

# Visão Computacional

Luiz Romário Santana Rios<sup>1</sup>

<sup>1</sup>Departamento de Ciência da computação – Universidade Federal da Bahia (UFBA)  
Salvador, Bahia, Brasil

luizromario@dcc.ufba.br

**Abstract.** *Computer Vision is the field in computing that studies ways of extracting information from the objects in an image, such as their shapes, speeds, etc.. It was theorized a long time ago, but just recently, when computers could process large amounts of data like images, it started to develop for practical use. Still, there's no standard formulation or solution for “the Computer Vision problem”.*

**Resumo.** *A Visão Computacional é o campo da computação que estuda maneiras de extrair informação dos objetos de uma imagem, tais como suas formas, velocidades, etc.. Foi teorizado há bastante tempo, mas começou a ser empregado na prática apenas recentemente, quando os computadores se tornaram capazes de processar quantidades de dados tão grandes quanto imagens. Ainda assim, não há formulação ou solução padrões para “o problema da Visão Computacional”.*

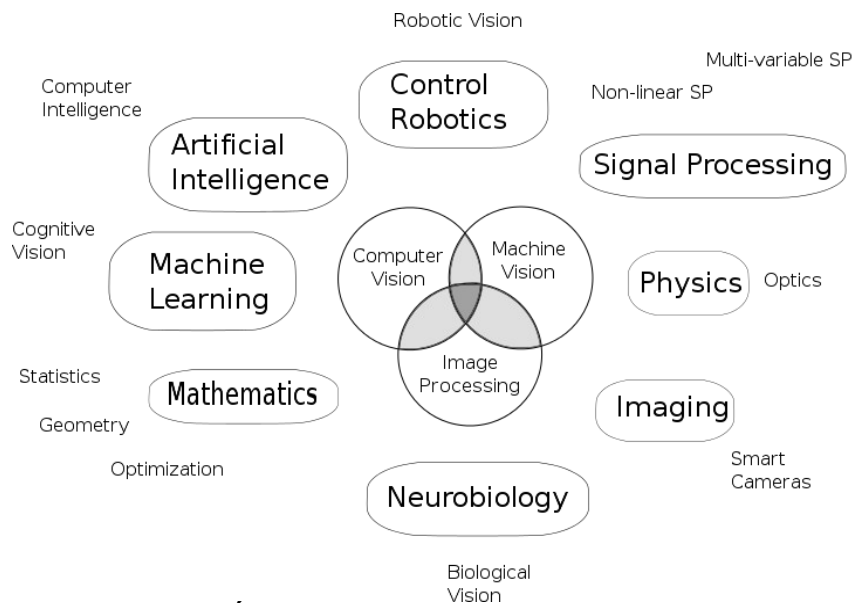
## 1. Introdução

Visão computacional é o estudo da extração de informação de uma imagem; mais especificamente, é a construção de descrições explícitas e claras dos objetos em uma imagem [Ballard and Brown, 1982]. Difere do processamento de imagens porque, enquanto ele se trata apenas da transformação de imagens em outras imagens, ela trata explicitamente da obtenção e manipulação dos dados de uma imagem e do uso deles para diferentes propósitos.

### 1.1. Campos relacionados

Diversos campos estão relacionados à visão computacional. O mais notável é a própria visão biológica, no qual a visão computacional é baseado e que, logicamente, é diretamente relacionado. Na verdade, é justo dizer que essas são áreas análogas, uma para a computação, outra, para a biologia. Afinal, enquanto o último estuda os processos psicológicos envolvidos na formação e percepção de imagens pelos seres vivos, o primeiro estuda os processos e algoritmos usados por máquinas para enxergar. Estudos interdisciplinares nas áreas vêm se mostrando bastante proveitosos.

Além desse campo, o campo de inteligência artificial, principalmente as áreas de aprendizado de máquina e de reconhecimento de padrões, é muito empregado para o entendimento da informação gerada a partir da imagem. Outro campo importante é a própria física, que explica como radiação numa certa frequência sensibiliza sensores óticos e como sua trajetória se comporta ao atingir um anteparo. Outros campos também relacionados são o processamento de sinais, processamento de imagens, visão de máquina, entre outros. A figura 1 mostra um esquema dos relacionamentos da visão computacional com outros campos.



