizam técnicas computacionais para deficientes visuais).
- **Revisão bibliográfica sobre algumas técnicas para estimar o TTC:** foram pesquisados na literatura métodos para determinar o TTC que não utilizam visão monocular, bem como métodos para estimação desse tempo a partir da visão monocular.
- **Desenvolvimento e implementação e validação do algoritmo:** esta tarefa consiste nos passos a seguir.
 - ✓ Testar e selecionar um método de fluxo óptico que atenda à tarefa proposta - Nesta etapa foram analisados alguns algoritmos para o cálculo de fluxo óptico, sendo que alguns foram implementados e testados para verificar qual apresenta melhor desempenho. A partir daí foi selecionado o mais adequado para aplicações em tempo real, de modo a satisfazer o modelo desenvolvido.
 - ✓ Detectar objetos no ambiente a partir do fluxo óptico – Com o método de fluxo óptico selecionado foi implementado um algoritmo para detecção de obstáculos.
 - ✓ Encontrar uma métrica para estimar o TTC e o risco de colisão com objetos – Com base no cálculo do fluxo óptico, foram pesquisadas métricas para estimar o TTC utilizando apenas visão monocular. Após a pesquisa, verificou-se que o modelo proposto seria viável. Então, a partir do modelo de uma câmera pinhole¹, desenvolveu-se uma equação para estimar esse tempo. Dessa forma, a partir dos valores desse tempo, foi possível determinar quais objetos estão mais próximos e assim determinar o risco de uma possível colisão.

¹ “... se baseia na análise física do comportamento óptico das câmaras escuras. A câmara escura consiste em uma caixa preta com um pequeno orifício em um dos seus lados. No lado oposto é formada a imagem invertida da cena à frente da pequena abertura.” (Aires, 2010 p. 9).

Disponível em: <http://pee.ufrj.br/teses/textocompleto/2010032301.pdf>

- ✓ Implementar o algoritmo, realizar os experimentos – a partir do modelo da câmera pinhole o algoritmo para estimar o TTC e detectar colisões foi implementado. Em seguida foram realizados os experimentos.
- ✓ Performance e validação do algoritmo – Foram aplicadas métricas para verificar o desempenho do algoritmo para detecção de colisões e validar o TTC.
- **Considerações finais e trabalhos futuros:** nesta etapa são discutidas as considerações finais da dissertação, bem como as limitações, principais contribuições, trabalhos futuros e propostas de melhorias que poderão ser agregadas ao modelo desenvolvido.

1.5. Estrutura da dissertação

Capítulo 1 – Introdução: Aborda uma introdução geral sobre o tema, a motivação, os objetivos e a metodologia aplicada.

Capítulo 2 – Deficiência visual: Traz uma visão básica sobre os tipos de deficiência, dando ênfase a alguns aspectos considerados relevantes para a compreensão de assuntos relacionados aos deficientes visuais.

Capítulo 3 – Visão Computacional: São apresentados conceitos básicos de Visão Computacional, Biblioteca OpenCV, Sistema visual humano, visão estéreo, visão monocular, fazendo uma analogia entre o sistema visual humano e o computacional.

Capítulo 4 – Tecnologia Assistiva: Apresenta uma classificação sobre os tipos de TA para as diferentes deficiências, trazendo alguns exemplos utilizados por deficientes visuais, dando ênfase às tecnologias computacionais.

Capítulo 5 – Tempo de Colisão com Visão Monocular: Apresenta alguns métodos para determinar o TTC com visão monocular.

Capítulo 6 – O algoritmo para detecção de colisões: Detalhes das etapas do algoritmo são apresentados, bem como todo o equacionamento.

Capítulo 7 – Experimentos e resultados: Destaca-se os principais resultados de cada experimento, bem como as métricas utilizadas para avaliar o algoritmo.

Capítulo 8 – Considerações finais e trabalhos futuros: As considerações finais, contribuições e propostas de melhorias são destacadas.

Referências – Exibe as principais referências utilizadas.

CAPÍTULO 2 – DEFICIÊNCIA VISUAL

2.1. Introdução

Segundo a Lei Nº 13.146, de 6 de julho de 2015, que foi instituída como “a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência), destinada a assegurar e a promover, em condições de igualdade, o exercício dos direitos e das liberdades fundamentais por pessoa com deficiência, visando a sua inclusão social e cidadania”, uma pessoa com deficiência é:

“aquela que tem impedimento de longo prazo de natureza física, mental, intelectual ou sensorial, o qual, em interação com uma ou mais barreiras, pode obstruir sua participação plena e efetiva na sociedade em igualdade de condições com as demais pessoas.” (BRASIL, 2015)

Neste contexto, as deficiências podem ser congênitas ou adquiridas por consequência de doenças, acidentes, entre outros fatores agravantes. De acordo com o Decreto nº 3.298, de 20 de Dezembro de 1999 (BRASIL, 1999), uma pessoa é considerada com deficiência quando se enquadra em uma das cinco categorias apresentadas a seguir:

I - Deficiência física (ou motora) - alteração completa ou parcial de um ou mais segmentos do corpo humano, acarretando o comprometimento da função física, apresentando-se sob a forma de paraplegia, paraparesia, monoplegia, monoparesia, tetraplegia, tetraparesia, triplicia, triparésia, hemiplegia, hemiparesia, ostomia, amputação ou ausência de membro, paralisia cerebral, nanismo, membros com deformidade congênita ou adquirida, exceto as deformidades estéticas e as que não produzam dificuldades para o desempenho de funções.

II - Deficiência auditiva - perda bilateral, parcial ou total, de quarenta e um decibéis (dB) ou mais, aferida por audiograma nas frequências de 500HZ, 1.000HZ, 2.000Hz e 3.000Hz.

III - Deficiência visual - cegueira, na qual a acuidade visual é igual ou menor que 0,05 no melhor olho, com a melhor correção óptica; a baixa visão, que significa acuidade visual entre

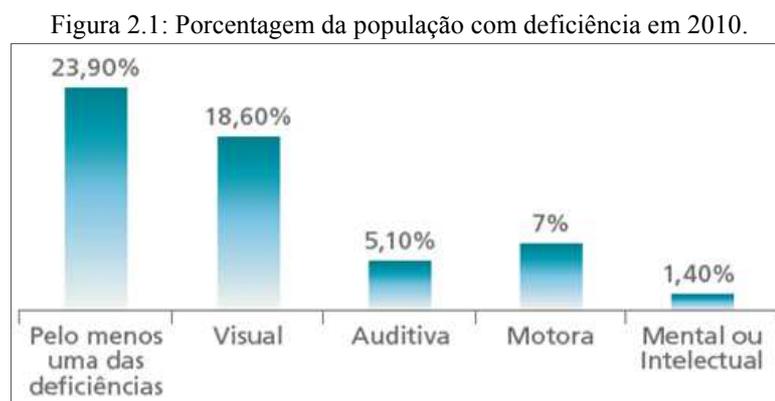


0,3 e 0,05 no melhor olho, com a melhor correção óptica; os casos nos quais a somatória da medida do campo visual em ambos os olhos for igual ou menor que 60°; ou a ocorrência simultânea de quaisquer das condições anteriores.

IV - Deficiência mental – funcionamento intelectual significativamente inferior à média, com manifestação antes dos dezoito anos e limitações associadas a duas ou mais áreas de habilidades adaptativas, tais como: comunicação, cuidado pessoal, habilidades sociais, utilização dos recursos da comunidade, saúde e segurança, habilidades acadêmicas, lazer e trabalho.

V - Deficiência múltipla – associação de duas ou mais deficiências.

Em 2010, o Brasil tinha uma população residente de 190.755.799 de pessoas, sendo que 23,9% deste total têm pelo menos algum dos tipos de deficiência listados acima. Isso significa que 45.623.910 brasileiros apresentavam alguma deficiência. A prevalência da deficiência variou de acordo com a natureza delas, sendo que a deficiência visual apresentou a maior ocorrência, afetando 18,6%, a deficiência motora em 7%, a auditiva em 5,10% e da deficiência mental ou intelectual em 1,40%. Pelo exposto, nota-se que a deficiência visual atingiu a maior porcentagem entre todas as demais deficiências. Conforme ilustra o gráfico da Figura 2.1. (DO CENSO, 2012)



Fonte: DO CENSO (2012).

No Brasil, segundo o IBGE (Instituto Brasileiro de Geografia e Estatística) no Censo de 2010, a deficiência visual foi a que mais afetou os brasileiros perfazendo uma parcela de 18,6% da população e totalizando 35.791.488 pessoas, onde 6,6 milhões de pessoas com grave



comprometimento visual (visão cega ou baixa), dos quais 506,3 mil eram totalmente cegos. De acordo com dados da OMS, o número de pessoas com deficiência visual no mundo é de aproximadamente 285 milhões, em pesquisa divulgada em agosto de 2012, sendo 246 milhões com perda parcial (baixa visão) e 39 milhões com perda total da visão (cegos).

Esses números ajudam a compreender o tamanho do desafio envolvido na construção de uma sociedade inclusiva, em que as diferenças sejam deixadas de lado e o direito à cidadania seja posto acima de tudo, pois os deficientes possuem os mesmos direitos que as demais pessoas, conforme reforça o trecho abaixo:

“As pessoas com deficiência têm os mesmos direitos que as pessoas sem deficiência. No entanto, elas nem sempre conseguem clamar por seus direitos nas mesmas condições em que o fazem as pessoas sem deficiência, devido a desvantagens impostas pela restrição de funcionalidades e pela sociedade, que lhes impõe barreiras físicas, legais e de atitude. Essas barreiras são responsáveis pelo distanciamento que existe na realização dos direitos das pessoas com e sem deficiência.” (DO CENSO, 2012)

Visto que a deficiência visual é um dos grandes desafios da atualidade, pelo menos quando comparadas com as demais deficiências, é de extrema importância que essa parcela significativa da população receba mais atenção e respeito por parte de toda a população. Além disso, eles não são incapazes ou incompetentes como muitas vezes são taxados, são sim pessoas que possuem dificuldades para realizar determinadas tarefas em relação a outras.

Essas dificuldades muitas vezes poderiam ser reduzidas ou evitadas, como reforça a OMS que em 2010

“considerou a deficiência visual um grave problema de saúde global, e chamou atenção para o fato de que com o conhecimento e tecnologias atuais, 80% das deficiências visuais advêm de causas evitáveis. Apesar disso, milhões de pessoas continuam em risco de perda visual devido à falta de cuidados dos olhos”. (ÁVILA et al., 2015, p. 109)

2.2. Deficiência visual e principais causas

A visão é um dos sentidos mais incríveis e que ajudam a compreender o mundo ao redor, e quando este sentido é privado de alguma pessoa, suas capacidades visuais e a noção do mundo ao seu redor tornam-se comprometidas.

A deficiência visual é definida como a perda total ou parcial da visão e pode ser de origem congênita (hereditária, adquirida antes do nascimento ou logo nos primeiros meses de

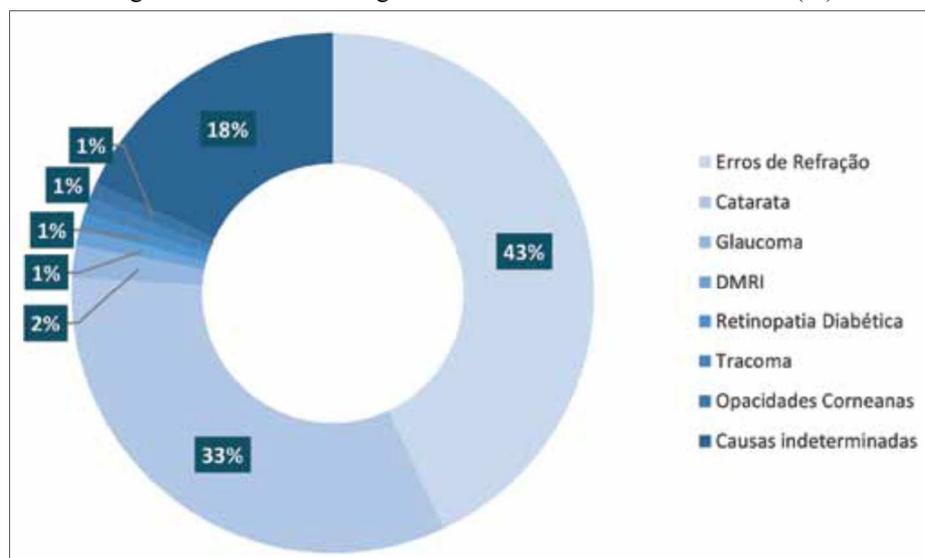


vida) ou adquirida (ao longo da vida). Os graus de visão vão desde a cegueira total até a visão perfeita, enquanto que a deficiência visual se refere à perda total da visão (cegueira) até a baixa visão (visão subnormal).

A cegueira é considerada a perda total da visão ou pouquíssima capacidade visual (cegueira parcial). Na perda total da visão ou cegueira total (Amaurose – termo utilizado para cegueira total), a pessoa não possui nenhuma percepção de luminosidade. Já na cegueira parcial, existe a percepção de luz que permite apenas diferenciar a direção da fonte luminosa e o claro do escuro. Enquanto a visão subnormal é a alteração da capacidade visual dada pelo comprometimento funcional dos olhos, em que a pessoa possui acuidade visual reduzida (baixo grau de aptidão do olho para discriminar os detalhes espaciais), e não pode ser corrigido com uso de óculos convencionais, lentes de contato e nem mesmo com intervenções cirúrgicas. Com o auxílio de recursos ópticos (com lentes especiais), a pessoa fica com a visão embaçada e consegue apenas distinguir vultos, claridade, ou objetos bastante próximos.

De acordo com o Conselho Brasileiro de Oftalmologia, a pesquisa *Global Data on Visual Impairments* (WHO, 2012 apud ÁVILA et al., 2015), aponta que globalmente a deficiência visual pode ser ocasionada por diferentes fatores, sendo as principais causas ilustradas na Figura 2.2. Onde mostra que erros de refração não corrigidos e catarata são os principais fatores com 43% e 33%, respectivamente, seguidos por glaucoma (2%), degeneração macular relacionada à idade (DMRI), retinopatia diabética, tracoma e opacidades corneanas (1% cada) e 18% são por causas indeterminadas.

Figura 2.2: Causas de Cegueira e Deficiência Visual no Mundo (%).



Fonte: Ávila et al. (2015, p. 109).



2.3. Visão humana

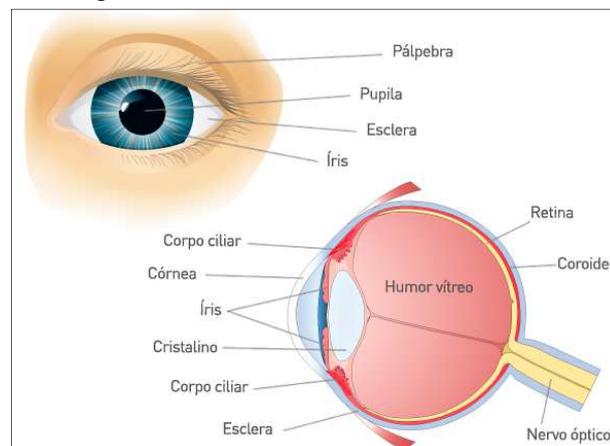
Como a deficiência visual está diretamente relacionada com a visão humana, cabe apresentar alguns detalhes básicos relacionados ao sistema visual humano.

Naturalmente a maioria das pessoas são dotadas de cinco sentidos fundamentais que lhes permitem interagir com o ambiente no qual estão situadas. Esses sentidos são: Audição (capacidade de ouvir os sons), Paladar (capacidade de sentir sabor), Olfato (capacidade de sentir o cheiro), Tato (sensibilidade de coisas em contato com a pele) e a Visão (capacidade de visualizar, de enxergar).

Dos cinco sentidos apresentados pode-se destacar a visão, pois permite detectar, localizar, interpretar e reconhecer objetos do ambiente ao redor (PEDRINI e SCHWARTZ, 2008). Sendo o sistema visual humano o responsável por este sentido, que se inicia pelo olho, sendo que qualquer tipo de lesão ou doença nesse órgão pode implicar na queda da capacidade visual, podendo acarretar desde uma pequena redução da acuidade visual até a deficiência visual (cegueira ou visão subnormal), tendo como consequências a perda ou a redução da percepção visual do mundo ao redor. Dessa forma, torna-se interessante apresentar alguns conceitos acerca da visão humana.

O olho (globo ocular) é um órgão do corpo humano que tem sua estrutura dividida na parte externa e na parte interna. A Figura 2.3 ilustra sua constituição básica. Ele atua como um sensor, captando a luz refletida pelos objetos convertendo-a em sinais que são enviados ao cérebro, de modo que esses sinais sejam processados e a cena interpretada.

Figura 2.3: Estrutura básica do olho humano.



Fonte: Cerpo (2014)

Essa luz captada pelo olho inicialmente atinge a córnea, em seguida atravessa a pupila e atinge o cristalino, convergindo para um ponto focal na retina. Na retina existem células



sensíveis à luz, então os estímulos luminosos são convertidos em impulsos eletroquímicos que são enviados ao cérebro pelo nervo óptico. No cérebro, o córtex visual processa as informações das imagens recebidas pelos dois olhos e completa a sensação visual.

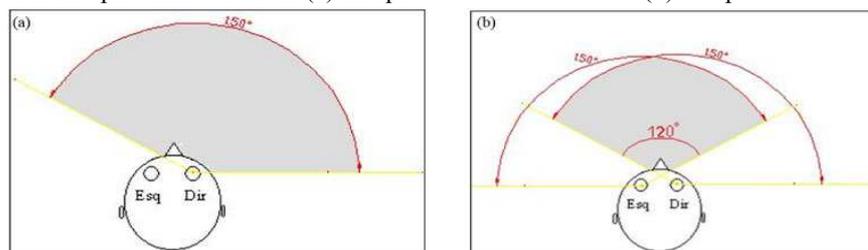
Quando a imagem é interpretada pelo cérebro através dos dois olhos, tem-se a visão binocular ou estéreo, então é possível perceber cenas tridimensionais com a percepção de profundidade de cada objeto no ambiente.

Por outro lado, quando a imagem é interpretada apenas por um dos olhos (a visão normal em um olho e cegueira no outro) a percepção de tridimensionalidade e a noção de profundidade não existem mais, então a informação que chega ao cérebro é de uma cena em 2 dimensões (em um único plano). Desse modo, tem-se a visão Monocular, onde as pessoas reconhecem formas, cores e tamanhos, porém têm incapacidade ou dificuldade em obter informações de profundidade que uma pessoa com visão estéreo possui.

Além disso, há uma alteração no campo visual. O campo visual pode ser entendido como uma extensão angular do ambiente no qual um objeto pode ser percebido, quando os olhos (ou o olho, para visão monocular) estiverem observando frontalmente e com a cabeça imóvel. Esse campo pode ser monocular ou binocular.

Na visão monocular, o campo visual é aproximadamente 150° (Figura 2.4a). Já na visão binocular cada um dos olhos possui um campo visual independente de 150° , contudo, o campo visual binocular se dá pela área de sobreposição dos campos visuais de cada uma das câmeras, formando um ângulo aproximado de 120° , conforme a Figura 2.4b.

Figura 2.4: Campo visual humano. (a) campo de visão monocular. (b) campo de visão binocular.



Fonte: Delfin; De Jesus (2011).

2.4. Orientação e Mobilidade

Seja uma pessoa cega ou com visão subnormal, independentemente do grau de sua deficiência visual, a noção e a capacidade de se orientar e se movimentar com segurança torna-se comprometida à medida que o espaço em que se situa é pouco conhecido ou desconhecido.

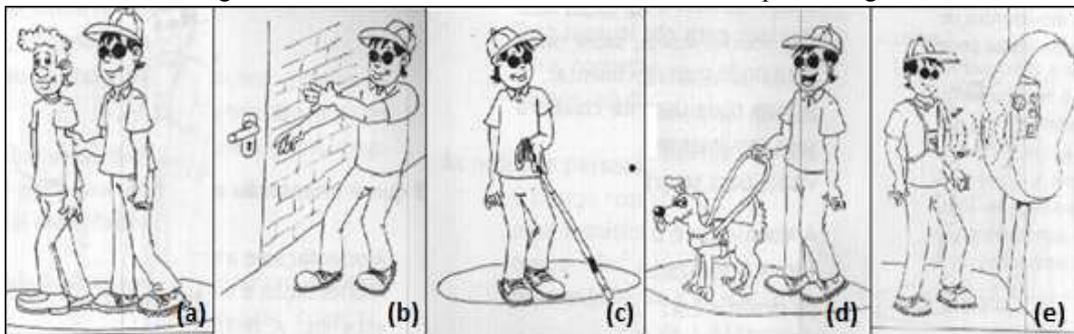


Deste modo, como destaca Bueno (1992), “um dos maiores problemas que os deficientes visuais enfrentam é o da orientação e mobilidade”.

Para Bueno (1992), a orientação é a percepção do indivíduo em relação à sua localização em um ambiente, e a mobilidade refere-se à capacidade de uma pessoa em se locomover. Em outras palavras, conforme Felipe (2001), a orientação é o aprendizado no uso dos sentidos para obter informações do ambiente, onde a pessoa deficiente visual pode usar outros sentidos como por exemplo, auditivos, tato e olfato para se orientar; enquanto que a mobilidade é o aprendizado para o controle dos movimentos de forma organizada e eficaz.

A mobilidade do deficiente visual está diretamente relacionada com sua capacidade de locomoção, podendo ser alcançada por meio de treinamentos que envolvem a utilização de recursos mecânicos como bengalas, ópticos (lentes especiais), eletrônicos como laser e ultrassom, cães treinados (Bueno, 1992). Estes recursos têm o objetivo de auxiliar no desenvolvimento das capacidades motoras e de percepção dos deficientes. Felipe (2001) complementa que existem cinco técnicas para que uma pessoa deficiente visual possa se movimentar em um ambiente. Estas estão ilustradas na Figura 2.5 e são brevemente comentadas logo em seguida.

Figura 2.5: As cinco técnicas de Mobilidade de uma pessoa cega.



