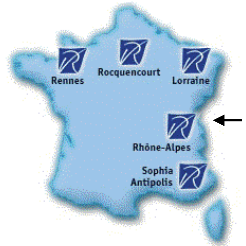
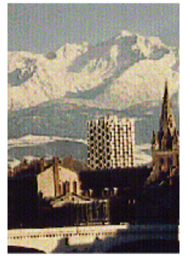
Recognition and Matching

based on local invariant features

Cordelia Schmid
INRIA, Grenoble

Cordelia Schmid
LEAR research team
INRIA, Rhône-Alpes





LEAR research team

Learning and Recognition in Vision

- INRIA team created in July 2003
- 4 researchers, 9 PhDs, 3 postdocs, 4 engineers, ….
- Goal : scene interpretation, object and activity recognition
- http://lear.inrialpes.fr

LEAR team

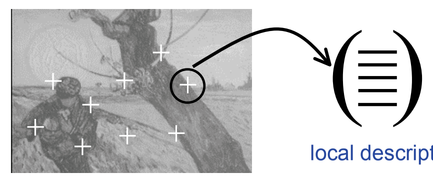
- Research interests
  - Local invariant image description
  - Image correspondence, indexing, object recognition
  - Machine learning for visual classification
  - Human detection, pose & motion from single images & videos
- Collaborations
  - Industrial: Xerox, MBDA, Thales, Bertin
  - Academic: U.Illinois, UC Berkeley, UBC Vancouver, U.Oxford, Australian Nat U
- Contracts
  - 2 EU projects on visual recognition
  - Core member of EU Network on machine learning & applications
  - 2 French projects, on recognition and data collection

Overview

- Introduction to local features
- Scale & affine invariant interest points/regions detectors
- Evaluation and comparison of different detectors
- Region descriptors and their performance
- Object recognition, demonstration of a system

Introduction

Local invariant photometric features



Local : robust to occlusion/clutter + no object segmentation
Photometric : distinctive
Invariant : to image transformations + illumination changes

History - Matching

Matching based on line segments

⇒Not very discriminant & lack of robustness

⇒Solution : matching with interest points & correlation

⇒Interest points are discriminant features at 2D signal changes

[A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Z. Zhang, R. Deriche, O. Faugeras and Q. Luong. Artificial Intelligence 1995.]

Approach [Zhang et al.'95]

- Extraction of interest points with the Harris detector
- Comparison of points with cross-correlation
- Verification with the fundamental matrix

Harris detector

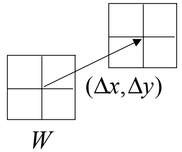
Based on the idea of auto-correlation

Important difference in all directions => interest point

Harris detector

Auto-correlation function for a point (x, y) and a shift $(\Delta x, \Delta y)$

$$a(x,y) = \sum_{(x_k, y_k) \in W(x,y)} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2$$



$$a(x,y) \begin{cases} \text{small in all directions} & \rightarrow \text{uniform region} \\ \text{large in one directions} & \rightarrow \text{contour} \\ \text{large in all directions} & \rightarrow \text{interest point} \end{cases}$$

Harris detector

Discret shifts are avoided based on the auto-correlation matrix

with first order approximation

$$I(x_k + \Delta x, y_k + \Delta y) = I(x_k, y_k) + (I_x(x_k, y_k) \quad I_y(x_k, y_k)) \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

$$a(x,y) = \sum_{(x_k, y_k) \in W(x,y)} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2$$

$$= \sum_{(x_k, y_k) \in W} \left(\begin{pmatrix} I_x(x_k, y_k) & I_y(x_k, y_k) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right)^2$$

Harris detector

$$= (\Delta x \quad \Delta y) \begin{bmatrix} \sum_{(x_k, y_k) \in W} (I_x(x_k, y_k))^2 & \sum_{(x_k, y_k) \in W} I_x(x_k, y_k) I_y(x_k, y_k) \\ \sum_{(x_k, y_k) \in W} I_x(x_k, y_k) I_y(x_k, y_k) & \sum_{(x_k, y_k) \in W} (I_y(x_k, y_k))^2 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

