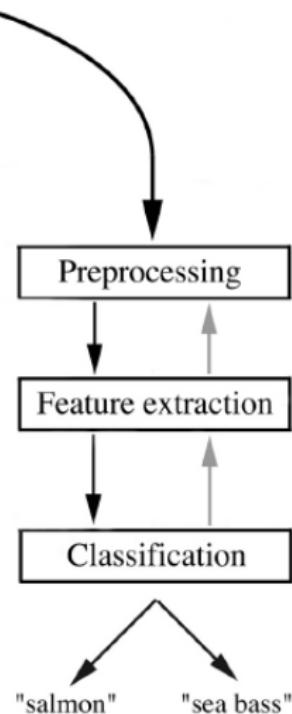


# Reconhecimento de Padrões em Imagens Características

Dainf - UTFPR

Profa. Leyza Baldo Dorini

# Motivação



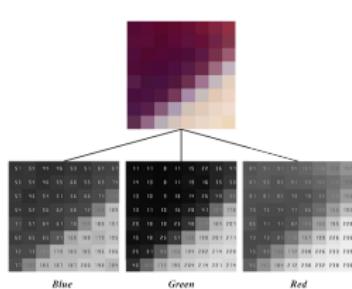
## Características: *detection* × *description*

- **Detecção (extração)**: dada uma imagem ou região de interesse, busca-se identificar características relevantes, que permitam a um classificador determinar a qual classe ela pertence.
  - Tipicamente envolve uma etapa de pré-processamento, visando deixar a imagem de entrada mais adequada.
  - Além de agrupar objetos de uma mesma classe, as características devem permitir a separação inter-classes.
- **Descrição**: é um aspecto fundamental (forma com que as características serão representadas computacionalmente).
  - Não necessariamente uma maior quantidade de características conduz a um melhor desempenho.
  - Dependendo do problema e do algoritmo de reconhecimento, algumas representações são mais adequadas.

# Características

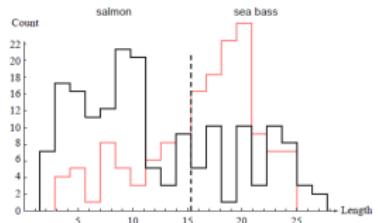
Quais características utilizar para representar uma imagem?

- Intensidade ou cor (histogramas).
- Coleção de componentes que representam regiões de interesse, tal como arestas e blobs.
- Pontos característicos.

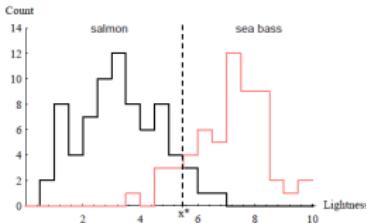


# Características

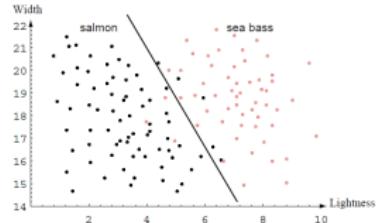
- Segmentar o objeto de interesse e determinar brilho, forma, tamanho...



(a) Tamanho



(b) Brilho



(c) 2D feature space

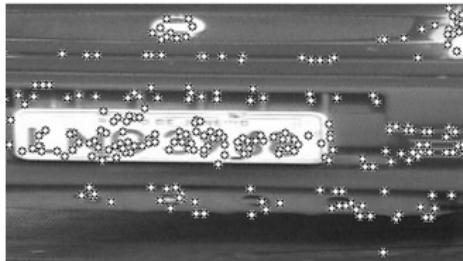
## Importante

Em todos os casos, é preciso considerar a presença de ruídos, variações intra-classe (inerentes à própria população ou ao processo de aquisição) e transformações que não são relevantes (rotação, escala, etc).

# Características: categorias

De forma simplificada:

- 1 Baixo (cor, textura, forma) e alto nível (velocidade de um carro - relação com informação semântica).



(a) *Low level*



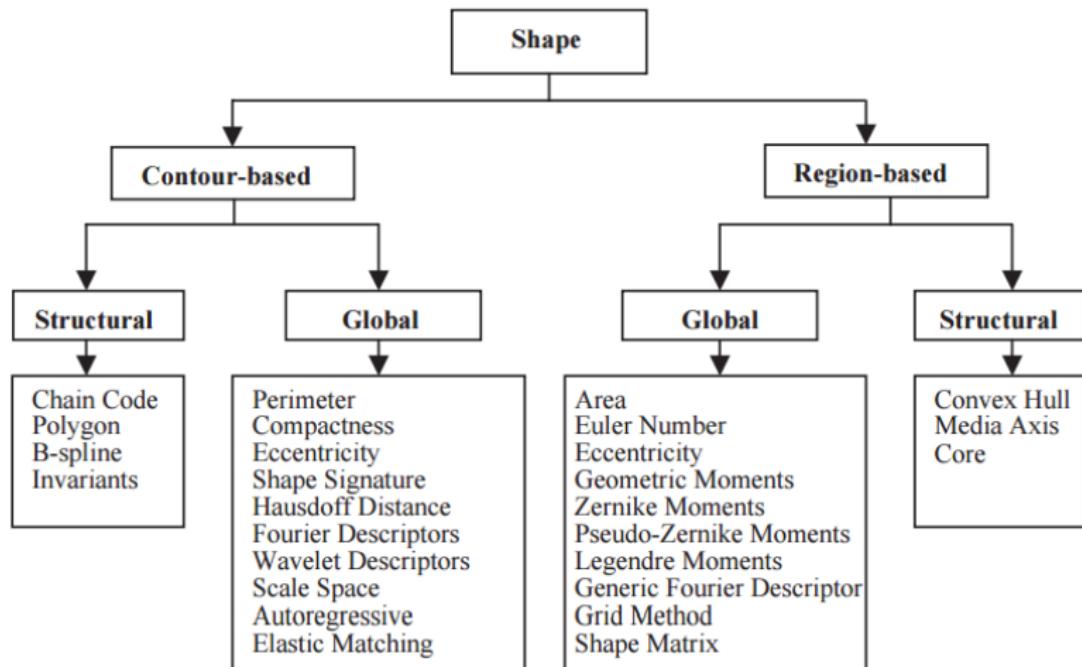
43.5 km/h (real 44.6 km/h)