**Figura 1. Áreas relacionadas à visão computacional**

## 1.2. Dificuldades e desafios

O reconhecimento de imagens requer bastante conhecimento prévio sobre o mundo e bastante capacidade de julgamento para se fazer conclusões sobre os dados da imagem e, então, tomar uma ação adequada às necessidades do agente. Por exemplo, um agente de controle de qualidade precisa ter um conhecimento prévio sobre o que é uma peça fora da especificação – e precisa, para isso, conhecer a especificação – e, também, saber que ação tomar para essa peça – descarte, conserto, etc..

Mas, para julgar os objetos e dados numa imagem, o agente precisa, antes, extraí-los da imagem. E isso inclui, entre outras coisas, saber remover o ruído das imagens, saber lidar com diferentes configurações de iluminação e sombra, muitas vezes numa mesma imagem – por exemplo, saber que o cinza escuro na sombra e o branco no claro são a mesma cor, a noção de perspectiva, perceber profundidade a partir de uma visão estereoscópica e das configurações das sombras, entre outros. Mas, principalmente, reconhecer e isolar os objetos, os elementos realmente relevantes ao contexto, de uma dada cena.

Como ainda se sabe pouco sobre o funcionamento dessas técnicas de visão nos seres vivos, é um enorme desafio pensar em maneiras de implementá-las de maneira viável usando um computador. E, apesar de todo o estudo realizado na área, ainda não existe uma solução definitiva, única, para “o problema da visão computacional”. Tudo o que se tem são algumas definições gerais e soluções esparsas e específicas para algum problema (leitura de texto, reconhecimento facial, etc.), pois, novamente, pouco se sabe sobre como a visão funciona.

Serão explicadas, a seguir, as diversas técnicas conhecidas pela computação para implementar visão computacional em agentes inteligentes e, então, diversas aplicações práticas. É bom notar que aqui será feita apenas uma análise qualitativa da visão computacional e de seus métodos, já que se trata de um campo bastante extenso e relativamente complexo.

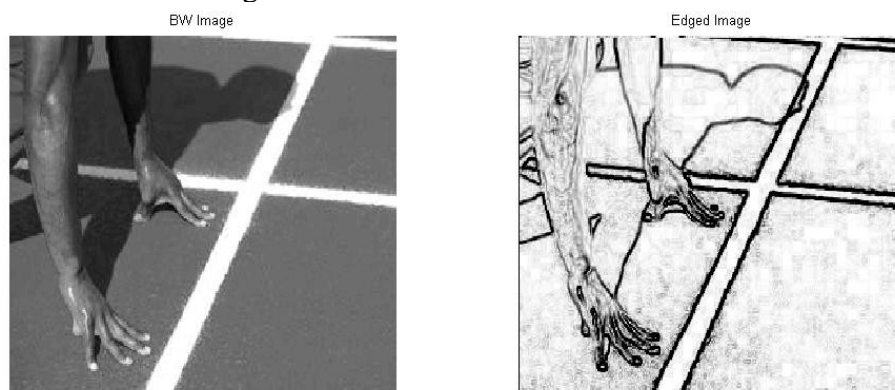
## 2. Problemas da visão computacional

### 2.1. Detecção de bordas

Um dos problemas mais importantes no reconhecimento de objetos em imagens é a detecção de bordas. Basicamente, consiste em detectar regiões da imagem nas quais ocorre uma mudança abrupta de brilho na imagem, que, geralmente, representam uma mudança das características do que está sendo visto.

É um problema importante porque mudanças abruptas no brilho da imagem, sob a percepção natural humana da visão, podem representar descontinuidade de profundidade – uma parede atrás de outra, por exemplo, geralmente é mais escura; descontinuidade da orientação da superfície – uma face da parede que está mais perpendicular à iluminação é mais clara que uma face que está paralela; mudanças nas propriedades do material – pedras pretas e brancas no chão são pedras diferentes – e variações na iluminação da cena – a pedra cinza do lado de fora da casa e a pedra preta do lado de dentro são, na verdade, da mesma cor, além de outras características como reflexão e refração, por exemplo.

Em relação a esses aspectos, uma representação por bordas é bem fiel às propriedades físicas do mundo. Além disso, a representação do mundo bidimensional por linhas unidimensionais tem a vantagem de ser compacta, pois leva em conta apenas os detalhes relevantes da imagem.



