A Contrario Cluster Detection and Application to Planar Shape Recognition

P. Musé F. Sur F. Cao J.M. Morel

IPAM 2005

fcao@irisa.fr

A. Almansa, J. Delon, A. Desolneux, Y. Gousseau, J.L. Lisani, L. Moisan

Supported by CNRS, INRIA, CNES, ONR, Ministère de la Recherche

What are Shapes?

- Gestaltists:
  shapes are the result of a grouping process.
- “Naive” point of view:
  shapes are objects in images that can be recognized.

A more simple question:

Do two given images have shapes in common?

Shape recognition $\neq$ generic object recognition or categorization.

- No *a priori* (precise) model.
- Automatic decision for detection and recognition.

A Decision Tool

A contrario detection

- No a priori model, but a statistical *background model*, assuming a partial independence of the data.
- Relevant events are those with a very small probability of occurence in the background model.
- Threshold the Number of False Alarms (NFA)=expected number of occurences of a class of events in the background model.
- Decision weakly depends on the threshold (log).
- Able to automatically compute a rejection threshold.
- Has been used for the detection of
    - alignments
    - modes in histograms
    - contrasted edges
    - vanishing points

[Desolneux, Moisan, Morel 00, Almansa *et al.* 04, Musé Sur *et al.* 03]

A Subproblem in Clustering

A particular application of a contrario detection

1. A set of points being given, do they constitute a relevant group? (cluster validity)
2. Are two small groups better than a single big one? (merging criterion)

Particularly adapted to hierarchical clustering method.

What we are not interested in:
designing a new clustering method.

[Duda Hart 73, Bock 85, Dubes 87, Gordon 99]

Perception of Shapes

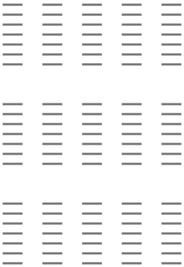
Gestalt Theory. Laws determining perception are grouping laws:

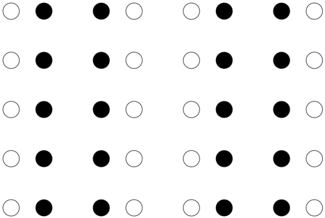
- alignment
- good continuation
- proximity
- similarity
- color
- symmetry
- convexity
- width constancy
- past experience
- ...

about 20 elementary laws (*partial gestalts*) interacting recursively.

[Wertheimer 23, Köhler 35, Kanizsa 79, Kandinsky 17, Desolneux, Moisan, Morel 04]

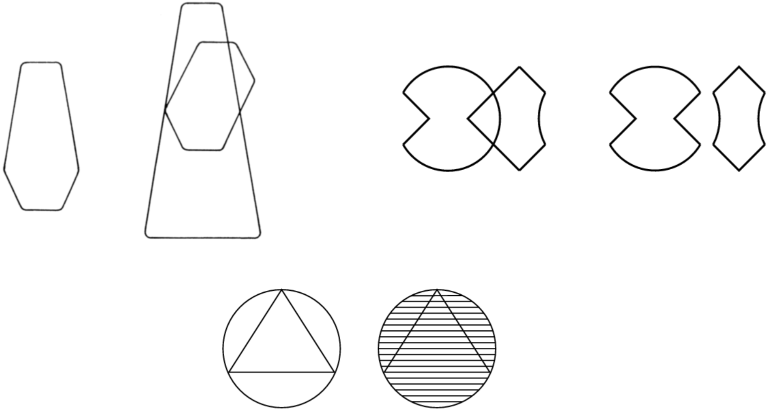
Examples





Masking Phenomenon

Partial gestalts may conflict with each other.



From Kaniza, *Vedere e Pensare*

Computational Gestalt Theory

A partial program (no masking phenomena)

- Form groups of data points in $\mathbb{R}^d$, with respect to some common partial gestalt.
- $d \leq 20$ (one dim./partial gestalt).
- data are sparse (between 100 and 1000 datapoints).
- Many outliers.
- no precise prior model.
- fully unsupervised decision.
- number of groups is unknown (can be zero).

