

Image Descriptors: local features

Image Processing — scc0251

`www.icmc.usp.br/~moacir` — `moacir@icmc.usp.br`

ICMC/USP — São Carlos, SP, Brazil

2011

Agenda

- 1 Introdução
- 2 Características Locais
- 3 Detectores de pontos-chave
 - Hessian
 - Harris
- 4 SIFT

Agenda

- 1 Introdução
- 2 Características Locais
- 3 Detectores de pontos-chave
 - Hessian
 - Harris
- 4 SIFT

Descritores visuais

Problema — computar de forma eficiente valores que possam descrever uma imagem (ou parte de uma imagem)

- **Discriminativos**
- **Robusto com relação à:** transformações na imagem e nos objetos, ponto de visão e oclusão.

É importante que os descritores sejam comparáveis, por alguma métrica de similaridade ou dissimilaridade.

Descritores visuais

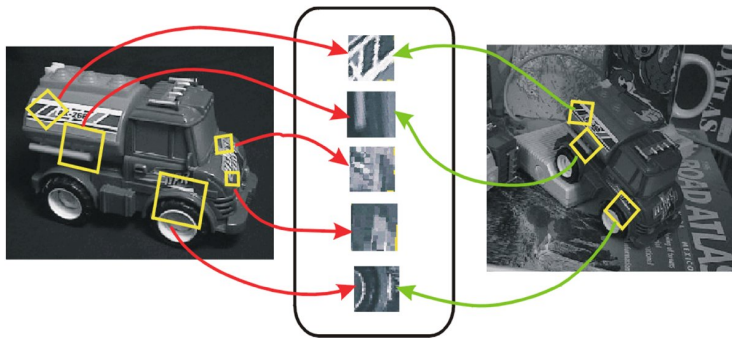
- Cor
- Textura
- **Características Locais**
- Dicionários/palavras visuais

Agenda

- 1 Introdução
- 2 Características Locais
- 3 Detectores de pontos-chave
 - Hessian
 - Harris
- 4 SIFT

Características Locais

O conteúdo da imagem é transformado em coordenadas de características locais que são invariantes (ou covariantes) a rotação, escala, entre outros parâmetros.

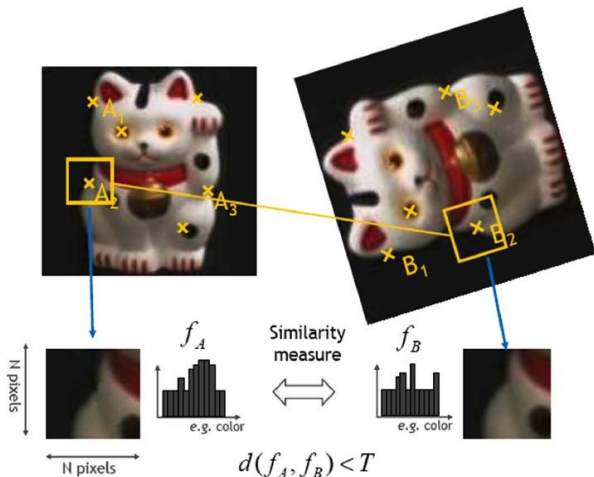


Características Locais: componentes

- Pontos de interesse:
 - especificam pontos repetíveis
 - posição (x,y) e escala
 - Exemplos: cantos, regiões (*blobs*)
- Descritores locais (pontos-chave)
 - definem a representação da característica em torno de um ponto de interesse
 - Exemplos: valores de intensidade ou um histograma de gradientes na vizinhança de um ponto de interesse

Características Locais: abordagem

- 1 Encontra um conjunto de pontos-chave distintos
- 2 Define uma região ao redor de cada ponto-chave
- 3 Extrai e normaliza o conteúdo da região
- 4 Computa um descritor local da região normalizada
- 5 Casa os descritores locais