**Figura 2. Exemplo de reconhecimento de borda: à esquerda, imagem original; à direita, as bordas da imagem**

Entretanto, o reconhecimento de bordas claramente não é trivial. Uma imagem digital comum tirada do mundo é discreta, geralmente. Logo, não há uma noção clara do que é borda ou não, já que borda significa descontinuidade da imagem e não existe uma definição natural do que é descontinuidade em uma função discreta. Um exemplo é a Figura 3 abaixo. À medida que as imagens são borradas, fica cada vez mais difícil definir suas bordas, já que imagens mais borradas são mais “contínuas” que imagens nítidas. Por causa disso, bordas extraídas de imagens reais estão a problemas como a fragmentação, quando as bordas de uma mesma curva estão desconectadas, e bordas falsas, quando são criadas bordas para elementos irrelevantes. Esses problemas podem afetar a forma como os dados extraídos são analisados, frequentemente levando a interpretações falsas sobre o conteúdo.

## 2.2. Reconhecimento de objetos

O reconhecimento de objetos é a tarefa de reconhecer, numa cena, um objeto predefinido na base de conhecimento ou um aprendido. Existem diversas formas de reconhecer objetos numa cena, mas, geralmente, os métodos empregados usam templates para gerar o conjunto de bordas do objeto requerido e, então, compara suas bordas com as bordas da imagem (métodos baseados em aparência). Outro conjunto de métodos bastante comum busca por semelhanças entre as características do modelo do objeto e as da imagem (métodos baseados em característica).

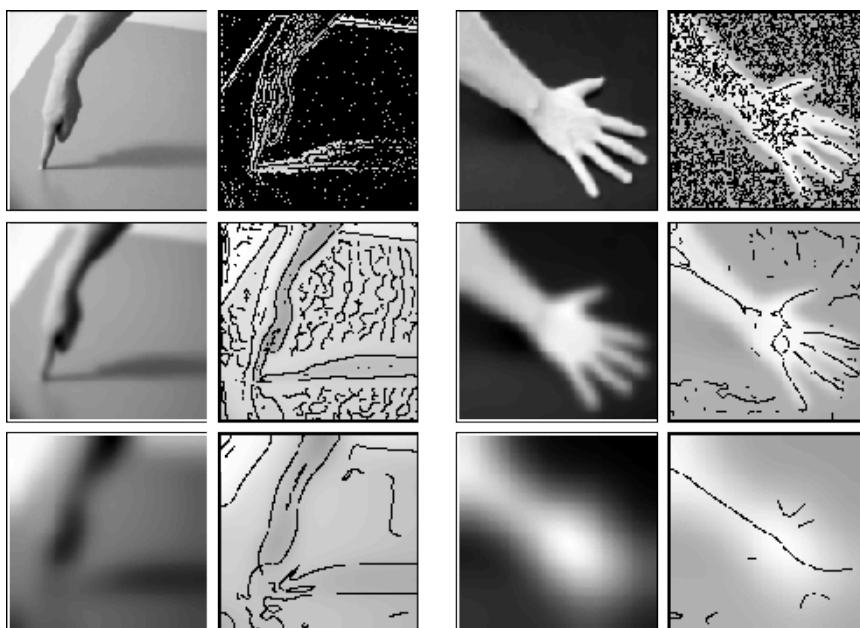


Figura 3. Duas imagens diferentes borradas a diferentes intensidades e suas respectivas bordas.



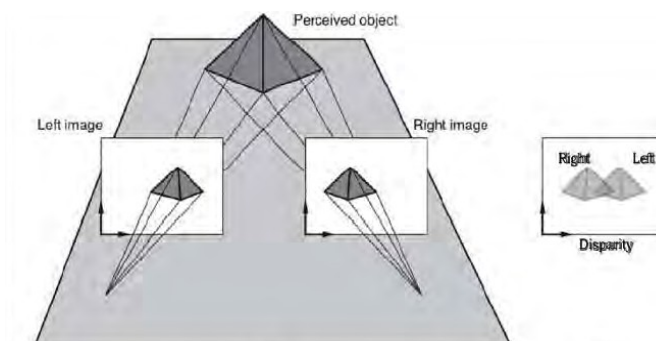
Figura 4. Reconhecimento de objetos por meio de método baseado em característica.

