Automatic image analysis : a challenge for Computer Vision

Agnès DESOLNEUX, Lionel MOISAN, Jean-Michel MOREL,

Centre de Mathématiques et Leurs Applications
Ecole Normale Supérieure de Cachan, 61 av. du Président Wilson,
94235 Cachan Cedex, France
desolneu@cmla.ens-cachan.fr, moisan@cmla.ens-cachan.fr
morel@cmla.ens-cachan.fr

Résumé – Nous exposons une méthode récente pour trouver des structures géométriques arbitraires dans une image, sans information ou modèle a priori. Par un principe de la perception dû à Helmholtz, une structure géométrique est perceptuellement significative si l'espérance de son nombre d'apparitions dans une image aléatoire est très petite. Les structures géométriques sont donc caractérisées comme de grandes déviations par rapport au bruit. Ce principe permet de définir et calculer des structures telles qu'alignements, bords, groupes dans une image, par une méthode libre de paramètres. On définira les structures géométriques "maximales significatives" et on les comparera avec les mêmes structures calculées par des algorithmes plus classiques.

Abstract – We expose a recently introduced method for computing geometric structures in a digital image, without any a priori information. According to a basic principle of perception due to Helmholtz, an observed geometric structure is perceptually "meaningful" if its number of occurences would be very small in a random situation: geometric structures are characterized as large deviations from randomness. This leads us to define and compute alignments, edges, and clusters in an image by a parameter-free method. Maximal meaningful objects are defined, computed, and the results compared with the ones obtained by classical algorithms.

1 Helmholtz principle

In statistical methods for image analysis, one of the main problems is the choice of an adequate prior. For example, in the Bayesian model [6], given an observation "obs", the aim is to find the original "model" by computing the Maximum A Posteriori (MAP) of

$$\text{Prob}(\text{model}|\text{obs}) = \frac{\text{Prob}(\text{obs}|\text{model}) \times \text{Prob}(\text{model})}{\text{Prob}(\text{obs})}.$$

The term Prob(obs|model) represents the degradation (superimposition of a gaussian noise for example) and the term Prob(model) is called the prior. This prior plays the same role as the regularity term in the variational framework. This prior has to be fixed and it is generally difficult to find a good prior for a given class of images. It is also probably impossible to give an all-purpose prior!

In [2], [3] and [7], we have outlined a different statistical approach, based on phenomenological observations coming from Gestalt theory [5]. According to a perception principle which seems to go back to Helmholtz, every large deviation from a "uniform noise" image should be perceptible, provided this large deviation corresponds to an *a priori* fixed list of geometric structures (lines, curves, closed curves, convex sets, spots, local groups,...). Thus, there still is an a priori geometric model, but, instead of being quantitative, this model is merely qualitative. Let us illustrate how this should work for "grouping" black dots in a white sheet. Assume we have a white image with black dots spread out. If some of them form a cluster, say, in the center of the image, then, in order to decide whether this cluster indeed is a group of points, we compute the expectation of this grouping event happening by chance **if** the dots were uniformly distributed in the image. If this expectation happens to be very low, we decide that the group in the center is meaningful. Thus, instead of looking for objects as close as possible to a given prior model, we consider a "wrong" and naive model, actually a random uniform distribution, and then define the "objects" as large deviations from this generic model. One can find in [1] a very close formulation of computer vision problems.

We may call this method Minimal A Posteriori Expectation, where the prior for the image is a uniform random noise model. Indeed, the groups (geometric structures, gestalts) are defined as the best *counterexamples*, i.e. the least expected. Those counterexamples to the uniform noise assumption are taken in a restricted geometric class. Notice that not all such counterexamples are valid: the Gestalt theory fixes a list of perceptually relevant geometric structures which are supposedly looked for in the perception process. The computation of their expectation in the uniform noise model validates their detection: the least expected in the uniform noise model, the more perceptually meaningful they will be.

A main claim in favour of the Minimum a Posteriori is its reduction to a single parameter, the meaningfulness of a geometric event depending only on the difference between the logarithm of the false alarm rate and the logarithm of the image size! We just have to fix this false alarm rate and the dependance of the outcome is anyway a log-dependence on this rate, so that the results are very insensitive to a change.

