Interest Point Detectors and Descriptors

IPAM - UCLA
July 24, 2013

Iasonas Kokkinos
Center for Visual Computing
Ecole Centrale Paris / INRIA Saclay
Introduction to Interest Point Detectors and Descriptors

Image processing
Image to Image

Computer Vision
Image to Symbols
Low-level symbolic representations

- Primal sketch

Low-level symbolic representations

- Primal sketch

Low-level symbolic representations

- Primal sketch

Goals for a low-level representation

- **Compact**
  - Reduce the number of processed image locations
    
    \[
    000000011111100000111111111000 = 000000011111100000111111111000
    \]
    
    \[
    = 7x0 \quad 5x1 \quad 5x0 \quad 7x1 \quad 3x0
    \]

- **Sufficient**
  - No need to look back into the image
Boundaries

Object/Surface Boundaries
A brief analogy with text

- What matters is what happens on word boundaries

- Mocpera iwht htsi sesm
  (compare with this mess)

- Evidence of our visual system employing boundary detection
Signal-level challenges

- Poor contrast
- Shadows
- Texture
Fundamental challenges: can humans do it?
Fundamental challenges: can humans do it?
Fundamental challenges: can humans do it?
Learning-based approaches

Boundary or non-boundary?

Use human-annotated segmentations

Use any visual cue as input to the decision function.
Use decision trees/logisitic regression/boosting/… and learn to combine the individual inputs.

Boundaries or edges?


Boundaries or edges?


Contours or points?

Introduction to Interest Point Detectors and Descriptors

Application: Image Stitching

Slide credit: Darya Frolova, Denis Simakov
Application: Image Stitching

- Procedure:
  - Detect feature points in both images

Slide credit: Darya Frolova, Denis Simakov
Application: Image Stitching

- Procedure:
  - Detect feature points in both images
  - Find corresponding pairs

Slide credit: Darya Frolova, Denis Simakov
Application: Image Stitching

- Procedure:
  - Detect feature points in both images
  - Find corresponding pairs
  - Use these pairs to align the images

Slide credit: Darya Frolova, Denis Simakov
Common Requirements

- Problem 1:
  - Detect the same point *independently* in both images

No chance to match!

Slide credit: Darya Frolova, Denis Simakov
Common Requirements

- Problem 1:
  - Detect the same point *independently* in both images

- Problem 2:
  - For each point correctly recognize the corresponding one

Slide credit: Darya Frolova, Denis Simakov
Laplacian-of-Gaussian

\[ \nabla^2 g_\sigma(x, y) = \frac{\partial^2 g_\sigma(x, y)}{\partial x^2} + \frac{\partial^2 g_\sigma(x, y)}{\partial y^2} \]
Early edge detection research

- Zero-crossings of LoG operator at increasing scales
- Different take: go for the maxima/minima
Finding blobs

Filtering = inner product between image patch and filter: template matching

\[ |I - f|^2 = \langle I - f, I - f \rangle \]
\[ = \langle I, I \rangle + \langle f, f \rangle - 2\langle f, I \rangle \]
\[ = C - 2\langle I, f \rangle \]
Scale selection

- First idea: convolve with Laplacians at several scales and find maximum in scale
- Observation: Laplacian decays as scale increases:

![Unnormalized Laplacian response plots](image)
Scale normalization

- The response of a derivative of Gaussian filter to a perfect step edge decreases as $\sigma$ increases.

$$\frac{1}{\sigma \sqrt{2\pi}}$$
Scale normalization

- The response of a derivative of Gaussian filter to a perfect step edge decreases as $\sigma$ increases.
- To keep response the same (scale-invariant), must multiply Gaussian derivative by $\sigma$.
- Laplacian is the second Gaussian derivative, so it must be multiplied by $\sigma^2$. 
Effect of scale normalization
Blob detection in 2D

Laplacian of Gaussian: Circularly symmetric operator for blob detection in 2D

\[ \nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \]
Blob detection in 2D