(b) *High level*

- 2 Globais, locais e regionais.

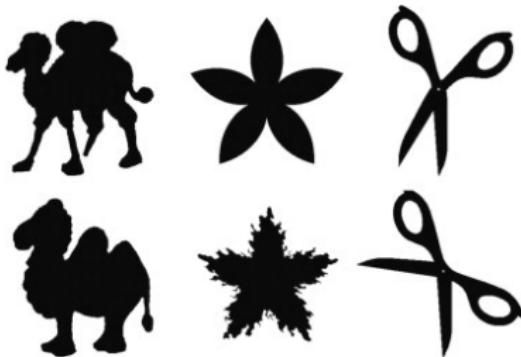
# Características: forma

Para características relacionadas à forma, a seguinte classificação pode ser considerada.



## Características: forma

Quais características considerar: área, perímetro, fecho convexo, excentricidade, etc? Podem ser muito simples.



Exemplos de aplicação: caracteres, *trademarks* e contornos pré-segmentados.



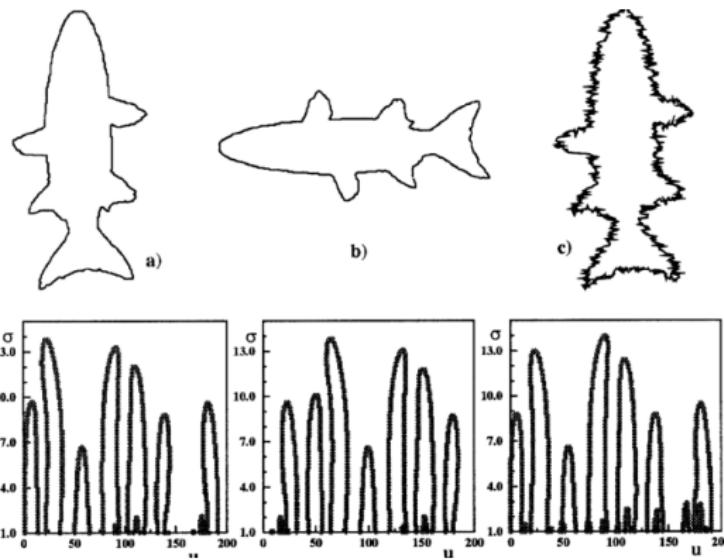
(a) MNIST



(b) MPEG7

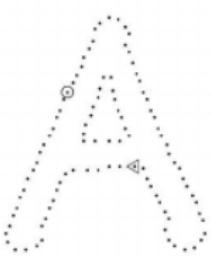
## Características: forma (*Curvature Scale-Space*)

“The CSS image consists of several arch-shape contours representing the inflection points of the shape as it is smoothed. The maxima of these contours are used to represent a shape. The CSS representation is robust with respect to scale, noise and change in orientation”.



Demo: <http://www.ee.surrey.ac.uk/CVSSP/demos/css/demo.html>

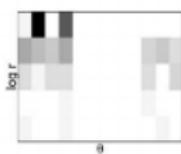
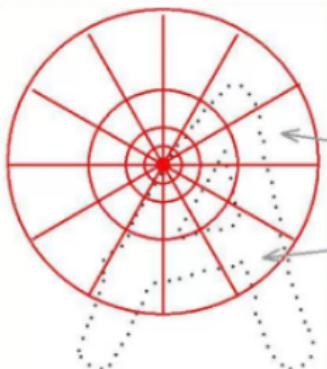
# Características: forma (*Shape Context*)



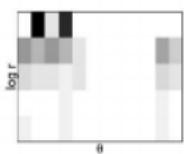
(a)



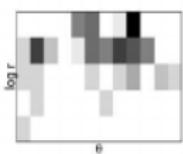
(b)



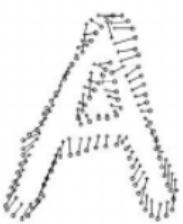
1  
(d)



(e)



(f)



(g)

1

<sup>1</sup>Belongie et al. Shape Matching and Object Recognition Using Shape Contexts. PAMI 2002.

## Características estatísticas: Momentos Invariantes

Também chamadas momentos estatísticos, se caracterizam por valores numéricos calculados a partir da imagem previamente segmentada e que descrevem a distribuição espacial dos pontos contidos na imagem ou em uma região.

- Característica global e invariante.
- Medidas puramente estatísticas da distribuição dos pontos.

