

**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**



DISSERTAÇÃO DE MESTRADO

**Dispositivo de Tecnologia Assistiva Baseado no Processamento
em Tempo Real de Imagens do Globo Ocular**

Tiago Figueiredo Vieira

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**Dispositivo de Tecnologia Assistiva Baseado no Processamento
em Tempo Real de Imagens do Globo Ocular**

por:

Tiago Figueiredo Vieira

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da
Universidade Federal de Pernambuco como parte dos requisitos para a obtenção do grau de
Mestre em Engenharia Elétrica.

ORIENTADOR: Prof. Eduardo Fontana, PhD.

Recife, Abril de 2009.

© Tiago Figueiredo Vieira, 2009

V657d

Vieira, Tiago Figueiredo.

Dispositivo de tecnologia assistiva baseado no processamento em tempo real de imagens do globo ocular / Tiago Figueiredo Vieira. – Recife: O Autor, 2009.

xiii, 110 folhas, il : figs., tabs.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Elétrica, 2009.

Inclui Referências Bibliográficas e Apêndice.

1. Engenharia Elétrica. 2. Rastreamento ocular. 3. LED Infravermelho. 4. CCD. 5. ELA. 6. Síndrome do Encarceramento. I. Título.

UFPE

621.3

CDD (22. ed.)

BCTG/2009-145



Universidade Federal de Pernambuco

Pós-Graduação em Engenharia Elétrica

PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE
DISSERTAÇÃO DO MESTRADO ACADÊMICO DE

TIAGO FIGUEIREDO VIEIRA

TÍTULO

**“DISPOSITIVO DE TECNOLOGIA ASSISTIVA BASEADO NO
PROCESSAMENTO EM TEMPO REAL
DE IMAGENS DO GLOBO OCULAR”**

A comissão examinadora composta pelos professores: EDUARDO FONTANA, DES/UFPE, RAFAEL DUEIRE LINS, DES/UFPE e TSANG ING REN, CIN/UFPE sob a presidência do prof. Eduardo Fontana, Coordenador do Programa de Pós-Graduação em Engenharia Elétrica, consideram o candidato **TIAGO FIGUEIREDO APROVADO.**

Recife, 29 de abril de 2009.

EDUARDO FONTANA

Orientador e Membro Titular Interno

RAFAEL DUEIRE LINS

Membro Titular Interno

TSANG ING REN

Membro Titular Externo

Aos meus pais,
Amilton e Carmem
ao meu irmão,
Luiz Paulo
e à minha noiva,
Marcelle.

Agradecimentos

Primeiramente a Deus pela vida, saúde e felicidade.

Ao Professor Eduardo Fontana pela oportunidade e apoio oferecidos.

À minha família, pelo amor e dedicação incondicionais.

Finalmente, gostaria de agradecer à CAPES pelo suporte financeiro e ao Programa de Pós-Graduação em Engenharia Elétrica da UFPE.

TIAGO FIGUEIREDO VIEIRA

Universidade Federal de Pernambuco

13 de Abril de 2009

Resumo da dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica

Resumo

Dispositivo de Tecnologia Assistiva Baseado no Processamento em Tempo Real de Imagens do Globo Ocular

Recife, Abril de 2009.

Orientador: Prof. Eduardo Fontana, PhD

Área de concentração: Fotônica

Palavras-chaves: rastreamento ocular, LED infravermelho, CCD, ELA, síndrome do encarceramento

Número de Páginas: 110

Nesta dissertação são investigadas algumas técnicas de rastreamento ocular para o desenvolvimento de um dispositivo de tecnologia assistiva simples e de baixo custo, para indivíduos com deficiências motoras severas. O dispositivo desenvolvido permite que o usuário controle a posição de um cursor em uma tela gráfica e acione programas de computador com seus movimentos oculares. Dessa forma, os indivíduos vítimas de doenças degenerativas do sistema nervoso ou de acidentes vasculares cerebrais, com pouca ou nenhuma liberdade de movimento dos membros e/ou da cabeça, podem exercer alguma atuação no mundo à sua volta se comunicando com outras pessoas. Os sistemas desenvolvidos utilizam *webcams* de baixo custo e iluminação do objeto com LEDs infravermelhos. Algumas técnicas de detecção de formas e de calibração de câmeras foram investigadas para localização, em tempo real, da pupila do globo ocular de um indivíduo. Os algoritmos de processamento de imagens desenvolvidos para funcionamento da interface foram implementados inicialmente em Matlab[®] e posteriormente em C. Este último forneceu um ganho substancial em velocidade de processamento. Foi desenvolvida uma primeira configuração de dispositivo de tecnologia assistiva, em formato de armação de óculos, que funciona sem necessidade de re-calibração, desde que o usuário permaneça com a cabeça imóvel. São também descritos resultados preliminares obtidos no âmbito de desenvolvimento de uma segunda configuração de dispositivo, para uso remoto, que permite rastrear os movimentos oculares independentemente dos movimentos da cabeça do indivíduo.

Abstract of Dissertation presented to UFPE as a partial fulfillment of the requirements for the degree of
Master in Electrical Engineering

Abstract

Assistive Technology Device Based on Real Time Image Processing of the Eyeball

Recife, April 2009.

Supervisor: Prof. Eduardo Fontana, PhD

Concentration Area: Photonics

Keywords: eye tracking, infrared LED, CCD, ELA, locked-in-syndrome

Number of Pages: 110

This dissertation reports a number of eye tracking techniques for the development of a simple, low-cost, assistive technology device, for severely impaired individuals. The device allows the user to control the position of a cursor on a computer screen and to activate computer programs, with his/her eyeball movements, thus enabling the possibility for those individuals who suffer the sequels of degenerative diseases of the nervous system or cerebrovascular accidents, with minimum, if any, movements of the limbs and/or head, to exert some action on the outer world and to communicate with other individuals. The developed systems employ low-cost webcams and infrared LEDs. Some techniques of shape detection and camera calibration were investigated for real-time location of the pupil on the eyeball of an individual. The image processing algorithms developed for operation of the device were first implemented in Matlab[®] and then written in C language, the latter yielding executables with a substantial gain in processing speed. The first implementation of the assistive technology device was configured in an eyeglass, head mounted, format and works without the need of recalibration, provided the user maintains his/her head still. Preliminary results related to the implementation of a second device configuration that allows remote eye-tracking, without recalibration, independently of the user's head movements, are also reported.

Sumário

Lista de figuras	ix
Lista de tabelas	xiii
1. INTRODUÇÃO	1
1.1. Rastreamento ocular	3
1.2. Estado da arte.....	5
1.3. Escopo deste trabalho	8
1.4. Organização da dissertação	9
2. A VISÃO	11
2.1. O globo ocular	11
2.2. Fontes de iluminação e limites de exposição do olho humano.....	15
3. SENSORES DE IMAGEM E MODELAMENTO MATEMÁTICO DE CÂMERAS DIGITAIS	19
3.1. Tecnologias CCD e CMOS	20
3.2. Modelo matemático da câmera.....	24
3.3. Calibração das câmeras	30
4. ALGUMAS TÉCNICAS DE PROCESSAMENTO DE IMAGENS	38
4.1. Limiarização de imagens	38
4.2. Filtragem e detecção de contornos	41
4.3. Algumas técnicas de detecção de formas	47
4.4. Detecção de elipses.....	55
5. RASTREAMENTO OCULAR	57
5.1. Modelo matemático	57
5.2. Determinação do eixo visual	61
5.3. Situações específicas	64
6. SISTEMA DE RASTREAMENTO OCULAR 1 – SRO1	66
6.1. Descrição do sistema	66

6.2.	Processo de calibração e transformação das coordenadas	70
6.3.	Algoritmos desenvolvidos.....	71
6.4.	Resultados	75
7.	SISTEMA DE RASTREAMENTO OCULAR 2 – SRO2	89
7.1.	Procedimento de calibração	89
7.2.	Implementação do sistema SRO2 e resultados preliminares	95
8.	CONSIDERAÇÕES FINAIS	99
8.1.	Conclusões	99
8.2.	Sugestões para trabalhos futuros	100
	APÊNDICE A – PROCEDIMENTO DE CALIBRAÇÃO DO SISTEMA SRO2	103
	REFERÊNCIAS BIBLIOGRÁFICAS	107

Lista de figuras

Figura 1.1 – Indivíduo utilizando equipamento de tecnologia assistiva baseado no rastreamento do globo ocular realizado por eletrodos.	7
Figura 1.2 – Imagens: (a) pupila clara, (b) pupila escura, e (c) subtração.	8
Figura 2.1 – Seção transversal do olho humano [30].	11
Figura 2.2 – Espectro de absorção da retina [32].	13
Figura 2.3 – Absorção espectral do meio ocular humano na região do ultravioleta próximo e infravermelho médio.	14
Figura 2.4 – Valores de MPE para exposição ocular direta à radiação visível e no infravermelho próximo [33].	17
Figura 3.1 – Componente básico de um sensor de imagem.	20
Figura 3.2 – Arquitetura CCD e CMOS [37].	21
Figura 3.3 – Responsividade espectral do fotodiodo de silício [38].	21
Figura 3.4 – Modelo da câmera de orifício com projeção em perspectiva.	25
Figura 3.5 – Padrão típico utilizado na calibração da câmera [41].	31
Figura 4.1 – Imagem original (a), e binarizações resultantes com limiares 70 (b), 90 (c) e 150 (d).	39
Figura 4.2 – Histograma normalizado $p(r_k)$ da imagem da Figura 4.1a.	40
Figura 4.3 – Exemplo de aplicação de filtragem. Imagem original (a), resultado da aplicação das máscaras: $sobel_u$ (b), $sobel_v$ (c) e laplaciano' (d).	47
Figura 4.4 – (a) Plano da imagem uv e (b) plano de Hough ab	48
Figura 4.5 – Representação normal de uma linha (a) e subdivisão do espaço de Hough em células (b).	49
Figura 4.6 – Plano de Hough e respectiva representação 3D para o intervalo $19 \leq c_3 \leq 21$ (a) e (b), e análise idêntica para o intervalo $15 \leq c_3 \leq 60$ (c) e (d).	51
Figura 4.7 – Imagem original (a), obtenção das bordas através do filtro de Sobel (b), aplicação da Transformada Circular de Hough para $R = 21$ (c) e $R = 30$ pixels (d).	52

Figura 4.8 – Detecção errônea do centro da pupila usando a Transformada Circular de Hough. Imagem original (a), contornos de Sobel (b), soma dos planos correspondentes à variação no raio (c) e centróides deslocados (d).....	54
Figura 4.9 – Exemplo de obtenção do centróide da pupila através do rastreamento de elipses. Imagens originais (a) e (d), contornos da binarização (b) e (e), e elipses resultantes (c) e (f), respectivamente.	56
Figura 5.1 – Disposição geométrica para o modelo matemático de rastreamento ocular [51].	58
Figura 5.2 – Eixos óptico e visual para o olho direito e seus ângulos de desvio horizontal e vertical θ e φ , e $\theta+\alpha$ e $\varphi+\beta$, respectivamente. Todos os vetores \vec{o} , \vec{p} e \vec{t} são relativos ao sistema de cena xyz	62
Figura 6.1 – Configuração do Sistema de Rastreamento Ocular 1 – SRO1.....	67
Figura 6.2 – (a) Espectros de diferentes LEDs. (b) Curvas de transmissão de diversos filtros ópticos.....	68
Figura 6.3 – Imagem do olho iluminado com IR (a) e binarização e centróide da pupila (b).	69
Figura 6.4 – Transformação espacial de coordenadas de imagem para tela.	71
Figura 6.5 – Fluxograma do programa escrito em Matlab [®] R2007a.	72
Figura 6.6 – Fluxograma simplificado do programa escrito em C.....	73
Figura 6.7 – Imagem do olho (a) e binarização e centróide da pupila (b).	76
Figura 6.8 – Teste de calibração (a) e erro diferencial (b).	77
Figura 6.9 – Rastreamento ocular baseado na leitura de um texto.....	78
Figura 6.10 – Circunferência teste (R=200 pixels) (a) e coordenadas polares (r, θ) das medidas (b).....	78
Figura 6.11 – Funcionamento do programa de rastreamento ocular.....	83
Figura 6.12 – Teste de calibração (a) e erro diferencial (b).	84
Figura 6.13 – Circunferência teste (R=256 pixels) (a) e coordenadas polares (r, θ) dos dados adquiridos (b).....	85
Figura 6.14 – Imagem original (a), binarização (b) e elipses aproximadas aos contornos fechados.	86
Figura 7.1 – Padrão planar utilizado na calibração.	90
Figura 7.2 – Imagens utilizadas na calibração da câmera.	91

Figura 7.3 – Delimitação interativa dos vértices (a) e obtenção dos vértices por processamento de imagens (b).....	91
Figura 7.4 – Parâmetros extrínsecos para uma imagem.	94
Figura 7.5 –Imagem original com distorção (a), e corrigida (b).	95
Figura 7.6 – Sistema para rastreamento ocular remoto.	96
Figura 7.7 – Imagens do olho fitando o centro e os quatro cantos da tela.	97
Figura 7.8 – Múltiplas reflexões decorrentes da utilização de óculos de grau.....	98

Lista de tabelas

Tabela 2.1 – Valores típicos dos parâmetros geométricos e das constantes ópticas de um olho normal humano[31].	12
Tabela 2.2 – Fatores de correção C dependentes do comprimento de onda [33]......	18
Tabela 5.1 – Parâmetros da Figura 5.1.	59
Tabela 5.2 – Parâmetros típicos do olho humano [51]......	63
Tabela 6.1 – Resumo das estatísticas dos testes (todas as unidades em pixels)......	79
Tabela 6.2 – Erros médios e valores máximos.	80
Tabela 6.3 – Estatísticas dos testes realizados com 1 sujeito.	86
Tabela 6.4 – Resumo das comparações relevantes às duas aplicações desenvolvidas.....	87
Tabela 7.1 – Parâmetros da câmera obtidos em 4 calibrações com 20 imagens.	93

1. INTRODUÇÃO

*"Nenhum homem é uma ilha, isolado em si mesmo;
todos somos partes de um mesmo continente".*

John Donne, poeta inglês.

O ser humano tem a necessidade de se comunicar. A fala, a escrita, os gestos, as expressões artísticas fazem com que cada um se sinta parte de um todo, inserindo e recebendo informações em uma sociedade contemporânea, que tende, cada vez mais, a interagir. Infelizmente, existem pessoas que, apesar de preservarem sua inteligência, possuem certas limitações que lhes impossibilita exercer a comunicação em sua plenitude. Os cegos, surdos, mudos, tetraplégicos, deficientes físicos, aqueles com problemas de dicção, sofrem com a limitada interação pessoal imposta por suas deficiências.

Uma condição extrema de incapacidade de comunicação é a de indivíduos que sofrem da síndrome do encarceramento, conhecida na literatura como *locked-in-syndrome* (LIS). Essa patologia é causada mais comumente por acidentes vasculares cerebrais (AVCs) isquêmicos do tronco cerebral [1] ou doenças como a Esclerose Lateral Amiotrófica (ELA) [2]. Pacientes nessa condição usam os seus movimentos voluntários limitados para se comunicar com a família, amigos, e outras pessoas envolvidas em seus cuidados. Quando também perdem a fala, algumas pessoas podem ainda mover a cabeça, talvez o nariz ou a língua, mas outras têm como únicos movimentos voluntários restantes o piscar dos olhos e os movimentos verticais e horizontais do globo ocular.

Um caso bem conhecido de um indivíduo, vítima de AVC do tronco cerebral, que se comunicava com o movimento dos olhos é o do editor chefe da Revista Elle francesa, Jean Dominique Bauby, que no estado LIS foi capaz de escrever o livro “O escafandro e a borboleta”, lançado em filme em 2007 [3]. O livro relata o cotidiano do autor de maneira poética e bem humorada, apesar das adversidades, e foi escrito com o auxílio de sua

fonoaudióloga que ditava o alfabeto letra por letra. Bauby informava a letra desejada através de um piscar da pálpebra esquerda, um dos poucos movimentos voluntários que lhe restara.

Outro caso é o do cientista Stephen William Hawking, que ocupa o posto de professor lucasiano do Departamento de Física Teórica e Matemática Aplicada da Universidade de Cambridge, na Inglaterra, desde 1979 [4]. Hawking, portador de ELA, foi gradualmente perdendo o movimento de sua musculatura voluntária, incluindo a força para manter sua cabeça erguida. Mesmo sob limitações físicas excessivas, Hawking tem sido capaz de contribuir significativamente para o desenvolvimento da astronomia e representa um notável exemplo da produtividade alcançável por indivíduos com LIS, que apesar de serem fisicamente debilitados, mantêm sua integridade intelectual.

As doenças do neurônio motor, incluindo ELA, afetam mais frequentemente pessoas entre 40 e 70 anos, com aproximadamente 120 mil novos casos por ano no mundo (328 novos casos por dia) [5]. As estatísticas são baseadas na incidência; número de novos casos por período de tempo de geralmente um ano, e prevalência; número de casos existentes que permanece constante em qualquer período de tempo. A incidência dessas doenças é de 2 a cada 100 mil pessoas, enquanto a prevalência é de 6 a cada 100 mil pessoas no mundo. No Brasil, a prevalência é de 2.500 pessoas por ano. Essas doenças afetam cada paciente de maneira diferenciada e pode ter um impacto devastador em sua família, carreira e amigos. A natureza progressiva da degeneração motora requer adaptação constante aos níveis crescentes de deficiência que, por sua vez, fazem com que o paciente necessite cada vez mais de amparo.

Este trabalho tem como objetivo o desenvolvimento de um dispositivo de tecnologia assistiva para indivíduos com severas limitações de atuação e de comunicação. Especificamente, o trabalho trata do estudo e desenvolvimento de um sistema capaz de possibilitar ao sujeito o acesso a todos os recursos que um microcomputador pode oferecer, através apenas do rastreamento dos movimentos do seu globo ocular¹. Em outras palavras, o usuário deve ser capaz de movimentar o cursor do *mouse* na tela do equipamento para

¹ Rastreamento ocular (*eye tracking*) é um conceito ligeiramente diferente de rastreamento do olhar (*gaze tracking*), entretanto, o primeiro termo será utilizado para referenciar ambos.

acionar programas e gerar informação, apenas com o movimento dos olhos. A melhoria na qualidade de vida do indivíduo acontece de forma imediata e evidente, em vista da perspectiva de interação com outros indivíduos e até da possibilidade de realização de algumas atividades profissionais.

Na subseção seguinte é feito um breve relato histórico do conhecimento gerado a respeito das características inerentes ao movimento ocular e do desenvolvimento de equipamentos de tecnologia assistiva baseados no rastreamento ocular.

1.1. Rastreamento ocular

Os movimentos oculares no processo de leitura de um texto não são contínuos, mas compostos por paradas curtas (fixações) intercaladas por rápidos saltos. Sua observação levanta questões sobre leitura como, quais as características das palavras que chamam a atenção do sujeito, por quanto tempo e, quando é necessário visualizar algo por mais de uma vez. Essas informações são utilizadas em técnicas contemporâneas para a extração de padrões de visão de um usuário da *internet*, por exemplo, conforme descrito na Seção 1.2.

Além da mera observação do comportamento visual, os métodos iniciais para rastreamento da fixação ocular eram bastante invasivos, envolvendo contato mecânico com a córnea. Dodge *et al.*[6], desenvolveram a primeira técnica de rastreamento ocular precisa e não invasiva, usando luz refletida na córnea. O sistema gravava apenas a posição horizontal numa placa fotográfica móvel e a cabeça do sujeito deveria permanecer imóvel. Pouco tempo depois, em 1905, Judd, McAllister & Steel [7] utilizaram fotografia móvel para gravar a variação temporal dos movimentos dos olhos em duas dimensões. Esse método utilizava um pequeno cisco branco de material refletor inserido no olho do sujeito ao invés de luz refletida diretamente da córnea. Esses e outros pesquisadores interessados nos movimentos oculares contribuíram com avanços nos sistemas de rastreamento durante a primeira metade do século XX, mesclando técnicas de reflexão corneana e filmagem.

Em 1947, Paul Fitts *et al.*[8] começaram a usar câmeras de vídeo para estudar os padrões de visão de pilotos enquanto estes utilizavam os controles e instrumentos da cabine do avião durante o procedimento de aterrissagem. Esses estudos representaram a

primeira aplicação do rastreamento ocular como engenharia de usabilidade – o estudo sistemático da interação dos usuários com os produtos de forma a melhorar o projeto desses produtos.

Hartridge e Thompson [9] inventaram o primeiro equipamento de tecnologia assistiva que seria montado na cabeça do usuário. Apesar de grosseiro para os padrões atuais, essa inovação contornava muitas restrições impostas pelo movimento da cabeça. Nos anos 1960, sistemas com essa configuração evoluíram se tornando menos invasivos e reduzindo as restrições na liberdade de movimento da cabeça do indivíduo.

Em outro avanço relevante à aplicação do rastreamento ocular para interação homem-máquina, Mackworth *et al.*[10] conceberam um sistema que registrava os movimentos dos olhos superpostos à cena sendo visualizada pelo usuário.

Pesquisas sobre o movimento e rastreamento ocular floresceram nos anos 70 com grandes avanços tanto na tecnologia de rastreamento quanto na teoria da psicologia para fazer a ligação entre os dados de rastreamento e o processo cognitivo. Muitos dos trabalhos foram voltados para a psicologia e fisiologia e exploraram como o olho humano opera e o que isso pode revelar sobre os processos de percepção e cognição. Mas as publicações dos anos 1970 indicam uma diminuição na atividade associada a equipamentos de tecnologia assistiva. Isso pode ter ocorrido devido ao esforço envolvido na aquisição e análise dos dados. Naquela época não era incomum alguns dias de duração para processar dados que levavam apenas minutos para serem coletados. Trabalhos em vários laboratórios, particularmente aqueles ligados à aviação militar, se concentravam em solucionar as deficiências na tecnologia de rastreamento ocular e na análise de dados.

Muitos dos trabalhos relevantes na década de 1970 foram voltados para o desenvolvimento das técnicas existentes, a fim de aumentar a exatidão, a precisão, e de reduzir o impacto dos protótipos nos usuários. A descoberta que reflexões múltiplas no olho poderiam ser usadas para precisar o posicionamento do globo ocular de um indivíduo independentemente do movimento da cabeça [11] levou a um aumento da precisão. Usando essa descoberta, instituições militares americanas desenvolveram um equipamento remoto que reduziu drasticamente o tamanho do protótipo, e conseqüentemente as

limitações do usuário [12]. Essas parcerias, juntamente com o advento dos computadores pessoais, contribuíram fortemente para a automação da análise dos dados provenientes do rastreamento, tornando possível o rastreamento em tempo real.

Apesar de amplamente utilizado nos estudos de percepção e processo cognitivo, somente com a ampla utilização dos computadores nos anos 1980 que o rastreamento ocular foi aplicado como um instrumento de interação homem-máquina. Nesse período, as propostas iniciais focavam aplicações de assistência para indivíduos com necessidades especiais [13].

1.2. Estado da arte

As técnicas de rastreamento ocular atuais são globalmente organizadas segundo seus objetivos e métodos. Segundo o objetivo, cada técnica se enquadra em uma das definições a seguir.

Aquisição e/ou análise de padrões de visão

Visto que os padrões de visão estão intimamente ligados ao foco de atenção de um sujeito, essas análises são muito usadas em ciências cognitivas como psicologia, para inferências sobre percepção, atenção e disfunções de aprendizado [14]. Essa abordagem é também utilizada com o intuito de treinamento de decisões táticas para pilotos de caça, baseado em padrões de visão extraídos de pilotos mais experientes [15], [16]. Outra aplicação muito frequente atualmente é a determinação de interesse dos internautas, que fornece informação sobre o quanto o design de um *website* é bom ou não, quais as principais preferências entre homens e mulheres [17], e quais padrões de *merchandising* são realmente impactantes ou não, e se maximizam ou prejudicam as vendas [18].

Desenvolvimento de interface homem-máquina

Existem sistemas onde a máquina consegue reproduzir as feições humanas sendo captadas por uma câmera, usando algoritmos de visão computacional e rastreamento ocular [19]. Outros exemplos são os dispositivos de assistência para possibilitar a comunicação de sujeitos com deficiências motoras severas [20], [21], com a interface sendo implementada através do uso de um computador.

Com relação à metodologia, as técnicas podem pertencer aos casos em que:

(a) O sistema é preso à cabeça do usuário e o olhar é extraído em relação ao posicionamento da cabeça. O ponto de mira (do termo inglês *Point of Gaze* – POG) na tela pode ser extraído se o usuário mantiver sua cabeça parada, caso contrário, algum método de rastreamento de cabeça deve ser usado para corrigir a referência global entre as coordenadas de câmera e olhar.

(b) O sistema é remoto, *i.e.*, nenhum aparato é acoplado ao usuário de forma que o olhar é adquirido em relação à cena sendo observada. O POG é obtido com alguma liberdade de movimento da cabeça. O processo de calibração é mais simples quanto mais complexo for o aparato de captação de imagens. Mais detalhes são tratados no Capítulo 5.

A seguir são citadas algumas técnicas analisadas para a realização do rastreamento ocular.

Técnicas

Uma das primeiras técnicas utilizadas é aquela que usa lentes de contato especiais, que possuem um espelho acoplado ou um sensor de campo magnético [22]. Esse tipo de sistema, apesar de ser um tanto desconfortável ao usuário, fornece a obtenção dos movimentos oculares com extrema sensibilidade, sendo, portanto, a escolha de pesquisadores envolvidos em estudos relacionados à fisiologia dos movimentos do olho.

Outra técnica faz uso dos potenciais elétricos medidos com eletrodos de contato posicionados ao redor dos olhos, como ilustrado na Figura 1.1. A diferença de potencial elétrico gerada pela contração dos músculos oculares é medida por cada eletrodo, tornando possível a correlação entre diferentes movimentos do olho e o ponto de concentração na tela [23]. A desvantagem desse sistema é a imprecisão no caso de pequenos movimentos que geram sinais abaixo do limite de sensibilidade dos eletrodos e o desconforto ao usuário causado pela adesão dos eletrodos à pele. No entanto, esses sensores são úteis quando o objetivo é medir variações bruscas na rotação do globo ocular, devido ao rápido tempo de resposta e alta sensibilidade às variações de tensão inerentes a esses movimentos.



Figura 1.1 – Indivíduo utilizando equipamento de tecnologia assistiva baseado no rastreamento do globo ocular realizado por eletrodos.

Atualmente, as vantagens alcançadas com as técnicas de captura e processamento de imagens tornaram esses métodos predominantes em rastreamento ocular, especialmente no desenvolvimento de interfaces homem-máquina para possibilitar acesso ao computador. Apesar de existirem casos onde se usa apenas uma câmera [24], a utilização de técnicas de iluminação melhora consideravelmente o desempenho do sistema quando se trata de robustez e precisão. Essas técnicas geralmente realizam o rastreamento ocular usando processamento de imagem para extração de informações de posicionamento do olho como centro da pupila e um ou mais reflexos de fontes de luz. C. H. Morimoto, *et al.*[25], utilizaram um sistema de iluminação infravermelha constituído pela alternância entre iluminação “no eixo” (próximo do eixo óptico da câmera) e “fora do eixo” (distante do eixo óptico da câmera). A radiação emitida quando os LEDs próximos à câmera estão ativados, incide na retina e, devido à sua propriedade de retro-reflexão², retorna ao CCD causando o efeito pupila-clara, como mostrado na Figura 1.2-(a). Esse é o mesmo efeito que ocorre em fotografias efetuadas em ambientes escuros, onde o diâmetro da pupila é grande e a radiação do *flash* da máquina incide na retina, refletindo a irrigação sanguínea dessa camada, causando o conhecido efeito “olho-vermelho”. Quando a iluminação “fora do eixo” é ativada, obtém-se o efeito “pupila-escura”, como mostrado na Figura 1.2-(b). Essa radiação, no entanto, é captada apenas como o reflexo na córnea, pois a radiação que incide na retina não é captada pela câmera, devido à distância do eixo óptico desta ao

² Uma superfície possui a propriedade de retro-reflexão quando a radiação refletida acontece prioritariamente na direção da radiação incidente.

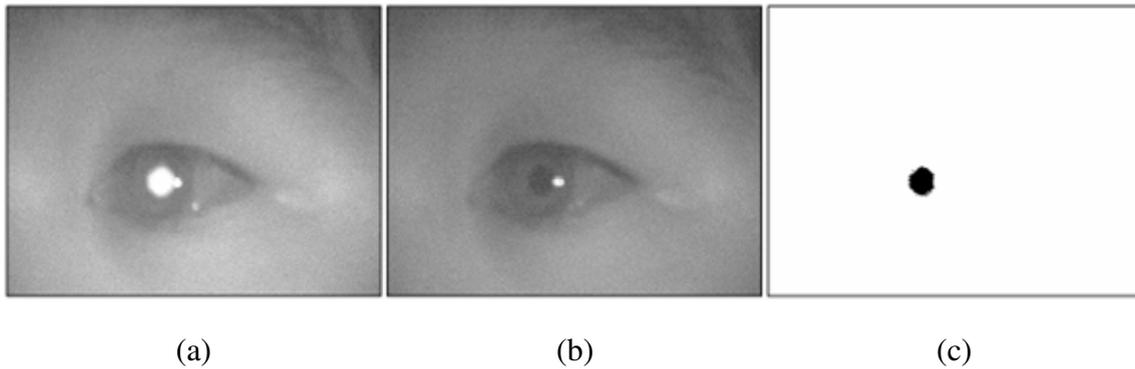


Figura 1.2 – Imagens: (a) pupila clara, (b) pupila escura, e (c) subtração.

ponto de iluminação “fora do eixo”. A subtração das imagens resultantes é mostrada na Figura 1.2-(c), onde é possível extrair as coordenadas do centróide da pupila através de um algoritmo de cálculo centro de massa, como tratado na Seção 6.1.

Esse método possui duas grandes vantagens. A primeira é tornar a captura de imagens robusta com relação aos ruídos provenientes da intensidade de iluminação ambiente, cujos níveis de cinza permanecem aproximadamente iguais nas imagens “pupila-clara” e “pupila-escura”. Além disso, o processamento de imagens necessário à resolução do centróide da pupila é preciso e computacionalmente barato. Apesar de possuir essas vantagens, o resultado alcançado pelos autores é insatisfatório por apresentar uma grande intermitência do mouse na tela.

1.3. Escopo deste trabalho

Como descrito anteriormente, o objetivo deste trabalho é possibilitar ao usuário o controle do mouse na tela do computador através do rastreamento ocular. Para tanto, os protótipos desenvolvidos são formados exclusivamente por câmeras USB populares (*webcams*) e iluminação infravermelha. Isso implica em dizer que a comunicação homem-máquina é realizada prioritariamente através dos softwares escritos especificamente para cada caso (cabeça estática e cabeça móvel), tendo os sensores de imagens e fontes de radiação infravermelha apenas como parâmetros de entrada dos sistemas.

A principal motivação é proporcionar a interatividade àqueles indivíduos que apresentam dificuldades medianas ou severas em utilizar um mouse tradicional, através de

um sistema de baixo custo controlado pelos movimentos do globo ocular. Existem soluções comerciais sofisticadas que, em sua maioria, fazem uso de técnicas de processamento de imagens, e podem ser importadas com um custo superior a 10 mil dólares [26], [27]. Há também o mouse ocular nacional que utiliza Eletromiógrafos como os da Figura 1.1, que seria disponibilizado comercialmente por R\$ 150,00 em novembro de 2006 [28]. Entretanto, a comercialização desse produto não foi efetivada pela empresa responsável. Atualmente no Brasil, não se tem notícia de um sistema de rastreamento ocular que seja satisfatório (possibilite um controle preciso do cursor do computador) e acessível.

Essa abordagem qualifica os protótipos desenvolvidos como pertencentes à classe dos sistemas de rastreamento ocular que utilizam iluminação infravermelha, tendo o desenvolvimento de uma interface homem-máquina como objetivo.

1.4. Organização da dissertação

Esta dissertação está organizada em oito capítulos. O Capítulo 2 trata da fisiologia do olho humano. Os tipos de lesão ocular passíveis de ocorrência quando o olho é submetido a radiações IR do espectro eletromagnético são apresentados. Por fim, a metodologia que garante uma potência óptica segura de exposição é apresentada.

O Capítulo 3 apresenta o sensor de imagens como elemento inicial do sistema. As características das tecnologias CCD e CMOS são analisadas e comparadas para obtenção do melhor resultado desejável dando prioridade à velocidade e à qualidade das imagens. É descrita a modelagem utilizada em visão computacional, com destaque à reconstrução de coordenadas tridimensionais a partir de coordenadas bidimensionais do plano imagem. É também descrita a teoria matemática necessária para obtenção dos parâmetros intrínsecos e extrínsecos da câmera, para relacionar as coordenadas de imagem com as coordenadas de cena.