Laplacian of Gaussian: Circularly symmetric operator for blob detection in 2D

\[ \nabla^2_{\text{norm}} g = \sigma^2 \left( \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} \right) \]

Scale-normalized:
Scale selection

- Characteristic scale: peak of normalized Laplacian response

Scale invariance using scale selection
Scale-space blob detector

1. Convolve image with scale-normalized Laplacian at several scales
2. Find maxima of squared Laplacian response in scale-space
Scale-space blob detector: Example
Scale-space blob detector: Example

sigma = 3.1296

Slide credit: S. Lazebnik
Scale-space blob detector: Example

sigma = 4.8972

Slide credit: S. Lazebnik
Scale-space blob detector: Example

sigma = 7.6631

Slide credit: S. Lazebnik
Scale-space blob detector: Example

sigma = 11.9912

Slide credit: S. Lazebnik
Scale-space blob detector: Example

Blob coordinates: \((x, y, \text{scale})\)

Laplacian of Gaussian $\sim$ Difference of Gaussian

- We can efficiently approximate the Laplacian with a difference of Gaussians:

\[
L = \sigma^2 \left( G_{xx}(x, y, \sigma) + G_{yy}(x, y, \sigma) \right)
\]

(Laplacian)

\[
DoG = G(x, y, k\sigma) - G(x, y, \sigma)
\]

(Difference of Gaussians)

Efficient Computation (SIFT)

- Computation in Gaussian scale pyramid

Keypoint Detection (SIFT)

(a) 233x189 image
(b) 832 DoG extrema
(c) 729 left after peak value threshold
(d) 536 left after testing ratio of principle curvatures (removing edge responses)

**Second Moment Matrix**

\[ J = G_\rho \ast \left[ (\nabla G_\sigma \ast u)^T (\nabla G_\sigma \ast u) \right] \]

- Eigenvectors \( w_+, w_- \): directions of maximal and minimal variation of \( u \)
- Eigenvalues: amounts of minimal and maximal variation \( u \)

Interpreting the eigenvalues

Classification of image points using eigenvalues of the second moment matrix:

- **Corner**
  - $\lambda_1$ and $\lambda_2$ are large,
  - $\lambda_1 \sim \lambda_2$

- **Edge**
  - $\lambda_2 \gg \lambda_1$
  - $\lambda_1 \gg \lambda_2$

- **Flat** region
  - $\lambda_1$ and $\lambda_2$ are small

Slide credit: K. Grauman
Harris Detector: Steps
Harris Detector: Steps
Compute corner response $R$
Harris Detector: Steps

Find points with large corner response: $R > \text{threshold}$
Harris Detector: Steps
Take only the points of local maxima of $R$
Harris-Laplace [Mikolajczyk ‘01]

1. Initialization: Multiscale Harris corner detection

Slide adapted from Krystian Mikolajczyk
Harris-Laplace [Mikolajczyk ‘01]

1. Initialization: Multiscale Harris corner detection
2. Scale selection based on Laplacian
   (same procedure with Hessian ⇒ Hessian-Laplace)
SIFT computation

Slide credit: E. Tola
SIFT computation
Orientation Histogram

- 4x4 spatial bins (16 bins total)
- 8-bin orientation histogram per bin
- $8 \times 16 = 128$ dimensions total
- Normalized to unit norm
SIFT descriptor

- Image patch descriptor
- Location and characteristic scale: Blob/Corners detector
- Find orientation from orientation histogram

![Diagram of SIFT descriptor process]
SIFT invariances

- Spatial binning: tolerance to small shifts in location
- Orientation normalization
- Photometric normalization by making all vectors unit norm
- Orientation histogram: robustness to small local deformations
Application: Image Matching
Assumption: images undergo global deformations with a few degrees-of-freedom (e.g. scaling, rotation)

Correspondences of a few points suffice (found e.g. with SIFT)

Richard Szeliski’s talk, tomorrow
Open source implementation: www.vlfeat.org

This section features a number of tutorials illustrating some of the main algorithms implemented in VLFeat. The tutorials can be divided into three categories. The first class of algorithms detect and describe image regions (features). The second class of algorithms cluster.