Fonte: Adaptado de Felipe (2001).

a) Com ajuda de outra pessoa - guia vidente: é bastante comum o uso de um guia vidente, mas apesar da dependência em se locomover, é importante a participação efetiva do deficiente visual nas decisões do que estiver ocorrendo durante o seu deslocamento.

b) Usando seu próprio corpo – autoproteções: a pessoa utiliza partes do próprio corpo (por exemplo: mãos, pernas, braços) como ajuda para se deslocar com eficiência, segurança e independência. Normalmente é utilizada em ambientes internos, muitas vezes por já serem conhecidos ou serem menores do que ambientes externos que apresentam maior risco a integridade da pessoa. De acordo com Felipe (2001), “as autoproteções podem ser utilizadas



em conjunto com outras habilidades e sistemas de Orientação e Mobilidade como o guia vidente, a bengala longa, o cão-guia e as ajudas eletrônicas”.

c) Usando uma bengala - bengala longa: é um instrumento indispensável para a locomoção de uma pessoa cega, pois funciona como uma extensão táctil da mão que é usada para transmitir à pessoa uma riqueza de informações como se ela caminhasse passando a mão no chão. Com tempo e treinamento é possível desenvolver a percepção para detectar desníveis, buracos e outros obstáculos ao nível do chão. Também pode ser usada em conjunto com outras formas de mobilidade.

d) Usando um animal - cão-guia: é responsável por auxiliar um deficiente visual a se locomover à qualquer lugar e normalmente é aceito em locais públicos. Devido a essa responsabilidade o cão guia deve garantir segurança e integridade do deficiente visual, recebendo um treinamento rigoroso para que siga diversas regras, como por exemplo, não sofrer com distrações de outros cães.

e) Usando a tecnologia - algum aparato eletrônico: o deficiente visual utiliza algum recurso baseado em tecnologias eletrônicas e/ou computacionais para se locomover nos ambientes. São inúmeros aparelhos ou métodos que vêm sendo desenvolvidos para auxiliar nessa locomoção, sendo assim, alguns serão apresentados e discutidos no Capítulo 4.

2.5. Considerações finais do capítulo

Este capítulo apresentou alguns conceitos gerais sobre a deficiência visual, que pode ser congênita ou adquirida e os graus de visão, que vão da cegueira total até a visão subnormal. Bem como alguns conceitos sobre a visão humana, destacando o globo ocular e o campo visual do homem (monocular ou binocular).

Trouxe os conceitos de Orientação e Mobilidade, que é algo fundamental para os deficientes visuais, pois objetiva auxiliar em ações cotidianas e visa facilitar a sua integração social com maior independência e autonomia, com o intuito de melhorar a qualidade de vida e inseri-los como pessoas cidadãs na sociedade. Ainda traz as técnicas de Mobilidade apresentadas por Felipe (2001), onde comenta as formas que um deficiente visual pode se locomover.

CAPÍTULO 3 – VISÃO COMPUTACIONAL

3.1. Introdução

A área da Visão Computacional pode ser definida como um conjunto de métodos e técnicas que tornam sistemas computacionais capazes de extrair e interpretar informações de imagens digitais, podendo ser considerado um forte recurso computacional para auxiliar em sistemas de TA para deficientes visuais

Nesse sentido, os sistemas de Visão Computacional normalmente utilizam câmeras de vídeo para tentar emular a visão humana e obter a percepção do espaço tridimensional. Porém, enquanto o homem utiliza os olhos para fazer a aquisição das imagens, a visão por computador normalmente utiliza câmeras de vídeo que podem ser por visão estéreo (binocular - um ou mais pares de câmeras) ou por visão monocular (apenas uma câmera).

Para auxiliar em tarefas relacionadas à visão por computador, algumas bibliotecas vêm sendo desenvolvidas, como por exemplo a OpenCV que é uma biblioteca de código aberto desenvolvida para aplicações de visão computacional em tempo real, e foi utilizada na implementação do método desenvolvido.

Para a aplicação desenvolvida, se faz necessário detectar movimentos entre o observador (usuário) e os objetos contidos na cena, logo optou-se por utilizar a visão computacional para esta tarefa. Um dos métodos para percepção de movimentos é o baseado em fluxo óptico, que tem seu princípio de funcionamento fundamentado a partir da análise de uma sequência de quadros (*frames*) de imagens adquiridas ao longo do tempo.

Foi por volta da década de 1980 que começaram a entrar os primeiros algoritmos puramente computacionais para processamento de vídeo e estimação de movimento, dos quais citam-se os algoritmos de Lucas e Kanade, Horn e Schunck e posteriormente Farneback, que são algoritmos para estimação de movimento por fluxo óptico (KUIASKI, 2012).

Os algoritmos de Lucas-Kanade, Horn-Schunck e Farneback, apresentam características importantes para o método desenvolvido, pois de acordo com Aires (2009), facilitam o processamento em tempo real, que é algo imprescindível quando trata-se de tarefas que necessitam de respostas rápidas à variação do movimento, como é o caso da estimação do Tempo de Colisão.

A seguir serão apresentados alguns conceitos relacionados à Visão Computacional.

?

3.2. Visão Computacional

É uma área da Computação que pode ser definida como um conjunto de métodos e técnicas que tornam sistemas computacionais capazes de extrair e interpretar informações de imagens digitais. Algoritmos capazes de extrair informações presentes em imagens a partir de processos de aquisição de imagens e de sua percepção pelo sistema visual. Em outras palavras, a visão por computador tenta emular a visão humana, auxilia na resolução de problemas altamente complexos, buscando imitar a habilidade do ser humano em tomar decisões de acordo com as informações contidas na imagem (PEDRINI e SCHWARTZ, 2008).

?

?

3.2.2. Biblioteca OpenCV

A OpenCV é uma biblioteca de código aberto desenvolvida para aplicações de visão computacional em tempo real. Possui diversas funções pré-definidas facilitando a implementação de algoritmos, por exemplo: detecção de bordas, filtros, extração de características, limiarização (*thresholding*) e histograma.

Pode ser utilizada em diversas e diferentes áreas, como em Interface Humano Computador, identificação de objetos, segmentação e reconhecimento, reconhecimento de faces, reconhecimento de gestos, rastreamento de movimentos, compreensão de movimentos, robótica móvel entre outras relacionadas ao processamento de imagens digitais. A biblioteca inicialmente foi desenvolvida pela Intel a partir do ano 2000, com sua primeira versão liberada em outubro de 2006; atualmente ela é gratuita e de código aberto (*open source*) e está disponível tanto para uso acadêmico como comercial. Possui interfaces para C++, C, Python e Java, além de suportar os Sistemas Operacionais Windows, Linux, Mac OS, iOS e Android.

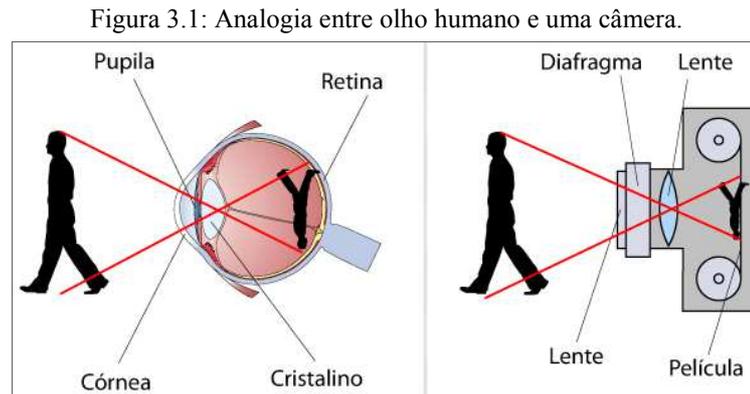
3.3. Sistema de visão por computador

Os sistemas de visão consistem em métodos de percepção do ambiente. Como exemplos cita-se a visão humana (ver no capítulo 2) e a visão por computador. Esses dois tipos de visão se assemelham, visto que as técnicas de Visão Computacional têm como objetivo tentar emular a visão humana, porém enquanto o homem utiliza os olhos para fazer a aquisição das



imagens, a visão por computador normalmente utiliza câmeras de vídeo que podem ser estéreo (duas ou mais câmeras) ou monocular (apenas uma câmera).

O princípio de funcionamento de uma câmera fotográfica ou de vídeo, é análogo ao sistema de visão humano, visto que a córnea e o cristalino funcionam como lentes, a pupila assemelha-se ao diafragma, e a retina funciona como uma película ou sensor de imagem. A Figura 3.1 demonstra essa analogia.



Fonte: Cerpo (2014).

Nesse sentido, os sistemas de Visão Computacional normalmente utilizam câmeras de vídeo para tentar simular a visão humana e obter a percepção do espaço tridimensional. Esses sistemas basicamente podem se dividir em dois modelos, um por meio de visão estéreo e o ou por visão monocular. Assim, esses dois modelos são comentados a seguir.

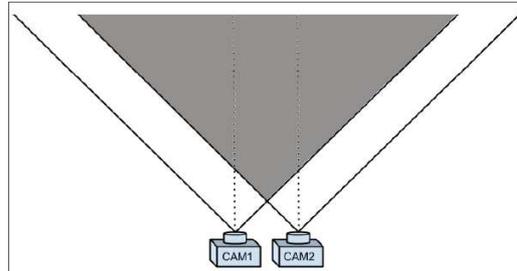
3.3.1. Visão Estéreo

O modelo baseado em Visão Estéreo utiliza câmeras aos pares, e funciona de maneira análoga ao sistema visual humano, onde as duas câmeras capturam duas imagens semelhantes. Otuyama (1998) complementa que é um “ramo da Visão Computacional que analisa o problema da reconstrução da informação tridimensional de objetos a partir de um par de imagens capturadas simultaneamente, mas com um pequeno deslocamento lateral”; e que esse deslocamento gera diferenças nas imagens que são praticamente imperceptíveis quando observadas separadamente, são essas diferenças que permitem a percepção de tridimensionalidade e conseqüentemente obter a profundidade entre os objetos e as câmeras, e com certa facilidade é possível estimar o TTC.



Assim como a visão binocular humana, o campo visual da visão por câmeras é dado pela área de sobreposição dos campos visuais de cada uma das câmeras, como pode ser visto na Figura 3.2, onde a área em cinza representa o campo visual por meio da visão estéreo.

Figura 3.2: Par de câmeras posicionadas paralelamente, direcionadas para uma mesma cena.



Fonte: Pflieger (2013).

Contudo, sistemas baseados em visão estéreo apresentam alguns inconvenientes, como por exemplo a necessidade de: conhecer algumas características intrínsecas das câmeras, conhecer a distância entre as câmeras, calibração, utilizar duas câmeras, custo financeiro e computacional maior que com uma única câmera. Desse modo, uma das alternativas é o uso da visão monocular.

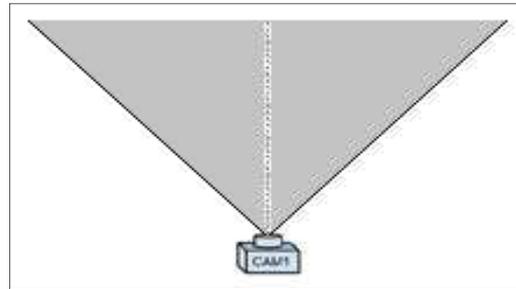
3.3.2. Visão Monocular

O sistema de visão monocular utiliza uma única câmera para adquirir as imagens do ambiente, e diferente da visão estéreo que fornece informações tridimensionais, a monocular retorna apenas informações da cena em duas dimensões, não fornecendo informações sobre a profundidade. A Figura 3.3 ilustra uma câmera monocular com seu campo visual, que assim como na visão monocular humana, é maior que o da visão estéreo, entretanto, a visão monocular não fornece informações sobre a profundidade em relação aos objetos.

Porém, existem métodos para extrair características da cena tridimensional, assim como estimar a profundidade e o TTC unicamente com visão monocular. Como exemplo cita-se os métodos baseados em fluxo óptico, que a partir de uma sequência de imagens pode extrair informações importantes para estimar a profundidade e o TTC, necessários para evitar colisões entre o observador (câmera) e um obstáculo em potencial.



Figura 3.3: Câmera monocular com seu campo visual em cinza.



Fonte: Adaptada de Pflieger (2013).

3.4. Considerações finais do capítulo

Este Capítulo abordou brevemente alguns conceitos relacionados à Visão Computacional. Além disso, fez uma analogia entre o sistema visual humano e o por câmeras, bem como uma comparação entre a visão estéreo e a monocular.

CAPÍTULO 4 – TECNOLOGIA ASSISTIVA

4.1. Introdução

Diversas tecnologias encontradas no cotidiano das pessoas podem ser utilizadas em diferentes atividades a fim de suprir suas necessidades. Para as pessoas com deficiência, essas tecnologias recebem o nome de TA e servem como instrumentos auxiliares que visam proporcionar uma certa independência e autonomia em algumas tarefas cotidianas.

Para amparar os deficientes, em novembro de 2006, a Secretaria Especial dos Direitos Humanos da Presidência da República - SEDH/PR, através da portaria nº 142, instituiu o Comitê de Ajudas Técnicas – CAT (Ajudas Técnicas é o antigo termo para designar o atual termo TA), onde reuniu um grupo de especialistas brasileiros e representantes de órgãos governamentais para criar diretrizes relacionadas à TA, que propôs o seguinte conceito:

“Tecnologia Assistiva é uma área do conhecimento, de característica interdisciplinar, que engloba produtos, recursos, metodologias, estratégias, práticas e serviços que objetivam promover a funcionalidade, relacionada à atividade e participação de pessoas com deficiência, incapacidades ou mobilidade reduzida, visando sua autonomia, independência, qualidade de vida e inclusão social”. (BERSCH, 2013)

De acordo com Bersch e Tonolli (2006) o termo TA foi criado em 1988 como elemento de uma Lei que regula os direitos dos cidadãos com deficiência nos EUA (Estados Unidos da América), sendo classificada segundo as diretrizes gerais da ADA (*American with Disabilities Act*) em 11 categorias. Em 24 de Outubro de 2012 essa classificação foi adaptada e utilizada pelo Ministério da Fazenda, Ciência, Tecnologia e Inovação e pela Secretaria Nacional de Direitos Humanos da Presidência da República do Brasil, em uma Portaria que trata sobre a linha de crédito subsidiado para aquisição de bens e serviços de TA destinados às pessoas com deficiência e sobre o rol dos bens e serviços (BERSCH, 2013).

A seguir, na Tabela 4.1, são apresentadas as categorias de classificação de TA, que assim como o texto explicativo de cada uma, essas foram adaptadas e extraídas de Bersch (2013).



Tabela 4.1: Quadro de classificação de TA.

Categoria	Aplicação
1 - Auxílios para a vida diária e vida prática	Materiais e produtos que favorecem desempenho autônomo e independente em tarefas rotineiras ou facilitam o cuidado de pessoas em situação de dependência de auxílio, nas atividades como se alimentar, cozinhar, vestir-se, tomar banho e executar necessidades pessoais. Ex: Fixador de talher à mão e cadarço mola.
2 - Comunicação Aumentativa e Alternativa	Destinada a atender pessoas sem fala ou escrita funcional ou em defasagem entre sua necessidade comunicativa e sua habilidade em falar e/ou escrever. Ex: Prancha de comunicação impressa e vocalizadores de mensagens gravadas.
3 - Recursos de acessibilidade ao computador	Conjunto de hardware e software especialmente idealizado para tornar o computador acessível a pessoas com privações sensoriais (visuais e auditivas), intelectuais e motoras. Ex: Mouses, teclado Braille e mapa tátil com impressão em relevo.
4 - Sistemas de controle de ambiente	São sistemas em que as pessoas com limitações motoras podem controlar aparelhos eletroeletrônicos. Ex: Sistema controle de ambiente a partir do controle remoto.
5 - Projetos arquitetônicos para acessibilidade	Projetos de edificação e urbanismo que garantem acesso, funcionalidade e mobilidade a todas as pessoas, independentemente de sua condição física e sensorial. Ex: Acessibilidade no banheiro.
6 - Órteses e próteses	Próteses são peças artificiais que substituem partes ausentes do corpo. Órteses são colocadas junto a um segmento corpo, garantindo-lhe um melhor posicionamento, estabilização e/ou função. Ex: Próteses e órteses.
7 - Adequação Postural	São recursos que servem para garantir uma postura estável e confortável, que quando utilizados precocemente podem auxiliam na prevenção de deformidades corporais. Ex: Poltrona postural e estabilizador ortostático.
8 - Auxílios de mobilidade	A mobilidade pode ser auxiliada por bengalas, muletas, andadores, carrinhos, cadeiras de rodas manuais ou elétricas, scooters e qualquer outro veículo, equipamento ou estratégia utilizada na



	melhoria da mobilidade pessoal. Ex: Cadeira de rodas motorizada, andador com freio e carrinho de transporte infantil.
9 - Auxílios para qualificação da habilidade visual e recursos que ampliam a informação de pessoas com baixa visão ou cegas	São utilizados para auxiliar pessoas cegas ou com visão subnormal. Ex: Lupas manuais, lupa eletrônica, aplicativos para celulares com retorno de voz.
10 - Auxílios para pessoas com surdez ou com déficit auditivo	Incluem vários equipamentos para auxiliar pessoas com algum tipo de deficiência auditiva. Ex: Aparelho auditivo.
11 - Mobilidade em veículos	Acessórios que possibilitam uma pessoa com deficiência física dirigir um automóvel, facilitadores de embarque e desembarque como elevadores para cadeiras de rodas, rampas para cadeiras de rodas. Ex: Adaptação para dirigir somente com as mãos e elevador para cadeiras de rodas.
12 - Esporte e Lazer	Recursos que favorecem a prática de esporte e participação em atividades de lazer. Ex: Cadeira de rodas adaptada e bola sonora.

4.2. Tecnologia Assistiva para deficientes visuais

Nos dias de hoje existem inúmeros recursos e produtos que podem ser utilizados por pessoas com qualquer tipo de deficiência ou idosos, porém nem todos são comercializados no Brasil, e muito menos disponibilizados para a população menos amparada. Para regularizar a distribuição e comercialização foi criado o Catálogo Nacional de Produtos de TA (MCTI, 2015), que surgiu de uma iniciativa do Ministério da Ciência, Tecnologia e Inovação, através da Secretaria de Ciência e Tecnologia para Inclusão Social, desenvolvido e realizado em parceria com o Instituto de Tecnologia Social.



O Catálogo é um serviço de informação na internet que apresenta e possibilita a realização de buscas sobre os produtos de TA, distribuídos no Brasil, sejam eles nacionais ou importados. Seu objetivo não é o de distribuir ou comercializar produtos, e sim o de oferecer informações sobre os produtos que podem contribuir para maior autonomia e qualidade de vida aos deficientes e idosos (MCTI, 2015).

Tratando-se de deficiência visual, o Catálogo conta com um total de 520 produtos cadastrados (em agosto de 2015) para diversas aplicações diferentes, alguns são baseados em tecnologias eletrônicas e/ou computacionais, enquanto outros não. Alguns exemplos podem ser visualizados na Figura 4.1. Cada um deles é brevemente comentado conforme está descrito no Site do Catálogo.

Figura 4.1: Alguns exemplos de TA para cegos.



Fonte: Adaptado de MCTI (2015)

(a) Leitor Autônomo: Reconhece o texto que o documento contiver e o lê em voz, em português, após poucos segundos do reconhecimento, permitindo gravar para futura audição; (b) Mesa de Relevos Táteis: Permite a criação de diagramas táteis, possibilitando que os símbolos, ícones, regiões e toda superfície possa falar através de uma voz sintetizada. (c) Impressora Braille: Impressão de textos em Braille; (d) Leitor de textos: Trabalhando junto com um scanner de mesa comum, converte materiais impressos em imagens digitais cujo conteúdo textual é reconhecido e convertido em texto para ser falado por um sintetizador de voz; (e) Máquina de escrever: Máquina de escrever portátil em Braille; (f) Termômetro frontal falante: Mede a temperatura corporal e reproduz em fala humana os dados mostrados na tela; (g) Colocador linha e agulha



automático: Permite passar linha em qualquer tipo de agulha; (h) Indicador de nível de líquido: Indica mediante um sinal sonoro quando o nível de líquido está chegando o topo do recipiente; (i) Relógio falante: Fala as horas ao apertar um botão; (j) Leitor de telas para celular: software leitor de telas que dá acessibilidade a praticamente todas as funções e aplicativos de um celular.

Os exemplos acima serviram apenas para dar uma visão geral sobre alguns produtos de TA que são comercializados no Brasil. Contudo, muitos estudiosos têm pesquisado métodos e/ou desenvolvido sistemas que visam facilitar a vida dos deficientes. Dessa forma, a seguir são apresentados alguns dos sistemas computacionais e/ou eletrônicos que foram encontrados na literatura durante a revisão bibliográfica. São sistemas de apoio aos deficientes visuais que servem para aplicações de navegação em ambientes e reconhecimento de objetos.

4.2.1. Métodos computacionais de Tecnologia Assistiva para deficientes visuais

Durante a revisão bibliográfica constatou-se que a grande maioria das pesquisas encontradas na literatura exibiam métodos para navegação e/ou reconhecimento de objetos para deficientes visuais. A partir dessa revisão observou-se que há uma distinção entre os métodos, dividindo-os nas seguintes categorias: apoio à navegação não contextualizada e apoio à navegação contextualizada.

Os sistemas com informações não contextualizadas fornecem apenas a informação de orientação e interação com o utilizador, não sendo dada qualquer informação sobre o ambiente. Enquanto que os sistemas com informações contextualizadas, fornecem algumas informações do ambiente, utilizam informações visuais com contexto do ambiente para dar suporte necessário para a navegação dos deficientes visuais com maior autonomia e segurança (FERNANDES, 2010).

4.2.1.1. Apoio à navegação para deficientes visuais com informações não contextualizadas

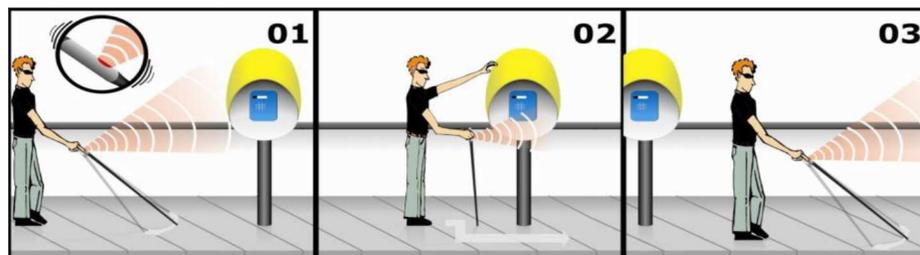
Um dos recursos mais utilizados pelos deficientes visuais é a bengala longa que serve para locomoção e funciona como uma extensão do sentido tátil, detectando obstáculos no chão



ou bem próximos a ele. Entretanto, para obstáculos acima da linha da cintura, como por exemplo: lixeiras, galhos de árvores, caçambas de lixo, telefones públicos, etc, uma bengala tradicional não pode detectar uma possível colisão e conseqüentemente expõem ao risco a integridade física do deficiente visual. Para contornar esse tipo de problema, muitos pesquisados propuseram bengalas com dispositivos eletrônicos adaptados.

Da Silva et al. (2012) descreve o projeto de uma Bengala Longa Eletrônica, que utiliza um sistema eletrônico com um sensor ultrassônico para detectar obstáculos acima da linha da cintura e calcular a profundidade. A Figura 4.2 ilustra uma sequência onde uma resposta tátil é transmitida ao usuário através de um pequeno motor vibratório em sua pega (local onde ele segura a bengala). Conforme o obstáculo vai se aproximando, a vibração do motor aumenta e a resposta tátil torna-se mais intensa pulsando rapidamente.

