

# Visão Computacional

01 - Introdução



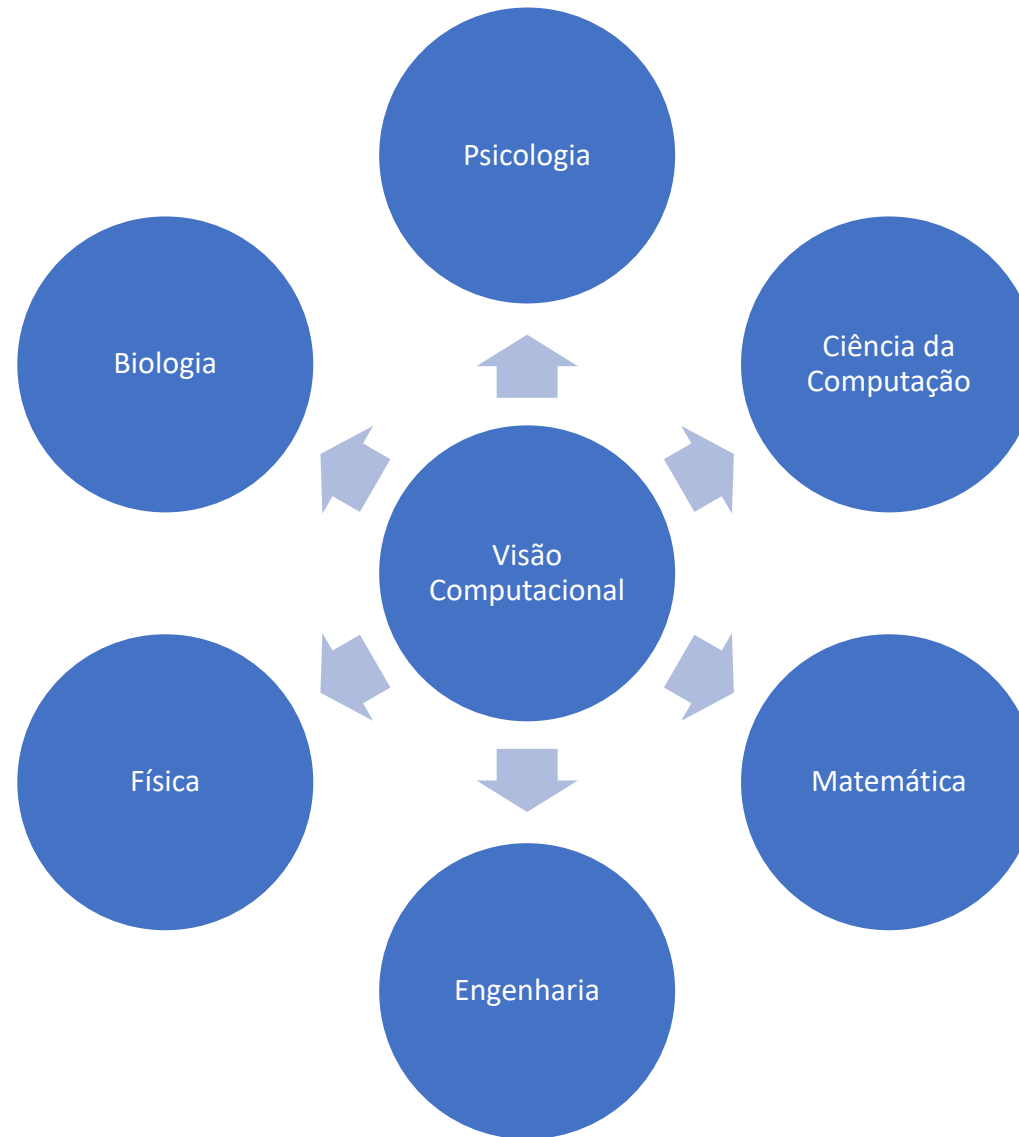
# O que é Visão Computacional

- É o estudo de dados visuais.
- Construir algoritmos que consigam interpretar imagens.
  - A partir de uma imagem, extrair alguma informação com significado sobre ela (o que se passa nela).
  - Ter um imagem como entrada e uma informação como saída.

# Motivação

- A quantidade de dados visuais no mundo cresce exponencialmente.
  - Em sua grande parte, provocado pelo aumento da quantidade de sensores.
  - Crescimento de mídias sociais e outros serviços que usam fotos e vídeos.
  - Por exemplo, 500 horas de vídeos são subidas para o youtube por minuto (2020).
  - A grande maioria dos dados da internet são de conteúdo visual.
- Existe a necessidade de analisar esses dados de forma eficiente.
  - É impossível para pessoas conseguirem os dados na velocidade que surgem.
- No entanto, compreender um dado visual não é uma tarefa visual para os computadores.