Features

- **Covariant detectors.** An introduction to computing co-variant features like Harris-Affine.
- **Histogram of Oriented Gradients (HOG).** Getting started with this ubiquitous representation for object recognition.
- **Scale Invariant Feature Transform (SIFT).** Getting started with this popular feature detector / descriptor.
- **Dense SIFT (DSIFT) and PHOW.** A state-of-the-art descriptor for image categorization.
- **Maximally Stable Extremal Regions (MSER).** Extracting MSERs from an image.
- **Image distance transform.** Compute the image distance transform for fast part models and edge matching.

Clustering

- **Integer optimized k-means (IKM).** A quick overview of VLFeat fast k-means implementation.
- **Hierarchical k-means (HIKM).** Create a fast k-means tree for integer data.
- **Agglomerative Information Bottleneck (AIB).** Cluster discrete data based on the mutual information between the data and cl
- **Quick shift.** An introduction which shows how to create superpixels using this quick mode seeking method.
- **SLIC.** An introduction to SLIC superpixels.

Other

- **Pegasos SVM.** Learn a binary classifier and check its convergence plotting the energy value.
- **Forests of kd-trees.** Approximate nearest neighbor queries in high dimensions using an optimized forest of kd-trees.
- **Plotting functions for rank evaluation.** Learn how to plot ROC, DET, and precision-recall curves.
- **MATLAB Utilities.** A list of useful MATLAB functions bundled with VLFeat.

© 2007-13 The authors of VLFeat
Further reading (literature ‘seeds’)

- **Compact Codes & Large-scale Retrieval**
  - A. Babenko and V. Lempitsky, The Inverted Multi-Index, CVPR 12
  - R. Arandjelović, A. Zisserman, All about VLAD, CVPR 2013

- **Fast/Compact Descriptors**
  - SURF, FAST, ORB, FREAK,...
**Further reading (literature ‘seeds’)**

**Feature encoding**
- The devil is in the details: an evaluation of recent feature encoding methods, K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman, BMVC, 2011

**Descriptor Learning**
- Descriptor Learning for Efficient Retrieval, J. Philbin, M. Isard, J. Sivic, A. Zisserman, ECCV 10
Dense descriptors

- Interest point detection revisited

D. Lowe, Perceptual Organization and Visual Recognition, 1985
F. Jurie, B. Triggs, Sampling strategies for bag-of-work classification, 2005
Dense descriptors

- Interest point detection revisited: there is nothing special about corners

D. Lowe, Perceptual Organization and Visual Recognition, 1985
F. Jurie, B. Triggs, Sampling strategies for bag-of-work classification, 2005
Histogram of Orientated Gradients (HOG) descriptor

- Dalal and Triggs, ICCV 2005
  - Like SIFT descriptor, but for arbitrary box aspect ratio, and computed over all image locations and scales
  - Highly accurate detection using linear classifier

Feature vector $f = \left[ \ldots, \ldots, \ldots \right]$
Part score computation

\[ s[x] = \sum_y \langle h[x + y], w[y] \rangle \]
Part score

\[ h[x] \quad s[x] = \sum_y \langle h[x + y], w[y] \rangle \]
Dense descriptors: motivation

Narrow baseline: Pixel Difference + Graph Cuts

DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo, E. Tola, V. Lepetit, P. Fua, PAMI, 10
Dense descriptors: motivation

Wide baseline: Pixel Difference + Graph Cuts

DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo, E. Tola, V. Lepetit, P. Fua, PAMI, 10
Dense descriptors

Wide baseline: **SIFT Descriptor** + Graph Cuts

DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo, E. Tola, V. Lepetit, P. Fua, PAMI, 10
Fast dense descriptors