Detectores de pontos-chave

- Hessian (Beaudet, 1978) (Harris and Stephens, 1988)
- Laplacian, Difference of Gaussians (Lindeberg, 1998) (Lowe, 1999)
- Harris/Hessian – Laplace (Mikolajczyk and Schmid, 2002)
- Harris/Hessian – Affine (Mikolajczyk and Schmid, 2004)
- MSER (Matas et al, 2002)
- EBR / IBR (Tuytelaars and van Gool, 2004)

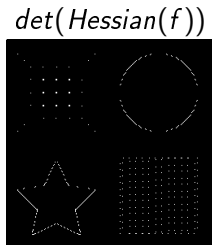
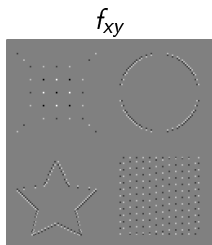
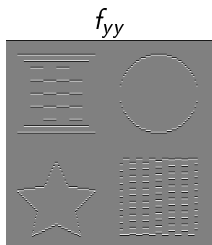
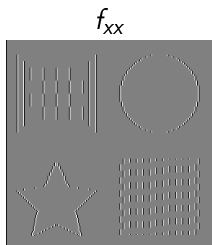
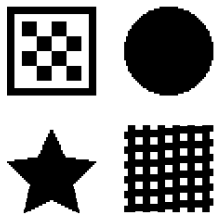
Agenda

- 1 Introdução
- 2 Características Locais
- 3 Detectores de pontos-chave**
 - Hessian
 - Harris
- 4 SIFT

Agenda

- 1 Introdução
- 2 Características Locais
- 3 Detectores de pontos-chave
 - Hessian
 - Harris
- 4 SIFT

Detectores de pontos-chave: Hessian



$$\text{Hessian}(f) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix}$$

$$\det(\text{Hessian}(f)) = f_{xx} \cdot f_{yy} - f_{xy}^2$$

Detectores de pontos-chave: Hessian

Responde nas principais curvaturas: cantos e também áreas com textura forte



Agenda

- 1 Introdução
- 2 Características Locais
- 3 Detectores de pontos-chave
 - Hessian
 - Harris
- 4 SIFT

Detectores de pontos-chave: Harris

Matriz do segundo momento – relacionada à autocorrelação.

$$A(\mathbf{x}) = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2(\mathbf{x}) & I_x I_y(\mathbf{x}) \\ I_x I_y(\mathbf{x}) & I_y^2(\mathbf{x}) \end{bmatrix}$$

onde I_x e I_y são as respectivas derivativas nas direções x e y no ponto \mathbf{x} .

- Matriz A descreve a forma da medida de autocorrelação por deslocamentos locais.
- Descreve como essa medida muda localmente e assim tenta extrair os principais pontos de curvatura.
- Função de peso $w(x,y)$ (tipicamente Gaussiana isotrópica e circular).

$$w(x,y) = g(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{\left(-\frac{x^2+y^2}{2\sigma^2}\right)}$$

Detectores de pontos-chave: Harris

Outra interpretação:

$$A(\mathbf{x}) = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2(\mathbf{x}) & I_x I_y(\mathbf{x}) \\ I_x I_y(\mathbf{x}) & I_y^2(\mathbf{x}) \end{bmatrix}$$

a matrix $A(\mathbf{x})$ é de mesmo tamanho da imagem. Para cada pixel da imagem A possui 4 valores representados pelos elementos da matriz.

- A partir da análise dos autovalores de A em cada pixel, é possível interpretar cantos, regiões planas e bordas.

Detectores de pontos-chave: Harris

Matriz do segundo momento – relacionada à autocorrelação.

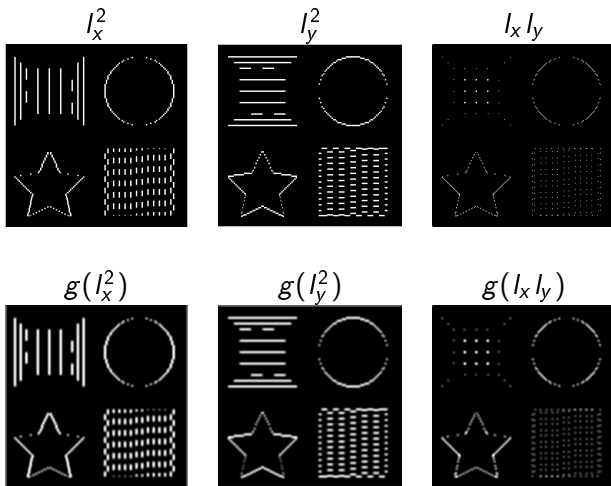
$$Harris(I) = \det(A(\mathbf{x})) - \alpha(\text{tr}(A(\mathbf{x})))$$