A Detection Principle

Helmholtz Principle. An observation is meaningful when the probability that it occurs by chance is very small.

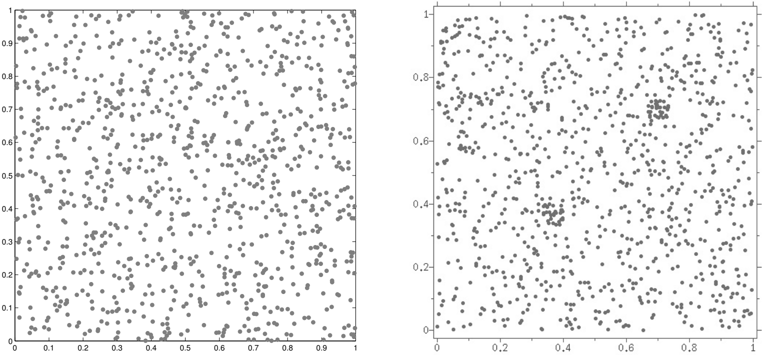
(conspiracy of randomness)

What is

- a *small* probability?
- *chance*?

[Helmholtz 1867, Attneave 54, Lowe 85, Grimson/Huttenlocher 91, Lindenbaum 96, Watson&Watson 96, Olson Huttenlocher 97, Pennec 98, DMM 00, Chapple *et al.* 01]

A Contrario Detection



Left: realization of “chance”=1000 i.i.d points, uniform in $[0, 1]^2$. Right: 1000 data points.

A Contrario Decision

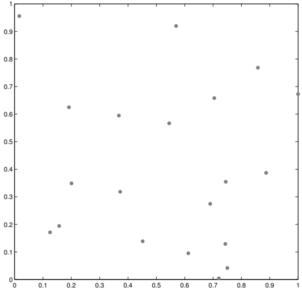
Assume that the dataset S contains M data points in $\mathbb{R}^d$, i.d. following a known distribution π.
Let $R \in \mathbb{R}^d$. Assume *a contrario* that the datapoints are independent. Then

$$\begin{aligned} P(|S \cap R| \geq k) &= \mathcal{B}(M, k, \pi(R)). \\ &= \sum_{j=k}^{M} \binom{M}{j} \pi(R)^j (1 - \pi(R))^{M-j}. \end{aligned}$$

Very small if

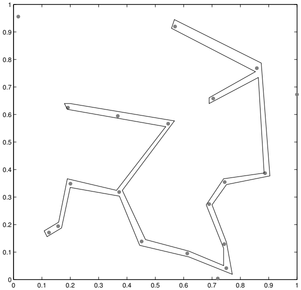
- $k \ggg M \cdot \pi(R)$.
- $\pi(R)$ is small.

A Contrario Decision



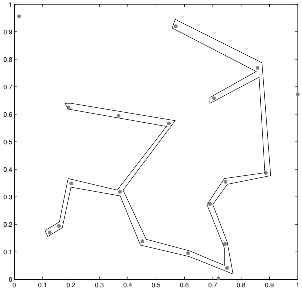
20 points i.i.d., uniform in $[0, 1]^2$

A Contrario Decision



A small set containing most of data points.

A Contrario Decision



A small set containing most of data points.

- A set of *a priori* test regions must be given, independently of the data.

Validity of a Region

Let $\mathcal{R}$ be a *finite* (but large) set of regions containing the origin. (*e.g.* hyperrectangles with different sidelengths.)

Definition

Consider a region $R = x_i + R'$, $R' \in \mathcal{R}$. Let $k = |R \cap S|$. We call number of false alarms

$$NFA(R) = M \cdot |\mathcal{R}| \cdot \mathcal{B}(M-1, k-1, \pi(R)).$$

We say that R is ε-meaningful if $NFA(R) \leq \varepsilon$.