Wide baseline: **DAISY Descriptor** + Graph Cuts

DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo, E. Tola, V. Lepetit, P. Fua, PAMI, 10
SIFT-> DAISY

DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo, E. Tola, V. Lepetit, P. Fua, PAMI, 10
Introduction to Interest Point Detectors and Descriptors

**SIFT -> DAISY**

<table>
<thead>
<tr>
<th>Detectors</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SIFT</strong></td>
<td>+ Good Performance</td>
</tr>
<tr>
<td></td>
<td>- Not suitable for dense computation</td>
</tr>
<tr>
<td><strong>Sym.SIFT</strong></td>
<td>+ Gaussian Kernels: Suitable for Dense Computation</td>
</tr>
<tr>
<td><strong>GLOH</strong></td>
<td>+ Good Performance</td>
</tr>
<tr>
<td></td>
<td>+ Better Localization</td>
</tr>
<tr>
<td></td>
<td>- Not suitable for dense computation</td>
</tr>
</tbody>
</table>

Introduction to Interest Point Detectors and Descriptors

SIFT -> DAISY

DAISY

Sym.SIFT

GLOH

Suitable for dense computation
+ Improved performance:
  + Precise localization
  + Rotational Robustness

+ Suitable for Dense Computation

+ Good Performance
  + Better Localization

- Not suitable for dense computation

* S. Winder and M. Brown. *Learning Local Image Descriptors* in CVPR’07
Daisy computation

DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo, E. Tola, V. Lepetit, P. Fua, PAMI, 10
Daisy computation

DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo, E. Tola, V. Lepetit, P. Fua, PAMI, 10
Daisy computation

- Rotating the descriptor only involves reordering the histograms.
- The computation mostly involves 1D convolutions, which is fast.

DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo, E. Tola, V. Lepetit, P. Fua, PAMI, 10
Scale- and rotation- invariance through Fourier Transforms

Lecture 1: Modulation property: \[ f(x) \leftrightarrow F(\omega) \quad f(x)e^{j\omega_c x} \leftrightarrow F(\omega - \omega_c) \]

Fact 1: Fourier shifting property: \[ f[i - n_i, j - n_j] \rightarrow F \exp \left(-j \left( n_i \frac{2\pi}{N} + n_j \frac{2\pi}{K} \right) \right) \]

**Signal translation does not affect the signal’s Fourier Transform Magnitude**

Fact 2: the log-polar sampling turns image scaling and rotation to translation:

Fact 1+2: the Fourier Transform Modulus of log-polar descriptors is invariant

D. Casasent and D. Psaltis, Rotation and scale-invariant optical correlation, Applied Optics, 1976
Dense Scale-Invariant Descriptors

D. Casasent and D. Psaltis, Rotation and scale-invariant optical correlation, Applied Optics, 1976
Dense Scale-Invariant Descriptors

\[ \omega_n = 1, \omega_k = 1, d \in \{1, 2, 3\} \]

D. Casasent and D. Psaltis, Rotation and scale-invariant optical correlation, Applied Optics, 1976
Dense Scale-Invariant Descriptors

\[ \omega_n = 1, \omega_k = 2, d \in \{1, 2, 3\} \]

Dense Scale-Invariant Descriptors

Query similarity to red (left) and green (right) reference points.

I. Kokkinos, M. Bronstein and A. Yuille, Dense Scale-Invariant Descriptors for Images and Surfaces, Technical report 2012
D. Casasent and D. Psaltis, Rotation and scale-invariant optical correlation, Applied Optics, 1976
Segmentation-aware descriptors

- We improve descriptors by suppressing the background with soft segmentations.

Segmentation masks: $w^{[i]} = \exp\left(-\lambda \cdot d(x, G^{[i]}(x))\right)$

- Simple, generic and fast.
- Scale- and rotation- invariant, if need be.