$$Harris(I) = [g(I_x^2)g(I_y^2) - g(I_x I_y)^2] - \alpha(g(I_x^2) + g(I_y^2))^2$$

onde $g(I)$ é um filtro gaussiano.

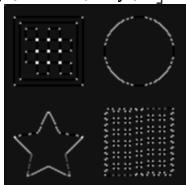
- Após computar H é preciso encontrar os máximos locais (pontos maiores que seus vizinhos em janelas 3×3 e 5×5).
- Uma dica importante é remover a resposta nos cantos da imagem, para facilitar o processamento

Detectores de pontos-chave: Harris

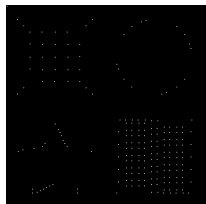


Detectores de pontos-chave: Harris

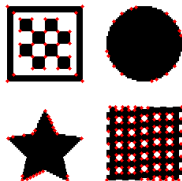
$$Harris(I) = [g(I_x^2)g(I_y^2) - g(I_x I_y)^2] - 0.04(g(I_x^2) + g(I_y^2))^2$$



Local Maxima



Detection



Detectores de pontos-chave: Harris

A partir dos autovalores da matriz A permite interpretar: regiões planas, bordas e cantos.



Hessian



Harris

Agenda

- 1 Introdução
- 2 Características Locais
- 3 Detectores de pontos-chave
 - Hessian
 - Harris
- 4 SIFT

Espaço-Escala



Detectores de pontos-chave: Difference of Gaussians

Considerando uma série de imagens no espaço-escala:

$$L(x, y, \sigma) = g_{\sigma}(f(x, y))$$

- O detector DoG (Difference of Gaussians) se baseia na diferença entre as escalas:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

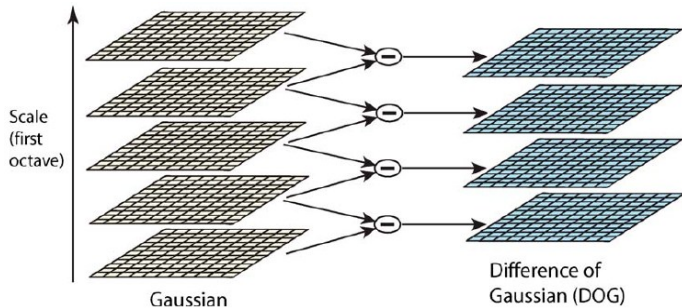
- A operação comumente é feita usando $\sigma = \sqrt{2}$ e:

$$L_1(x, y, \sigma) = g_{\sigma}(f(x, y))$$

$$L_2(x, y, \sigma) = g_{\sigma}(L_1(x, y))$$

$$DoG_1(x, y, \sigma) = L_2 - L_1$$

Detectores de pontos-chave: Difference of Gaussians



Detectores de pontos-chave: Difference of Gaussians



Hessian



Harris

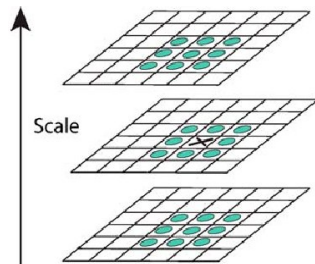


DoG

SIFT: Scale Invariant Feature Transform

Identifica máximos e mínimos locais

- 1 Cada ponto é comparado com os 26 vizinhos no espaço-escala.
- 2 Identificar se é mínimo ou máximo: ponto-chave candidato.
- 3 É possível aplicar filtros para remover respostas de baixo contraste e bordas (baseado em *Hessian*)



