

INSTITUTO FEDERAL DO ESPÍRITO SANTO  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**GABRIEL CARDOSO DE AVELAR**

**UTILIZAÇÃO DE VISÃO ESTEREOSCÓPICA PARA RECONSTRUÇÃO 3D**

Cachoeiro de Itapemirim

2023

**GABRIEL CARDOSO DE AVELAR**

**UTILIZAÇÃO DE VISÃO ESTEREOSCÓPICA PARA RECONSTRUÇÃO 3D**

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Sistemas de Informação do Instituto Federal do Espírito Santo, Campus Cachoeiro de Itapemirim, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Dr. Ricardo Maroquio Bernardo

Cachoeiro de Itapemirim

2023

## **RESUMO**

Mapeamento geográfico, ressonâncias magnéticas e veículos autônomos. São esses, apenas três exemplos de áreas que empregam uma crescente e relevante aplicação de Visão Computacional: a reconstrução 3D. Esse nome é dado ao processo pelo qual se faz possível a obtenção de um modelo gráfico detalhado em três dimensões de uma determinada cena ou sequência de cenas objetivadas. Para isso, faz-se uso de dispositivos de capturas de imagem aliado a alguma técnica computacional para desenvolvimento do modelo. Visto isso, neste trabalho, é proposto a utilização de pares de imagens provindas de um dispositivo estereoscópico aliado a técnicas de calibragem e visão estereoscópica para obtenção de modelos reconstruídos em três dimensões.

Palavras-chave:

## **ABSTRACT**

Geographic mapping, MRI and entities. These are just three examples of application areas that employ a growing and relevant Computational vision: a 3D reconstruction. This name is given to the process by which it is possible to configure a detailed graphic model in three dimensions of a given scene or sequence of scenes. For this, make use of image capture devices combined with some computational technique to develop the model. In this work, it is proposed the use of depth images from a stereoscopic device combined with stereoscopic triangulation techniques to view reconstructed models in three dimensions.

Keywords:

## SUMÁRIO

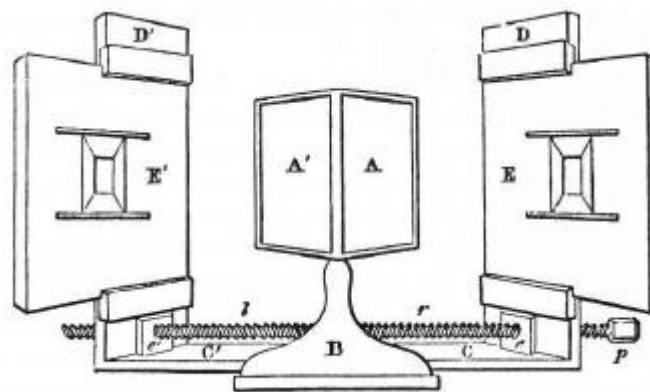
<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>5</b>
1.1	Descrição do Problema . . . . .	6
1.2	Objetivos . . . . .	6
1.3	Justificativa . . . . .	7
<b>2</b>	<b>REFERENCIAL TEÓRICO . . . . .</b>	<b>8</b>
2.1	Visão Humana . . . . .	8
2.2	Visão Computacional . . . . .	9
<b>2.2.1</b>	<b>Aquisição de Imagem . . . . .</b>	<b>11</b>
<b>2.2.2</b>	<b>Processamento . . . . .</b>	<b>11</b>
<b>2.2.3</b>	<b>Interpretação ou compreensão . . . . .</b>	<b>12</b>
<b>2.2.4</b>	<b>Aplicação . . . . .</b>	<b>13</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>14</b>

## 1 INTRODUÇÃO

A reconstrução 3D por meio de visão estereoscópica é uma técnica que tem recebido crescente interesse da comunidade científica e da sociedade em geral, permitindo a criação de modelos tridimensionais a partir de cenas 2D. Segundo (SCHARSTEIN; SZELISKI, 2002), a técnica se baseia na utilização de duas imagens bidimensionais capturadas por câmeras com posições ligeiramente diferentes, permitindo a reconstrução de um modelo 3D do objeto em questão.