Introduction to Interest Point Detectors and Descriptors

Application-1: large displacement optical flow

SIFT-flow for optical flow estimation
JHU/MOSEG benchmark for evaluation

MOSEG: Cars 1

Frame 1  DSIFT  SID

Frame 2  SLS  Ours (SID SS-Gb)

IN RED: ground truth segmentation (when available)


Application 2: Wide-baseline stereo

Dense wide-baseline stereo reconstruction with Graph Cuts
Code

- http://vision.mas.ecp.fr/Personnel/iasonas/descriptors.html
- http://www.iri.upc.edu/people/etrulls/#code
Features for 3D data

- Intensity data: Histogram of Gradients, SIFT etc.

- Depth/surface data?

Kinect

Kinect Fusion
Dense invariant surface descriptors

- Recent advances in surface analysis
- Combination with computer vision techniques

M. Bronstein, I. Kokkinos, Scale-Invariant Heat Kernel Signatures, CVPR 2010
I. Kokkinos, M. Bronstein, R. Litman, A. Bronstein, Intrinsic Shape Context, CVPR 2012

Slides from:

http://www.cs.technion.ac.il/~mbron/teaching.html
http://tosca.cs.technion.ac.il/book/course_milano08.html
**Diffusion geometry**

Heat equation

\[ \left( \Delta_S + \frac{\partial}{\partial t} \right) u(t, x) = 0. \]

where

- \( \Delta_S \) - Laplace-Beltrami operator

\( u \) - heat distribution

**Fundamental solution:** solution for initial conditions \( u(x, 0) = \delta(x - y) \)

Represented in the Laplace-Beltrami eigenbasis as

\[
k_t(x, y) = \sum_{i=0}^{\infty} e^{-t\lambda_i} \phi_i(x) \phi_i(y)
\]

Coefficients decay fast: approximation by **truncated sum**

\[
k_t(x, y) \approx \sum_{i=0}^{N} e^{-t\lambda_i} \phi_i(x) \phi_i(y)
\]
Chladni’s patterns
http://www.youtube.com/watch?v=cT30XOfd1yl
Laplace-Beltrami eigenfunctions
Laplace-Beltrami eigenfunctions: invariance
From the heat kernel to heat kernel signature

Multiscale local shape descriptor

\[ p(x) = (K_{t_1}(x, x), \ldots, K_{t_n}(x, x)) \]

Interpretation: multiscale gaussian curvature (time = scale)
Heat kernel signature

Heat kernel signatures represented in RGB space

J. Sun, M. Ovsjanikov, L. Guibas, SGP 2009
Heat kernel descriptors

Invariant to isometric deformations

Localized sensitivity to topological noise
Scale invariance?

Original shape
\[ \lambda, \phi \]
\[ \text{HKS} = K_t(x, x) \]

Scaled by \( 1/\alpha \)
\[ \alpha^2 \lambda, \alpha \phi \]
\[ \text{HKS} = \alpha^2 K_{\alpha^2 t}(x, x) \]

Not scale invariant!
Scale-invariant heat kernel signature

Log scale-space

\[ K_{\alpha^\tau} \rightarrow \alpha^2 K_{\alpha^{\tau+2}} \]

Log + \( d/d\tau \)

\[ \log \alpha^2 + \log K_{\alpha^{\tau+2}} \]

\[ \frac{d}{d\tau} \log K_{\alpha^{\tau+2}} \]

Fourier transform magnitude

\[ \mathcal{F} \frac{d}{d\tau} \log K_{\alpha^{\tau+2}} = e^{2i\omega\pi} \mathcal{F} d/d\tau \log K_{\alpha^\tau} \]

Scaling = shift and multiplicative constant in HKS

Undo scaling

Undo shift

Michael Bronstein, Iasonas Kokkinos, CVPR 2010
Introduction to Interest Point Detectors and Descriptors

Invariance

- Heat kernel signature (HKS)
  - Rigid: ✓
  - Scale: ×
  - Inelastic: ✓
  - Topology: ✓