SIFT: Scale Invariant Feature Transform

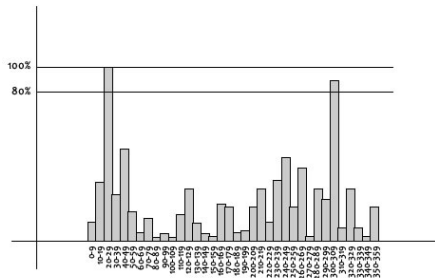
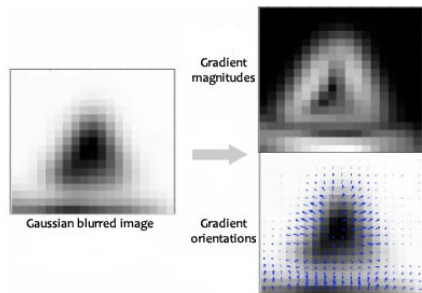
- Computa o Gradiente para cada imagem borrada

$$m(x, y) = \sqrt{L(x + 1, y) - L(x - 1, y)^2 + L(x, y + 1) - L(x, y - 1)^2}$$

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \right)$$

- Para a região ao redor de cada ponto-chave:
 - 1 Criar histograma com 36 orientações diferentes (deslocamento de 10 graus) acumulando as magnitudes.
 - 2 O valor máximo do histograma irá representar sua orientação.
 - 3 Esse valor é interpolado usando uma parábola em torno dos 3 vizinhos mais próximos: invariância à rotação.
- Cada ponto chave tem: localização, escala e orientação.

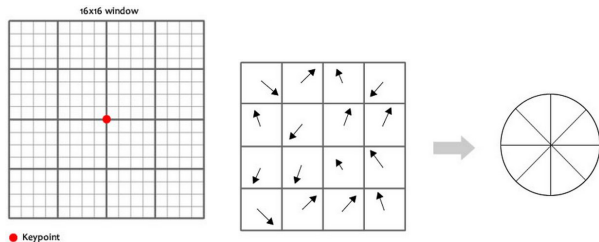
SIFT: Scale Invariant Feature Transform



- No histograma, para cada orientação acima de 80% do pico um novo ponto-chave é criado, com mesma escala e posição, mas orientação diferente.

SIFT: descritor local

- Para descrever cada ponto chave são amostradas as orientações (gradiente) ao redor do ponto
 - uma janela de 16×16 ao redor do ponto será analisada.
 - cada sub-janela de 4×4 formará um histograma de orientações.
 - ao somar uma orientação ao histograma, a magnitude é ponderada por uma gaussiana de forma que orientações referentes a pontos mais distantes tenham menor peso.



SIFT: descritor local

- O descritor final de cada ponto tem 128 números (8 orientações \times 4×4 janelas). Esses números formam um vetor de características para cada ponto.
 - O descritor deve ser normalizado para melhorar a invariância a iluminação

Referências

- J. Matas, O. Chum, M. Urban and T. Pajdla, Robust wide baseline stereo from maximally stable extremum regions. *British Machine Vision*, pp. 384-393, 2002.
- T. Tuytelaars and L. Van Gool, Matching widely separated views based on affine invariant regions . In *IJCV* 59(1):61-85, 2004
- K. Mikolajczyk and C. Schmid, Scale and Affine invariant interest point detectors. In *IJC V* 60(1):63-86, 2004.
- D. Lowe, Distinctive image features from scale invariant keypoints. In *IJCV* 60(2):91-110, 2004.
- T. Tuytelaars and K. Mikolajczyk, Local Invariant Feature Detectors: Survey. In *CVG*, 3(1):1-110, 2008.
- H. Bay, T. Tuytelaars and L. van Gool. SURF: Speeded Up Robust Features. In *9th ECCV*, LNCS 3951, pp. 404-417, 2006.
- C. Harris and M. Stephens. A combined corner and edge detector. *Proc. 4th Alvey Vision Conference*, pp.147-151, 1988.
- Affine Covariant Features: <http://www.robots.ox.ac.uk/~km>