O uso dessa técnica remonta ao início do século XIX, quando Charles Wheatstone desenvolveu o primeiro estereoscópio(1838), um aparelho capaz de criar a percepção de profundidade a partir de fotografias bidimensionais (WHEATSTONE, 1838). Desde então, a visão estereoscópica evoluiu consideravelmente, com o avanço da tecnologia de câmeras e computação gráfica.

Figura 1 – Estereoscópio de Charles Wheatstone.



Fonte: (ZONE, 2014).

Atualmente, a reconstrução 3D por meio de visão estereoscópica tem uma ampla variedade de aplicações em diferentes áreas, como medicina, engenharia e arquitetura. Conforme apontado por (NAM et al., 2012), tal abordagem pode fornecer uma melhor compreensão e análise a partir de modelos tridimensionais dos órgãos humanos, permitindo a realização de cirurgias menos invasivas e mais precisas. Além disso, a capacidade de fornecer informações exatas e detalhadas sobre objetos e ambientes, aliada ao barateamento e à acessibilidade da tecnologia, implica em um grande crescimento de demanda social por este tipo de aplicação.

## 1.1 DESCRIÇÃO DO PROBLEMA

Embora a visão binocular tenha se mostrado um método eficaz e promissor para a obtenção de modelos tridimensionalmente reconstruídos, ainda existem desafios a serem superados, especialmente no que se refere à precisão do modelo obtido. Isso se deve em parte à complexidade dos algoritmos matemáticos utilizados para extrair informações de profundidade das imagens, que podem ser afetadas por fatores como a iluminação e a qualidade da imagem. Além disso, o processo de reconstrução pode ser custoso e requer conhecimentos técnicos avançados, o que torna a técnica inacessível para muitas pessoas e limita seu uso em aplicações práticas.

## 1.2 OBJETIVOS

Diante do exposto, este trabalho tem como objetivo geral a elaboração de uma ferramenta que permita a reconstrução de modelos 3D com precisão e eficiência, explorando as possibilidades e limitações da visão estéreo.

Para atingir esse objetivo geral, torna-se necessário a realização dos seguintes objetivos específicos:

1. Estudar os fundamentos teóricos da visão estereoscópica e da reconstrução 3D;
2. Estudar a documentação do OpenCV, uma biblioteca de código aberto e multiplataforma que é utilizado para o desenvolvimento de aplicações na área de Visão computacional;
3. Desenvolver uma configuração própria para a aquisição dos dados.
4. Elaboração do processo de obtenção do modelo reconstruído em 3D;
5. Realizar experimentos e testes para avaliar a eficácia e eficiência da solução proposta.

### 1.3 JUSTIFICATIVA

Com o avanço constante das tecnologias e o surgimento de experiências digitais inovadoras, como o metaverso, o estudo relacionado à reconstrução e modelagem em três dimensões tem se tornado cada vez mais relevante. Segundo as estatísticas apresentadas pela (SkyQuest Technology Consulting Pvt. Ltd., 2021), a projeção é que o mercado global de reconstrução 3D alcance um valor de 1300 milhões de dólares em 2028.

A pesquisa em 3D é essencial para o aprimoramento de sistemas e a compreensão das diferentes perspectivas de objetos e ambientes. Dessa forma, espera-se que a elaboração desta ferramenta contribua para o avanço das pesquisas nessa área de estudo, ampliando suas diversas aplicações práticas e atendendo aos objetivos propostos neste trabalho.