Proposition

Assume that $(X_1, \dots, X_M)$ are i.i.d. following π. Then the expected number of ε-meaningful regions is less than ε.

Comparison with Classical Hypothesis Testing

- $\mathcal{H}_0$: data points are independent

$$P(|S \cap R| \geq k | \mathcal{H}_0) = \mathcal{B}(M-1, k-1, \pi(R)).$$

- $\mathcal{H}_1$: data points form a group and are not independent

$$P(|S \cap R| \geq k | \mathcal{H}_1)???$$

- Empirical evaluation of π is sound and does not need to be very accurate.

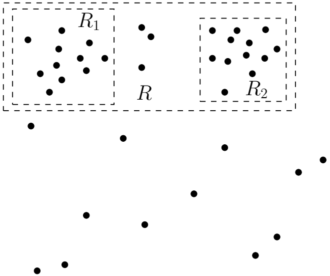
Comparison with Classical Hypothesis Testing (2)

- Classical tests (Bayes, Neyman-Pearson, min-max) are based on the likelihood ratio
$$R(k) = \frac{P(|S \cap R| \geq k|\mathcal{H}_0)}{P(|S \cap R| \geq k|\mathcal{H}_1)},$$
but we don't know how to compute the denominator!
- A contrario hypothesis allows to reject $\mathcal{H}_0$ surely (up to ε false detection).
- The threshold on the probability depends on the number of data points.
- Detection only based on the probability of false alarm.
- No ROC curves.
- Because of the independence hypothesis detection thresholds depend on $\log \frac{\varepsilon}{M|\mathcal{R}|}$.
- Automatic thresholds since in practice, we can always take $\varepsilon = 1$.

Merging Criterion (1)

Let R_1 and R_2, $R_1 = x_1 + R_1'$, $R_2 = x_2 + R_2'$, with R_1' and R_2' in $\mathcal{R}$.

- either R_1 and R_2 must remain separate regions.
- or R_1 and R_2 should be embedded in a single region R.



Merging Criterion (2)

For $(i,j) \in \{(1,2),(2,1)\}$, let

- $k_i = |S \cap (R_i \backslash R_j)|$,
- $\pi_i = \pi(R_i \backslash R_j)$.

Definition

$$NFA_{gg}(R_1, R_2) = \frac{M^3 |\mathcal{R}|^2}{2} \mathcal{M}(M-2, k_1 - 1, k_2 - 1, \pi_1, \pi_2).$$

(tail of trinomial law)
The pair of regions (R_1, R_2) is ε-meaningful if $NFA_{gg}(R_1, R_2) \leq \varepsilon$.

Proposition

If the M data points are i.i.d. following π, the expected number of ε-meaningful pairs of regions is less than ε.

Indivisibility

Definition

Let R be a region containing the points of R_1 and R_2. We say that R is indivisible with respect to (R_1, R_2) iff

$$NFA(R) \leq NFA_{gg}(R_1, R_2).$$

Algorithmic Considerations

- Construct the minimal spanning tree of the data points.
- For each node G (candidate group), call $R(G)$ the region $x + R'$, $x \in G$, $R' \in \mathcal{R}$, containing all the points of the node and with minimal NFA.
- Compute the NFA of all nodes.
- For all node G with children G_1, G_2, compute $NFA_{gg}(R(G_1), R(G_2))$ and compare it to the NFA of G.

