SIFT: Scale Invariant Feature Transform

Robert B. Fisher
School of Informatics
University of Edinburgh

SIFT Features

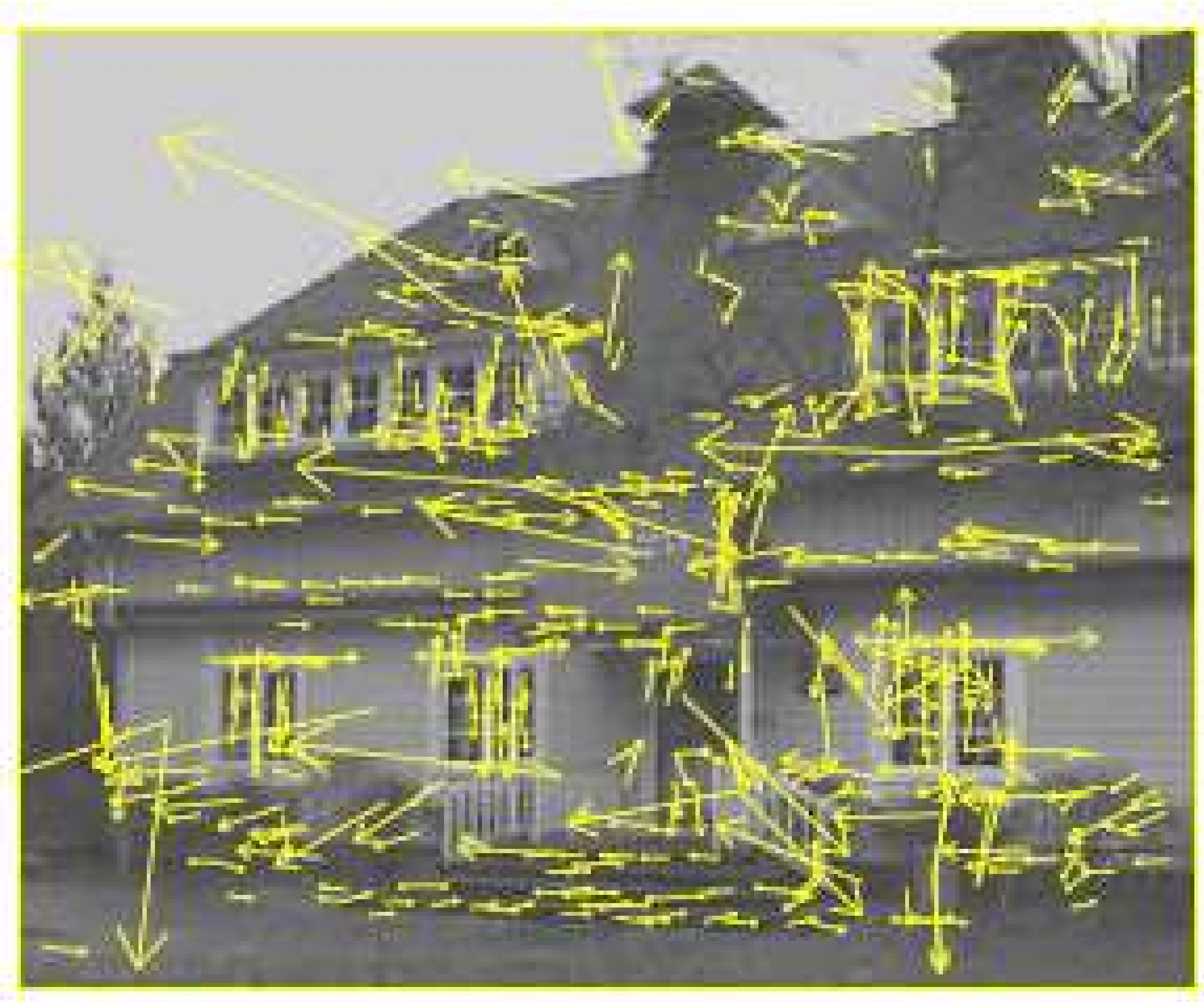
SIFT: Scale Invariant Feature Transform

Image points + local description
(128 vector)

Sparse, reasonably distinguishable points

Invariant to translation, rotation, scale,
some 3D

Example feature locations



Matching Applications

Matchable features for:

- Object recognition
- Model-data alignment
- Image registration
- Stereo matching

Four Step Algorithm

1. Detect extremal points in scale space
2. Accurate keypoint subpixel localization
3. Feature orientation estimation
4. Keypoint descriptor calculation

Scale Space Smoothing

Gaussian smoothing via convolution

$$L(x, y, \sigma) = G(x, y, \sigma) \circ I(x, y)$$

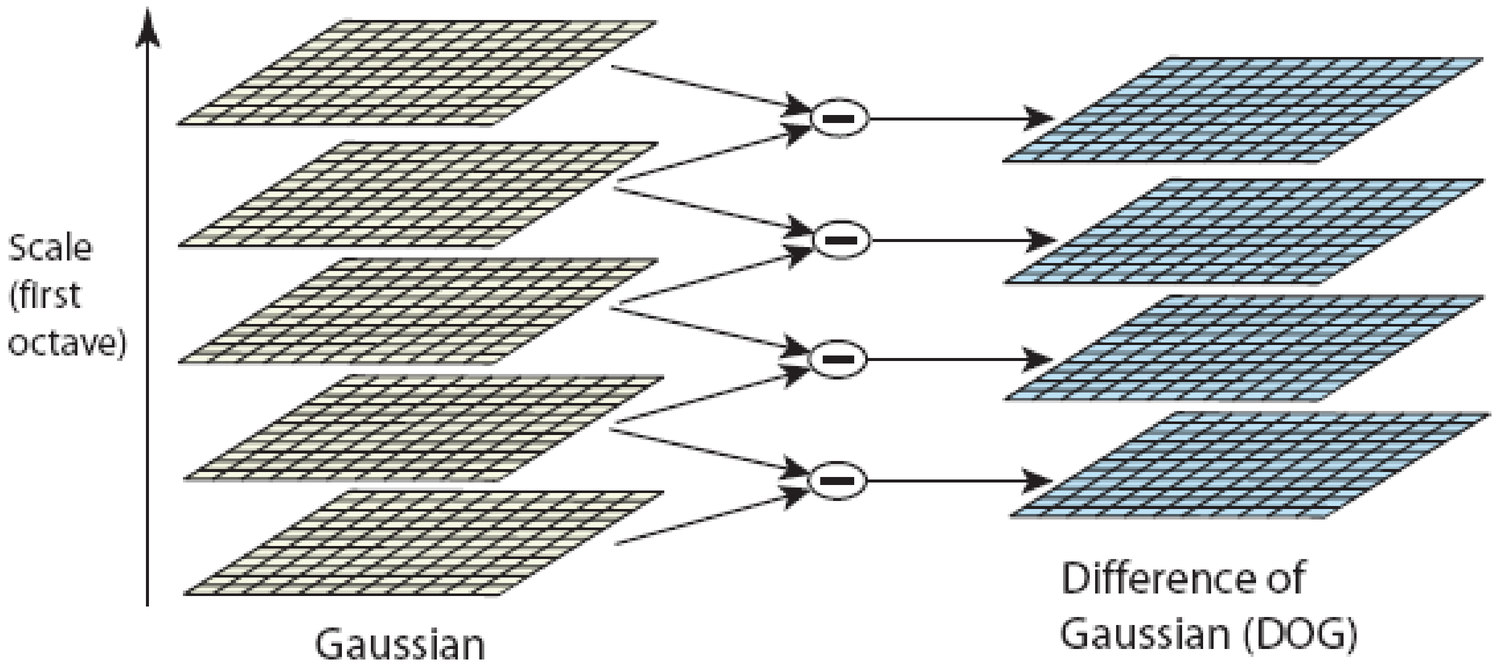
$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

