

Visão Computacional: Uma introdução e aplicações no mundo atual

Adriner Maranhão de Andrade
Centro de Ciências Exatas e
Naturais
FURB - Fundação Universidade
Regional de Blumenau
Pomerode, Brasil
adriner_andrade@hotmail.com

Jorge Guilherme Kohn
Centro de Ciências Exatas e
Naturais
FURB - Fundação Universidade
Regional de Blumenau
Gaspar, Brasil

Fábio Luiz Fischer
Centro de Ciências Exatas e
Naturais
FURB - Fundação Universidade
Regional de Blumenau

Resumo—Este trabalho apresenta alguns conceitos relacionados à visão computacional e sua aplicação no cenário atual. Há cada ano que passa, a computação se moderniza, sendo a visão computacional um dos ramos de grande impacto neste processo de modernização. A extração de informação através de imagens pode ajudar

a identificar tumores, produzir carros autônomos e agilizar processos que antes eram feitos somente de maneira orgânica e manual.

Palavras-chave—visão computacional, inteligência artificial, computação, análise de imagens

I. INTRODUÇÃO

A visão computacional é um ramo da computação que começou desde a década de 1960, com Larry Roberts. Porém, vem se popularizando e ganhando ainda mais força recentemente à medida que o poder de processamento dos computadores aumenta, mais ferramentas são construídas, e assim, consequentemente, as possibilidades de aplicações no mundo real. O principal objetivo da área, é conseguir transformar uma imagem em informação. O processamento de imagens, podemos dizer então, que é uma etapa dentro de um processo extremamente complexo. Já se é possível atingir resultados notórios, mas o almejado é conseguir atingir um nível onde seja possível compreender informações da mesma maneira, senão melhor, que somos capazes com a arquitetura visual orgânica presente em nosso corpo.

desenvolvido e apresenta grande capacidade de reconhecimento de padrões e de interpretação em cima daquilo que vemos. Isso nos coloca em vantagem em diversas situações, porém, em alguns casos, podemos ser superados pelas máquinas. Isso porque o foco de uma máquina em determinada atividade pode ser maior, e todos os seus recursos direcionados para uma determinada atividade, o que é muito difícil de acontecer com o corpo humano que lida com diversas atividades ocorrendo em paralelo. Além disso, existem cenários em que o cérebro é de certo modo limitado ou acaba se confundindo na tentativa de aplicar mecanismos de correção ou interpretar a realidade. As “fig. 1” e “fig.2” abaixo, representam bem este cenário.

II. O QUE É VISÃO COMPUTACIONAL

Pode ter ser subentendida em duas vertentes. A primeira, a partir de um ponto de vista biológico, se trata de uma implementação artificial do mecanismo presente no nosso corpo. Já de um ponto de vista de engenharia da computação, se trata de a criação de mecanismos que podem performar as mesmas atividades que o olho humano [1]. Basicamente, a visão computacional, visa a interpretação do conteúdo das imagens. O nosso mecanismo visual, é altamente

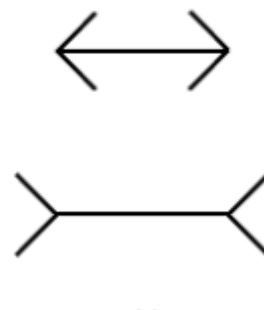


Fig. 1. As duas retas horizontais possuem o mesmo comprimento, mas a percepção é de que possuem tamanhos diferentes [2].
Extraído de Richard Szeliski, *Computer Vision: Algorithms and Applications*.

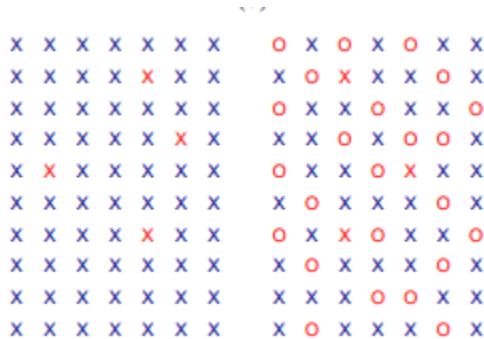


Fig. 2. Contar a quantidade de 'X' vermelhos no lado esquerdo é uma tarefa muito mais fácil do que no lado direito devido a grande quantidade de percepções paralelas [2]. Isso é uma tarefa que certamente a máquina poderia desempenhar um melhor papel que o cérebro humano. Extraído de Richard Szeliski, *Computer Vision: Algorithms and Applications*.