## Características estatísticas: Momentos de Hu

Considere a imagem binária de um objeto  $M \times N$  onde  $I(x, y)$  representa o estado do pixel  $(x, y)$  - preto ou branco. Um momento regular de ordem  $(p + q)$  é definido por:

$$m_{pq} = \sum_{x=1}^N \sum_{y=1}^M x^p y^q I(x, y), \text{ para } p, q = 0, 1, 2$$

Por exemplo, o momento de ordem 0 ( $m_{00}$ ) representa a superfície, enquanto os momentos de ordem 1 ( $m_{01}$  e  $m_{10}$ ) definem o centro de gravidade ( $x_g$  e  $y_g$ ) da imagem.

$$x_g = \frac{m_{10}}{m_{00}}, \quad y_g = \frac{m_{01}}{m_{00}}$$

## Características estatísticas: Momentos de Hu

Com o intuito de serem invariantes a rotação e translação, Hu definiu os momentos centrais  $n_{pq}$ :

$$n_{pq} = \sum_{x=1}^N \sum_{y=1}^M (x - x_g)^p (y - y_g)^q I(x, y), \text{ para } p, q = 0, 1, 2$$

Os momentos centrais de ordem 2, por exemplo permitem achar os eixos principais de inércia, os prolongamentos e as orientações da forma.

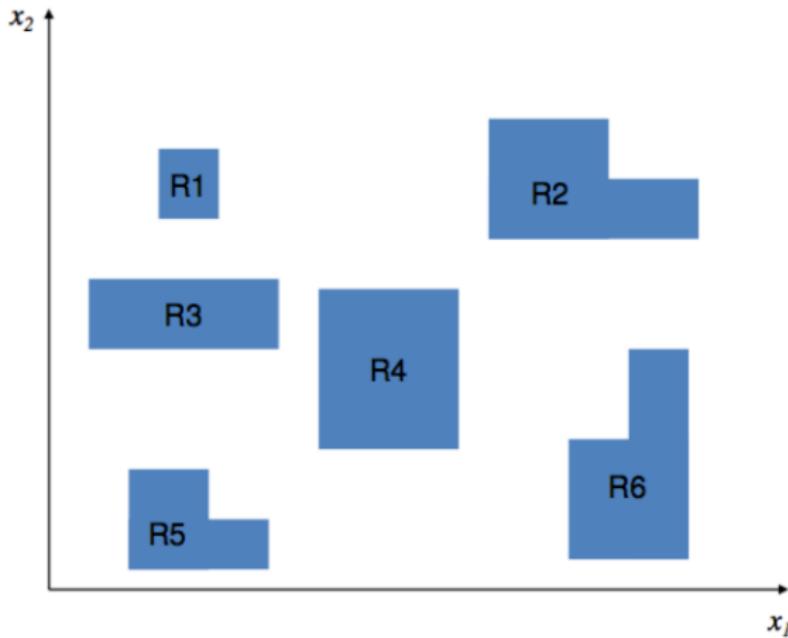
## Características estatísticas: Momentos de Hu

Os momentos mais utilizados são os 7 momentos invariantes de HU, (de ordem 2 e 3):

- $\Phi(1) = (\mu_{20} + \mu_{02})$
- $\Phi(2) = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2$
- $\Phi(3) = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2$
- $\Phi(4) = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2$
- $\Phi(5) = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$
- $\Phi(6) = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03})$
- $\Phi(7) = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$

## Momentos de Hu: exemplo

Considere a seguinte imagem de entrada.



## Momentos de Hu: exemplo

Considere a seguinte imagem de entrada.

Momento	R1	R2	R3	R4	R5	R6
1	1.67E-01	1.94E-01	2.08E-01	1.67E-01	1.94E-01	1.94E-01
2	0.00E+00	6.53E-03	1.56E-02	0.00E+00	6.53E-03	6.53E-03
3	0.00E+00	1.02E-03	0.00E+00	0.00E+00	1.02E-03	1.02E-03
4	0.00E+00	4.56E+05	0.00E+00	0.00E+00	4.56E+05	4.56E+05
5	0.00E+00	4.25E-09	0.00E+00	0.00E+00	4.25E-09	4.25E-09
6	0.00E+00	1.70E+06	0.00E+00	0.00E+00	1.70E+06	1.70E+06
7	0.00E+00	-8.85E+09	0.00E+00	0.00E+00	-8.85E+09	-8.85E+09

R1 e R4, R2 e R5 são diferentes escalas do mesmo objeto R6 é a versão rotacionada de R2 e R5.

## Características: cor

A cor é uma característica amplamente considerada em aplicações de reconhecimento e recuperação de padrões.

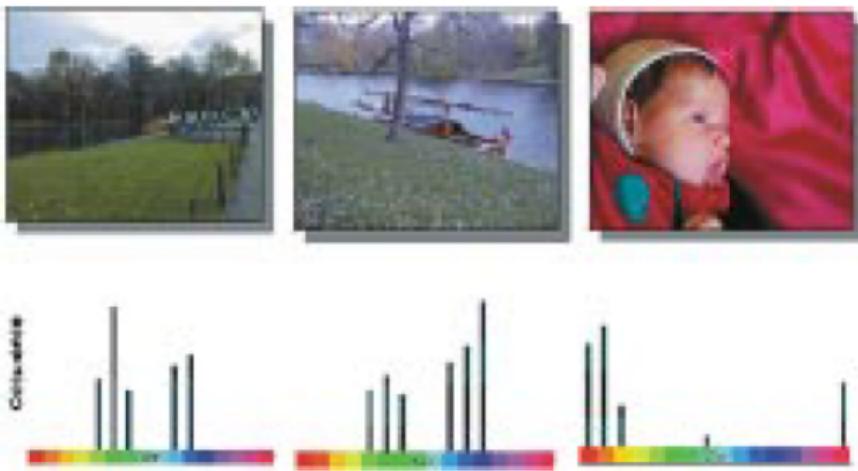


Fig. 2. Three color images and their MPEG-7 histogram color distribution, depicted using a simplified color histogram. Based on the color distribution, the two left images would be recognized as more similar compared to the one on the right.