- Scale-invariant HKS (SI-HKS)
  - Rigid: ✓
  - Scale: ✓
  - Inelastic: ✓
  - Topology: ✓
Geometric vocabulary
Bags of features

\[ p(x) \]

Geometric vocabulary \( p_1, \ldots, p_V \)

\[ k^* = \arg \min_{i=1,\ldots,V} \| p(x) - p_i \| \]

Nearest neighbor in the descriptor space

\[ \theta_k(x) = \begin{cases} 
1 & k = k^* \\
0 & \text{else}
\end{cases} \]

M. Ovsjanikov, M Bronstein A. Bronstein, L. Guibas, 2009
Bags of features

Statistics of different geometric words over the entire shape

\[ f(X) = \int_X \theta(x) \, dx \]

Shape distance = distance between bags of features

\[ d_{\text{BOF}}(X, Y) = \| f(X) - f(Y) \| \]
Bags of features
## ShapeGoogle with HKS descriptor (mAP %)

<table>
<thead>
<tr>
<th>Transform</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Isometry</td>
<td>100.00</td>
</tr>
<tr>
<td>Topology</td>
<td>100.00</td>
</tr>
<tr>
<td>Holes</td>
<td>100.00</td>
</tr>
<tr>
<td>Micro holes</td>
<td>100.00</td>
</tr>
<tr>
<td>Scale</td>
<td>0.98</td>
</tr>
<tr>
<td>Local scale</td>
<td>100.00</td>
</tr>
<tr>
<td>Sampling</td>
<td>100.00</td>
</tr>
<tr>
<td>Noise</td>
<td>100.00</td>
</tr>
<tr>
<td>Shot noise</td>
<td>100.00</td>
</tr>
<tr>
<td>Partial</td>
<td>7.54</td>
</tr>
<tr>
<td>Mixed</td>
<td>53.13</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>94.94</strong></td>
</tr>
</tbody>
</table>
Introduction to Interest Point Detectors and Descriptors

<table>
<thead>
<tr>
<th>Transform</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Isometry</td>
<td>100.00</td>
</tr>
<tr>
<td>Topology</td>
<td>96.15</td>
</tr>
<tr>
<td>Holes</td>
<td>100.00</td>
</tr>
<tr>
<td>Micro holes</td>
<td>100.00</td>
</tr>
<tr>
<td>Scale</td>
<td>91.03</td>
</tr>
<tr>
<td>Local scale</td>
<td>100.00</td>
</tr>
<tr>
<td>Sampling</td>
<td>100.00</td>
</tr>
<tr>
<td>Noise</td>
<td>100.00</td>
</tr>
<tr>
<td>Shot noise</td>
<td>100.00</td>
</tr>
<tr>
<td>Partial</td>
<td>17.43</td>
</tr>
<tr>
<td>Mixed</td>
<td>56.47</td>
</tr>
<tr>
<td>Average</td>
<td>97.05</td>
</tr>
</tbody>
</table>

ShapeGoogle with SI-HKS descriptor (mAP %)

M. Bronstein, I.K., CVPR 2010
Scale-invariant retrieval

Scale 1.3

Heat Kernel Signature

Scale-Invariant Heat Kernel Signature

M. Bronstein, I.K., 2010
Scale-invariant retrieval

Local scale

Heat Kernel Signature

Scale-Invariant Heat Kernel Signature

M. Bronstein, I.K., 2010
Intrinsic Shape Contexts

Goal: use context information for descriptor construction

Naïve: work at larger feature scale (more smoothing)
Better: stack together neighbors (meta-descriptor)

Surface processing: local surface charting

Intrinsic: invariant to isometric deformations
Intrinsic Shape Context Construction

Uniformly sample directions on the tangent plane of a point
Shoot and track geodesics outwards from the point
Construct soft angular membership functions based on distance from geodesics
Construct soft radial membership functions based on distance from point
Intrinsic Shape Contexts: higher discrimination

IK, M. Bronstein, R. Littman, A. Bronstein, ISC, CVPR 2012