Maximality

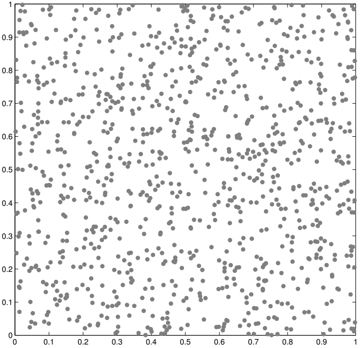
Definition

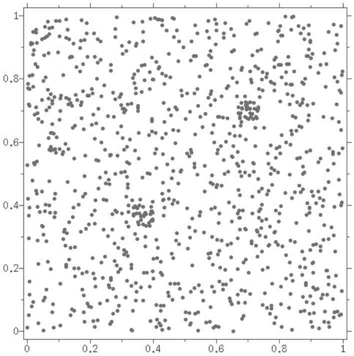
A node region $R = R(G)$ is maximal ε-meaningful if and only if

1. $NFA_g(R) \leq \varepsilon$,
2. R is indivisible in $\mathcal{R}$,
3. for all indivisible descent R', $NFA_g(R') \geq NFA_g(R)$,
4. for all indivisible ascent R', either $NFA_g(R') > NFA_g(R)$ or there exists an indivisible descent R'' of R' such that $NFA_g(R'') < NFA_g(R')$.

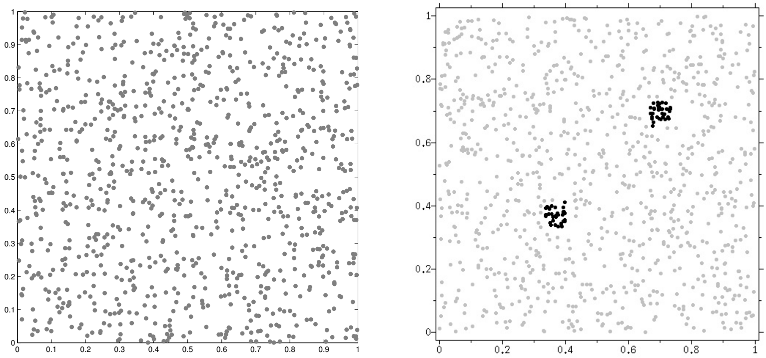
We say that G is a maximal ε-meaningful if $R(G)$ is.

Experiments



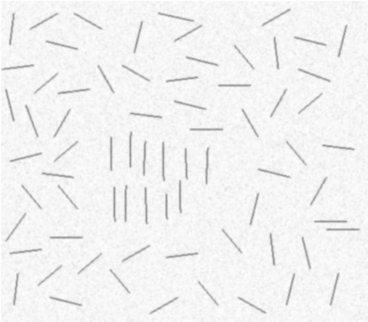


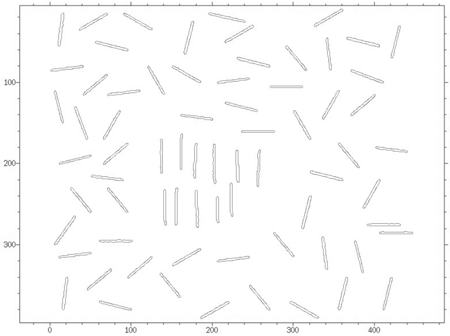
Experiments



2 maximal meaningful groups, $NFA = 10^{-11}$ and $NFA = 10^{-8}$.