# Contexto





# Computação Gráfica

- Os modelos utilizados, em geral, são desenvolvidos na física e em computação gráfica.
  - As duas áreas buscam modelar como os objetos se movem, animam, como a luz reflete nas suas superfícies, como a luz refrata através das lentes de câmeras (e olhos humanos) e como projetar isso em uma tela plana.
- Ainda não existe perfeição, mas os avanços na área são impressionantes: renderizar cenas complexas do mundo real contendo objetos parados ou em movimento.

# Visão humana

- Enquanto humanos, nós conseguimos perceber a estrutura tridimensional do mundo sem problemas.
- Facilmente conseguimos observar diferenças de iluminação, transparência, cores, textura...
- Como nós enxergamos é tema de pesquisas por muito tempo e, apesar de termos evoluído absurdamente, ainda não chegamos a uma solução completa.

# Visão Computacional

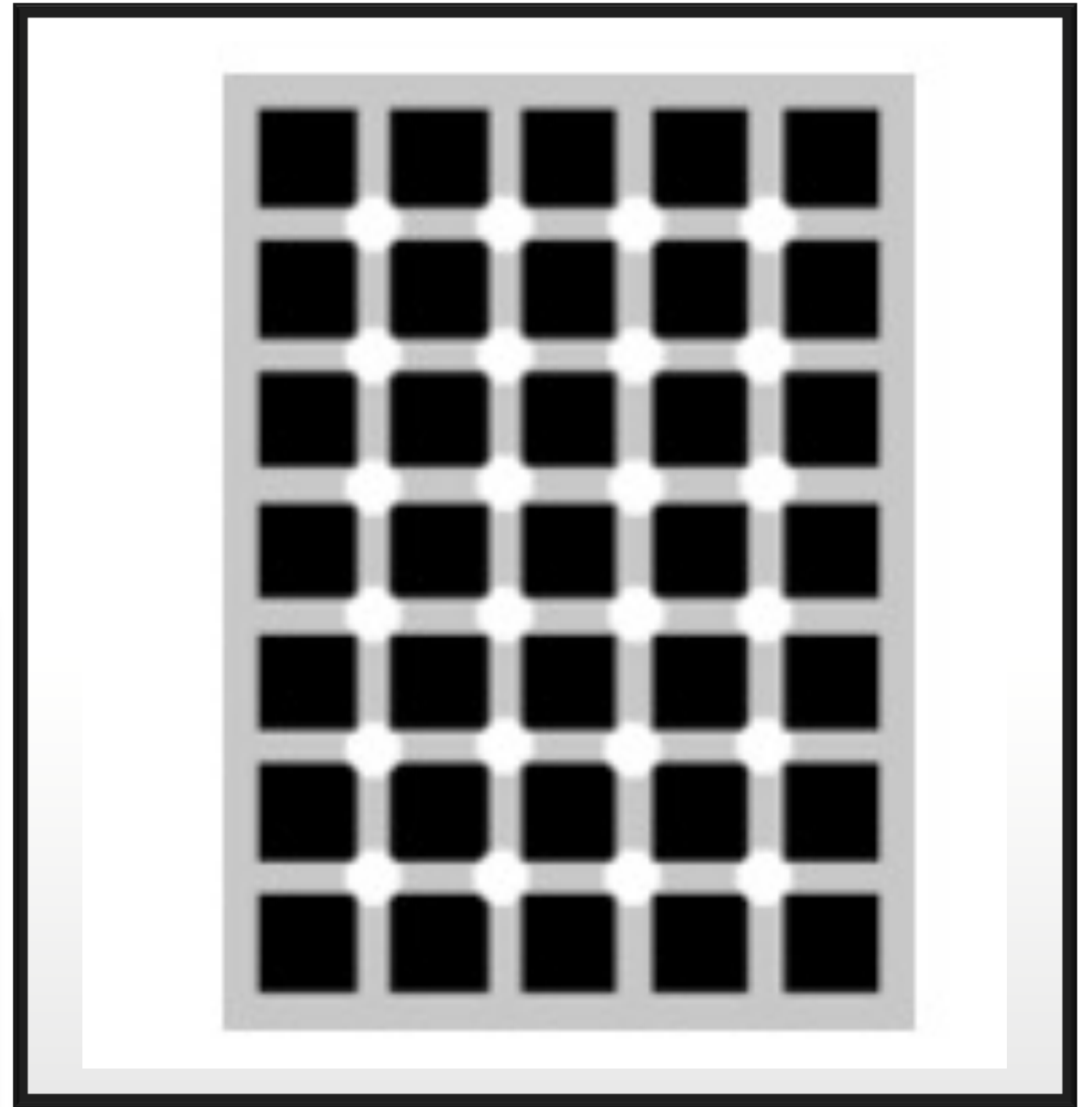
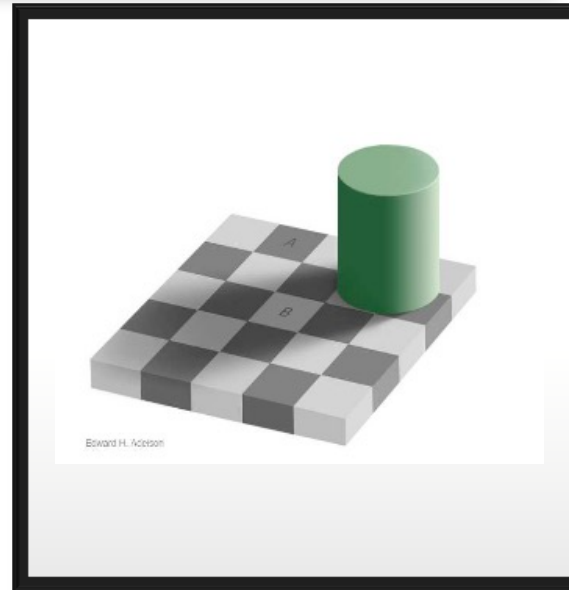
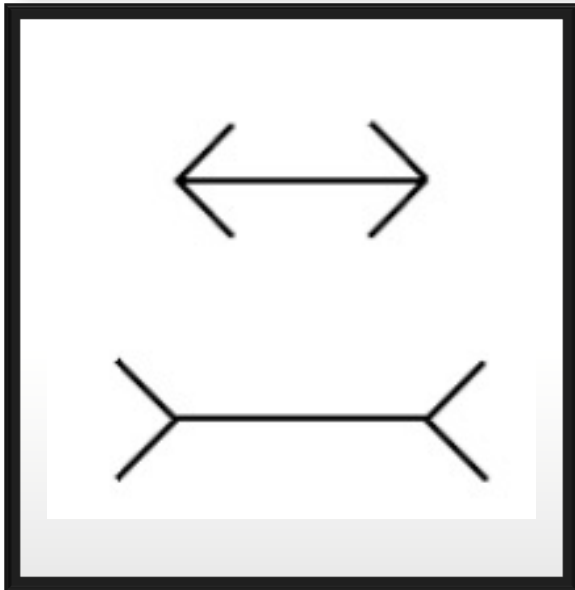
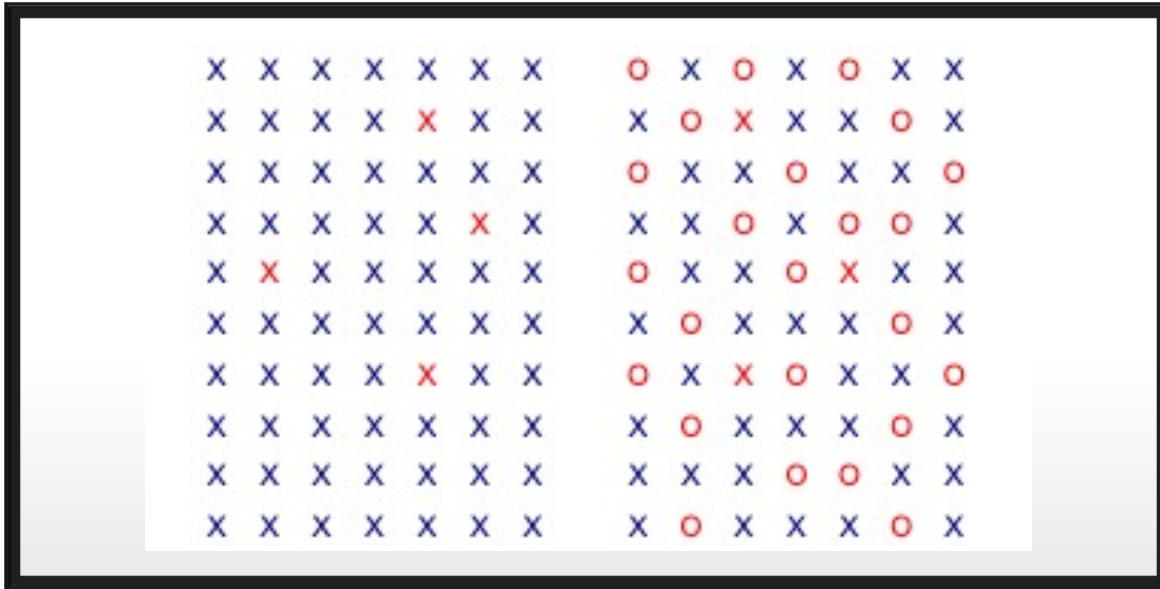
- Na Visão Computacional, estamos tentando fazer o inverso: descrever o mundo como vemos para reconstruir suas propriedades
  - Iluminação, formato, cores
- Em animais, essa tarefa é simplesmente “transparente”
- Em máquinas, essa tarefa é complexa e “error prone”
- É uma tarefa extremamente complexa e é comumente enxergada como simples.
- Existe uma popularização da área nos últimos anos pela utilização em tarefas cotidianas.

