

# Visão Computacional

Breve introdução com foco na tarefa de reconhecimento

# Sumário

## O gap semântico

### Visão Computacional

Para além de pixels

Reconhecimento em Visão Computacional

Técnicas envolvidas

Abordagem tradicional

Abordagem com Deep Learning

Aplicações

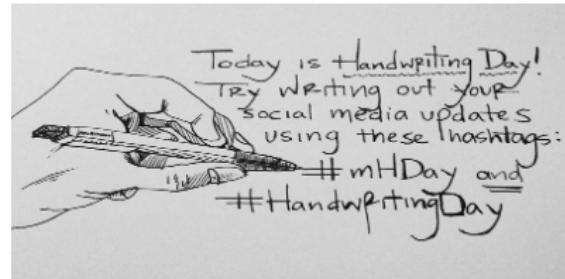
# Percepção humana

O que você percebe nestas imagens?



# Percepção humana

O que você percebe nestas imagens?



A facilidade com que respondemos a essa pergunta se deve ao nosso sistema visual **nativo** extremamente poderoso!

# A percepção do computador

As matrizes de *pixels*

À primeira vista...

# A percepção do computador

## As matrizes de *pixels*

À primeira vista...

```
[[ 42  23  19 ... , 21  29  25]
 [ 40  40  36 ... , 24  24  21]
 [ 28  30  36 ... , 30  13  27]
 ...
 [[115  78  45 ... , 28  36  17]
 [ 67  78 192 ... , 35  31  36]
 [ 67  79 104 ... , 34  32  31]]]
```

```
[[222  224  224 ... , 204  201  200]
 [223  225  223 ... , 201  203  204]
 [226  226  226 ... , 204  202  205]
 ...
 [[210  203  208 ... , 192  188  189]
 [206  206  207 ... , 190  188  189]
 [210  208  210 ... , 191  193  185]]]
```

```
[[138 137 137 ... , 107 107 107]
 [135 134 134 ... , 107 107 107]
 [130 129 129 ... , 107 107 107]
 ...
 [[145 145 146 ... , 142 142 142]
 [146 145 144 ... , 144 144 145]
 [147 146 144 ... , 145 145 146]]]
```

```
[[ 48  45  40 ... , 28  29  31]
 [ 45  46  43 ... , 28  29  30]
 [ 41  43  43 ... , 27  27  29]
 ...
 [[101 101 103 ... , 64  51  32]
 [ 98  97  99 ... , 63  71  57]
 [ 97  97  97 ... , 38  57  65]]]
```

# A percepção do computador

## As matrizes de pixels

À primeira vista...

```
[[ 42  23  19 ... , 21  29  25]
 [ 40  40  36 ... , 24  24  21]
 [ 28  30  36 ... , 30  13  27]
 ...
 [[115  78  45 ... , 28  36  17]
 [ 67  78 192 ... , 35  31  36]
 [ 67  79 104 ... , 34  32  31]]]
```

```
[[222  224  224 ... , 204  201  200]
 [223  225  223 ... , 201  203  204]
 [226  226  226 ... , 204  202  205]
 ...
 [[210  203  208 ... , 192  188  189]
 [206  206  207 ... , 190  188  189]
 [210  208  210 ... , 191  193  185]]]
```

```
[[138 137 137 ... , 107 107 107]
 [135 134 134 ... , 107 107 107]
 [130 129 129 ... , 107 107 107]
 ...
 [[145 145 146 ... , 142 142 142]
 [146 145 144 ... , 144 144 145]
 [147 146 144 ... , 145 145 146]]]
```

```
[[ 48  45  40 ... , 28  29  31]
 [ 45  46  43 ... , 28  29  30]
 [ 41  43  43 ... , 27  27  29]
 ...
 [[101 101 103 ... , 64  51  32]
 [ 98  97  99 ... , 63  71  57]
 [ 97  97  97 ... , 38  57  65]]]
```

## Imagens digitais monocromáticas

Matrizes  $I_j \in \mathbb{M}_{w_j \times h_j}([0, \dots, 255])$  ou funções

$f_j : \{1, \dots, w_j\} \times \{1, \dots, h_j\} \rightarrow [0, 255]$ , onde  $w_j$  é a largura e  $h_j$  é a altura da imagem  $j$ .

