Approche locale “a contrario”
pour la reconnaissance multiple d’objets

Julien Rabin
CNRS, Université Paris Dauphine (CÉRÉMADE)

GdR ISIS, 7 Mai 2010, Télécom Paristech

Avec
Julie Delon, CNRS (LTCI)
Yann Gousseau, Télécom ParisTech (LTCI)
Lionel Moisan, Université Paris Descartes (MAP5)
Most of computer vision problems are based on **local representation** of images: 
*Image registration, indexation, classification, mosaicing, motion segmentation, object detection, object recognition, camera calibration, etc.*

Two methodologies:

- Bag of features,
- Geometric grouping of local features
Object Recognition example

Is there any similar objects between these images?
Preliminary step: Local representation of images

- Extract local and invariant descriptors, SIFT [Lowe 1999], Shape Context [Belongie and Malik, 2000], MSER [Matas et al. 2002], etc.
Step 1: Local features comparison

Define a robust dissimilarity measure between local features
Step 2: Local features matching

- Local Representation
- Features Comparison
- Features Matching
- Features Grouping

Select **reliable (multiple)** correspondences between local features
Step 3: Local features grouping

- Detect a group and estimate the geometrical model's parameters related to the recognized object
Objectives and outline

➢ How to set automatically reliable detection thresholds for multiple object recognition?

Part I. Local features detection

Part II. A contrario methodology

Part III. Multiple local features matching

Part IV. Multiple local features grouping
Part I

Local descriptors

Local Representation → Features Comparison → Features Matching → Features Grouping
Local features

1. **Discrete image** $u$

2. **Scale-space representation**
   \[ \forall \sigma, \quad u_\sigma = g_\sigma \ast u \]

3. **Local extrema** $(\vec{x}, \sigma)$ in space and scale of $\sigma^2 \Delta u_\sigma$

4. Harris (or Hessian) multiscale criterion to eliminate edge points → interest points $(\vec{x}, \sigma)$.

5. **Main orientations** (directions of $\nabla u_\sigma$) assigned at each point → interest points $(\vec{x}, \sigma, \theta)$. 

J. Rabin (CNRS, CÉRÉMADE)
Local features

1. Discrete image $u$
2. Scale-space representation
   $\forall \sigma, \quad u_{\sigma} = g_{\sigma} * u$
3. Local extrema $(\vec{x}, \sigma)$ in space and scale of $\sigma^2 \Delta u_{\sigma}$
4. Harris (or Hessian) multiscale criterion to eliminate edge points $\rightarrow$ interest points $(\vec{x}, \sigma)$.
5. Main orientations (directions of $\nabla u_{\sigma}$) assigned at each point $\rightarrow$ interest points $(\vec{x}, \sigma, \theta)$. 
Local features

1. Discrete image $u$
2. Scale-space representation
   \[ \forall \sigma, \ u_\sigma = g_\sigma \ast u \]
3. Local extrema $(\vec{x}, \sigma)$ in space and scale of $\sigma^2 \Delta u_\sigma$
4. Harris (or Hessian) multiscale criterion to eliminate edge points $\rightarrow$ interest points $(\vec{x}, \sigma)$.
5. Main orientations (directions of $\nabla u_\sigma$) assigned at each point $\rightarrow$ interest points $(\vec{x}, \sigma, \theta)$. 
Local features

1. Discrete image \( u \)
2. Scale-space representation
   \[ \forall \sigma, \quad u_\sigma = g_\sigma * u \]
3. Local extrema \((\vec{x}, \sigma)\) in space and scale of \(\sigma^2 \Delta u_\sigma\)
4. Harris (or Hessian) multiscale criterion to eliminate edge points \(\rightarrow\) interest points \((\vec{x}, \sigma)\).
5. Main orientations (directions of \(\nabla u_\sigma\)) assigned at each point \(\rightarrow\) interest points \((\vec{x}, \sigma, \theta)\).
Local features

1. Discrete image $u$

2. Scale-space representation
   \[ \forall \sigma, \quad u_\sigma = g_\sigma * u \]

3. Local extrema $(\vec{x}, \sigma)$ in space and scale of $\sigma^2 \Delta u_\sigma$

4. Harris (or Hessian) multiscale criterion to eliminate edge points \( \rightarrow \) interest points $(\vec{x}, \sigma)$.