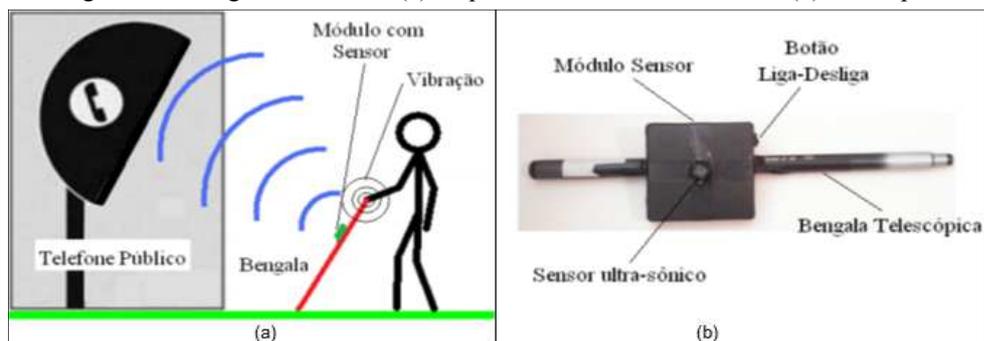
Figura 4.2: A bengala eletrônica alerta sobre obstáculos acima da linha da cintura.



Fonte: Da Silva et al. (2012).

Outro exemplo que emprega sensor de ultrassom é o de Santos et al. (2010) que utiliza um sistema eletrônico (com microcontrolador PIC) acoplado em uma bengala telescópica (Figura 4.3(b)). Quando um obstáculo é detectado a uma certa distância de segurança, um motor é acionado e sua vibração infere no corpo da bengala como um alerta da presença de um obstáculo (Figura 4.3(a)).

Figura 4.3: Bengala eletrônica. (a) Esquema básico funcionamento. (b) Protótipo.



Fonte: Santos et al. (2010).



Gonzatto et al. (2009) apresenta um protótipo de óculos com um sensor ultrassônico acoplado que detecta obstáculos a distâncias entre 0,2 e 1,5 metros. O sistema utiliza um circuito com microcontrolador PIC, onde são conectados o sensor ultrassônico e um fone de ouvido. A medida que o usuário se aproxima de um obstáculo, um sinal sonoro emitido pelo fone de ouvido aumenta de intensidade, representando o grau do risco de colisão. A ilustração da Figura 4.4 mostra um deficiente visual em rota de colisão detectada e o protótipo do óculos.

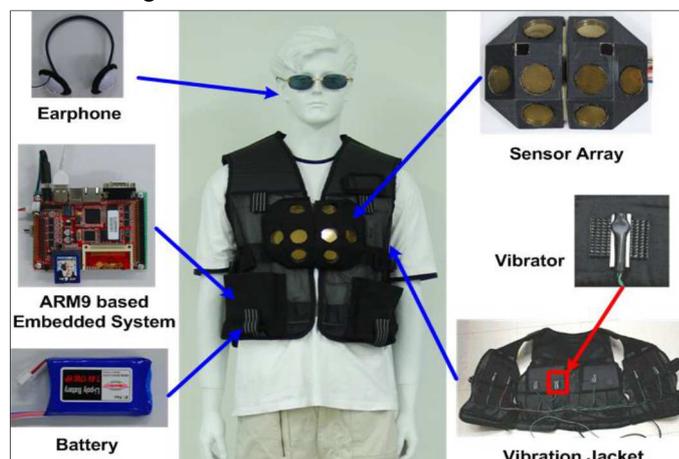
Figura 4.4: Ilustração de um deficiente visual em rota de colisão e o protótipo do óculos.



Fonte: Gonzatto et al. (2009).

Shin e Lim (2007) propuseram um sistema de detecção de obstáculo composto por um hardware embarcado com um microcontrolador ARM, onde é acoplado um fone de ouvido e um *array* com oito sensores de ultrassom que detectam a posição dos obstáculos. Então as informações da posição dos obstáculos são transmitidas aos fones de ouvido por comandos de voz e pela vibração de motores instalados no colete, dando o sentido de orientação para o deficiente visual. A Figura 4.5 mostra as partes constituintes do sistema.

Figura 4.5: Partes constituintes do sistema.



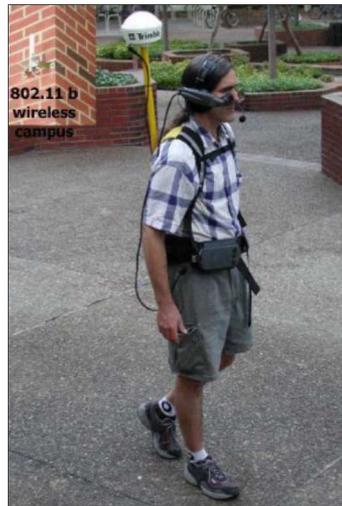
Fonte: Shin e Lim (2007).



Ran et al. (2004) apresenta um protótipo do *Drishti*, um sistema de navegação para deficientes visuais. Consiste em um sistema de localização e posicionamento que integra diferentes tecnologias, incluindo computador, reconhecimento e síntese de voz, redes sem fio, Sistema de Informação Geográfica (GIS) e Sistema de Posicionamento Global (GPS). Funciona de forma similar a um sistema tradicional com GPS, contudo a rota pode ser modificada de acordo com informações do GIS. Além disso utiliza *Differential* GPS, cuja precisão em relação ao posicionamento é superior a um GPS comum, pois utiliza estações de rádio na superfície da Terra para corrigir possíveis imprecisões do sinal captado dos satélites. O sistema constantemente guia o usuário pela rota escolhida através de sinais de voz e fornece informações sobre obstáculos dinâmicos.

Um grande empecilho se dá pelo tamanho do protótipo, visto que é bastante volumoso, dificultando o seu uso por um tempo muito prologando. Também utiliza fones de ouvido que podem interferir na percepção de sons do ambiente, colocando em risco o usuário. A Figura 4.6 ilustra uma pessoa utilizando o sistema.

Figura 4.6: Usuário com o protótipo do sistema *Drishti* em um ambiente externo.

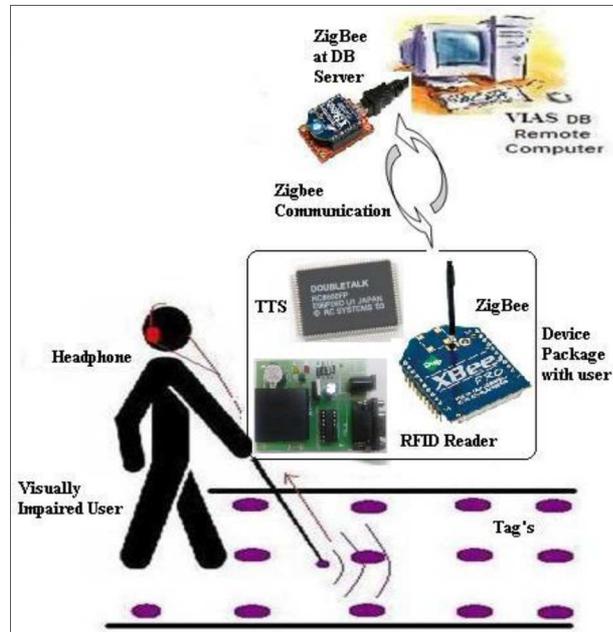


Fonte: Ran et al. (2004).

Varpe e Wankhade (2013) emprega *Radio Frequency Identification* (RFID) para identificar certos caminhos construídos (com etiquetas RFID) para navegação do deficiente visual. Também fornece certas características tais como identificação de objetos, informações do caminho, etc, por meio da leitura/identificação de etiquetas de RFID distribuídas ao longo das rotas e coladas nos objetos. A Figura 4.7 ilustra a proposta do sistema que é composto por: módulo leitor de RFID (*RFID Reader*), transmissor/receptor *ZigBee* para enviar/receber informações das etiquetas via wireless para/de um Servidor, e um módulo TTS para informar

ao usuário (por meio de fones de ouvido) sobre o caminho a ser seguido e os objetos que estiverem no ambiente (desde que estejam com etiquetas RFID).

Figura 4.7: A proposta do sistema baseado em RFID.



Fonte: Varpe e Wankhade (2013).

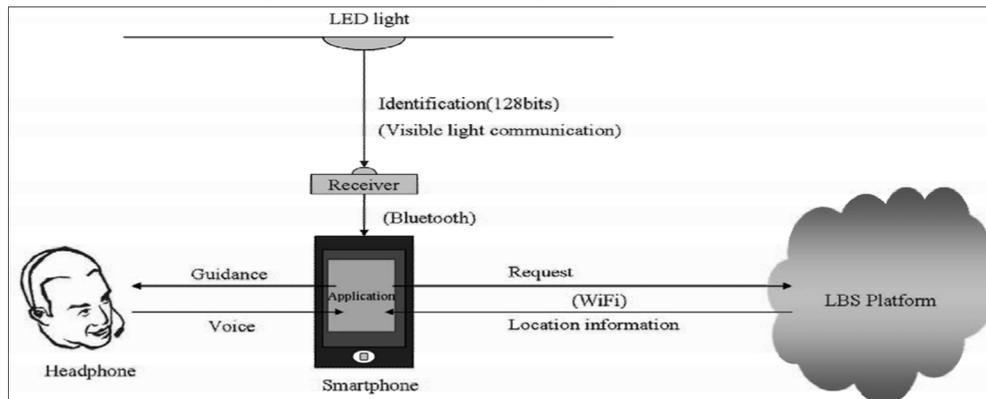
Cha et al. (2013) propõe um sistema de navegação usando um *Smartphone*. O aplicativo usa o Google Map e Google TTS (*Text-to-Speech*, converte texto em voz) para conduzir o usuário até o destino. O sistema está basicamente dividido em três etapas. Primeiro, o usuário deve informar o destino, então o reconhecimento de voz e o Google TTS são aplicados. Se o usuário for verdadeiro, ele vai para a próxima etapa. Após, a rota é pesquisada e traçada via Google Maps, então o mapa traçado é salvo em uma base de dados do celular. Finalmente, a pessoa é guiada pela rota traçada através de instruções de voz para o seu destino.

Nakajima e Haruyama (2012) apresentam um sistema de navegação para deficientes visuais baseado na tecnologia de Comunicação por Luz Visível (VLC - *Visible Light Communication*) e sensoriamento geomagnético integrados em um *smartphone*. Através de um receptor do *smartphone*, a tecnologia utiliza luz visível da iluminação ambiente (LED) para determinar a localização do usuário. O sensor geomagnético serve para guiar a direção de deslocamento (orientação).

O sistema consiste em luzes LED, um *smartphone* com receptor Bluetooth integrado, fones de ouvido e um ambiente de nuvem Panasonic (*LBS Platform*) que é referido como base de informações da posição (armazena as rotas). O Receptor abstrai a identificação da luz visível

por Bluetooth para encontrar a localização do usuário, via WiFi recebe da nuvem as informações de posição e a orientação é dada através de fones de ouvido. O usuário deve indicar o destino através de reconhecimento de voz e o *smartphone* solicita a rota para a nuvem, então o usuário é guiado por todo o caminho. A medida que a pessoa vai se movendo, as informações são atualizadas até que o destino seja alcançado. A Figura 4.8 ilustra o sistema.

Figura 4.8: Sistema de navegação por luz visível.



Fonte: Nakajima e Haruyama (2012).

Kaiser e Lawo (2012) propuseram um sistema para auxiliar na navegação de ambientes desconhecidos, seja interno ou externo. O sistema é baseado em Localização e Mapeamento Simultâneo (*SLAM – Simultaneous Localization and Mapping*) normalmente utilizado em robótica móvel, que é responsável por rastrear e mapear a posição do deficiente visual no ambiente, bem como criar um mapa para guiar o usuário de maneira eficiente por uma rota segura e com risco de colisão reduzido. O sistema conta com um scanner laser de curto alcance para detectar objetos, um computador para processar as informações, uma Unidade de Medição Inercial (*IMU – Inertial Measurement Unit*) para garantir a orientação do usuário, e fones de ouvido com tecnologia *bone conduction* (em que o áudio é transmitido aos ouvidos por meio da condução do som pelos ossos do crânio) utilizado para informar a rota.

4.2.1.2. Apoio à navegação para deficientes visuais com informações contextualizadas

Os sistemas de apoio à navegação que fornecem informações visuais do ambiente têm uma grande vantagem em relação aos que não fornecem esse tipo de informação, pois oferecem elementos de contexto do ambiente com base em algumas características, proporcionando um



suporte necessário para a navegação dos deficientes visuais com mais autonomia e segurança. Dessa forma, a seguir são brevemente comentados alguns dos exemplos encontrados na literatura e que utilizam visão por computador para dar suporte à navegação de deficientes visuais, no qual é dada uma breve descrição sobre cada um.

Molton et al. (1998, *apud* FERNANDES 2010) demonstra um sistema baseado em visão estéreo e sensor de ultrassom para auxiliar na navegação de deficientes visuais. Para obter informações do ambiente o sistema possui sensores de ultrassom no peito e cinto do usuário, além de duas câmeras para detectar objetos de grande dimensão (Figura 4.9). O modelo de visão por computador é baseado no algoritmo *Ground Plane Obstacle Detection (GPOD)*, que usa duas câmeras para caracterizar o chão através de uma parametrização da medição de disparidade entre as câmeras. Para estimar o movimento do chão foi também criado um modelo baseado no filtro de *Kalman*, que têm uma boa precisão, contudo não consegue funcionar em tempo real.

Figura 4.9: Sistema com visão estéreo e sensores de ultrassom.

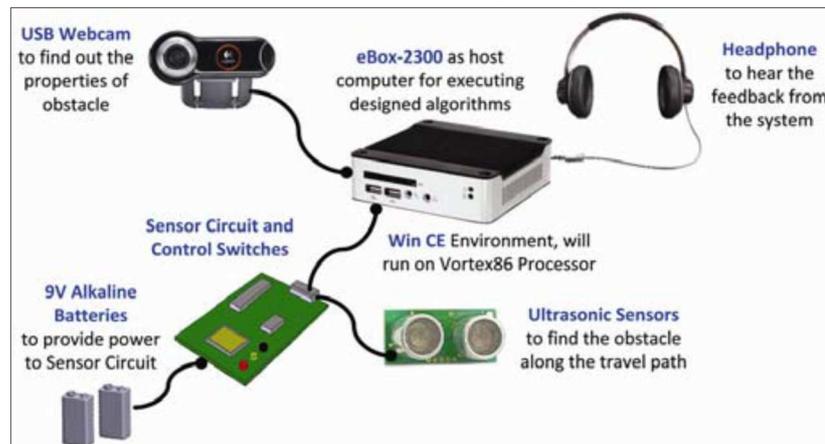


Fonte: Molton et al. (1998, *apud* FERNANDES 2010).

Kumar et al. (2013) desenvolveu um protótipo para navegação e detecção de obstáculos com sensor ultrassônico, que conta ainda com uma etapa que detecta a presença de pessoas e a textura de roupas. O sistema proposto detecta obstáculos até 3 metros por meio do sensor de ultrassom e envia feedbacks de áudio (por fones de ouvido) para informar a pessoa sobre a sua localização. Além disso, uma webcam USB é conectada com eBox 2300™ (Sistema Embarcado) para capturar e detectar a textura de roupas e a presença de pessoas por um método de detecção de face. Uma prévia do protótipo do sistema pode ser observada na Figura 4.10.



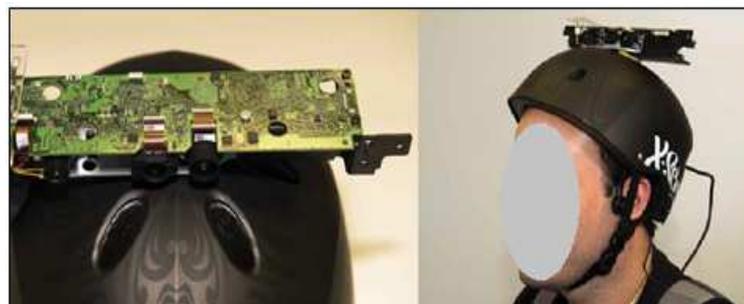
Figura 4.10: Prévia do protótipo.



Fonte: Kumar et al. (2013).

Outra aplicação consiste em um sistema de auxílio à navegação proposto por Ereno (2015) que utiliza um Kinect da Microsoft para converter vídeos em informações sonoras para auxiliar deficientes visuais. Foi desenvolvido um protótipo que teve como suporte um capacete de *skate*, onde foi colocado um sensor Kinect (que retorna o valor de profundidade), conectado a um laptop, além de um giroscópio, um acelerômetro e uma bússola, dispositivos que registram, em conjunto, mudanças de direção da cabeça (Figura 4.11). O usuário carrega o laptop dentro de uma mochila. Os dados do ambiente são capturados pela câmera (Kinect), processados pelo computador e transmitidos por fones de ouvido com tecnologia *bone conduction*. O sistema conta ainda com uma etapa para reconhecimento de pessoas que utiliza uma técnica de reconhecimento de face.

Figura 4.11: Detalhes do experimento de Ereno.



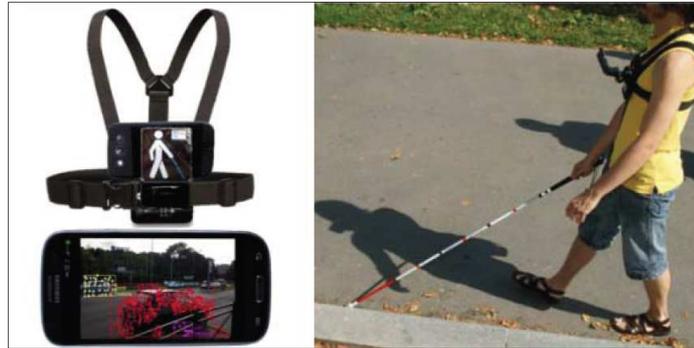
Fonte: Ereno, 2015.

Tapu et al. (2013) apresenta um sistema que divide-se basicamente em detecção de obstáculos (estáticos e dinâmicos) e classificação de objetos para ajudar a navegação para deficientes visuais através da câmera de um *smartphone*. Tem como embasamento técnicas de visão computacional para detecção de obstáculos informando o seu grau de risco de acordo

?

com a proximidade do usuário, e a classificação de objetos é realizada por técnicas para reconhecimento de objetos que comparam o objeto detectado com uma base de imagens no próprio celular. A Figura 4.12 ilustra o sistema proposto.

Figura 4.12: Teste do protótipo do sistema para smartphone.

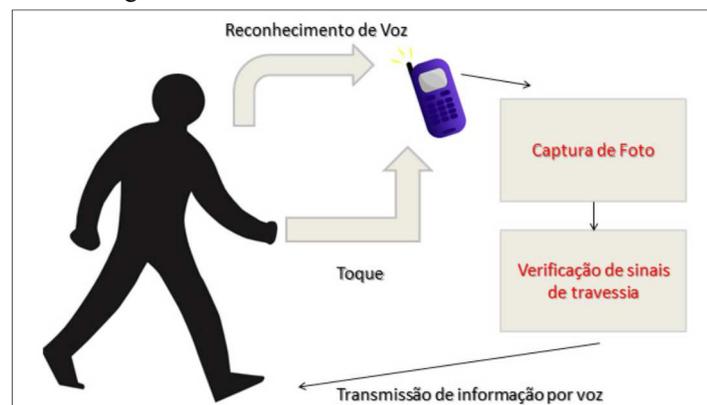


Fonte: Tapu et al. (2013).

?

Sousa e Marengoni (2012) mostram um sistema implementado em um *smartphone* baseado em técnicas de visão computacional, utilizado na detecção de faixa de pedestres para auxiliar na travessia de deficientes visuais. Na Figura 4.13 pode-se observar fluxo de funcionamento do sistema, o usuário ativa o sistema por comando de voz ou por toque no aparelho, então ocorre o reconhecimento do ambiente em busca da faixa de pedestres. Quando o sistema identifica uma faixa o usuário é informado através da vibração do aparelho ou por processamento de voz a condição de travessia na faixa de pedestre, se é segura ou não.

Figura 4.13: Fluxo de funcionamento do sistema.



Fonte: Sousa e Marengoni (2012).

Paisos (2012) desenvolveu um sistema para celular como auxílio de deficientes visuais, que conta com quatro diferentes aplicações. A primeira é o auxílio à navegação, que



fornece ao usuário informações para locomoção em ambientes internos desconhecidos. A segunda apresenta vários métodos de protótipos de entrada de texto em um telefone celular com tela sensível ao toque, são baseadas no alfabeto Braille e assim são convenientes para usuários com deficiência visual. A terceira aplicação é um sistema para reconhecimento de cédulas de dólar por meio da câmera. A última é uma ferramenta que utiliza a câmera para determinar quais peças de roupa combinam, com base em atuais modas sociais de combinação de cores.

4.3. Considerações finais do capítulo

A condição de vida de uma pessoa com deficiência acaba dificultando, ou até impedindo, sua participação em atividades cotidianas comuns para muitas pessoas sem nenhum tipo de deficiência. Desse modo, para ajudar essas pessoas foram desenvolvidos produtos ou serviços de TA que visam ampará-las, possibilitando sua inserção no contexto social.

Entretanto, muitos produtos são importados, caros e de difícil acesso para a maioria das pessoas deficientes. Tratando-se de tecnologias para auxiliar na navegação de deficientes visuais, existem poucos produtos no mercado nacional, em contrapartida existem muitos protótipos vinculados a projetos de pesquisa em universidades, nacionais ou internacionais, que sofrem pela falta de recursos e não chegam no mercado.

Verificou-se que muitos recursos e métodos podem ser incorporados em um sistema de navegação, entre eles: sensor de ultrassom, sensor laser, câmera estéreo, câmera monocular, GPS, GIS, RFID, IMU, Google Map, WiFi, Kinect, iluminação LED, Bluetooth, entre outros. Cada um com suas particularidades, vantagens e desvantagens. Além disso, cabe destacar a importância de um sistema de apoio à navegação, pois este deve fornecer informações necessárias, rápidas e precisas com a menor taxa de erro possível ao deficiente visual, de modo que consiga se deslocar pelo ambiente com autonomia e segurança.

Apesar das pesquisas citadas exibirem seus resultados, poucos reportam sua taxa de erro ou precisão, apenas Souza e Maregone (2012) apresentam um desempenho geral na ordem de 76% em seu sistema de detecção de faixa de pedestres; e Tapu (2013) apresentou uma precisão para detecção de obstáculos, obtendo a taxa de precisão média de 91,5%. Pelo exposto, percebe-se uma carência na apresentação de resultados, pois, apenas 2 das 15 pesquisas citadas reportam o seu desempenho, o que dificulta identificar a eficiência de um método.

A seguir um resumo com detalhes das TA citadas neste Capítulo é apresentada na Tabela 4.2.

Tabela 4.2: Resumo das TA para deficientes visuais apresentadas.