III. HISTÓRIA DA VISÃO COMPUTACIONAL

A visão computacional vem se desenvolvendo desde a década de 1960, onde em uma tese de Ph.D. de Larry Roberts, levantou a possibilidade de extrair informações geométricas tridimensionais a partir de um plano de somente duas dimensões contendo blocos geométricos de faces planas (poliedros). Diversos pesquisadores se basearam no seu trabalho e se desenvolveram dentro do ramo de visão computacional, trazendo novos trabalhos para a área. Técnicas desempenhadas nesse trabalho como a detecção de vértices e segmentação foram essenciais na análise de imagens do mundo real [1]. Segmentação trata-se da capacidade de distinguir objetos diferentes na imagem, e conforme o próprio nome, segmentá-los como dado de um plano 2D.

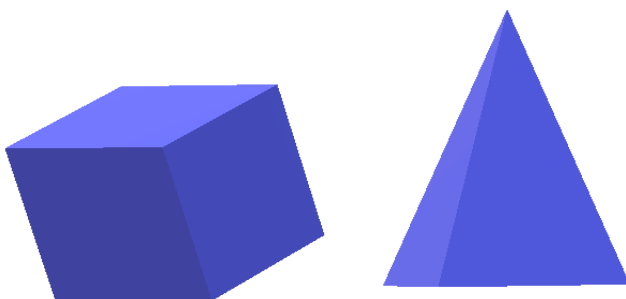


Fig. 3. Cubos e triângulos são exemplos de poliedros. Elaborado pelos autores.

Atualmente, as pesquisas do ramo estão altamente focadas na aplicação de aprendizado de máquina juntamente com técnicas de visão computacional. Para auxiliar este processo, também está crescendo a quantidade de dados caracterizados (ou rotulados) disponíveis pela internet, permitindo aumentar a capacidade de análise de imagens sem supervisão humana [2]. Com a melhoria na análise de objetos, vem se trabalhando também na área, a análise de cenários. Neste caso, se buscaria compreender a relação entre os objetos e seus possíveis estados, como por exemplo, diferenciar uma pessoa pulando de uma pessoa parada.

IV. FUNÇÕES TÍPICAS

Quando visualizamos algo uma série de fatores são iniciados, a retina humana encaminha imagens para o córtex cerebral que por sua vez efetua funções dentre elas o reconhecer, analisar e obter padrões para tomar a melhor decisão, passamos bastante tempo de nossas vidas repetindo este processo e o aperfeiçoando. O qual para Jitendra Malik [3] são funções típicas de qualquer visão computacional, podemos entender melhor na "fig. 4" apelidado de "Os três R's da visão", são eles *Reconhecimento*, *Reorganização* e *Reestruturação*, não necessariamente nesta ordem.

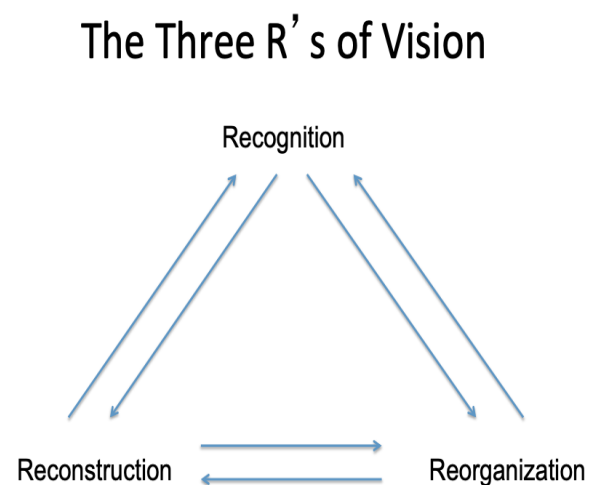


Fig 4. The three R's of visio, elaborada pelo Jitendra Malik [3].
Extraído de <http://www-inst.eecs.berkeley.edu/~cs280/sp15/index.html>.