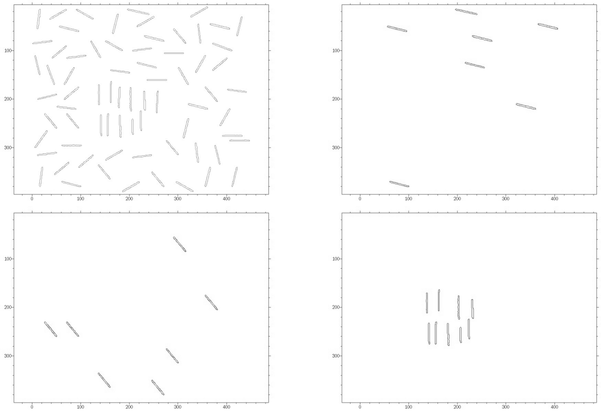
Experiments



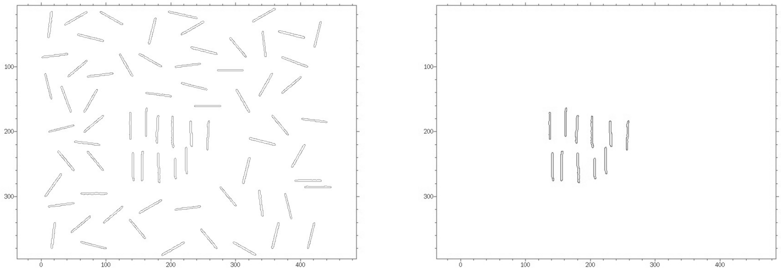


A grey level image and its ε-meaningful level lines.

Experiments

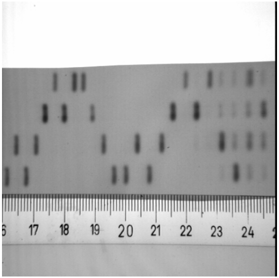


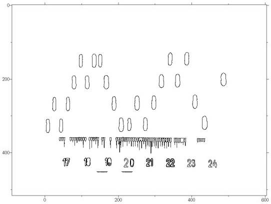
Three maximal meaningful groups with respect to orientation. (smallest $NFA = 10^{-2}$.)

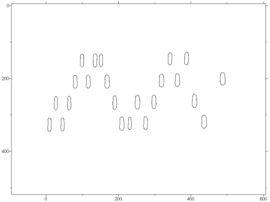


Only one maximal meaningful group with respect to orientation and position.
($NFA = 10^{-4}$.)

Experiments







A level grey image, its meaningful level lines. The only meaningful group, with respect to diameter, elongation, orientation. ($NFA = 10^{-88}$.)

Planar Shape Identification

Given two images, do they have shapes in common?

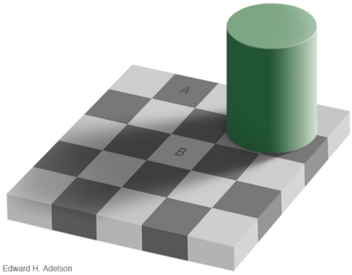
- No a priori information on the images (what is shape?)
- Do not assume *a priori* a positive answer.

Psychophysical requirements for shape recognition

- Compactness of information.
- Affine (projective) invariance.
- Robustness to occlusions.
- Invariance to contrast change.

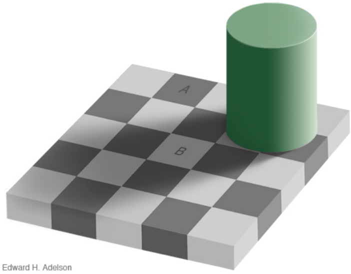
[Attneave 54, Kanizsa 79]

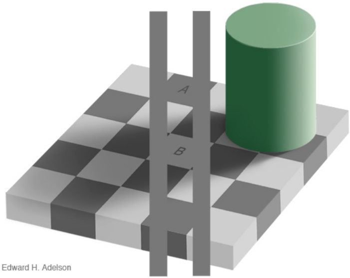
Grey Level is not a Reliable Perceptual Information



From E.H. Adelson

Grey Level is not a Reliable Perceptual Information





From E.H. Adelson

Extracting Shape Elements from Images

Definition (and Theorem)

Topographic map of a grey level image u: set of its level lines, $\partial\{u \geq \lambda\}$.
For a.e. level, level lines are Jordan curves.
The topographic map is a complete, contrast invariant representation of images.

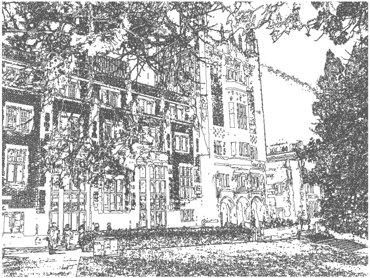
Original image

[Matheron 75, Morel-Guichard 95, Monasse 00]

Extracting Shape Elements from Images

Definition (and Theorem)

Topographic map of a grey level image u: set of its level lines, $\partial\{u \geq \lambda\}$.
For a.e. level, level lines are Jordan curves.
The topographic map is a complete, contrast invariant representation of images.