## 2.3. Visão estereoscópica

A visão estereoscópica computacional é a extração de dados 3D de uma imagem digital. Ao analisar diversas imagens diferentes do mesmo objeto tiradas de ângulos diferentes, é possível, ao analisar as mudanças nas perspectivas de uma imagem para a outra, obter uma representação tridimensional limitada do objeto, dando a sensação de profundidade a ele. Esse processo é análogo à estereopsia nos animais, que são capazes de perceber profundidade combinando as imagens que chegam em ambos os olhos.

Tradicionalmente, são usadas duas câmeras, em analogia aos olhos humanos, para capturar imagens levemente diferentes do mesmo objeto. Como as duas imagens geradas estão em pontos diferentes do espaço, diferentes pontos do mesmo objeto estarão em localizações diferentes dependendo da sua distância ao referencial de

observação – objetos mais próximos se movem mais rapidamente com o movimento da câmera. A diferença das posições dos diversos pontos do objeto é, então, usada para gerar o modelo tridimensional do objeto. É bom notar que, para medir essa diferença, é necessário resolver o problema da correspondência, isto é, determinar qual ponto em uma imagem é resultado da projeção do mesmo ponto na outra imagem [Russel and Norvig, 2010].



**Figura 5. Duas projeções diferentes do mesmo objeto formam uma representação tridimensional dele, de acordo com as disparidades entre elas.**

Uma forma parecida de perceber profundidade nas imagens, mas que não precisa de mais de uma câmera, é a paralaxe. É parecida com a estereoscopia, mas, em vez de usar diversas projeções do mesmo objeto no mesmo instante, são usadas diversas projeções que ocorreram em tempos diferentes em que a câmera estava em movimento. Como as imagens mostrarão o mesmo objeto em posições diferentes, o restante do processo é o mesmo que no processo usando duas câmeras.

## 2.4. Detecção de movimento

Detectar que objetos ou áreas estão se movendo numa cena é importante para várias aplicações. Existem diversos métodos diferentes de se fazê-lo em relação a diversos referenciais, mas todos partem do mesmo princípio de observar quais pontos se moveram na imagem e com que velocidade.

O fluxo óptico descreve a velocidade e a direção do movimento de cada pixel em relação à imagem anterior. Por medir um movimento na imagem, não no mundo, é medido em pixels, não na unidade de velocidade do objeto – mesmo porque, não é possível saber a velocidade do objeto, já que esse método não detecta o objeto. Com o fluxo óptico é possível inferir algumas informações úteis sobre a cena. Por exemplo, objetos mais distantes se movem mais lentamente na imagem que objetos próximos. Logo, é possível deduzir a velocidade de um objeto dependendo de sua velocidade da imagem. O conceito de fluxo óptico vem sendo bastante desenvolvido em áreas como a compactação de vídeo.



**Figura 6. Dois quadros de um vídeo mostrando uma tenista mexendo sua raquete e o fluxo óptico extraído das duas primeiras figuras em seguida.**

Uma maneira melhor para se seguir o movimento de um determinado objeto seria localizá-lo no vídeo, usando técnicas de reconhecimento de objetos a cada quadro que se passa nele. Esse método se chama *video tracking*. Com o vídeo tracking é possível detectar movimentos de translação, bem como rotação, de um objeto em específico. Geralmente, um modelo bidimensional, ou tridimensional, se necessário, é usado para detectar o objeto nas imagens do vídeo, o que torna possível saber o deslocamento absoluto do objeto no ambiente. Entretanto, esse método é mais complicado para objetos que se movem muito rapidamente em relação à taxa de quadros por segundo. Além disso, um modelo tridimensional é obrigatório caso haja muitas mudanças no sentido do movimento.

### **3. Aplicações práticas da visão computacional**

Por ser um campo bastante amplo e por trata do mecanismo de percepção do ambiente mais fundamental que existe na natureza, a visão computacional tem bastante emprego em diversas áreas, não só em informática e robótica, como, também, na medicina, na área de digitalização de dados, segurança e em todo outro campo no qual se tenha que perceber algo visualmente. Entretanto, como já foi afirmado, esse é um campo que começou a se desenvolver, de certa forma, recentemente e que não é muito maduro ainda; logo, há muito para se estudar. Mas, por enquanto, serão mostradas as diversas aplicações já existentes que utilizam das técnicas de visão computacional.