# O *gap* semântico

Percepção humana × percepção da máquina

## *Gap* semântico

Diferença entre a maneira como o ser humano **percebe** o conteúdo de uma imagem e como a imagem é **representada** de forma manipulável no computador [Ros17].

# O *gap* semântico

Percepção humana × percepção da máquina

## *Gap* semântico

Diferença entre a maneira como o ser humano **percebe** o conteúdo de uma imagem e como a imagem é **representada** de forma manipulável no computador [Ros17].

Em outras palavras: enquanto vemos gatos e cachorros, o computador apenas percebe números.



# O *gap* semântico

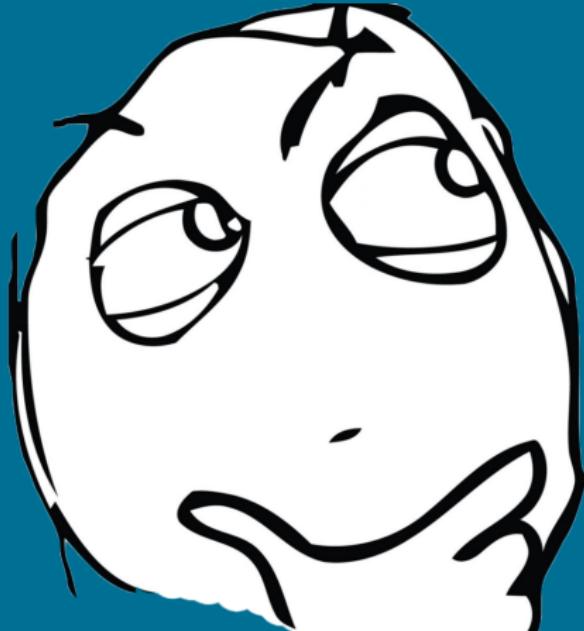
Percepção humana × percepção da máquina

## *Gap* semântico

Diferença entre a maneira como o ser humano **percebe** o conteúdo de uma imagem e como a imagem é **representada** de forma manipulável no computador [Ros17].

Em outras palavras: enquanto vemos gatos e cachorros, o computador apenas percebe números.





Será que podemos vencer o *gap semântico*?

# Sumário

O gap semântico

## Visão Computacional

Para além de pixels

Reconhecimento em Visão Computacional

Técnicas envolvidas

Abordagem tradicional

Abordagem com Deep Learning

Aplicações

# Visão Computacional

O que é?

## Visão Computacional

Visão Computacional é uma área da Ciência da Computação cujo propósito é capacitar os computadores para extraírem informações de imagens, ou seja, permitir que tenham um entendimento visual do mundo.

# Visão Computacional

O que é?

## Visão Computacional

Visão Computacional é uma área da Ciência da Computação cujo propósito é capacitar os computadores para extraírem informações de imagens, ou seja, permitir que tenham um entendimento visual do mundo.

Entre as tarefas compreendidas nesse âmbito, estão principalmente [Sze11]:

# Visão Computacional

O que é?

## Visão Computacional

Visão Computacional é uma área da Ciência da Computação cujo propósito é capacitar os computadores para extraírem informações de imagens, ou seja, permitir que tenham um entendimento visual do mundo.

Entre as tarefas compreendidas nesse âmbito, estão principalmente [Sze11]:

- ▶ Análise de movimento;

# Visão Computacional

O que é?

## Visão Computacional

Visão Computacional é uma área da Ciência da Computação cujo propósito é capacitar os computadores para extraírem informações de imagens, ou seja, permitir que tenham um entendimento visual do mundo.

Entre as tarefas compreendidas nesse âmbito, estão principalmente [Sze11]:

- ▶ Análise de movimento;
- ▶ Reconstrução de cenas;

# Visão Computacional

O que é?