## Características: cor (estudo comparativo)

São diversos os descritores baseados em cor. Exemplos<sup>2</sup>

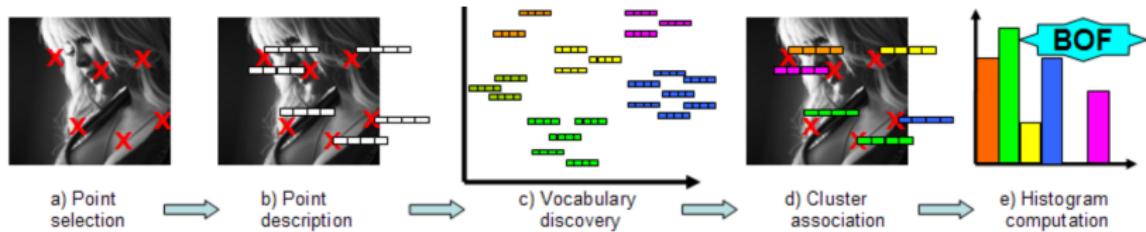
- Histogramas: diferentes abordagens (RGB, Hue, Opponent Color).
- Momentos de cor: informações fotométricas e espaciais.
- Color SIFT: combina descritor do SIFT com espaços de cor utilizados em outros descritores de cor.

No artigo, os descritores são calculados a partir de uma vizinhança de pontos selecionados com o detector de Harris. Um dicionário visual é construído utilizando k-means e SVM é utilizado para classificação. Conclusão: os descritores baseados em SIFT apresentam desempenho superior aos histogramas.

---

<sup>2</sup>van de Sande, et al. Evaluating Color Descriptors for Object and Scene Recognition. PAMI 2010.

# Características: cor (*Nude Detection*)



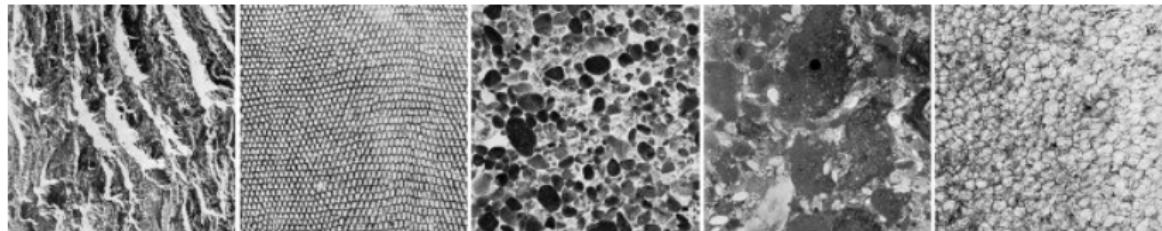
3

---

<sup>3</sup>A BAG-OF-FEATURES APPROACH BASED ON HUE-SIFT  
DESCRIPTOR FOR NUDE DETECTION Ana P. B. Lopes, Sandra E. F. de  
Avila, Anderson N. A. Peixoto, Rodrigo S. Oliveira and Arnaldo de A. Araújo

## Características: textura

*"Texture refers to the visual patterns that have properties of homogeneity or not, that result from the presence of multiple colors or intensities in the image. It is a property of virtually any surface, including clouds, trees, bricks, hair, and fabric. It contains important structural information of surfaces and their relationship to the surrounding environment."*



Padrão visual que possui algumas propriedades de homogeneidade que não resultam simplesmente de uma cor ou intensidade.

## Descritores de textura

Buscam detectar padrões de homogeneidade que não estão presentes em uma simples cor ou intensidade. Comumente utilizados em problemas de classificação e também na recuperação de imagens. Por exemplo, imagens médicas e imagens de satélite.  
Algumas abordagens:

- Momentos do Histograma
- *Grey-Level Co-Ocurrence Matrix (GLCM)*
- Fractais
- Filtros de Gabor
- *Local Binary Patterns (LBP)*

## Características: textura (*Grey-Level Co-Ocurrence Matrix*)

GLCM é uma abordagem que visa caracterizar padrões (texturas) utilizando um conjunto de medidas estatísticas extraídas a partir de uma matriz que descreve as ocorrências de pares de pixels com intensidades e em uma relação espacial específicas<sup>4</sup>.

---

<sup>4</sup>Sugestão: <http://www.fp.ucalgary.ca/mhallbey/tutorial.htm>

## Características: textura (*Grey-Level Co-Ocurrence Matrix*)

Considere por exemplo uma imagem com 3 níveis de cinza,  $z_0 = 0, z_1 = 1$  e  $z_2 = 2$ . Suponha que a relação espacial,  $P$ , seja dada por: um pixel à direita e um pixel abaixo.

0	0	0	1	2
1	1	0	1	1
2	2	1	0	0
1	1	0	2	0
0	0	1	0	1

- Sendo assim, a matriz de coocorrência,  $A$ , teria dimensão  $3 \times 3$ .
- $a_{00}$  é o número de vezes que um ponto com nível de cinza 0 aparece abaixo e a direita de outro píxel com o nível 0. No exemplo, é igual a 4.