O Capítulo 4 trata de algumas técnicas de processamento de imagens digitais. As teorias relevantes utilizadas nos projetos desenvolvidos, como binarização de imagens, filtragem e detecção de bordas, transformadas de Hough para detecção de círculos e

circunferências e detecção de elipses são descritas. As vantagens e desvantagens dos métodos empregados para rastreamento ocular são especialmente analisadas.

O Capítulo 5 apresenta a formulação matemática que torna possível o rastreamento ocular. Nesse capítulo o sistema é descrito, com a apresentação do ambiente desenvolvido para a aplicação da teoria geral de rastreamento usando iluminação infravermelha. Trata-se aí do modelo matemático geral, para diversas configurações de sistemas de rastreamento, e as vantagens e desvantagens de cada possível sistema são analisadas. É então discutida a metodologia que foi implementada neste trabalho, para possibilitar o controle do mouse do computador pelo olho do usuário.

O Capítulo 6 apresenta os resultados obtidos com a implementação do Sistema de Rastreamento Ocular 1 (SRO1) com funcionamento baseado no posicionamento fixo da cabeça do usuário. Os detalhes da montagem do dispositivo são apresentados, bem como as técnicas de iluminação infravermelha e os fluxogramas dos algoritmos desenvolvidos em Matlab[®] R2007a e na linguagem de programação C. As vantagens e desvantagens proporcionadas por cada linguagem são especialmente analisadas.

O Capítulo 7 apresenta a implementação preliminar do Sistema de Rastreamento Ocular 2 (SRO2) que visa obter o ponto de visão do usuário de maneira remota e com certa liberdade no movimento da cabeça. Resultados iniciais inerentes à montagem e calibração do sistema são apresentados, bem como os problemas relativos aos dispositivos aplicados.

No Capítulo 8 encontram-se as considerações finais referentes ao trabalho apresentado e sugestões e planejamento de métodos a serem desenvolvidos em trabalhos futuros.

2. A VISÃO

"A visão é o tato do espírito."

Fernando Pessoa.

2.1. O globo ocular

A Figura 2.1 mostra uma seção horizontal simplificada do olho humano. As dimensões médias desse órgão estão listadas na Tabela 2.1. O globo ocular é aproximadamente uma esfera de diâmetro médio de 20 mm, e é encoberto por três

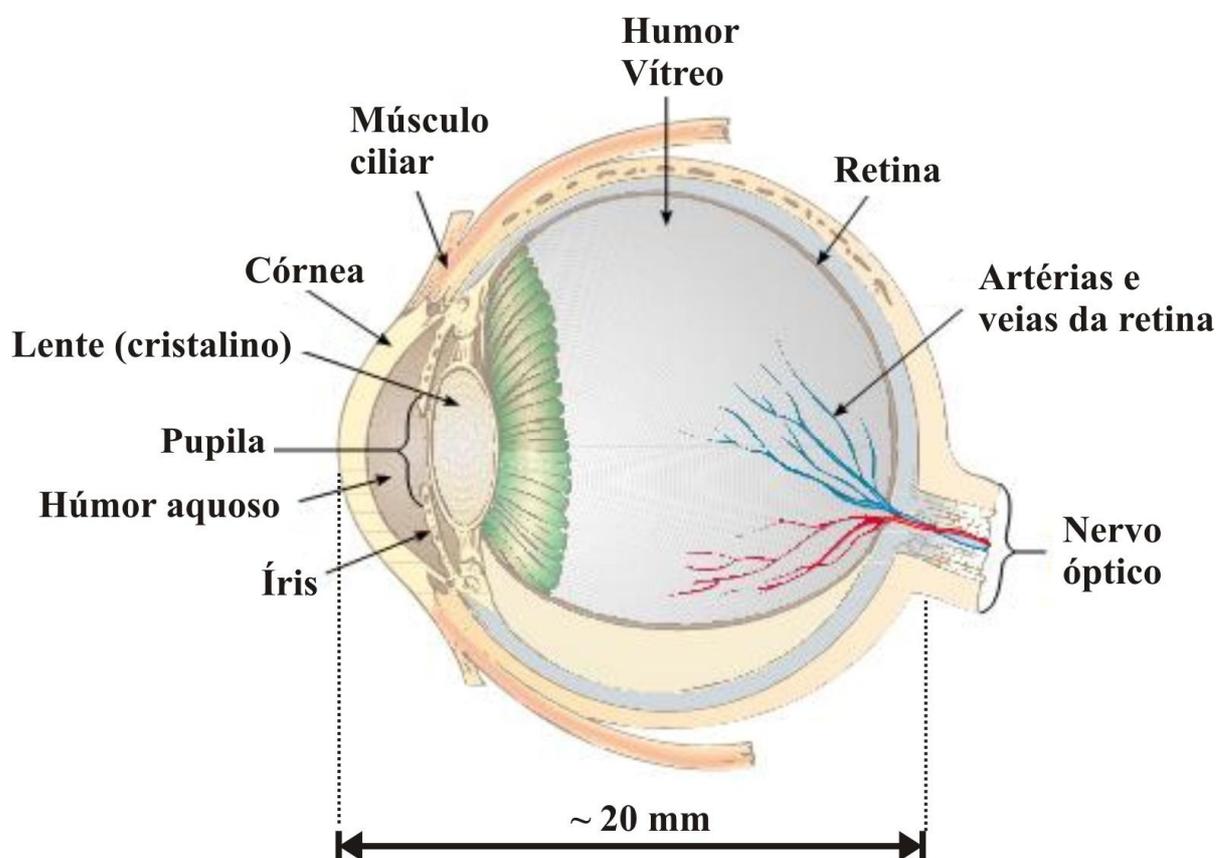


Figura 2.1 – Seção transversal do olho humano [30].

membranas: a camada externa é constituída pela córnea e pela esclera, e serve para proteção; acamada média é formada pela íris, pela coróide, e pelo corpo ciliar e acamada interna é constituída pela retina que é a parte nervosa, sensível à luz.

Íris

A quantidade de luz que pode entrar no globo ocular é regulada pela íris, a estrutura que circunda a pupila. Em função da iluminação ambiente, o diâmetro pupilar pode variar desde o mínimo de 2 mm (com luz intensa) até 7 mm (com luz muito fraca) [29]. Isso corresponde a uma variação de área de aproximadamente 16 vezes, o que permite ampla faixa de ajuste para a intensidade da iluminação incidente na retina. Esse controle é realizado pelo músculo ciliar. Essas modificações do diâmetro pupilar – chamadas de reflexo de acomodação – são acompanhadas por variações da forma do cristalino que também dependem da atuação do músculo ciliar. A pupila ainda apresenta outra resposta reflexa, chamada de reflexo consensual ou fotomotora em que qualquer variação abrupta da iluminação ambiente produz alteração correspondente no diâmetro pupilar. Por exemplo, o aumento súbito da iluminação ambiente provoca constrição das duas pupilas, mesmo quando a luz incide sobre apenas um dos olhos.

Tabela 2.1 – Valores típicos dos parâmetros geométricos e das constantes ópticas de um olho normal humano[31].

Superfície	Raio de curvatura [mm]	Índice de refração		Distância da superfície anterior da córnea [mm]	Poder de refração [D]
		Anterior	Posterior		
Córnea anterior	7.8	1.00 (ar)	1.376	0	+ 48.2
Córnea posterior	6.8	1.376	1.336 (aquoso)	0.5	- 5.9
Cristalino anterior	10.0	1.336	1.386	3.6	+ 5.0
Cristalino posterior	- 6.0	1.386	1.336 (vítreo)	7.2	+ 8.3
Retina				24.0	

Retina

Em cada retina há cerca de 100 milhões de fotorreceptores (cones e bastonetes) que emitem moléculas neurotransmissoras a uma taxa que é máxima na escuridão, e que diminui de forma logarítmica com o aumento da intensidade luminosa [31]. Tipicamente, a retina dos seres humanos é sensível a uma faixa delimitada da radiação eletromagnética compreendida entre os comprimentos de onda de 400 e 700 nanômetros, correspondentes aos limites das regiões ultravioleta e infravermelha, respectivamente. Contudo, não é igualmente sensível a todos os comprimentos de onda do espectro visível.

A Figura 2.2 mostra a dependência espectral da sensibilidade do olho humano à radiação eletromagnética. Como pode ser observado, a sensibilidade máxima depende do nível de iluminação ambiente, situando-se na faixa de comprimentos de onda entre 510 e 550 nm. O deslocamento na sensibilidade ocorre por que os cones e bastonetes são responsáveis pela resposta à luz. Em ambientes claros, os cones, responsáveis pela percepção de cor, são prioritários na percepção da luz. Nessa situação, a sensibilidade é maior em 555 nm. Em ambientes escuros, a curva representativa da sensibilidade é mostrada à esquerda. Nesse caso, os bastonetes são mais ativos e o olho é mais sensível à luz e possui menor habilidade da distinção de cores. Como a radiação infravermelha

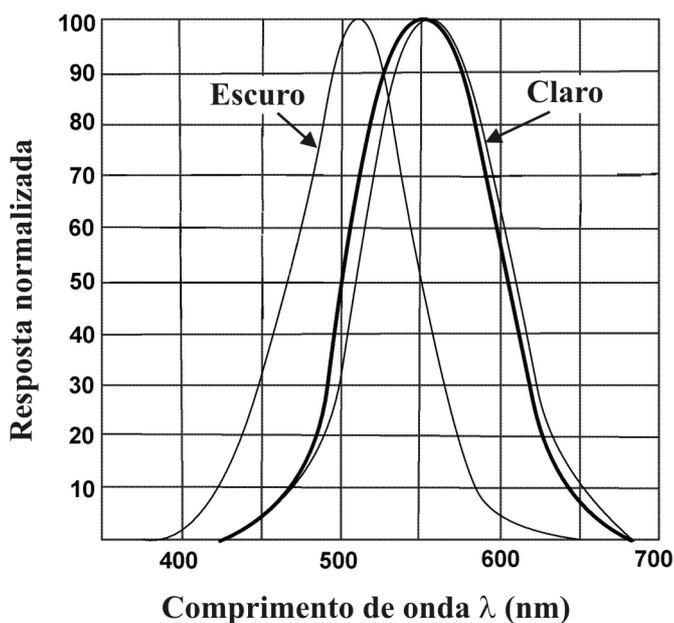


Figura 2.2 – Espectro de absorção da retina [32].

sensibiliza apenas fracamente as células fotorreceptoras da retina, tal radiação é invisível ao olho humano. Por isso, deve-se tomar cuidado com a exposição da retina a altos níveis de intensidade luminosa nesses comprimentos de onda, pois, por serem invisíveis, não ativam o fechamento involuntário da pupila através do músculo ciliar.

Apesar de a retina ser sensível apenas a luz visível, outros tecidos constituintes do globo ocular absorvem radiações fora dessa região, o que influencia nos limites seguros de exposição, como será discutido adiante, na Seção 2.2. A Figura 2.3 mostra os espectros de absorção da córnea, do cristalino (lente), do humor vítreo e do aquoso nas regiões do ultravioleta próximo (0.3 a 0.4 μm) e parte do infravermelho médio (0.8 a 2.4 μm).

Córnea e Esclera

Em função de sua curvatura, a córnea é o mais importante elemento refrativo do

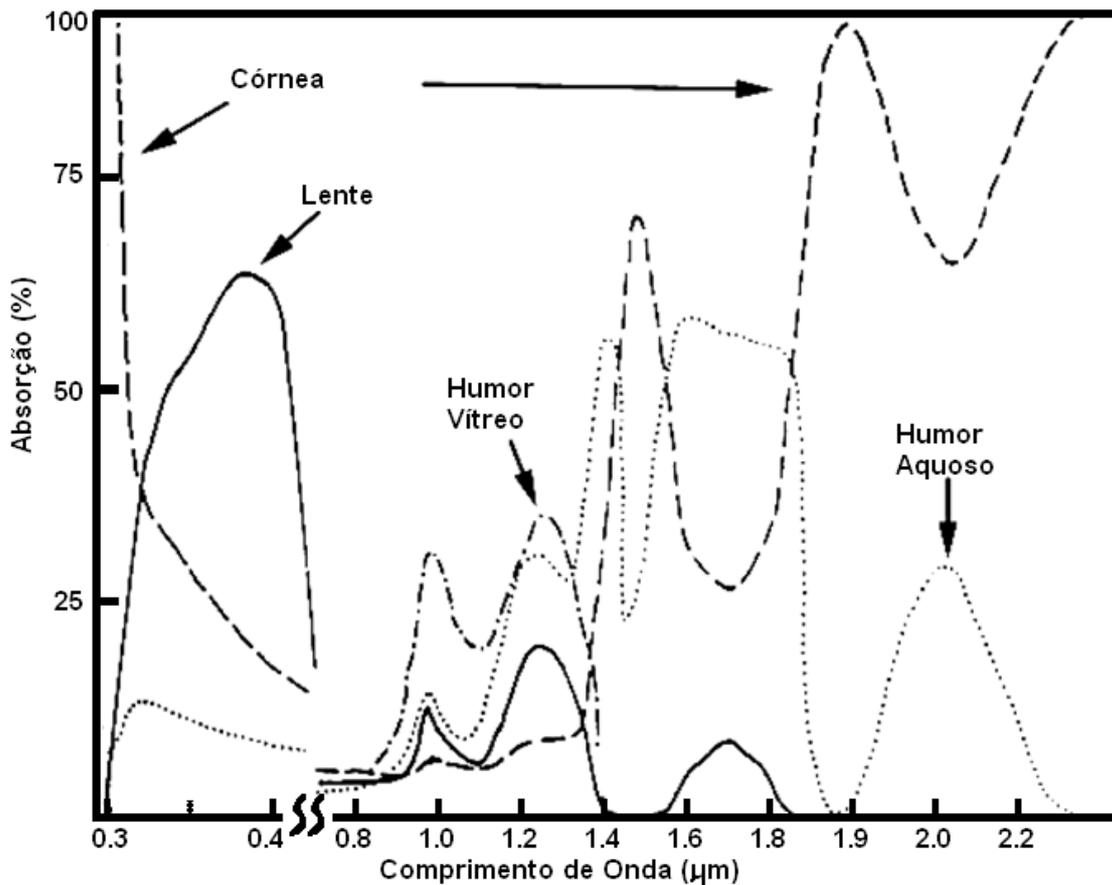


Figura 2.3 – Absorção espectral do meio ocular humano na região do ultravioleta próximo e infravermelho médio.

olho. A córnea é capaz de transmitir luz com comprimento de onda desde 300 nm, no limite do ultravioleta, até 2000 nm, já no infravermelho [33].

Cristalino

O cristalino tem a forma aproximada de uma lente biconvexa, com maior curvatura em sua face posterior. Sua forma pode ser modificada pela contração das fibras musculares ciliadas radiais, o que provoca o relaxamento de sua cápsula, permitindo maior convexidade de sua face anterior, sendo isso parte do processo de acomodação para a visão de perto. De forma inversa, a contração do músculo ciliar circular aumenta a tensão da cápsula, diminuindo sua convexidade, o que é igualmente importante para a visão de longe. Esses efeitos ocorrem, predominantemente, na face anterior do cristalino [29].

2.2. Fontes de iluminação e limites de exposição do olho humano

Como descrito no Capítulo 1, alguns dispositivos de rastreamento ocular utilizam iluminação infravermelha. Essa escolha permite filtrar a informação obtida da iluminação do objeto com luz infravermelha daquela emitida pelo ambiente, tipicamente ocorrendo no espectro visível. Geralmente são utilizadas fontes de luz emitindo na região do infravermelho próximo (de 0.7 a 1.1 μm), em face da disponibilidade em larga escala e a baixo custo de diodos emissores de luz comerciais, operando nos comprimentos de onda de 0.85 μm e 0.95 μm . Uma vez que a luz emitida por essas fontes é imperceptível aos seres humanos, como pode ser inferido da Figura 2.2, isso prejudica o controle involuntário da entrada de luz no olho, feito pela variação no diâmetro da pupila, podendo levar a danos nos tecidos oculares.

Efeitos nocivos da radiação no globo ocular

A exposição à radiação óptica não-ionizante pode causar queimaduras na córnea e/ou na retina, dependendo do comprimento de onda da radiação. Opacidades na córnea e/ou no cristalino podem ocorrer por exposição crônica a níveis elevados [33].

O potencial de lesão ocular geralmente depende de qual estrutura do olho absorve mais energia radiante por volume de tecido. Os efeitos na retina são possíveis quando os comprimentos de onda da radiação ocorrem nas regiões espectrais do visível e do infravermelho próximo (0.4 a $1.4\mu\text{m}$). A luz proveniente de um laser ou reflexão especular nessa região espectral quando entra no olho, pode ser focalizada numa imagem extremamente pequena na retina. A densidade de potência incidente na córnea pode ser amplificada por até 100.000 vezes na retina devido à focalização provocada pela córnea e pelo cristalino.

Emissões nas regiões ultravioleta e infravermelho distante produzem efeitos oculares prioritariamente na córnea. Entretanto, radiações em certos comprimentos de onda nessas regiões podem danificar o cristalino. A radiação UVB, nos comprimentos de onda de 180 nm a 315 nm é absorvida na córnea e pode causar fotoqueratite, que é um efeito que ocorre após exposição à radiação ultravioleta intensa, ocorrendo relativamente ao sol apenas em ambientes altamente refletivos como a neve ou praia. A soldadura a arco também pode provocar fotoqueratite. O epitélio da córnea absorve prioritariamente a radiação UV-B, deixando danos visíveis clinicamente e que provocam olhos vermelhos, sensação de corpo estranho, lacrimejo excessivo e maior sensibilidade à luz. A radiação no ultravioleta próximo (UVA) entre 315 nm e 400 nm é absorvida na lente e pode resultar em catarata.

Radiações no visível (400 nm a 780 nm) e na região do infravermelho próximo (780 nm a 1400 nm) são transmitidas através do meio ocular com pouca perda de intensidade e são focalizadas na retina num local de $10\mu\text{m}$ a $20\mu\text{m}$ de diâmetro. Essa focalização pode causar intensidades altas o suficiente e provocar danos. Por essa razão, a faixa de 400 nm a 1400 nm é denominada região de perigo para a retina. Comprimentos de onda visíveis entre 400 nm e 550 nm são particularmente danosos à retina para exposições longas, durando minutos ou horas.

Limites de exposição ocular

Do ponto de vista de limites de exposição segura à radiação, os LEDs têm sido tratados como lasers [34] e como lâmpadas [35]. Os lasers de diodo são construídos com

cavidades ressonantes com ganho óptico (como estruturas Fabry-Perot) onde ocorre emissão estimulada, produzem uma largura de banda bastante estreita, alcançam durações de pulso mais curtos, e podem emitir potências radiantes mais intensas do que os LEDs [33].

Segundo o Instituto Americano de Padrões Nacionais (ANSI), a Máxima Exposição Permitida (MPE) às radiações laser, depende do tempo de exposição e do comprimento de onda [36]. Na Figura 2.4 são apresentados alguns valores típicos de MPE. Os respectivos valores para exposição da pele são usualmente maiores, já que a pele não é tão sensível como a retina.

Para pulsos mais curtos do que 1 ns, o limite de dano na escala de densidade de energia depende da raiz quadrada da duração do pulso, aproximadamente [33]. Por exemplo, para calcular o limite de exposição para pulsos de laser com duração de 10 ps, as densidades de energia listadas na Figura 2.4 para pulsos de 1 ns, devem ser multiplicadas por um fator de $1/\sqrt{100} = 1/10$.

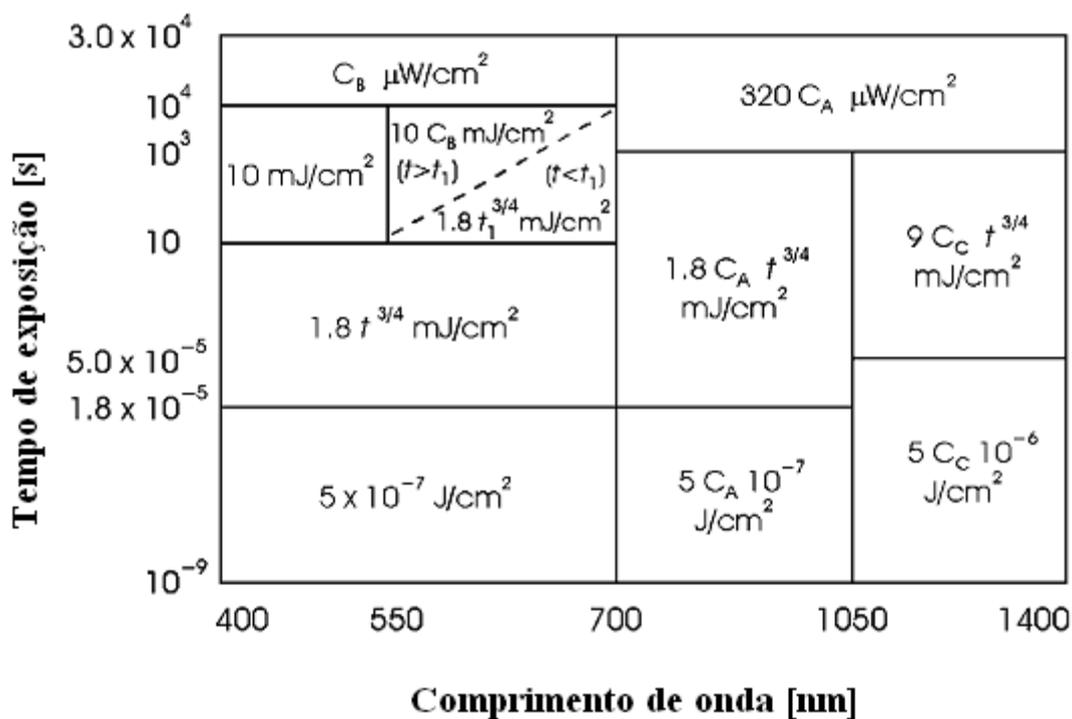


Figura 2.4 – Valores de MPE para exposição ocular direta à radiação visível e no infravermelho próximo [33].

Na Figura 2.4, os fatores de correção C_A , C_B , e C_C dependem do comprimento de onda. A Tabela 2.2 mostra como esses fatores devem ser calculados, de acordo com a região do espectro da radiação laser em questão.

Para o tipo de LED utilizado neste trabalho, operando no comprimento de onda $\lambda = 0.85\mu m$, pela Figura 2.4, a MPE é dada por

$$MPE = 320 \cdot C_A \mu W/cm^2, \quad (2.1)$$

onde

$$C_A = 10^{2(\lambda-0.7)}. \quad (2.2)$$

Assim,

$$MPE = 320 \times 10^{2(\lambda-0.7)} \mu W/cm^2, \quad (2.3)$$

que fornece uma densidade de potência laser máxima de $630mW/cm^2$. Considerando esse limite de exposição, todas as correntes dos LEDs utilizados nos projetos desenvolvidos neste trabalho foram dimensionadas de modo a não ultrapassar o limite adotado de $310mW/cm^2$ (metade do MPE). As potências emitidas por esses LEDs foram sempre medidas com um medidor de potência óptica, como precaução.

Tabela 2.2 – Fatores de correção C dependentes do comprimento de onda [33].

Fator	Comprimento de onda (μm)
$t_1 = 10 \times 10^{20(\lambda-0.550)}$	0.550 – 0.700
$C_A = 10^{2(\lambda-0.700)}$	0.700 – 1.050
$C_A = 5$	1.050 – 1.400
$C_B = 1$	0.400 – 0.550
$C_B = 10^{15(\lambda-0.550)}$	0.550 – 0.700
$C_C = 1$	1.050 – 1.150
$C_C = 10^{18(\lambda-1.150)}$	1.150 – 1.200
$C_C = 8$	1.200 – 1.400

3. SENSORES DE IMAGEM E MODELAMENTO MATEMÁTICO DE CÂMERAS DIGITAIS

*“O homem muda a tecnologia;
a tecnologia muda o homem”.*

Fernando Pessoa.

Imagens digitais podem ser obtidas de diversas maneiras e em diversas aplicações. Genericamente, essas imagens podem ser obtidas pela combinação de uma fonte de “iluminação” com a reflexão ou absorção da energia dessa fonte pelos elementos da “cena” sendo observada. É importante notar que os conceitos “iluminação” e “cena” são mais gerais do que a situação comum em que uma fonte de luz visível ilumina uma cena tridimensional. Por exemplo, a iluminação pode ser originária de uma fonte de energia eletromagnética em frequências de microondas, na região infravermelha ou mesmo com emissores de raios-X, ou mesmo de fontes menos tradicionais como ultrassom, por exemplo.

A Figura 3.1 mostra o componente básico de um sensor de imagens. A energia incidente no componente é selecionada através do filtro e transformada em tensão pela combinação da potência elétrica de entrada com o sinal gerado pelo material que compõe o sensor, responsável pela captação da energia específica sendo detectada. A forma de onda da tensão de saída é a resposta do sensor e é digitalizada para formar uma parte da imagem. O mais comum sensor desse tipo é o fotodiodo de silício, cuja tensão de saída é proporcional à intensidade da luz incidente.

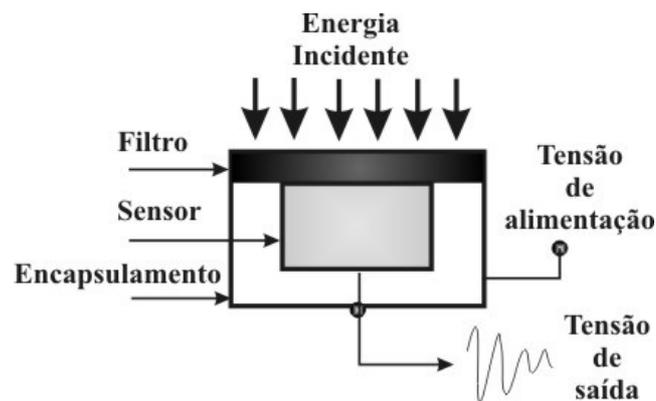


Figura 3.1 – Componente básico de um sensor de imagem.

3.1. Tecnologias CCD e CMOS

Os sensores de imagens que operam na região espectral do visível e infravermelho próximo são conjuntos bidimensionais de fotodiodos fabricados em um único substrato semicondutor e usam tecnologia CCD ou CMOS, e são largamente utilizados na fabricação de câmeras digitais. Ambos os sensores são semicondutores de metal-óxido (MOS). Cada fotodiodo (ou cada pixel), quando exposto à iluminação, acumula carga proporcionalmente à intensidade luminosa local, proporcionando a obtenção da imagem através de uma função de amostragem espacial.

Quando a exposição é concluída (aquisição de um quadro, ou *frame*, quando se trata de uma aplicação em tempo real), um sensor do tipo CCD, como mostrado na Figura 3.2, transfere cada pacote de carga de forma sequencial para uma estrutura comum de saída, que converte cada valor de carga em um valor correspondente de tensão, que é, posteriormente, amplificado e binarizado. Em um sensor CMOS, a conversão carga-tensão ocorre em cada pixel. Essas diferenças nas técnicas de leitura e captação têm implicações significativas para a capacidade, limitações e arquitetura dos sensores.

Alguns dos parâmetros utilizados para caracterizar diferentes tipos de sensores e fotodiodos estão sumarizados a seguir [37].

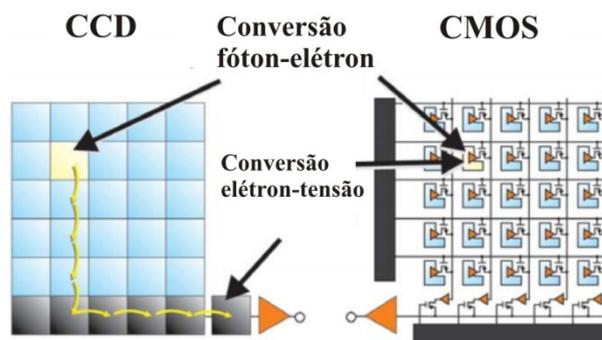


Figura 3.2 – Arquitetura CCD e CMOS [37].

Responsividade

A corrente gerada por unidade de potência óptica incidente em A/W define a responsividade de um fotodiodo. Esse parâmetro varia com o comprimento de onda. A dependência espectral da responsividade de fotodiodos de silício para diferentes temperaturas está ilustrada na Figura 3.3, em que se pode observar que a responsividade máxima ocorre para comprimentos de onda em torno de $1\ \mu\text{m}$.

Sobre esse aspecto, o sensor CMOS é ligeiramente superior ao CCD, em geral, devido à facilidade de integração dos elementos de ganho. No CMOS, os transistores complementares possibilitam amplificação com alto ganho, enquanto a amplificação no CCD é limitada por sua arquitetura.

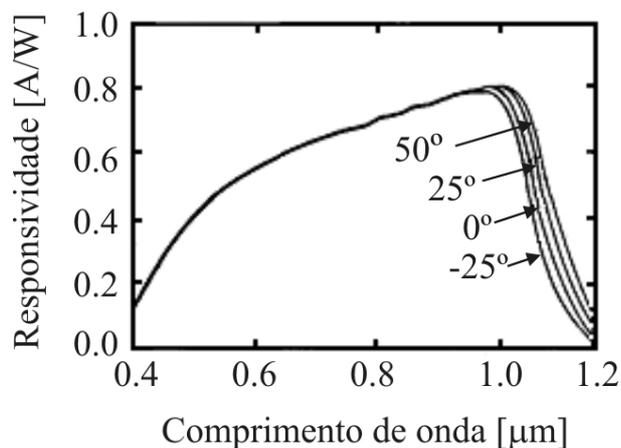


Figura 3.3 – Responsividade espectral do fotodiodo de silício [38].

Faixa dinâmica

A faixa dinâmica de uma grandeza física representa a razão entre o valor dessa grandeza que satura o dispositivo de medida e o valor mínimo dessa grandeza que representa o limite inferior de detecção do dispositivo. Em circunstâncias comparáveis, a faixa dinâmica de sensores CCD é duas vezes maior do que aquela de dispositivos CMOS. Além disso, os sensores CCD têm ainda desempenho melhor do que aquele do dispositivo CMOS devido a um número inferior de circuitos no substrato, a uma maior tolerância às variações na capacitância inerente às trilhas, etc.

Tempo de resposta

Sensores de imagem com tecnologia CMOS exibem tempos de resposta inferiores àqueles de dispositivos CCD, pois todas as funções da câmera em um dispositivo CMOS podem ser integradas no sensor de imagem. Assim, as distâncias percorridas pelos sinais e potências são menores, resultando em menor indutância, capacitância e atrasos.

Janelamento

O janelamento é uma característica exclusiva da tecnologia CMOS e permite a leitura de apenas uma porção da região ativa do sensor. Isso permite a análise em mais altas taxas de quadros para pequenas regiões de interesse, muito atrativa em certas aplicações como rastreamentos altamente precisos de objetos em tempo real.

Anti-saturação

A anti-saturação é uma característica que, quando incorporada ao dispositivo, permite drenar a exposição localizada excessiva sem comprometer a aquisição no resto do sensor. O CMOS geralmente tem imunidade natural à saturação. O CCD, por outro lado, requer engenharia de prototipação específica para alcançar essa capacidade.

Confiabilidade

As confiabilidades em ambos os sensores são relativamente iguais na maioria das aplicações comerciais e industriais. Em situações adversas (como movimentação brusca,

alta temperatura e pressão), os sensores CMOS têm a vantagem de possuir todas as funções de circuito integradas num único chip, minimizando condutores e soldas, causas comuns de falhas em circuitos submetidos a esses ambientes.

Uniformidade de resposta

A uniformidade de resposta é aquela que seria obtida em pixels distintos de um mesmo *chip*, submetido a uma iluminação uniforme. Essa característica do sensor CMOS é comprometida devido à amplificação se dar em cada pixel (o que implica em diferentes ganhos), fazendo com que os níveis de sinais de saída sejam ligeiramente diferentes para um mesmo nível de iluminação.

Em resumo, os sensores CMOS oferecem melhor capacidade de integração, maior taxa de aquisição, menores dissipação de potência e tamanho do sistema ao custo da flexibilidade e qualidade da imagem (particularmente em ambientes com baixa iluminação). Eles são mais atrativos para aplicações onde os requisitos de qualidade de imagem são baixos, como câmeras de segurança, videoconferência, máquinas de fax, *scanners* comerciais, brinquedos, e algumas aplicações em biometria e sensores automotivos. Já os sensores CCD oferecem maior qualidade de imagem (menor ruído e maior sensibilidade em ambientes com baixa iluminação) e flexibilidade a custo da dissipação de potência, tamanho do sistema, capacidade de integração e tempo de aquisição. Eles representam uma melhor opção para aplicações industriais, médicas e científicas que requerem alta qualidade de imagem, em fotografia digital profissional e propagação de sinais de TV. Os custos dos dois sensores são equivalentes.

Neste trabalho, foram utilizadas duas câmeras USB comerciais simples, com sensores CMOS, uma VIMICRO ZCO301 PLH, uma TECLASER PC CAMERA e uma mini câmera analógica TECVOZ com sensor CCD. As câmeras digitais trabalham a uma taxa de 30 quadros por segundo (fps) com uma resolução de 640 por 480, enquanto a câmera analógica trabalha com uma resolução de 420 linhas com taxa de aquisição dependendo da placa de captura utilizada.