Autor	Título	Recursos/Técnicas
Da Silva et al. (2012)	Locomoção independente: contribuição à locomoção do cego em espaços urbanos.	Bengala Longa, utiliza um sistema eletrônico com sensor de ultrassom para detectar obstáculos.
Santos et al. (2010)	Desenvolvimento de uma bengala eletrônica para locomoção de pessoas com deficiência visual.	Bengala telescopia, um sistema microcontrolado com sensor de ultrassom detecta obstáculos.
Gonzatto et al. (2009)	Óculos sonar para deficientes Visuais.	Óculos com sistema microcontrolado e sensor de ultrassom detecta obstáculos.
Shin e Lim (2007)	Obstacle detection and avoidance system for visually impaired people.	Colete com <i>array</i> de sensores ultrassônicos conectados em um sistema microcontrolado para detectar obstáculos.
Ran et al. (2004)	Drishti: an integrated indoor/outdoor blind navigation system and service.	Sistema de navegação, guia o usuário. Utiliza: computador, síntese de voz, redes sem fio, GPS, GIS.
Varpe e Wankhade (2013)	Visually Impaired Assistive System.	Sistema de navegação que guia o usuário por caminhos pré-definidos marcados com etiquetas RFID.
Cha et al. (2013)	Design and Implementation of a Smart Phone Navigation Application for Blind Persons.	Aplicativo implementado em um <i>Smartphone</i> para guiar o usuário. Utiliza o Google Map e o Google TTS.
Nakajima e Haruyama (2012)	Indoor navigation system for visually impaired people using visible light communication and compensated geomagnetic sensing.	Sistema de navegação em um <i>Smartphone</i> baseado em VLC. Utiliza luzes LED, Bluetooth, plataforma LBS, WiFi.
Kaiser e Lawo (2012)	Wearable navigation system for the visually impaired and blind people.	Sistema de navegação, utiliza SLAM, IMU e tecnologia <i>bone conduction</i> .



Molton et al. (1998)	Robotic sensing for the partially sighted.	Sistema de navegação com visão estéreo e sensor de ultrassom.
Kumar et al. (2013)	An embedded system for aiding navigation of visually impaired persons.	Utiliza um Sistema embarcado eBox 2300™ para navegação e detecção de obstáculos com ultrassom, também detecta pessoas e textura de roupas por uma webcam.
Ereno (2015)	Reconhecimento de ambiente.	Sistema de auxílio à navegação que utiliza Kinect conectado a um laptop, um giroscópio, acelerômetro e uma bússola.
Tapu et al. (2013)	A smartphone based obstacle detection and classification system for assisting visually impaired people.	Sistema de navegação em um <i>smartphone</i> para detectar obstáculos e classificar objetos, utiliza a câmera do celular para capturar as imagens.
Sousa e Marengoni (2012)	Uso de Visão Computacional em Dispositivos Móveis para o Reconhecimento de Faixa de Pedestres.	Sistema para detecção de faixa de pedestres implementado em um <i>smartphone</i> , utiliza a câmera do celular para capturar as imagens.
Paisos (2012)	Mobile Accessibility Tools for the Visually Impaired.	Sistema implementado em um <i>smartphone</i> com 4 aplicações: auxílio à navegação, método de entrada de textos, reconhecimento de cédulas e, combinação de roupas de acordo com a moda.

CAPÍTULO 5 – TEMPO DE COLISÃO COM VISÃO MONOCULAR

5.1. Introdução

Na literatura existem diferentes métodos e pesquisas para estimar o TTC e/ou detectar obstáculos, muitos, por exemplo, utilizam sensores laser ou ultrassom de modo a adquirir subsídios relevantes para obter uma estimativa desse tempo. Outros, com técnicas de Visão Computacional que se baseiam no tratamento de imagens adquiridas por sistemas de visão por câmeras, como por exemplo: câmera monocular, câmera estéreo, Kinect, sonar visual. Entretanto, nesta dissertação são apresentados apenas algumas técnicas baseadas somente em visão monocular.

Um dos primeiros estudos sobre o assunto foi proposto por Lee (1976), que em sua primeira abordagem definiu TTC como sendo a distância entre um observador e um obstáculo (também conhecida como profundidade - z) dividida pela velocidade constante (V) de deslocamento do observador, conforme foi representado na Equação (1.1) do Capítulo 1. Porém, com a evolução das técnicas computacionais e a necessidade em obter resultados mais precisos e confiáveis, novos algoritmos têm sido desenvolvidos.

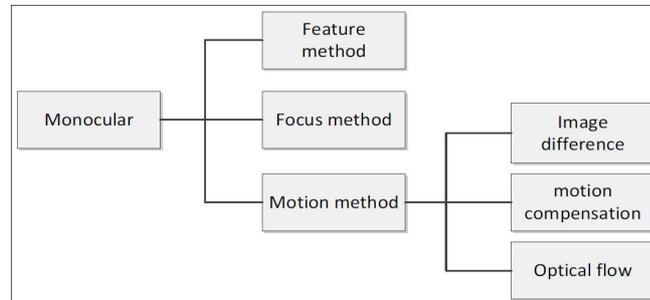
Contudo, antes de conhecer alguns métodos para estimar o TTC propriamente dito, torna-se interessante ter uma visão básica sobre métodos para detecção de obstáculos com visão monocular, que são brevemente comentados a seguir.

5.2. Algoritmos para detecção de obstáculos com visão monocular

As técnicas para calcular o TTC servem como base para determinar o risco de colisão com algum tipo de obstáculo, sendo que a primeira etapa para se calcular o TTC consiste na detecção do obstáculo. Dessa forma, antes de apresentar algumas técnicas para estimar o TTC, é necessário conhecer alguns métodos para detectar obstáculos, pois estes são a base de um sistema de navegação autônoma e segura. Desse modo, uma revisão bibliográfica foi feita em Wu et al. (2014). A Figura 5.1 ilustra um diagrama que representa uma classificação dos

algoritmos para detecção de obstáculos baseados em visão monocular. De acordo com Wu et al. (2014), os métodos para detecção de obstáculos a partir da visão monocular podem ser:

Figura 5.1: Classificação dos algoritmos para detecção de obstáculos com visão monocular.



Fonte: Wu et al. (2014).

a. Com base em correspondência de características (*Feature Matching*): utilizam alguns algoritmos para extração de características e correspondência de características de imagens.

b. Com base na distância focal (*Focal Distance*): baseiam-se em técnicas de variação da distância focal da câmera.

c. Com base no movimento (*Motion*): analisam o movimento de objetos em uma sequência de imagens, podendo ser de três tipos: diferença entre imagens, compensação do movimento ou fluxo óptico.

Para esta dissertação foram utilizados métodos baseados no movimento, mais especificamente métodos de fluxo óptico. Dessa forma, optou-se por fazer uma breve descrição baseada em Wu *et al.* (2014).

5.2.1. Método baseado no fluxo óptico

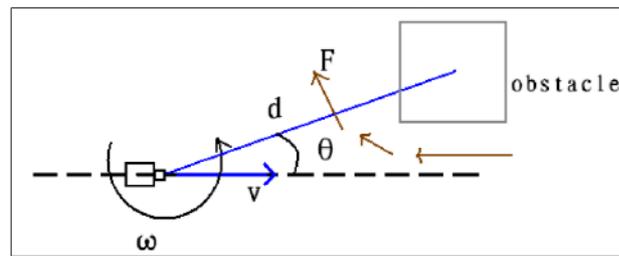
São amplamente utilizados na detecção de obstáculos, pois permitem rastrear movimentos entre o observador e os objetos da estrutura tridimensional do ambiente que o cerca. Existem vários algoritmos para determinar o fluxo óptico, entretanto os mais citados na literatura são os clássicos algoritmos de Lucas-Kanade e Horn-Shunck, ver detalhes em: Lucas e Kanade (1981) e, Horn e Shunck (1981).

O fluxo óptico pode ser gerado por dois tipos de movimento da câmera: $flow_{ir}$ (gerado pelo movimento de translação) e $flow_{rot}$ (gerado pelo movimento de rotação). Quando a distância entre o objeto no campo de visão da câmera é ' d ' e o ângulo entre o objeto e a direção

de movimento de translação é ' θ ', a câmera se move com uma velocidade de translação ' v ' e uma velocidade angular ' ω '. Então fluxo óptico gerado pelo objeto pode ser calculado pela Equação (5.1). A Figura 5.2 ilustra as variáveis de um processo de detecção de obstáculos.

$$F = \omega + \left(\frac{v}{d}\right) \sin \theta \quad (5.1)$$

Figura 5.2: Esquema do vetor de fluxo óptico para detecção de obstáculos.



Fonte: Wu et al. (2014).

Os pontos característicos que estão mais perto da câmera, geram um vetor maior de fluxo óptico. Ou seja, (quanto maior o tamanho do vetor de fluxo óptico, menor a distância entre o objeto e a câmera. A principal desvantagem da detecção de obstáculo a partir do fluxo óptico, se dá pelo fato de que a variação da luminosidade tem uma grande influência sobre este método, podendo causar erros na estimativa do TTC. Entretanto, técnicas poder ser utilizadas para reduzir esses erros.

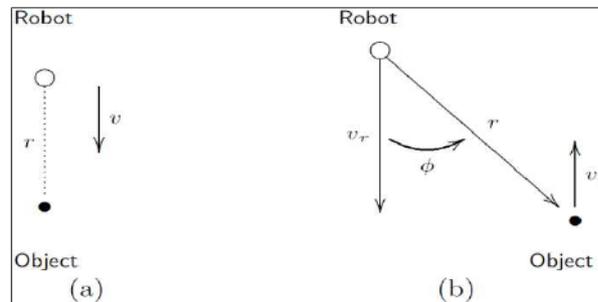
5.3. Técnicas para estimar o TTC por Visão Monocular

Conhecer o TTC entre um observador e um objeto é algo fundamental para qualquer sistema de navegação, seja robótica ou humana. Para determinar esse tempo muitas técnicas podem ser empregadas, essas técnicas podem ser divididas nas que são baseadas em visão (utilizam algum tipo de câmera como sensor) e as que não são (utilizam outro tipo de sensor). Diversos autores têm apresentado suas pesquisas fundamentadas em diferentes algoritmos ou algumas variações de algoritmos existentes. Entretanto, a seguir são apresentados apenas os alguns métodos baseados em visão monocular.

A teoria do TTC foi apresentada pela primeira vez por David N. Lee, quando conduziu muitos estudos sobre seres humanos e aves mostrando evidências de que o TTC é um componente crítico durante o movimento. Posteriormente esta abordagem passou a ser utilizada

na área da robótica móvel. Low e Wyeth (2005) mostraram essa ideia, onde a partir de uma única câmera obteve o TTC, conforme ilustra a Figura 5.3.

Figura 5.3: Abordagem de Lee - (a) robô em rota de colisão; (b) robô fora da rota de colisão.



Fonte: Low e Wyeth (2005).

Conforme a Figura 5.3(a), em que um robô se locomove com velocidade constante v a uma distância r de um dado obstáculo (objeto). Então o TTC é dado pela Equação (5.2).

$$TTC = \frac{r}{v} \quad (5.2)$$

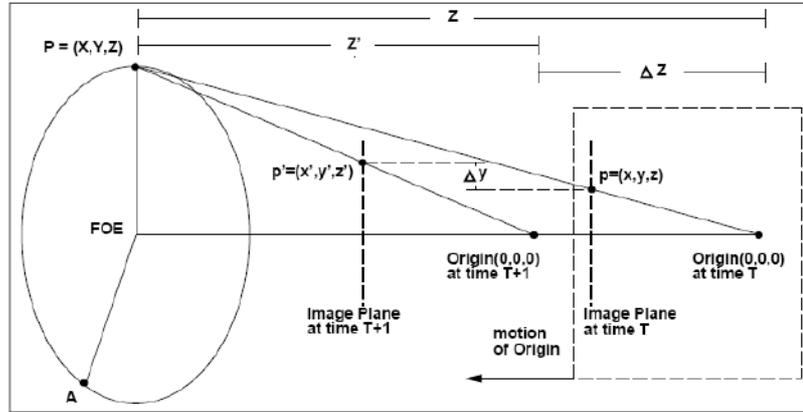
Uma outra abordagem proposta por David N. Lee para expressar o TTC é ilustrada na Figura 5.3(b). Onde utiliza o vetor velocidade v e o vetor de direção r para um ponto no objeto, formando um ângulo ϕ entre esses dois vetores. Dessa forma, a Equação (5.3) mostra uma das maneiras para estimar o TTC sem a necessidade de informações de velocidade e profundidade. Mais detalhes podem ser vistos em Lee (1976) e Low e Wyeth (2005).

$$TTC = \frac{\cos \phi \cdot \sin \phi}{\dot{\phi}} \quad (5.3)$$

Camus (1995) utiliza o fluxo óptico e apresenta uma geometria óptica para estimar o TTC, conforme Figura 5.4. Dado um ponto de interesse P com coordenadas (X, Y, Z) , projetado através do foco de projeção centrado na origem do sistema de coordenadas $(0, 0, 0)$, no espaço físico P é fixo e não se move. Se a câmera está voltada para a mesma direção que a direção de movimento, tem-se o foco de expansão (FOE), como sendo um ponto a partir do qual o fluxo óptico diverge. Assumindo que a direção da câmera aponta para a mesma direção no movimento de translação, o plano na imagem é fixado a uma distância z frontal a origem; por conveniência definiu $z=1$ (O valor real de z depende de fatores como a distância focal da câmera). Este plano

da imagem se move junto com a origem; P projeta sobre o ponto p neste plano. Como o plano da imagem se aproxima de P , a posição de p no plano da imagem também muda.

Figura 5.4: Geometria óptica para o TTC.



Fonte: Camus (1995).

A partir do triângulo equilátero da Figura 5.4, deduziu a Equação (5.4):

$$\frac{y}{z} = \frac{y}{1} = \frac{Y}{Z} \quad (5.4)$$

Diferenciando em relação ao tempo (em que \dot{a} representa a derivada temporal da/dt para uma dada variável a), analogamente tem-se:

$$\dot{y} = \frac{\dot{Y}}{Z} \cdot Y \left(\frac{\dot{Z}}{Z^2} \right) \quad (5.5)$$

Desde que P seja fixo, definido $\dot{Y} = 0$ e substituindo (yZ) por Y :

$$\dot{y} = \frac{\dot{Y}}{Z} \cdot y \left(\frac{\dot{Z}}{Z} \right) \quad (5.6)$$

Finalmente dividindo por y , tem-se o TTC:

$$TTC = \frac{\dot{y}}{y} = \frac{\dot{Z}}{Z} \quad (5.7)$$

O algoritmo apresentado em Camus (1995) e abordado por Baratoff et al. (1998), Sagrebin (2008) e em Liyanage e Perera (2012), consiste nos seguintes passos:

1. Calcular o FOE a partir de uma sequência de imagens consecutivas.
2. Encontrar características correspondentes em duas imagens consecutivas. Para cada uma destas correspondências de TTC.
3. Calcular os comprimentos dos vetores de disparidade ou vetores de fluxo óptico formado por duas características correspondentes. Os comprimentos dos vetores de disparidade fornecem uma estimativa para a divergência de um determinado ponto para o FOE.
4. Para cada característica relevante calcula-se a sua distância para o FOE e com isso o TTC.

Souhila e Karim (2007), Liyanage e Perera (2012) calcularam o TTC a partir do fluxo óptico extraído de uma sequência de imagens adquiridas durante o movimento. A velocidade da imagem pode ser descrita como uma função dos parâmetros da câmera e dividida em dois termos dependendo das componentes de rotação (V_t) e de translação (V_r) da velocidade (V) da câmera. A componente de rotação pode ser calculada a partir de informações de rotação e distância focal da câmera. Com o fluxo óptico global já calculado, então V_t é determinada pela subtração de V e V_r ($V_t = V - V_r$). A partir do movimento gerado no fluxo óptico devido à translação, o TTC é calculado pela Equação (5.8). Onde \bullet_i é a distância entre o ponto considerado (x_i, y_i) sobre o plano da imagem e o FOE. E $|\vec{V}_t|$ é a magnitude do fluxo óptico em movimento de translação.

$$TTC = \frac{\bullet_i}{|\vec{V}_t|} \quad (5.8)$$

Caldeira (2002) propõe um algoritmo baseado no fluxo óptico, onde as imagens são adquiridas considerando apenas o movimento de translação com velocidade linear constante. Então, os objetos são detectados e com base no TTC e uma estimativa da distância de cada objeto para a câmera é obtida. Dessa forma, o TTC pode ser calculado usando a Equação (5.9).

$$TTC = \frac{(x \bullet x_{FOE})}{u} = \frac{(y \bullet y_{FOE})}{v} = \frac{\sqrt{(x \bullet x_{FOE})^2 + (y \bullet y_{FOE})^2}}{\sqrt{u^2 + v^2}} \quad (5.9)$$

O foco de expansão é caracterizado por apresentar as componentes do fluxo óptico iguais a zero ($u = 0$ e $v = 0$). Porém, pode-se obter informação de profundidade na cena usando-se o FOE e o fluxo óptico para estimar o TTC, onde x e y são as coordenadas do *pixel* em consideração, x_{FOE} e y_{FOE} são as coordenadas do FOE, u e v são os valores das componentes do vetor de fluxo óptico neste *pixel*. A utilização do TTC como medida da profundidade relativa dos objetos na cena, parte do pressuposto que o campo de fluxo óptico é semelhante ao campo de movimento.

O procedimento implementado para determinar o FOE consiste em verificar, linha por linha de regiões, em que coluna da imagem o vetor de fluxo óptico passa de um ângulo maior a um menor que 90° , ou vice-versa, ou então de um ângulo maior a um menor que 270° , ou vice-versa. O valor de x_{FOE} pode ser, então, obtido como a média dos valores de coluna encontrados para cada linha. De maneira similar, o valor de y_{FOE} pode ser estimado verificando-se, coluna por coluna de regiões, em que linha da imagem o vetor de fluxo óptico passa de um ângulo maior a um menor que 0° , ou vice-versa, ou de um ângulo maior a um menor que 180° , ou vice-versa. Finalmente, o valor de y_{FOE} pode ser calculado como a média dos valores de linha obtidos para cada coluna.

Conhecendo as coordenadas do FOE, calculou o TTC usando a terceira das igualdades da Equação (5.8), em uma imagem dividida em regiões de 20×20 pixels. Sendo assim, o TTC foi calculado para cada região da imagem. Posteriormente, um único valor de TTC foi determinado para representar cada um dos objetos obtidos por um processo de segmentação. E o TTC resultante é determinado pelo valor mediano dos TTC das regiões pertencentes ao objeto.

Alenyà et al. (2009) propôs um método para mensurar o TTC baseado no contorno ativo, *Active Contour Affine Scale* (ACAS). Ainda faz uma comparação com outros dois métodos baseados em: escala invariante - *Scale Invariant Ridge Segments* (SIRS), e derivadas do brilho da imagem - *Image Brightness Derivatives* (IBD). Entretanto, apenas os métodos ACAS e IBD são descritos a seguir.

1) *Active Contour Affine Scale* (ACAS). Apresenta um método para estimar do TTC, baseado no rastreamento do contorno ativo. Dada uma deformação de um conjunto de pontos projetados entre dois frames diferentes Q e Q' , esta pode ser parametrizada com uma transformação afim. Então, a partir da Equação (5.10), o TTC pode ser estimado.

$$Q \bullet \bullet Q = WS \quad (5.10)$$

onde Q é um vetor formado com as coordenadas de N_Q pontos, primeiro todas as coordenadas x de Q_x , em seguida todas as coordenadas y de Q_y , e W é a forma de uma matriz composta por Q^x , Q^y , e os vetores dimensionais $0=(0,0, \dots,0)^T$ e $1=(1,1,\dots,1)^T$.

$$W = \begin{bmatrix} 1 & 0 & Q^x & 0 & 0 & Q^y \\ 0 & 1 & 0 & Q^y & Q^x & 0 \end{bmatrix} \quad (5.11)$$

$$S = (t_x, t_y, M_{11} \bullet 1, M_{22} \bullet 1, M_{21}, M_{12}) \quad (5.12)$$

Em que S é a forma de um vetor de 6 dimensões vector que codifica a deformação da imagem a partir da primeira para a segunda visualização. A translação dimensionada em profundidade da câmara para o objeto pode ser obtida com:

$$\frac{T_z}{Z_0} = \frac{1}{\sqrt{\bullet_1}} \bullet 1 \quad (5.13)$$

Sendo \bullet_1 o maior valor obtido pela multiplicação da matriz 2×2 M pela sua transposta M^T , formada a partir dos elementos que resultaram em uma nova matriz MM^T . Considerando o período de amostragem constante, então pode ser usado a diferença entre frames consecutivos $i-1$, onde i é uma estimativa para a velocidade da mudança de escala. Definindo a profundidade no frame i como:

$$\bullet_i = \frac{T_z}{Z_0} + 1 \quad (5.14)$$

Em seguida, a diferença de escala em dois frames consecutivos é:

$$\bullet_i \bullet_{i+1} = \frac{T_{z_i} \bullet T_{z_{i+1}}}{Z_0} \quad (5.15)$$

Nas equações surgiu uma distância desconhecida Z_0 , na prática, o autor considerou $Z_0=1$, que irá dimensionar todas as translações entre 1 e 0. Assim, o TTC é calculado pela Equação (5.16).

$$TTC = \frac{\dot{I}_i}{\dot{I}_i \dot{I}_{i+1}} = \frac{T_{Z_i} + Z_0}{T_{Z_i} \cdot T_{Z_{i+1}}} \quad (5.16)$$

2) Image Brightness Derivatives (IBD) - Estudos apresentados em Horn et al. (2007, *apud* ALENYÀ et al. 2009) mostram que o TTC pode ser estimada utilizando apenas derivadas espaciais e temporais do brilho da imagem. O método baseia-se na ideia de que o brilho é constante e descarta o uso de técnicas que se baseiam em características ou detecção de objetos, rastreamento de características, estimativa de fluxo óptico, por exemplo. Assim, partindo da Equação (5.17), o TTC pode ser obtido.

$$\frac{d}{dt} E(x, y, t) = 0 \quad (5.17)$$

Assumindo que o brilho na imagem é constante, ou seja, o brilho de um ponto da imagem na cena não se altera significativamente, então a derivada de (5.17) pode ser expandida:

$$uE_x + vE_y + E_t = 0 \quad (5.18)$$

onde $u=dx/dt$, $v=dy/dt$ são as componentes x e y do movimento do campo na imagem, enquanto E_x , E_y , E_t são as derivadas parciais do brilho x , y , e t .