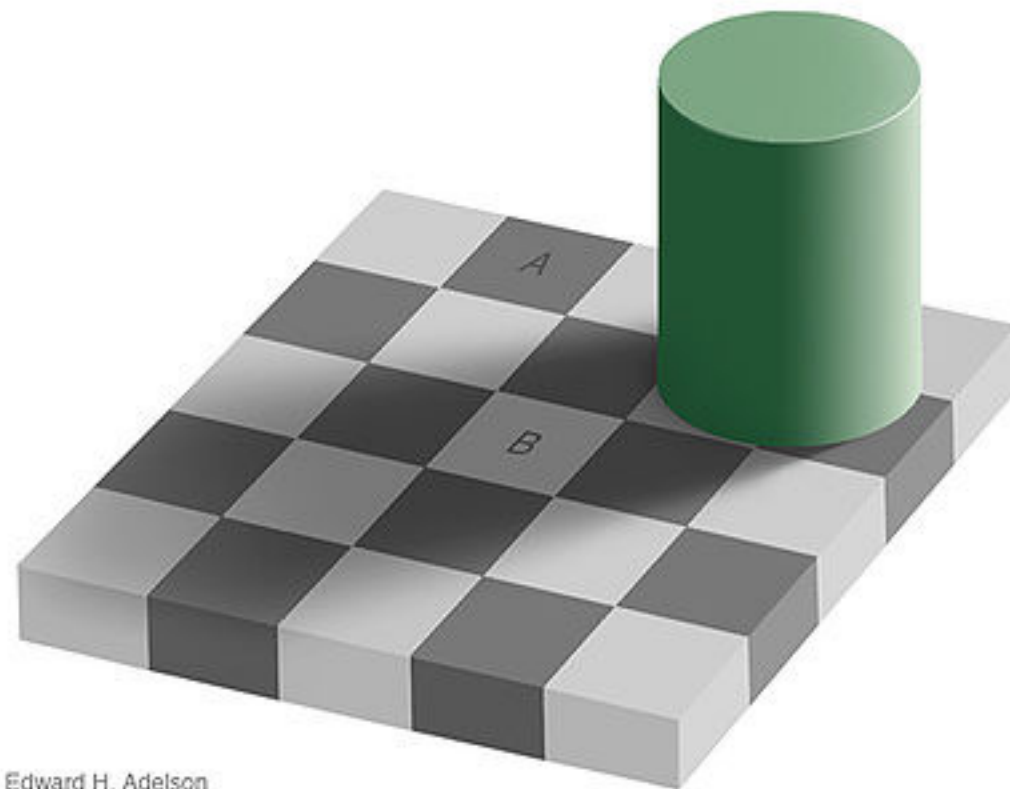
# Problema inverso

- Por que é tão difícil atribuir visão a uma máquina?
  - Nós buscamos recuperar informações incompletas (e desconhecidas) para especificar um problema.
  - Por isso, aprendizado de máquina acaba sendo um forte aliado.
  - Modelar um problema de visão computacional é mais complexo que um problema de reconhecimento de som, por exemplo.