2 Definitions

2.1 Meaningful groups, or clusters

This first example is the seminal one in Gestalt theory. Assume we see a set of dots on a white sheet and those dots happen to present one or several groups, separated by desert regions. In order to characterize this as an event with very low probability, we shall make all computations with the *a contrario* or *background* model that the dots have been uniformly distributed over the white sheet. This amounts to consider the dots as distributed over the sheet by a Poisson distribution. We then call A the simply connected region, with area σ, containing a given observed cluster of dots and $1-\sigma$ is the normalized area of the sheet. Assume that we observe k points in A and $M-k$ outside. Then the "cluster probability" of observing at least k points among the M inside A is given by

$$P(k,A)=\sum_{i=k}^{M}C_M^i\sigma^i(1-\sigma)^{M-i}. \tag{1}$$

It is easily checked by large deviations estimates that if $\frac{k}{M}$ exceeds significantly σ, this probability can become very small. Now, the event is not a generic event in that we have fixed a posteriori the domain A. The real a priori event we can define is "There is a simply connected domain A, with area σ, containing at least k points". Since the number of such domains A is huge, we see that the expectation of such an event is by no means small. In the following, we shall therefore consider a more restrictive set of domains $\mathcal{D}$ with cardinality $N_{\mathcal{D}}$.

Definition : We say that a group of dots is ε-meaningful if $N_{\mathcal{D}}P(k,A)\le\varepsilon$.

In order to define $\mathcal{D}$ in a realistic way, we have to *sample* the set of simply connected domains by encoding their boundaries as "low resolution" Jordan curves. We consider a low resolution grid in the image, which for a sake of low complexity we take to be hexagonal, with mesh m. The number of curves with length lm starting from a point and supported by the grid is bounded from above by 2^l. The overall number of low resolution curves with length lm is bounded by $N^2 2^l$. Thus, we can consider several resolutions in logarithmic scale $m_1,\dots,m_q$, each one larger than 1, the pixel mesh and the larger one m_q proportional to the image size, so that q is actually small. Our set of domains will be the set of all Jordan curves at all given resolutions and all with discrete length measured in the corresponding mesh less than a fixed length L. Thus, the overall number of possible low resolution curves is simply N^2q2^L. Notice that all numbers here are relatively small since we shall never allow for a very intrincated cluster. Thus, L will always be smaller than, say, 20 in practice. We therefore define a meaningful cluster as a set of points contained in a low resolution curve defined as above, and such that $N^2q2^LP(k,A)\le\varepsilon$. The experiment below is based on this definition. It can also happen that a cluster is not overcrowded, but only fairly isolated from the other dots. In such a case, we can find a low resolution curve surrounding the cluster and such that some dilate of the curve does not contain any point. Accordingly, we can modify the probability of the event : *this cluster contained in A with area σ is surrounded by an empty thick curve C with area σ'*. In such a case the definition of ε-meaningfulness for an isolated cluster becomes

$$N^2q2^L\sum_{i=k}^{M}C_M^i\sigma^i(1-\sigma-\sigma')^{M-i}\le\varepsilon. \tag{2}$$

Our first experiment displays such clusters.

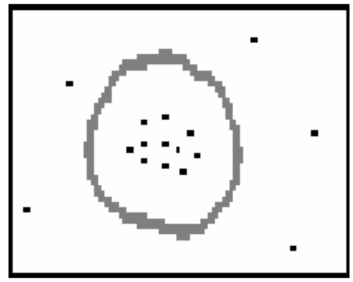

FIG. 1: *Meaningful isolated cluster, surrounded by an empty low resoltion curve.*

2.2 Meaningful boundaries

Let u be a discrete image, of size $N\times N$. We consider the level lines at quantized levels $\lambda_1,\dots,\lambda_k$. The quantization step q is chosen in such a way that level lines make a dense covering of the image: if e.g. this quantization step q is 1 and the natural image ranges 0 to 256, we get such a dense covering of the image. A level line can be computed as a Jordan curve contained in the boundary of a level set with level λ,

$$\chi_\lambda=\{x/u(x)\le\lambda\}\quad\text{and}\quad\chi^\lambda=\{x/u(x)\ge\lambda\}.$$

Notice that along a level line, the gradient of the image must be everywhere above zero. Otherwise the level line contains a critical point of the image and is highly dependent upon the image interpolation method. Thus, we consider in the following only level lines along which the gradient is not zero.