## 2 REFERENCIAL TEÓRICO

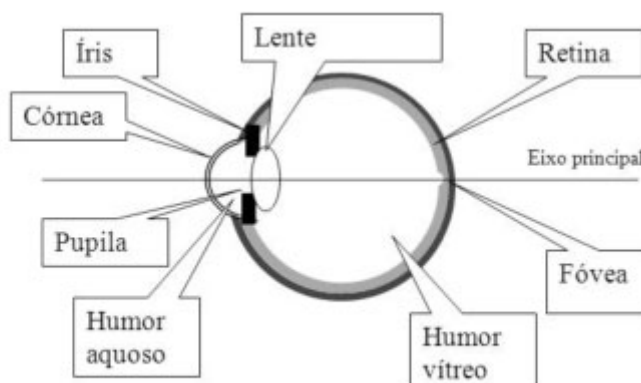
Nesta seção, serão apresentados os conceitos essenciais referentes às técnicas que fundamentaram o desenvolvimento do trabalho.

### 2.1 VISÃO HUMANA

Dos cinco sentidos – visão, audição, olfato, paladar e tato – a visão é, sem dúvida, aquele de que o homem passou a depender acima de todos os outros, e de fato aquele que fornece a maioria dos dados que ele recebe (DAVIES, 2012). Tais dados são capturados em forma de impulsos luminosos através dos olhos humanos e posteriormente processados em imagens pelo cérebro.

A recepção dos dados externos do ambiente, como dito, é a primeira etapa do sistema visual humano. Tal tarefa é atribuída a dois importantes órgãos: os olhos. Trata-se de pequenos órgãos constituídos de estruturas transparentes que permitem a máxima transmissão ininterrupta à luz. Essas estruturas, a saber, a córnea, a pupila, o cristalino e os humores (aquoso e vítreo) do olho, determinam coletivamente a ótica do olho (RAMAMURTHY; LAKSHMINARAYANAN, 2015).

Figura 2 – Exemplo estrutural do olho humano.



Fonte: Scielo.

A córnea é o principal componente refrator do olho, portanto, é a primeira parte do olho humano que a luz encontra. Sua transparência e avascularidade proporcionam uma ótima transmissão de luz (REMINGTON, 2012).

A esclera mantém o formato do globo, oferecendo resistência a impactos internos e forças externas, e fornece uma fixação para as inserções do músculo extraocular (REMINGTON, 2012).

A pupila e a íris em conjunto formam um componente chamado uvea e são responsáveis por regular a quantidade de luz incidente no olho humano. A íris é uma estrutura fina e circular localizada anterior à lente. A pupila, está localizada ligeiramente nasal, inferior ao centro da íris e é responsável por regular a iluminação da retina. Em condições de muita luz a pupila é muito pequena (miótica) e bastante grande (midriático) em iluminação fraca (REMINGTON, 2012).

As lentes ou cristalino é uma estrutura elíptica avascular e transparente que auxilia na focalização dos raios de luz na retina. Tal estrutura é maleável, e a contração do músculo ciliar pode causar uma alteração na forma da lente, aumentando o poder dióptrico do olho (REMINGTON, 2012).

A distância entre a superfície posterior da córnea e a íris é chamada a câmara anterior do olho e é preenchido com um meio opticamente claro chamado humor aquoso. A pressão exercida pelo humor aquoso ajuda a manter a integridade estrutural do olho e é chamada de pressão intraocular (PIO). Já o espaço entre a lente e a retina é a câmara posterior e é preenchida com um meio gelatinoso incolor chamado de humor vítreo. Tanto o humor aquoso quanto o vítreo absorvem menos de 10 por cento da luz incidente (RAMAMURTHY; LAKSHMINARAYANAN, 2015).

## 2.2 VISÃO COMPUTACIONAL

A visão computacional é uma área da informática que existe desde a década de 1960, mas devido a evolução tecnológica e o barateamento dos sistemas de imagem, a necessidade de pesquisa e desenvolvimento acarretou um grande crescimento nessa área de atuação (FORSYTH; PONCE, 2011).