5. Main orientations (directions of $\nabla u_\sigma$) assigned at each point \( \rightarrow \) interest points $(\vec{x}, \sigma, \theta)$. 
Local features

1. Discrete image $u$
2. Scale-space representation $\forall \sigma, \ u_\sigma = g_\sigma \ast u$
3. Local extrema $(\vec{x}, \sigma)$ in space and scale of $\sigma^2 \Delta u_\sigma$
4. Harris (or Hessian) multiscale criterion to eliminate edge points $\rightarrow$ interest points $(\vec{x}, \sigma)$.
5. Main orientations (directions of $\nabla u_\sigma$) assigned at each point $\rightarrow$ interest points $(\vec{x}, \sigma, \theta)$. 
Local features: example of SIFT descriptors [Lowe, 1999]

Construction of a **local descriptor** $a$ at each interest point $(\vec{x}, \sigma, \theta)$.

**Mask** (e.g. a square, a disk) around $\vec{x}$:
- $M$ sectors,
- size proportional to $\sigma$.
- orientation given by $\theta$

**Descriptor** $a = (a_1, \ldots a_M)$

$a_m$ = normalized histogram of the gradient orientation $(*)$, weighted by the gradient magnitude, in the $m^{th}$ sector.

$(*)$ Orientations defined with respect to the reference direction $\theta$. 

---

J. Rabin (CNRS, CÉRÉMADE)

Robust multiple object recognition

7 Mai 2010 8 / 30
Part II

A Contrario detection theory
Gestalt theory and the Helmholtz Principle
Gestalt theory and the Helmholtz Principle
Basic principles of *a contrario* methodology [Desolneux, Moisan and Morel, 00]:

- Detect groups of features that are very unlikely under the hypothesis that features are independent (*null hypothesis*)

- Unlikeness ensured by controlling the expected number of false alarms.

Many applications since [Desolneux, Moisan, Morel, 00].

Number of False Alarms and $\varepsilon$-meaningful events

$$\text{NFA}(X) = N_{\text{test}} \cdot P_{\mathcal{H}_0}(\chi \geq X) \leq \varepsilon$$
Part III

Local features matching

1. Local Representation
2. Features Comparison
3. Features Matching
4. Features Grouping
Matching framework with SIFT-like descriptors

**Query descriptors**: \( a^1, \ldots a^{N_Q} \).
- each descriptor \( a^i \) is made of \( M \) histograms, \( a^i = (a^i_1, \ldots, a^i_M) \)

**Candidate descriptors** (database): \( b^1, \ldots b^{N_C} \).

**Distance** \( D \) between descriptors. We assume that

\[
D(a^i, b^j) = \sum_{m=1}^{M} d(a^i_m, b^j_m)
\]

(not restrictive: \( d \) can be any distance between histograms, Euclidean, Manhattan, CEMD, etc.).

**Problem**: Among all \( N_Q \times N_C \) possible matches, which ones are relevant?  
Need of a decision criterion \( \leftrightarrow \) thresholds on \( D \).
Classical criteria

**Usual criteria used to validate matches:**

1. **DT**: Global (unique) threshold
2. **NN-DT**: Global threshold + restriction to the nearest neighbor (avoids many false detections); [Baumberg 00], [Jia and Tang 08]
3. **NN-DR** [Lowe, 00]: Global threshold on the ratio between the distances to the two closest neighbors. [Chum and Matas 05], [Perona and Moreels 07], [Snavely et al. 08]

**Limitations:**

1. Choice of the threshold
2. **DT**: Impossible to define an optimal threshold.
3. **NN-criteria**: No multiple detections.
4. **NN-DR**: Not robust to self similarity and multiple occurrences [Zhang and Kosecka 06], [Noury, Sur and Berger 10]
Objectives

- Multiple correspondences (non nearest neighbor restriction);
- Adaptive thresholds on dissimilarity measure, depending on each query descriptor and on the database.
Let \( \mathbf{b} \) be a **random descriptor**, and \( a^i \) any query descriptor.