Como descrito nos Capítulos 6 e 7, em diferentes implementações, as imagens são adquiridas a altas taxas em ambiente com boa iluminação infravermelha, em especial no

projeto onde a câmera fica fixada aos óculos de segurança, captando imagens próximas ao olho. Nesse caso, o sensor CMOS foi escolhido pelos seus atributos supracitados. Numa outra abordagem, a câmera fica presa ao monitor, a certa distância do olho, implicando em aquisição de imagem com baixa iluminação, já que um filtro óptico de bloqueio da radiação visível foi utilizado para melhorar a imunidade do sistema às variações da intensidade de iluminação ambiente. Neste caso, um sensor CCD foi utilizado devido ao seu menor ruído e à sua maior sensibilidade a baixas variações de iluminação.

3.2. Modelo matemático da câmera

Um ponto no espaço pode ser especificado por três coordenadas (x, y, z) . Uma imagem gerada em um sensor de imagens é planar e requer duas coordenadas (u, v) para representar um ponto. Assim, uma dimensão é perdida no processo de projeção. Um dos importantes objetivos em Visão Computacional é recuperar essa dimensão perdida.

A principal abordagem para solucionar esse problema é determinar um modelo matemático de câmera a ser utilizado que possibilite relacionar as coordenadas tridimensionais da cena com as coordenadas bidimensionais correspondentes da imagem [39]. Isso é necessário para se obter um rastreamento ocular satisfatório, através da metodologia tratada no Capítulo 5. Com o emprego de um processo de calibração, pode-se obter a matriz da câmera \tilde{M} (ou matriz de projeção) que relaciona essas coordenadas de cena com as coordenadas de imagem.

Modelo da câmera de orifício

O modelo geométrico mais comum de uma câmera é o modelo da câmera de orifício [40]. A Figura 3.4 ilustra a geometria utilizada na construção desse modelo, consistindo em um plano imagem π (área ativa do sensor de imagem), e de um ponto O' , representando o centro ou foco de projeção. Esse foco de projeção é definido por um sistema de lentes de entrada no sensor de imagem. A distância entre π e O' é o comprimento focal f . A reta que passa por O' e é perpendicular a π é o eixo visual z' . O ponto C , representando a interseção do eixo óptico com o plano π é chamado de ponto principal (também chamado de ponto nodal da câmera) ou centro da imagem. Como mostrado na Figura 3.4, I é a imagem ou projeção do ponto P no plano uv . Considerando o

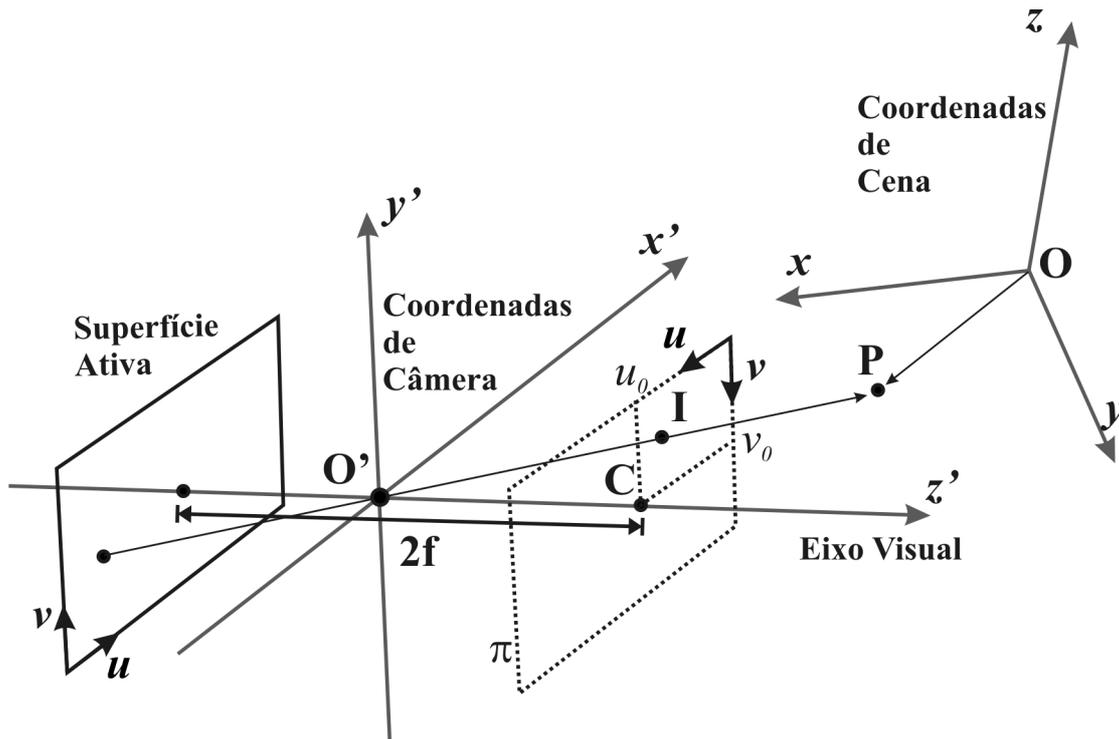


Figura 3.4 – Modelo da câmera de orifício com projeção em perspectiva.

sistema de coordenadas de referência da câmera cuja origem é o foco de projeção O' , o eixo z' coincide com o eixo óptico, o plano $x'y'$ é paralelo ao plano uv , os pontos em questão são dados por $P = (x', y', z')$ e $I = (x'_I, y'_I, z'_I)$ e têm a seguinte interdependência:

$$x'_I = f \frac{x'}{z'}, \quad (3.1a)$$

$$y'_I = f \frac{y'}{z'}, \quad (3.1b)$$

$$z'_I = f. \quad (3.1c)$$

As coordenadas de I no plano imagem (u, v) são dadas no sistema de coordenadas da câmera pelas relações

$$x'_I = -s_u (u - u_0), \quad (3.2a)$$

$$y'_I = -s_v (v - v_0). \quad (3.2b)$$

Nessa equação u_0 e v_0 são as coordenadas do centro da imagem C , deslocado do centro da superfície sensora da câmera por, geralmente, alguns pixels devido a incertezas de posicionamento do sistema de lentes, que ocorrem no estágio de fabricação da câmera. Os parâmetros s_u e s_v representam os comprimentos de cada pixel nas direções horizontal e vertical, respectivamente. A razão s_u/s_v , fornece informação sobre a distorção imposta pela forma ligeiramente retangular dos sensores constituintes do CCD. Na prática essa razão é unitária para a maioria das câmeras comerciais. Os valores de f , u_0 , v_0 , s_u e s_v , são os cinco parâmetros intrínsecos do modelo de câmera em perspectiva.

A relação entre o vetor coluna que representa as coordenadas do ponto P no sistema de referência da cena $\tilde{X} = (x \ y \ z)^T$ e o vetor coluna das coordenadas desse ponto no sistema de referência da câmera $\tilde{X}' = (x' \ y' \ z')^T$, define os parâmetros extrínsecos desta.³ Essa relação pode ser representada de forma geral por uma rotação seguida de uma translação do sistema de referência da cena, na forma

$$\tilde{X}' = \tilde{T}\tilde{X} + \tilde{X}_0, \quad (3.3)$$

com

$$\tilde{T} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \quad (3.4)$$

representando uma matriz unitária de rotação [40] e

$$\tilde{X}_0 = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}, \quad (3.5)$$

a matriz coluna das coordenadas de O no sistema $x'y'z'$. É importante observar que há seis relações independentes entre os nove elementos da matriz de rotação \tilde{T} e esta pode,

³ Nesta dissertação grandezas matriciais são denotadas com o símbolo (\sim) sobreposto à representação literal da grandeza. O sobrescrito T nessa representação indica a transposta da matriz.

portanto, ser especificada em termos de três parâmetros ou, equivalentemente, três ângulos de rotação [40].

Substituindo as equações (3.1) e (3.2) na equação (3.3) obtém-se

$$u = u_o - \frac{f}{s_u} \cdot \frac{\tilde{T}_1 \tilde{X} + x_0}{\tilde{T}_3 \tilde{X} + z_0}, \quad (3.6a)$$

$$v = v_o - \frac{f}{s_v} \cdot \frac{\tilde{T}_2 \tilde{X} + y_0}{\tilde{T}_3 \tilde{X} + z_0}, \quad (3.6b)$$

com

$$\tilde{T}_i \equiv (t_{i1} \quad t_{i2} \quad t_{i3}), \quad i = 1, 2, 3. \quad (3.7)$$

Definindo-se

$$w = \tilde{T}_3 \tilde{X} + z_0 \quad (3.8)$$

as relações (3.6) e (3.8) podem ser postas na forma matricial

$$\tilde{W} = \tilde{M}_{int} \tilde{M}_{ext} \begin{pmatrix} \tilde{X} \\ 1 \end{pmatrix}, \quad (3.9)$$

com

$$\tilde{W} \equiv \begin{pmatrix} uw \\ vw \\ w \end{pmatrix} \quad (3.10)$$

e

$$\tilde{M}_{int} \equiv \begin{pmatrix} -\frac{f}{s_u} & 0 & u_0 \\ 0 & -\frac{f}{s_v} & v_0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (3.11)$$

$$\tilde{M}_{ext} \equiv (\tilde{T} \quad \tilde{X}_0), \quad (3.12)$$

representando matrizes que dependem dos parâmetros intrínsecos e extrínsecos da câmera, respectivamente.

Distorção não-linear

Em muitos casos, o sistema óptico formado pelas lentes introduz distorções na imagem que se tornam evidentes nas periferias desta, ou até em outros lugares quando se usa óptica com um campo de visão largo. Lentes podem inserir distorções radiais e tangenciais. Estas devido ao deslocamento do sistema óptico em relação ao centro do sensor de imagens.

A distorção radial pode ser modelada segundo as relações

$$\tilde{S} = \tilde{S}_d (1 + k_1 r^2 + k_2 r^4 + \dots), \quad (3.13)$$

com

$$\tilde{S}_d = \begin{pmatrix} u_d \\ v_d \end{pmatrix} \quad (3.14)$$

e

$$r^2 = u_d^2 + v_d^2 = \tilde{S}_d^T \tilde{S}_d \quad (3.15)$$

Como essas relações indicam, nesse modelo, a distorção é um desvio radial e isotrópico dos pontos na imagem. Os coeficientes de distorção k_1 e k_2 são parâmetros intrínsecos usualmente nulos e, portanto, ignorados quando uma alta exatidão nas periferias da imagem não é prioridade. A magnitude da distorção radial depende da qualidade das lentes utilizadas e, segundo Trucco e Verri [40], para sensores de imagem

CCD de qualidade média com 500 por 500 pixels, uma alta distorção de aproximadamente 5 pixels pode ser esperada nas periferias da imagem. Nessas circunstâncias, um modelo com $k_2 = 0$ ainda é preciso.

O método de obtenção direta dos parâmetros da câmera, tratado na Seção 3.3, não leva em consideração os coeficientes de distorção. Esse método fornece resultados insatisfatórios para os parâmetros da câmera (apresentados no Capítulo 7) e, portanto, outro modelo de câmera, utilizado no processo não-linear de obtenção dos parâmetros é apresentado a seguir.

Modelo de distorção afim

Outro modelo de câmera é o de distorção afim, que leva em consideração os coeficientes de distorção radial e tangencial como parâmetros intrínsecos da câmera, além daqueles citados anteriormente, e o ângulo θ entre os eixos u e v do sensor de imagem [41]. Em geral $\theta \approx \pi / 2$.

Seja o ponto P no espaço, de coordenadas $\tilde{X}' = (x' \ y' \ z')^T$ no sistema da câmera, como mostrado na Figura 3.4. Em consonância com a notação adotada anteriormente, a projeção desse ponto no plano imagem π é dada por

$$\tilde{r}_I = \begin{pmatrix} x_I \\ y_I \end{pmatrix} = f \begin{pmatrix} x'/z' \\ y'/z' \end{pmatrix}. \quad (3.16)$$

Considerando a coordenada radial obtida de

$$r_I^2 = \tilde{r}_I^T \tilde{r}_I, \quad (3.17)$$

o efeito de distorção radial e tangencial do sistema de lentes da câmera, é levado em conta com a definição da matriz coluna de 5 elementos

$$\tilde{q} = (q_1 \quad q_2 \quad q_3 \quad q_4 \quad q_5)^T, \quad (3.18)$$

de forma que as novas coordenadas de P na imagem, podem ser escritas na forma matricial

$$\tilde{r}_d = \left(1 + q_1 r_l^2 + q_2 r_l^4 + q_3 r_l^6\right) \tilde{r}_l + \tilde{D}, \quad (3.19)$$

onde \tilde{D} é o vetor de distorção tangencial dado por

$$\tilde{D} = \begin{pmatrix} 2q_3 x_l y_l + q_4 (r_l^2 + 2x_l^2) \\ q_3 (r_l^2 + 2y_l^2) + 2q_4 x_l y_l \end{pmatrix}. \quad (3.20)$$

O efeito de distorção devido a um ângulo $\theta \neq \pi/2$ entre os eixos u e v do sensor de imagem é levado em conta, em primeira ordem, com a inclusão da projeção do eixo vertical no eixo horizontal. O efeito da projeção do eixo horizontal no eixo vertical é um efeito de 2ª ordem e pode ser desprezado. Com a inclusão dessa distorção as coordenadas da imagem, expressas em unidades de coordenadas do sensor de imagem podem ser expressas na forma matricial

$$\tilde{S}_d = \tilde{K} \tilde{r}_l, \quad (3.21)$$

com a matriz coluna de coordenadas distorcidas da câmera, \tilde{S}_d , já definida em (3.13) e com

$$\tilde{K} = \begin{pmatrix} -f/s_u & -f \cos \theta / s_u & u_0 \\ 0 & -f/s_v & v_0 \end{pmatrix}. \quad (3.22)$$

3.3. Calibração das câmeras

Os métodos de calibração de câmeras consistem inicialmente na obtenção da correlação entre pontos de coordenadas de cena conhecidas e suas respectivas coordenadas no plano imagem [40]. Para tanto, utiliza-se um padrão, que pode ser coplanar ou não, de

dimensões conhecidas como mostrado na Figura 3.5. Utilizando processamento de imagens, é possível obter as coordenadas do plano imagem correspondentes a cada ponto de coordenadas cena (a partir da construção do padrão). Para padrões não coplanares, apenas uma imagem é suficiente para o processamento e consequente obtenção dos parâmetros da câmera, enquanto para padrões coplanares, mais de uma imagem é necessária [42].

Método direto

Com base no padrão da Figura 3.5, considerando o plano do padrão como sendo o plano- xy , perpendicular ao eixo z , formando o sistema de coordenadas de cena, em que o vértice de cada quadrado seja o ponto $\tilde{X}_i = (x_i \ y_i \ z_i)^T$, obtêm-se as coordenadas imagem $\tilde{S}_i = (u_i \ v_i)^T$ de forma que

$$\tilde{W}_i = \tilde{M} \begin{pmatrix} \tilde{X}_i \\ 1 \end{pmatrix}, \quad (3.23)$$

com \tilde{W}_i sendo definido em consonância com (3.10), i.e.,

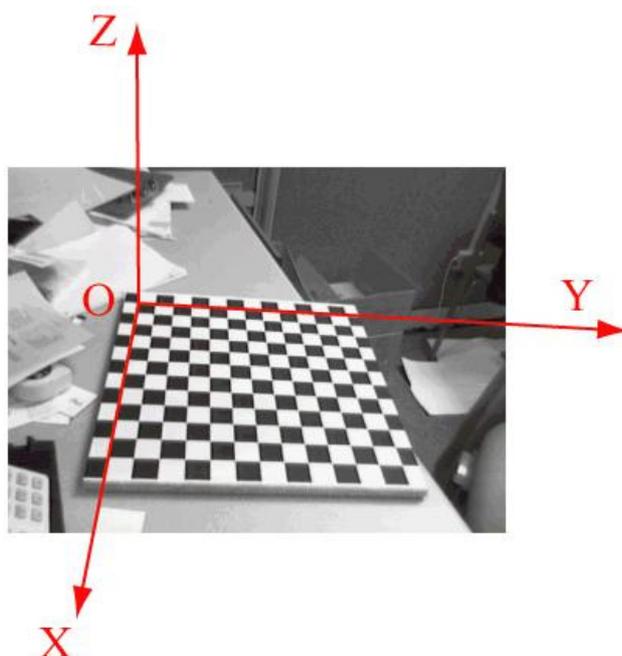


Figura 3.5 – Padrão típico utilizado na calibração da câmera [41].

$$\tilde{W}_i = \begin{pmatrix} u_i w_i \\ v_i w_i \\ w_i \end{pmatrix} \quad (3.24)$$

e

$$\tilde{M} = \tilde{M}_{int} \tilde{M}_{ext} \quad (3.25)$$

representa a matriz de projeção da câmera com \tilde{M}_{int} e \tilde{M}_{ext} dados por (3.11) e (3.12), respectivamente.

A matriz \tilde{M} pode ser escrita na forma

$$\tilde{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \quad (3.26)$$

e as duas primeiras linhas de (3.21) fornecem o par de equações homogêneas

$$m_{11}x_i + m_{12}y_i + m_{13}z_i + m_{14} - m_{31}u_i x_i - m_{32}u_i y_i - m_{33}u_i z_i - m_{34}u_i = 0, \quad (3.27a)$$

$$m_{21}x_i + m_{22}y_i + m_{23}z_i + m_{24} - m_{31}v_i x_i - m_{32}v_i y_i - m_{33}v_i z_i - m_{34}v_i = 0. \quad (3.27b)$$

Usando padrões de calibração como mostrado na Figura 3.5, para N pontos, $2N$ equações da forma (3.27) são obtidas e o sistema pode ser escrito na forma

$$\tilde{A}\tilde{m} = 0, \quad (3.28)$$

com

$$\tilde{A} = \begin{pmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -u_1 x_1 & -u_1 y_1 & -u_1 z_1 & -u_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -v_1 x_1 & -v_1 y_1 & -v_1 z_1 & -v_1 \\ x_2 & y_2 & z_2 & 1 & 0 & 0 & 0 & 0 & -u_2 x_2 & -u_2 y_2 & -u_2 z_2 & -u_2 \\ 0 & 0 & 0 & 0 & x_2 & y_2 & z_2 & 1 & -v_2 x_2 & -v_2 y_2 & -v_2 z_2 & -v_2 \\ \vdots & \vdots \\ x_N & y_N & z_N & 1 & 0 & 0 & 0 & 0 & -u_N x_N & -u_N y_N & -u_N z_N & -u_N \\ 0 & 0 & 0 & 0 & x_N & y_N & z_N & 1 & -v_N x_N & -v_N y_N & -v_N z_N & -v_N \end{pmatrix} \quad (3.29)$$

e

$$\tilde{m} = (m_{11} \ m_{12} \ m_{13} \ m_{14} \ m_{21} \ m_{22} \ m_{23} \ m_{24} \ m_{31} \ m_{32} \ m_{33} \ m_{34})^T. \quad (3.30)$$

A solução da equação (3.28) pode ser obtida expressando-se uma variável em função das onze restantes e aplicando-se, por exemplo, o método dos mínimos quadráticos ou da decomposição singular de matrizes [40]. Com isso, a solução para os doze parâmetros da matriz \tilde{M} é obtida a menos de um fator de escala. Uma vez que, a matriz \tilde{M} depende de fato de apenas 11 parâmetros, i.e., 5 parâmetros intrínsecos e 6 extrínsecos (3 ângulos de rotação e 3 coordenadas de translação), esse fator de escala pode ser obtido levando-se em conta as propriedades da matriz de rotação, conforme descrito a seguir

Obtenção dos parâmetros da câmera a partir da matriz \tilde{M}

Podem-se explicitar os parâmetros da câmera na matriz \tilde{M} , com base em (3.9)-(3.12), o que fornece

$$\tilde{M} = \begin{pmatrix} -f_u t_{11} + u_0 t_{31} & -f_u t_{12} + u_0 t_{32} & -f_u t_{13} + u_0 t_{33} & -f_u x_0 + u_0 z_0 \\ -f_v t_{21} + v_0 t_{31} & -f_v t_{22} + v_0 t_{32} & -f_v t_{23} + v_0 t_{33} & -f_v y_0 + v_0 z_0 \\ t_{31} & t_{32} & t_{33} & z_0 \end{pmatrix}. \quad (3.31)$$

Onde $f_u \equiv f/s_u$ e $f_v \equiv f/s_v$ são as distâncias focais efetivas nas direções horizontal e vertical, respectivamente. Seguindo o formalismo de Trucco e Verri [40], considere a obtenção de uma estimativa \tilde{B} de \tilde{M} , através do método dos mínimos quadráticos. A condição de normalização de um dos elementos de \tilde{M} equivale a definir essa matriz a menos de um fator de escala γ e, portanto,

$$\tilde{B} = \gamma \tilde{M}, \quad (3.32)$$

Seja,

$$\tilde{B} = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \end{pmatrix} \quad (3.33)$$

e as matrizes coluna

$$\tilde{B}_1 = (b_{11} \ b_{12} \ b_{13})^T, \quad (3.34a)$$

$$\tilde{B}_2 = (b_{21} \ b_{22} \ b_{23})^T, \quad (3.34b)$$

$$\tilde{B}_3 = (b_{31} \ b_{32} \ b_{33})^T, \quad (3.34c)$$

$$\tilde{B}_4 = (b_{14} \ b_{24} \ b_{34})^T. \quad (3.34d)$$

Uma vez que a matriz de rotação \tilde{T} é unitária, seus elementos satisfazem às relações de ortogonalidade

$$\sum_{j=1}^3 t_{ij} t_{kj} = \sum_{j=1}^3 t_{ji} t_{jk} = \delta_{ik}, \quad (3.35)$$

onde δ_{ik} representa o delta de kronecker. Uma relação obtida de (3.32)-(3.35) pode ser posta na forma

$$b_{31}^2 + b_{32}^2 + b_{33}^2 = \gamma^2 (t_{31}^2 + t_{32}^2 + t_{33}^2), \quad (3.36)$$

e com o uso de (3.34c) e (3.35) esta última fica expressa na forma

$$\gamma^2 = \tilde{B}_3^T \tilde{B}_3. \quad (3.37)$$

Da última linha de (3.32) tem-se

$$z_0 = \frac{b_{34}}{\gamma} \quad (3.38)$$

e

$$t_{3i} = \frac{b_{3i}}{\gamma}, \quad (3.39)$$

com $i = 1, 2, 3$. Assim, a magnitude do parâmetro γ é determinada de (3.37) e seu sinal é obtido de (3.38), pois o valor estimado de b_{34} tem um sinal definido e uma vez definidas as origens dos sistemas de coordenadas da câmera e da cena, o sinal da coordenada z_0 se torna univocamente determinado.

Tomando os produtos escalares de $\tilde{\mathbf{B}}_3$, com $\tilde{\mathbf{B}}_1$ e $\tilde{\mathbf{B}}_2$, e utilizando-se a propriedade de ortogonalidade (3.35), podem-se obter as coordenadas do centro da imagem u_0 e v_0 , ou seja,

$$u_0 = \frac{\tilde{\mathbf{B}}_1^T \tilde{\mathbf{B}}_3}{\tilde{\mathbf{B}}_3^T \tilde{\mathbf{B}}_3}, \quad (3.40)$$

$$v_0 = \frac{\tilde{\mathbf{B}}_2^T \tilde{\mathbf{B}}_3}{\tilde{\mathbf{B}}_3^T \tilde{\mathbf{B}}_3}. \quad (3.41)$$

Fazendo os produtos escalares $\tilde{\mathbf{B}}_1^T \tilde{\mathbf{B}}_1$ e $\tilde{\mathbf{B}}_2^T \tilde{\mathbf{B}}_2$ com o emprego de (3.37) e as relações de ortogonalidade (3.35), obtém-se

$$f_u = \sqrt{\frac{\tilde{\mathbf{B}}_1^T \tilde{\mathbf{B}}_1}{\tilde{\mathbf{B}}_3^T \tilde{\mathbf{B}}_3} - u_0^2}, \quad (3.42)$$

$$f_v = \sqrt{\frac{\tilde{\mathbf{B}}_2^T \tilde{\mathbf{B}}_2}{\tilde{\mathbf{B}}_3^T \tilde{\mathbf{B}}_3} - v_0^2}. \quad (3.43)$$

Com base em (3.32), e utilizando (3.39), obtém-se, para $i = 1, 2, 3$,

$$t_{1i} = \frac{1}{f_u \gamma} (u_0 b_{3i} - b_{1i}), \quad (3.44)$$

$$t_{2i} = \frac{1}{f_v \gamma} (v_0 b_{3i} - b_{2i}), \quad (3.45)$$

Finalmente, com base em (3.32), e utilizando (3.38), obtém-se, para $i = 1, 2, 3$,

$$x_0 = \frac{1}{f_u \gamma} (u_0 b_{34} - b_{14}), \quad (3.46)$$

$$y_0 = \frac{1}{f_v \gamma} (v_0 b_{34} - b_{24}). \quad (3.47)$$

Esse método possibilita a obtenção dos parâmetros da câmera de orifício de uma forma não iterativa. Entretanto, apesar de rápido, o método direto tem ao menos duas desvantagens. A distorção não-linear imposta pelo sistema óptico não pode ser incorporada e, portanto, os efeitos de distorção não são corrigidos. Por exemplo, Shih *et al.* [43], usaram um método onde a estimativa dos coeficientes de distorção radial é transformada num problema de autovalor. A segunda desvantagem é mais comprometedora. Com o objetivo de obter algoritmos não-iterativos, as restrições nos parâmetros intermediários não são consideradas. Conseqüentemente, na presença de ruído, a solução intermediária não satisfaz essas restrições, e a exatidão da solução final é relativamente pobre [44], como constatado em experimentos cujos resultados são apresentados no 6.4.

Para resolver esses problemas, os parâmetros obtidos através desse método são usados como valores iniciais de um processo iterativo não-linear de minimização de erros proposto por Heikkilä [42]. Jean-Yves Bouquet [41] desenvolveu um conjunto de ferramentas (*Câmera Calibration Toolbox for Matlab*[®]) atualmente disponível na *internet*, que aplica a metodologia proposta por Heikkilä e fornece os parâmetros intrínsecos e extrínsecos da câmera (modelo afim) com poucas iterações e erro desprezível. A matemática formal utilizada por Heikkilä e no *toolbox* está fora do escopo deste trabalho,

que utiliza o processo de calibração das câmeras apenas como ferramenta para tornar possível o rastreamento ocular.

4. ALGUMAS TÉCNICAS DE PROCESSAMENTO DE IMAGENS

“Uma imagem vale mais do que mil palavras”

Autoria desconhecida.

O tão conhecido ditado acima citado sugere a quantidade de informação que pode estar contida em uma imagem. As pinturas e gravuras rupestres, como exemplo, fornecem indícios sobre cultura, crenças, estilo de vida, de civilizações que datam de 4000 a.C., através de imagens de caça, dança e rituais. Atualmente, diversos tipos de fotos tiradas com vários tipos de fotosensores nos permitem saber se um procedimento cirúrgico será recomendado, ou indicar se um dia existiu água em Marte. A visão é o mais aguçado dos sentidos humanos. Entretanto, ao contrário do homem, cuja visão é limitada à região visível do espectro infravermelho, os sensores de imagem atuais cobrem quase todo o espectro eletromagnético, desde raios gama a ondas de rádio, possibilitando a obtenção de informações em regiões do espectro invisíveis ao olho humano.

Os métodos de processamento de imagens aplicam-se, principalmente a duas áreas, a saber, melhoria de informação pictórica para interpretação humana (realce, contraste, etc.) e processamento visando armazenamento, transmissão e interpretação autônoma de máquina [45]. Nas seções seguintes são brevemente revisadas algumas técnicas de processamento de imagens utilizadas no desenvolvimento deste trabalho.

4.1. Limiarização de imagens

Considere uma imagem digital monocromática representada pela função bidimensional $f(u,v)$ com $u = 0,1,2,3,\dots,W - 1$ e $v = 0,1,2,3,\dots,H - 1$. O valor de pixel

$\xi_{uv} = f(u,v)$ pertence à escala de cinza de $L = 256$ valores definida pelo padrão UINT8 (unsigned integer) de oito bits, *i.e.*, o intervalo discreto dos inteiros de 0 (preto) a 255 (branco).

O histograma da imagem f é a função discreta $h(r_k) = n_k$, onde r_k é o k -ésimo nível de cinza e n_k é o número de pixels com nível de cinza r_k . É comum se normalizar o histograma dividindo seus valores pelo número total de pixels da imagem $n = W \times H$. Assim, o histograma normalizado é dado por $p(r_k) = n_k/n$, para $k = 0, 1, \dots, L-1$. Com essa definição, o parâmetro $p(r_k)$ fornece uma estimativa da probabilidade de ocorrência do nível de cinza r_k , *i.e.*, $\sum_{k=0}^{L-1} p(r_k) = 1$ [45]. Para o exemplo da imagem mostrada na Figura

4.1-(a), obtém-se o histograma mostrado na Figura 4.2.

Uma técnica de isolamento de objeto de interesse da imagem é selecionar um limiar de valor de escala de cinza ξ que separe o grupo formado pelos níveis de cinza predominantes do objeto. Dessa forma, qualquer pixel satisfazendo à condição $\xi_{uv} > \bar{\xi}$ é considerado *ponto objeto* e, caso contrário, se trata de um *ponto de fundo*. A imagem binarizada $g(u,v)$ é definida por

$$g(u,v) = \begin{cases} 1 & f(u,v) > \bar{\xi} \\ 0 & f(u,v) \leq \bar{\xi} \end{cases} \quad (4.1)$$

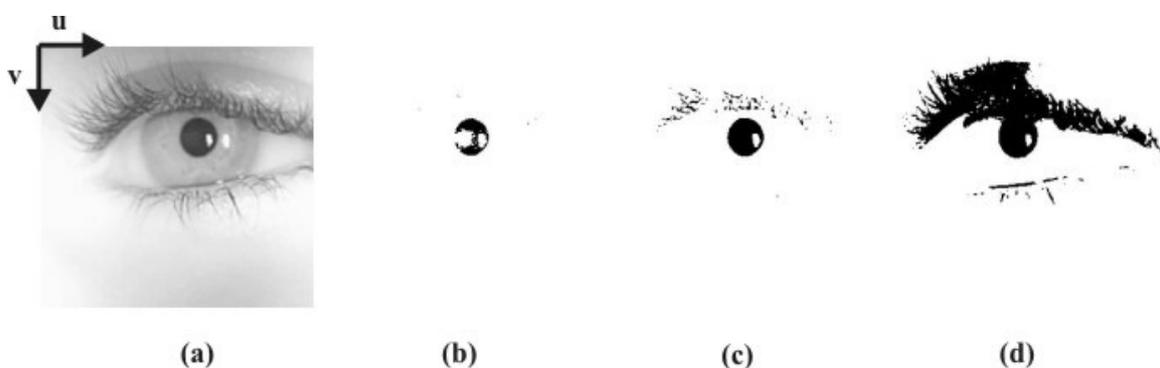


Figura 4.1 – Imagem original (a), e binarizações resultantes com limiares 70 (b), 90 (c) e 150 (d).

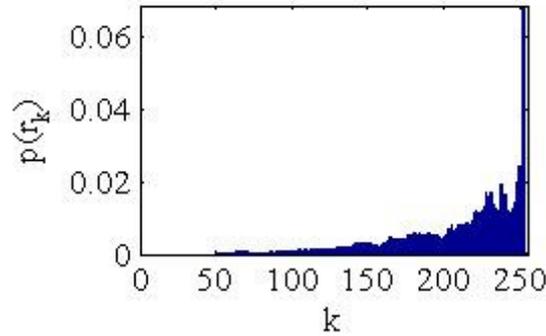


Figura 4.2 – Histograma normalizado $p(r_k)$ da imagem da Figura 4.1a.

Portanto, os pixels mapeados em 1 (ou em algum nível de cinza conveniente) correspondem ao objeto, enquanto os pixels mapeados em 0 (ou em outro nível, diferente, atribuído ao objeto) pertencem ao plano de fundo.

O conceito de limiarização pode ser estendido para dois ou mais objetos de interesse que apresentem seus respectivos níveis de cinza relativamente separados em seus grupos específicos. Portanto, o limiar $\bar{\xi}$ pode ser visto como o resultado de uma função da forma

$$\bar{\xi} = \bar{\xi}(u, v, p(u, v), f(u, v)), \quad (4.2)$$

onde $f(u, v)$ é o nível de cinza do ponto (u, v) e $p(u, v)$ alguma propriedade desse ponto – por exemplo, a média dos níveis de cinza dos pixels vizinhos a esse ponto.