Ao utilizar um modelo de projeção em perspectiva, uma relação pode ser encontrada entre o movimento de campo e o TTC e, então, uma relação direta entre as derivadas de imagem e TTC. No caso especial de movimento de translação ao longo do eixo óptico da câmera em direção a um plano perpendicular ao eixo óptico, (5.18) pode ser reformulada como:

$$\frac{x E_x + y E_y}{TTC} + E_t = 0 \quad (5.19)$$

Ou

$$\frac{G}{TTC} + E_t = 0 \quad (5.20)$$

Onde G é o gradiente radial ($xE_x + yE_y$), x e y são as coordenadas do pixel em consideração. Um método de mínimos quadrados aplicado em todos os pixels pode ser usado para estimar uma média do TTC, resultando na Equação (5.21).

$$TTC = \frac{\sum G^2}{\sum GE_t} \quad (5.21)$$

Watanabe et al. (2015) propõe um método para calcular o TTC a partir de informações fotométricas em imagens, não precisando de qualquer informação adicional. Quando uma fonte de luz se move na cena 3D, uma intensidade pode ser observada e muda de acordo com o movimento ou variação da fonte de luz. A partir daí acontecem mudanças na intensidade da imagem e informações fotométricas nas imagens são usadas para estimar o TTC. Este método apresenta um diferencial em relação a Horn et al. (2007, *apud* ALENYÀ et al. 2009), pois permite variações na intensidade luminosa da cena.

Pundlik et al. (2011) apresenta uma abordagem para determinar o TTC e o risco de colisão com obstáculos. O TTC é calculado utilizando a razão do tamanho do obstáculo e a sua derivada no tempo. Para calcular essa relação foram utilizadas informações locais de mudança de escala e movimento, obtidas a partir da detecção e rastreamento de pontos característicos em uma sequência de imagens. A partir dessas mesmas informações, uma medida para prever o risco de colisão foi obtida para obstáculos que se deslocam ao longo do eixo óptico da câmera. À medida que um obstáculo se afasta do eixo óptico da câmera, o risco diminui.

A detecção e rastreamento de pontos característicos é executada na sequência de imagens de entrada. A detecção de pontos característicos baseia-se no algoritmo de Shi-Tomasi, já no rastreamento desses pontos foi usado o algoritmo de fluxo óptico Lucas-Kanade Piramidal, onde apenas os pontos característicos acima de um *threshold* foram selecionados para serem rastreados. O cálculo da mudança de escala é realizado na vizinhança de cada ponto característico, e um conjunto de pontos característicos onde há um aumento na escala local entre dois frames de uma sequência é obtido. Para esse conjunto de pontos característicos, um modelo de movimento afim² é provido, levando a um grupo de características associadas com um obstáculo potencial. A utilização de pontos característicos e do modelo de movimento afim

² “O modelo de movimento afim é uma boa aproximação para a imagem do movimento 3D projetado em 2D quando a distância da câmera em relação aos objetos é grande se comparada com o tamanho dos objetos”. (SILVA, 2011 p. 21). Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/39108/000824188.pdf?sequence=1>



proporciona flexibilidade para representar um obstáculo em movimento. A partir de características associadas com o obstáculo, é estimado TTC e risco de colisão.

Apesar dos resultados terem sido considerados satisfatórios pelo autor, algumas propostas de melhoria foram deixadas. Como trabalho futuro incluiu uma extensão do algoritmo para lidar com cenas complexas do mundo real, envolvendo múltiplos obstáculos, movendo-se a câmera em diferentes posições e sentidos de rotação e translação, e em condições de iluminação variáveis. Algumas dessas melhorias foram aprimoradas e apresentadas em Pundlik et al. (2013).

Em 2013 propôs um sistema para detecção de colisão em tempo real desenvolvido para deficientes visuais. O sistema foi aprimorado da versão de 2011 e calcula o fluxo óptico esparsos em uma sequência de imagens, porém a rotação da câmera é compensada usando sensor giroscópio externo. Dessa forma, o TTC e o risco de colisão são estimados com base no movimento.

Uma vez que a estimativa de movimento foi realizada, o próximo passo é estimar o risco de colisão. O autor utilizou duas quantidades para determinar o risco de colisão: TTC (aspecto temporal), e ponto de colisão (o aspecto espacial). A fim de executar computações locais usando o movimento de imagem, uma vizinhança local é estabelecida com base em uma triangulação *Delaunay* dos pontos localizados e rastreados. Com as características detectadas e rastreadas, os pontos vizinhos característicos são encontrados usando a triangulação. Para cada ponto característico, é obtido um conjunto de vizinhos espaciais imediatos na imagem que compartilham uma borda da triangulação com essa característica.

Dado p_i , como sendo um ponto i^{th} rastreado no primeiro frame para um ponto q_i no próximo frame. Em geral, existem N pontos característicos rastreados para um determinado par de frames ($i = 0, 1, \dots, N-1$). Para cada par de pontos característicos rastreados (p_i, q_i), um conjunto de pontos característicos que são seus vizinhos imediatos nos frames correspondentes, são dados pela triangulação *Delaunay* $D(p_i, q_i)$.

Com os pontos característicos rastreados e uma vizinhança local estabelecida, o próximo passo consiste em calcular a mudança de escala local para cada conjunto de pontos característico. A medida que o objeto se aproxima da câmera, seu tamanho aumenta gradualmente. Considerando que o objeto se aproxima movendo-se ao longo do eixo óptico da câmera, a distância entre a vizinhança e os pontos característicos também aumenta. Essa mudança de escala local entre dois frames de uma sequência é dada pela Equação (5.22).

$$e_i = \frac{\sum_{p_k, q_k \in D(p_i, q_i)} (\|q_i \cdot q_k\| \cdot \|p_i \cdot p_k\|)}{\sum_{p_k, q_k \in D(p_i, q_i)} (\|p_i \cdot p_k\|)} \quad (5.22)$$

onde e_i é o valor de escala local normalizado para o ponto característico em análise. Quando esse valor é positivo, significa que a vizinhança de um ponto se expande, caso contrário, para um valor negativo, essa vizinhança se aproxima.

O TTC, calculado localmente, está inversamente relacionado com a expansão local (e_i). Assim, o valor do TTC para cada ponto característico da vizinhança associado com o movimento é dado por $t_i = 1/e_i$. Para um grupo de características, o valor do TTC foi aproximado pela média dos valores de TTC obtidos a partir dos pontos característicos pertencentes ao grupo. Esse valor foi convertido em segundos usando uma taxa de frames por segundo adquiridos pela câmera (o autor usou 30 frames por segundo). O Ponto de colisão, foi então calculado localmente com base na relação de mudança de escala e o movimento de translação.

O risco de colisão poderia ser obtido a partir de um limiar para o valor do TTC, tal que abaixo desse valor o risco é provável. Entretanto, essa aproximação tende a considerar qualquer obstáculo detectado no campo visual da câmera (com um TTC abaixo desse limiar), podendo detectar obstáculos que não representam um risco real.

A fim de avaliar o risco representado por colisões com obstáculos de diferentes trajetórias, Pundlik et al. (2011) propôs o que chamou de fator de risco de colisão. Esse fator é baseado na razão entre a mudança de escala local e o movimento local entre dois frames. Para um obstáculo movendo-se ao longo do eixo óptico da câmera essa relação seria elevada, à medida que esse obstáculo se afasta do eixo óptico esse fator vai diminuindo. Para um ponto característico i^{th} , a razão de mudança de escala local e movimento é dada por $d_i^{(j)} = e_i^{(j)} / v_i$, onde j^{th} é um frame e $v_i = \|p_i^{(j)} \cdot p_i^{(j-b)}\|$ é a magnitude do vetor de movimento do ponto característico i^{th} entre os frames j^{th} e $j^{(j-b)}$. Para um obstáculo no frame j^{th} de uma sequência de imagens, o fator de risco de colisão é dado pela Equação (5.23), em que m é o número de características que pertencem ao objeto em movimento (PUNDLIK, 2011).

$$C^{(j)} = \frac{1}{m \cdot p_i^{(j)} \cdot F^{(j)}} \sum d_i^{(j)} \quad (5.23)$$

5.4. Considerações finais do capítulo

Conhecer o TTC entre um observador e um objeto é fundamental para qualquer sistema de navegação, seja robótica ou humana. Para determinar esse tempo muitas técnicas podem ser empregadas, e elas podem ser divididas nas que utilizam visão (algum tipo de câmera como sensor) e as que não utilizam a visão (empregam outro tipo de sensor). Diversos autores têm apresentado suas pesquisas fundamentadas em diferentes algoritmos ou algumas variações de algoritmos existentes. Entretanto, este capítulo apresentou apenas alguns dos métodos baseados em visão monocular.

O Capítulo seguinte irá apresentar detalhes do algoritmo desenvolvido para estimar o TTC e detectar colisões.

CAPÍTULO 6 – O ALGORITMO PARA DETECÇÃO DE COLISÕES

6.1. Introdução

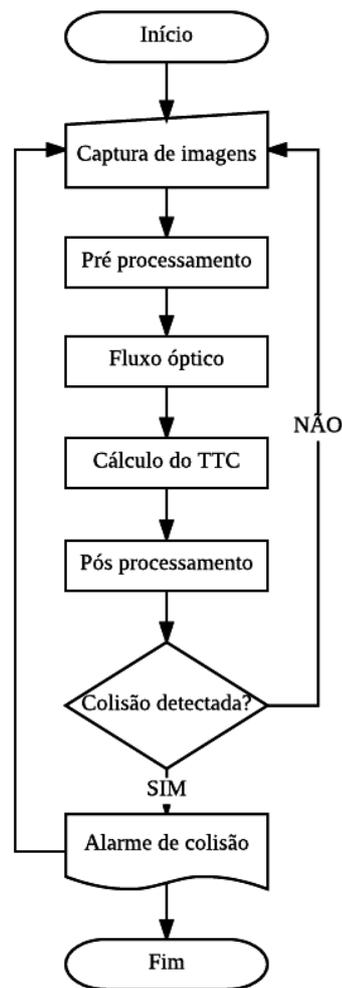
No Capítulo anterior, menciona-se que, segundo Wu et al. (2014), os métodos para detecção de obstáculos com visão monocular podem ser classificados em três tipos: com base em correspondência de características, com base na distância focal, e com base no movimento. Sendo que os métodos com base em movimento podem ser dados pela diferença entre os pontos de imagens, compensação do movimento ou fluxo óptico (ver Fig. 5.1). Entre os métodos citados, para o algoritmo desenvolvido, utilizou-se um modelo baseado no fluxo óptico.

Dos diversos algoritmos para estimar movimentos, empregou-se o algoritmo de Gunnar Farnebäck, que detecta movimentos por meio do fluxo óptico denso (FARNEBÄCK, 2012). O algoritmo de Farnebäck consiste em um método diferencial que, segundo estudos feitos em Aires (2009), possuem características que facilitam o processamento em tempo real, o que é extremamente necessário para a estimação do *TTC* e detecção de obstáculos, uma vez que a determinação desse tempo é imprescindível para evitar uma possível colisão. O algoritmo se resume na Equação (6.1) que define o frame atual (*next*) como sendo o frame anterior (*prev*) acrescido do deslocamento do fluxo óptico $flow_x$ e $flow_y$ (deslocamentos em x e y , respectivamente, ou u e v no plano da imagem), ou seja, ele utiliza o frame anterior como referência.

$$frame_{prev}(x,y) \bullet frame_{next}(x + flow_x, y + flow_y) \quad (6.1)$$

No entanto, apenas com o fluxo óptico não é possível determinar o *TTC* entre a câmera e um obstáculo. Desta forma, foi desenvolvida uma equação fechada a partir do modelo de uma câmera pinhole para estimar esse tempo. Então, foi desenvolvido um algoritmo que é basicamente apresentado no fluxograma da Figura 6.1. Detalhes de cada etapa são descritos logo em seguida, bem como o equacionamento do *TTC* a partir do modelo de uma câmera pinhole e do fluxo óptico. Detalhes dos experimentos realizados e seus resultados serão apresentados e discutidos no Capítulo 7.

Figura 6.2: Fluxograma básico do algoritmo de detecção de colisões.



Fonte: O autor.

6.2. Etapas do algoritmo

Conforme o fluxograma da Figura 6.1, o algoritmo segue as seguintes etapas:

1. Captura de imagens - Aquisição de uma sequência de *frames*.

As imagens são adquiridas por meio de uma câmera monocular RGB ou em escala de cinzas (mais detalhes no próximo Capítulo).

2. Pré-processamento

Visando diminuir o tempo de processamento as imagens foram redimensionadas (reduzidas) e em seguida convertidas para escala de cinzas (no caso da câmera RGB). Além disso, para suavizar a imagem e reduzir possíveis interferências causadas por ruídos aleatórios



provenientes do processo de aquisição e, que podem interferir nas etapas posteriores, foi aplicado um filtro Gaussiano com desvio padrão de 1,1 com uma máscara de 15x15 (parâmetros recomendados pelo OpenCV).

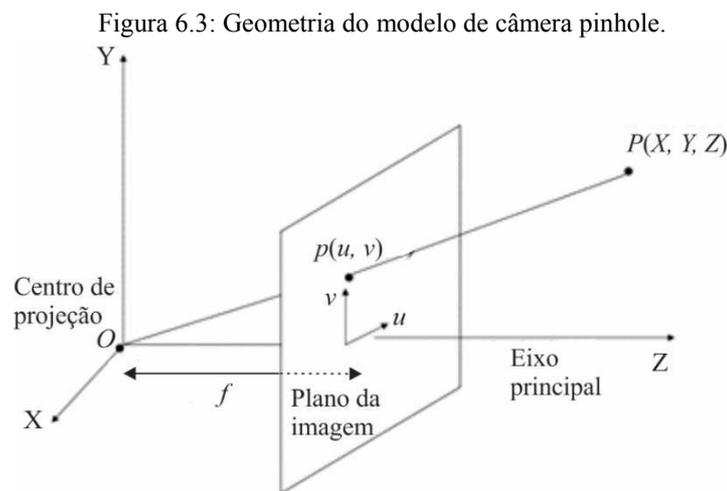
3. Fluxo óptico

Dados os frames atual ($frame_{next}$) e anterior ($frame_{prev}$) utilizado como referência, nos instantes (t) e ($t-1$), respectivamente, o algoritmo Farneback foi aplicado para calcular o fluxo óptico denso, retornando como resultado um vetor de fluxos $[flow_x; flow_y]$ nas direções x e y para cada ponto da imagem no plano. Esse vetor representa o deslocamento de cada ponto em dois frames consecutivos, e será utilizado em uma das etapas para o equacionamento do TTC .

4. Cálculo do TTC

O TTC é calculado a partir de um modelo de uma câmera pinhole e da variação do fluxo óptico para estimar os movimentos dos pontos, conforme detalhes do método a seguir.

Dado um modelo ideal de uma câmera pinhole, pode ser descrita uma relação geométrica entre as coordenadas de um ponto no espaço tridimensional 3D $P(X, Y, Z)$ e sua projeção perspectiva correspondente mapeada no plano da imagem 2D $p(u, v)$. A Figura 6.2 ilustra essa relação. A partir dessa relação, foi obtida uma equação fechada para calcular o TCC .



Fonte: adaptação do autor.

Observa-se o centro óptico da câmera em 'O', como sendo o centro de projeção no qual tem-se o sistema de coordenadas nos eixos x, y, z , bem como, a origem desse sistema. A distância entre o centro óptico e o plano de projeção 2D (plano da imagem) é a distância focal f . A linha perpendicular ao plano da imagem que passa pelo centro óptico é o eixo óptico (eixo

principal), sendo esse colinear ao eixo z do sistema de coordenadas. O ponto de interseção do plano da imagem com o eixo óptico é chamado de ponto principal.

Em seguida, pode-se modelar uma projeção em perspectiva de pontos 3D no plano 2D. O modelo da câmera pinhole descreve a relação entre um ponto no espaço 3D e suas coordenadas correspondentes no plano da imagem 2D, no qual a projeção do ponto 3D $P(X, Y, Z)$ é dada pela posição correspondente do ponto $p(u, v)$ no plano da imagem 2D. Então, por semelhança de triângulos encontra-se o ponto correspondente, conforme Equação 6.2.

$$\frac{f}{Z} = \frac{u}{X} = \frac{v}{Y} \quad (6.2)$$

Isolando as coordenadas no ponto, tem-se:

$$u = \frac{f \times X}{Z}, \quad v = \frac{f \times Y}{Z} \quad (6.3) \quad (6.4)$$

Logo, a projeção correspondente ao ponto P é mapeada e conseqüentemente, sua posição correspondente do ponto $p(u, v)$ é obtida, conforme a Equação 6.5.

$$P(X, Y, Z) \cdot p\left(\frac{f \times X}{Z}, \frac{f \times Y}{Z}\right) \quad (6.5)$$

Utilizando coordenadas homogêneas a relação também pode ser expressa em notação de matriz. Referindo-se ao modelo de câmera pinhole, essa matriz é usada para denotar um mapeamento projetivo de coordenadas 3D para coordenadas do ponto em 2D. Da matriz tem-se $\lambda = Z$, que representa o fator de escalas homogêneas. Observar a Equação 6.6.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (6.6)$$

Considerando os parâmetros intrínsecos e extrínsecos da câmera, tem-se a Equação 6.7. Na qual, a posição da projeção do ponto no plano da imagem é dada pelo produto dos parâmetros da câmera (K) pela posição do ponto no espaço 3D.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (6.7)$$

Sendo K uma matriz que contém 5 parâmetros intrínsecos da câmera, conforme a Equação 6.8. Esses parâmetros envolvem a distância focal, formato do sensor de imagem e o ponto principal.

$$K = \begin{bmatrix} \bullet & \bullet & u_0 \\ 0 & \bullet & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6.8)$$

Os parâmetros α e β representam a distância focal da câmera em termos de pixel nas direções x e y respectivamente, podem ser reescritos como f_x e f_y . O parâmetro γ é o coeficiente *skew* da câmera, está relacionado com os ângulos entre os eixos nas direções x e y da imagem, sendo nulo em casos onde essas direções são perpendiculares. Os parâmetros u_0 e v_0 representam o ponto principal que idealmente encontra-se no centro da imagem, ou seja, na origem do eixo óptico $(0,0)$. Além dos parâmetros intrínsecos da câmera, podem ser considerados dois parâmetros extrínsecos: movimentos de rotação e translação, no qual, R representa a matriz de rotação e T é a matriz de translação.

Dessa forma, considerando os parâmetros da câmera, a Equação 6.7 pode ser reescrita:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \bullet & \bullet & u_0 \\ 0 & \bullet & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} & m_{13} & t_x \\ m_{21} & m_{22} & m_{23} & t_y \\ m_{31} & m_{32} & m_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (6.9)$$

Visando simplificar o modelo e reduzir o número de variáveis no equacionamento, algumas considerações iniciais foram assumidas: i) a origem do eixo óptico (u_0, v_0) em $(0, 0)$; ii) os parâmetros $\alpha = \beta = f$; $\gamma = 0$; iii) a matriz de rotação igual a matriz identidade ($R = I$), assumindo que o algoritmo seja robusto a rotações; iv) o movimento de translação só ocorre no eixo z ($t_z \neq 0$), logo $t_x = t_y = 0$; Assim, tem-se a Equação 6.10.

$$\bullet \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (6.10)$$

O produto das matrizes correspondentes aos parâmetros intrínsecos e extrínsecos da câmera e das variáveis de rotação e translação, $[K] \times [R \ T]$, resultando na Equação 6.11.

$$\bullet \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (6.11)$$

Uma nova multiplicação é realizada entre a resultante do produto dos parâmetros e as coordenadas do ponto P , como resultado segue a Equação 6.12.

$$\bullet \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} fx \\ fy \\ z + t_z \end{bmatrix} \quad (6.12)$$

Da Equação 6.12, para cada deslocamento de translação (no eixo z) de um ponto no espaço 3D, encontra-se o ponto correspondente no plano da imagem 2D. Ou seja, considerando o movimento de translação em z em dois instantes de tempo consecutivos, encontra-se a posição relativa do ponto nos instantes $(t-1)$ e (t) . Essa posição e os intervalos de tempo são consideradas no equacionamento do algoritmo.

Até o momento foi realizada uma análise simplificada. No entanto, novas considerações devem ser realizadas e analisadas. Conforme segue:

Como os movimentos de translação em x e y são desprezados ($t_x = t_y = 0$), considera-se que um ponto se desloca em direção ao centro óptico da câmera apenas no eixo z ($t_z \neq 0$). Desse modo, em um primeiro momento, considerou-se um cenário simplificado em que um objeto se move apenas em linha reta em direção à câmera. Assim, uma análise é feita a partir da translação em z (t_z) para um mesmo ponto em 3D, em dois instantes de tempo consecutivos. Detalhes do equacionamento são apresentados a seguir.

Da física, tem-se que o deslocamento (d) é a variação da velocidade (V) no tempo (t), $d = V \times t$. Transcrevendo essa abordagem para o modelo desenvolvido, tem-se que o deslocamento em z (t_z) é igual a d ; V está relacionada com a velocidade dos pontos no plano da imagem dada pelo fluxo óptico (utiliza-se V_z , pois representa o deslocamento de um ponto P no eixo z , que tem sua projeção no plano da imagem p , no qual a variação dessa posição corresponde a velocidade do ponto); t está relacionado ao t_{tc} ou a dt . Onde dt é a diferença de tempo entre em dois frames consecutivos nos instantes ($t-1$) e (t), e está relacionado com a taxa de frames por segundo. Então:

$$t_z = V_z \times dt \quad (6.13)$$

$$z = V_z \times t_{tc} \quad (6.14)$$

Isolando o t_{tc} :

$$t_{tc} = \frac{z}{V_z} \quad (6.15)$$

Considerando os instantes ($t-1$) e (t), as equações 6.3 e 6.4 podem ser reescritas. Onde (u_{t-1} , v_{t-1}) e (u_t , v_t) representam a posição do ponto na imagem em dois instantes de tempo consecutivos, conforme as equações 6.16, 6.17 e 6.18, 6.19.

$$u_{(t-1)} = \frac{f \times x}{z}, \quad v_{(t-1)} = \frac{f \times y}{z} \quad (6.16) \quad (6.17)$$

$$u_{(t)} = \frac{f \times x}{z + t_z}, \quad v_{(t)} = \frac{f \times y}{z + t_z} \quad (6.18) \quad (6.19)$$

Para determinar o t_{tc} , uma relação é estabelecida pela Equação 6.20. Essa relação possibilita mapear o deslocamento de um mesmo ponto e a sua posição correspondente no plano da imagem. Analogamente uma relação pode ser determinada entre $v_{(t-1)}$ e $v_{(t)}$.

$$\frac{u_{(t)}}{u_{(t-1)}} = \frac{\frac{f \times x}{z + t_z}}{\frac{f \times x}{z}} = \frac{z}{z + t_z} \quad (6.20)$$

Substituindo 6.13 e 6.14 em 6.20,

$$\frac{u_{(t)}}{u_{(t+1)}} = \frac{ttc \times \mathcal{V}_z}{ttc \times \mathcal{V}_z + V_z \times dt} = \frac{ttc}{ttc + dt} \quad (6.21)$$

Isolando o ttc , considerando as direções u e v , tem-se as equações 6.22 e 6.23,

$$ttc_{(u)} = \bullet \frac{u_{(t)}}{u_{(t)} \bullet u_{(t+1)}} \times dt \quad (6.22)$$

Da mesma forma em relação a v ,

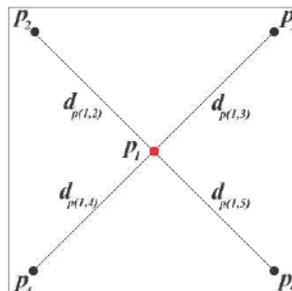
$$ttc_{(v)} = \bullet \frac{v_{(t)}}{v_{(t)} \bullet v_{(t+1)}} \times dt \quad (6.23)$$

A Equação 6.22 consiste na análise realizada para um único ponto, considerando as variáveis na direção u . Uma equação análoga para ttc é obtido se forem consideradas as variáveis na direção v , conforme Equação 6.23. No entanto, essas equações ainda não são suficientes para determinar o TTC , pois são para um único ponto p .