**Definition (Null hypothesis)**

\[ H_0^i : \text{"the distances } d(a^i_m, b_m) (m \in \{1, \ldots M\}) \text{ are mutually independent random variables."} \]

Then

\[
\mathbb{P} \left( D(a^i, \mathbf{b}) \leq \delta \mid H_0^i \right) = \int_{-\infty}^{\delta} \prod_{m=1}^{M} p^i_m(x) \, dx,
\]

where \( p^i_m \) is the pdf of the random variable \( d(a^i_m, b_m) \).

**In practice**: for every \( m \in \{1, \ldots M\} \), \( p^i_m \) is **empirically** estimated over the database \( \{b^1, \ldots b^{N_C}\} \),

\[
p^i_m(x) = \frac{1}{N_C} \# \left\{ \mathbf{b}^i ; d(a^i_m, \mathbf{b}^i_m) = x \right\}.
\]

\[ \Rightarrow \text{for each } i \text{ and each value of } \delta, \text{ one can estimate } \mathbb{P} \left( D(a^i, \mathbf{b}) \leq \delta \mid H_0^i \right). \]
Local features matching

A Contrario Criterion

Meaningful matches

Definition:

\[ \text{NFA}(a^i, \delta) := N_Q N_C \mathbb{P}(D(a^i, b) \leq \delta | \mathcal{H}_0^i) \]

Adaptive threshold computation:

\[ \tilde{\delta}_i(\varepsilon) = \arg \max_{\delta} \left\{ \text{NFA}(a^i, \delta) \leq \varepsilon \right\} \]

A Contrario (AC) Matching criterion: A match between \( a^i \) and \( b^j \) is

- **validated** if \( \text{NFA}(a^i, D(a^i, b^j)) \leq \varepsilon \iff D(a^i, b^j) \leq \tilde{\delta}_i(\varepsilon) \),
- **rejected** if \( \text{NFA}(a^i, D(a^i, b^j)) > \varepsilon \iff D(a^i, b^j) > \tilde{\delta}_i(\varepsilon) \).

Control of false alarms:

Proposition

Under the null hypothesis, the expected number of matches among the \( N_C \times N_Q \) possible matches is smaller than \( \varepsilon \).
Can experiment

Small dataset illustration
Can experiment

**NN-DT matching criterion**
Can experiment

**NN-DR matching criterion**
Can experiment

AC matching criterion
Part IV

Local features grouping and model selection

- Local Representation
- Features Comparison
- Features Matching
- Features Grouping
How to group matches?

Direct estimation (Least Square estimators) from selected correspondences is not robust, due to irrelevant data (outliers).

Two different alternatives:

1. Grouping matches in parameter space: **Hough transform** [Hough 1959]
   - ✔ Multiple group detection
   - ✗ Time complexity, precision

2. Grouping matches by geometric consensus: **RANSAC** algorithm [Fischler and Bolles 1981]
   - ✔ Fast
   - ✗ Single group detection, parameters tuning

▷ How to use RANSAC to detect several transformations?
How to group matches?

Direct estimation (Least Square estimators) from selected correspondences is not robust, due to irrelevant data (outliers).

Two different alternatives:

1. Grouping matches in parameter space: Hough transform [Hough 1959]
   - ✓ Multiple group detection
   - ✗ Time complexity, precision

2. Grouping matches by geometric consensus: RANSAC algorithm [Fischler and Bolles 1981]
   - ✓ Fast
   - ✗ Single group detection, parameters tuning

▷ How to use RANSAC to detect several transformations?
How to group matches?

Direct estimation (Least Square estimators) from selected correspondences is not robust, due to irrelevant data (outliers).