No exemplo:

$$A = \begin{pmatrix} 4 & 2 & 1 \\ 2 & 3 & 2 \\ 0 & 2 & 0 \end{pmatrix}$$

## Características: textura (*Grey-Level Co-Ocurrence Matrix*)

Característica	Descrição
Probabilidade Máxima	Indica a direção mais importante da textura a ser examinada.
Entropia	Mede a informação contida em $p$ ; muitos valores nulos representam pouca informação.
Momento de diferenças de ordem $k$	Distorção da imagem. Apresenta valores pequenos se $p$ tiver valores maiores na diagonal principal.
Momento inverso de diferenças de ordem $k$	Inverso do contraste. Apresenta valores maiores se $p$ tiver valores pequenos na diagonal principal.

Figura : Exemplos de descritores de Haralick (1979). No total, são 14.

## Características: textura (LBP)

*Local Binary Pattern* (LBP): descrito inicialmente em 1994, é um descriptor que apresenta bons resultados para classificação de texturas. Além de simples, tem uma propriedade muito interessante: é robusto a alterações de intensidade causadas, por exemplo, por variações de iluminação.

example	thresholded	weights
6 5 2 7 6 1 9 8 7	1 0 0 1 0 0 1 1 1	1 2 4 128 8 64 32 16

$$\text{Pattern} = \mathbf{11110001}$$

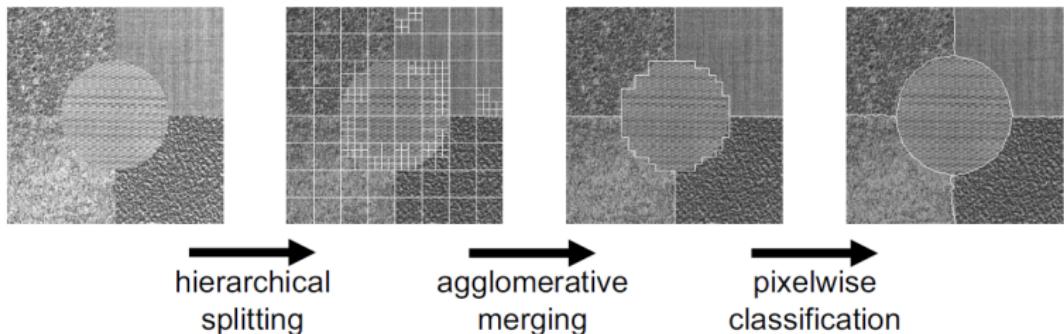
$$\mathbf{LBP} = 1 + 16 + 32 + 64 + 128 = \mathbf{241}$$

$$\mathbf{C} = (6+7+8+9+7)/5 - (5+2+1)/3 = \mathbf{4.7}$$

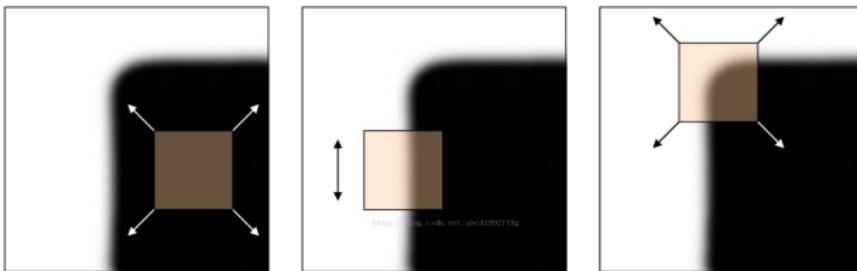
# Características: textura (LBP)

Outros exemplos:

- Ahone, et al. Face Description with Local Binary Patterns: Application to Face Recognition. PAMI, 2006
- Hanusiak et al. Writer verification using texture-based features. IJDAR, 2012.
- Costa et al. Music genre classification using LBP textural features. Signal Processing, 2012.
- Ojala e Pietikainen. Unsupervised texture segmentation using feature distributions. PR, 1999.



## Características: *edges* e *corners* (*interest points*)



- Uma aresta (*edge*), no caso ideal, consiste em um conjunto de curvas conectadas que representam as fronteiras dos objetos. Não ocorre variação ao longo da direção principal.
- Um canto (*corner*) pode ser definido como uma intersecção de arestas (*edges*) que possui ao menos duas diferentes direções dominantes em uma determinada vizinhança.
  - Nem todos os cantos são pontos de interesse (*interest points*): é preciso também possuir uma posição bem definida e ser detectado de forma robusta. Em reconhecimento de padrões, esse é tipicamente o caso.