Estes fatores mostrados acima se assemelham bastante com o processo humano de reconhecer e quantificar os objetos, ou seja, o computador recebe uma imagem e a transforma em vários fragmentos e fragmentos de fragmentos que são conhecidos como camadas e após existe uma quantificação, processo conhecido como CNN (redes neurais convolucionais), proposto por Yann LeCun [4] em 1998. Esta arquitetura é demonstrada na "fig. 5", a qual os melhores algoritmos e visão computacional são baseados. Conseguimos identificar na figura o processo em que uma visão computacional quantificar dados e consegue reconhecer objetos, por meio desta rede neural profunda.

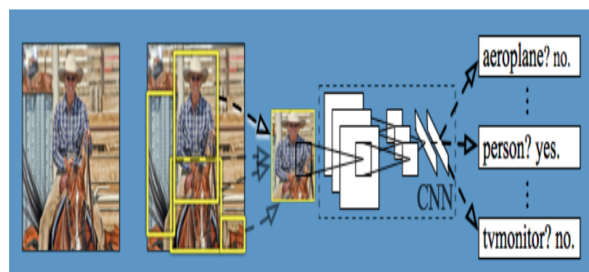


Fig 5. R-CNN: Regions with CNN features
Girshick, Donahue, Darrell & Malik (CVPR 2014). Extraído de <http://www-inst.eecs.berkeley.edu/~cs280/sp15/index.html>.

Assim que um objeto é quantificado conseguimos identificar padrões, aplicar transformações, reconhecimento e todos os outros fatores que existem em uma visão, de forma igual ou em algumas vezes até de forma melhor que a visão humana, percebendo assim quem está ao nosso redor ou até mesmo o que está acontecendo, podendo obter sentimentos que os olhos humanos não são capazes de perceber.

V. APLICAÇÕES NA ATUALIDADE

A maioria dos sistemas já utiliza visão computacional, seja para validar não repúdio ou localizar uma pessoa em seu sistema de forma fácil e amigável. Mas a visão vai muito além disso, com carro autônomos, fazer uma imagem falar ou para reconhecer células cancerígenas de forma eficaz, em vezes melhor do que a visão humana. Um bom exemplo são os algoritmos de detecção em massa de

câncer de mama, que segundo a pesquisadora *Simara Vieira da Rocha* [5], é a maior causa de morte de mulheres entre os 35 a 55 anos. Com a utilização do sistema CAD, Computer-Aided Diagnostic [6], propõe reduções significativas de falsos positivos e com ajuda de processamento de imagem e reconhecimento de padrão identificar fases iniciais destes cânceres. Mas o sistema CAD não é limitado a apenas um tipo de câncer consegue identificar com processamento de imagem qualquer tipo de câncer, seja ele pulmonar, cerebral entre outro. Na "fig. 6" temos imagens de uma identificação com o sistema CAD.

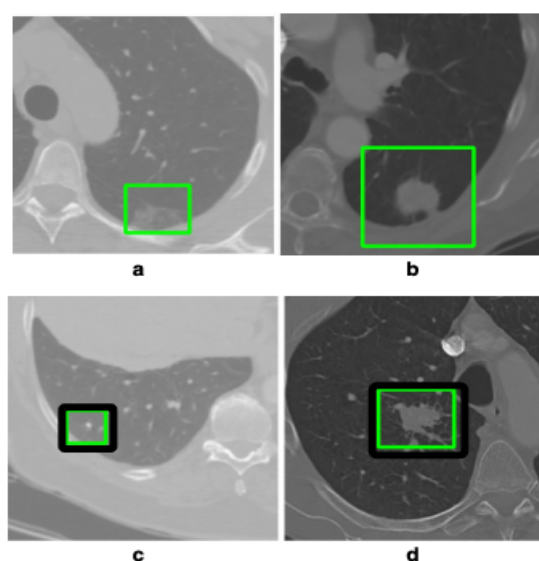


Fig. 6: Exemplos de diferentes tipos de nódulos pulmonares. **A** um nódulo em vidro fosco com forma irregular, **B** sólido nódulo juxtavascular em forma ovóide, **C** nódulo esférico sólido com 4 mm de diâmetro e **D** dérmico juxtavascular nódulo em forma ovóide.
Extraído de <https://core.ac.uk/download/pdf/81846540.pdf>.