Quando $\bar{\xi}$ depende apenas dos níveis de cinza $f(u, v)$, a limiarização é chamada *global*. Se $\bar{\xi}$ depende de $f(u, v)$ e $p(u, v)$, a limiarização é chamada *local* e é denominada *dinâmica* ou *adaptativa* quando, depende adicionalmente de u e v .

As imagens da Figura 4.1(b), (c) e (d) mostram os resultados da binarização global da imagem f , com os limiares $\bar{\xi}$ iguais a 70, 90 e 150, respectivamente. A intenção nesse processo é determinar até que ponto a pupila pode ser isolada do resto da imagem. A imagem (b) contém prioritariamente pixels pertencentes à pupila, enquanto (c) e (d)

apresentam um ruído significativo, sendo que na última os pixels da pupila se aglomeram aos dos cílios, devido baixa seletividade do limiar.

4.2. Filtragem e detecção de contornos

Outra forma de extrair contornos numa imagem é através do uso de filtros lineares no domínio espacial determinado definido pelas coordenadas (u, v) . Filtragens no domínio da frequência também são comuns [45]. Uma vez que apenas técnicas de filtragem espacial são utilizadas neste trabalho, técnicas no domínio da frequência não são tratadas nesta secção.

A filtragem espacial de uma imagem $f(u, v)$ de tamanho $W \times H$ através do uso uma máscara $w(u, v)$ de tamanho $W' \times H'$, gerando uma imagem filtrada $g(u, v)$, pode ser realizada com o emprego da *convolução* bi-dimensional

$$g(u, v) = \sum_{s=-a}^{s=a} \sum_{t=-b}^{t=b} w(s, t) f(u + s, v + t), \quad (4.3)$$

onde $a = (W' - 1)/2$ e $b = (H' - 1)/2$.

Para a obtenção da imagem filtrada $g(u, v)$, a equação (4.3) deve ser computada para $u = 0, 1, 2, \dots, W - 1$ e $v = 0, 1, 2, \dots, H - 1$. Nesse contexto, as máscaras devem ser de tamanho ímpar mínimo de 3×3 .

É importante considerar o comportamento da filtragem nas bordas externas da imagem. Considerando, sem perda de generalidade, uma máscara de tamanho quadrado ímpar $W' \times W'$, quando o centro desta está a uma distância menor do que $(W' - 1)/2$ de qualquer canto da imagem, alguma linha ou coluna da máscara estará fora dos limites da imagem. Para solucionar este problema, a filtragem pode ser computada de forma a resultar numa imagem de dimensões menores do que a original, ou pode ser calculada na

imagem original expandida através do uso de linhas e colunas externas nulas, ou seguindo algum padrão específico (replacões dos cantos da imagem original, etc.).

A filtragem é muito utilizada para proporcionar uma melhoria na qualidade da imagem. Existem diferentes tipos de filtros aplicáveis tanto no domínio das coordenadas espaciais quanto no domínio de frequências espaciais. Neste trabalho serão tratados alguns tipos mais usados de filtros passa-alta, responsáveis pelo aguçamento da imagem e consequente melhoria na visualização de seus contornos.

Operação gradiente para detecção de contornos

Em forma matricial, o gradiente de uma função $f(u,v)$ é o resultado da atuação do operador matricial

$$\tilde{\nabla} \equiv \left(\frac{\partial}{\partial u} \quad \frac{\partial}{\partial v} \right)^T, \quad (4.4)$$

sobre $f(u,v)$ [45] o que fornece a matriz coluna

$$\tilde{G} = (G_u \quad G_v)^T = \left(\frac{\partial f}{\partial u} \quad \frac{\partial f}{\partial v} \right)^T. \quad (4.5)$$

Para funções de duas variáveis, o gradiente é interpretado como um vetor no plano, ortogonal às curvas de nível $f = \text{constante}$. A magnitude do gradiente,

$$G = \sqrt{\tilde{G}^T \tilde{G}}, \quad (4.6)$$

correspondente à máxima taxa de variação direcional de f no plano uv . Devido a essa propriedade, o gradiente tem importância na detecção de contornos entre elementos de imagem, estas correspondendo a curvas delimitadores onde as variações da escala de cinza são acentuadas.

Outra informação importante fornecida pelo gradiente é sua direção relativamente ao eixo u no plano uv , dada por

$$\alpha(u, v) = \tan^{-1} \left(\frac{G_v}{G_u} \right). \quad (4.7)$$

que permite obter a direção de um contorno ou borda no plano uv , uma vez que esta é perpendicular à direção do vetor gradiente nesse ponto.

O cálculo do gradiente de uma imagem é baseado na obtenção das derivadas parciais de primeira ordem $\partial f / \partial u$ e $\partial f / \partial v$ em todos os pixels dessa imagem, com emprego da definição discretizada da derivada de primeira ordem de uma função unidimensional de variável discreta x ,

$$\frac{\partial f}{\partial x} \rightarrow f(x+1) - f(x) \quad (4.8)$$

Em processamento de imagem, é comum a utilização apenas da magnitude do gradiente [45]. Entretanto, a equação (4.6) é computacionalmente custosa por envolver o cálculo de potências e de uma raiz quadrada. Uma forma de diminuir o custo computacional é utilizar uma versão de magnitude de gradiente próxima à original, definida por

$$G' = \nabla' f = |G_u| + |G_v|, \quad (4.9)$$

que ainda contém informação sobre mudanças rápidas nos níveis de cinza da imagem.

Para se definir as máscaras que correspondam à aplicação do gradiente, no sentido mais restrito definido em (4.9), seja uma região 3×3 de uma imagem centrada em z_5 dada pela matriz

$$f = \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \quad (4.10)$$

Uma aproximação para o cálculo de cada derivada aparecendo em (4.9) é calcular o valor médio aproximado, por exemplo, para a direção horizontal,

$$\left| \frac{\partial f}{\partial u} \right| \rightarrow \frac{|(f_{13} - f_{11})/2 + (f_{23} - f_{21})/2 + (f_{33} - f_{31})/2|}{3},$$

ou equivalentemente,

$$\left| \frac{\partial f}{\partial u} \right| \rightarrow \frac{|(f_{13} + f_{23} + f_{33}) - (f_{11} + f_{21} + f_{31})|}{6}.$$

Em geral, acentua-se a contribuição da linha central da matriz incorporando-se um fator 2 nos termos de derivada ao longo dessa linha. Notando-se também que qualquer fator de escala na derivada pode ser normalizado, tem-se a versão modificada para $|\partial f / \partial u|$, a menos de um fator de escala,

$$\left| \frac{\partial f}{\partial u} \right| \rightarrow |(f_{13} + 2f_{23} + f_{33}) - (f_{11} + 2f_{21} + f_{31})| \quad (4.11)$$

Definição semelhante para a derivada $|\partial f / \partial v|$ fornece

$$\left| \frac{\partial f}{\partial v} \right| \rightarrow |(f_{31} + 2f_{32} + f_{33}) - (f_{11} + 2f_{12} + f_{13})| \quad (4.12)$$

e com essas definições, (4.9) assume a forma

$$G' = |(f_{13} + 2f_{23} + f_{33}) - (f_{11} + 2f_{21} + f_{31})| + |(f_{31} + 2f_{32} + f_{33}) - (f_{11} + 2f_{12} + f_{13})| \quad (4.13)$$

correspondente a duplicar o peso usando valores absolutos do gradiente no ponto z_5 na região definida.

As mascaras de Sobel são definidas pelas matrizes

$$sobel_v = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \quad (4.14)$$

$$sobel_v = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}. \quad (4.15)$$

Note-se que a soma dos coeficientes das máscaras de Sobel é sempre nula, fornecendo um valor nulo em regiões de nível de cinza constante, como o comportamento esperado de operadores diferenciais.

Se a filtragem com máscaras de Sobel é realizada com a aplicação do gradiente, cada ponto da imagem é substituído pela grandeza G' dada por (4.13). Convolvendo cada máscara isoladamente com a imagem, o módulo resultante em cada caso é dado por

$$G'_u = |(f_{13} + 2f_{23} + f_{33}) - (f_{11} + 2f_{21} + f_{31})|, \quad (4.16)$$

$$G'_v = |(f_{31} + 2f_{32} + f_{33}) - (f_{11} + 2f_{12} + f_{13})|. \quad (4.17)$$

Laplaciano para detecção de contornos

O Laplaciano de uma função de duas variáveis $f(u,v)$ é dado por

$$\nabla^2 f = \frac{\partial^2 f}{\partial u^2} + \frac{\partial^2 f}{\partial v^2} \quad (4.18)$$

Considerando a versão discretizada da segunda derivada parcial nas direções u e v , tem-se

$$\frac{\partial^2 f(u,v)}{\partial u^2} \rightarrow f(u+1,v) + f(u-1,v) - 2f(u,v), \quad (4.19)$$

$$\frac{\partial^2 f(u,v)}{\partial v^2} \rightarrow f(u,v+1) + f(u,v-1) - 2f(u,v), \quad (4.20)$$

e a equação (4.18) assume a versão discretizada

$$\nabla^2 f \rightarrow f(u+1,v) + f(u-1,v) + f(u,v+1) + f(u,v-1) - 4f(u,v), \quad (4.21)$$

que pode ser implementada através da convolução da imagem f com a máscara definida por

$$\text{laplaciano} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad (4.22)$$

que é uma máscara isotrópica, *i.e.*, invariante perante a operação de transposição, ou equivalentemente, de uma rotação de 90° . Uma forma de incorporar as transições na direção diagonal na definição do Laplaciano digital é através do uso da máscara

$$\text{laplaciano}' = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad (4.23)$$

que é isotrópica com relação a rotações de 45° e auxilia na detecção de bordas diagonais, além da horizontal e vertical.

A Figura 4.3-(a) mostra uma imagem original f , e as imagens resultantes da convolução de f com as máscaras $sobel_u$ (b), $sobel_v$ (c) e $\text{laplaciano}'$ (d).

As bordas nas direções u e v são bastante destacadas, em especial as da pupila, nas imagens (b) e (c), respectivamente, enquanto (d) mostra uma imagem mais uniforme, contendo o contorno completo da pupila. Uma forma de obter o centróide desta é através da transformada circular de Hough, que é tratada na Seção 4.3.

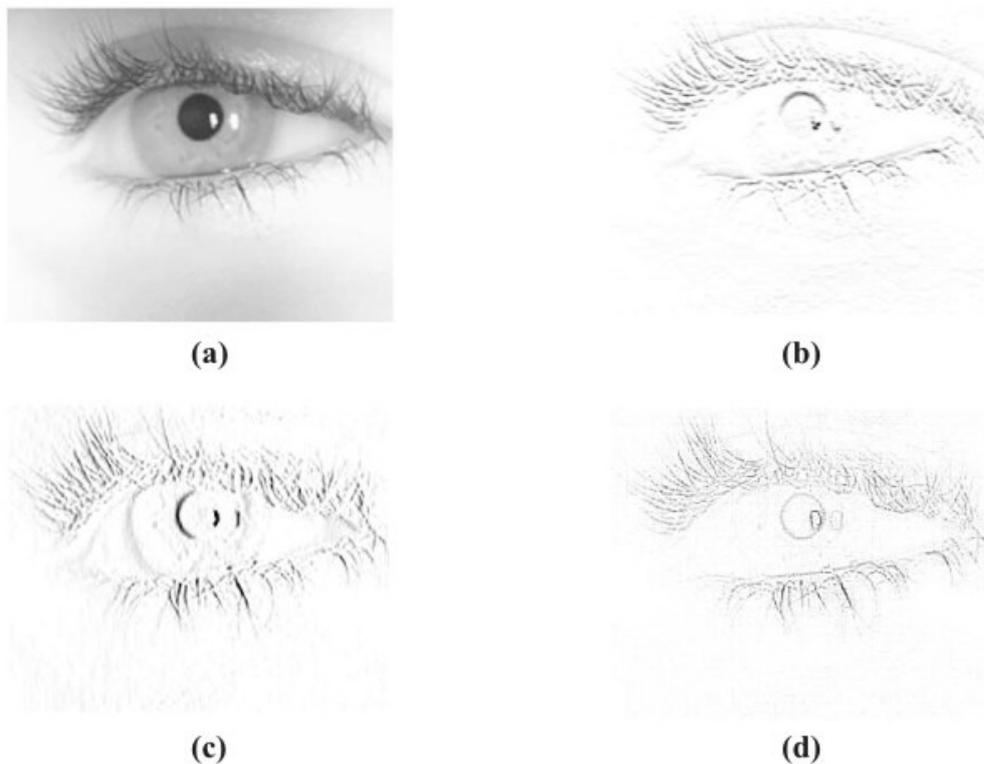


Figura 4.3 – Exemplo de aplicação de filtragem. Imagem original (a), resultado da aplicação das máscaras: $sobel_u$ (b), $sobel_v$ (c) e laplaciano' (d).

4.3. Algumas técnicas de detecção de formas

Transformada linear de Hough

A transformada linear de Hough permite realçar linhas retas em uma imagem. Uma possível solução é primeiramente encontrar todas as linhas determinadas por cada par de pontos e para cada linha obtida, o subconjunto de pontos pertencentes a essa linha. O problema é que esse procedimento envolve encontrar $N(N-1)/2 \sim N^2$ linhas e então realizar $N \times N \times (N-1)/2 \sim N^3$ comparações de todos os pontos com todas as linhas. Essa abordagem é computacionalmente custosa e por isso proibitiva na maioria das aplicações.

Hough [46] propôs uma abordagem alternativa que se tornou largamente conhecida nos dias atuais como a *Transformada de Hough*[45]. A Figura 4.4-(a) mostra o ponto $p_i = (u_i, v_i)$, e a uma linha reta que passa por esse ponto definida por

$$v_i = au_i + b. \quad (4.24)$$

Todas as linhas retas que passam por p_i , satisfazem à equação (4.24) para algum valor de a e b . Considere a equação $b = -u_i a + v_i$ que define uma única reta no plano ab de Hough como mostra a Figura 4.4-(b) onde a e b são as variáveis enquanto u_i, v_i são as constantes. Considere um segundo ponto $p_j = (u_j, v_j)$ no plano- uv da imagem que também possui uma única reta equivalente no plano de Hough definida por $b = -u_j a + v_j$. O ponto de interseção entre essas duas retas no plano de Hough define a inclinação a_0 e o termo b_0 da equação da reta que intercepta os pontos p_i e p_j no plano da imagem.

Um problema em trabalhar com equações do tipo $v = au + b$ para representar as linhas, é que $a \rightarrow \infty$ quando a reta é paralela ao eixo v e com isso o plano de Hough pode ter extensão infinita. Uma maneira de solucionar este problema é representar as retas envolvidas por coordenadas polares. Conforme ilustrado na Figura 4.5-(a), a forma polar da equação da reta que passa a uma distância mínima da origem ρ_0 , e ortogonal à linha radial rotacionada de θ_0 do eixo u é

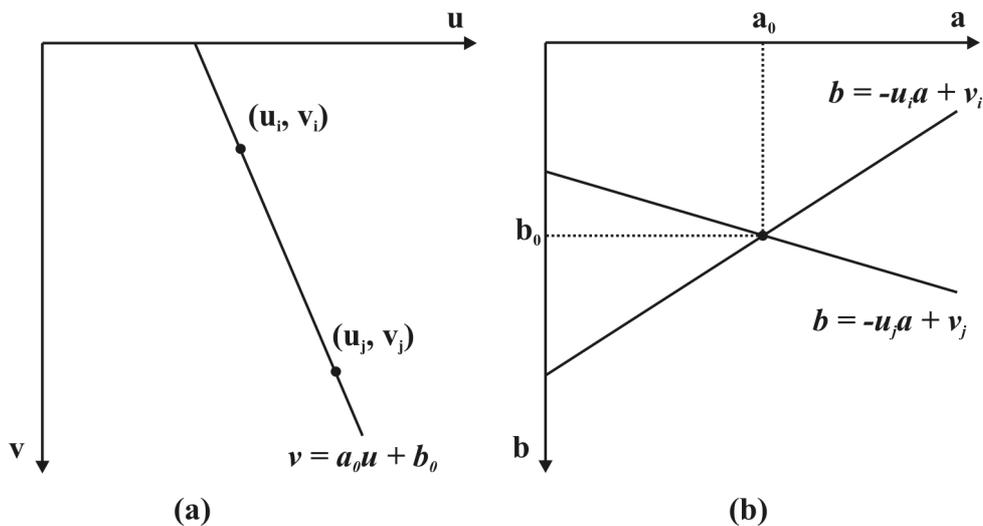


Figura 4.4 – (a) Plano da imagem uv e (b) plano de Hough ab .

$$\rho_0 = \rho \cos(\theta - \theta_0), \quad (4.25)$$

e dado que $u = \rho \cos \theta$, $v = \rho \sin \theta$, a equação (4.25) fornece

$$\rho_0 = u \cos \theta_0 + v \sin \theta_0. \quad (4.26)$$

Assim, os pontos no plano uv agora são mapeados em senóides no plano de Hough (ρ_0, θ_0) e as faixas de valores possíveis para ρ_0 e θ_0 são finitas, uma vez que $u \geq 0$ e $v \geq 0$, o ângulo polar satisfaz a $0 \leq \theta_0 < 2\pi$ e a distância ρ_0 satisfaz a $0 \leq \rho_0 \leq \sqrt{u_{\max}^2 + v_{\max}^2}$. A interseção entre duas senóides no espaço de Hough tem, portanto, apenas uma solução (ρ_0, θ_0) que define a condição de ambos os pontos pertencerem à mesma reta (definida no plano imagem por essas constantes), como ilustrado na Figura 4.5-(b).

A atratividade da Transformada de Hough é a subdivisão do espaço (ou plano) de Hough em células de acumulação, como ilustrado na Figura 4.5-(b). A célula no espaço de Hough de coordenadas (i, j) , correspondente aos valores (ρ_i, θ_j) tem a ela associado o valor de acumulação $A(i, j)$. O valor inicial de todas as células é nulo. A partir daí, para

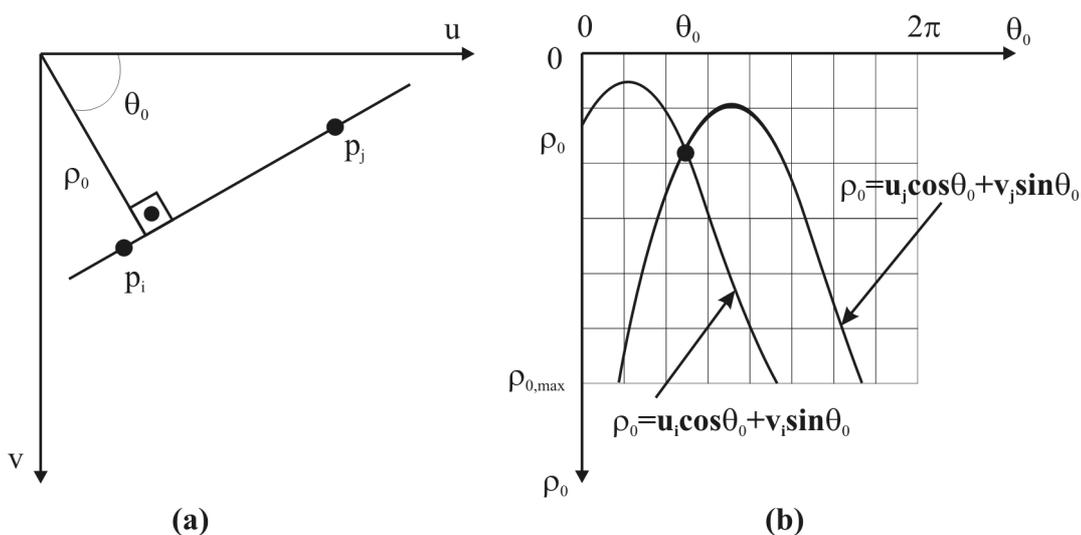


Figura 4.5 – Representação normal de uma linha (a) e subdivisão do espaço de Hough em células (b).

cada ponto (u_k, v_k) no plano da imagem, todos os possíveis pares (ρ_i, θ_j) são calculados de (4.26) e os respectivos valores de acumulação são incrementados. No final do processamento, um valor Q de $A(i, j)$ corresponde a Q pontos no plano uv , pertencentes à reta $\rho_i = u \cos \theta_j + v \sin \theta_j$. A largura da grade de subdivisões do plano $\rho\theta$ determina a exatidão da colinearidade desses pontos.

Uma subdivisão do eixo θ em k incrementos fornece, para cada ponto (u_k, v_k) , k valores correspondentes de ρ , calculados de $\rho = u_k \cos \theta + v_k \sin \theta$ para cada incremento de θ . Para N pontos na imagem, este método envolve $N \times k$ operações, tornando este procedimento linearmente dependente de N e, mais computacionalmente viável do que aquele mencionado no início desta seção, a menos que k seja próximo ou maior do que N .

Transformada circular de Hough

A Transformada de Hough é aplicável a qualquer função da forma $g(\tilde{X}, \tilde{C}) = 0$, onde \tilde{X} é uma matriz coluna de coordenadas e \tilde{C} , uma matriz coluna de coeficientes. Por exemplo, para a circunferência de centro em (c_1, c_2) e raio c_3 , definida por

$$(u - c_1)^2 + (v - c_2)^2 = c_3^2, \quad (4.27)$$

o espaço de Hough torna-se o espaço tridimensional (c_1, c_2, c_3) com células cúbicas e acumuladores da forma $A(i, j, k)$. O procedimento para determinação das circunferências na imagem é discretizar o espaço de Hough, incrementar c_1 e c_2 e calcular o valor de c_3 que satisfaça a (4.27) para cada ponto (u, v) no plano da imagem, e incrementar o acumulador associado à tripla (c_1, c_2, c_3) . Claramente, a complexidade da Transformada de Hough é proporcional ao número de coordenadas e à quantidade de coeficientes relacionados à função matemática associada à detecção da curva de interesse. E, apesar da resolução da pupila estar mais atrelada à detecção de elipses (5 parâmetros), algoritmos de Hough para detecção de círculos em imagens em escala de cinza e circunferências em imagens binarizadas (3 parâmetros) foram testados visando testar a viabilidade computacional dos mesmos.

Como dito anteriormente, após a delimitação dos parâmetros, incrementa-se c_1 e c_2 (coordenadas do centro do círculo), e resolve-se (4.27) para c_3 (raio) para os pixels com mesmo nível de cinza [47]. As imagens (a) e (c) na Figura 4.6 mostram o plano de Hough para a soma na variação do raio entre $19 \leq c_3 \leq 21$ e $15 \leq c_3 \leq 60$, respectivamente. Esses intervalos foram escolhidos para englobar o valor do raio real da pupila e testar a velocidade do algoritmo no caso do raio ser desconhecido. Os gráficos 3D representam a dependência da variável de acumulação como função de c_1 e c_2 para a soma das variações de c_3 de $15 \leq c_3 \leq 21$ e $15 \leq c_3 \leq 60$, e são mostrados nas imagens (b) e (d), respectivamente. É possível observar a susceptibilidade a erro do algoritmo dependente da escolha do intervalo do raio, em especial do processo de detecção de máximos. Dependendo do intervalo escolhido para o raio, mais ou menos picos podem ser detectados, comprometendo a resolução final para o centróide da pupila. Esses intervalos

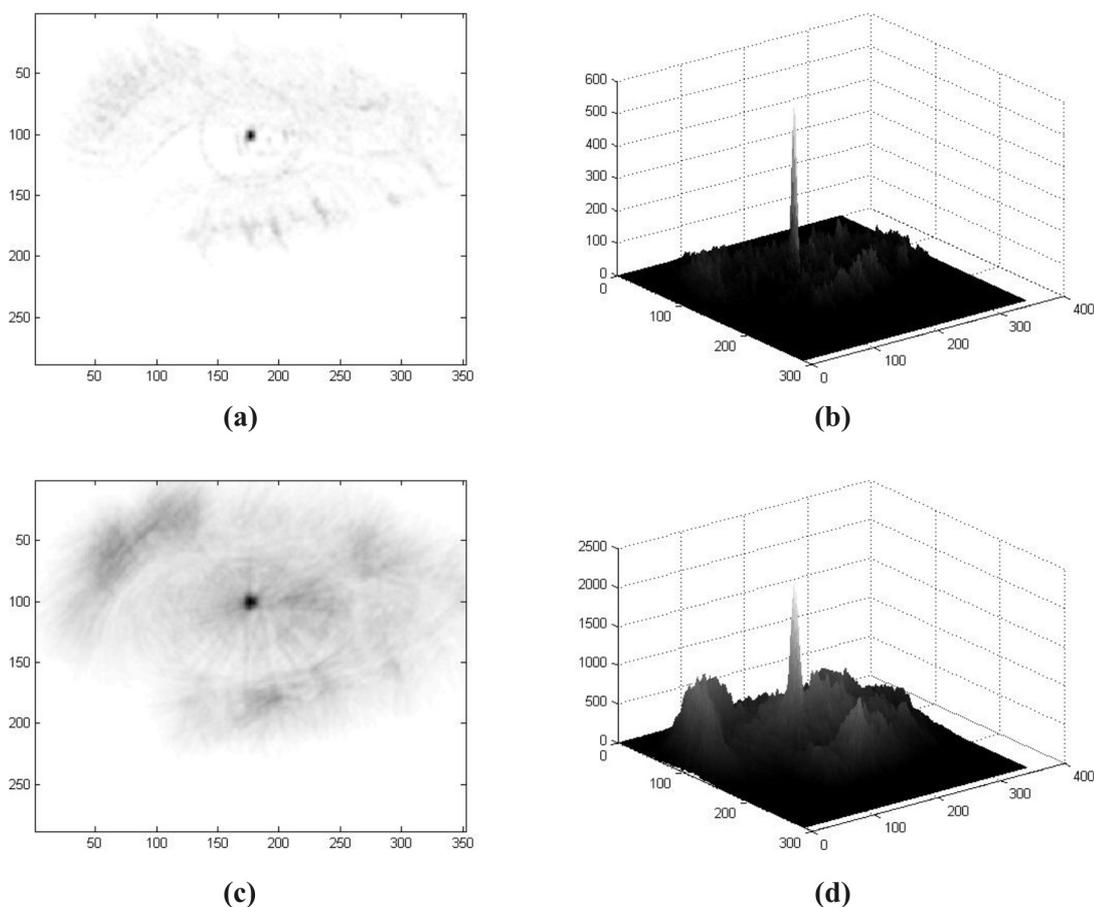


Figura 4.6 – Plano de Hough e respectiva representação 3D para o intervalo $19 \leq c_3 \leq 21$ (a) e (b), e análise idêntica para o intervalo $15 \leq c_3 \leq 60$ (c) e (d).

forneem bons resultados para o centróide da pupila quando aplicados à imagem original mostrada na Figura 4.3-(a) de 352×288 pixels.

Essa simulação é baseada no algoritmo apresentado por Tao Peng [47] e é implementada em Matlab[®], onde cada ponto (u, v) da imagem é considerado o centro de uma circunferência de raio c_3 variável. Para o intervalo de $15 \leq c_3 \leq 21$ pixels, o algoritmo apresenta velocidade de 2 fps, enquanto para $15 \leq c_3 \leq 60$, a velocidade é de 0.5 fps, ambas insatisfatórias para uma implementação em tempo real.

Uma maneira equivalente e mais rápida de calcular a Transformada Circular de Hough é ilustrada na Figura 4.7. A imagem apresentada na Figura 4.7-(a) é a imagem original da Figura 4.3-(a) e a convolução desta com o filtro de Sobel e posterior binarização é mostrada na Figura 4.7-(b). As bordas servem de imagem de entrada ao

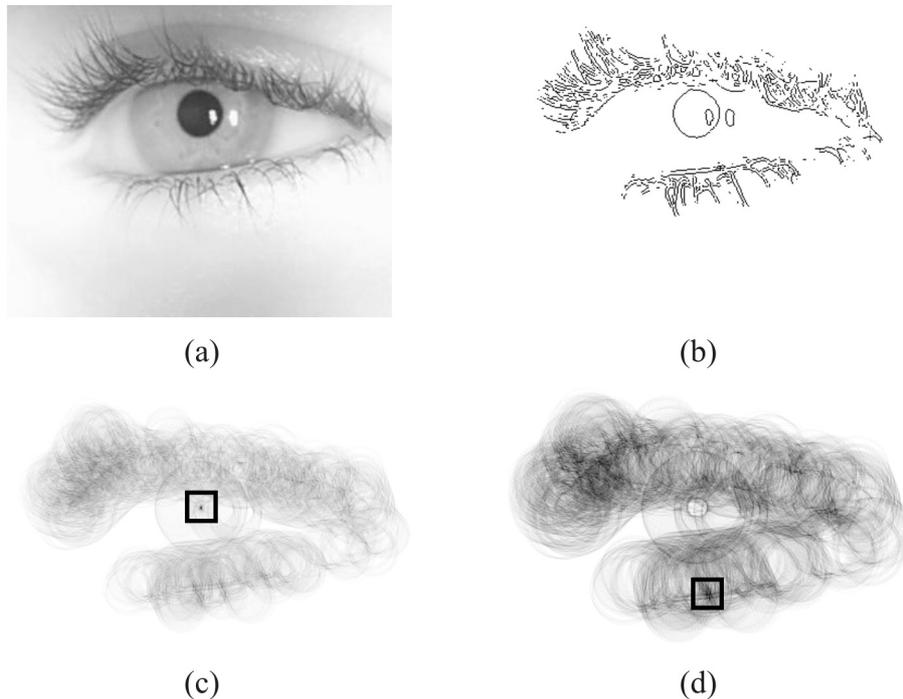


Figura 4.7 – Imagem original (a), obtenção das bordas através do filtro de Sobel (b), aplicação da Transformada Circular de Hough para $R = 21$ (c) e $R = 30$ pixels (d).

algoritmo (também em Matlab[®] e disponível no site *mathworks* sendo de autoria de Tao Peng [47] – função *houghcircle2*). A imagem tem dimensões de 352×288 pixels e cada ponto de borda (u, v) da imagem binarizada mostrada na Figura 4.7-(b) é considerado o centro de uma circunferência de raio fixo c_3 . As imagens mostradas na Figura 4.7-(c) e (d) mostram o plano de Hough para $c_3 = 21$ e $c_3 = 30$ pixels, respectivamente. Após obtenção da função de acumulação $A(c_1, c_2, c_3)$ no plano c_3 fixo, obtêm-se as células que (c_1, c_2, c_3) que contêm os máximos locais mais significantes no espaço de Hough, fornecendo como informação de saída o centro da circunferência obtida para cada caso. Na Figura 4.7, as imagens (c) e (d) mostram os centros das circunferências obtidas. As velocidades de processamento foram de 15 e 5 fps, para os raios de 21 e 30 pixels, respectivamente.

Essa taxa pode ser aumentada no processamento em tempo real, reduzindo o campo de visão da câmera, com a atualização em tempo real desse campo diretamente no algoritmo.

A partir dos resultados de simulações desenvolvidas neste trabalho, concluímos que a detecção e localização de objetos circulares em uma imagem com a transformada circular de Hough tem, ao menos, três desvantagens:

- (i) É computacionalmente custosa, possibilitando taxas de processamento de, no máximo 1 fps em Matlab[®], que é insatisfatória para aplicações em tempo real.
- (ii) O raio não pode ser fixo, pois a abertura da pupila varia dependendo da intensidade da iluminação ambiente. Além de ser necessário o uso do intervalo para o raio, este deve conter as amplitudes de abertura máxima e mínima da pupila, acarretando em erros.
- (iii) A pupila tem um formato mais elipsoidal do que circular, fonte de erro constante quando se aplica a transformada circular de Hough em tempo real, sobretudo quando o sensor de imagens está posicionado próximo ao olho.