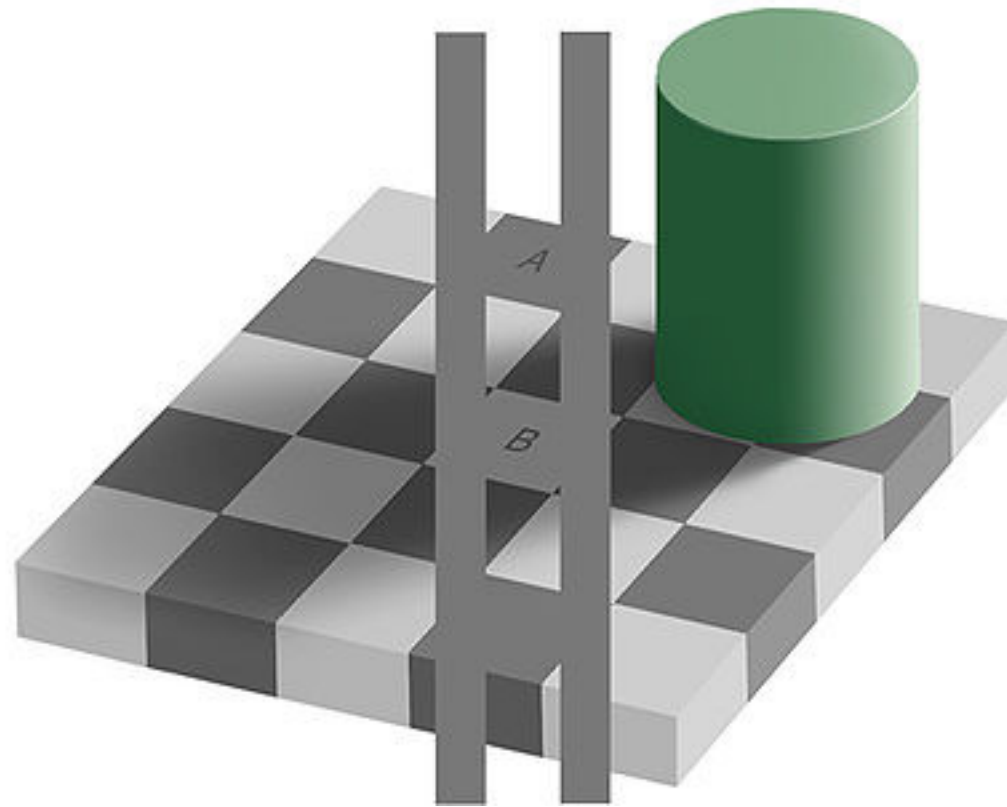








Edward H. Adelson





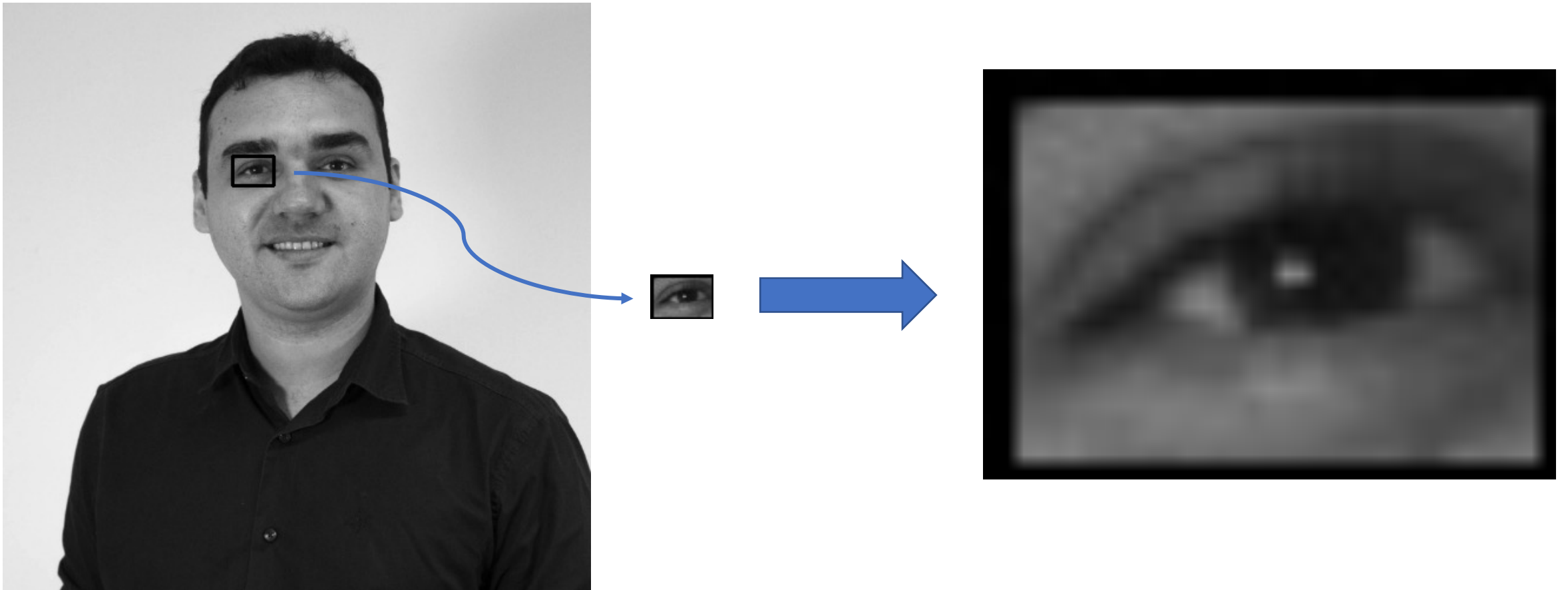
## Gap semântico

---

- Existe uma diferença de percepção entre o que a gente entende e o que a máquina entende.



# Gap semântico



# Gap semântico

[illegible]

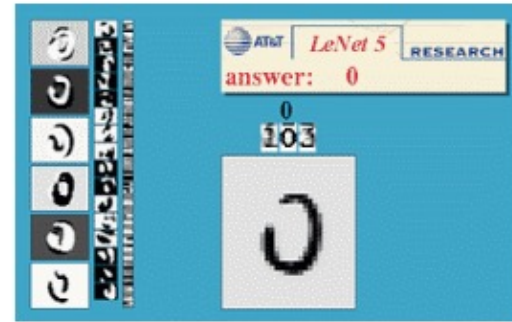


# Aplicações

- OCR
- Inspeção de máquinas
- Comércio
- Logística
- Medicina