Um *frame* é composto por centenas ou milhares de pontos e, cada ponto corresponde à projeção de cada um dos pontos da cena tridimensional; assim a análise de um único ponto não é suficiente para determinar o TTC de um objeto com a câmera. Pois, o objeto é representado por um conjunto de pontos mapeados no plano da câmera.

No entanto, uma equação de forma fechada para o TTC ainda pode ser encontrada quando mais pontos são considerados. Então, uma heurística foi utilizada para estimar esse tempo, conforme demonstrado a seguir:

Figura 6.4: Ilustração da relação entre os pontos p_1 e p_a .



Fonte: O autor.

Seja $p_a = \{p_2, p_3, p_4, p_5\}$ um conjunto de pontos equidistantes em torno de um ponto central p_1 , como ilustrado na Figura 6.3. Com base nas distâncias entre p_1 e p_a ($d_{p(1,2)}$, $d_{p(1,3)}$, $d_{p(1,4)}$, $d_{p(1,5)}$), as equações 6.24 e 6.25 são definidas, no qual determina-se a distância entre p_1 e os outros pontos nos instantes $(t-1)$ e (t) considerando as direções u e v .

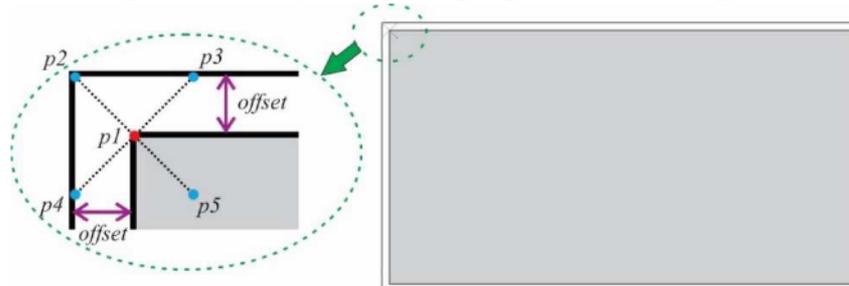
$$d_{t-1}(p_1, p_a) = \sqrt{(u_{p_1(t-1)} \cdot u_{p_a(t-1)})^2 + (v_{p_1(t-1)} \cdot v_{p_a(t-1)})^2} \quad (6.24)$$

$$d_t(p_1, p_a) = \sqrt{(u_{p_1(t)} \cdot u_{p_a(t)})^2 + (v_{p_1(t)} \cdot v_{p_a(t)})^2} \quad (6.25)$$

Das equações acima, para todos os pontos p_a , $d_{t-1}(p_1, p_a)$ é a distância euclidiana entre p_1 e p_a no instante $(t-1)$. Da mesma forma, no instante (t) , porém são considerados os vetores de fluxo óptico calculados pelo algoritmo de Farneback, então a distância euclidiana $d_t(p_1, p_a)$ considera a posição do ponto no instante $(t-1)$ acrescida dos valores dos vetores nas direções u e v para cada ponto p_1 .

Antes de continuar com o equacionamento, cabe destacar detalhes. Observando as Figuras 6.4 e 6.3, algumas considerações são feitas.

Figura 6.5: Ilustração dos pontos p_1 e p_a considerando o *offset*.



Fonte: O autor.

Dada a Figura 6.4, observa-se um retângulo que corresponde a um *frame*, a parte cinza é a região de interesse a ser varrida para determinar o *TTC* de cada ponto p_1 a partir da relação das distâncias euclidianas entre p_1 e p_a nos instantes $(t-1)$ e (t) , conforme a Equação 6.26. A parte em branco desse retângulo não é considerada. Uma ampliação é ilustrada nessa mesma Figura, no qual observam-se os pontos p_1 a p_a , bem como o termo *offset*. Com o *offset* delimitou-se a região do *frame* a ser varrida, uma vez que o total de pontos p_1 a serem analisados é igual a n , sendo n igual ao total de pixels do *frame* menos o *offset*. Pelos experimentos obteve-se melhores resultados com *offset*=20 pixels. Além disso, foi possível eliminar possíveis erros decorrentes das bordas do *frame*.

$$ratio_d(p_1, p_a) = \frac{d_{(t)}(p_1, p_a)}{d_{(t-1)}(p_1, p_a)} \quad (6.26)$$

Substituindo 6.24 e 6.25 em 6.26

$$ratio_d(p_1, p_a) = \frac{\sqrt{(u_{p_1(t)} \cdot u_{p_a(t)})^2 + (v_{p_1(t)} \cdot v_{p_a(t)})^2}}{\sqrt{(u_{p_1(t-1)} \cdot u_{p_a(t-1)})^2 + (v_{p_1(t-1)} \cdot v_{p_a(t-1)})^2}} \quad (6.27)$$

Equações análogas às 6.16, 6.17 e 6.18, 6.19 são obtidas, onde 6.28 e 6.29 representam a posição dos pontos no instante $(t-1)$, 6.30 e 6.31 corresponde a esse mesmo ponto em (t) , já incluindo o fluxo óptico. Observa-se que nessas equações são considerados os movimentos de translação em x , y e z .

$$(u_{p_1(t-1)} \cdot u_{p_a(t-1)}) = \frac{f(x_{R_1} \cdot x_{P_a})}{z} \quad (6.28)$$

$$(v_{p_1(t-1)} \cdot v_{p_a(t-1)}) = \frac{f(y_{R_1} \cdot y_{P_a})}{z} \quad (2.29)$$

$$(u_{p_1(t)} \cdot u_{p_a(t)}) = \frac{f(x_{R_1} + t_x)}{z + t_z} \cdot \frac{f(x_{P_a} + t_x)}{z + t_z} = \frac{f(x_{R_1} \cdot x_{P_a})}{z + t_z} \quad (6.30)$$

$$(v_{p_1(t)} \cdot v_{p_a(t)}) = \frac{f(y_{R_1} + t_y)}{z + t_z} \cdot \frac{f(y_{P_a} + t_y)}{z + t_z} = \frac{f(y_{R_1} \cdot y_{P_a})}{z + t_z} \quad (6.31)$$

Substituindo 6.28, 6.29 e 6.30, 6.31 em 6.27, tem-se:

$$ratio_d(p_1, p_a) = \frac{\sqrt{\left[\frac{f(x_{R_1} \cdot x_{P_a})}{z + t_z}\right]^2 + \left[\frac{f(y_{R_1} \cdot y_{P_a})}{z + t_z}\right]^2}}{\sqrt{\left[\frac{f(x_{R_1} \cdot x_{P_a})}{z}\right]^2 + \left[\frac{f(y_{R_1} \cdot y_{P_a})}{z}\right]^2}} \quad (6.32)$$

$$ratio_d(p_1, p_a) = \frac{\frac{1}{z+t_z} \sqrt{[f(x_{R_1} \cdot x_{P_a})]^2 + [f(y_{R_1} \cdot y_{P_a})]^2}}{\frac{1}{z} \sqrt{[f(x_{R_1} \cdot x_{P_a})]^2 + [f(y_{R_1} \cdot y_{P_a})]^2}} \quad (6.33)$$

Resultando na Equação 6.34

$$ratio_d(p_1, p_a) = \frac{z}{z+t_z} \quad (6.34)$$

Como, das equações 6.13 e 6.14 tem-se que $t_z = V_z \cdot dt$ e $z = ttc \cdot V_z$. O valor de $ratio_d(p_1, p_a)$, ou simplesmente $ratio_d$, é apresentado em 6.35.

$$ratio_d = \frac{ttc \cdot V_z}{ttc \cdot V_z + V_z \cdot dt} = \frac{ttc}{ttc + dt} \quad (6.35)$$

Isolando o ttc :

$$ttc = \frac{ratio_d}{(ratio_d \cdot 1)} \cdot dt \quad (6.36)$$

Esse tempo é estimado para cada uma das quatro relações dos pontos (p_1, p_a) , isso significa que para cada ponto p_1 são retornados quatro valores de ttc . Agora pode-se chamar ttc como ttc_{p_1} , uma vez que a análise é feita para cada ponto p_1 , conforme Equação 6.37.

$$ttc_{p_1}(p_1, p_a) = \frac{ratio_d}{(ratio_d \cdot 1)} \cdot dt \quad (6.37)$$

Quando um objeto está se afastando da câmera, o valor retornado em ttc_{p_1} é negativo. No entanto, o interesse está em detectar apenas objetos que estão se movendo em direção à câmera, uma vez que esses podem representar um risco de colisão.

Dessa forma, um novo valor para ttc_{p_1} pode ser definido, para um tempo igual a $\lim_{ttc_{p_1}}$. Isso poderá acontecer em algumas situações de acordo com o tempo calculado em 6.37, conforme apresenta a Equação 6.38. Caso o ttc_{p_1} calculado em 6.37 seja menor (ou igual) que zero ou maior que $\lim_{ttc_{p_1}}$, o valor de ttc_{p_1} será igual a $\lim_{ttc_{p_1}}$. Caso contrário, ttc_{p_1} é o mesmo valor calculado em 6.37, conforme a Equação 6.39. Esse limite também foi definido pois em

casos onde o $t_{tc_{p1}}$ é muito alto, significa que não há risco de colisão para o ponto p_1 . Nos experimentos usou-se $\lim_{t_{ccp1}} = 4$ segundos, pois acredita-se que este é um tempo seguro para uma pessoa tentar evitar alguma colisão com o objeto.

$$t_{tc_{p1}}(p_1, p_a) = \lim_{t_{ccp1}}, \quad \text{se } \{t_{tc_{p1}}(p_1, p_a) \cdot 0 \text{ ou } t_{tc_{p1}}(p_1, p_a) > \lim_{t_{ccp1}}\} \quad (6.38)$$

$$t_{tc_{p1}}(p_1, p_a) = t_{tc_{p1}}(p_1, p_a), \quad \text{se } \{t_{tc_{p1}}(p_1, p_a) > 0 \text{ ou } t_{tc_{p1}}(p_1, p_a) \cdot \lim_{t_{ccp1}}\} \quad (6.39)$$

O $t_{tc_{p1}}$ é para um único ponto, e pode apresentar ruídos e valores incorretos quando o objeto está se deslocando ao longo dos eixos x e/ou y , pois nas bordas do objeto os valores retornados pelo fluxo óptico podem não ser confiáveis. Para evitar esse possível problema, é definido uma estimativa final para p_1 , no qual TTC_j é igual ao maior valor entre os quatro calculados para $t_{tc_{p1}}$, ver Equação 6.40.

$$TTC_j = \max\{t_{tc_{p1}}(p_1, p_a)\} \quad (6.40)$$

onde j é o ponto correspondente para p_1 . Em termos práticos, isso significa que se algum dos pontos p_1 não indicar colisão, então o TTC_j final também não indicará colisão para o mesmo ponto.

Para determinar o TTC entre um obstáculo e a câmera, todos os pontos da região de interesse (ver Figura 6.4) devem ser considerados, então a média no TTC é calculada pela Equação 6.41, onde R é o conjunto de pontos pertencentes a região de interesse e n é o número de pontos em análise. Em termos práticos, o tempo estimado na equação abaixo representa o TTC do objeto no instante considerado.

$$TTC = \frac{\sum_{j \in R} TTC_j}{n} \quad (6.41)$$

5. Pós processamento

Nesta etapa, são aplicadas técnicas de processamento de imagens baseadas em uma heurística para remover possíveis erros das estimativas do TTC . Quando foi calculado o TTC_j para cada ponto, pequenos picos isolados podem ser ocasionados por problemas durante a estimativa do fluxo óptico, que por sua vez são causados principalmente por descontinuidades de objetos e ruídos provenientes do movimento e da variação de luminosidade.



Para compensar esses picos, são aplicadas operações morfológicas de erosão e dilatação visando eliminar pequenas regiões e “fechar buracos” na imagem. As regiões conectadas são segmentadas a partir da imagem e filtradas com base no seu tamanho, então uma nova imagem é formada e é aplicada uma operação de limiar para que regiões contendo menos de 10% do número total de pixels do *frame* sejam removidas. Cada região resultante representa um possível obstáculo.

6. Colisão detectada

Um limiar foi utilizado para decidir quais regiões resultantes da etapa anterior representam risco de colisão. Esse limiar deve ser definido de tal forma que permita ao usuário evitar a colisão após receber o aviso. Logo, o *TTC* é comparado com o limiar para detectar uma colisão. Nos experimentos foram realizados testes com limiares de 3 e 2,5 segundos. Quando o *TTC* for maior que esse limiar não há risco de imediato e novos *frames* são capturados, caso contrário segue para o passo seguinte.

7. Alarme de colisão

Quando o *TTC* de uma determinada região estiver abaixo do limiar indica que um obstáculo foi detectado, então um aviso sonoro é emitido indicando uma possível colisão. A frequência desse som é diretamente proporcional ao *TTC*, dando uma indicação se o risco vai aumentando ou diminuindo. Dessa forma, uma ação de desvio pode ser tomada pelo usuário evitando uma possível colisão com um obstáculo.

6.3. Considerações finais do capítulo

Neste Capítulo foram apresentados detalhes do algoritmo desenvolvido. No qual, a partir de um modelo ideal de uma câmera pinhole, obteve-se uma equação fechada para estimar o *TTC* e conseqüentemente detectar colisões entre o usuário e um obstáculo. Para detectar movimentos foi aplicado um algoritmo de fluxo óptico denso de Farnerbäck, que analisa todos os pixels do *frame*, com isso toda a região da imagem foi considerada (exceto a região de *offset*, conforme Figura 6.4). Nas etapas de pré-processamento e pós-processamento, técnicas de processamento de imagens foram aplicadas para minimizar possíveis erros provenientes da etapa de aquisição das imagens. No próximo Capítulo, detalhes dos experimentos e seus principais resultados serão apresentados e discutidos.

CAPÍTULO 7 – EXPERIMENTOS E RESULTADOS

7.1 Introdução

O algoritmo foi desenvolvido em linguagem C/C++, foi utilizada a biblioteca de visão computacional OpenCV, implementado no ambiente Qt Creator e compilado no Visual Studio. Os vídeos foram adquiridos por um *smartphone* com câmera de 1280x720 pixels de resolução e uma taxa de 30 ou 25 frames por segundo, conforme os experimentos. Foi processado em um computador com processador Intel® Core™ i7-2630QM 2.0 GHz e 4GB de RAM.

Três experimentos foram realizados para testar e validar o algoritmo. Antes de explicar cada um deles, cabe salientar que ao verificar as Figuras referentes aos vídeos utilizados em cada experimento, percebe-se que foram gravados em ambientes diversos, com diferentes tamanhos de objetos (obstáculos) com formas variadas e em condições sem controle do ambiente (salvo o primeiro experimento). Optou-se por realizar dessa forma, para obter uma aproximação das condições reais de um usuário durante sua locomoção. No entanto, sabe-se que testes com usuários deficientes visuais deverão ser realizados para avaliar o método e fazer eventuais ajustes.

7.2 Primeiro Experimento

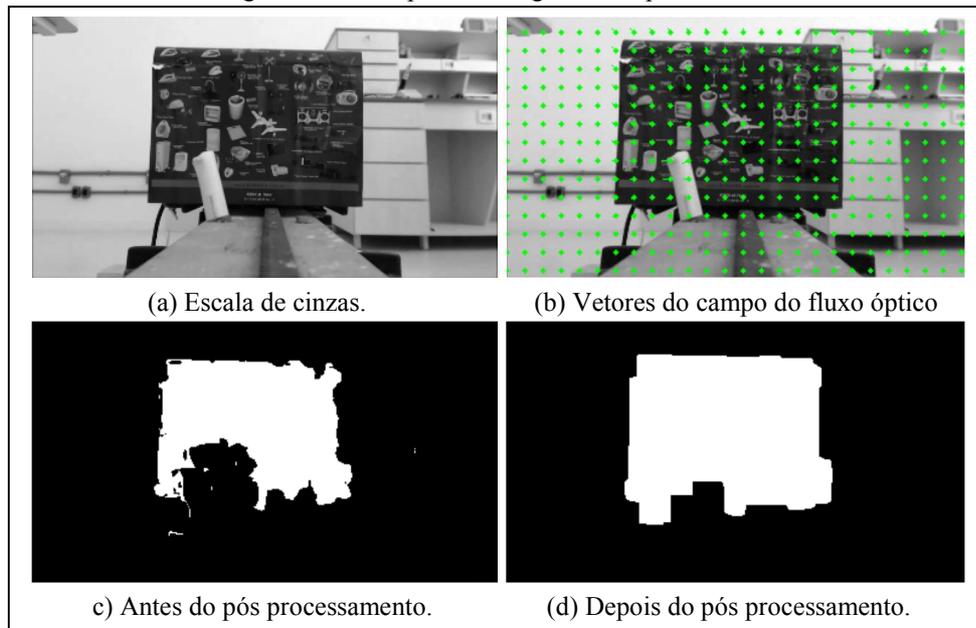
Este experimento foi conduzido em uma situação controlada, em um ambiente interno, no qual a variação de luminosidade e movimentos laterais (em x e y) foram evitados. O objetivo consiste em avaliar o desempenho do algoritmo em um cenário “ideal”, onde um objeto se desloca em linha reta na direção da câmera. O experimento resume-se em um robô móvel portando um objeto e percorrendo um trilho com 180 cm de comprimento e velocidade constante de 0,3 metros por segundo; a câmera foi fixada ao final do trilho de modo a capturar o vídeo do objeto se aproximando. Como são conhecidas a velocidade e a posição inicial e final do objeto, obteve-se o *TTC* verdadeiro (TTC_{real}) para cada um dos *frames*. Esse TTC_{real} é importante pois serve como *ground truth* para avaliar o experimento. Ao todo foram gravados 5 vídeos a uma taxa de 30 *fps*.

A Figura 7.1 mostra um exemplo em que o objeto parte em direção à câmera. A primeira Figura apresenta a imagem original em escala de cinzas (a); seguida pelos vetores de



campo do fluxo óptico (b); a terceira traz a imagem resultante antes do pós processamento (c) e; a quarta o resultado após o pós processamento (d). Comparando imagens (c) e (d) percebe-se nitidamente o resultado da aplicação das operações morfológicas de erosão e dilatação, conforme comentado no capítulo anterior.

Figura 7.1: Exemplos de imagens do experimento 1.



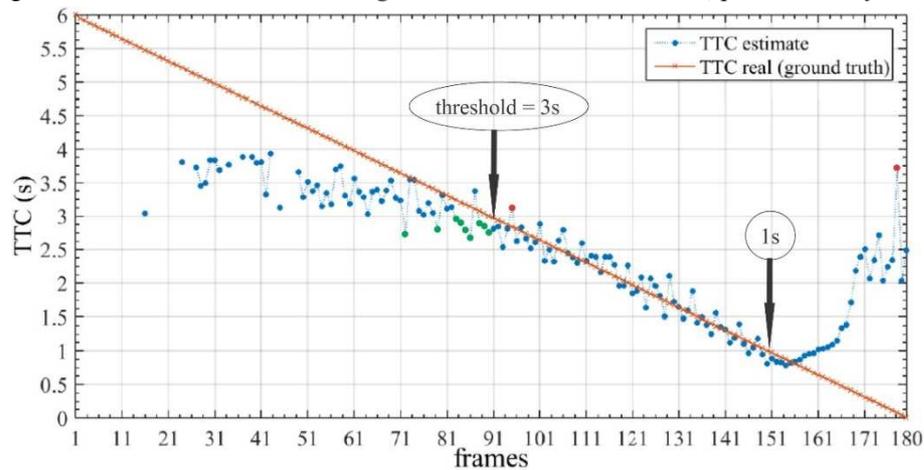
Fonte: autor.

Uma observação importante a ser feita em relação a Figura 7.1b: nota-se que os vetores de fluxo óptico não estão sendo representados para todos os pontos do *frame*, no entanto, optou-se por ilustrar dessa forma apenas para visualização, pois os vetores são para todos os pontos.

Para demonstração gráfica dos resultados apresentados neste experimento foi utilizado apenas um dos vídeos, porém para os cálculos foram utilizados todos os cinco. Dessa maneira, o resultado de um único vídeo é mostrado na Figura 7.2, onde o *TTC* estimado é comparado com o *TTC_{real}* (*ground truth*) para todos os *frames* do vídeo. Pode ser visto que para tempos abaixo de 1 segundo a estimativa do *TTC* começa a apresentar um desempenho insatisfatório quando comparado com o *ground truth*. Pois quando um objeto está muito próximo da câmera, o fluxo óptico pode não ser muito preciso, levando a estimativas precárias de *TTC*. Por esse motivo, concluiu-se que o range de operação do método pode ser confiável para tempos acima de 1 segundo. Além disso, foi definido um *threshold* de 3 segundos, esse valor representa o limite superior no qual um alarme é emitido sempre que o tempo estimado for menor.



Figura 7.2: Gráfico *TTC* estimado e *ground truth* de um dos vídeos, para todos os *frames*.



Fonte: autor.

Todos os pontos (azuis, verdes e vermelhos), referentes ao *TTC* estimado que aparecem no gráfico da Figura 7.2, correspondem a tempos calculados abaixo \lim_{tccp1} para cada *frame*. Percebe-se que em alguns *frames* não há nenhum ponto. Isso é nítido para tempos maiores, ocorre pois o tempo estimado é maior que \lim_{tccp1} , como no *frame* 11, por exemplo.

Ainda em relação ao *TTC* estimado, os pontos vermelhos representam colisões que deveriam ter sido detectadas e não foram, pois estão acima do *threshold*; os verdes são colisões detectadas que não deveriam ter sido, pois o tempo estimado é menor do que os 3 segundos e na verdade deveriam ser tempos maiores, de acordo com o TTC_{real} , os azuis são os demais tempos estimados. Essas informações são importantes, pois serão utilizadas para definir a performance do algoritmo para detecção de colisões, que é dada em termos de precisão, revocação e acurácia, bem como o erro médio entre o tempo calculado e o real.