A Figura 4.8 ilustra as conseqüências decorrentes do problema descrito no item (iii) do parágrafo anterior. Observe que os pixels da união entre a pálpebra e a íris, logo acima da pupila (Figura 4.8-(b)), são levados em consideração no algoritmo de detecção de

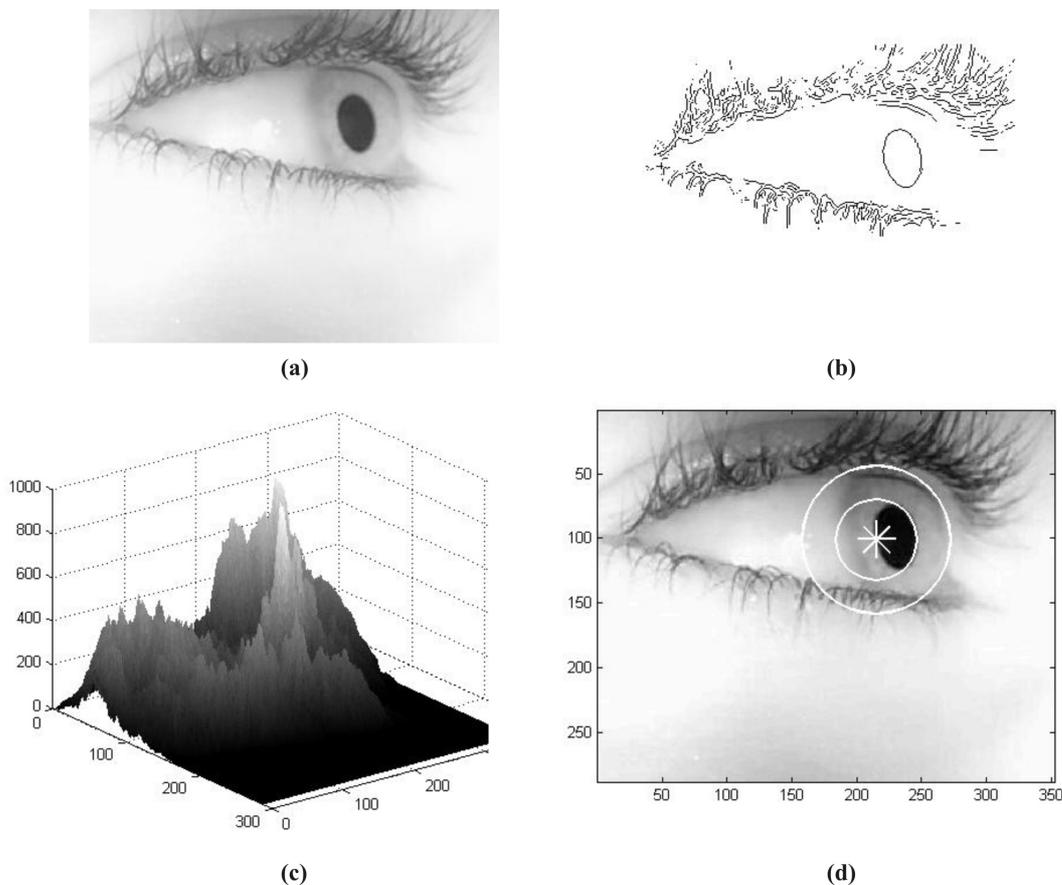


Figura 4.8 – Detecção errônea do centro da pupila usando a Transformada Circular de Hough. Imagem original (a), contornos de Sobel (b), soma dos planos correspondentes à variação no raio (c) e centróides deslocados (d).

máximos e duas circunferências são obtidas. O centróide da pupila é deslocado na Figura 4.8-(d) devido à prioridade dada aos pixels pertencentes à junção pálpebras/ globo ocular por serem mais provavelmente coincidentes com a circunferência detectada. Nesse caso, onde a pupila apresenta claro comportamento elipsoidal, a transformada circular de Hough se apresenta insatisfatória.

Pode-se fazer uso de uma transformada elipsoidal de Hough para detecção de elipses. Neste caso, trata-se da determinação de 5 parâmetros para resolução da pupila e não seria mais rápida do que aquela de detecção de raio fixo que atingiu velocidade máxima de 15 fps, como mostrado anteriormente. Por esse motivo, uma técnica de detecção de elipses foi aplicada, como tratado na Seção seguinte.

4.4. Detecção de elipses

A transformada de Hough, apresentada na seção anterior, não é um procedimento de desempenho robusto para obtenção de parâmetros em tempo real. Portanto, uma metodologia mais eficaz deve ser empregada. Para o processamento de imagens para fins de rastreamento ocular, uma técnica de detecção de elipses permitiria detectar o centróide da pupila em qualquer situação (frontal ou ligeiramente lateral). Uma técnica eficiente, proposta por W. Gander, *et al.* é baseado na aproximação da melhor elipse que se adapte à um conjunto de pontos, através do método dos mínimos quadráticos[48].

Apesar de teoricamente correto, o método apresentado em [48] é instável, podendo fornecer resultados infinitos e/ou complexos, devido às singularidades impostas pela formulação matemática utilizada.

Para solucionar este problema, Radim H. e Jan Flusser propuseram um algoritmo numericamente estável [49] para detecção de elipses, implementado em Matlab[®]. O formalismo matemático utilizado está fora do escopo deste trabalho e não será abordado aqui. O fato relevante é que o algoritmo proposto se mostrou muito satisfatório dos pontos de vista de confiabilidade e tempo de processamento. As imagens da Figura 4.9-(a) e (d) apresentam duas imagens originais em escala de cinza (352×288 pixels). As imagens da Figura 4.9-(b) e (e) apresentam os contornos das originais (a) e (d), respectivamente, e as imagens (c) e (f) mostram as elipses que melhor se adaptam aos conjuntos de pontos dos respectivos contornos (b) e (e), obtidos com o algoritmo proposto em [49]. A velocidade de processamento é superior a 300 fps, muito superior à máxima velocidade de 2 fps obtida com a transformada de Hough, que ainda resultava em erros para imagens ligeiramente laterais, e, portanto, elípticas da pupila.

Vale ressaltar ainda que, para o processamento de uma imagem com diversos contornos, uma segmentação prévia é necessária, *i.e.*, os contornos devem ser rotulados (*clustering*) antes de considerados no algoritmo de aproximação de elipses. Dessa forma, apenas os pixels pertencentes ao contorno da pupila são avaliados e aproximados, sem levar em consideração quaisquer contornos provenientes de outras características da imagem. Nas imagens da Figura 4.9, os limiares usados para obtenção dos contornos (b) e

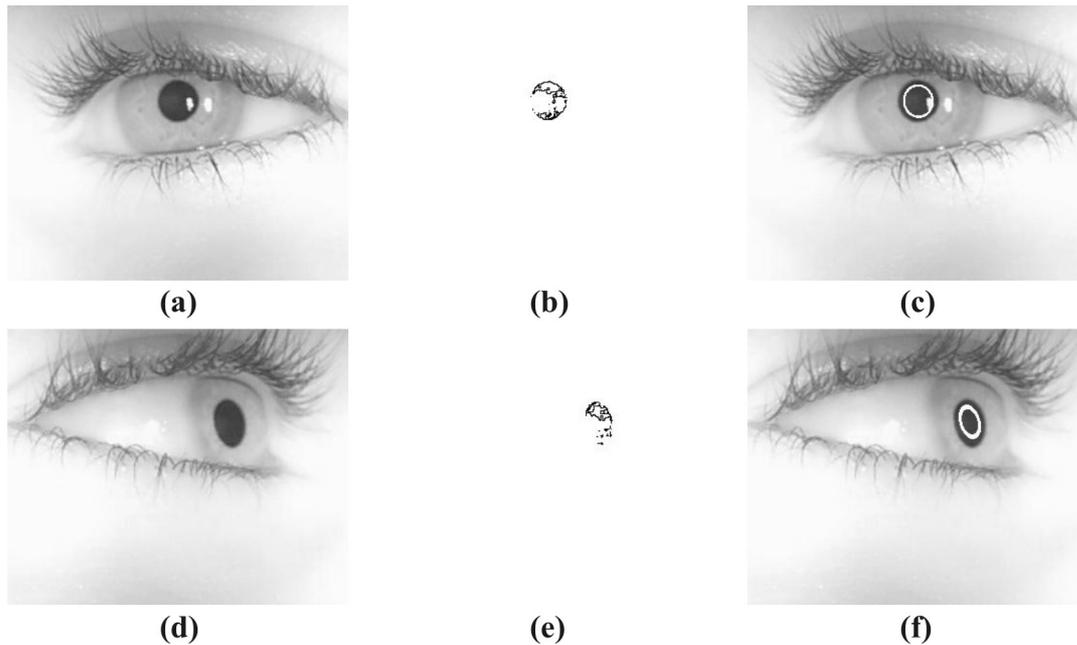


Figura 4.9 – Exemplo de obtenção do centróide da pupila através do rastreamento de elipses. Imagens originais (a) e (d), contornos da binarização (b) e (e), e elipses resultantes (c) e (f), respectivamente.

(e) foram tais que apenas pixels do contorno da pupila são considerados, retirando-se o erro que estará presente na implementação em tempo real.

Este algoritmo é implementado em C através da função *cvFitEllipse* pertencente ao conjunto de funções do código aberto OpenCV[®] [50], da Intel[®]. Na implementação em tempo real, o código em C fornece velocidade de 160 fps, por que a aproximação de elipses é feita para os diversos *clusters* existentes na imagem de tamanho 352×288 pixels, enquanto a implementação em Matlab foi testada para um único *cluster*.

5. RASTREAMENTO OCULAR

*“Na arte como na ciência, no fazer como no agir,
tudo está em apanhar claramente um assunto
e tratá-lo de conformidade com sua natureza”.*

Göthe.

Neste capítulo, será descrito um modelo matemático geral, proposto por Elias D. Guestrin, e Moshe Eizenman, [51] para sistemas de estimação remota ponto de mira (POG) que utilizam o centro da pupila e reflexões de uma ou mais fontes de luz. O modelo geral engloba todas as possíveis configurações do sistema, do mais simples, composto por uma câmera de vídeo e uma fonte de luz, aos mais complexos, compostos por múltiplas câmeras e múltiplas fontes. A intenção é implementar gradualmente os diferentes níveis de complexidade do sistema, a fim de estabelecer comparativos entre os mesmos e o sistema implementado com a cabeça fixa.

5.1. Modelo matemático

O modelo matemático desenvolvido por Guestrin e Eizenmann [51] é baseado na disposição geométrica de elementos ilustrada na Figura 5.1. Nesse modelo, a córnea é considerada uma superfície esférica de raio R de centro localizado em \vec{o} . O eixo óptico ocular é determinado pelo centro de rotação do olho \vec{d} e pelo centro da pupila \vec{p} , enquanto o eixo visual é desviado do eixo óptico. O objetivo do procedimento matemático é reaver o POG do usuário, formalmente definido como o ponto de interseção do eixo visual com o plano de interesse, este em geral o monitor de um computador.

A Figura 5.1 mostra a câmera j em modelo de câmera de orifício, uma fonte pontual de luz i e o globo ocular. Os vetores $\vec{l}_i, \vec{d}, \vec{o}, \vec{p}, \vec{c}_j, \vec{u}_{ij}$ e \vec{v}_j estão definidos na Tabela 5.1.

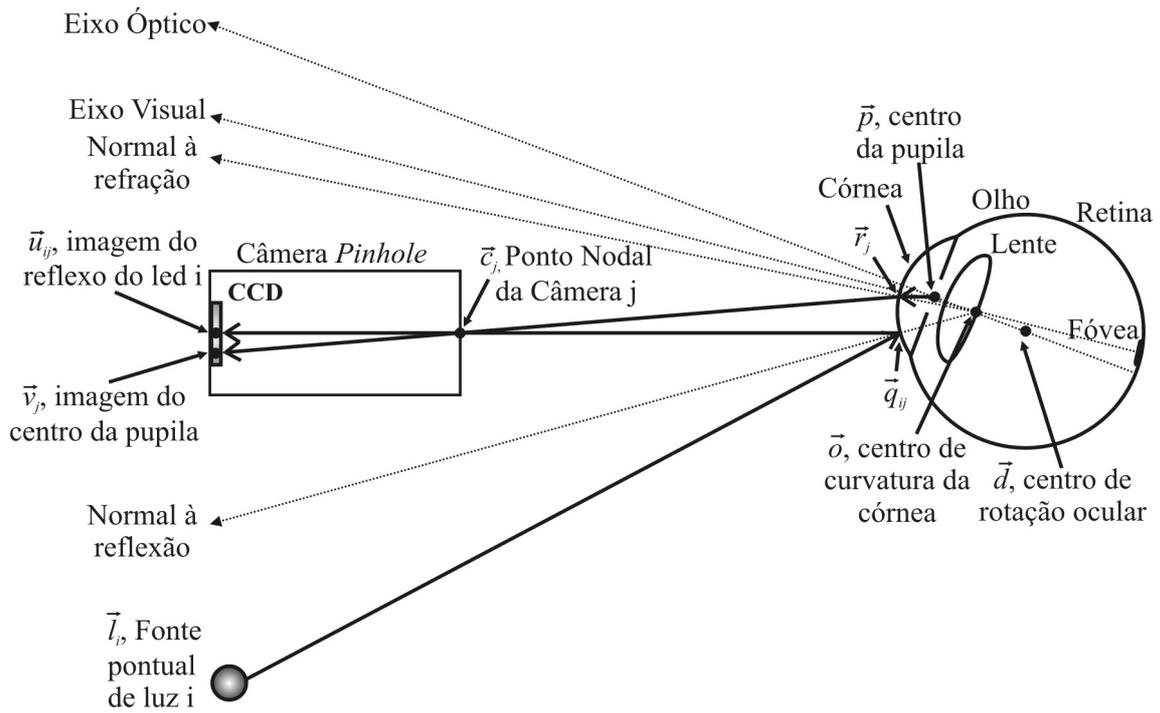


Figura 5.1 – Disposição geométrica para o modelo matemático de rastreamento ocular [51].

Todos esses vetores são baseados no sistema de coordenadas de cena (x,y,z) , como tratado na seção 3.2. Os vetores \vec{r}_j e \vec{q}_{ij} em coordenadas de cena determinam o ponto de refração na superfície da córnea do raio proveniente da pupila e o ponto de reflexão do raio proveniente da fonte pontual i , respectivamente.

Assume-se que os vetores \vec{l}_i e \vec{c}_j sejam previamente conhecidos e que os vetores \vec{v}_j e \vec{u}_{ij} sejam obtidos no sistema de referência da câmera, por processamento de imagem, e posteriormente transformados em coordenadas de cena com o uso das matrizes de rotação \tilde{T} e translação \tilde{X}_0 da equação (A.10) do Apêndice A.

Na Figura 5.1, o raio refletido proveniente de \vec{l}_i , intercepta a câmera j em \vec{c}_j , gerando a imagem \vec{u}_{ij} . O que permite escrever

$$\vec{q}_{ij} - \vec{c}_j = \lambda_{ij} (\vec{c}_j - \vec{u}_{ij}), \quad (5.1)$$

Tabela 5.1 – Parâmetros da Figura 5.1.

Vetor	Descrição	Aplicação no modelo matemático
\vec{l}_i	Posição da i -ésima fonte de luz	Constante conhecida
\vec{d}	Centro de rotação do globo ocular	Variável
\vec{o}	Centro de curvatura da córnea	Variável
\vec{p}	Centro da pupila	Variável
\vec{c}_j	Ponto nodal da j -ésima câmara	Constante conhecida
\vec{u}_{ij}	Imagem do reflexo da i -ésima fonte na j -ésima câmara	Proc. de imagem + transformação
\vec{v}_j	Imagem do centro da pupila na câmara j .	Proc. de imagem + transformação
\vec{r}_j	Ponto de origem, na córnea, do raio refratado da pupila	Variável
\vec{q}_{ij}	Ponto de reflexão da i -ésima fonte cuja imagem é \vec{u}_{ij}	Variável

para algum escalar λ_{ij} . De forma semelhante, para o raio refratado proveniente do ponto \vec{r}_j passando por \vec{c}_j e gerando a imagem \vec{v}_j obtém-se

$$\vec{r}_j - \vec{c}_j = \sigma_j (\vec{c}_j - \vec{v}_j), \quad (5.2)$$

para algum escalar σ_j . Na , pelas leis da reflexão [30], os raios incidente e refletido estão no mesmo plano que contém a direção normal, i.e.,

$$(\vec{l}_i - \vec{c}_j) \times (\vec{q}_{ij} - \vec{c}_j) \cdot (\vec{o} - \vec{c}_j) = 0 \quad (5.3)$$

e estão igualmente desviados da direção normal, o que corresponde a

$$\frac{(\vec{l}_i - \vec{q}_{ij}) \cdot (\vec{q}_{ij} - \vec{o})}{|\vec{l}_i - \vec{q}_{ij}|} = \frac{(\vec{c}_j - \vec{q}_{ij}) \cdot (\vec{q}_{ij} - \vec{o})}{|\vec{c}_j - \vec{q}_{ij}|}. \quad (5.4)$$

Duas outras condições são obtidas das leis de refração [30]. Uma condição é que o raio proveniente de \vec{p} , atingindo a córnea em \vec{r}_j esteja no mesmo plano do raio refratado na direção do vetor $(\vec{r}_j - \vec{c}_j)$ e da direção normal definida pelo vetor $(\vec{r}_j - \vec{o})$, i.e.,

$$(\vec{r}_j - \vec{c}_j) \times (\vec{o} - \vec{c}_j) \bullet (\vec{p} - \vec{c}_j) = 0. \quad (5.5)$$

A segunda condição é a Lei de Snell [30] que relaciona o seno do ângulo entre as direções $(\vec{r}_j - \vec{c}_j)$ e $(\vec{r}_j - \vec{o})$ e o seno do ângulo entre as direções $(\vec{r}_j - \vec{p})$ e $(\vec{r}_j - \vec{o})$. Sendo n o índice de refração do humor aquoso e assumindo que haja refração desprezível na interface entre esse meio e a córnea, a lei de Snell pode ser posta na forma vetorial

$$n \left| (\vec{r}_j - \vec{o}) \times \frac{(\vec{p} - \vec{r}_j)}{|\vec{p} - \vec{r}_j|} \right| = \left| (\vec{r}_j - \vec{o}) \times \frac{(\vec{c}_j - \vec{r}_j)}{|\vec{c}_j - \vec{r}_j|} \right|, \quad (5.6)$$

Sendo R o raio de curvatura da córnea, conforme ilustrado na Figura 5.1, para que os pontos \vec{q}_{ij} e \vec{r}_j estejam em sua superfície é necessário que

$$|\vec{q}_{ij} - \vec{o}| = R, \quad (5.7)$$

$$|\vec{r}_j - \vec{o}| = R. \quad (5.8)$$

Além disso, considerando a distância do centro da pupila ao centro da córnea como sendo uma constante K , tem-se

$$|\vec{p} - \vec{o}| = K. \quad (5.9)$$

Como o eixo óptico do olho passa pelo centro da pupila \vec{p} e pelo centro de curvatura da córnea \vec{o} , se esses vetores forem obtidos através do sistema de equações (5.1) a (5.9), o eixo óptico do olho do usuário pode ser reconstruído. Mas, como o POG é definido como a interseção da cena com o eixo visual e não com o eixo óptico, a relação

entre estes também deve ser modelada. O eixo visual é definido pelo centro de curvatura da córnea \vec{o} e pelo centro da fóvea (região da retina de maior acuidade visual) [51]. Já o eixo óptico é definido pelo centro de curvatura da córnea \vec{o} e pelo centro de rotação do olho \vec{d} , como ilustrado na Figura 5.1.

5.2. Determinação do eixo visual

O sistema formado pelas equações (5.1) a (5.9) pode ser resolvido para o centro de curvatura da córnea \vec{o} e para o centro da pupila \vec{p} , possibilitando a obtenção do eixo óptico do olho no espaço de cena. Entretanto, para se obter o POG, é preciso relacionar o desvio do eixo visual em relação ao eixo óptico definido por \vec{o} e \vec{p} . Considere o sistema de coordenadas de cena (x,y,z) , mostrado na Figura 5.2 com origem no monitor, o plano xy correspondendo ao plano do monitor e com o eixo z dirigido no sentido da superfície do monitor para a cena. Considere os ângulos θ e φ que definem a orientação do eixo óptico definido por $\vec{p}-\vec{o}$ em relação ao sistema ocular $x_e y_e z_e$, transladado do sistema xyz , conforme ilustrado na Figura 5.2. Com essa definição, θ e φ podem ser calculados a partir da equação

$$\frac{\vec{p}-\vec{o}}{|\vec{p}-\vec{o}|} = \begin{pmatrix} \cos\varphi\sin\theta \\ \sin\varphi \\ -\cos\varphi\cos\theta \end{pmatrix}, \quad (5.10)$$

e o eixo visual pode ser determinado pela forma paramétrica

$$\vec{t}-\vec{o} = k \begin{pmatrix} \cos(\varphi+\beta)\sin(\theta+\alpha) \\ \sin(\varphi+\beta) \\ -\cos(\varphi+\beta)\cos(\theta+\alpha) \end{pmatrix}, \quad (5.11)$$

para qualquer k , onde α e β são os ângulos de desvio horizontal e vertical entre os eixos óptico e visual, respectivamente, e \vec{t} é o vetor olhar, obtido pela interseção do eixo visual com o plano da tela, conforme ilustrado na Figura 5.2. Tipicamente, $\alpha < 0$ para o olho direito (Figura 5.2), $\alpha > 0$ para o olho esquerdo e $\beta > 0$ para ambos [51].

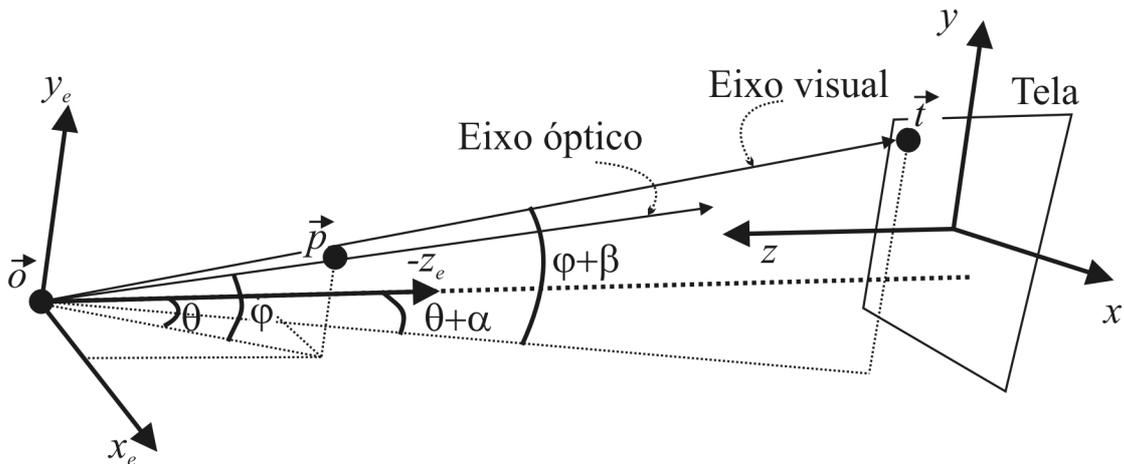


Figura 5.2 – Eixos óptico e visual para o olho direito e seus ângulos de desvio horizontal e vertical θ e ϕ , e $\theta + \alpha$ e $\phi + \beta$, respectivamente. Todos os vetores \vec{o} , \vec{p} e \vec{t} são relativos ao sistema de cena xyz .

Como no plano de cena, $z = 0$, o ponto de visão é dado pela equação (5.11) para um valor k de forma que a componente t_z do vetor \vec{t} seja nula, ou seja,

$$k = \frac{o_z}{\cos(\phi + \beta)\cos(\theta + \alpha)} \quad (5.12)$$

As equações (5.11) e (5.12) possibilitam obter o eixo visual a partir do conhecimento prévio dos respectivos ângulos de desvio horizontal e vertical α e β da fóvea em relação ao eixo óptico, inerentes a cada usuário, e dos vetores \vec{o} e \vec{p} que definem o eixo óptico do olho.

Obtenção dos parâmetros oculares

Os parâmetros oculares do usuário são obtidos através de um processo de calibração inicial, realizado uma vez para cada sujeito. Para tanto, com a cabeça relativamente imóvel, o indivíduo é solicitado a fixar seu olhar em nove pontos distribuídos de forma retangular num padrão 3×3 na tela. A partir das imagens obtidas, estimativas do POG são calculadas a partir das equações (5.1) a (5.12) com valores iniciais dos parâmetros típicos do olho humano, listados na Tabela 5.2[51]. Os parâmetros são então otimizados iterativamente, com o emprego das equações (5.1) a (5.12), através do

Tabela 5.2 – Parâmetros típicos do olho humano [51].

	Descrição	Valor típico
R	Raio de curvatura da córnea	7.8 mm
K	Distância do centro da pupila ao centro de curvatura da córnea	4.2 mm
n	Índice de refração da córnea e do humor aquoso combinados	1.3375
α	Ângulo horizontal entre os eixos óptico e visual do olho	-5° dir./ 5° esq.
β	Ângulo vertical entre os eixos óptico e visual do olho	1.5°

método dos mínimos quadráticos. Comumente se usam várias medidas [51] para cada um dos nove pontos na tela e, uma média das medidas é usada como base para os cálculos dos POGs.

Obtenção do ponto de mira

O conjunto de equações da seção anterior possibilita realizar o rastreamento ocular remoto sob certas condições:

(i) A câmera e as fontes pontuais são afixadas ao monitor e suas posições são previamente conhecidas após uma calibração adequada. Com isso, obtém-se o ponto nodal da câmera j , \vec{c}_j , e a posição de cada fonte \vec{l}_i , todos baseados no sistema de coordenadas de cena (x,y,z) , com origem no monitor.

(ii) O raio de curvatura da córnea R , a distância entre o centro da pupila e o centro de curvatura da córnea K , o índice de refração n , os desvios α e β são parâmetros que dependem de cada usuário e devem ser previamente obtidos através de um processo de calibração, como descrito anteriormente.

(iii) A obtenção do eixo óptico definido pelo centro de curvatura da córnea \vec{o} , e pelo centro da pupila \vec{p} depende da configuração do sistema, i.e., número de câmeras e de fontes pontuais.

(iv) Uma vez que o centro de curvatura da córnea \vec{o} e o centro da pupila \vec{p} são conhecidos, a reconstrução do eixo visual do olho é independente da configuração do sistema e depende apenas do conhecimento prévio dos desvios α e β .

Como descrito anteriormente, o rastreamento pode ser feito com o uso de uma câmera e uma fonte pontual, com uma câmera e múltiplas fontes, ou com múltiplas câmeras e múltiplas fontes. Com uma câmera e uma fonte, o POG só pode ser obtido após um processo de calibração para o caso em que a cabeça do usuário fica parada em relação ao sistema. Com uma câmera e múltiplas fontes o POG pode ser obtido após um processo de calibração com vários pontos na tela, que objetiva a obtenção dos parâmetros inerentes ao olho de cada usuário. E para múltiplas câmeras e múltiplas fontes o rastreamento é possível após uma calibração simples com apenas um ponto na tela. Quanto maior a complexidade do sistema, mais simples é o processo de calibração. As seções a seguir tratam de duas situações específicas de importância para o desenvolvimento deste trabalho.

5.3. Situações específicas

Uma câmera e uma fonte pontual

Dado que cada equação vetorial da seção anterior equivale a três equações escalares, as equações (5.1) a (5.9) fornecem 13 equações. Admitindo conhecidos R , K , n , α e β , há 14 incógnitas a serem determinadas. Assim sendo, o problema não pode ser resolvido a menos que uma condição adicional seja especificada. Essa condição pode ser a fixação da distância da cabeça do usuário em relação à câmera, o que equivale a conhecer-se, de forma indireta o valor de $|\vec{c}_1 - \vec{o}|$. Caso contrário, se houver movimento da cabeça do usuário, esse parâmetro tem de ser estimado através de algum sistema que proporcione o seu rastreamento. Portanto, com a hipótese de o usuário manter a cabeça imóvel é possível resolver o sistema de equações e rastrear seu ponto de mira. Nessa situação é necessária a realização da calibração com pelo menos 7 pontos para obtenção dos parâmetros dos oculares. Esse sistema foi desenvolvido e testado no âmbito desta Dissertação e os resultados estão apresentados no Capítulo 6.

Uma câmera e múltiplas fontes pontuais

O caso de uma câmera e múltiplas fontes pontuais propicia a solução do sistema formado pelas equações (5.1) a (5.9) com $i=1\dots N$, e $j=1$, se os parâmetros oculares R , K e n são conhecidos. Conforme demonstrado por Guestrin e Eizenman [51] esse conjunto de equações fornece $(5N+3)$ equações escalares e $(4N+4)$ incógnitas. Para o caso específico de duas fontes pontuais, tratado neste trabalho, tem-se 13 equações para a determinação de 12 incógnitas. Isso permitiria assim o rastreamento ocular com liberdade de movimento de cabeça. Nesse sistema ainda é necessária uma calibração prévia, conforme descrito na seção 5.2 para a obtenção dos parâmetros R , K , n , α e β .

6. SISTEMA DE RASTREAMENTO OCULAR 1 – SRO1

“Talento é 1% inspiração e 99% transpiração”.

Thomas Edison.

O sistema de rastreamento ocular 1 (SRO1) foi desenvolvido com o intuito de se obter um mouse ocular simples e de baixo custo para uso do computador por pessoas deficientes. O dispositivo resultante é, de fato, barato, e é montado na cabeça do usuário (*head-mounted*) de forma que impossibilita liberdade de movimento da cabeça. Neste Capítulo serão abordados os algoritmos desenvolvidos em Matlab[®] e em C, destacando-se as vantagens e desvantagens de cada um, no tocante à estabilidade e ao desempenho dos mesmos. O Capítulo 7 contém resultados preliminares da implementação da 2ª versão do sistema de rastreamento ocular (SRO2), obtido com o uso de uma câmera acoplada ao monitor que capta imagens do indivíduo, e duas fontes pontuais de iluminação \vec{l}_1 e \vec{l}_2 , posicionadas lateralmente à tela. Esta montagem tenciona oferecer algum grau de liberdade para o movimento da cabeça do usuário que não há possui qualquer dispositivo acoplado.

6.1. Descrição do sistema

O Sistema de Rastreamento Ocular 1, desenvolvido no âmbito deste trabalho, possibilita o controle do mouse a partir do movimento ocular do usuário e foi implementado com uma câmera e uma fonte de luz infravermelha. Os componentes do sistema são mostrados na Figura 6.1, e suas especificações são:

(i) Suporte do dispositivo que possibilita o uso por uma pessoa que use óculos de grau.

(ii) Haste de suporte flexível o suficiente para possibilitar o ajuste do campo de visão da câmera (para uma boa obtenção da imagem do olho) e rígida o suficiente para evitar oscilações de imagem provenientes de vibrações mecânicas. A câmera é posicionada abaixo do olho, de forma a não bloquear o campo de visão do usuário de forma significativa, *i.e.*, as regiões de observação do monitor e teclado ficam desobstruídas. No posicionamento é importante que haja pouca interferência dos cílios ou pálpebras no campo de visão da íris e da pupila.

(iii) Câmera USB Vimicro[®] ZC0- 301PLH, de resolução 300K, de visão noturna, com lente adaptada para captação de imagens apenas no infravermelho.

(iv) LED infravermelho emitindo em 850 nm, para iluminação do olho.

O sistema foi testado num Laptop Dell Latitude D520 com plataforma Windows[®], com processador Core 2 Duo de 1.66 GHz Intel[®], 2 MB de cache e 2 GB de memória RAM. O software utilizado para aquisição e processamento de imagem foi inicialmente o Matlab[®] R2007ae posteriormente o código computacional foi implementado em C.

A montagem desse sistema é simples e de baixo custo. A maioria das câmeras USB comercialmente disponíveis possuem um filtro de bloqueio de infravermelho em suas

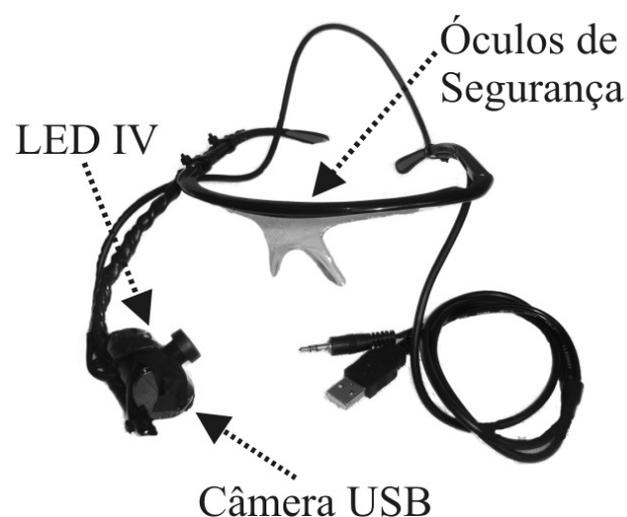


Figura 6.1 – Configuração do Sistema de Rastreamento Ocular 1 – SRO1.

lentes, para evitar saturação do sensor de imagens, pois nessa região, os fotodiodos constituintes do sensor apresentam maior responsividade (ver seção 3.1). A câmera Vimicro utilizada, por ser de visão noturna, já possui a lente sem bloqueio de infravermelho e 6 LEDs embutidos emitindo no comprimento de onda 850 nm. O espectro de emissão desse tipo de LED está mostrado na Figura 6.2-(a), juntamente com os espectros de outros LEDs comerciais.

Para aumentar a robustez do sistema com relação ao nível de intensidade de iluminação ambiente, um filtro óptico de bloqueio da luz visível foi utilizado. A Figura 6.2-(b) mostra as curvas de transmissão de quatro filtros testados. Os filtros VCR1, VCR2 e DVD, são, respectivamente, usados em aparelhos de vídeo-cassete e DVDs convencionais para limitação de ruído na captação de sinal de controle-remoto. Outra maneira ainda mais acessível de se obter um filtro óptico de bloqueio do visível é utilizar uma ou mais camadas de filmes fotográficos já revelados, cuja curva de transmissão característica também é mostrada na Figura 6.2-(b). Opcionalmente há filtros ópticos disponíveis no mercado com faixas estreitas de transmissão em torno do comprimento de onda de interesse.