Auto-correlation matrix

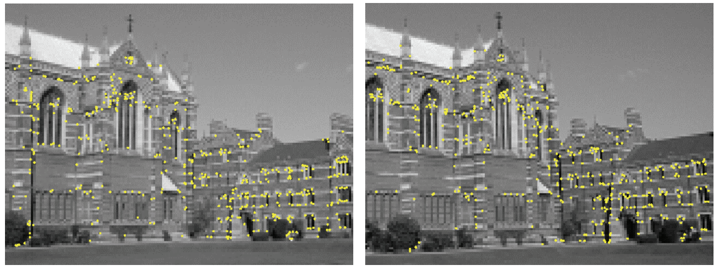
the sum can be smoothed with a Gaussian

$$= (\Delta x \quad \Delta y) G \otimes \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

Harris detector

- Auto-correlation matrix
  - captures the structure of the local neighborhood
  - measure based on eigenvalues of this matrix
    - 2 strong eigenvalues => interest point
    - 1 strong eigenvalue => contour
    - 0 eigenvalue => uniform region
- Interest point detection
  - threshold on the eigenvalues
  - local maximum for localization

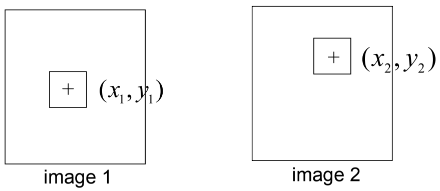
Harris detector



Interest points extracted with Harris (~ 500 points)

Cross-correlation

Comparison of the intensities in the neighborhood of two interest points

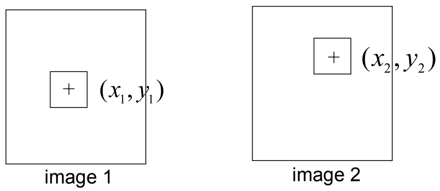


SSD : sum of square difference

$$\frac{1}{(2N+1)^2}\sum_{i=-N}^{N}\sum_{j=-N}^{N}(I_1(x_1+i, y_1+j)-I_2(x_2+i, y_2+j))^2$$

Cross-correlation

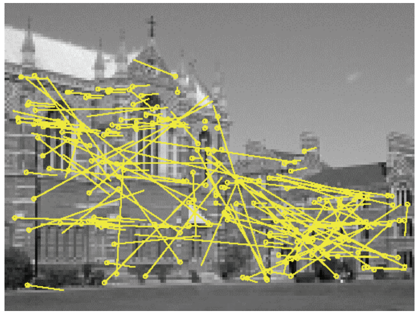
Comparison of the intensities in the neighborhood of two interest points



ZNCC: zero normalized cross correlation

$$\frac{1}{(2N+1)^2}\sum_{i=-N}^{N}\sum_{j=-N}^{N}\left(\frac{I_1(x_1+i, y_1+j)-m_1}{\sigma_1}\right)\cdot\left(\frac{I_2(x_2+i, y_2+j)}{\sigma_2}\right)$$

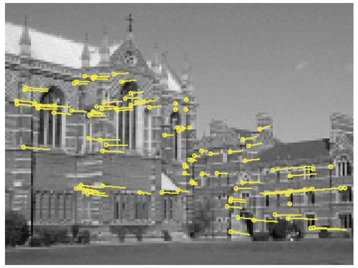
Cross-correlation matching



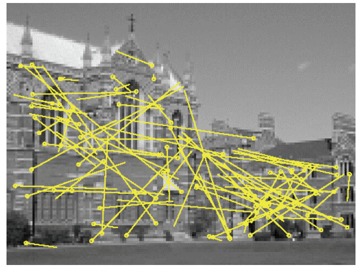
Initial matches (188 pairs)

Global constraints

Robust estimation of the fundamental matrix



99 inliers



89 outliers

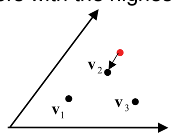
Summary of the approach

- Very good results in the presence of occlusion and clutter
  - local information
  - discriminant greyvalue information
  - robust estimation of the global relation between images
  - for limited view point changes
- Solution for more general view point changes
  - local scale & affine invariant photometric descriptors

History - Recognition

Eigenimages [Turk 91]

- Each face vector is represented in the eigenimage space
  - eigenvectors with the highest eigenvalues = eigenimages



- The new image is projected into the eigenimage space
  - determine the closest face

⇒ not robust to occlusion, requires object segmentation, not invariant, discriminant

History - Recognition

Geometric invariants [Rothwell 92]

- Function with a value independent of the transformation

  $f(x', y') = f(x, y)$ where $(x', y')^t = T(x, y)^t$

- Invariant for image rotation : distance between two points

- Invariant for planar homography : cross-ratio

=> local and invariant, not discriminant, requires sub-pixel extraction of primitives