Difference of Gaussians:

$$D(x, y, n) = L(x, y, 2^{\frac{n}{S}}) - L(x, y, 2^{\frac{n-1}{S}})$$

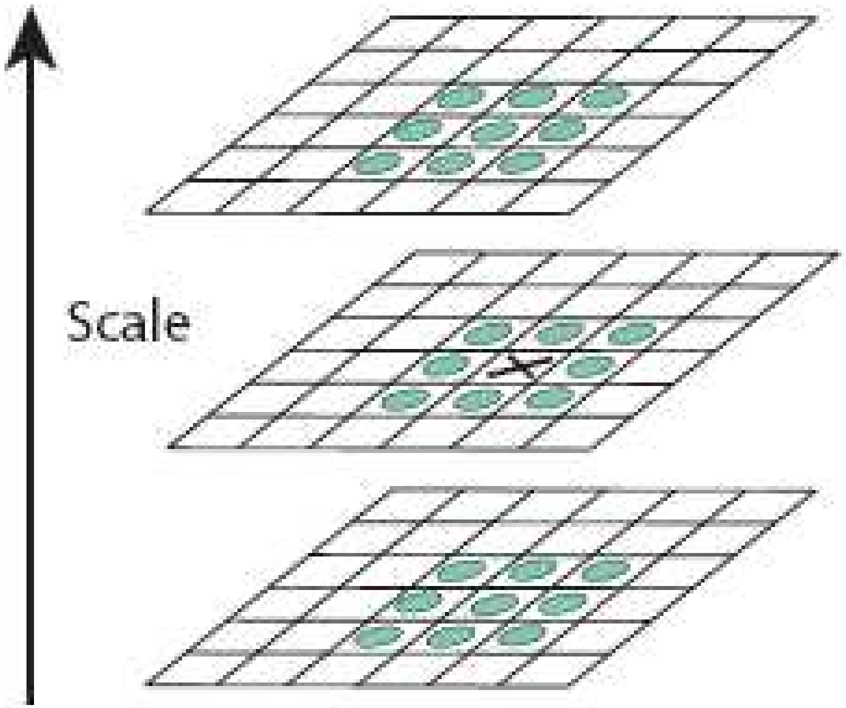
where $n = 1 \ldots N$
$S = 3$ best

Scale Space



Point Extrema

Pick extremal points larger/smaller than their 26 neighbours:



Subpixel Extrema Localization

Hessian:

$$\mathrm{H}_3 = \begin{bmatrix} \partial^2 D/\partial x^2 & \partial^2 D/\partial x \partial y & \partial^2 D/\partial x \partial \sigma \\ \partial^2 D/\partial x \partial y & \partial^2 D/\partial y^2 & \partial^2 D/\partial y \partial \sigma \\ \partial^2 D/\partial x \partial \sigma & \partial^2 D/\partial y \partial \sigma & \partial^2 D/\partial \sigma^2 \end{bmatrix}$$

Optimal position is $(x, y, \sigma) + \hat{x}$, where

$$\hat{x} = -\mathrm{H}_3^{-1} \begin{bmatrix} \partial D / \partial x \\ \partial D / \partial y \\ \partial D / \partial \sigma \end{bmatrix}$$

Low Contrast Extrema Pruning

Predict DoG value at subpixel extrema:

$$p = \mid D(x, y, \sigma) + \frac{1}{2}\left[\frac{\partial D}{\partial x}, \frac{\partial D}{\partial y}, \frac{\partial D}{\partial \sigma}\right]\hat{x} \mid$$

Reject if $p < 0.03$

Unstable Point Extrema Pruning

Let

$$\mathrm{H}_2 = \begin{bmatrix} \partial^2 D/\partial x^2 & \partial^2 D/\partial x \partial y \\ \partial^2 D/\partial x \partial y & \partial^2 D/\partial y^2 \end{bmatrix}$$

Reject if $det(H_2) < 0$ or

$$\frac{trace(\mathrm{H}_2)^2}{det(\mathrm{H}_2)} > \tau \ (e.g.12)$$

Rejects points that can slide along an edge

Getting Rotation Invariance

Local orientation $\hat{\theta}$ estimation

Use keypoint scale σ

Let $\vec{v} = \nabla L(r, s, \sigma)$ for $(r, s) \in neigh(x, y)$
Compute strength $m = \mid \vec{v} \mid$ and
$\theta = direction(\vec{v})$
Compute histogram of θ values weighted by m