## Visão Computacional

Visão Computacional é uma área da Ciência da Computação cujo propósito é capacitar os computadores para extraírem informações de imagens, ou seja, permitir que tenham um entendimento visual do mundo.

Entre as tarefas compreendidas nesse âmbito, estão principalmente [Sze11]:

- ▶ Análise de movimento;
- ▶ Reconstrução de cenas;
- ▶ Restauração de imagens;

# Visão Computacional

O que é?

## Visão Computacional

Visão Computacional é uma área da Ciência da Computação cujo propósito é capacitar os computadores para extraírem informações de imagens, ou seja, permitir que tenham um entendimento visual do mundo.

Entre as tarefas compreendidas nesse âmbito, estão principalmente [Sze11]:

- ▶ Análise de movimento;
- ▶ Reconstrução de cenas;
- ▶ Restauração de imagens;
- ▶ **Reconhecimento.**

# Reconhecimento em Visão Computacional

O que é?

## Reconhecimento

Tarefa de determinar se um certo fato (ocorrência ou atividade) está presente em uma imagem ou vídeo.

# Reconhecimento em Visão Computacional

O que é?

## Reconhecimento

Tarefa de determinar se um certo fato (ocorrência ou atividade) está presente em uma imagem ou vídeo.

Apresenta-se geralmente para três principais propósitos:

# Reconhecimento em Visão Computacional

O que é?

## Reconhecimento

Tarefa de determinar se um certo fato (ocorrência ou atividade) está presente em uma imagem ou vídeo.

Apresenta-se geralmente para três principais propósitos:

- ▶ Classificação de objetos;

# Reconhecimento em Visão Computacional

O que é?

## Reconhecimento

Tarefa de determinar se um certo fato (ocorrência ou atividade) está presente em uma imagem ou vídeo.

Apresenta-se geralmente para três principais propósitos:

- ▶ Classificação de objetos;
- ▶ Identificação;

# Reconhecimento em Visão Computacional

O que é?

## Reconhecimento

Tarefa de determinar se um certo fato (ocorrência ou atividade) está presente em uma imagem ou vídeo.

Apresenta-se geralmente para três principais propósitos:

- ▶ Classificação de objetos;
- ▶ Identificação;
- ▶ Detecção.

# Reconhecimento em Visão Computacional

## Desafios

Imagine que se queira reconhecer os gatos em uma imagem.

Quais os desafios envolvidos nisso?

# Reconhecimento em Visão Computacional

## Desafios

Imagine que se queira reconhecer os gatos em uma imagem.

Quais os desafios envolvidos nisso?

### Variação de ponto de vista

Não importa sob qual ângulo se fotografe um gato: ele continuará sendo um gato.

### Variação de escala

Não importa a que distância o gato estará da câmera: ele continuará sendo um gato.

### Deformação

Um gato pode estar esticando suas pernas ou/e contorcendo seu pescoço para se lambiar, e isso não o faz ser outro ser além de um gato na imagem.

# Visão Computacional

## Desafios

### Oclusão

Um gato pode estar espiando o mundo ao redor de dentro de uma caixa, apenas com a cabeça de fora, e ele continuará sendo um gato à lente de uma câmera em frente à caixa.

### Iluminação

Um gato num estacionamento mal iluminado ainda é um gato.

### Ruído de fundo

Um gato em frente à uma tela de TV repleta de ruído ainda é um gato.

### Variações intra-classe

Gatos de diversas raças, cores e tamanhos serão sempre gatos.

# Técnicas envolvidas

## Processamento de imagens digitais

### De nível baixo

Operações primitivas para remoção de ruídos, melhoria de contraste e nitidez. Aqui, a entrada e a saída do sistema são sempre imagens.

### De nível médio

Segmentação e descrição dos objetos em uma forma adequada para o processamento e classificação de objetos individuais. Nesta, a entrada é uma imagem, e a saída geralmente é um conjunto de atributos extraídos da imagem.

### De nível alto

Reconhecimento de objetos individuais, geralmente por meio de técnicas de reconhecimento de padrões (*Machine Learning*).