Uma imagem típica do olho do autor sob iluminação infravermelha, obtida com o filtro DVD acrescentado à câmera VIMICRO, é mostrada na Figura 6.3-(a), enquanto a

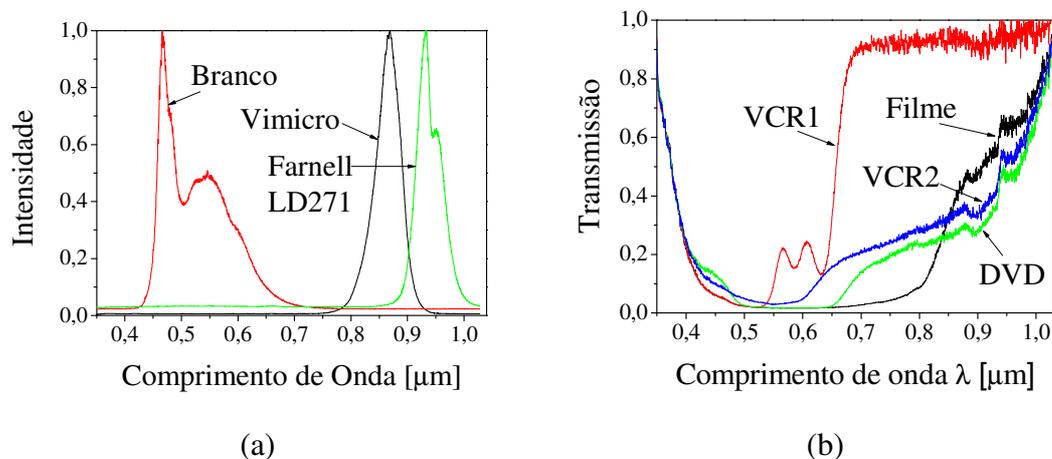


Figura 6.2 – (a) Espectros de diferentes LEDs. (b) Curvas de transmissão de diversos filtros ópticos.

Figura 6.3-(b) mostra a binarização sob um limiar fixo $\xi = 90$ da primeira. Pode-se observar a boa separação entre os níveis de cinza da pupila em relação ao resto da imagem.

Observe que esse método utiliza apenas a iluminação “fora do eixo” da câmera ou, equivalentemente, a imagem obtida é somente a imagem “pupila escura” descrita na seção 1.2.

O método de obtenção do centro de massa consiste em calcular as coordenadas do centróide da pupila, $\tilde{X}_C \equiv (x_c \ y_c)^T$ a partir das médias

$$\tilde{X}_c = \frac{1}{N} \sum_{i=1}^N \tilde{X}_i, \quad (6.1)$$

dos N pixels pretos (nulos), de coordenadas $\tilde{X}_i \equiv (x_i \ y_i)^T$, na imagem binarizada. Esse procedimento é simples e rápido para obtenção do centro de massa. Vale ressaltar que os testes iniciais citados neste trabalho foram efetuados em Matlab[®], que se mostrou uma ferramenta custosa e insatisfatória para o tratamento de imagens em tempo real. Para superar essas limitações, algoritmos equivalentes foram desenvolvidos em C, apresentando desempenhos muito superiores aos iniciais.

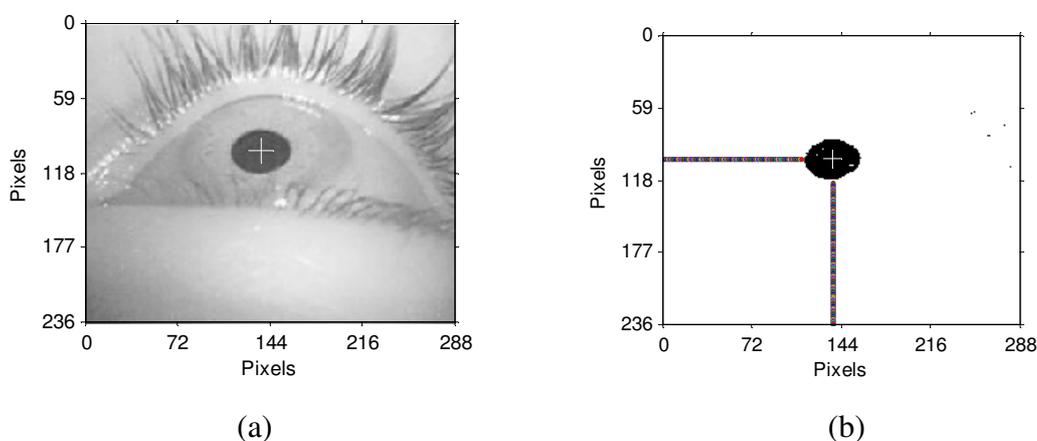


Figura 6.3 – Imagem do olho iluminado com IR (a) e binarização e centróide da pupila (b).

6.2. Processo de calibração e transformação das coordenadas

Uma vez que o algoritmo calcula o centro da pupila com rapidez, é necessário haver um processo inicial de calibração para possibilitar a previsão do ponto da tela para o qual o usuário olha e conseqüentemente posiciona o mouse. Esse procedimento torna possível a transformação das coordenadas do plano imagem da câmera, definido pelo vetor \tilde{S} , para o plano da tela, definido por \tilde{E}

$$\tilde{E} = \begin{pmatrix} x_e \\ y_e \end{pmatrix} \quad (6.2)$$

$$\tilde{S} = \begin{pmatrix} x_s \\ y_s \end{pmatrix} \quad (6.3)$$

Esse processo, ilustrado na Figura 6.4, é muito usado em restaurações de imagens e faz uso de pontos-âncora, presentes em ambas os planos, e que servem para determinação dos coeficientes das equações de interdependência entre as coordenadas. Existe um compromisso entre a precisão que se deseja obter e o número de pontos necessários para alcançar esse objetivo. Quanto maior a precisão, maior a quantidade de pontos, mais demorado o processo e maior a inconveniência para o usuário. Outra observação é a utilização ou não do método dos mínimos quadráticos para obtenção dos coeficientes. Se o número de coeficientes for igual ao número de pontos usados, temos um sistema fechado e as soluções dos pontos-âncora devem coincidir em ambos os planos. O método de mapeamento polar [52] também foi testado, porém apresentou pior precisão do que o procedimento de calibração com nove pontos, adotado aqui.

Com esta metodologia, o usuário é solicitado a olhar para nove pontos na tela, organizados em uma malha 3×3 . A tela tem resolução de 800×600 pixels, e os 9 pontos correspondem aos cruzamentos das linhas 20, 300 e 580 (pixels) com as colunas 20, 400 e 780 (pixels). Cada ponto \tilde{S} é mapeado em um ponto \tilde{E} que vai da origem da imagem ao centro da pupila. A transformação de \tilde{E} para \tilde{S} é modelada pelas funções quadráticas

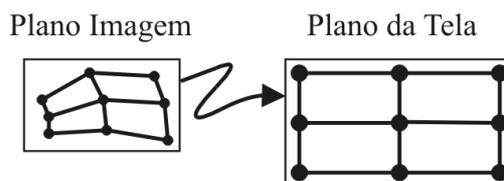


Figura 6.4 – Transformação espacial de coordenadas de imagem para tela.

$$x_s = a_1 + a_2x_e + a_3y_e + a_4x_ey_e + a_5x_e^2 + a_6y_e^2, \quad (6.4)$$

$$y_s = a_7 + a_8x_e + a_9y_e + a_{10}x_ey_e + a_{11}x_e^2 + a_{12}y_e^2. \quad (6.5)$$

Cada ponto \tilde{E} fornece duas equações para as duas coordenadas do ponto \tilde{S} e, portanto 18 equações são obtidas para 12 incógnitas. Os coeficientes polinomiais para as duas coordenadas de \tilde{S} podem ser obtidos independentemente, de modo que dois sistemas mais simples podem ser resolvidos usando o método dos mínimos quadráticos.

6.3. Algoritmos desenvolvidos

A Figura 6.5 mostra o fluxograma simplificado do algoritmo implementado em Matlab[®] R2007a. A funcionalidade do programa é simples e os passos relevantes são descritos a seguir:

1. Início do programa

2. Busca o dispositivo de captura, se falhar o programa é terminado. Inicializa a variável inteira $i=1$, que representa o número de imagens de calibração adquiridas. Mostra o padrão de calibração 3×3 (como descrito na seção 6.2) na tela.

3. Aguarda o pressionamento da tecla *enter*.

4. Quando a tecla *enter* é pressionada, o programa captura a imagem de calibração i e incrementa o contador ($i = i + 1$).

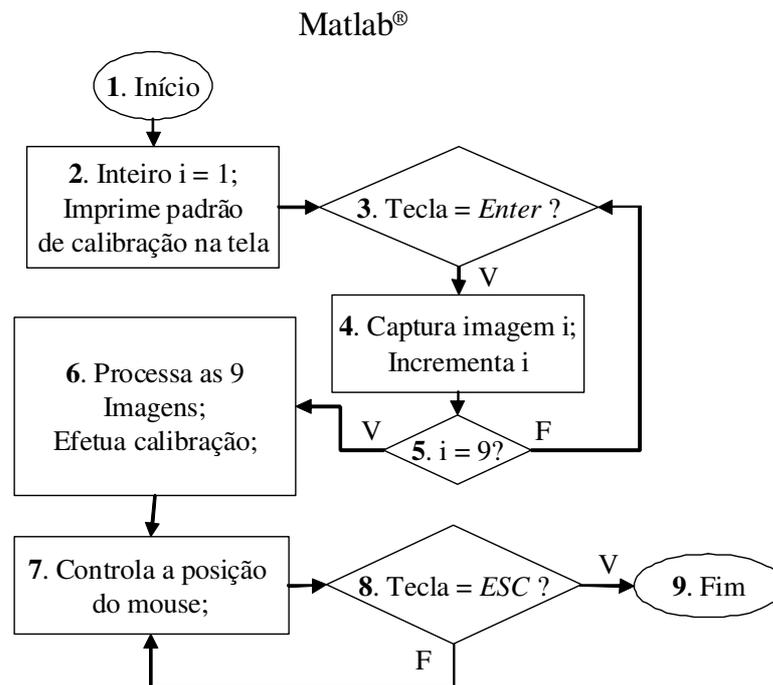


Figura 6.5 – Fluxograma do programa escrito em Matlab® R2007a.

5. Testa o número de imagens de calibração obtidas.

6. Se as nove imagens de calibração (9 correspondências Imagem \times Tela entre os pontos âncora como na disposição 3×3 na Figura 6.4) foram adquiridas, o programa obtém os coeficientes a_i como descrito na seção 6.2, se não, o programa volta a aguardar pela tecla *enter*.

7. Se a tecla *esc* é pressionada, então o programa é finalizado.

A limitação do programa escrito em Matlab fica clara quando se trata de flexibilidade do sistema. A calibração só pode ser efetuada uma vez a cada inicialização do programa. No caso em que os coeficientes a_i sejam mal calculados, o programa deve ser reinicializado. A interface não é amigável, visto que uma *interface gráfica para usuário* (GUI – *Graphic User Interface*) não foi implementada.

A Figura 6.6 apresenta o fluxograma do algoritmo implementado em C.

Os passos são descritos a seguir:

1. Início do programa. Declaração de variáveis globais. Dentre elas, a variável *selection*, do tipo *CvRect* (retângulo), que contém as coordenadas do retângulo que deve conter a pupila e excluir os ruídos provenientes dos cílios. A variável *selection* é inicializada com as coordenadas {50, 50, 100, 100} e pode ser ajustada interativamente

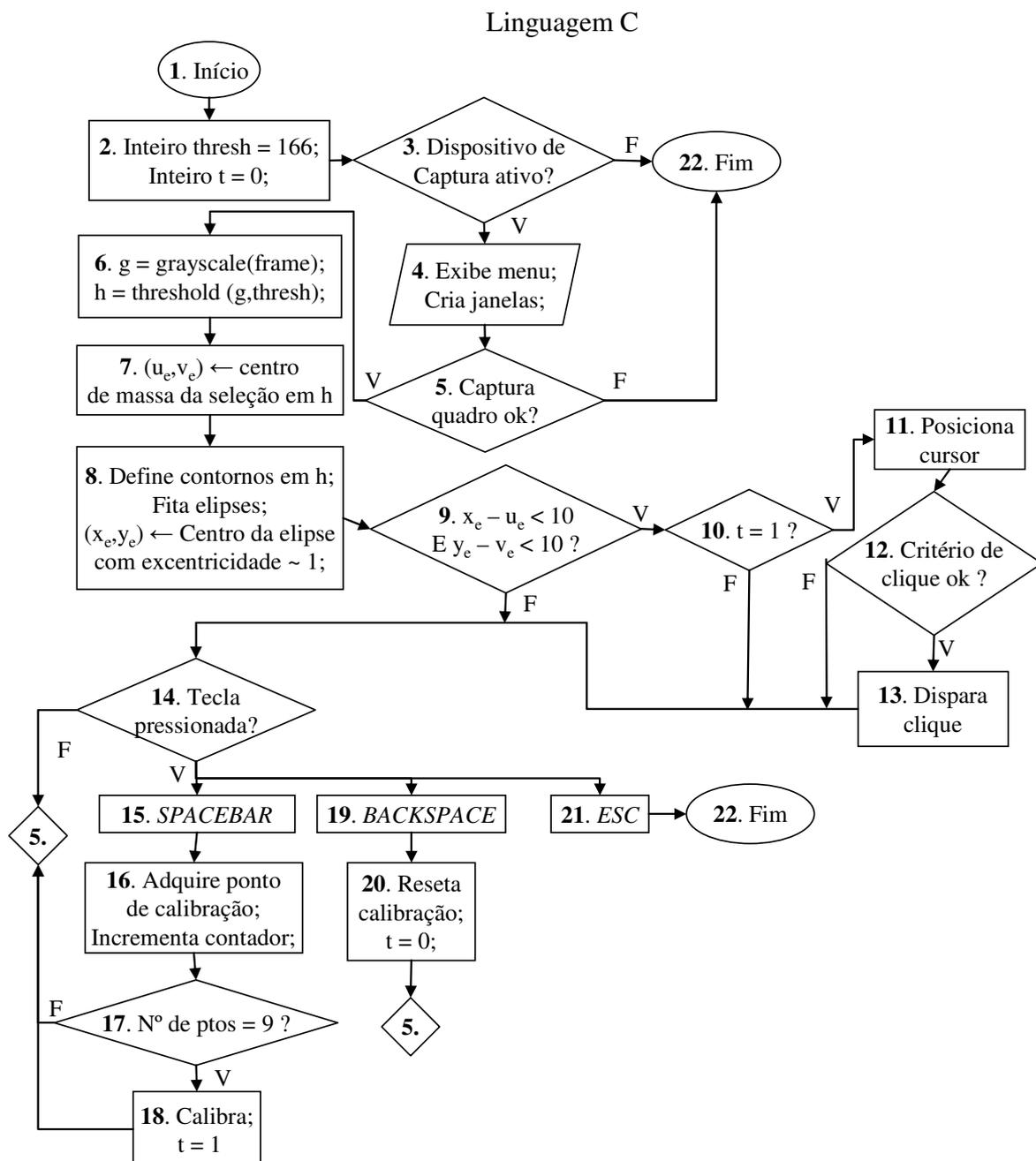


Figura 6.6 – Fluxograma simplificado do programa escrito em C.

pelo mouse, clicando e arrastando na janela que contém a imagem original. Isto pode ser feito durante qualquer momento durante a utilização do programa. Os algoritmos de centro de massa e ajuste de elipses são aplicados apenas no retângulo selecionado.

2. Declaração de variáveis locais. Dentre elas, o limiar de binarização *thresh* é inicializado com o valor 166 e pode ser ajustada interativamente através de uma barra de rolagem na janela da imagem binarizada. A variável *t* (de *toggle*) habilita/ desabilita o controle do mouse pelo algoritmo de processamento de imagens para $t = 1 / t = 0$. Ela é inicializada com o valor nulo.

3. Busca o dispositivo de captura (câmera). Se inativo finaliza o programa.

4. Exibe um menu no console contendo as teclas de atalho para controle do programa e inicializa as janelas que conterão as imagens.

5. Testa se houve sucesso na captura de um quadro (*frame*) no *stream* de vídeo da câmera USB. Se houver falha, o programa é finalizado.

6. Converte o *frame* capturado de RGB para escala de cinza e binariza, criando duas novas imagens *g* e *h*, respectivamente. O programa é incapaz de configurar o parâmetro “espaço de cor” da câmera por ser uma característica de somente leitura. Se o *stream* de vídeo fosse configurado para retornar imagens já em escala de cinza, a conversão de RGB poderia ser evitada, tornando o programa ligeiramente mais rápido.

7. Aplica o algoritmo de centro de massa descrito na seção 6.1 e obtém o centróide da pupila (u_e, v_e) .

8. Define os contornos de *h*, agrupa os contornos fechados e aplica o algoritmo *cvFitEllipse* em cada contorno fechado. Retorna o centro da elipse de excentricidade mais próxima de 1 nas coordenadas (x_e, y_e) .

9. Testa se as coordenadas (u_e, v_e) estão a uma distância algébrica menor que 10 pixels das coordenadas (x_e, y_e) . Se negativo, indica que a imagem binarizada contém muito ruído e o limiar *thresh* deve ser ajustado, e por isso o controle do mouse é desativado fazendo-se $t = 0$.

10, 11. Se o controle do mouse está ativo ($t = 1$), o ponteiro é posicionado de acordo com as relações (6.4) e (6.5). Este posicionamento é realizado após uma média entre 25 frames consecutivos, de forma a amenizar a intermitência do mouse ocular. Esta característica não está destacada no fluxograma, que é apresentado na forma simplificada.

12, 13. O critério de clique é satisfeito se o usuário permanece olhando fixamente para uma região quadrangular de aresta 40 pixels durante um intervalo de tempo (*default* de 0.75 segundo). Se este critério é satisfeito, um clique é disparado.

14. Retorna a tecla pressionada através da função *cvWaitKey*.

15, 16, 17, 18. Se o usuário pressiona a tecla Barra de Espaço, uma coordenada de calibração é adquirida. Este procedimento se repete até que 9 coordenadas sejam obtidas o sistema seja calibrado segundo o procedimento descrito na seção 6.2. Com a obtenção dos coeficientes a_i , o controle do mouse pode ser iniciado ($t = 1$) segundo as equações (6.4) e (6.5).

19, 20. Se *Backspace* é pressionada, as coordenadas e coeficientes de calibração são anulados, de modo que o processo de calibração é reinicializado.

21, 22. Se a tecla *Esc* é pressionada, o programa é finalizado.

6.4. Resultados

Implementação no Matlab® R2007a

A Figura 6.7-(a) mostra o plano da imagem do olho obtida com iluminação infravermelha e a Figura 6.7-(b) sua respectiva binarização e centróide da pupila. Levando

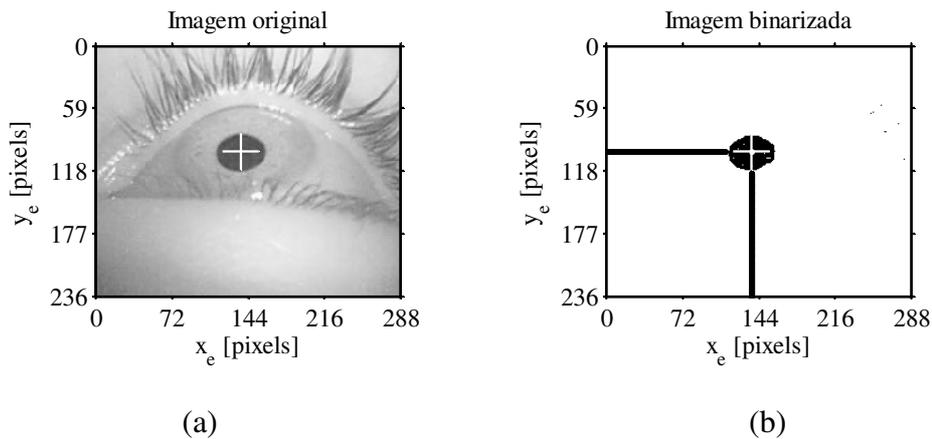


Figura 6.7 – Imagem do olho (a) e binarização e centróide da pupila (b).

em consideração a simplicidade do algoritmo, a implementação em Matlab forneceu um desempenho razoável de 18 fps [53]. Apesar de esta velocidade ser 4 vezes maior do que a obtida por Abu-Faraj *et al.*[54], que usa correlação sem iluminação infravermelha, a utilização do mouse a essa velocidade apresentou uma intermitência desconfortável em muitas aplicações, dificultando o uso de softwares de assistência como teclados virtuais, e browsers. A minimização desse problema é realizada através do uso de um filtro temporal que calcula a posição do mouse como a média de 3 medidas consecutivas do parâmetro \tilde{X}_e , o que reduz a velocidade do sistema para 6 fps. Ainda assim há um acréscimo de 50 % na velocidade alcançada em relação ao trabalho de Abu-Faraj *et al.* [54].

Pode se perceber, na Figura 6.7-(b), a presença de um pequeno ruído na binarização proveniente dos pixels escuros dos cílios pouco iluminados, cujos níveis de cinza são próximos aos níveis de cinza da pupila. Entretanto, esse ruído é significativo apenas no caso de fraca iluminação e foi retirado no algoritmo implementado em C com a implementação de uma função para análise de uma região mais restrita da imagem. Outra desvantagem do aplicativo em Matlab[®] foi a dificuldade de implementar uma função de acionamento do clique do mouse.

Como descrito anteriormente, após a efetiva resolução do centróide da pupila, uma calibração é realizada para obtenção dos coeficientes das relações quadráticas entre os vetores \tilde{X}_e e \tilde{X}_s . Essa calibração é realizada com os nove pontos definidos pelas interseções das linhas 20, 300 e 580, com as colunas 20, 400, 780, da tela com resolução de

800×600 pixels, como mostrado na Figura 6.8-(a). Após a realização da calibração com esses nove pontos, a precisão do sistema é testada com os 25 pontos, quando o usuário é solicitado a fixar seu olhar em cada um deles por aproximadamente 35 segundos (ou 100 estimativas para cada ponto). Essas estimativas são apresentadas na Figura 6.8-(a), na grade de 800×600 pixels da tela. Há um notável erro na última linha horizontal, onde as estimativas atingem altos valores negativos ($y \approx -50px$). Isso ocorreu nos testes iniciais devido à instabilidade da haste usada, que apresentava alta sensibilidade a vibrações mecânicas. Esse problema foi minimizado posteriormente com a utilização de outra haste mais rígida.

Cada ponto da Figura 6.8-(b) representa a distribuição de erros absolutos horizontal Δx e vertical Δy obtidos nos vários pontos da grade. Os valores médios obtidos foram $\overline{\Delta x} = -4px$ e $\overline{\Delta y} = 2px$, e os desvios padrão $\sigma_x = 14$ e $\sigma_y = 25px$.

Para se ter uma idéia prática da eficiência do rastreamento ocular, a Figura 6.9 mostra as coordenadas na tela estimadas pelo algoritmo quando o usuário é solicitado a ler um texto específico, mostrado em tela cheia. É notável o erro horizontal na previsão da direção do olhar para linhas na parte inferior na tela, decorrente da baixa qualidade da calibração efetuada com a haste de flexibilidade excessiva.

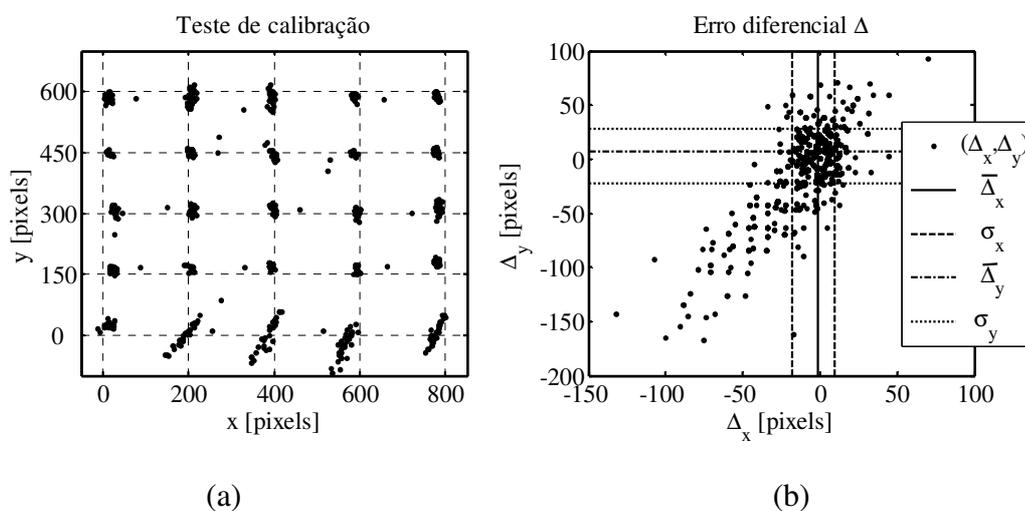


Figura 6.8 – Teste de calibração (a) e erro diferencial (b).

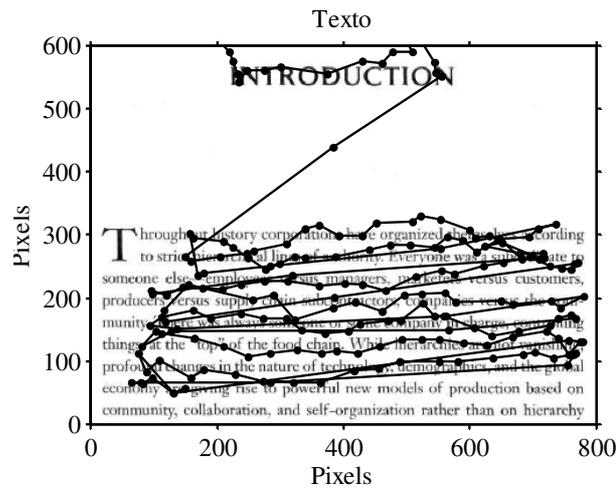


Figura 6.9 – Rastreamento ocular baseado na leitura de um texto.

Apesar de fornecer informação sobre a precisão do sistema, os erros horizontal e vertical são perceptivelmente maiores quando se trata do rastreamento ocular quando o usuário visualiza uma curva, devido à descontinuidade inerente ao movimento do olho. Uma maneira de mensurar o erro radial presente no rastreamento, é analisar as coordenadas estimadas pelo programa quando o objeto de interesse na tela é uma circunferência de raio específico, como mostrado na Figura 6.10-(a). Nesse teste, o usuário é solicitado a olhar fixamente para o mouse deliberadamente posicionado sobre a circunferência de raio 200 pixels. O mouse percorre em sentido horário e velocidade crescente, determinada pelo algoritmo. A Figura 6.10-(b) mostra as estimativas do POG em coordenadas polares (r, θ) , bem como a média e o desvio padrão radial $\bar{r} = 199$ e $\sigma_r = 31$, em pixels, respectivamente.

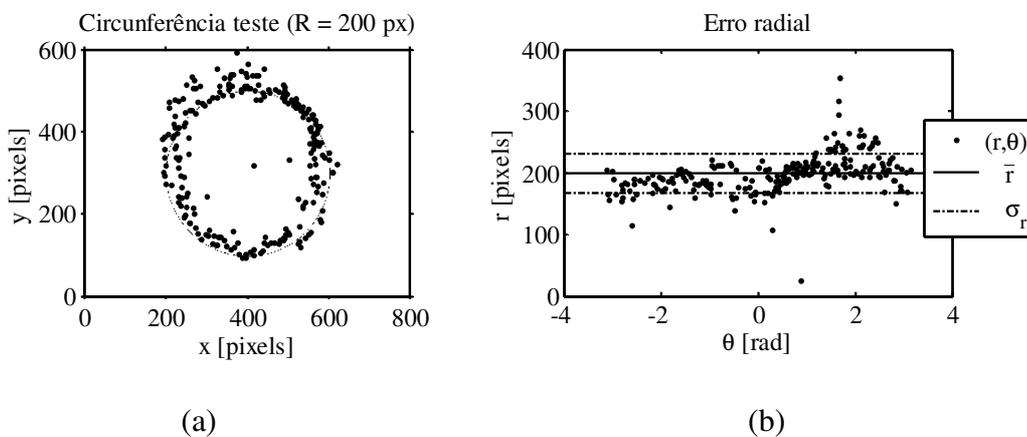


Figura 6.10 – Circunferência teste (R=200 pixels) (a) e coordenadas polares (r, θ) das medidas (b).

Os resultados das medidas descritas anteriormente são resumidos na Tabela 6.1, que apresenta os valores mínimo, médio, máximo e o erro quadrático médio dos erros horizontal, vertical e radial dos experimentos realizados e ilustrados na Figura 6.8-(b) e na Figura 6.10-(b). Apesar da falta de generalidade da unidade “pixels” (px), os resultados da Tabela 6.1 são mais bem compreendidos quando expressos nessa unidade, considerando os valores negativos que ultrapassam os limites do monitor, calculados pelo algoritmo de rastreamento. Não obstante, considerando as dimensões de 304.1×228.1 mm (0.38 milímetros/pixel para uma resolução de 800×600) do display efetivo do notebook segundo o fabricante, os valores em pixels podem ser obtidos na dimensão mais geral, milímetros.

Utilizando essa abordagem, outros testes idênticos foram realizados com 6 sujeitos (3 homens e 3 mulheres, com idades entre 27 e 56 anos). A Tabela 6.2, mostra os erros típicos do sistema para os resultados obtidos. Observa-se que o erro radial Δr é sempre maior do que os erros vertical e horizontal, uma vez que no caso radial a visualização é feita em velocidade crescente e, como discutido anteriormente, ocorrem saltos durante o rastreamento. A velocidade de rastreamento e o erro radial são diretamente proporcionais, e, à medida que a velocidade aumenta o erro radial Δr diverge do valor esperado geometricamente

$$\Delta r_g = \sqrt{\Delta x^2 + \Delta y^2} \quad (6.6)$$

com Δ_x e Δ_y representando os erros quadráticos médios associados às dimensões horizontal e vertical, respectivamente. Nos testes realizados, ficou notório que $\Delta r \approx \Delta r_g$

Tabela 6.1 – Resumo das estatísticas dos testes (todas as unidades em pixels).

Valor	Horizontal (Δ_x)	Vertical (Δ_y)	Radial (Δr)
Mínimo	- 132	- 167	-175
Médio	- 4.2	2.7	-1.1
Máximo	70	94	154
rms	14.001	25.077	31.080

quando o olhar rastreia a circunferência em baixa velocidade.

Os erros na Tabela 6.2, expressos em milímetros, possibilitam uma melhor definição da precisão do sistema. Com um erro radial máximo de ~24 milímetros é possível lançar mão de todas as ferramentas oferecidas pelo computador com apenas uma adaptação na acessibilidade do sistema operacional. O *assistente de acessibilidade* do Windows® XP possibilita um ligeiro aumento no tamanho dos ícones, barras de rolagem e de tarefa, fonte e barra de menus de janelas para deficientes visuais. Esses pequenos ajustes tornam possível o uso do sistema de forma confortável, mesmo com a leve intermitência do mouse ocular. Um defeito da implementação em Matlab® é a impossibilidade do disparo do clique do mouse, que foi obtido apenas com a posterior implementação em C, como explicado adiante.

Outra medida de confiabilidade é o erro angular na previsão do POG, dado por:

$$\Delta\alpha = \tan^{-1}\left(\frac{\Delta}{L}\right) \quad (6.7)$$

onde L é a típica distância de visualização entre a cabeça do usuário e o monitor e Δ é o erro quadrático médio (rms) na tela. Para $L \approx 50\text{cm}$ e levando em consideração os valores máximos da Tabela 6.2, obtém-se $\Delta\alpha \approx 1.2, 2.2$ e 2.7 graus para os erros angulares horizontal, vertical e radial, respectivamente.

Os erros, horizontal e vertical, obtidos acima são representativos de situações, por exemplo, em que o usuário lê um texto na tela do computador. Para situações mais gerais, em que o rastreamento de elementos na tela é mais aleatório, o erro obtido pelo rastreamento da circunferência com velocidade variável é mais representativo para se definir os limites de

Tabela 6.2 – Erros médios e valores máximos.

Erro	Horizontal Δ_x [mm]	Vertical Δ_y [mm]	Radial Δ_r [mm]
Médio	8.6	13	21.2
Desv. Padrão	1.6	4	2.1
Máximo	10.6	19.0	23.6

confiabilidade do sistema.

O desempenho obtido com o programa implementado em Matlab[®] é satisfatório. As desvantagens deste sistema podem ser organizadas nos seguintes tópicos:

(i) A cabeça do usuário deve permanecer estática durante todo o processo de calibração e utilização do sistema, pois a origem do sistema de coordenadas muda com o movimento. No caso de pacientes completamente imóveis, com deficiência motora severa, isso não seria um problema frequente, embora seja passível de ocorrer em convulsões, tosses do paciente, etc.

(ii) O Matlab[®] é um software restrito por ser de alto custo e seu funcionamento requer instalação de várias ferramentas com outros propósitos e alocação de memória que poderia ser utilizada para outros fins.