O grande papel da visão computacional é extrair descrições do mundo a partir de imagens ou sequências de fotos. Sendo assim, a ideia desse campo de estudo é a obtenção de dados para tomada de decisão. Isso, é feito, a partir de interpretações de

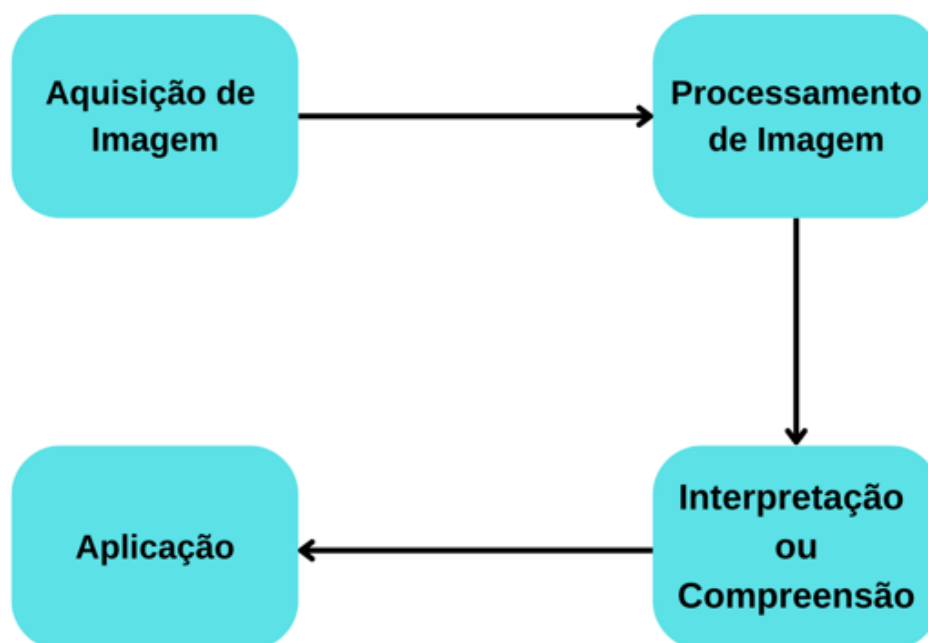
imagens baseado em modelos construídos com o auxílio da geometria, física e teoria da aprendizagem (FORSYTH; PONCE, 2011).

O que a visão computacional se propõe a fazer é inequivocamente útil. Sendo assim, as possibilidades de aplicação são muito amplas e se estendem a diversos ramos da sociedade e do mundo, como: na medicina, para fins militares, fins ambientais, etc (FORSYTH; PONCE, 2011).

Além das diversas possibilidades de aplicação, também ocorre a possibilidade da divisão de subáreas de pesquisa, por exemplo: análise de movimento, detecção de objetos, visão estéreo, entre outras. Isso, ocorre principalmente, devido a vasta capacidade de extração de diversas informações diferentes a partir de uma determinada imagem, tornando inviável a unificação em uma única área de pesquisa (TRUCCO; VERRI, 1998).

Um sistema computacional baseado nessa área de estudo, em termos gerais, possui tipicamente, quatro etapas principais, como mostrado na Figura 2. Sendo que cada uma delas serão exploradas com mais detalhes nos tópicos seguintes.

Figura 3 – Fluxograma das etapas de um sistema de Visão Computacional.



Fonte:

Adaptado de (ĆIRIĆ et al., 2016).

### **2.2.1 Aquisição de Imagem**

A etapa de aquisição é o primeiro passo de um sistema de visão computacional, e é nela que será feita a captura, coleta ou geração das imagens e vídeos que serão utilizados posteriormente.

De acordo com (FERREIRA; ANJOS, 1988), o conceito de imagem pode ser definido como a representação de uma pessoa ou uma coisa por meio de algo, como pintura, imitação, cópia, entre outros. No âmbito computacional, imagens são armazenadas e representadas em estruturas de matrizes e com isso podem ser divididas em pequenas porções de pixels, cujo cada pixel tem uma respectiva cor e posição representada por um par de coordenadas (KUMAR; VERMA, 2010).