# Técnicas envolvidas

## *Machine Learning*

### *Machine Learning*

Área voltada à construção de modelos computacionais capazes de aprimorar seu desempenho na resolução de problemas a partir da experiência [Mit97].

# Técnicas envolvidas

## *Machine Learning*

### *Machine Learning*

Área voltada à construção de modelos computacionais capazes de aprimorar seu desempenho na resolução de problemas a partir da experiência [Mit97].

Entre as técnicas mais comuns nesse contexto, destacam-se:

# Técnicas envolvidas

## *Machine Learning*

### *Machine Learning*

Área voltada à construção de modelos computacionais capazes de aprimorar seu desempenho na resolução de problemas a partir da experiência [Mit97].

Entre as técnicas mais comuns nesse contexto, destacam-se:

- ▶ Support Vector Machines;

# Técnicas envolvidas

## *Machine Learning*

### *Machine Learning*

Área voltada à construção de modelos computacionais capazes de aprimorar seu desempenho na resolução de problemas a partir da experiência [Mit97].

Entre as técnicas mais comuns nesse contexto, destacam-se:

- ▶ Support Vector Machines;
- ▶ Random Forests;

# Técnicas envolvidas

## *Machine Learning*

### *Machine Learning*

Área voltada à construção de modelos computacionais capazes de aprimorar seu desempenho na resolução de problemas a partir da experiência [Mit97].

Entre as técnicas mais comuns nesse contexto, destacam-se:

- ▶ Support Vector Machines;
- ▶ Random Forests;
- ▶ Multilayer Perceptron;

# Técnicas envolvidas

## *Machine Learning*

### *Machine Learning*

Área voltada à construção de modelos computacionais capazes de aprimorar seu desempenho na resolução de problemas a partir da experiência [Mit97].

Entre as técnicas mais comuns nesse contexto, destacam-se:

- ▶ Support Vector Machines;
- ▶ Random Forests;
- ▶ Multilayer Perceptron;
- ▶ Convolutional Neural Networks (Deep Learning).

# Abordagem tradicional

Baseada em extração de descritores

A abordagem tradicional para a tarefa de reconhecimento combinava as etapas:

1. aquisição do conjunto de imagens;
2. divisão entre conjuntos de treino e testes;
3. **extração de descritores;**
4. treinamento de um modelo de aprendizagem.
5. avaliação do modelo.

As imagens, nessa abordagem, perdem seu formato original de matrizes de pixels e se tornam **vetores de características** produzidos com o objetivo de **descrever os objetos** de interesse.

# Abordagem com Deep Learning

## Convolutional Neural Networks

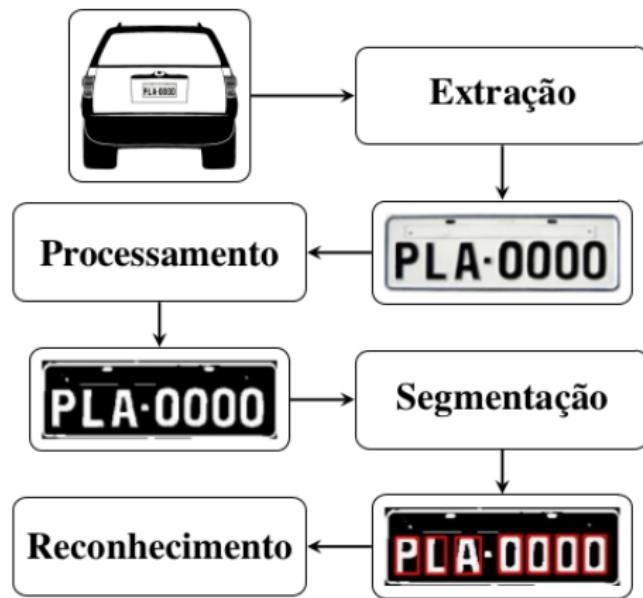
Com as Deep Neural Networks, mais especificamente, as Redes Neurais Convolucionais, o processo demanda menos pré-processamento: a rede neural recebe **toda a matriz de pixels** e, em suas camadas, abstrai as características dos objetos, **produzindo elas mesmas os descritores.**

Com isso, o *framework* geral nessa abordagem se resume a:

1. aquisição das imagens;
2. divisão entre conjuntos de treino e testes;
3. treinamento de um modelo de aprendizagem.
4. avaliação do modelo.