- Veículos autônomos
- Modelagem 3D de prédios
- Vigilância
- Biometria

- *Stitching*
- *Controle de exposição*
- *Morphing*
- *Modelagem 3D*
- *Estabilização e matching de vídeo*
- *Navegação de ambientes*
- *Detecção facial*
- *Autenticação visual*





(a)



(b)



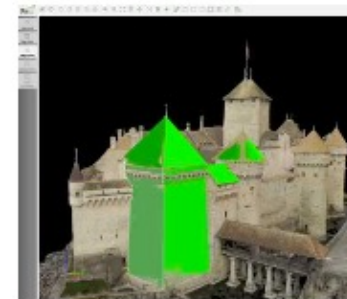
(c)



(d)



(e)



(f)



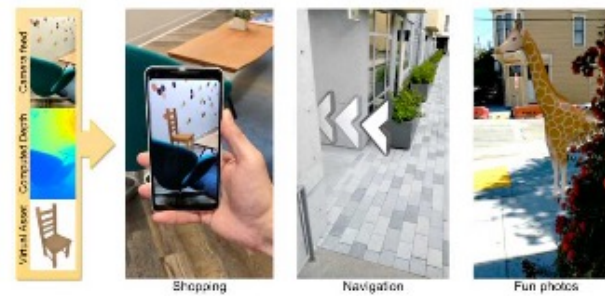
(a)



(b)