Let L be a level line of the image u. We denote by l its length counted in independent points. In the following, we will consider that points at a geodesic distance (along the curve) larger than 2 are independent (i.e. the contrast at these points are independent random variables). Let x_1, x_2,...x_l denote the l considered points of L. For a point $x\in L$, we will denote by $c(x)$ the contrast at x. It is defined by

$$c(x)=|\nabla u|(x), \tag{3}$$

where ∇u is computed by a standard finite difference on a 2×2 neighborhood [2]. For $\mu\in\mathbf{R}_+^*$, we consider the event: for all $1\le i\le l$, $c(x_i)\ge\mu$, i.e. each point of L has a contrast larger than μ. From now on, all computations are performed in the Helmholtz framework explained in the introduction: we make all computations as though the contrast observations at x_i were mutually independent. Since the l points are independent, the probability of this event is $\mathrm{Prob}(c(x_1)\ge\mu)\times\mathrm{Prob}(c(x_2)\ge\mu)\times$

... $\times$ Prob$(c(x_l) \geq \mu) = H(\mu)^l$, where $H(\mu)$ is the probability for a point on any level line to have a contrast larger than μ. An important question here is the choice of $H(\mu)$. Shall we consider that $H(\mu)$ is given by an *a priori* probability distribution, or is it given by the image itself (i.e. by the histogram of gradient norm in the image)? In the case of alignments, we shall take by Helmholtz principle the orientation at each point of the image to be a random, uniformly distributed variable on $[0, 2\pi]$. Here, in the case of contrast, it does not seem sound at all to consider that the contrast is uniformly distributed. In the following, we will consider that $H(\mu)$ is given by the image itself, which means that

$$H(\mu) = \frac{1}{M}\#\{x \,/\, |\nabla u|(x) \geq \mu\}. \tag{4}$$

where M is the number of pixels of the image where $\nabla u \neq 0$. In order to define a meaningful event, we have to compute the expectation of the number of occurrences of this event in the observed image. Thus, we first define the number of false alarms.

Definition :[Number of false alarms] Let L be a level line with length l, counted in independent points. Let μ be the minimal contrast of the points x_1,..., x_l of L. The number of false alarms of this event is defined by

$$NF(L) = N_{ll} \times [H(\mu)]^l, \tag{5}$$

where N_{ll} is the number of level lines in the image. We call interval of level lines a set of level lines such that each one is enclosed in only one, and contains only another one. In such a case, we shall only display in the experiments, the "maximal meaningful" level line of the interval, i.e. the one for which NF is minimal over the interval. This is can be compared to a global Canny filter.

Notice that the number N_{ll} of level lines is provided by the image itself. We now define ε-meaningful level lines.

Definition :[ε-meaningful boundary] A level line L with length l and minimal contrast μ is ε-meaningful if

$$NF(L) = N_{ll} \times [H(\mu)]^l \leq \varepsilon. \tag{6}$$

See Figure 2, for an example of meaningful and maximal meaningful boundaries.

2.3 Meaningful alignments

We assume that the accuracy of a measured gradient direction at a point is equal to $p\pi$ radians. This means that a casual alignment of a direction with a prefixed one happens with probability p. In practice, $p = \frac{1}{16}$ is the best we can hope from usually noisy and aliased digital images. We consider the following event : "on a discrete segment of the image, joining two pixel centers, and with length l counted in points at Nyquist distance, at least k points have the same direction as the segment with precision p." (The direction at each point is computed as the direction of the gradient rotated by $\frac{\pi}{2}$.)

Definition : Consider a segment S of length l containing k aligned points. We call number of false alarms of S,

$$NF(S) = N^4 \sum_{j=k}^{l} C_l^j p^j (1-p)^{l-j}. \tag{7}$$

We say that S is ε-meaningful if $NF(S) \leq \varepsilon$.

If on a straight line we have found a very meaningful segment S, then by enlarging slightly or reducing slightly S, we still find a meaningful segment. This means that meaningfulness cannot be a univoque criterion for detection, unless we can point out the "best meaningful" explanation of what is observed as meaningful. This is done by the following definition, which can be adapted as well to meaningful edges [3], meaningful modes in a histogram [7] and clusters (to appear).

Definition : We say that an ε-meaningful geometric structure A is maximal meaningful if

- it does not contain a strictly more meaningful structure : $\forall B \subset A$, $NF(B) \geq NF(A)$.
- it is not contained in a more meaningful structure : $\forall B \supset A$, $NF(B) > NF(A)$.