# Aplicações

## Reconhecimento Automático de Placas



(Fonte: Autoral)

# Aplicações

## Reconhecimento de Faces (Expressões)



(Fonte: Mobile Vision - <https://developers.google.com/vision/>)

# Aplicações

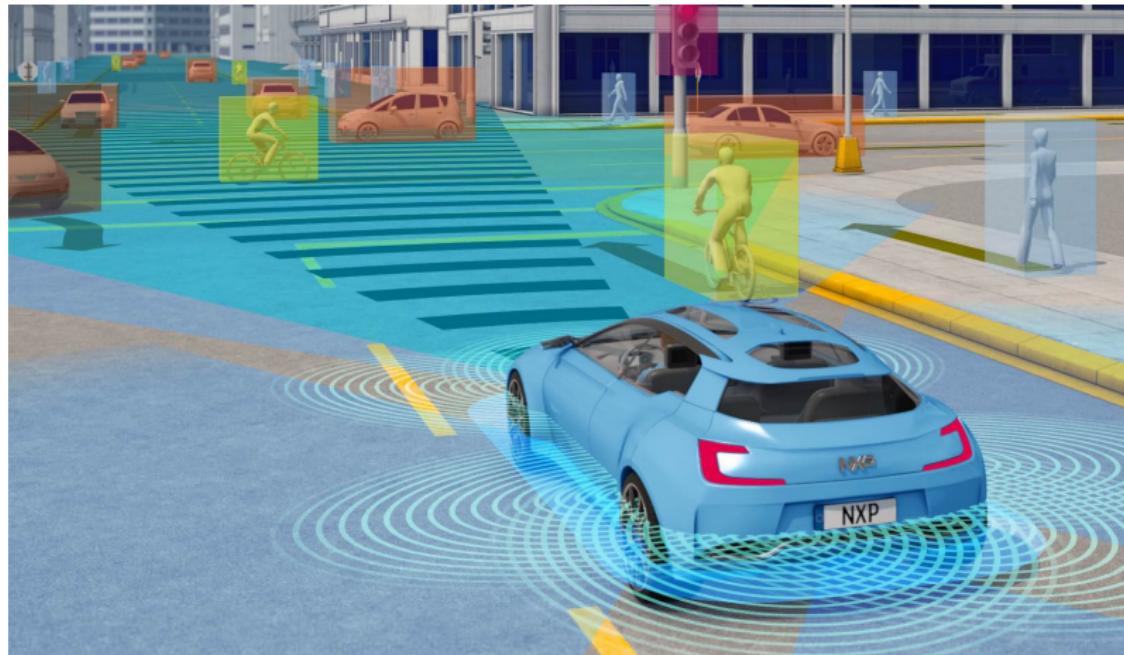
## Reconhecimento de gestos



(Fonte: Wikipedia - [https://en.wikipedia.org/  
wiki/Gesture\\_recognition](https://en.wikipedia.org/wiki/Gesture_recognition))

# Aplicações

## Veículos Autônomos



(Fonte: Electronics Weekly -

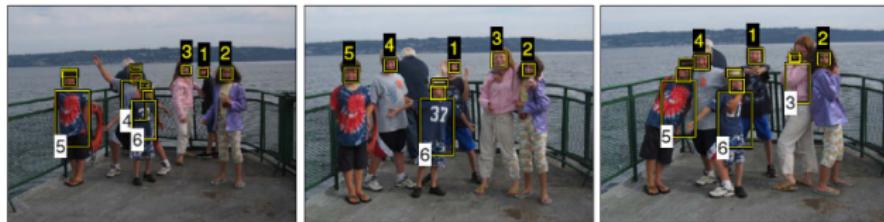
<https://www.electronicsweekly.com/market-sectors/automotive-electronics/ces-autonomous-cars-sensors-make-safe-2017-01/>)