Level lines (multiple of 40)

[Matheron 75, Morel-Guichard 95, Monasse 00]

Meaningful level lines

Meaningful level lines = $P(|Du|$ is large everywhere$)$ small.
(A contrario detection)

An 1024×768 image encoded on 256 grey levels, 1,055,684 level lines.

[Desolneux *et al.* 01]

Meaningful level lines

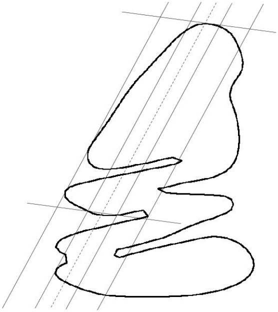
Meaningful level lines = $P(|Du|$ is large everywhere$)$ small.
(A contrario detection)

2398 meaningful level lines

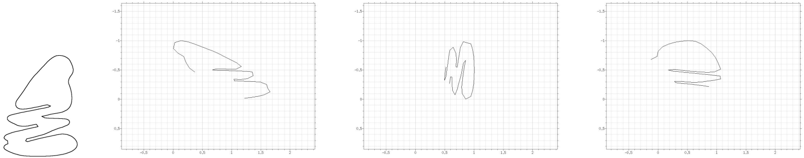
[Desolneux *et al.* 01]

Invariant local encoding

Compute local invariant frames on each level lines.



$\longrightarrow$ local, invariant, redundant shape elements.



Recognition and Grouping of Shape Elements

Given two images u_1 and u_2,

- Compute meaningful level lines.
- Smooth level lines by affine scale space (limit the number of shape elements).
- Compute all the shape elements in both images (N_1, N_2 resp.)
- Match shape elements by an *a contrario* decision rule: s' matches s iff

$$N_1 \cdot N_2 \cdot P(d(s, S') < d(s, s')) < \varepsilon$$

- Each match uniquely determines an affine transformation, i.e. a point in $\mathbb{R}^6$.
- Group these affine transformations by an a contrario decision.

[Sapiro Tannenbaum 93, Alvarez Guichard Lions Morel 93, Musé, Sur *et al* 03]

Clustering of Affine Transformations

- Rigidity hypothesis: matching shape elements of a single shape defines the same transformation.
- Define a dissimilarity between affine transformation, and apply the a contrario rules.
- Similar to Hough Transform except that
  - No a priori quantization of the space of transformations.
  - can compute a rejection/acception threshold.

[Ballard 81, Grimson-Huttenlocher 91]

Experiments (0)

Experiments (1)

Experiments (1)

First maximal group: 5 matches, $-\log(NFA) = 6.85$.

Experiments (1)

Registration w.r.t. first group; pieces of level lines of length 40 that are at distance less than 4.

Experiments (1)

Second maximal group: 7 matches, $-\log(NFA) = 9.27$.

Experiments (1)

Registration w.r.t. second group

Experiments (1)

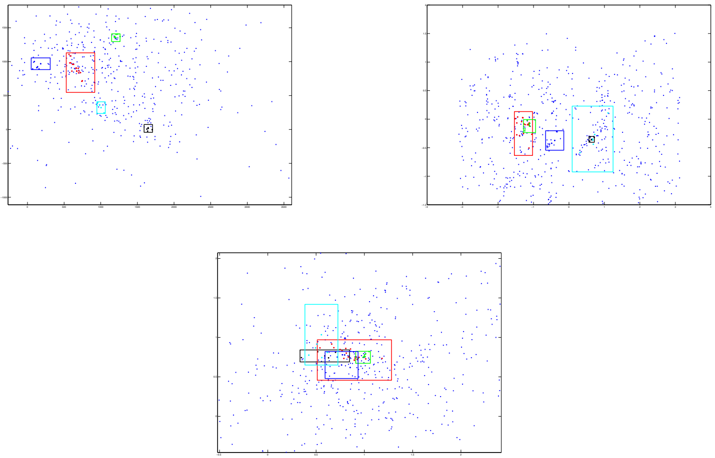
A non maximal group: $-\log(NFA) = 3.85$ (11 meaningful non maximal groups)