Além disso, pelo gráfico da Figura 7.2, sabe-se que o limite para detecção de colisões foi definido em 3 segundos. A partir daí algumas taxas aplicadas na métrica de avaliação da performance são definidas. São elas: qualquer colisão detectada à esquerda do *threshold* com *TTC* menor que 3 segundos é um falso positivo (*FP*) – pontos verdes; as estimativas à direita que estão acima de 3 segundos são falsos negativos (*FN*) – pontos vermelhos. Os demais pontos à esquerda e a direita do *threshold*, são verdadeiros negativos (*TN*) e verdadeiros positivos (*TP*), respectivamente.

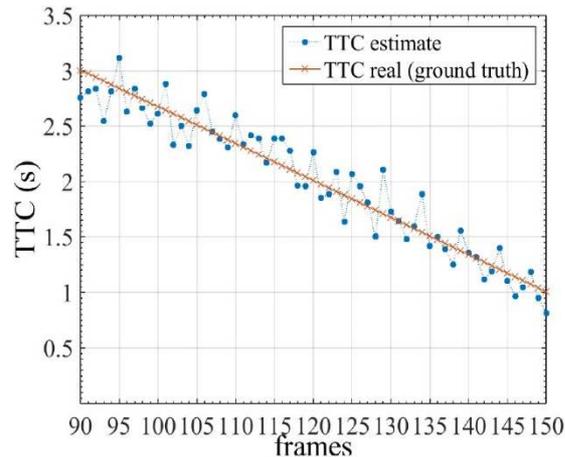
Para esse experimento, o resultado médio obtido em todos os 5 vídeos gravados foi: $TP = 84$, $FN = 6$, $TN = 81$ e $FP = 9$. Essas taxas são utilizadas para definir a performance do algoritmo. No entanto, para definir o erro entre o TTC_{real} e o estimado, definiu-se o intervalo de tempo entre 1 e 3 segundos, uma vez que a maioria das estimativas de *TTC* fora desse intervalo



não são muito confiáveis quando comparadas com o *ground truth*, além disso, apenas os tempos estimados iguais ou menores que 3 segundos são considerados como colisão. Para avaliar o método em relação a esse erro, alguns resultados obtidos são apresentados, conforme segue.

O gráfico da Figura 7.3, ilustra o comportamento dos tempos para o intervalo selecionado para um único vídeo tomado como amostra deste experimento.

Figura 7.3: Erro médio para um dos vídeos do primeiro experimento (em segundos).



Fonte: O autor.

Observando o gráfico da Figura 7.3, nota-se que os valores de TTC_{real} acima do *threshold* de 3 segundos e abaixo de 1 segundo não são considerados. A partir desse intervalo, obteve-se um erro médio entre o TTC estimado e o TTC_{real} para todos os *frames* desse intervalo. Esse erro é dado como uma média dos erros instantâneos obtidos em cada *frame*.

A Tabela 7.1 exibe os erros mínimo ($erro_{min}$), máximo ($erro_{max}$) e médio ($erro_{med}$) para cada um dos vídeos e o valor do $erro_{médio}$ para todos os vídeos. Observa-se que o $erro_{médio}$ foi de aproximadamente 0,18 segundos, o que pode ser considerado uma boa aproximação entre o TTC estimando e o TTC_{real} .

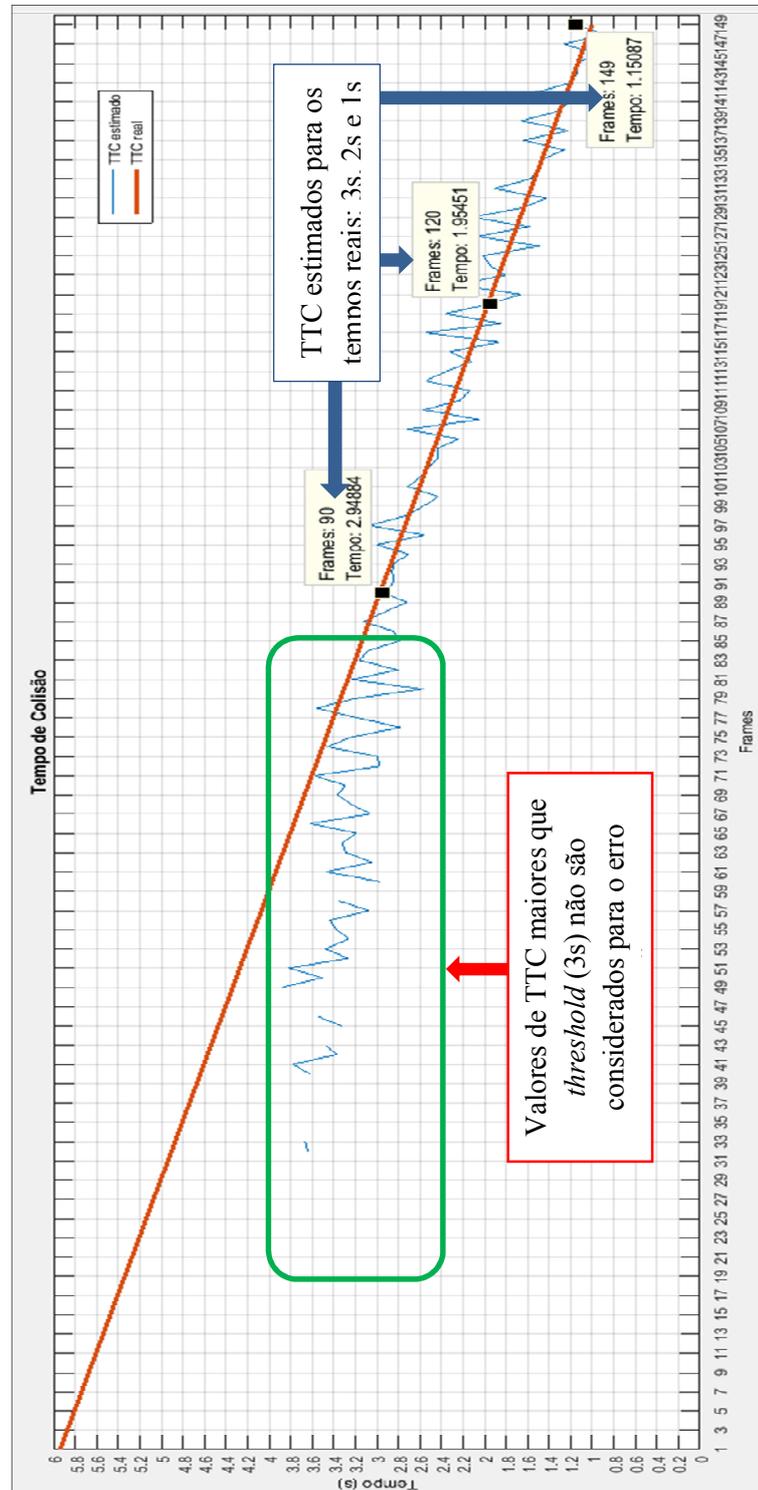
Tabela 7.1: Erros mínimo, máximo e médio do experimento.

Vídeo	$erro_{min}$	$erro_{max}$	$erro_{med}$
1	0,001023	0,470005	0,160474
2	0,005742	0,403066	0,144991
3	0,010294	0,392261	0,188112
4	0,001946	0,429479	0,173847
5	0,011539	0,490372	0,230721
$erro_{médio}$			0,179629



Apenas para comparar o TTC estimado com o TTC_{real} , foram destacados esses tempos para os instantes de 1, 2 e 3 segundos, conforme o gráfico da Figura 7.4. Destaca-se que esse gráfico não é relevante para as métricas de avaliação, serve apenas como ilustração.

Figura 7.4: Gráfico do TTC real x TTC estimado para os instantes de 1,2 e 3 segundos.

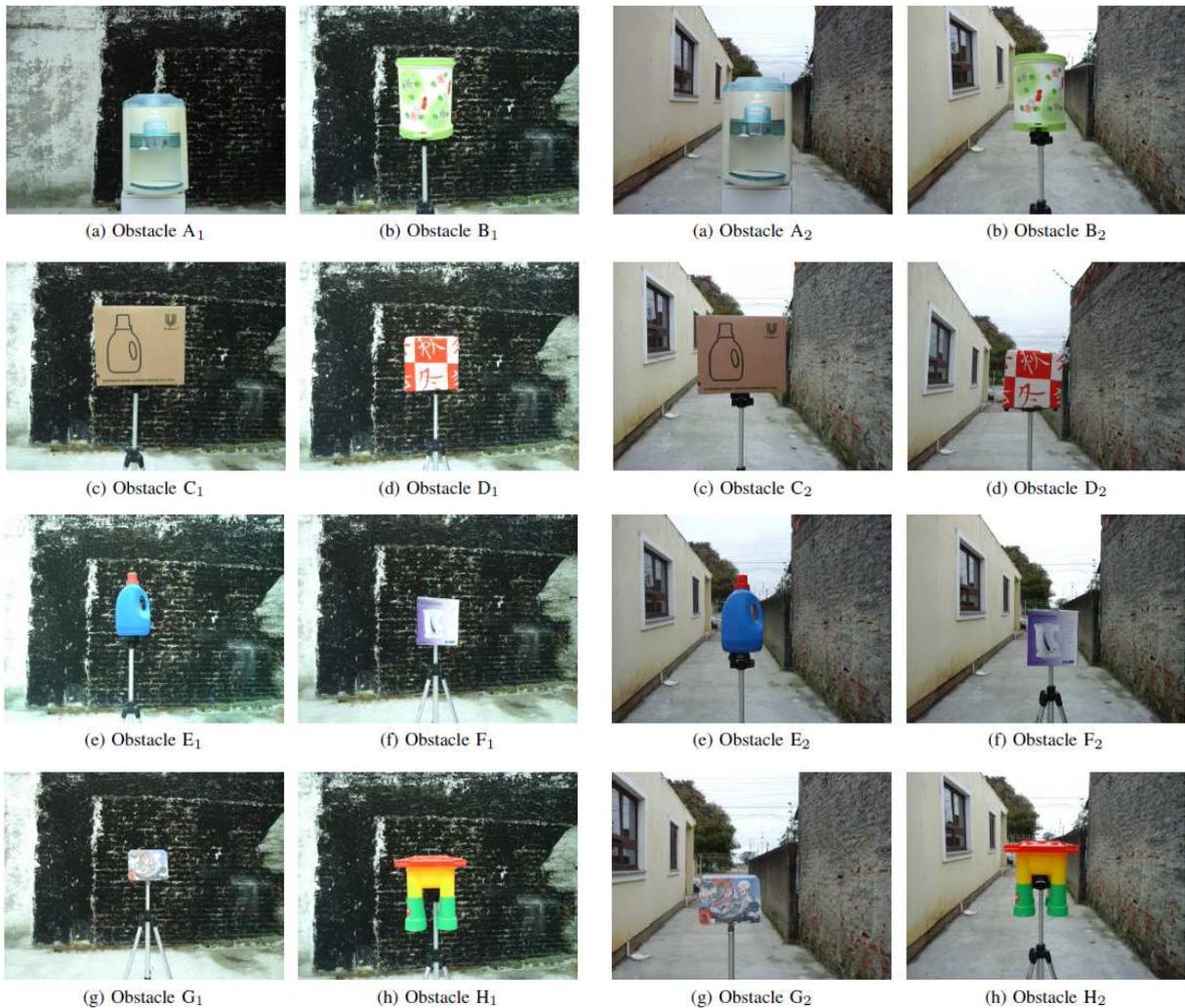


Fonte: o autor.

7.3. Segundo Experimento

Neste experimento, o método foi avaliado em ambientes externos e não controlados que continham alguns obstáculos estáticos, no qual um usuário, com um *smartphone* preso ao peito, se desloca em direção a um obstáculo com velocidade aproximada de 0,4 m/s. Um total de 80 vídeos de aproximadamente 7,6 segundos foram gravados com objetos diferentes, sendo 5 vídeos para cada um dos 16 obstáculos da Figura 7.5, a taxa de amostragem foi de 25 *fps*. Os obstáculos foram divididos em dois grupos GRUPO 1 = {A₁, B₁, C₁, D₁, E₁, F₁, G₁, H₁} e GRUPO 2 = {A₂, B₂, C₂, D₂, E₂, F₂, G₂, H₂}. No primeiro há uma parede irregular ao fundo do obstáculo, no segundo, o obstáculo foi colocado em um corredor de estacionamento.

Figura 7.5: Exemplos de imagens do experimento 2.



Fonte: Autor.

Para esse experimento foi utilizado um *threshold* de 2,5 segundos e a análise para determinar as taxas de *FP*, *FN*, *TN* e *TP*, e o erro médio, foram as mesmas aplicadas no



experimento 1. Dessa forma, as Tabelas 7.2 e 7.3 foram obtidas para cada um dos grupos. Os valores apresentados correspondem às médias entre os vídeos para cada um dos grupos.

Tabela 7.2: Resultados para o Grupo 1.

Obstáculo	<i>FN</i>	<i>TP</i>	<i>FP</i>	<i>TN</i>	<i>erro</i>_{médio} (s)
A ₁	9	55	5	121	0,44
B ₁	13	51	0	126	0,69
C ₁	7	57	1	125	0,39
D ₁	4	60	0	126	0,19
E ₁	20	44	0	126	0,81
F ₁	8	56	7	119	0,35
G ₁	10	54	1	125	0,56
H ₁	7	57	2	124	0,39

Pela Tabela 7.2, observa-se que os piores e melhores resultados são apresentados para os obstáculos dos grupos (D) e (E), respectivamente.

Tabela 7.3: Resultados para o Grupo 2.

Obstáculo	<i>FN</i>	<i>TP</i>	<i>FP</i>	<i>TN</i>	<i>erro</i>_{médio} (s)
A ₂	9	55	4	122	0,40
B ₂	8	56	1	125	0,25
C ₂	6	58	1	125	0,24
D ₂	4	60	0	126	0,19
E ₂	17	47	1	125	0,66
F ₂	6	58	4	122	0,27
G ₂	7	57	2	124	0,31
H ₂	11	53	3	123	0,51

A Tabela 7.3 também apresenta como melhor e pior resultado os obstáculos dos grupos (D) e (E). Comparando as Tabelas 7.2 e 7.3, destaca-se que, independente do grupo, os obstáculos que apresentaram melhor e pior resultados são o (D) e o (E) para cada conjunto de vídeos correspondente a cada tipo de obstáculo. Também, observa-se que esses resultados não são apenas em relação ao *erro*_{médio}, ao mesmo tempo pode-se comparar as taxas entre os melhores e piores casos de cada grupo, conforme segue: D₁ - *FN*=4, *TP*=60, *FP*=0, *TN*=126 e D₂ - *FN*=4, *TP*=60, *FP*=0, *TN*=126; E₁ - *FN*=20, *TP*=44, *FP*=0, *TN*=126 e E₂ - *FN*=17, *TP*=47, *FP*=1, *TN*=125. Comparando as colisões que deveriam ter sido detectadas, mas não foram (*FN*) ou as que foram detectadas de forma correta (*TP*), destaca-se que os obstáculos D e E continuam apresentando o melhor e o pior resultado. Quanto as taxas de *FP*, que são colisões detectadas

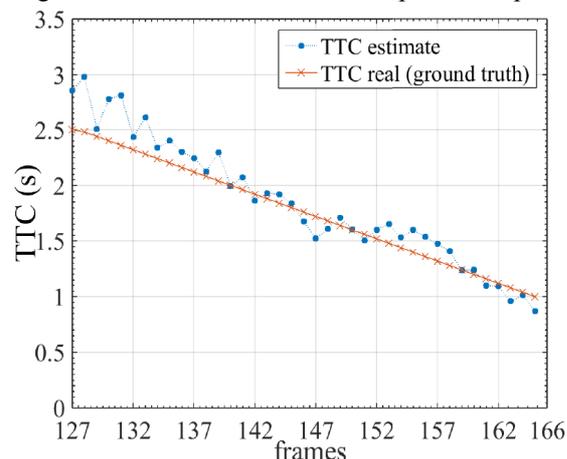


erroneamente, não têm tanta importância quanto aos FN , uma vez que para um usuário é mais importante detectar colisões com falsos positivos do que não as detectar com falsos negativos.

Outro ponto a ser observado, se dá pelo fato de que independente do cenário no qual se encontra cada obstáculo, os resultados dispostos nas tabelas são muito próximos, em outras palavras, os resultados de cada obstáculo são semelhantes, independente do Grupo. Isso leva a concluir, que para esse segundo experimento, o ambiente não influenciou significativamente nos resultados do método.

A seguir são apresentados os resultados dos melhores e piores casos para cada grupo. Em termos de erro médio do TTC em relação ao TTC_{real} , a Figura 7.6 mostra um gráfico correspondente a um vídeo do melhor caso para o Grupo 1. Foi obtido com um dos vídeos do obstáculo D_1 , onde $FN = 7$, $TP = 57$, $FP = 0$, $TN = 126$, e o $erro_{médio}$ de 0.13 segundos.

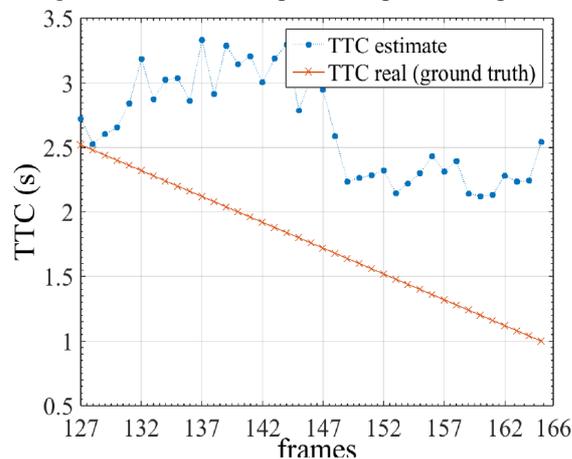
Figura 7.6: Gráfico do melhor caso para o Grupo 1.



Fonte: Autor.

O gráfico da Figura 7.7 corresponde a um vídeo do pior caso dado por um obstáculo E_1 do Grupo 1, com $FN = 25$, $TP = 39$, $FP = 1$, $TN = 125$ e o $erro_{médio}$ de 0.86 segundos.

Figura 7.7: Gráfico do pior caso para o Grupo 1.

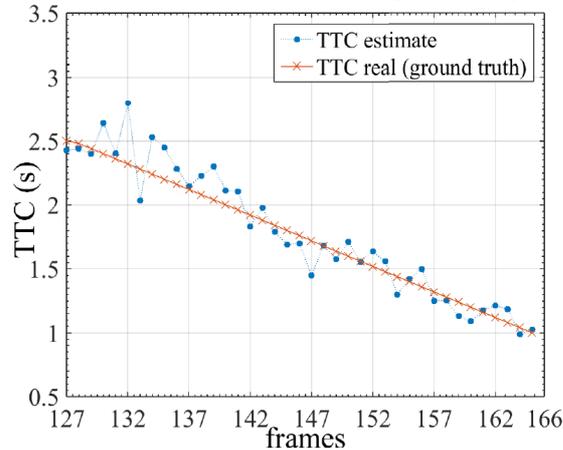


Fonte: Autor.



A Figura 7.8 apresenta o gráfico correspondente a um vídeo do melhor caso do Grupo 2. Foi obtido em um dos vídeos do obstáculo D_2 , onde $FN = 0$, $TP = 64$, $FP = 0$, $TN = 126$, e o *erro médio* de 0.15 segundos.

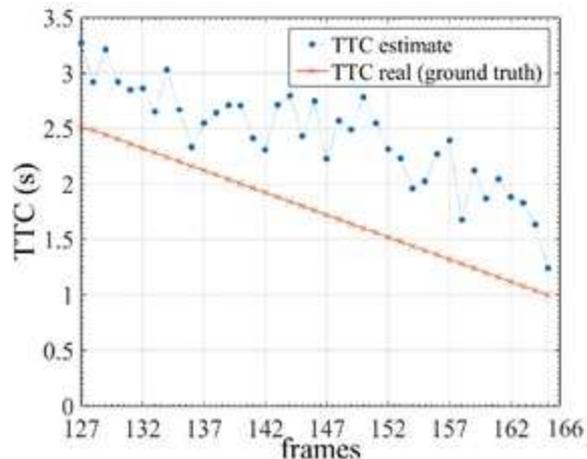
Figura 7.8: Gráfico do melhor caso para o Grupo 2.



Fonte: Autor.

Na Figura 7.9 observa-se o gráfico correspondente ao pior caso do Grupo 2. Foi em relação ao obstáculo E_2 , com $FN = 23$, $TP = 41$, $FP = 4$, $TN = 122$ e o *erro médio* de 0.66 segundos.

Figura 7.9: Gráfico do pior caso para o Grupo 2.



Fonte: Autor.

Observa-se que para ambos os grupos, os obstáculos com piores e melhores resultados são D e E, respectivamente. Considerando os resultados apresentados nas Tabelas 7.2 e 7.3, a Tabela 7.4 traz os resultados em termos de FN , TP , FP e TN , e a média do *erro médio* para todos os vídeos dos dois grupos, ou seja, para este experimento.

Tabela 7.4: Resultados para todos os vídeos.

Obstáculo	FN	TP	FP	TN	<i>erro médio</i> (s)
Todos	146	878	32	1984	0,41



7.4. Terceiro experimento

Neste experimento foram gravados um total de 20 vídeos com 450 *frames* cada a uma taxa de 25 *fps*, sendo 5 vídeos para cada um dos 4 ambientes diferentes ilustrados na Figura 7.10. Um usuário com um *smartphone* preso ao peito caminhou livremente por um percurso sem obstáculos em sua trajetória durante 18 segundos. A velocidade com que ele se deslocou e a distância que percorreu são desconhecidas e podem variar para cada vídeo. Além disso, nenhum obstáculo foi colocado no caminho do usuário.

A Figura 7.10 mostra o experimento em um ambiente interno (Vídeo A), ambiente externo (Vídeo B), estacionamento para carro em ambiente externo (Vídeo C) e uma parede com fundo irregular em um ambiente externo (Vídeo D).

Figura 7.10: Exemplos de imagens do experimento 3.



Fonte: Autor.

Como o objetivo do experimento foi o de avaliar a robustez do método em relação à mobiliade do usuário em um ambiente livre de obstáculos, o ideal é que nenhum obstáculo seja detectado, ou seja, nenhum valor de *TTC* igual ou abaixo do *threshold* deveria ser retornado. Assim, não existem verdadeiros positivos (que indicam corretamente colisão) e falsos negativos (que deveriam indicar colisão e não indicam). Desse modo, apenas as taxas de *FP* e *TN* para cada ambiente puderam ser obtidas, uma vez que taxas *FP* correspondem às colisões que foram detectadas mas não deveriam ter sido, já *TN* ocorre quando não houve detecção de colisões, como é o esperado para esse experimento. A Tabela 7.5 apresenta essas taxas. Vale destacar que os valores desta tabela são uma média para cada um dos 5 vídeos de cada ambiente.