# Aplicações

## Coleções de Fotos Pessoais



(a)

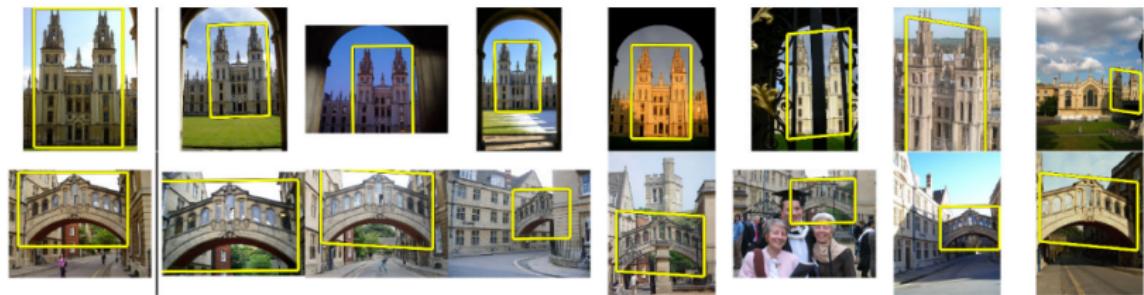


(b)

(Fonte: Computer Vision - Algorithms and Applications - Richard Szeliski)

# Aplicações

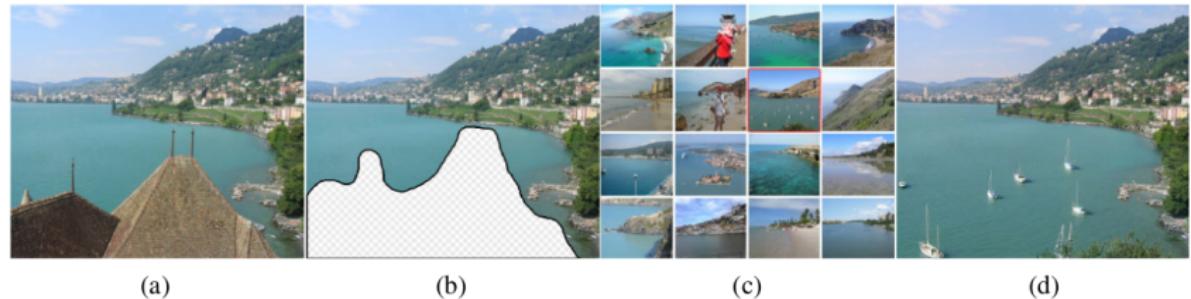
## Reconhecimento de Locais



(Fonte: Computer Vision - Algorithms and Applications - Richard Szeliski)

# Aplicações

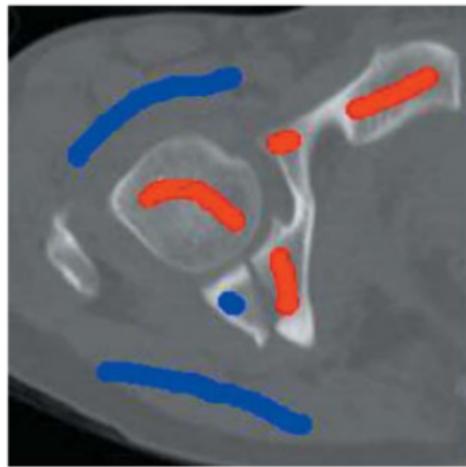
## Edições Inteligentes



(Fonte: Computer Vision - Algorithms and Applications - Richard Szeliski)

# Aplicações

## Segmentação de Imagens Médicas



(a)