#### **3.1. Aplicações na medicina**

A medicina é uma das áreas em que a visão computacional tem mais importância. O uso dessa naquela é, geralmente, receber uma imagem de um paciente e realizar o diagnóstico. Diferentes imagens podem ir desde um raio X até uma tomografia computadorizada. E os dados que podem ser extraídos dessas imagens são, órgãos, que podem vir de diversos tipos de imagens (endoscopia, ultrassom, etc.), anomalias nesses órgãos entre outros. Outro exemplo de aplicação é o estudo do movimento do corpo humano por visão computacional, para ajudar pacientes com deficiências motoras.

As aplicações da visão computacional na medicina são inúmeras e bastante espalhadas: praticamente qualquer área que esteja relacionada à análise visual de algo – diagnóstico, estudos em neurobiologia e até mesmo a própria visão – são potenciais campos de uso da visão computacional.

#### **3.2. Realidade aumentada**

O conceito de realidade aumentada é bastante novo e só começou a ser empregado em larga escala recentemente. Realidade aumentada é o conceito de adicionar mais uma camada de análise à realidade usando a visão computacional, em tempo real, permitindo que um usuário consiga ver dados com exatidão e clareza impossíveis de serem atingidas apenas com sua visão e raciocínio convencionais.

Um exemplo simples é um carro que reconhece as bordas de uma estrada e avisa quando o carro sai da sua faixa. Sem o reconhecimento da pista pelo carro, o controle do carro depende exclusivamente do raciocínio do motorista e sua capacidade de ver a pista. Mas o carro, nesse caso, aumenta os dados que o motorista recebe: de apenas imagens, ele passa a ser informado da configuração da pista.

#### **3.3. Reconhecimento ótico de caracteres (OCR)**

OCR é o processo de ler texto impresso, escrito ou datilografado e transformá-lo em

texto digital. É usado com diversos objetivos, incluindo: informatizar um centro empresarial que tem todos os seus arquivos armazenados num sistema físico, escrito em papéis; preservar um acervo de livros antigos, digitalizando todos os livros e mantendo cópias digitais de cada um; distribuir um livro via internet, permitindo, dentre outras vantagens, uma reimpressão mais limpa, facilidade de busca de termos etc..

Apesar de requerer calibração para ler bem uma dada fonte, já são comuns, atualmente, sistemas “inteligentes” de OCR, que reconhece a maioria das fontes conhecidas com acurácia bastante satisfatória, chegando bastante perto da impressão original, escaneando, inclusive, imagens, tabelas e outros elementos. Para chegar a esse nível de precisão, o reconhecedor de caracteres usa uma tabela de palavras conhecidas para corrigir a possível criação de palavras inexistentes pelo sistema.

A área de reconhecimento de texto escrito à mão, por outro lado, está num estado bem mais imaturo que o reconhecimento de texto digitado, pois a forma dos caracteres escritos à mão é bem menos legível que caracteres digitados, que já são predefinidos. Para atingir um nível satisfatório de precisão, é necessário, nesse caso, saber qual o contexto do texto que está sendo lido e conhecer o assunto sobre o qual o texto trata. Determinar esse tipo de informação não é trivial num nível mais generalizado, já que envolve, além de uma base enorme de conhecimento, o reconhecimento de linguagens naturais, que é outro problema bastante complexo da inteligência artificial.

### **3.4. Visão de máquina**

A visão de máquina é empregada geralmente nos setores industriais e de engenharia de controle de qualidade. Usa bastante da visão computacional para detectar componentes fora do padrão desejado e descartá-los ou executar outra ação, dependendo das necessidades.