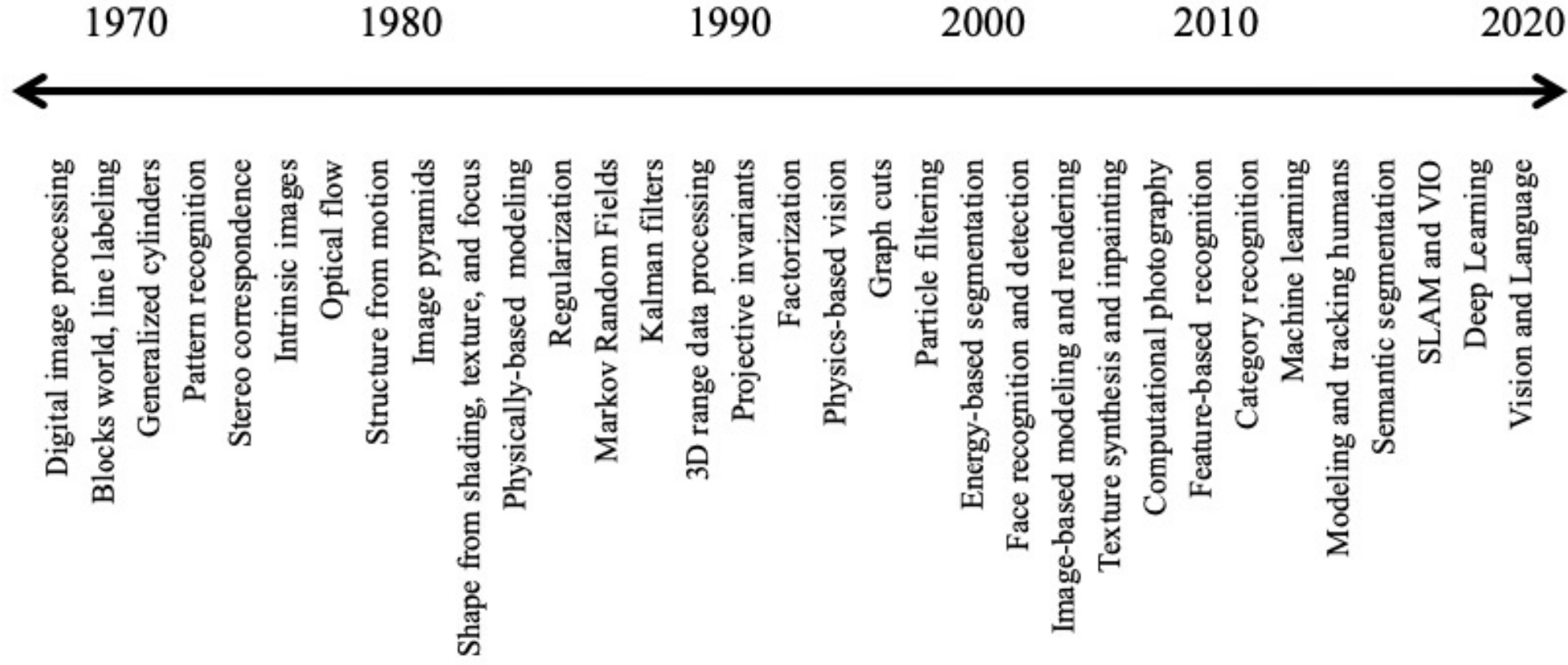


(c)



(d)

# Evolução

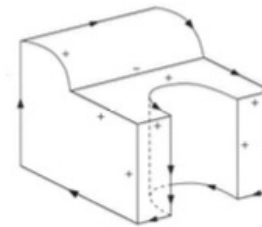


# 1970

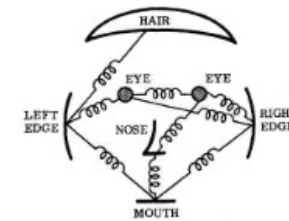
Digital image processing  
Blocks world, line labeling  
Generalized cylinders  
Pattern recognition  
Stereo correspondence  
Intrinsic images  
Optical flow



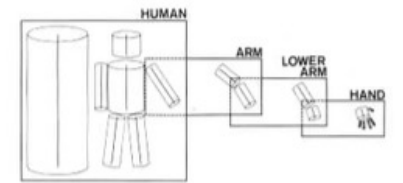
1970



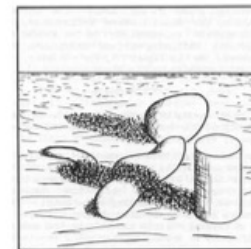
(a)



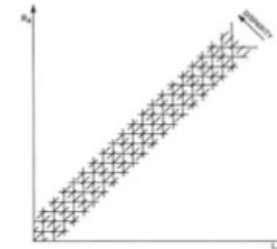
(b)



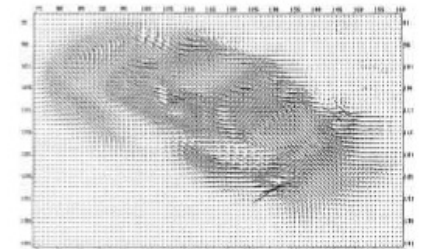
(c)



(d)



(e)



(f)

# 1980

Image pyramids  
Shape from shading, texture, and focus  
Physically-based modeling  
Regularization  
Markov Random Fields  
Kalman filters  
3D range data processing

80

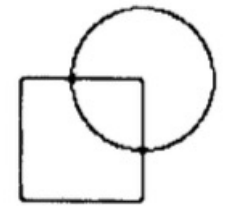
15



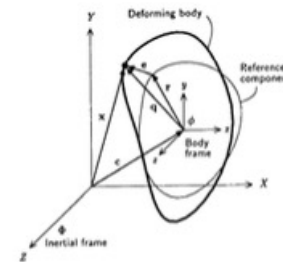
(a)