(iii) A implementação no Matlab[®] é altamente instável do ponto de vista de alocação de memória. Os testes realizados foram algumas vezes interrompidos por este motivo, com a consequente geração de um log de erro sobre alocação inadequada de memória cache. Este problema acarretou duas vezes em danos irreversíveis ao sistema operacional, que fizeram com que a máquina precisasse ser formatada.

Pelos motivos supracitados, a implementação deste mesmo procedimento foi realizada em C, bem como melhorias na obtenção do centróide da pupila, como detecção de elipses e definição da região de interesse para redução do ruído proveniente dos cílios.

Implementação em C

A implementação do programa de rastreamento ocular baseado na binarização da imagem original na linguagem de programação C permitiu diversas melhorias:

(i) A obtenção de um executável dedicado resolveu o problema de alocação inadequada de memória, possibilitando o uso ininterrupto do mouse ocular por diversas horas consecutivas. A instalação do sistema resume-se agora à instalação do driver da

câmera e das bibliotecas do tipo DLL abertas (*Open Source Computer Vision Libraries*) necessárias ao funcionamento do programa e disponibilizadas pela Intel[®].

(ii) O clique do mouse é implementado sem conflitos com drivers de mouses ópticos e/ou *mouse pad* paralelos que estejam sendo utilizados simultaneamente. O clique é disparado se o usuário mantiver seu olhar fixado numa região geométrica limitada (raio *default* de 20 pixels) por algum intervalo de tempo (*default* de 1.2 segundo). Esses parâmetros são facilmente ajustáveis através de atalhos no teclado, antes ou durante a utilização do sistema. A priori esses controles não são disponibilizados ao usuário, tendo em vista a simplicidade dos controles para o uso de um deficiente físico, e que esses valores padrão são razoáveis ao uso geral.

(iii) A velocidade de processamento é incomparável à obtida pelo Matlab[®], chegando a 300 fps inclusive para um *stream* de 2 câmeras captando em modo *stereo*, o que possibilita a utilização de algoritmos de PI bem mais sofisticados. Vale ressaltar que a velocidade máxima da câmera utilizada é de 30 fps quando funcionando à resolução de 640×480.

(iv) A definição da região de interesse e o limiar de binarização são feitos de modo interativo, *i.e.*, o usuário ou algum assistente, no caso de pacientes fisicamente incapazes, seleciona o limiar (de 0 a 255) e um retângulo que engloba os limites de movimento da pupila de forma a retirar o ruído proveniente dos cílios. Isto melhora sensivelmente a robustez do sistema com relação a este ruído.

O funcionamento do programa é ilustrado na Figura 6.11. O menu apresenta os comandos básicos que devem ser efetuados pelas teclas de atalho do teclado:

- **ESC – Sai do programa.**
 - **BARRA DE ESPAÇO – Adquire um ponto de calibração (de 1 a 9).**
 - **BACKSPACE – Inicializa/ reinicializa o processo de calibração.**
 - **T – Habilita/ desabilita o controle do mouse pelo PI. Quando desabilitado, o controle do mouse é entregue a outro dispositivo apontador presente no sistema.**
-

Além dos atalhos de controle, o console apresenta em tempo real as coordenadas \tilde{X}_e no plano imagem e coordenadas \tilde{X}_s estimadas, no caso de uma calibração já tiver sido efetuada. O console também apresenta os 9 pontos \tilde{X}_e de calibração adquiridos, bem como os coeficientes a_i obtidos, para verificação da qualidade da calibração. As três imagens distribuídas verticalmente na parte esquerda da tela mostram, respectivamente, a imagem original em escala de cinza, sua binarização, e a aproximação elíptica do contorno da pupila num close feito na região de interesse delimitada pelo retângulo branco. O usuário pode controlar o limiar de binarização e a região de interesse usando um mouse comum, previamente à utilização do sistema. Depois que estes valores são fixados, ajustes subsequentes são desnecessários.

A velocidade alcançada nos testes realizados foi de 14 fps com detecção de elipse. Esta velocidade é muito superior (~280 fps) quando o algoritmo de captação de um stream de vídeo é otimizado através da utilização de um conjunto de ferramentas disponíveis no

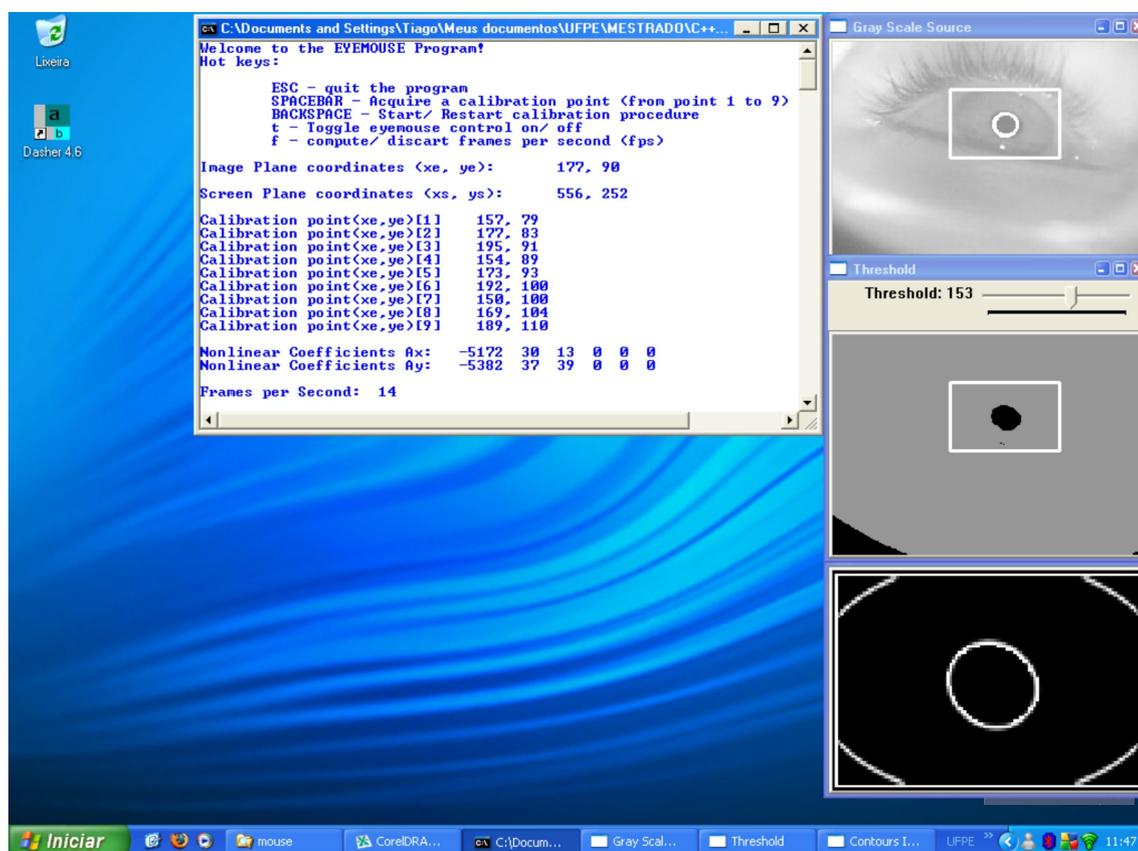


Figura 6.11 – Funcionamento do programa de rastreamento ocular.

cabeçalho *cvcam.h*, da biblioteca aberta de visão computacional (OPENCV) da Intel® [50]. Neste caso, pode-se melhorar a precisão do sistema utilizando-se as coordenadas \tilde{X}_s do mouse na tela como sendo resultante da média das coordenadas \tilde{X}_e de 25 frames seguidos, mesmo quando ocorrem frames repetidos. Assim, a velocidade do algoritmo se torna $280/25 \approx 11$ fps, que ainda é bastante satisfatória.

Com a mesma abordagem da aplicação realizada em Matlab®, os mesmos testes foram realizados para alguns usuários e os resultados analisados. Considerando o que foi implementado o mesmo processamento de imagem, a confiabilidade do sistema permaneceu equivalente à anterior, com uma melhoria devido ao aumento na resolução da tela de 800×600 para 1024×768 , como esperado. A Figura 6.12 mostra o teste de calibração, e o erro diferencial obtido das estimativas do POG, equivalentemente à Figura 6.8. Aqui, os nove pontos resultantes das interseções das linhas $(27,206,384,562,741)$, com as colunas $(27,269,512,755,997)$ são utilizados no método de calibração, e os 25 pontos mostrados na Figura 6.12-(a) apresenta os 25 pontos de teste. A Figura 6.12-(b) mostra os valores médios das diferenças horizontal $\overline{\Delta x}$ e vertical $\overline{\Delta y}$ entre as estimativas das coordenadas reais e as de tela, e os respectivos desvios padrão horizontal σ_x e vertical σ_y .

Com a mesma abordagem da seção anterior, o erro radial foi medido baseado no rastreamento em velocidade crescente de uma circunferência, como mostrado na Figura

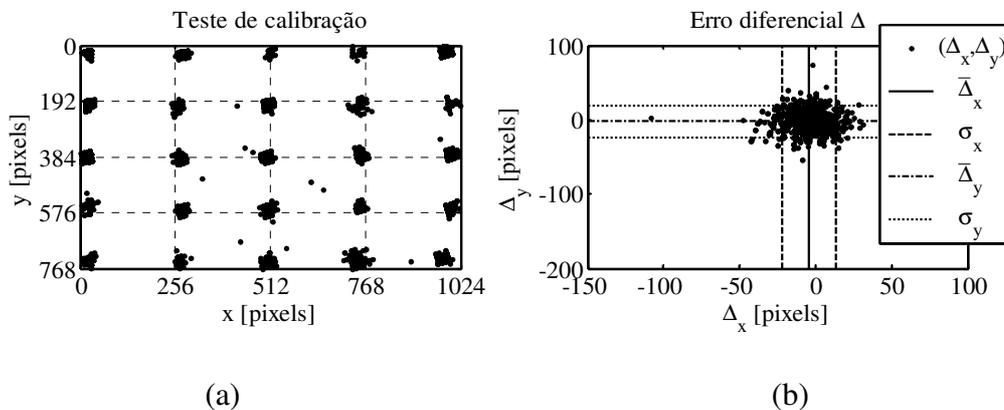


Figura 6.12 – Teste de calibração (a) e erro diferencial (b).

6.13. Desta vez, o raio é de 256 pixels, devido ao aumento na resolução da tela para 1024×768 . Devido a este aumento, os erros foram ligeiramente menores do que os obtidos anteriormente, como mostra a Tabela 6.3.

Levando em consideração os desvios padrão, mostrados na Tabela 6.3, e utilizando a equação de erro angular dada por (6.7), a relação de 0.30 milímetro/pixel (com a resolução de 1024×768 da tela), e uma distância típica de visualização de 50 cm, a implementação em C forneceu os desvios angulares $\Delta\alpha = 0,6^\circ$, $0,7^\circ$ e $0,9^\circ$ para os erros horizontal, vertical e radial, respectivamente. Novamente, o sistema possibilita a utilização de todas as ferramentas do computador quando é realizado um pequeno ajuste (aumento) no tamanho das fontes, menus e barras de rolagens com o assistente de acessibilidade do Windows[®] XP.

A detecção de elipses não apresentou melhoria na confiabilidade de detecção do centróide da pupila para um limiar bem determinado, como aquele mostrado na Figura 6.11, quando os centróides determinados pelo centro de massa da binarização e centro da elipse aproximada pelo contorno da pupila são praticamente idênticos. Sua aplicação é justificada para situações em que o limiar é, por algum motivo, deslocado do ideal, proporcionando a binarização de ruído proveniente da imagem original. A Figura 6.14 ilustra esse problema. A imagem original mostrada na Figura 6.14-(a) possui o centróide da pupila (circunferência branca) deslocado, devido à binarização de ruído proveniente dos cílios como mostra a Figura 6.14-(b), de forma que o algoritmo de centro de massa fica

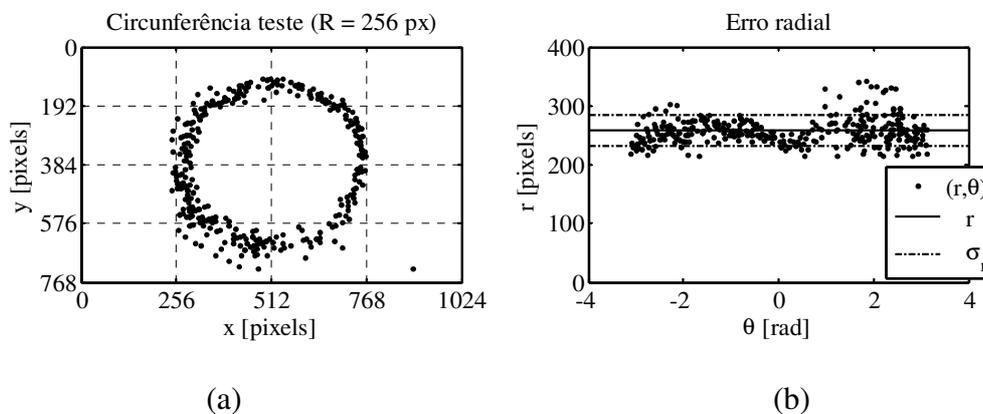


Figura 6.13 – Circunferência teste ($R=256$ pixels) (a) e coordenadas polares (r, θ) dos dados adquiridos (b).

Tabela 6.3 – Estatísticas dos testes realizados com 1 sujeito.

Valor	Erro radial [px]		Erro vertical [px]		Erro Radial [px]	
	Matlab [®]	C	Matlab [®]	C	Matlab [®]	C
Mínimo	-132	-261	-167	-270	-175	-42
Médio	-4.2	-4.3	2.7	-2.4	-1.1	2.2
Máximo	70	385	94	433	154	254
rms	14.001	17.57	25.077	20.901	31.080	26.903

prejudicado. Nesse caso, porém, a detecção de elipses mostra-se ideal para resolução do centróide, uma vez que os contornos fechados da Figura 6.14-(b) são aproximados por elipses, como mostra Figura 6.14-(c), e o centróide pode ser corrigido para o centro real da pupila. A elipse mais externa é sempre presente por que o algoritmo de aproximação de elipses considera a borda retangular (limites da imagem) na obtenção dos coeficientes da cônica. Esse problema pode ser resolvido forçando essa borda externa a assumir valores nulos, ou simplesmente desconsiderando os coeficientes constantes dessa elipse.

A Tabela 6.4 apresenta uma comparação entre as características principais dos programas desenvolvidos com o objetivo de destacar as vantagens e desvantagens de cada abordagem.

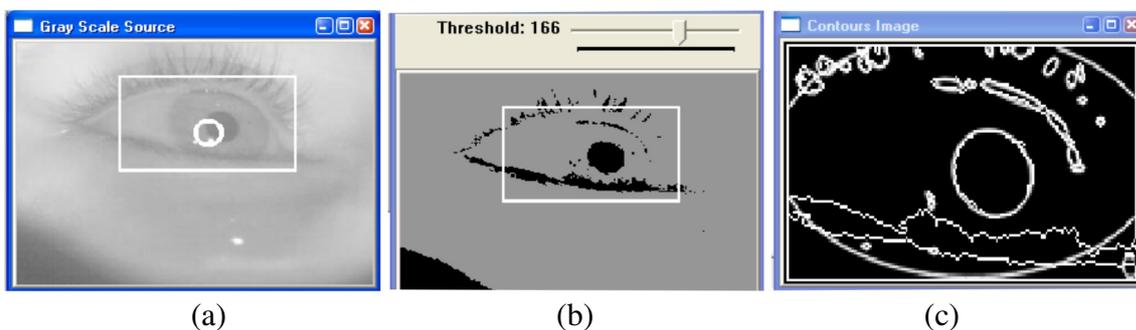


Figura 6.14 – Imagem original (a), binarização (b) e elipses aproximadas aos contornos fechados.

Algumas observações sobre as características listadas na Tabela 6.4:

- (i) O controle do clique do mouse não foi realizado em Matlab[®], e o controle da posição do apontador apresentou alta instabilidade devido ao conflito com o driver do mouse previamente instalado na máquina. O funcionamento do aplicativo resumia-se a poucos minutos de utilização antes de apresentar problemas.
- (ii) A velocidade com detecção de elipses em Matlab[®] não foi medida em tempo real, com uma imagem que apresentava apenas um contorno. No caso de imagens com múltiplos contornos a velocidade é ainda menor. Já em C a medida foi realizada em tempo real com processamento de imagem com múltiplos contornos, como mostra a Figura 6.14. No caso da utilização da ferramenta *cvcam.h*, que possibilita captação de 2 *streams* de vídeo no modo estéreo com velocidade de ~ 280 fps, com a média de 25 frames, essa velocidade chega a ~ 11 fps, que é muito mais rápida e precisa do que a obtida em Matlab[®].
- (iii) O ajuste interativo do *threshold* foi implementado em Matlab[®] mas também apresentou alta instabilidade.

Tabela 6.4 – Resumo das comparações relevantes às duas aplicações desenvolvidas.

Característica do aplicativo.	Matlab [®]	C
Estabilidade.	Péssima	Ótima
Controla a posição do mouse?	Sim	Sim
Controla o clique do mouse?	Não	Sim
Velocidade com detecção de elipses.	2 fps	14 fps
Ajuste interativo do <i>threshold</i> e da ROI	Não	Estável
Criação de um executável?	Não	Sim
Confiabilidade radial linear.	23.6 mm	8 mm
Confiabilidade radial angular.	2.7°	0.9°

(iv) O Matlab[®] possibilita a criação de executáveis em C a partir de seus códigos-fonte, porém, a suposição de que tais códigos seriam não otimizados implicou na implementação direta dos algoritmos em C.

(v) As confiabilidades dos dois sistemas podem ser diretamente comparadas mesmo com as diferenças nas resoluções dos planos da tela, que foram 800×600 em Matlab[®] e 1024×768 em C. Os erros são obtidos em milímetros levando em consideração o tamanho do monitor de 15 polegadas para efetiva comparação posterior.

7. SISTEMA DE RASTREAMENTO OCULAR 2 – SRO2

O sistema SRO2 visa realizar o rastreamento ocular de maneira remota, de forma a possibilitar alguma liberdade no movimento da cabeça do usuário. O sistema é formado por uma câmera acoplada ao monitor de modo que capta imagens do indivíduo. Duas fontes de iluminação infravermelha, cada uma contendo 8 LEDs (850 nm) são também acopladas ao monitor de forma a iluminar os olhos do usuário. Com este aparato, é possível implementar a metodologia usada por E. D. Guestrin, *et al.*[51], como abordado no Capítulo 5.

Neste capítulo os procedimentos iniciais para implementação do sistema SRO2 são tratados. Alguns resultados preliminares são mostrados, bem como os problemas encontrados e sugestões de soluções para os mesmos.

7.1. Procedimento de calibração

Na Seção 3.3 foi abordado um método direto de obtenção dos parâmetros da câmera a partir da matriz de projeção da mesma, e um método não-linear de otimização desses parâmetros foi citado. Como abordado na Seção 3.3, o método não-linear foi implementado através da utilização do *Toolbox* do Matlab[®], disponível em código aberto na *web* (*Câmera Calibration Toolbox for Matlab*[41]). A formulação matemática para obtenção desses parâmetros está fora do escopo deste trabalho e mais informações podem ser obtidas na referência [42].

Outra facilidade oferecida pelo *Toolbox* é a simplicidade na construção do padrão, que pode ser coplanar. O método direto necessita de um padrão não-coplanar (3D) para a calibração da câmera e os resultados são insatisfatórios. Com os dados do padrão 3D utilizado por *Heikkilä*[42], o método direto forneceu resultados absurdos por motivos desconhecidos.

O *Toolbox* utilizado neste trabalho proporciona a calibração da câmera a partir de várias imagens (20 foram utilizadas), de um padrão planar quadriculado, como mostrado na Figura 7.1. Essa imagem (7×5 quadrados internos) foi impressa de modo que todas as arestas medem 30 mm, permitindo que as coordenadas dos vértices (x,y,z) sejam facilmente obtidas a partir do conhecimento da origem do sistema de coordenadas de cena.

A calibração através do método utilizado necessita de mais de uma imagem para obtenção dos parâmetros. Para isso, 20 imagens foram obtidas a uma resolução de 640×480 , como mostra a Figura 7.2. A partir dessas imagens, podem-se obter as coordenadas em pixels dos vértices dos quadrados presentes no padrão, e fazer a correlação com as coordenadas de cena para efetiva realização da calibração da câmera.

A partir da Figura 7.2, o *Toolbox* obtém de forma interativa os quatro cantos externos dos pontos a serem calculados, *i.e.*, o usuário é solicitado a clicar nos vértices extremos, onde o 1º clique corresponderá à origem do sistema de coordenadas de cena, como mostra a Figura 7.3-(a), com esse ponto representado por O. A partir das escolhas desses cantos, as coordenadas internas (representadas por uma cruz “+”, na Figura 7.3-(a)) são obtidas pela simples divisão das arestas externas determinadas pelos 4 vértices fornecidos pelo usuário, pelo número de quadrados (contagem automática ou inserido pelo usuário) presentes nas direções horizontal e vertical. A partir dessas estimativas iniciais, as reais localidades dos vértices são obtidas através de processamento de imagem, em regiões

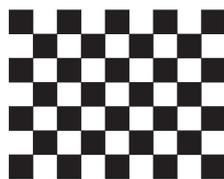


Figura 7.1 – Padrão planar utilizado na calibração.

Imagens de Calibração

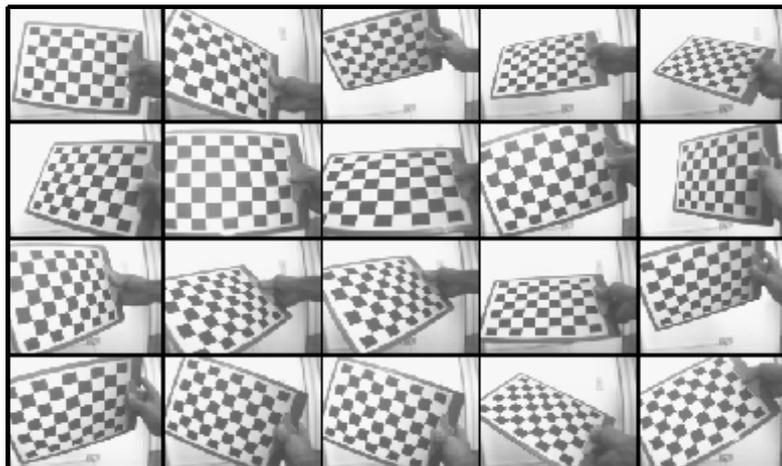


Figura 7.2 – Imagens utilizadas na calibração da câmera.

de análise quadrangulares (17×17 pixels), centradas nas estimativas iniciais, como mostra a Figura 7.3-(b). As coordenadas dos vértices assim obtidas são utilizadas no cálculo dos parâmetros da câmera. As coordenadas de imagem \tilde{X}_e na Figura 7.3-(a) são colineares, enquanto na Figura 7.3-(b), a distorção provocada pela lente é levada em consideração, de forma que, geralmente, esses pontos são não-colineares.

As coordenadas das 20 imagens obtidas levando em consideração a distorção

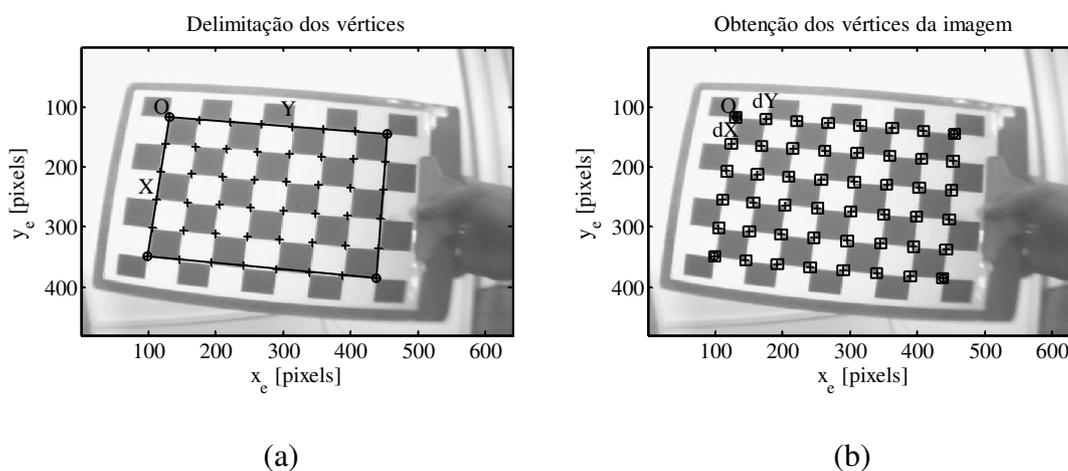


Figura 7.3 – Delimitação interativa dos vértices (a) e obtenção dos vértices por processamento de imagens (b).

proveniente da lente são utilizadas no algoritmo de otimização não-linear dos parâmetros através de iterações. A Tabela 7.1 mostra os parâmetros obtidos para 4 diferentes calibrações efetuadas com 20 imagens cada. São apresentados; a quantidade de iterações necessárias para convergência dos resultados, o comprimento focal f em milímetros, as coordenadas no plano imagem do ponto central $\tilde{C} = (u_0, v_0)$ em pixels, os coeficientes de distorção radial, k_1 , k_2 e k_5 e tangencial, k_3 , k_4 , presentes nas equações (3.19), (3.20) e os erros horizontal e vertical $(\Delta u, \Delta v)$ em pixels. O comprimento focal é fornecido pelo programa na forma de comprimentos focais efetivos (adimensionais) nas direções horizontal e vertical, dados respectivamente, por

$$f_u = \frac{f}{s_u} \quad (7.1)$$

e

$$f_v = \frac{f}{s_v} \quad (7.2)$$

onde s_u e s_v são os tamanhos dos pixels em milímetros nas direções horizontal e vertical, respectivamente. A razão entre os tamanhos horizontal e vertical dos sensores constituintes do CCD (*aspect ratio*) é definida por

$$ar = \frac{s_u}{s_v} = \frac{f_v}{f_u}. \quad (7.3)$$

Considerando a resolução da câmera Teclaser de 300k (ou 640×480 pixels), e as dimensões do sensor de 4×3 mm, cada pixel constituinte do sensor é, portanto, um quadrado, de forma que $s_u = s_v = 6.25 \mu\text{m}$ (valor típico para uma *webcam* comercial). O valor obtido a partir da relação (7.3) é, portanto $ar \approx 1$.

Na Tabela 7.1, os erros horizontal e vertical $(\Delta u, \Delta v)$, em pixels, são os valores rms das diferenças entre as coordenadas (u, v) estimadas inicialmente (sem distorção) e obtidas por processamento de imagem (com distorção), como ilustrado na Figura 7.3. A

confiabilidade do método é confirmada pela semelhança em todos os valores, para diferentes imagens de calibração. Os parâmetros obtidos estão muito próximos dos esperados para o modelo de câmera *pinhole*, com exceção do pequeno valor da coordenada horizontal u_0 do centro da imagem, que deveria ser próximo de 320 pixels, considerando uma resolução de imagem de 640×480 . Isso ocorreu por que a lente utilizada na calibração não foi a original da câmera, que foi trocada por outra, devido à presença de filtro de bloqueio do IR na primeira.

Os valores obtidos para os parâmetros extrínsecos que relacionam as coordenadas de cena (x,y,z) com as coordenadas de câmera (x',y',z') foram

$$\tilde{X}_0 = (-96,364274 \quad -48,342971 \quad 409,976656)^T, \quad (7.4)$$

$$\tilde{T} = \begin{pmatrix} -0,058725 & 0,991248 & -0,118228 \\ 0,704443 & -0,042767 & -0,708470 \\ -0,707327 & -0,124890 & -0,695767 \end{pmatrix}, \quad (7.5)$$

Tabela 7.1 – Parâmetros da câmera obtidos em 4 calibrações com 20 imagens.

Parâmetros \ Calibração	1	2	3	4
Iterações	21	16	18	18
<i>Aspect Ratio, Ar</i>	1.0098	1.0098	1.010	1.010
Comprimento focal, f [mm]	4.7	4.6	4.6	4.7
Ponto central, $\tilde{C} = (u_0, v_0)$ [pixels]	(251, 250)	(251, 250)	(251, 250)	(255, 250)
Coef. de dist. rad. k_1	-0.38804	-0.38482	-0.38586	-0.38888
Coef. de dist. rad. k_2	0.02151	0.00000	0.00000	0.00000
Coef. de dist. tang. k_3	-0.01035	-0.01037	-0.01001	-0.01008
Coef. de dist. tang. k_4	-0.00111	-0.00110	-0.00127	-0.00093
Coef. de dist. rad. k_5	0.00000	0.00000	0.00000	0.00000
Erro $(\Delta u, \Delta v)$ [pixels]	(0.22, 0.36)	(0.22, 0.36)	(0.22, 0.35)	(0.20, 0.36)

que permitem relacionar as coordenadas de câmera $\tilde{X}' = (x' \ y' \ z')^T$ e cena $\tilde{X} = (x \ y \ z)^T$ pela equação (3.3).

A unitariedade da matriz de rotação \tilde{T} é verificada a menos de um erro da ordem de $\approx 10^{-17}$, que pode ser desprezado. A Figura 7.4 mostra as coordenadas de cena (x,y,z) da imagem em questão (com o eixo z apontando no sentido da imagem para a câmera), e as coordenadas no plano imagem (u,v) , em pixels.

A Figura 7.5 mostra a correção de uma imagem distorcida, através do uso do modelo de distorção afim, abordado na Seção 3.2. A distorção radial é notável na parte inferior e na parte superior do padrão, na imagem original mostrada na Figura 7.5-(a), devido à distância ao centro. A Figura 7.5-(b) mostra a imagem original corrigida, onde o padrão se assemelha mais a uma figura realmente retangular.

A calibração da câmera é fundamental para obtenção das coordenadas de cena dos vetores \vec{u}_{ij} e \vec{v}_j referentes às imagens no sensor da câmera j do reflexo do LED i e do centro da pupila, respectivamente, segundo a discussão do Capítulo 5. Com a obtenção

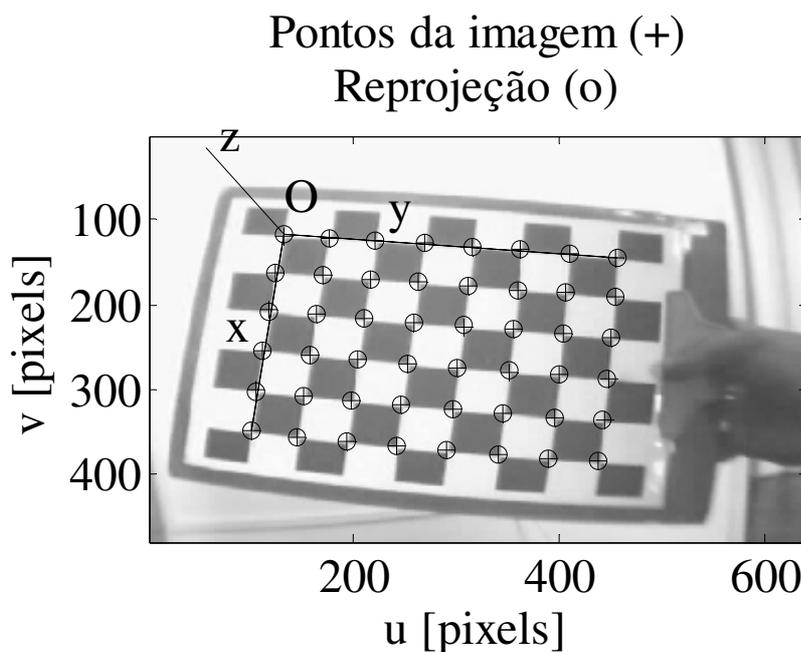


Figura 7.4 – Parâmetros extrínsecos para uma imagem.

dessas coordenadas e a abordagem matemática formalizada nesse capítulo é possível realizar o rastreamento ocular com liberdade de movimento de cabeça com a utilização de duas fontes de luz. A primeira implementação desse sistema é descrita a seguir.