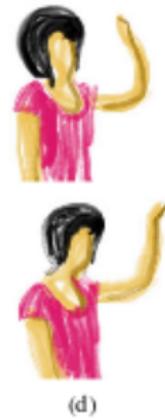
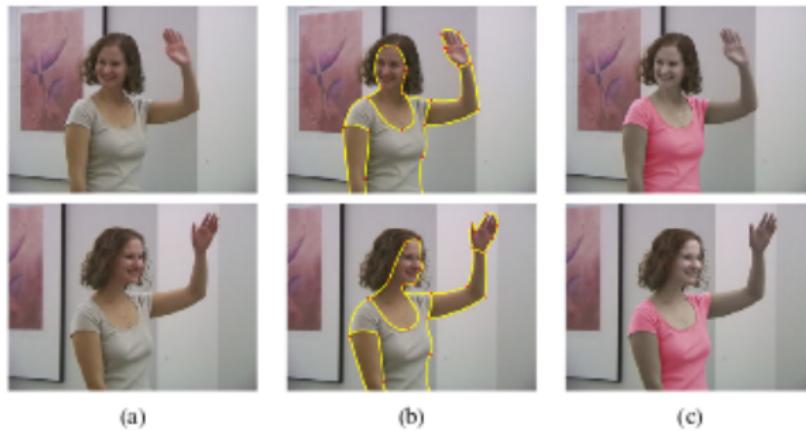


(b)

(Fonte: Computer Vision - Algorithms and Applications - Richard Szeliski)

# Aplicações

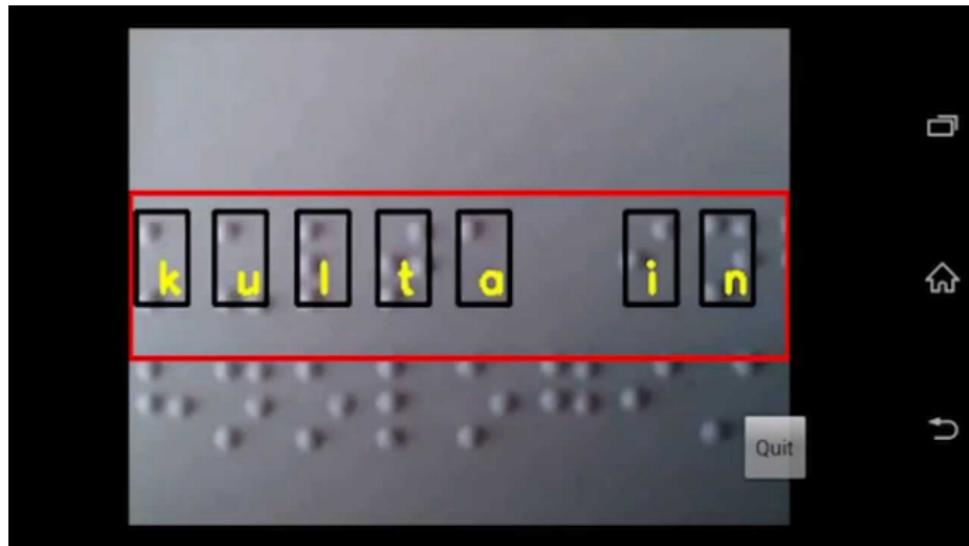
## Rastreamento de Contorno e Rotoscopia



(Fonte: Computer Vision - Algorithms and Applications - Richard Szeliski)

# Aplicações

## Reconhecimento Ótico de Braille



(Fonte: Youtube -  
<https://www.youtube.com/watch?v=X5kFdUsCkEA>)

# Aplicações

## Reconhecimento de Impressões Digitais



FINGERPRINT RECOGNITION

(Fonte: Dreamstime -

<https://thumbs.dreamstime.com/b/fingerprint-recognition-smartphones-smartphone-user-touching-screen-to-unlock-68633965.jpg>)

# Aplicações

## Reconhecimento de Iris



(Fonte: Iritech - <http://www.iritech.com/blog/iris-biometric-safe/>)

# Referências

-  Tom M. Mitchell, *Machine learning*, 1 ed., McGraw-Hill series in computer science, McGraw-Hill, 1997.
-  Adrian Rosebrock, *Deep learning for computer vision with python*, Online, 2017.
-  Richard Szeliski, *Computer vision: Algorithms and applications*, Texts in Computer Science, Springer, 2011.