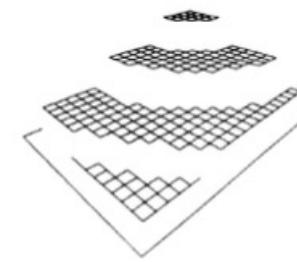
(b)



(c)



(d)



(e)



(f)

# 1990

3D range data processing  
Projective invariants  
Factorization  
Physics-based vision  
Graph cuts  
Particle filtering  
Energy-based segmentation  
Face recognition and detection

1990

2000



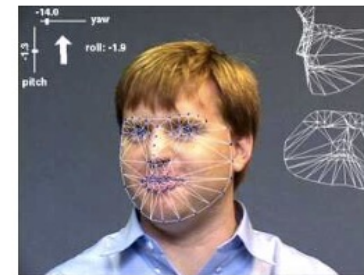
(a)



(b)



(c)



(d)



(e)



(f)



# 2000

Energy-based segmentation

Face recognition and detection

Image-based modeling and rendering

Texture synthesis and inpainting

Computational photography

Feature-based recognition

Category recognition

2000

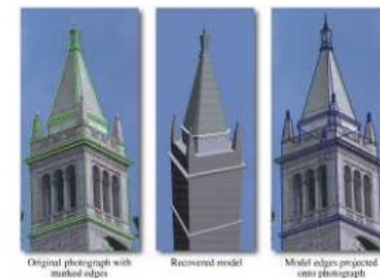
2010



(a)



(d)



(b)



(e)



(c)



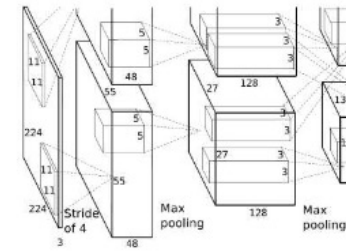
(f)

# 2010

Machine learning  
Modeling and tracking humans  
Semantic segmentation  
SLAM and VIO  
Deep Learning  
Vision and Language



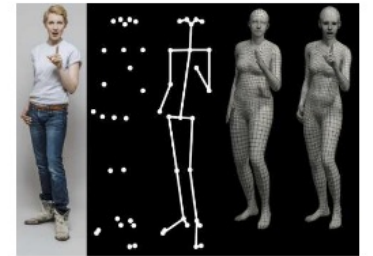
2020



(a)



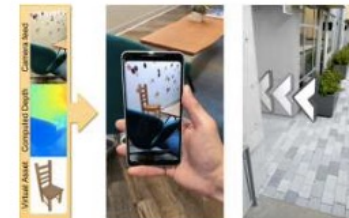
(b)



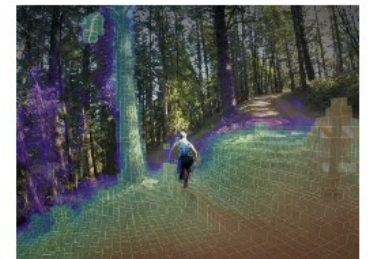
(c)



(d)



(e)



(f)



# Deep Learning

- A partir de 2012, com o surgimento mais evidente das redes convolucionais para o contexto de VC, a área cresceu exponencialmente.
  - O modelo SuperVision (AlexNet) ganhou o desafio [IMAGENET](#) ([Krizhevsky et al](#)).
- Desde então, o ganhador do desafio é uma nova rede neural mais eficiente.
  - Redes mais profundas.
  - ZFNet (2013)
  - Inception (GoogleNet) (2014)
  - VGG (Oxford) (2014)
  - ResNet (MSRA) (2015)

# Espaço para trabalho...

- Apesar de ter evoluído bastante, ainda existe muito a ser feito no âmbito da Visão Computacional:
  - Segmentação semântica (segmentar partes da imagem considerando o context)
  - Reconhecimento de atividade (o que cada pessoa/objeto faz na imagem)
  - Realidade virtual e aumentada
  - Descrição semântica de imagens (compreensão de cena completa de uma imagem)

# Referências

- Richard Szeliski. Computer Vision: Algorithms and Applications. 2nd Edition. 2021
  - <http://szeliski.org/Book/>
- Stanford CS231n: Convolutional Neural Networks for Visual Recognition Course
  - <http://cs231n.stanford.edu>
- Georgia Tech CS 6476: Computer Vision Course
  - <https://omscs.gatech.edu/cs-6476-computer-vision>