Outra importante utilização na atualidade é através de carro autônomos, empresas como a Tesla, Uber, já possuem carros com potencial de realizar trajetórias sem a intervenção humana, respeitando as leis de trânsito, sem causar danos a outros carros que estão ao seu redor. Segundo o G1 [7], a Tesla com os carros autônomos de modelo X e S, vem equipados com cerca de 8 câmeras que permitem uma visão de 360 graus e 12 sensores de diferentes tipos para capturar e reconhecer qualquer objeto. Gerando uma visão apresentada na "fig. 7", gerando espetacular uma imagem precisa para cálculo de trajetória ou desvio de

algum objeto que podem estar pelo caminho utilizando visão computacional.

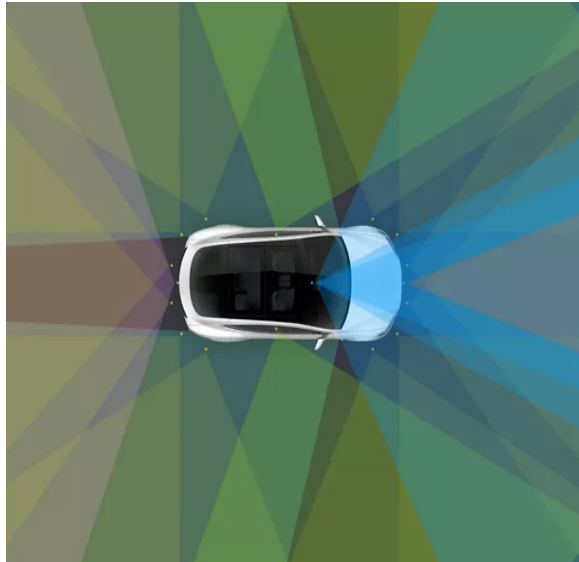


Fig. 7, Visualização de como os sensores do Tesla operam. Extraído de <https://www.newscientist.com/article/2109787-tesla-shows-off-fully-autonomous-car-in-new-video-demonstration>.

Apesar do reconhecimento de imagem já ser bem preciso ainda enfrentamos problemas para reconhecer algo muito parecido ou que pode ser feito de diferentes formas.

VI. DESENVOLVENDO UMA SOLUÇÃO DE VISÃO COMPUTACIONAL

Com o objetivo de aprofundamento no tema, implementamos um software que realiza o reconhecimento de placas de identificação de veículos brasileiros. O programa foi desenvolvido em Python e utiliza bibliotecas para funções de processamento de imagens e reconhecimento de texto. As entradas da aplicação são imagens frontais, traseiras ou diagonais de veículos automotores que contenham uma placa de identificação no padrão brasileiro. As imagens são, então, pré-processadas pela operação morfológica de dilatação e binarização, com o objetivo de remover imperfeições físicas desnecessárias e realçar as características desejadas para a análise do algoritmo.

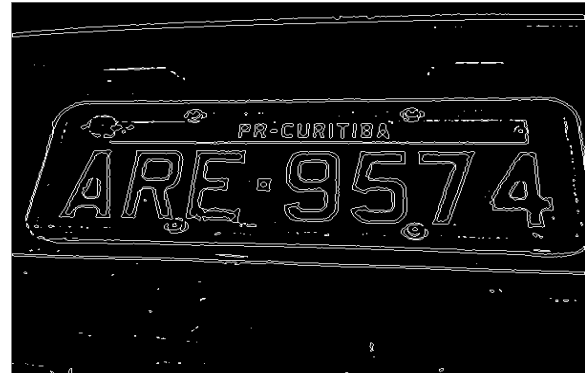


Fig. 8. Execução do pré-processamento e dos contornos encontrados de uma imagem contendo uma placa de identificação veicular.