7.2. Implementação do sistema SRO2 e resultados preliminares

Implementação inicial do sistema e análise preliminar de desempenho

O sistema SR2 implementado para a realização do rastreamento ocular com liberdade de movimento da cabeça do usuário está ilustrado na Figura 7.6 e é constituído de uma *webcam* USB Teclaser com velocidade de captura de 30 fps a uma resolução de 640×480 , e duas fontes de luz IR formadas por 8 LEDs cada, emitindo no comprimento de onda de 850 nm e com largura espectral de 55nm. As fontes IR são posicionadas nas posições laterais à tela, enquanto a câmera encontra-se abaixo dela, para permitir a captação de imagens da pupila sem a obstrução causada pelos cílios e pálpebras. A montagem foi realizada com uma moldura de plástico rígido, adaptada de uma moldura de um filtro de vidro para monitor comercial, de forma que as posições dos componentes podem ser ajustadas e fixadas de forma estável. A câmera foi montada numa estrutura fixa por parafusos e os LEDs podem ser ajustados por estarem presos a uma estrutura metálica (alumínio), para obtenção de uma iluminação satisfatória. O ajuste inicial é feito de forma que os eixos de iluminação das fontes de luz e o eixo óptico da câmera interceptam-se a uma distância típica de visualização do monitor, de aproximadamente 50 cm. Depois da

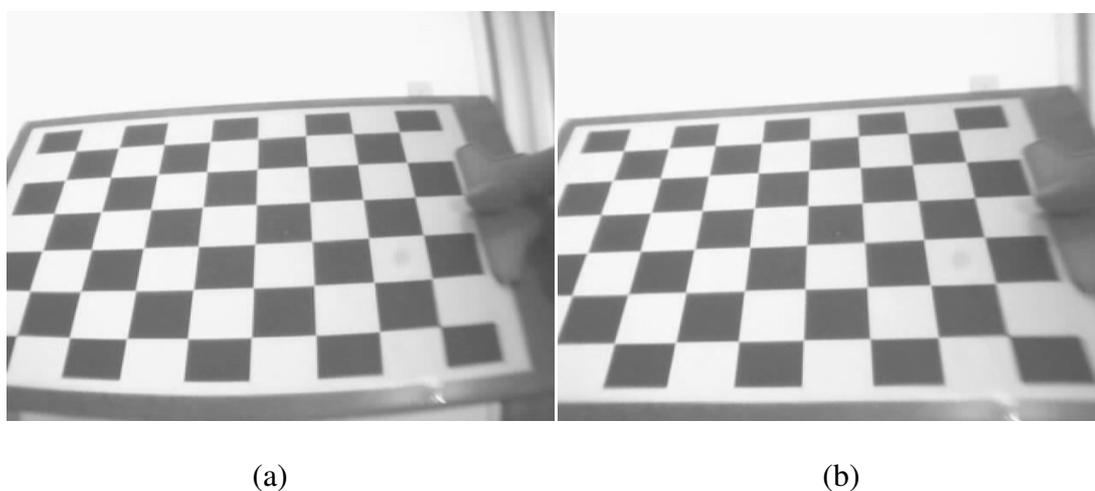


Figura 7.5 –Imagem original com distorção (a), e corrigida (b).

realização desse ajuste, o sistema é calibrado uma única vez segundo o procedimento tratado nas Seções 3.3, 7.1 e no Apêndice A, onde são obtidos os parâmetros da câmera e as posições das 2 fontes de luz, \vec{l}_1 e \vec{l}_2 , em relação ao sistema de cena com origem no monitor, como mostrado na Figura 7.6.

Após a calibração ter sido efetuada, foram obtidas imagens do olho fitando os quatro cantos da tela, com a cabeça na posição central e em posições levemente laterais, para uma realização prévia do processamento de imagens para obtenção do centro da pupila \vec{v} e dos centros dos reflexos \vec{u}_1 e \vec{u}_2 , das fontes de luz 1 e 2, respectivamente.

A Figura 7.7 apresenta 5 imagens do olho, quando a cabeça do usuário encontra-se na posição central de visualização do monitor, e esse usuário é solicitado a fitar os quatro cantos extremos da tela. Como a câmera trabalha com uma resolução de 640×480 , as sub-imagens de regiões menores do olho são retângulos de apenas 78×60 pixels.

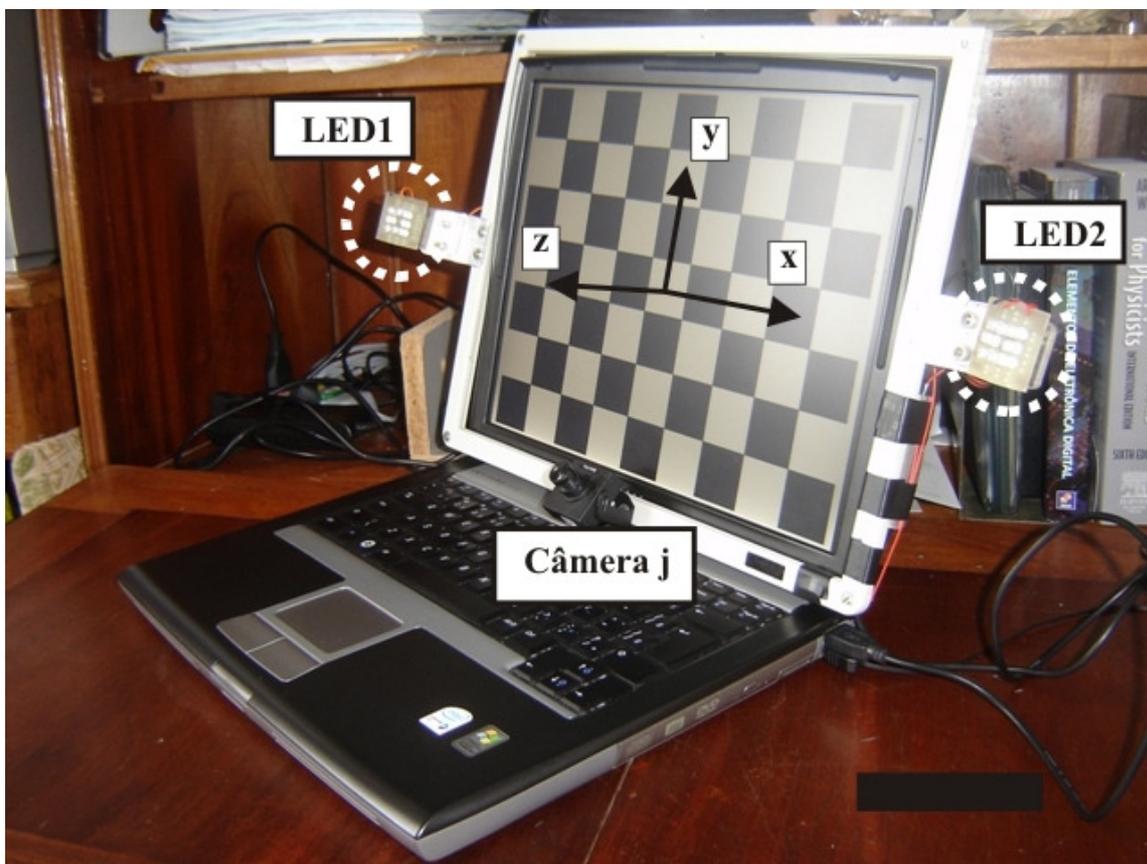


Figura 7.6 – Sistema para rastreamento ocular remoto.



Figura 7.7 – Imagens do olho fitando o centro e os quatro cantos da tela.

Alguns problemas decorrem da qualidade das imagens mostradas na Figura 7.7:

(i) A baixa resolução da imagem compromete a precisão do sistema, tornando-o mais susceptível a erro.

(ii) Os reflexos dos LEDs ocorrem com frequência fora da íris, passando para o globo ocular. Nessas situações a formulação matemática abordada na Seção 5.1 que leva em conta o raio de curvatura da córnea como base, passa a ser inválido. Esse problema é mais freqüente quando a posição da cabeça é levemente lateral.

(iii) Os níveis de cinza dos reflexos são próximos da saturação, como os de outras regiões do olho, o que dificulta a resolução dos centros apenas com limiarização. A utilização de filtros e detecção de bordas é necessária, bem como a técnica de detecção de elipses. Entretanto, devido ao problema (i), a presença de ruído devido à baixa resolução adiciona prejudica esses tipos de processamento.

(iv) Quando o usuário usa óculos de grau, a presença de múltiplas reflexões também é um problema que dificulta o processamento de imagens. No caso ilustrado na Figura 7.8, para excluir as múltiplas reflexões, pode-se considerar apenas aquelas duas mais próximas do

centróide da pupila. O problema maior é se, porventura, em um posicionamento extremo da cabeça, o reflexo venha a ocorrer sobre a própria pupila, retirando-se qualquer possibilidade de extração das coordenadas de interesse.

Devido aos problemas supracitados, o rastreamento ocular remoto utilizando o sistema SRO2 não foi implementado dentro do período de desenvolvimento deste trabalho. Não obstante, modificações no aparato SRO2 são propostas no Capítulo 8 como trabalhos futuros, para efetivação do sistema remoto e posteriores comparações entre diferentes sistemas, inclusive o que utiliza múltiplas câmeras (2 devem ser utilizadas), proporcionando a simplificação do processo de calibração para apenas um ponto, como proposto por Guestrin e Eizenman [55].



Figura 7.8 – Múltiplas reflexões decorrentes da utilização de óculos de grau.

8. CONSIDERAÇÕES FINAIS

*“Quem sabe concentrar-se numa coisa
e insistir nela como único objetivo,
obté, ao cabo, a capacidade de
fazer qualquer coisa”.*

Gandhi.

8.1. Conclusões

Neste trabalho foi desenvolvido o sistema SRO1 como uma solução de baixo custo e, portanto, passível de aplicação em larga escala, que possibilita o controle do mouse pelo olho, para pessoas que sofrem de deficiências motoras severas. As variáveis mais relevantes à funcionalidade do projeto foram consideradas, a saber; limite seguro de exposição ocular à radiação IR; técnicas utilizadas de PI; e formulação matemática que possibilita o rastreamento ocular remoto.

No sistema SRO1 o aplicativo desenvolvido em C executa a previsão do POG com eficácia e bom desempenho nos requisitos velocidade (possibilidade de alcançar 280 fps com médias de 25 observações ou 11 fps) e precisão (erro radial máximo de 8 mm). A desvantagem desse sistema é ser do tipo *head mounted* (HM), onde o olhar do usuário é obtido em relação à posição da cabeça, impossibilitando a liberdade de movimento desta. Em vista dessas limitações iniciamos a implementação do sistema de rastreamento remoto, SRO2, que extrai o POG em relação ao sistema de coordenadas de cena e, por esse motivo, oferece algum grau de liberdade no movimento da cabeça. Até o presente um primeiro protótipo foi montado e calibrado. Detectamos dificuldades em sua utilização quando os pontos de reflexão sobre a córnea das duas fontes utilizadas ficam situados fora da íris,

tornando-os imperceptíveis por processamento de imagens, impedindo, portanto a determinação do POG. Dificuldades adicionais também são causadas por múltiplas reflexões, quando o usuário é portador de óculos de grau.

Uma versão alternativa de realizar o rastreamento ocular ainda no sistema SRO1 desenvolvido é acrescentar uma câmera no monitor, que capte imagens da cabeça do usuário e, dois LEDs IR, um em cada lado dos óculos, apontando para a tela. Enquanto a câmera montada nos óculos capta imagens do olho, aquela montada no monitor obtém as imagens dos dois LEDs IR. Com a obtenção destes por processamento de imagem, é possível se obter a posição (e alguma inclinação) da cabeça do usuário, sabendo-se a distância [mm] entre os LEDs. Este sistema chegou a ser testado, mas não ofereceu bons resultados. Essa abordagem encarece um pouco a montagem por conter duas câmeras, e, foi descartada pela motivação principal de se obter um sistema no qual o usuário não precisa suportar nenhum aparato físico, característica extremamente desejável, sobretudo para utilização por pessoas portadoras de deficiências motoras severas.

8.2. Sugestões para trabalhos futuros

Para continuidade do trabalho desenvolvido, sugere-se:

- (i) Levando em consideração uma abordagem científica, com o intuito de se obter informação sobre o desempenho de diferentes sistemas, sugere-se desenvolver uma modelagem matemática para o rastreamento ocular com um sistema composto de duas fontes pontuais (LEDs) incorporados aos óculos de suporte e mais a câmera no monitor. Esse sistema, apesar pode fornecer uma precisão maior do que o sistema remoto SRO2, devido à confiabilidade na extração da posição da cabeça do usuário proporcionada pelos dois marcadores luminosos presentes nos óculos. Além disso, interferências devido a múltiplas reflexões ou desaparecimento dos pontos de reflexão na córnea, observados experimentalmente no sistema SRO2, seriam evitadas com essa nova configuração.
 - (ii) Modificação do sistema SRO2 com a substituição da câmera atual por outra de melhor resolução e maior sensibilidade. Além disso, incorporar um sistema eletrônico de chaveamento dos LEDs para permitir a captura de imagens do tipo “pupila-clara” e
-

“pupila-escura” de forma a simplificar consideravelmente a obtenção do centróide da(s) pupila(s). Com essas modificações no sistema SRO2, será possível realizar o rastreamento ocular remoto, com a vantagem de se desenvolver um sistema que não seja do tipo *head mounted*.

(iii) Desenvolver um sistema que utilize única e exclusivamente uma câmera do próprio computador e incorporar novas técnicas de reconhecimento de formas e algoritmos rápidos para permitir o rastreamento ocular sem a adição de qualquer componente na máquina.

APÊNDICE A – PROCEDIMENTO DE CALIBRAÇÃO DO SISTEMA SRO2

A Figura A.1 mostra a configuração experimental do sistema SRO2. As fontes de luz são compostas por dois conjuntos de 8 LEDs cada, emitindo no comprimento de onda de 850 nm e com largura espectral de 55 nm. A câmera é analógica CCD, TECVOZ, monocromática com 420 linhas horizontais de resolução (apesar de a maioria dos testes terem sido feitos com a câmera USB TECLASER a 640×480 de resolução a 30 fps). A Figura A.1 mostra também o padrão impresso no monitor usado na calibração do aparato, onde o centro da tela de 15 polegadas é considerado o centro do eixo de coordenadas de cena (x,y,z) , de forma que o plano xy coincide com o plano da tela e o eixo z é perpendicular e orientado para fora dele.

Para calibração do sistema, é preciso usar uma câmera auxiliar a para a obtenção da posição da câmera j em relação ao sistema de coordenadas do *display* (monitor). Essa câmera auxiliar é necessária por que a câmera j não capta imagens do padrão virtual impresso no *display* que determina o sistema cena (x,y,z) .

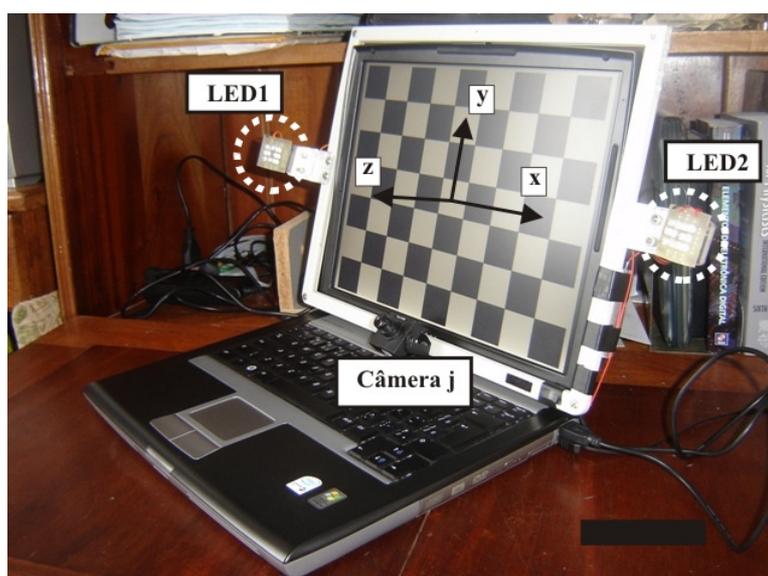


Figura A.1 – Montagem do sistema no notebook.

Seguindo o procedimento proposto Guestrin e Eizenmann [55] processo de calibração do aparato físico formado pela câmera j e pelos LEDs 1 e 2, ilustrado na Figura A, é feita de acordo com os seguintes passos:

PASSO 1 – As câmeras j e a são calibradas com o uso do padrão físico quadriculado. Cada quadrado do padrão mede 30 mm de lado e é impresso frente e verso como mostrado na Figura A.2-(a) e fica posicionado entre as câmeras. Dessa forma, ambas são calibradas segundo o sistema de coordenadas de cena 1 (x_1, y_1, z_1) , de acordo com a formulação tratada na 3.3. Assim, são obtidas as matrizes de rotação \tilde{T}_a e translação \tilde{Y}_a que relacionam o sistema de cena (x_1, y_1, z_1) com o sistema (x'_a, y'_a, z'_a) da câmera a , bem como as matrizes de rotação \tilde{T}_j e translação \tilde{Y}_j , que relacionam o sistema da cena com o sistema (x'_j, y'_j, z'_j) da câmera j . Utilizando a relação de transformação (3.3) tem-se

$$\tilde{X}_a = \tilde{T}_a \tilde{X}_1 + \tilde{Y}_a, \quad (\text{A.8.1})$$

$$\tilde{X}_j = \tilde{T}_j \tilde{X}_1 + \tilde{Y}_j, \quad (\text{A.8.2})$$

com

$$\tilde{X}_a = (x'_a \quad y'_a \quad z'_a)^T, \quad (\text{A.3})$$

$$\tilde{X}_1 = (x_1 \quad y_1 \quad z_1)^T, \quad (\text{A.4})$$

$$\tilde{Y}_a = (x_{a0} \quad y_{a0} \quad z_{a0})^T, \quad (\text{A.5})$$

$$\tilde{X}_j = (x'_j \quad y'_j \quad z'_j)^T \quad (\text{A.6})$$

e

$$\tilde{Y}_j = (x_{j0} \quad y_{j0} \quad z_{j0})^T. \quad (\text{A.7})$$

Inserindo (A.1) em (A.2) permite obter a relação entre as coordenadas das câmeras a e j ,

$$\tilde{X}_j = (\tilde{T}_j \tilde{T}_a^{-1})(\tilde{X}_a - \tilde{Y}_a) + \tilde{Y}_j \quad (\text{A.8})$$

PASSO 2 – O padrão é retirado do ambiente e é exibido em tamanho real no monitor, de forma que o centro do monitor seja agora a origem de um novo sistema de coordenadas de cena (x,y,z) , como mostra a Figura A.2-(b). As posições dos LEDs em relação ao centro desse novo sistema são medidas fisicamente com trenas, réguas etc. As matrizes de rotação \tilde{T}_2 e de translação \tilde{Y}_2 , que relacionam o novo sistema de cena e a câmera auxiliar a , podem ser obtidas a partir da calibração dessa câmera segundo esse novo padrão. Inserindo a relação entre $\tilde{X} = (x \ y \ z)^T$ e $\tilde{X}_a = (x'_a \ y'_a \ z'_a)^T$,

$$\tilde{X}_a = \tilde{T}_2 \tilde{X} + \tilde{Y}_2, \quad (\text{A.9})$$

em (A.8) resulta em

$$\tilde{X}_j = \tilde{T} \tilde{X} + \tilde{X}_0, \quad (\text{A.10})$$

onde,

$$\tilde{T} \equiv \tilde{T}_j \tilde{T}_a^{-1} \tilde{T}_2 \quad (\text{A.11})$$

e

$$\tilde{X}_0 = (\tilde{T}_j \tilde{T}_a^{-1})(\tilde{Y}_2 - \tilde{Y}_a) + \tilde{Y}_j. \quad (\text{A.12})$$

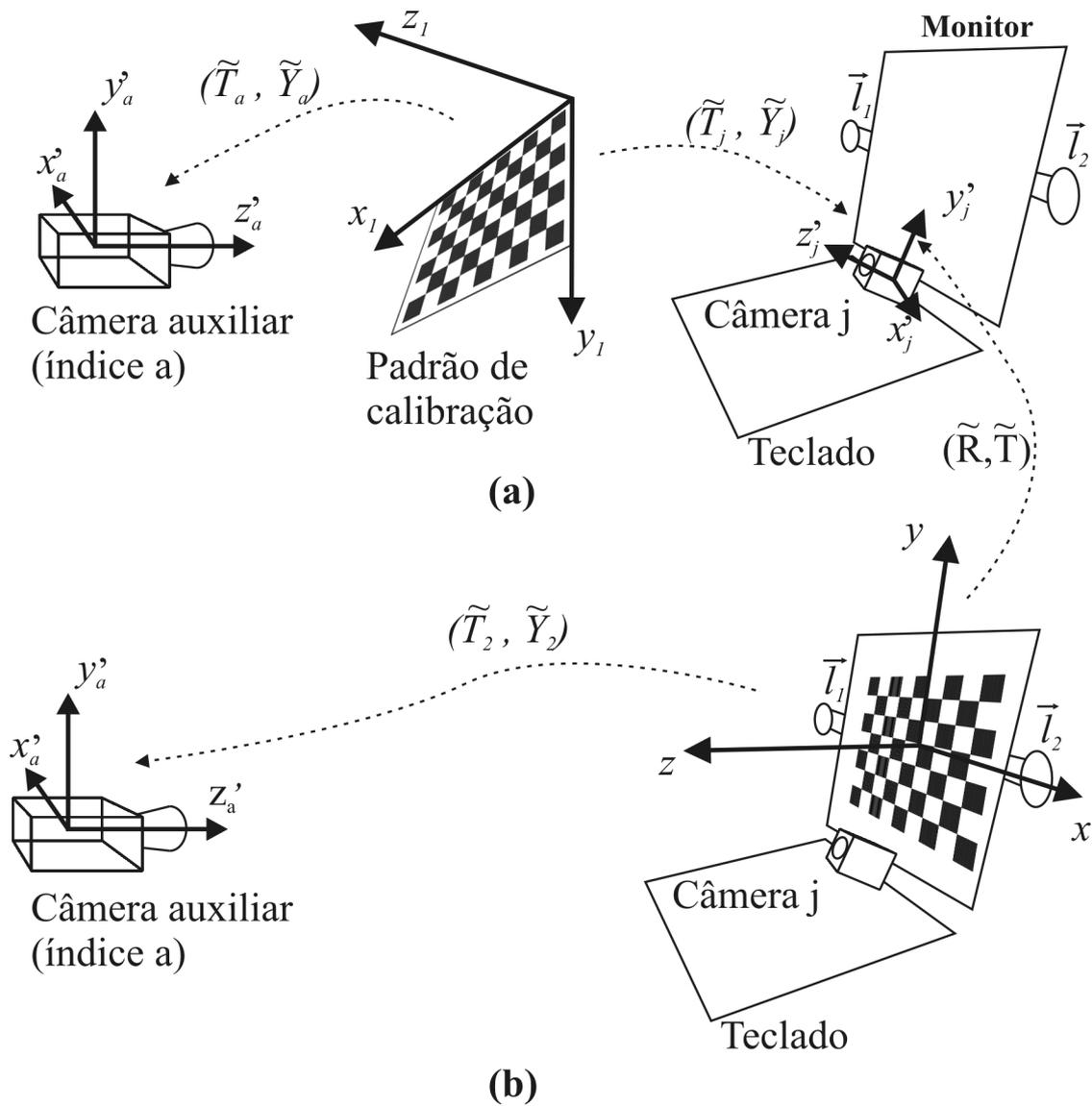


Figura A.2 – Esquema de calibração da montagem. Obtenção da relação entre as câmeras (a), e calibração final da câmera j e, relação ao sistema (x,y,z) (b).

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] P. Allain *et. al.*, “*Cognitive functions in chronic locked-in syndrome: a report of two cases*”. *Cortex*, Vol. 34, pp. 629-634 (1998).
- [2] J. R. Bach & J. A. Delisa, “*Evaluation and rehabilitation of patients with adult motor neuron disease*”. *Archives of Physical Medicine and Rehabilitation*, Vol. 80, pp. 951-963 (1999).
- [3] D. Bauby, “*The diving bell and the butterfly*”. Vintage Books, New York (1998).
- [4] S. W. Hawking, “*The Universe in a Nutshell*”. Bantam Press (2001).
- [5] Amiotrofic Lateral Sclerosis and Motor Neurone Disease Alliance, <http://www.alsmndalliance.org/>, (11 de Janeiro de 2009).
- [6] Dodge & Cline, “*The angle velocity of eye movements*”. *Psychological Review* 8, 145-157 (1901).
- [7] C.H. Judd & C. N. McAllister, “*General introduction to a series of studies of eye movements by means of kinoscopic photographs*”. *Psychological Review*, 7:1-16, The Review Publishing Company, Baltimore (1905).
- [8] P. M. Fitts, R. E. Jones & J. L. Milton, “*Eye movements of aircraft pilots during instrument-landing approaches*”. *Aeronautical Engineering Review*, 9(2), 24-29 (1950).
- [9] H. Hartridge & L. C. Thompson, “*Methods of investigating eye movements*”. *British Journal of Ophthalmology*, 32:581-591 (1948).
- [10] J. F. Mackworth & N. H. Mackworth, “*Eye fixations recorded on changing visual scenes by the television eye-marker*”. *Journal of the Optical Society of America*, 48:439-445 (1958).
- [11] Cornsweet & Crane, “*Accurate two-dimensional eye tracker using first and fourth Purkinje images*”. *Journal of the Optical Society of America*, 63:921-928 (1973).
- [12] R. A. Monty, “*An advanced eye-movement measuring and recording system*”. *American Psychologist* (1975).
- [13] J. L. Levine, “*An Eye-Controlled Computer*”. Research Report RC-8857, IBM Thomas Watson Research Center, Yorktown Heights, N.Y (1981).
- [14] C. Karatekin & R. F. Asarnow, “*Exploratory eye movements to pictures in childhood-onset schizophrenia and attention-deficit/hyperactivity disorder (ADHD)*”. *J. Abnorm. Child Psychology*, vol. 27, no. 1, pp.35–49 (1999).
-

- [15] J. G. Morrison, S. P. Marshall, R. T. Kelly & R. A. Moore, “*Eye Tracking in Tactical Decision Making Environments: Implications for Decision Support Evaluation*”. 3rd International Command and Control Research and Technology Symposium, National Defense University (1997).
- [16] P. A. Wetzel, G. Krueger-Anderson, C. Poprik, & P. Bascom, “*An Eye Tracking System for Analysis of Pilots’ Scan Paths*”. United States Air Force Armstrong Laboratory, Tech. Rep. AL/HR-TR-1996-0145 (1997).
- [17] S. Djamasbi, T. Tullis, “*Gender Preferences in Web Design: Usability Testing through Eye Tracking*”. America Conference on Information Systems, Pgs 1-8.
- [18] A. Bojko, “*Using Eye Tracking to Compare Web Page Designs: A Case Study*”. Journal of Usability Studies, Issue 3, Vol. 1, May 2006, pp. 112-120 (2006).
- [19] I. Haritaoglu *et al.*, “*Attentive Toys*”. IBM Almaden Research, San Jose, CA 95120, USA and Computer Vision Laboratory University of Maryland, College Park, MD 20742.
- [20] C. R. Cassemiro & C. G. Arce, “*Comunicação visual por computador na esclerose lateral amiotrófica*”. Arquivos Brasileiros de Oftalmologia, 67:295-300 (2004).
- [21] L. A. Frey, K. P. White, & T. E. Hutchinson, “*Eye-gaze word processing*”. IEEE Transactions on Systems, Man and Cybernetics, vol. 20, no. 4, pp. 944–950 (1990).
- [22] David A. Robinson, “*A Method of Measuring Eye Movement Using a Scleral Search Coil in a Magnetic Field*”. IEEE Transactions on Biomedical Electronics, issue 1963, Ps. 137-145 (1963).
- [23] A. F. Müller *et al.*, “*Dispositivo para emulação de mouse dedicado a pacientes tetraplégicos ou portadores de doença degenerativa do sistema neuromuscular*”. IX Congresso Brasileiro de Biomecânica, Gramado, Vol. 1. p. 66-71 (2001).
- [24] Abu-Faraj, O. Ziad *et al.*, “*Design and Development of a Low-Cost Eye Tracking System for the Rehabilitation of the Completely Locked-In Patient*”. Proceedings of the 28th IEEE EMBS Annual International Conference New York (2006).
- [25] C.H. Morimoto, D. Koons, A. Amir, M. Flickner, “*Frame-Rate Pupil Detector and Gaze Tracker*”. Dept. Ciência da Computação IME/USP (1999).
- [26] Eye Tech Digital Systems, Inc., <http://www.eyetechds.com/assistivetech/index.htm>, (23 de Março de 2009).
- [27] Tobii Technology, AB, http://www.tobii.com/scientific_research/home.aspx, (23 de Março de 2009).
-

- [28] <http://www2.uol.com.br/aprendiz/guiadeempregos/eficientes/noticias/ge150405.htm>, (23 de Março de 2009).
- [29] M. M. Aires, “*Fisiologia*”, Ed. Guanabara Koogan, (1991).
- [30] D. Halliday, R. Resnick, “*Fundamentals of Physics*”, Vol I, II, III & IV. 7th Ed.
- [31] G. Westheimer, “*Medical Physiology*”. V. B. Mountcastle, Vol. 1, 14^a Ed., St. Louis, Mosby, (1980).
- [32] J. Tombran-Tink & C. J. Barnstable, “*Visual Transduction and Non-Visual Light Perception*”. Humana Press.
- [33] M.H. Niemz, “*Laser Tissue Interactions*”. 3^a Ed., Springer.
- [34] International Commission on Non-Ionizing Radiation Protection, “*ICNIRP Statement on Light-Emitting Diodes (Leds) and Laser Diodes: Implication for Hazard Assessment*”. Health Physics Society (2000).
- [35] International Electrotechnical Commission, “*Safety of Laser Products*”. IEC, International Standard IEC 60825-1.1 (1998).
- [36] American National Standards Institute, “*American National Standard for the Safe Use of Lasers and Leds Used in Optical Fiber Transmission Systems*”, Orlando, FL: Laser Institute of America; ANSI Z136.2-1988; 1988 (1998).
- [37] D. Litwiller, “*CCD vs. CMOS: Facts and Fiction*”. Photonics Spectra (2005).
- [38] H. Yamagishi, Y. Suzuki & A. Hiraid, “*Precise Measurement of Photodiode Spectral Responses Using the Calorimetric Method*”. IEEE Transactions on Instrumentation and Measurement, VOL 38, NO 2 (1989).
- [39] M. Shah, “*Fundamentals of Computer Visison*”. Computer Science Department, University of Central Florida, Orlando, FL 32816, (1997).
- [40] E. Trucco & A. Verri, “*Introductory Techniques for 3D Computer Vision*”. Upper Saddle River, NJ: Prentice-Hall, pp. 123–138 (1998).
- [41] Jean-Yves Bouguet, <http://www.vision.caltech.edu/bouguetj/index.html> (06 de Março de 2009).
- [42] J. Heikkilä & O. Silvén, “*A Four-step Camera Calibration Procedure with Implicit Image Correction*”. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1106-1112, San Juan (1997).
- [43] S. W. Shih, Y. P. Hung & W. S. Lin, “*Accurate linear technique for camera calibration considering lens distortion by solving an eigenvalue problem*”. Optical Engineering 32(1): 138-149 (1993).
-

- [44] J. Weng, P. Cohen, & M. Herniou, “*Camera calibration with distortion models and accuracy evaluation*”. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-14(10): 965-980 (1992).
- [45] R. C. Gonzalez & Richard E. Woods, “*Digital Image Processing*”, 2nd ed., Prentice Hall.
- [46] P. V. C. Hough, “*Methods and Means for Recognizing Complex Patterns*”. U.S. Patent 3.069.654.
- [47] T. Peng, “*Circular Hough Transform Algorithm*”. Dept. of Mechanical Engineering, University of Maryland, College Park, Maryland 20742, USA pengtao@glue.umd.edu.
- [48] W. Gander, G. H. Golub & R. Strebler, “*Least-Squares Fitting of Circles and Ellipses*”. BIT 34(1994), 558–578.
- [49] H. Radim & J. Flusser, “*Numerically Stable Direct Least Squares Fitting of Ellipses*”. Dept. of Software Engineering, Charles University, Malostranské, Prague, Czech Republic.
- [50] “*Open Source Computer Vision Library*”, Reference Manual. Copyright © 2001 Intel Corporation All Rights Reserved Issued in U.S.A. Order Number: A77028-004 World Wide Web: <http://developer.intel.com/>.
- [51] E. D. Guestrin, & M. Eizenman, “*General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections*”. IEEE Transactions on Biomedical Engineering, Vol. 53, No. 6 (2006).
- [52] C. S. Lin *et al.*, “*Polar coordinate mapping method for an improved infrared eyetracking system*”. Biomedical Engineering Applications, Basis & Communications Vol. 17 No. 3, pp. 141-146 (2005).
- [53] T. F. Vieira & E. Fontana, “*Dispositivo de Rastreamento de Movimentos Oculares Baseado em Webcam e Iluminação com LED Infravermelho*”, 21º Congresso Brasileiro de Engenharia Biomédica, CBEB (2008).
- [54] Abu-Faraj, O. Ziad *et al.*, “*Design and Development of a Low-Cost Eye Tracking System for the Rehabilitation of the Completely Locked-In Patient*”. 28th IEEE EMBS Annual International Conference New York, Aug 30-Sept 3 (2006).
- [55] E. D. Guestrin, & M. Eizenman, “*Remote Point-of-Gaze Estimation with Free Head Movements Requiring a Single-Point Calibration*”. 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France (2007).
-