Com a popularização e barateamento das câmeras, esse meio se tornou o principal para a aquisição de tais representações externamente, contudo, não são os únicos, existem uma vasta gama de dispositivos externos, como: scanners, equipamentos de ressonância magnética e raio-x.

Apesar disso, o processo de aquisição não se limita somente a dispositivos externos, já que meios internos gerados pelo próprio computador também podem ser úteis para esta etapa, como por exemplo: ilustrações e capturas de tela.

### **2.2.2 Processamento**

Este tópico e o tópico anterior são fortemente dependentes, uma vez que é nessa etapa que serão aplicadas ações sobre os dados adquiridos anteriormente.

As ações são técnicas aplicáveis que visam um objetivo específico para o melhor manuseio e uso dos dados nas etapas seguintes. Exemplos dessas técnicas são: transformações para aumentar a performance como no caso de um redimensionamento da imagem, conversão de formatos por conta de compatibilidade, ajustes de características, como cor e luminosidade, correções de ruídos, aplicação de filtros, correções geométricas, segmentação e reconstrução.

Figura 4 – Antes e depois da aplicação de filtro de imagem.



Fonte: (MATLAB, 2022)

### 2.2.3 Interpretação ou compreensão

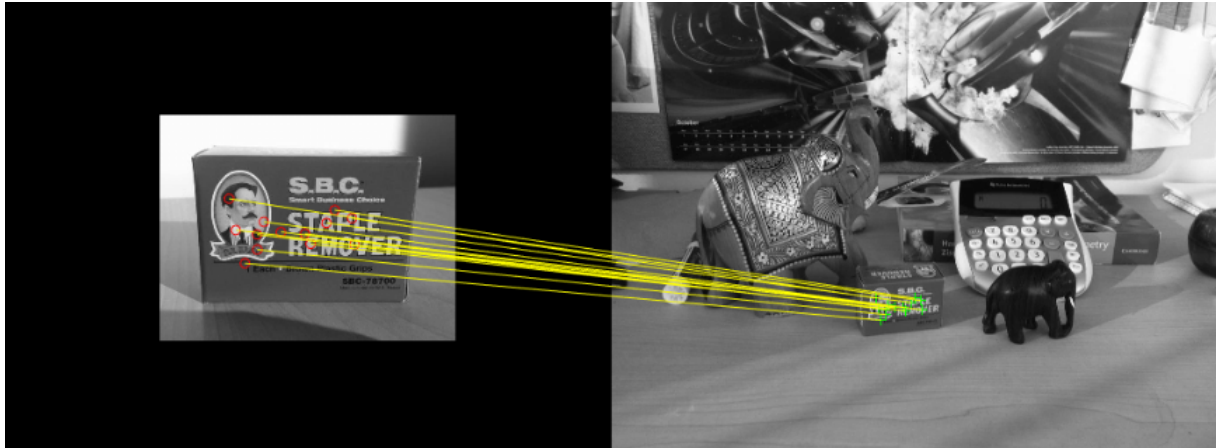
Pode-se dizer que esse passo é um intermediário, cuja principal função é a extração de recursos e características dos dados. A extração de recursos ou compreensão refere-se ao processo de transformação de dados brutos em recursos numéricos que podem ser processados, preservando as informações no conjunto de dados original (MATLAB, 2022).

A extração de recursos para dados de imagem representa as partes interessantes de uma imagem, como um vetor de feição compacto, por exemplo. A forma, textura, intensidade, tamanho e posição são alguns dos exemplos de recursos mensuráveis que podem ser extraídos de uma imagem. Para isso, existem diversas técnicas aplicáveis.