Pela tabela abaixo, pode-se observar que nos *frames* em que a colisão foi detectada, as taxas de colisões detectadas erroneamente (*FP*) foram baixas quando comparadas com o total de *frames*, visto que o pior caso foi para o Vídeo B, em que apenas 8,67% das estimativas de não haver colisão foram erradas. Isso ocorre, principalmente, por se tratar de um ambiente externo e com grande possibilidade de interferência, por exemplo, movimento de pessoas, carros, iluminação, levando à estimativas de fluxo óptico que podem não ser tão confiáveis e, conseqüentemente conduzir a uma falsa colisão, mesmo assim, considera-se um bom resultado. Por outro lado, no Vídeo C o método obteve 0% de colisões.

Tabela 7.5: Resultados para o experimento 3.

Ambiente	<i>FP</i>	<i>TN</i>
Vídeo A	22	428
Vídeo B	39	411
Vídeo C	0	450
Vídeo D	6	444

7.5. Performance e validação

Para avaliar a performance do algoritmo em relação à detecção de obstáculos, as métricas de validação foram dadas em termos de precisão (*precision*), revocação (*recall*) e acurácia (*accuracy*). Enquanto a validação do TTC, foi dada pelo erro médio absoluto entre o TTC estimado e o *ground truth*, conforme mencionado no Capítulo anterior.

A precisão é a taxa de colisões detectadas corretamente (*TP*) em relação a todas as colisões detectadas (*TP + FP*), ou seja, ela representa a probabilidade de uma predição positiva estar correta entre todas as predições classificadas como positivas, é dada pela Equação (7.1).

$$precision = \frac{TP}{TP + FP} \quad (7.1)$$

O termo revocação está relacionado com a taxa de colisões detectadas corretamente (*TP*) entre todas as que deveriam ter sido detectadas (*TP + FN*), ou seja, representa a probabilidade de uma predição ser positiva entre todas as predições que deveriam ter sido positivas. Também pode ser entendida como uma indicação do quanto do total de predições positivas foram (*TP*) retornadas corretamente. A Equação (7.2) retorna o valor da revocação.

$$recall = \frac{TP}{TP + FN} \quad (7.2)$$

Quanto a acurácia, corresponde à proporção de todas as predições corretas ($TP + TN$) em relação a todo o conjunto em análise ($TP + TN + FP + FN$), conforme a Equação (7.3). Ela apresenta o comportamento do algoritmo de modo geral, pois traz a taxa de acertos sobre o todo.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7.3)$$

A validação do TTC é dada pela Equação (7.4), que retorna o erro médio absoluto (Err_{mean}) entre o TTC estimado e o real. Onde Tf é o número total de frames analisados, N_V é o número de vídeos de cada experimento (se for analisar apenas um vídeo $N_V = 1$), th é o *threshold*, TTC_{real} é o *ground truth* e TTC_{est} é o tempo estimado.

$$err_{mean} = \frac{\sum_{n=1}^{N_V} \sum_{t=1}^{th} |TTC_{real}(t) \cdot TTC_{est}(t)|}{Tf} \quad (7.4)$$

A performance algoritmo e um resumo de seus resultados são expressos na Tabela 7.6.

Tabela 7.6: Performance e validação do algoritmo.

Experimento		Número de vídeos	Precisão <i>precision</i> (%)	Revocação <i>recall</i> (%)	Acurácia <i>Accuracy</i> (%)	Erro médio <i>err_{mean}</i> (s)
Ambiente						
1	Exp. 1	5	90,32	93,33	91,67	0,18
	Média		90,32	93,33	91,67	0,18
2	Grupo 1	40	96,44	84,76	93,81	0,47
	Grupo 2	40	96,52	86,72	94,47	0,35
	Média		96,48	85,74	94,14	0,41
3	Vídeo A	5	NA	NA	95,11	NA
	Vídeo B	5	NA	NA	91,33	NA
	Vídeo C	5	NA	NA	98,66	NA
	Vídeo D	5	NA	NA	100,00	NA
	Média		NA	NA	96,27	NA
Média geral			93,40	89,54	94,03	0,29



De acordo com a Tabela 7.6, algumas observações são realizadas:

- i) A sigla NA significa Não se Aplica;
- ii) Em relação à precisão, os dois primeiros experimentos apresentaram resultados superiores a 90%, o que indica que o método desenvolvido possui uma alta probabilidade em detectar colisões corretamente. Para os dois grupos do experimento 2, a precisão foi aproximadamente a mesma, sinalizando que os experimentos praticamente independem do meio no qual foram realizados. Na prática, quer dizer que 9,68% e 3,52% das colisões são falsas para os experimentos 1 e 2, respectivamente; isso indica falsos alarmes de colisão. No terceiro experimento, não foi possível obter a precisão e a revocação, como já comentado anteriormente.
- iii) Quanto à revocação, para o experimento 1 foi de 93,33% e para o 2 de 85,74%, há uma pequena diferença entre os dois. No entanto, a revocação corresponde a taxa de colisões detectadas entre todas que deveriam ter sido, dessa forma, o experimento 2 obteve um desempenho pior em relação ao primeiro. Na prática isso indica que 6,67% e 14,26% das colisões que deveriam ter sido detectadas não foram, para os experimentos 1 e 2, respectivamente; logo, a probabilidade de colisão detectada corretamente é maior para o experimento 1.
- iv) A acurácia representa o total de acertos de um todo, ou seja, são todos os TP e TN em relação a todos os *frames*. Para os três experimentos a acurácia foi superior a 90%. Em termos práticos, significa que mais de 90% das colisões e não colisões foram indicadas corretamente.
- v) Para avaliar o TTC, o método obteve um erro médio de 0,18 e 0,41 segundos para os experimentos 1 e 2, respectivamente, para o experimento 3 não se aplica. Isso corresponde ao erro em segundos entre o TTC estimado e o real. Apesar dessa diferença entre esses tempos, isso não significa que o erro para detecção de colisões foi esse, e sim que são média entre os valores instantâneos. Analisando as Figuras 7.2, 7.3, 7.4, 7.6, 7.7, 7.8 e 7.9, essa diferença entre os tempos instantâneos é facilmente percebida.

7.6. Considerações finais do capítulo

Neste Capítulo, foram apresentados detalhes dos experimentos realizados, bem como seus principais resultados. Foram realizados três experimentos em diferentes condições e cenários, possibilitando avaliar o desempenho do método verificando a sua performance para detecção de colisões e a validação do *TTC*. A performance foi dada em termos de precisão, revocação e acurácia. O erro médio possibilitou validar o *TTC*, no qual obteve-se resultados

?

considerados promissores. Além disso, a análise foi qualitativa, uma vez que os resultados foram obtidos a partir de uma pequena amostra de vídeos para cada cenário em cada experimento.

Embora os resultados tenham se mostrado bons de modo geral, eles também indicam algumas limitações do método. Por exemplo, o método não obteve bons resultados para tempos de TTC inferiores a 1 segundo. Pelo experimento 3, percebe-se que os piores resultados foram obtidos em um ambiente externo, mais especificamente no Vídeo B, onde apresentou um percentual de 91,33% de acurácia, que apesar de alto, é inferior em relação aos demais ambientes. No entanto, essas limitações poderão ser reavaliadas quando novos experimentos forem realizados com usuários deficientes visuais, pois novos ajustes no algoritmo provavelmente deverão ser realizados.

?

CAPÍTULO 8 – CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Os deficientes, em suas diferentes categorias e graus, cada vez mais necessitam de atenção, sensibilidade e ajuda da população e do governo. Entre eles estão os deficientes visuais, que no Brasil atinge 35,8 milhões de pessoas (IBGE, 2012). São inúmeros problemas e obstáculos que essa importante parcela da população enfrenta, entre eles, durante a locomoção em seu dia a dia, como postes, telefones públicos, placas de sinalização, caçambas de entulho, buracos, escadas, mesas, pedestres, veículos, entre outros, que podem estar em seu caminho podendo colocar em risco sua integridade devido a algum tipo de colisão.

Para tentar facilitar seu cotidiano, surgiram as TA, com o intuito de contribuírem para proporcionar ou ampliar habilidades funcionais de pessoas com deficiência, além de proporcionar maior independência, qualidade de vida e inclusão social, através da ampliação de sua comunicação, mobilidade, controle de seu ambiente, habilidades de seu aprendizado, trabalho e integração com a família, amigos e sociedade (BERSCH, 2013). Dessa forma, cada vez mais recursos de TA têm sido desenvolvidos e as pesquisas em nível acadêmico tem aumentado.

Nesse contexto, essa dissertação apresenta um método para detecção de colisões que pode ser utilizado para minimizar os riscos de colisão com obstáculos e que, quando implementado no sistema TA, visa proporcionar maior autonomia, independência, confiança, possibilitando sua locomoção com maior segurança aos deficientes visuais.

Para realizar essa tarefa, desenvolveu-se uma equação a partir do modelo ideal de uma câmera pinhole, que possibilitou estimar o Tempo de Colisão entre uma câmera monocular e os obstáculos e conseqüentemente detectar colisões a partir de um *threshold*. Detalhes do desenvolvimento e da implementação do algoritmo foram apresentados, bem com seus experimentos e principais resultados.

O algoritmo foi desenvolvido em linguagem C/C++, com técnicas de visão computacional e a biblioteca OpenCV. Três experimentos em condições e ambientes diferentes foram conduzidos a fim de avaliar a performance do método e validar o TTC. Os vídeos para cada experimento foram adquiridos por meio da câmera de um *smartphone* e o algoritmo foi implementado em um computador, conforme já mencionado. Dessa forma, o método utiliza



fluxo óptico denso, equações a partir do modelo ideal de uma câmera pinhole e, técnicas de processamento de imagem para detectar movimentos, estimar *TTC* em cada ponto e filtrar possíveis erros decorrentes do processo de aquisição das imagens.

De modo geral, a performance do método foi satisfatória, atingindo uma precisão de 93,40%, revocação de 89,54%, acurácia de 94,03% e erro médio absoluto de 0,29 segundos. Estes resultados podem ser considerados positivos, uma vez que indicam que o método pode detectar colisões com certa confiança, entre os limites estabelecidos. Também indicam que usuários com deficiência visual podem se beneficiar ao usar o método como uma tecnologia assistiva para evitar colisões, o que tem sido a aplicação alvo dessa dissertação.

Apesar de bons resultados, o método possui algumas limitações. Nas estimativas de *TTC*, se mostrou ruim para tempos inferiores a 1 segundo. No experimento 3, no Vídeo B, os resultados foram piores quando comparados com os demais ambientes, indicando que os ruídos provenientes de ambientes externos também podem interferir nos resultados do método.

Entre as principais contribuições citam-se: desenvolvimento de um algoritmo para detectar obstáculos e estimar o *TTC* por meio uma única câmera; possibilita a navegação em ambientes desconhecidos sem a necessidade de ter um mapa armazenado; fonte de pesquisa para futuras melhorias e aplicações; possibilidade de ser implementado em um aplicativo de TA para auxiliar deficientes visuais (trabalhos futuros).

Como trabalhos futuros e propostas de melhorias, destacam-se: realizar novos experimentos em diferentes condições e situações; desenvolver outra métrica para determinar o grau de risco de colisão de um obstáculo (utilizou-se apenas o *TTC* para determinar o risco), como dividir a imagem em quadrantes, por exemplo; implementar o algoritmo em um *smartphone* e testar com diferentes usuários videntes; testar o algoritmo usuários deficientes visuais e videntes em ambientes pré-estabelecidos e cenários do mundo real; reavaliar o algoritmo com os novos testes e experimentos; implementar o algoritmo no sistema TA. Além disso, futuramente pode-se testar com outros tipos de câmeras, com diferentes lentes e distâncias focais, etc; bem como implementar em um hardware embarcado ou desenvolver um hardware dedicado.

O método desenvolvido também serve como referência para outras aplicações, como a robótica móvel, auxílio para pessoas videntes, indústria, setor automotivo, entre outras aplicações que necessitem determinar o *TTC* e/ou detectar obstáculos.

REFERÊNCIAS

AIRES, K. R. T. *Segmentação de Planos em Homografia Afim, Fluxo Óptico e Reconstrução Métrica*. Tese de Doutorado em Ciências – Área: Engenharia de Computação, Universidade Federal do Rio Grande do Norte. Natal, 2009.

AIRES, G. de S. *Estimação de Movimento de uma Câmera via Geometria Epipolar*. Dissertação de Mestrado em Engenharia Elétrica, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2010.

ALENYÀ, G.; NÈGRE, A.; CROWLEY, J. L. *A comparison of three methods for measure of time to contact*. In: Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on. IEEE, p. 4565-4570, 2009.

ÁVILA, M.; DE FARIA, M. A. R.; MELLO, P. A. de A. *As Condições de Saúde Ocular no Brasil - 2015*. Conselho Brasileiro de Oftalmologia - CBO, 1 ed. São Paulo, 2015. Disponível em: <http://www.cbo.net.br/novo/publicacoes/Condicoes_saude_ocular_IV.pdf>. Acesso em: 04 mar 2016.

BARATOFF, G.; TOEPFER, C.; WENDE, M.; NEUMANN, H. *Real-time navigation and obstacle avoidance from optical flow on a space-variant map*. In: Intelligent Control (ISIC), 1998. Held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA), Intelligent Systems and Semiotics (ISAS), Proceedings. IEEE, p. 289-294, 1998.

BERSCH, R. *Introdução a Tecnologia Assistiva*. Porto Alegre, 2013.

BERSCH, R; TONOLLI, J. C. *Introdução ao Conceito de Tecnologia Assistiva e Modelos de Abordagem da Deficiência*. Jun. 2006. Disponível em: <<http://www.bengalalegal.com/tecnologia-assistiva>>. Acesso em: 25 abr 2015.

BRASIL. *Decreto nº 3.298, de 20 de dezembro de 1999. Regulamenta a Lei nº 7.853, de 24 de outubro de 1989, dispõe sobre a Política Nacional para a Integração da Pessoa Portadora de*



Deficiência, consolida as normas de proteção, e dá outras providências. Diário Oficial da União, 1999.

BRASIL. *Lei Nº 13.146, de 6 de julho de 2015. Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência).* Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/113146.htm>. Acesso em: 20 ago 2015.

BUENO, G. A. *Orientação e mobilidade na habilitação de deficientes visuais.* Revista da Faculdade de Educação da Universidade de São Paulo, n. 18, p. 205-215, 1992.

CALDEIRA, E. M. de O. *Navegação Reativa de Robôs Móveis com Base no Fluxo Óptico.* Tese de Doutorado em Engenharia Elétrica, Universidade Federal do Espírito Santo. Vitória, 2002.

CAMUS, T. *Calculating time-to-contact using real-time quantized optical flow.* National Institute of Standards and Technology - NISTIR, 1995.

CHA, J. S.; LIM, D. K.; SHIN, Y. N. *Design and Implementation of a Smart Phone Navigation Application for Blind Persons.* The 7th International Conference on Information Security and Assurance, v. 21, p. 149 - 151, 2013

DA SILVA, R. F. L.; LOPEZ, A. L.; RAMIREZ, A. R. G. *Locomoção independente: contribuição à locomoção do cego em espaços urbanos.* Anais do Computer on the Beach, p. 221-230, 2012.

DO CENSO, Cartilha 2010. *Pessoas com Deficiência.* Secretaria de Direitos Humanos da Presidência da República (SDH/PR) / Secretaria Nacional de Promoção dos Direitos da Pessoa com Deficiência (SNPD) / Coordenação-Geral do Sistema de Informações sobre a Pessoa com Deficiência. Brasília, 2012.

ERENO, D. *Reconhecimento de ambiente.* Revista Pesquisa – FAPESP, n. 232, p. 70-71, 2015.
FARNEBÄCK, G. *Polynomial expansion for orientation and motion estimation.* PhD dissertation, Linköping University, Computer Vision, The Institute of Technology, 2002.



FELIPPE, J. Á. De M. *Caminhando Juntos – Manual das habilidades básicas de Orientação e Mobilidade*. 2001. Disponível em: <<http://www.deficienciavisual.pt/txt-caminhandojuntos.htm>>. Acesso em: 23 jul 2015.

FERNANDES, H. R. M. *Orientação de Cegos usando Visão por Computador*. Dissertação de Mestrado em Informática, Universidade de Trás-os-Montes e Alto Douro. Vila Real, 2010.

GONZATTO, A.; SANTOS, C.; MELO, F.; RODRIGUES, G.; FARIA, J. R. *Óculos sonar para deficientes Visuais*. Artigo, XIII Encontro Latino Americano de Iniciação Científica e IX Encontro Latino Americano de Pós-Graduação – Universidade do Vale do Paraíba. 2009.

HORN, B. K.; SCHUNCK, B. G. *Determining optical flow*. In: 1981 Technical symposium east. International Society for Optics and Photonics, p. 319-331, 1981.

IBGE - Instituto Brasileiro de Geografia e Estatística. *Censo Demográfico 2010: Características gerais da população, religião e pessoas com deficiência*. 2012. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf> Acesso em: 10 jun. 2015.

KAISER, E. B.; LAWOW, M. *Wearable navigation system for the visually impaired and blind people*. In: Computer and Information Science (ICIS), 2012 IEEE/ACIS 11th International Conference on. IEEE, p. 230-233, 2012.

KUIASKI, J. R. *Segmentação de Movimento por Fluxo Óptico*. Dissertação de Mestrado em Ciências – Área: Engenharia de Computação, Universidade Federal do Paraná. Curitiba, 2012.

KUMAR, A.; PATRA, R.; MAHADEVAPPA, M.; MAJUMBA, A. K. *An embedded system for aiding navigation of visually impaired persons*. Current Science, v. 104, n. 3, p. 302-306, 2013.

LEE, D. N. *A theory of visual control of braking based on information about time-to-collision*. In: Perception, n. 5, p. 437- 459, 1976.



LIYANAGE, D. K.; PERERA, M. U. S. *Optical flow based obstacle avoidance for the visually impaired*. In: Business Engineering and Industrial Applications Colloquium (BEIAC), 2012 IEEE. IEEE, p. 284-289, 2012.

LOW, T.; WYETH, G. *Obstacle Detection using Optical Flow*. Australasian Conference on Robotics and Automation, 2005.

LUCAS, B.; KANADE, T. *An iterative image registration technique with an application to stereo vision*. In Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI' 81), p. 674-679., 1981.

HORN, B. K. P.; SCHUNK. B. G. *Determining Optical Flow*. Artificial Intelligence, vol. 17, p. 185-203, 1981.

MCTI - Ministério da Ciência, Tecnologia e Inovação. *Catálogo Nacional de Produtos de Tecnologia Assistiva, 2015*. Disponível em: <<http://assistiva.mct.gov.br/index.php>> Acesso em: 16 jul. 2015.

MÜLLER, Dennis et al. Time to contact estimation using interest points. In: Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on. IEEE, 2009. p. 1-6.

NAKAJIMA, M.; HARUYAMA, S. *Indoor navigation system for visually impaired people using visible light communication and compensated geomagnetic sensing*. In: Communications in China (ICCC), 2012 1st IEEE International Conference on. IEEE, p. 524-529, 2012.

OPENCV, Manual. *The OpenCV Reference Manual - Release 2.4.9.0*. 2014. Disponível em: <<http://docs.opencv.org/opencv2refman.pdf>>. Acesso em: 31 ago 2015.

OTUYAMA, J. M. *Visão Estéreo*. 1998. Disponível em: <<http://www.inf.ufsc.br/~visao/1998/otuyama/1.htm>>. Acesso em: 21 ago 2015.

PAISOS, N. *Mobile Accessibility Tools for the Visually Impaired*. PhD thesis, New York University, 2012.



PEDRINI, H.; SCHAWARTZ, W. R. *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. São Paulo: Thomson, 2008.

PFLEGER, S. G. *Sistema de Navegação Robótica por Visão Computacional*. Trabalho de Conclusão de Curso de Bacharel em Ciências da Computação, Universidade Federal de Santa Catarina. Florianópolis, 2013.

PUNDLIK, S.; PELI, E.; LUO, G. *Time to collision and collision risk estimation from local scale and motion*. In: *Advances in Visual Computing*. Springer Berlin Heidelberg, p. 728-737, 2011.

PUNDLIK, S.; TOMASI, M.; LUO, G. *Collision detection for visually impaired from a body-mounted camera*. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. IEEE, p. 41-47, 2013.

RAN, L.; HELAL, S.; MOORE, S. *Drishti: an integrated indoor/outdoor blind navigation system and service*. In: *Pervasive Computing and Communications, 2004. PerCom 2004. Proceedings of the Second IEEE Annual Conference on*. IEEE, p. 23-30, 2004.

SAGREBIN, M.; NOGLIK, A.; PAULI, J. *Robust time-to-contact calculation for real time applications*. In: *Proceedings of 18th International Conference on Computer Graphics and Vision*. p. 128-133, 2008.

SANTOS, D. R. de G.; FERREIRA, W. R. B.; BORGES, M. A.; GONÇALVES, R. S. *Desenvolvimento de uma bengala eletrônica para locomoção de pessoas com deficiência visual*. VI Congresso Nacional de Engenharia Mecânica – CONEM 2010. Paraíba, 2010.

SHIN, B.S.; LIM, C. S. *Obstacle detection and avoidance system for visually impaired people*. In: *Haptic and Audio Interaction Design*. Springer Berlin Heidelberg, p. 78-85, 2007.

SILVA, L. S. da. *Segmentação de Movimento Coerente Aplicada à Codificação de Vídeos Baseada em Objetos*. Tese de Doutorado em Ciência da Computação, Universidade Federal do Rio Grande do Sul. Porto Alegre, 2011.



SOUHILA, K.; KARIM, A. *Optical flow based robot obstacle avoidance*. International Journal of Advanced Robotic Systems, v. 4, n. 1, p. 13-16, 2007.

SOUSA, K.; MARENGONI, M. *Uso de Visão Computacional em Dispositivos Móveis para o Reconhecimento de Faixa de Pedestres*. 2012.

TAPU, R.; MOCANU, B.; BURSUC, A.; ZAHARIA, T. *A smartphone-based obstacle detection and classification system for assisting visually impaired people*. In: Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on. IEEE, p. 444-451, 2013.

VARPE, K. M.; WANKHADE, M. P. *Visually Impaired Assistive System*. In: International Journal of Computer Applications, v. 77, n. 16, 2013.

WATANABE, Y.; SAKAUE, F.; SATO, J. *Time-to-Contact from Image Intensity*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. p. 4176-4183. 2015.

WHO - World Health Organization. *Visual impairment and blindness*. Fact Sheet, n. 282, 2014. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs282/en/>>. Acesso em: 16 mar 2015.

WU, Q.; WEI, J.; LI, X. *Research Progress of Obstacle Detection Based on Monocular Vision*. In: Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference on. IEEE, 2014. p. 195-198.