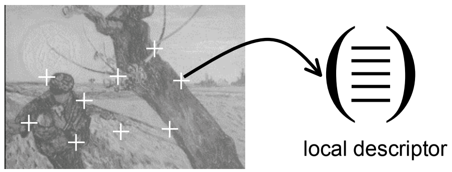
History - Recognition

Problems : occlusion, clutter, image transformations, distinctiveness

$\Rightarrow$ Solution : recognition with local photometric invariants

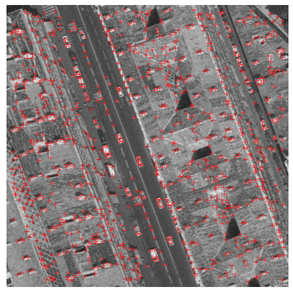
[Local greyvalue invariants for image retrieval, C. Schmid and R. Mohr, PAMI 1997]

Approach [Schmid & Mohr'97]



1) Extraction of interest points (characteristic locations)
2) Computation of local descriptors
3) Determining correspondences
4) Selection of similar images

Interest points

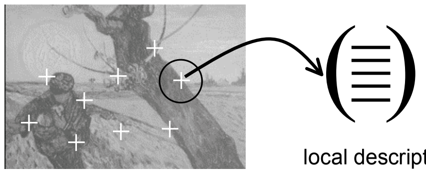


Geometric features
➡ repeatable under transformations

2D characteristics of the signal
➡ high informational content

Comparison of different detectors [Schmid98] ➡ Harris detector

Local descriptors



Descriptors characterize the local neighborhood of a point

Local descriptors

Greyvalue derivatives

$$\mathbf{v}(x,y) = \begin{pmatrix} I(x,y) * G(\sigma) \\ I(x,y) * G_x(\sigma) \\ I(x,y) * G_y(\sigma) \\ I(x,y) * G_{xx}(\sigma) \\ I(x,y) * G_{xy}(\sigma) \\ I(x,y) * G_{yy}(\sigma) \\ \vdots \end{pmatrix}$$

$$I(x,y) * G(\sigma) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x',y',\sigma) I(x-x',y-y') dx' dy'$$

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

Local descriptors

Greyvalue derivatives

$$\mathbf{v}(x,y) = \begin{pmatrix} I(x,y) * G(\sigma) \\ I(x,y) * G_x(\sigma) \\ I(x,y) * G_y(\sigma) \\ I(x,y) * G_{xx}(\sigma) \\ I(x,y) * G_{xy}(\sigma) \\ I(x,y) * G_{yy}(\sigma) \\ \vdots \end{pmatrix} = \begin{pmatrix} L(x,y) \\ L_x(x,y) \\ L_y(x,y) \\ L_{xx}(x,y) \\ L_{xy}(x,y) \\ L_{yy}(x,y) \\ \vdots \end{pmatrix}$$

Local descriptors

Invariance to image rotation : differential invariants [Koen87]

$$\begin{bmatrix} L \\ L_x L_x + L_y L_y \\ L_{xx} L_x L_x + 2 L_{xy} L_x L_y + L_{yy} L_{yy} \\ L_{xx} + L_{yy} \\ L_{xx} L_{xx} + 2 L_{xy} L_{xy} + L_{yy} L_{yy} \\ \cdots \\ \cdots \\ \cdots \\ \cdots \end{bmatrix}$$

Local descriptors

- Robustness to illumination changes

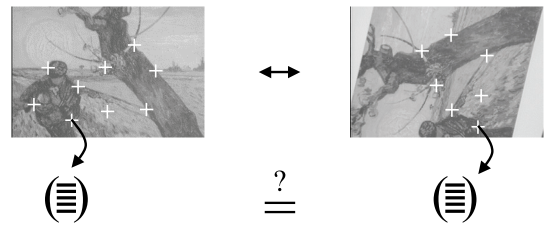
  in case of an affine transformation $I_1(\mathbf{x}) = aI_2(\mathbf{x}) + b$