## Edge detection (Sobel)

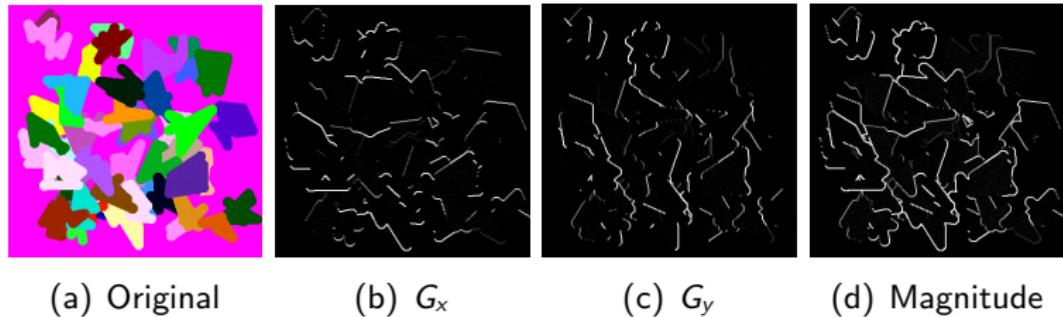
Operador que visa aproximar o gradiente de uma imagem em níveis de cinza com base na sua convolução com os seguintes filtros:

$$G_x = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \quad \text{e} \quad G_y = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}.$$

A magnitude e a direção do gradiente são dados, respectivamente, por:

$$G = \sqrt{G_x^2 + G_y^2} \quad \text{e} \quad \theta = \text{atan} \left( \frac{G_y}{G_x} \right)$$

## Edge detection (Sobel)

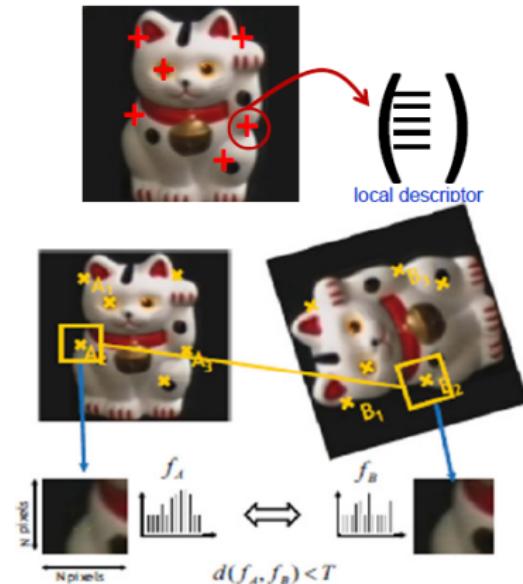


O gradiente é uma etapa essencial em diversos descritores. Sua utilização visa reduzir a quantidade de dados a ser processada, identificando regiões mais relevantes ao mesmo passo que preserva propriedades estruturais. Não é uma tarefa trivial (imagens coloridas, fragmentação, ruídos...).

# Corner (interest point) detection

De forma simplificada:

- Determinar pontos de interesse
- Selecionar uma região de interesse na vizinhança de cada ponto
- Calcular um descriptor com base em tais regiões (etapa de normalização pode ser necessária)



## Corner (*interest point*) detection (Harris)

Suponha que um determinada região (janela) da imagem sofre um deslocamento  $\Delta\mathbf{x}$ . A SSE (*Sum of Squared Differences*) em  $\mathbf{x}$  é dada por:

$$E(\mathbf{x}) = \sum_{\mathbf{x}_i \in W} [I(\mathbf{x}_i) - I(\mathbf{x}_i + \Delta\mathbf{x})]^2 \quad (1)$$

$$= \dots \quad (2)$$

$$= (\Delta\mathbf{x})^T A(\mathbf{x}) \Delta(\mathbf{x}) \quad (3)$$

em que a matriz de auto-correlação  $A$  é dada por:

$$A = \begin{pmatrix} \sum_W I_x^2 & \sum_W I_x I_y \\ \sum_W I_x I_y & \sum_W I_y^2 \end{pmatrix}$$

## *Corner (interest point) detection* (Harris)

Essa matriz captura o padrão de intensidade em  $W$ . A resposta do detector é dada por:

$$R(\mathbf{x}) = \det A - \alpha(\text{tr}A)^2$$

Os cantos podem ser determinados de diferentes formas:

- Alta resposta: localizações em que  $R(\mathbf{x})$  é maior que um determinado limiar.
- Máximo local: localizações em que  $R(\mathbf{x})$  é maior que os valores da vizinhança.

## *Corner (interest point) detection (Harris)*

Exemplo considerando a primeira regra. Observe que existe muita sobreposição (segundo critério pode ser interessante).



## Corner (interest point) detection (Shi e Tomasi)

Consideram um peso,  $w(\mathbf{x}_i)$ , na função de auto-correlação:

$$E(\mathbf{x}) = \sum_{\mathbf{x}_i \in W} w(\mathbf{x}_i) [I(\mathbf{x}_i) - I(\mathbf{x}_i + \Delta\mathbf{x})]^2$$

A matriz  $A$  é dada por:

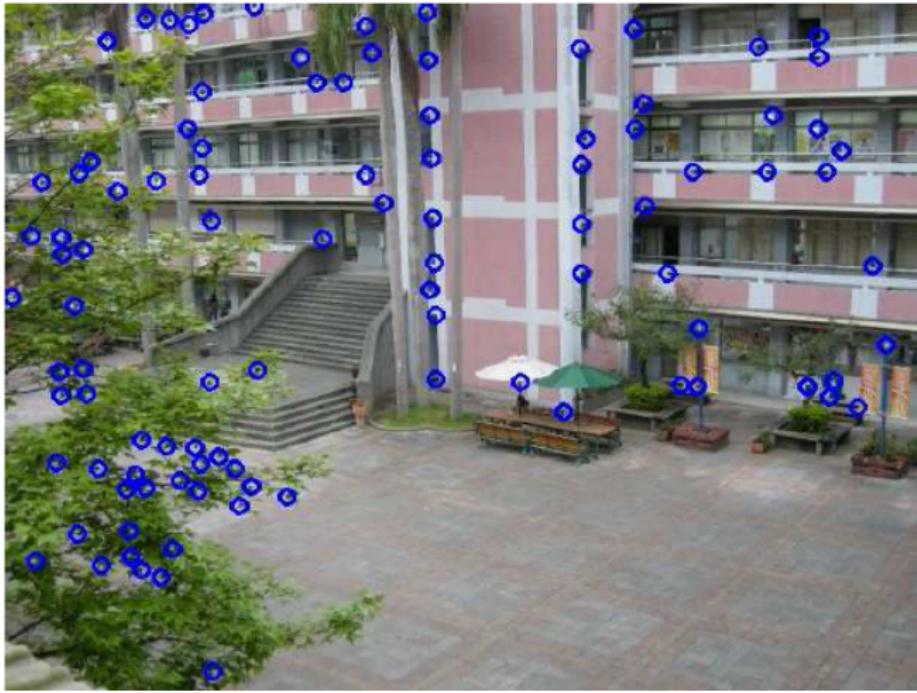
$$A = \begin{pmatrix} \sum_W wI_x^2 & \sum_W wI_x I_y \\ \sum_W wI_x I_y & \sum_W wI_y^2 \end{pmatrix}$$

Existem escalares  $\lambda_1, \lambda_2$  e vetores  $\mathbf{v}_1, \mathbf{v}_2$  tais que

$$A \mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, 2$$

Autovalores e autovetores!

## *Corner (interest point) detection (Shi e Tomasi)*



## *Corner (interest point) detection* (Shi e Tomasi)

Para escolher as “*Good Features to Track*”<sup>5</sup>, considera-se:

- Se os dois autovalores são pequenos, não há variação significativa na região (ou seja, ela é homogênea e o ponto não é bom).
- Se o maior autovalor é tal que  $\lambda_1 \gg \lambda_2$ , a variação mais significativa se dá apenas na direção de  $v_1$  (ou seja, é uma aresta).
- Se os dois autovalores são grandes, o ponto é selecionado. Como a imagem tem um valor máximo (255), existe um limite superior para os autovalores - ou seja, basta determinar se  $\min(\lambda_1, \lambda_2)$  é suficientemente grande.

---

<sup>5</sup>Shi e Tomasi. Good Features to Track. CVPR 1994.

# *Scale Invariant Feature Transform (SIFT)*

O método<sup>6</sup> pode ser dividido em duas etapas principais: detecção e descrição. Basicamente:

- Deteção de extremos no espaço-escala.
- Localização de pontos chave *keypoints*.
- Atribuição da orientação dos descritores.
- Construção do descriptor local.

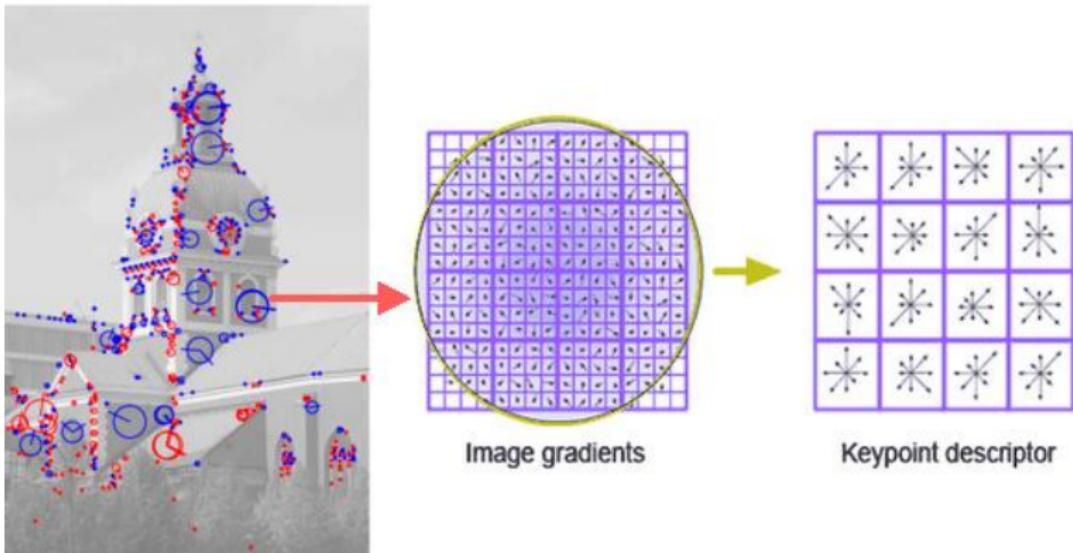
Diversas vantagens: invariância à escala, *distinctiveness* (grandes bases de dados), eficiência computacional, etc.

---

<sup>6</sup>Principais referências:

- Lowe, D. Object recognition from local scale-invariant features. ICCV, 1999.
- Lowe, D. Distinctive Image Features from Scale-Invariant Keypoints. IJCV, 2004.