Experiments (2)

Experiments (2)

All 633 matches between shape elements.

Experiments (2)



Projection of transformations in left: translations, middle: rotation/shear, right: zooms

Experiments (2)

5 matches, $-\log NFA = 10.2$

Experiments (2)

Experiments (2)

18 matches, $-\log NFA = 26.2$

Experiments (2)

Experiments (2)

7 matches, $-\log NFA = 18.4$

Experiments (2)

Experiments (2)

6 matches, $-\log NFA = 5.7$

Experiments (2)

Experiments (2)

6 matches, $-\log NFA = 1.9$

Experiments (2)

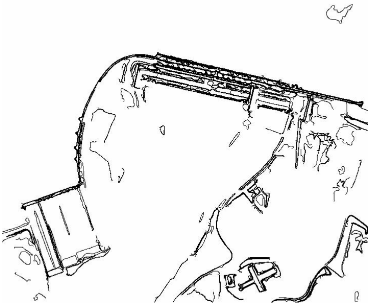
Experiments (3)

Itaipu Dam (original proportions)

Experiments (3)

Itaipu Dam (original proportions)

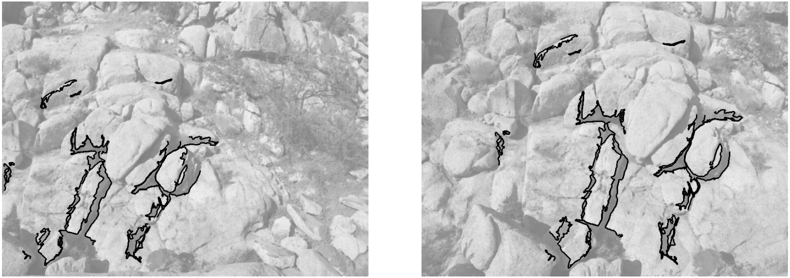
Experiments (3)

One single group, 16 matches, $-\log NFA = 28.3$.
Comparison of pieces of level lines after registration.

Experiments (4)

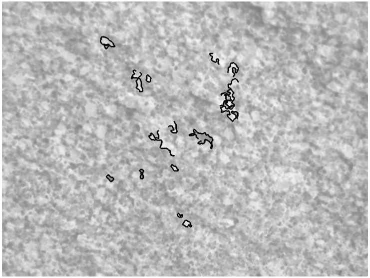
Experiments (4)

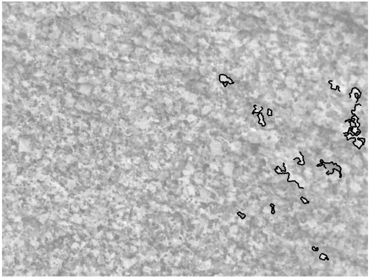


A single group, 38 matches, $-\log NFA = 68.03$.

Experiments (5)

Experiments (5)





A single group, 20 matches, $-\log NFA = 57.96$.

Conclusions and Perspectives

- Shapes exist! We can recognize them.
- A contrario decision framework is generic.
- Application to matching and grouping of shape elements
  - Make the algorithms faster (not brute force)
  - Apply the a contrario decision to other types of features (SIFT).
- More general deformations?
- Relation to machine learning?
- Book in progress (Musé, Sur, Cao, Lisani, Morel).

[Lowe 04, Trouvé *et al.* 05, Mémoli Sapiro 05]