It is proved in [7] that maximal structures cannot overlap, which one of the main theoretical outcomes validating the above definitions. We only display in the experiments those maximal meaningful alignments (see Figure 3).

Références

[1] Lowe, D. *Perceptual Organization and Visual Recognition.* Kluwer Academic Publishers, 1985.

[2] A. Desolneux, L. Moisan et J.M. Morel, Meaningful Alignments, *International Journal of Computer Vision*, Vol.40(1), pp.7-23, october 2000.

[3] A. Desolneux, L. Moisan et J.M. Morel, Edge Detection by Helmholtz Principle, à paraître dans *Journal of Mathematical Imaging and Vision*, Vol. 14(3), pp. 271-284, may 2001.

[4] Mumford, D. and J. Shah. Boundary detection by minimizing functionals *IEEE Conf. on Comp. Vision and Pattern Recognition*, San Francisco, 1985.

[5] Metzger, W. *Gesetze des Sehens.* Waldemar Kramer, 1975.

[6] Geman S. and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.* 6, pp. 721-741, 1984.

[7] A. Desolneux, L. Moisan et J.M. Morel, Maximal Meaningful Events and Applications to Image Analysis, preprint, CMLA, http://www.cmla.ens-cachan.fr/Cmla/Publications/2000 soumis en juillet 2000 à *Annals of Statistics.*

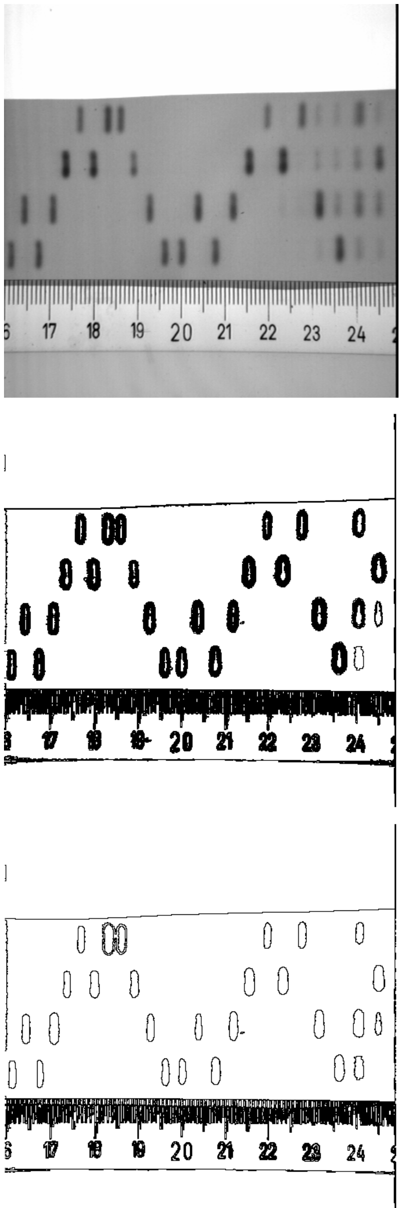

FIG. 2: *From top to bottom : 1. original image; 2. all meaningful boundaries; 3. maximal meaningful boundaries.*

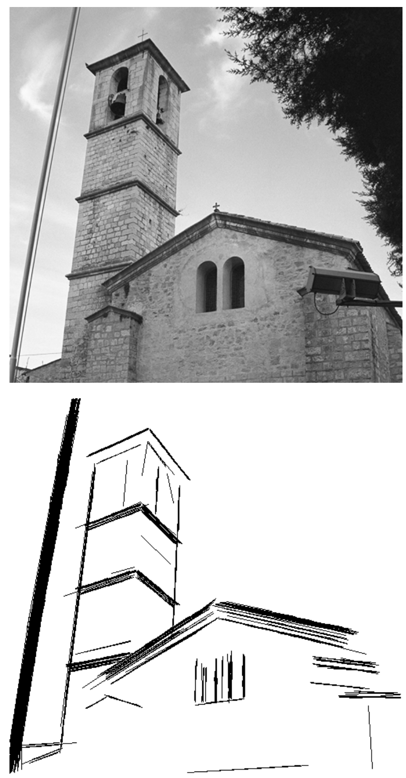

FIG. 3: *From top to bottom : 1. original image; 2. maximal meaningful alignments.*