# Scale Invariant Feature Transform (SIFT)



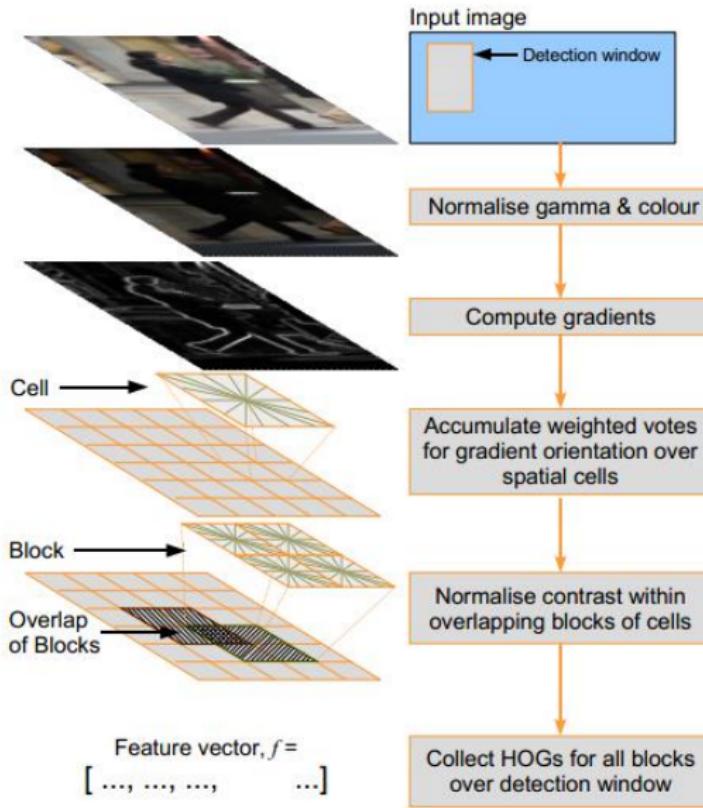
# SIFT - exemplos de aplicação

Nilsback and Zisserman. Automated flower classification over a large number of classes. Indian Conference on Computer Vision, 2006.

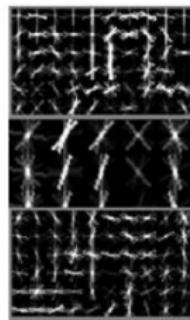
*"The low level features we use are colour, histogram of gradient orientations (HOG), and SIFT sampled on both the foreground region and its boundary".*



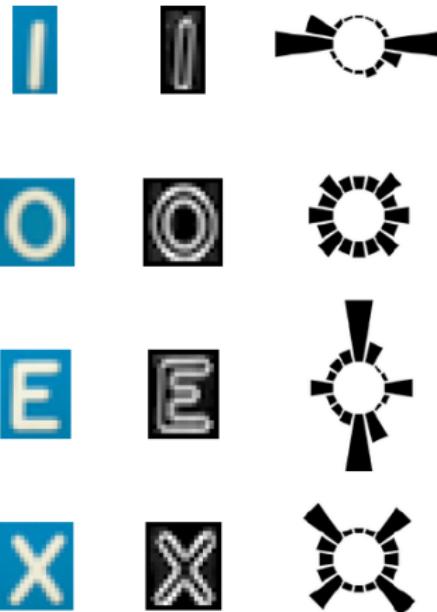
# Histogram of Oriented Gradients (HOG)



# HOG - exemplos de aplicação



(a) Pedestres



(b) Letras

7 8

<sup>7</sup>(a) Dalal e Triggs. HOG for human detection. CVPR 2005.

<sup>8</sup>(b) Minetto et al. T-HOG: an Effective Gradient-Based Descriptor for Single Line Text Regions. PR 2013.

# HOG - exemplos de aplicação

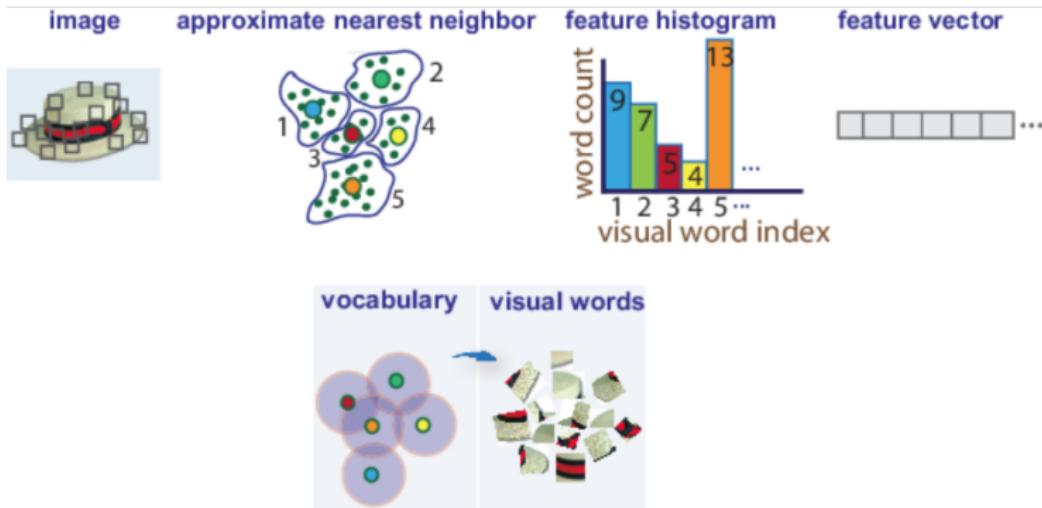


9

---

<sup>9</sup>Hu e Collomosse. A Performance Evaluation of Gradient Field HOG Descriptor for Sketch Based Image Retrieval. CVIU 2013.

# Bag of features



Inicialmente, são extraídas características da imagem, a partir das quais é construído um histograma de características (usando kNN). Uma determinada posição do histograma é incrementada com base na proximidade do descriptor ao centro de um cluster. O tamanho do histograma é o tamanho do vocabulário.