Two different alternatives:

1. Grouping matches in parameter space: **Hough transform** [Hough 1959]
   - ✓ Multiple group detection
   - ✗ Time complexity, precision

2. Grouping matches by geometric consensus: **RANSAC algorithm** [Fischler and Bolles 1981]
   - ✓ Fast
   - ✗ Single group detection, parameters tuning

How to use RANSAC to detect several transformations?
How to group matches?

Direct estimation (Least Square estimators) from selected correspondences is not robust, due to irrelevant data (outliers).

Two different alternatives:

1. Grouping matches in parameter space: Hough transform [Hough 1959]
   - ✔ Multiple group detection
   - ❌ Time complexity, precision

2. Grouping matches by geometric consensus: RANSAC algorithm [Fischler and Bolles 1981]
   - ✔ Fast
   - ❌ Single group detection, parameters tuning

▷ How to use RANSAC to detect several transformations?
RANSAC methodology

Let \( \{(m_i, m'_i)\}_{i=1,...,N} \) be matches between images.

Residuals \( r_i \) are defined for a given transformation \( T \) from geometrical errors
\[
r_i = \max\{d(m_i, T^{-1}m'_i), d(Tm_i, m'_i)\}.
\]

Fix a residual threshold \( r_{\text{max}} \) and a consensus threshold \( N_C \)

Then, iterate

- Draw \( n \) correspondences subset \( S' \) and infer transformation \( T_{S'} \);
- Let \( k \) be the number of pairs for which \( r_i \leq r_{\text{max}} \) (group of correspondences \( S \));
- \( T_{S'} \) is the best transformation if \( S \) is the biggest group found so far.

\( S \) is validated if \( \#S \geq N_C \)
RANSAC extensions

**Previous works:** MSAC, MLESAC [Torr and Zisserman 00], MINPRAN [Stewart 95], Coarse to fine RANSAC [Torr and Davidson, 03], PROSAC [Chum and Matas, 05], Multi-RANSAC [Zuliani et al. 05], A Contrario RANSAC without prior match [Noury, Sur and Berger, 07], J-linkage [Toldo and Fusiello, 08], Optimal Randomized Ransac [Chum and Matas, 08]

**Solution studied:**

* A Contrario RANSAC [Moisan, Stival 2004]

**Considered transformations:** similarity, affine transforms, projective transforms, epipolar geometry.
A contrario RANSAC (collaboration with L. Moisan)

Same sampling strategy as RANSAC, except that the quality measure of a group $S$ is defined by its NFA [Moisan, Stival 2004].

**Null hypothesis**: points $m_i$ and $m'_i$ are mutually independant and uniformly distributed.

**Rigidity** of group $S$ (cardinality $k$) according to a transformation $T_{S'}$:

$$\alpha(S, T_{S'}) = \max_{i=1,\ldots,k} \{ \hat{r}_i \}.$$  

where $\hat{r}_i$ represents the normalized residual error:

$$\hat{r}_i = \max \left\{ \frac{\pi}{A} d(m_i, T_{S'}^{-1} m'_i)^2, \frac{\pi}{A'} d(T_{S'} m_i, m'_i)^2 \right\},$$  

and where $A$ and $A'$ are the area of image $\mathcal{I}$ and $\mathcal{I'}$.

**Probability**: the probability of observing a group $S$ of $k$ points with rigidity $\alpha(S, T_{S'})$ under $\mathcal{H}_0$ is defined as

$$\mathbb{P}_{\mathcal{H}_0}(S) \leq \alpha(S, T_{S'})^k.$$
A contrario RANSAC

Quality measure of group $S$ with rigidity:

$$NFA(S, S') := (N - n) \binom{N}{k} \binom{N - k}{n} \alpha(S, T_{S'})^k,$$

with

- $N$ : the total number of correspondences,
- $k$ : the cardinality of the set $S$,
- $n = 2$ for similarity, $n = 3$ for affine transforms and $n = 4$ for homographies.

**Note**: a similar expression is obtained for epipolar geometry [Moisan, Stival 2004].

**Detection Decision**: S is validated if $NFA(S, S') \leq 1$.

**Note**: Some pre-processing stages are required to make sure that the null hypothesis is validated for random data.
RANSAC is devoted to the detection of only one group.