Com a grande evolução de algoritmos de inteligência artificial, o aprendizado de máquina profundo é predominante na análise de imagem e vídeo e tornou-se conhecido por sua capacidade de obter dados brutos de imagem como entrada, abrindo a possibilidade de ignorar essa etapa em alguns casos (MATLAB, 2022). Apesar disso, independente da abordagem adotada, os aplicativos de visão computacional exigem uma representação eficaz dos recursos de imagem – implicitamente pelas primeiras camadas de uma rede profunda ou aplicando explicitamente algumas das técnicas de

extração de recursos de imagem de longa data (MATLAB, 2022).

Figura 5 – Detecção de um objeto (à esquerda) em uma cena desordenada (à direita) usando uma combinação de detecção de recursos, extração de recursos e correspondência.



Fonte: (MATLAB, 2022)

#### 2.2.4 Aplicação

Essa etapa corresponde ao estágio final e é onde ocorre a aplicação das informações extraídas em um determinado contexto e visa um objetivo específico conforme o domínio de problema.

É muito comum, nesse tópico, a interdisciplinaridade com outras áreas de estudo, como o uso de redes neurais, que possibilita a aplicação de técnicas para obtenção de um determinado resultado. Podem ser citados como exemplos de técnicas aplicáveis nesse estágio: *Haar Cascade* para reconhecimento de imagens baseado em classificadores, *OCR* para reconhecimento de caracteres e a detecção de bordas para medição de mudanças bruscas no brilho de uma cena ou imagem.

## REFERÊNCIAS

DAVIES, E. R. *Computer and machine vision: theory, algorithms, practicalities*. [S.l.]: Academic Press, 2012.

FERREIRA, A. de H.; ANJOS, M. dos. *Dicionário Aurélio básico da língua portuguesa*. Editora Nova Fronteira, 1988. ISBN 9788520908266. Disponível em: <<https://books.google.com.br/books?id=Sb8uJQAACAAJ>>.

FORSYTH, D. A.; PONCE, J. *Computer vision: A modern approach. 2nd*. [S.l.]: Upper Saddle River: NJ: Prentice Hall, 2011.

KUMAR, T.; VERMA, K. A theory based on conversion of rgb image to gray image. *International Journal of Computer Applications*, Foundation of Computer Science, v. 7, n. 2, p. 7–10, 2010.

MATLAB. *version 9.13 (R2022b)*. Natick, Massachusetts: The MathWorks Inc., 2022.

NAM, K. W. et al. Application of stereo-imaging technology to medical field. *Healthcare informatics research*, Korean Society of Medical Informatics, v. 18, n. 3, p. 158–163, 2012.

RAMAMURTHY, M.; LAKSHMINARAYANAN, V. *Human vision and perception*. [S.l.]: Springer International Publishing, 2015. 1–23 p.

REMINGTON, L. *Clinical Anatomy and Physiology of the Visual System*. 3. ed. [S.l.]: Elsevier Health Sciences, 2012. ISBN 9781437719260.

SCHARSTEIN, D.; SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, Springer, v. 47, p. 7–42, 2002.

SkyQuest Technology Consulting Pvt. Ltd. *3D Reconstruction Technology Market - Growth, Trends, COVID-19 Impact, and Forecasts (2021 - 2026)*. 2021. Acesso em: 28 de fevereiro de 2023. Disponível em: <<https://skyquestt.com/report/3d-reconstruction-technology-market>>.

TRUCCO, E.; VERRI, A. *Introductory techniques for 3-D computer vision*. [S.l.]: Prentice Hall Englewood Cliffs, 1998. v. 201.

WHEATSTONE, C. Xviii. contributions to the physiology of vision.—part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical transactions of the Royal Society of London*, The Royal Society London, n. 128, p. 371–394, 1838.

ZONE, R. *Stereoscopic cinema and the origins of 3-D film, 1838-1952*. [S.l.]: University Press of Kentucky, 2014.

ĆIRIĆ, I. et al. Thermal vision based intelligent system for human detection and tracking in mobile robot control system. *Thermal Science*, v. 20, p. 1553–1559, 01 2016.