A seguir, o programa busca os principais contornos presentes na imagem, como apresentado na “fig. 8”. Cada um dos contornos encontrados são percorridos e filtrados de acordo com seu tamanho e probabilidade de ser um caráter. Além disso, caracteres muito distantes uns dos outros são separados em palavras diferentes, possibilitando a identificação de múltiplas placas veiculares na mesma imagem.

Assim que os conjuntos de caracteres são propriamente identificados, o algoritmo realiza os ajustes de rotação e recorta a imagem original para que apenas o conjunto de caracteres seja exibido. A imagem recortada, então, é utilizada para o reconhecimento do texto.

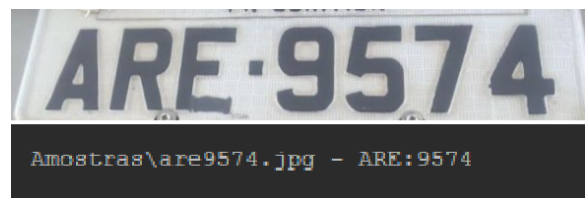


Fig. 9. Resultado da execução do algoritmo, contendo a imagem recortada para exibição do conjunto de caracteres e a identificação do texto.

Apesar de obter resultados aceitáveis, o algoritmo ainda sofre das dificuldades comuns de identificação e processamento de imagens, como por exemplo, iluminação de baixa qualidade. Além disso, existem casos de adulteração humana [8] nos números da placa de identificação que tornam indispensável a presença de um humano para a conclusão de uma análise consistente.



Fig 10. Exemplo de adulteração na placa de identificação veicular, onde os números reais são alterados com a utilização de uma fita isolante. Extraído de <http://www.corneliodigital.com/index.php?s=noticia&&id=33527>.

VII. CONCLUSÃO

A evolução da tecnologia está intrinsicamente ligada com a visão computacional. Isso pois, analisando de uma maneira mais profunda, estamos falando do desenvolvimento de uma área capaz de extrair informações de dados brutos gerados a partir da captura de fótons de luz (como uma imagem é formada). Não precisamos nos limitar somente ao que nossos olhos veem, mas a todo tipo de luz que possa ser capturada. É importante considerar também, que com o aprimoramento contínuo do ferramental disponível, será um tema cada vez mais difundido e utilizado também por soluções comerciais, podendo resultar em investimentos em novas pesquisas que irão alavancar a área.

REFERÊNCIAS

- [1] Huang, T. Vandoni, Carlo, E (ed.). *Computer Vision : Evolution And Promise*. 19th CERN School of Computing. Geneva: CERN. pp. 21–25, Novembro 1996. <http://cds.cern.ch/record/400313/files/p21.pdf>
- [2] SZELISKI, Richard. *Computer Vision: Algorithms and Applications*. New York City: Springer. pp. 3–19, outubro 2010.
- [3] MALIK, Jitendra. *Department of EECS - UC Berkeley*. <http://www-inst.eecs.berkeley.edu/~cs280/sp15/index.html>

- [4] LECUN, Yann. *Convolutional Networks for Images, Speech, and Time-Series*. <http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf>
- [5] ROCHA, Simara Vieira de *DETECÇÃO DE MASSAS EM IMAGENS DA MAMA USANDO ÍNDICES DE DIVERSIDADE E ALGORITMOS DE SEGMENTAÇÃO EM GRAFO*, <http://www.periodicoseletronicos.ufma.br/index.php/cadernosdesquisa/article/view/1757>.
- [6] Islam Reda; Ahmed Shalaby; Fahmi Khalifa ; Mohammed Elmogy ; Ahmed Aboulfotouh, *Computer-aided diagnostic tool for early detection of prostate cancer*, <https://ieeexplore.ieee.org/document/7532843>.
- [7] G1, *Tesla anuncia sistema 100% autônomo para todos os seus carros*, <http://g1.globo.com/carros/noticia/2016/10/tesla-anuncia-sistema-100-autonomo-para-todos-os-seus-carros.html>.
- [8] *Cornélio Digital, Mulher altera placas de carro com fita isolante para não receber multas no pedágio*, <http://www.corneliodigital.com/index.php?s=noticia&&id=33527>.