\( \triangleright \) Idea: Iteratively run A Contrario RANSAC while meaningful groups are found.

Advantages of a contrario framework for sequential RANSAC:
- Very robust stopping criterion;
- Automatic parameter selection;
- No over-segmentation;

... but still limitations!
- Fusion of several transformations
- Repeated detections of the same object
RANSAC is devoted to the detection of only one group.

▷ Idea: Iteratively run A Contrario RANSAC while meaningful groups are found

Advantages of a contrario framework for sequential RANSAC:

– Very robust stopping criterion;
– Automatic parameter selection;
– No over-segmentation;

... but still limitations!

– fusion of several transformations
– repeated detections of the same object
RANSAC is devoted to the detection of only one group.

▷ Idea: Iteratively run A Contrario RANSAC while meaningful groups are found

Advantages of a contrario framework for sequential RANSAC:

– Very robust stopping criterion;
– Automatic parameter selection;
– No over-segmentation;

... but still limitations!

– fusion of several transformations
– repeated detections of the same object
RANSAC is devoted to the detection of only one group.

▷ Idea: Iteratively run A Contrario RANSAC while meaningful groups are found

Advantages of a contrario framework for sequential RANSAC:
- Very robust stopping criterion;
- Automatic parameter selection;
- No over-segmentation;

... but still limitations!
- fusion of several transformations
- repeated detections of the same object
RANSAC is devoted to the detection of only one group.

▷ Idea: Iteratively run A Contrario RANSAC while meaningful groups are found

Advantages of a contrario framework for sequential RANSAC:

– Very robust stopping criterion;
– Automatic parameter selection;
– No over-segmentation;

... but still limitations!

– fusion of several transformations
– repeated detections of the same object
Some more experiments

Static scene with camera motion
With planar homography and epipolar geometry
Static scene with camera motion
With planar homography and epipolar geometry
Multiple occurrences of the same object
Multiple grouping with planar transformations
A complete and generic object recognition system based on a contrario methodology combined with local feature approach:

- Automatic detection thresholds definition without prior on the data
- Multiple object recognition algorithm
Merci de votre attention !
Bibliographie
Appendix
Fusion of transformations

A problem intrinsic to RANSAC [Stewart 1995, 1997]

Synthetic image pair example:

AC-RANSAC detects only one group $S_0$ with homography

Both "true" groups $S_1$ and $S_2$ are such that:

$$
\begin{align*}
\text{NFA}(S_0, S'_0) &< \min \{ \text{NFA}(S_1, S'_1), \text{NFA}(S_2, S'_2) \} \\
\alpha(S_0, T_{S'_0}) &\geq \max \{ \alpha(S_1, T_{S'_1}), \alpha(S_2, T_{S'_2}) \}
\end{align*}
$$
Appendix

Split algorithm review

The group $S_0$ is potentially divided into two groups:

- First group $S_1$ is detected as a minimum of NFA s.t. its cardinality is smaller than the half of $S_0$.
- Second group $S_2$ is composed of the remaining points.

The splitting search is iterated on the smallest group found, while

$$\text{NFA}(S_1, S'_1) \times \text{NFA}(S_2, S'_2) < \text{NFA}(S_0, S'_0).$$
Recursive split search

\[ S_0 \]

\[ k = 1 \]

\[ S_1^1 \quad \& \quad S_2^1 \]

\[ k = 2 \]

\[ S_1^2 \quad \& \quad S_2^2 \]
Folded card

Semi-rigid deformation
Folded card

With and without splitting criterion
Repeated detections

Repeated detection of the same object, or part of the same object (self similarity) [Lindenbaum 1997, Monasse 2000].

Several registrations are detected, but only one is correct.
Spatial filtering

The spatial filter should preserve **multiple** and **overlapping** objects recognition.
Spatial filtering

**Spatial exclusion principle:** for each group of validated points, a binary mask is constructed to eliminate remaining correspondences.

For each selected feature, a disk proportional to its **characteristic scale** is used.
Spatial filtering