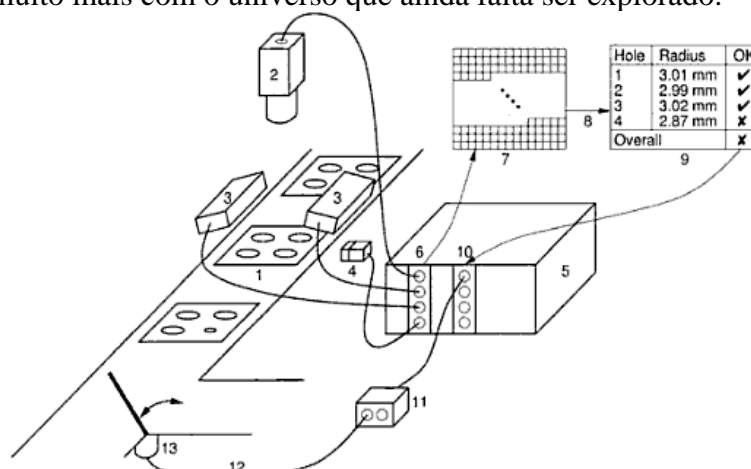
Ela pode ser usada para a identificação de diferentes tipos de objetos e para decidir que tipo de inspeções serão utilizadas. Isso pode ser feito utilizando códigos identificadores para cada objeto ou pode ser definido pelo próprio formato do objeto. Outro exemplo de visão de máquina é a inspeção de montagem de diferentes equipamentos. A máquina montadora precisa montar todos os componentes do equipamento na posição correta e fazer todas as inspeções necessárias das posições dos componentes – por exemplo, montagem dos chips em uma placa de circuito impresso, entre diversos outros exemplos de checagem ao adequamento a um certo padrão.

Assim como diversos outros campos da visão computacional, a visão de máquina é um campo especialmente multidisciplinar, pois envolve engenharia elétrica, eletrônica e de software, robótica, visão computacional, entre outros. É um dos campos mais antigos onde a visão computacional é usada – sua utilização remonta aos anos 90.

## **4. Conclusão**

A área de visão computacional ainda é relativamente imatura como um todo, é uma área cujos métodos e fundamentos não são simples, pois requerem um nível de abstração alto, além de ainda haver uma vastidão de aspectos não explorados ou inexplicados. Apesar disso, já há muitas soluções de visão computacional altamente confiáveis e bastante eficientes, bem como áreas bastante promissoras, sendo ativamente pesquisadas, que darão uma qualidade de vida muito maior à população no futuro. Mais ainda, estão surgindo e, provavelmente, surgirão muito mais áreas de pesquisa nesse

campo. Enfim, é um campo que, se já traz benefícios extraordinários com o pouco que se sabe, trará muito mais com o universo que ainda falta ser explorado.



**Figura 7. Modelo de um sistema de visão de máquina.** Uma imagem do objeto a ser inspecionado (1) é obtida pela câmera (2). O objeto é iluminado pela iluminação (3). Um sensor fotoelétrico (4) dá início à obtenção da imagem. Um computador (5) obtém a imagem por meio de uma comunicação câmera-computador (6). Um driver de dispositivo (7) monta a imagem na memória do computador. O software de visão de máquina (8) faz a inspeção dos objetos e fornece, então, uma avaliação dos objetos (9). O resultado da avaliação é comunicado a um controlador lógico (11) por meio de uma interface digital (10). O controlador controla o atuador (13) por meio de uma outra interface (12). O atuador, que pode ser um motor por exemplo, remove objetos defeituosos da linha de produção [Steger, Ulrich and Wiedemann, 2008].

## Referências

- Ballard, Dana H. and Brown, Christopher M. (1982). "Computer Vision". Prentice Hall. ISBN 0131653164.
- Azad, Gockel, and Dillmann, R. (2008). "Computer Vision – Principles and Practice". Elektor International Media BV. ISBN 0905705718.
- <http://eom.springer.de/E/e120040.htm>, acessado em 15 de novembro de 2010.
- <http://trfkad.tracker.prq.to:80/announce>, acessado em 15 de novembro de 2010.
- Elgammal, Ahmed. "CS 534: Computer Vision 3D Model-based recognition". Dept of Computer Science, Rutgers University
- Bradski, Gary and Kaehler, Adrian. "Learning OpenCV: Computer Vision with the OpenCV Library". O'Reilly.
- Russel, S. And Norvig, P. (2010). "Artificial Intelligence – A modern approach". Pearson. ISBN 0136042597.
- Steger, Carsten, Markus Ulrich and Christian Wiedemann (2008). "Machine Vision Algorithms and Applications". Weinheim: Wiley-VCH. ISBN 9783527407347.