Pick top peak direction $\hat{\theta}$ in histogram for feature orientation

Local Descriptor Computation

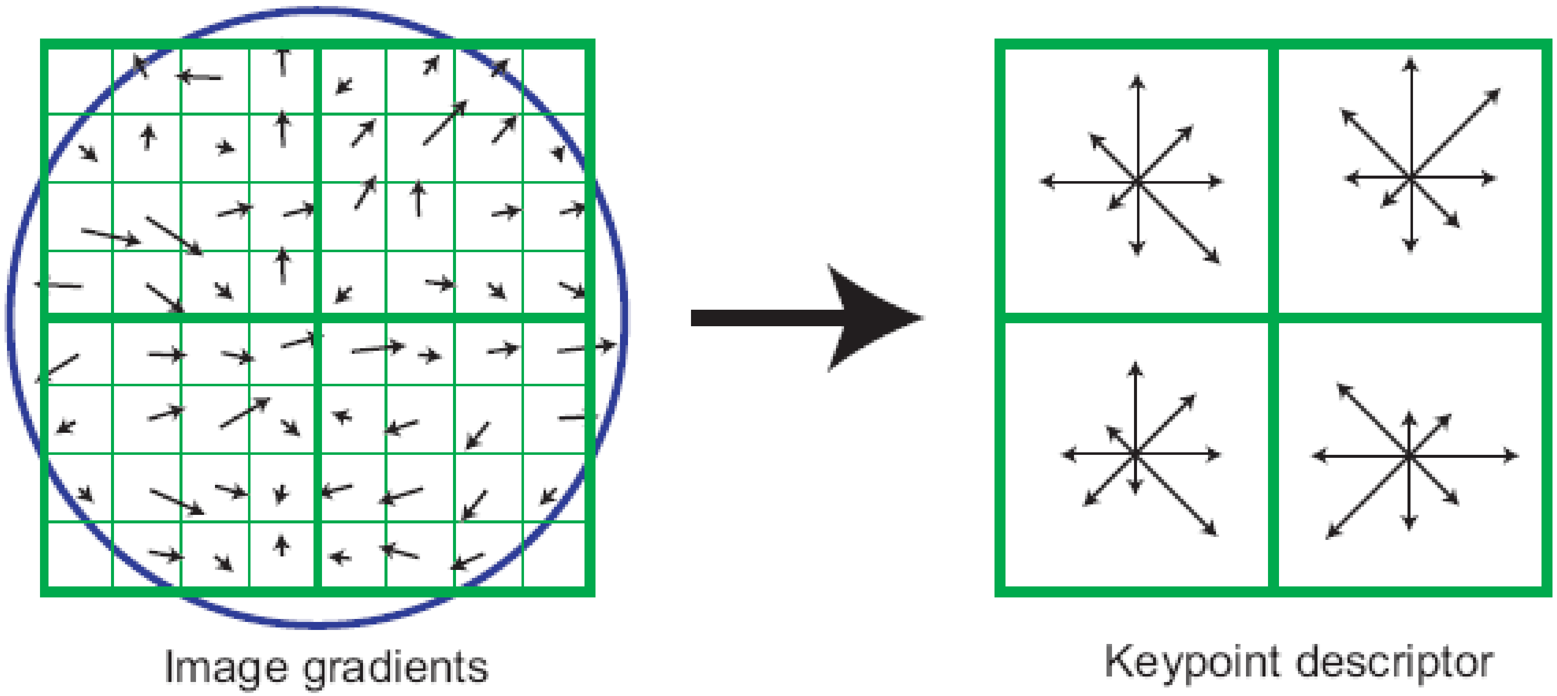
Use 16×16 neighbourhood about feature point subdivided into 16 4×4 pixel blocks
Create an 8 orientation histogram for each block
$\rightarrow$ 128 vector
Compute gradient orientation at each point
Rotate all orientations by $\hat{\theta}$ (for invariance)
Add to histogram weighted (details in paper)

Normalize 128 vector to unit length for illumination invariance

Descriptor similarity using Euclidean distance

Descriptor Example

4 histograms from 8 × 8 neighbourhood about feature point:



SIFT Summary

- Sparse, distinctive point features
- Translation independent by using local histogram
- Rotation independent by orientation adjustment
- Scale independent by extremal scale estimation
- Illumination independent by descriptor normalisation
- Widely used
- Real-time implementation possible

SIFT References

www.cs.ubc.ca/~lowe/papers/ijcv04.pdf

en.wikipedia.org/wiki/Scale-invariant_feature_transform