- Normalization of derivatives with gradient magnitude

$$(L_{xx} + L_{yy}) \Big/ \sqrt{L_x L_x + L_y L_y}$$

- Normalization of the image patch with mean and variance

Determining correspondences



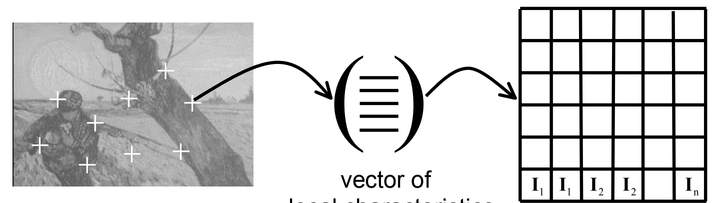
Vector comparison using the Mahalanobis distance

$$dist_M(\mathbf{p},\mathbf{q}) = \sqrt{(\mathbf{p}-\mathbf{q})^T \Lambda^{-1} (\mathbf{p}-\mathbf{q})}$$

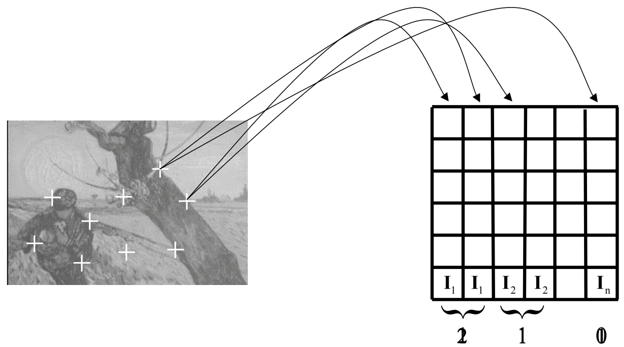
Selection of similar images

- In a large database
  - voting algorithm
  - additional constraints
- Rapid acces with an indexing mechanism
  - Kd tree

Voting algorithm



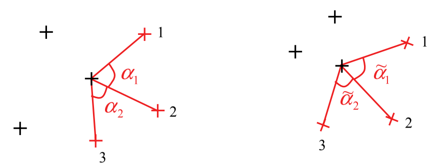
Voting algorithm



I_1 is the corresponding model image

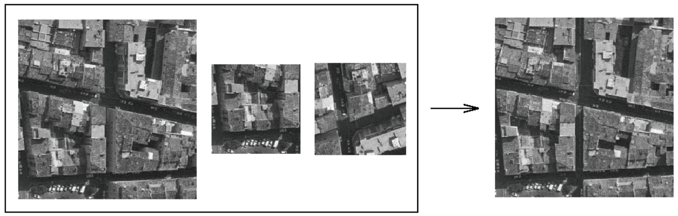
Additional constraints

- Semi-local constraints
  - neighboring points should match
  - angles, length ratios should be similar



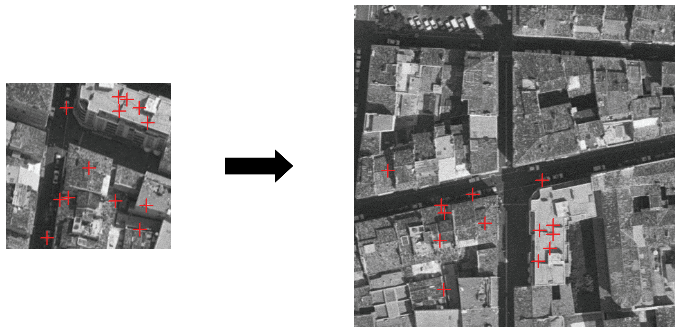
- Global constraints
  - robust estimation of the image transformation (homogaphy, epipolar geometry)

Results

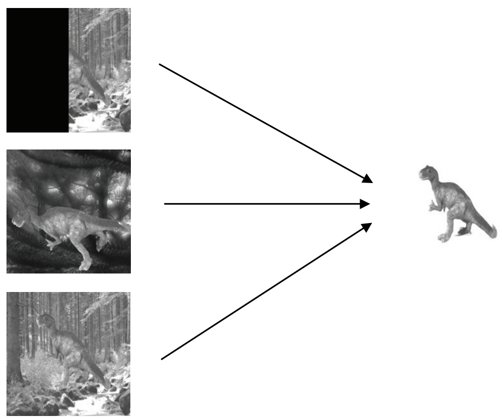


database with ~1000 images

Results



Results



Summary of the approach

- Very good results in the presence of occlusion and clutter
  - local information
  - discriminant greyvalue information
  - invariance to image rotation and illumination
- No invariance to scale and affine changes
- Solution for more general view point changes
  - local descriptors invariant to scale and affine transformations
  - extraction of invariant points and regions