Shape Recognition Based on an A Contrario Methodology

Pablo Musé[1], Frédéric Sur[1,4], Frédéric Cao[2], Yann Gousseau[3], and Jean-Michel Morel[1]

[1] CMLA, ENS de Cachan, 61 avenue du Président Wilson, 94235 Cachan Cedex, France, {muse,sur,morel}@cmla.ens-cachan.fr
[2] IRISA, INRIA Rennes, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France, fcao@irisa.fr
[3] TSI, CNRS UMR 5141, Télécom Paris, 46 rue Barrault, 75643 Paris Cedex 13, France, gousseau@tsi.enst.fr
[4] LORIA & CNRS, Campus Scientifique BP 239, 54506 Vandoeuvre-lès-Nancy Cedex, France.

1 Introduction

This chapter is concerned with the problem of visual recognition of two dimensional planar shapes. Shape recognition methods usually combine three stages: feature extraction, matching (the important point here being the definition of a distance or dissimilarity measure between features) and decision. The first two stages have been widely addressed in the literature (see for instance [41] or [43] and the references therein), and will be adressed in Sect. 2. On the contrary, the decision problem for shape matching has been much less studied, especially in a generic framework. Moreover, complete procedures starting from raw images and including this decision step are rarely exposed. In this chapter we show that this program is realistic and, even though entering all details is beyond the scope of this chapter, present the main ingredients of the proposed method.

Let us briefly describe the content of this contribution. In Sect. 2, we define the main atoms of our recognition procedure and explain how to extract them from images. Invariance and stability requirements lead us to consider *shape elements* that are suitably selected, normalized and encoded pieces of level lines. In Sect. 3, we introduce an abstract setting in which we build decision rules for pairing two shape elements. This section is quite general and the matching methodology is not restricted to the specific shape elements introduced in Sect. 2. The decision rule relies on a framework introduced by Desolneux, Moisan and Morel [12, 14], based on a statistical principle, the Helmholtz principle. The adaptation of this principle to the shape matching problem leads to an automatic decision rule. In Sect. 4, we present some experiments that show the validity of the proposed model. It is verified that the methodology satisfies the Helmholtz principle [15]: a meaningful match is a match

that is not likely to occur in a context where noise overwhelms the information. Experimental results are presented in Sect. 5. We conclude in Sect. 6.

Before proceeding, let us specify what we mean by an "automatic decision rule" for shape matching and say a few words about the methodology to be developed in Sect. 3. Assume we are looking for a query shape $\mathcal{S}$ in a shape database (usually extracted from an image or a set of images). A distance between shapes is available, so that the smaller the distance the more similar the shapes. The question is: what is the threshold value for that distance to ensure recognition? Given two shapes and an observed small distance δ between them, there are only two possibilities:

1. Both shapes lie at that distance because they 'match' (that is, they are similar because they are two instances of the same object, in the broadest sense).
2. The shape database is so large that, just by chance, one of these shapes is close to $\mathcal{S}$ (there is no underlying common cause between them, and they do not correspond to the same object).

Assume we are able to evaluate, for any δ, the probability of the second possibility. If this quantity happens to be very small for two shapes, then the first possibility is certainly a better explanation. Contrarily to classical approaches to hypothesis testing, we will see that we can build a decision rule only on the likelihood of the second possibility, which is usually more simple to model than the first one. This a contrario methodology will be detailed in Sect. 3, where we will also see that the shape elements selection principle of Sect. 2 follows the same principle.

2 From Images to Normalized Shape Elements

The recognition of shapes (in the widest sense) is invariant with respect to a large set of transformations, as global or non rigid deformation, contrast change, corruption by noise, scaling, local occlusion, etc... Therefore, the atoms of computational shape recognition should satisfy the same properties. An algorithm extracting pieces of Jordan curves corresponding to invariant local representations of shapes in images was proposed by Lisani et al. [24, 25], and mostly satisfies the above conditions. It proceeds with the following steps, that will be detailed below.

1. Extraction of meaningful level lines.
2. Affine invariant smoothing of the extracted level lines.
3. Semi local encoding of pieces of level lines after affine or similarity normalization.

The conjunction of these three stages was first introduced by Lisani et al. [24, 25]; the third stage is also based on the seminal work of Lamdan et al. [22], followed by Rothwell's work on invariant indexing [36]. For a more recent application of similar ideas, see Orrite et al. [35].

2.1 Extracting Meaningful Curves from Images

In computer vision, extraction of shape information from images dates back to Marr [27], but Attneave [5], as well as Wertheimer [42] and other Gestaltists had already remarked that information in images is concentrated along contours, and that shape perception is invariant to contrast changes (changes in the color and luminance scales). As we will see in the next paragraph, shapes can then be modeled as Jordan curves. However, as pointed out by Kanisza [21], in every day's vision most objects are partially hidden by other ones and despite this occlusion phenomenon humans still can recognize shapes in images. Consequently, the real atoms of shape representation should not be the whole Jordan curves corresponding to objects boundaries, but pieces of them. In this work we will adopt this atomic shape representation; we will call *shape element* any piece of Jordan curve.

Topographic Map and Tree of Level Lines

Following Mathematical Morphologists, the image information is completely contained in a family of binary images that are obtained by thresholding the images at given values [28, 38]. This is equivalent to considering level sets; the (upper) level set of u at the value λ is

$$\chi_\lambda(u) = \{x \in \mathbb{R}^2, \quad u(x) \geq \lambda\}. \tag{1}$$

Obviously, if we only consider a coarsely quantized set of different grey levels, information is lost, especially in textures. Nevertheless it is worth noticing how large shapes are already present with as few as 5 or 6 levels. As remarked by Serra [38], we can reconstruct an image from the whole family of its level sets by

$$u(x) = \sup\{\lambda \in \mathbb{R}, \quad x \in \chi_\lambda(u)\}.$$

Thus, the level sets provide a complete representation of images. Morphologists also noticed that boundaries of level sets fit parts of objects boundaries very well. They call level lines the topological boundaries of connected components of level sets, and topographic map of an image the collection of all its level lines. Remark that, if the image u is C^1, level lines coincide with isophotes in the neighborhood of x such that $Du(x) \neq 0$. The topographic map also enjoys several important advantages [10]:

- It is invariant with respect to contrast change (a contrast change is the composition of the image with an increasing real function). It is not invariant to illumination change, since in this case the image is really different although it represents the same scene. However, many pieces of level lines remain the same under such illumination changes.
- In general, edge detectors lead to disconnected pieces of curves which are too small to be individually relevant. A preliminary grouping step is necessary to get shape elements. On the contrary using level lines directly yields long curves since, for almost all gray level, level lines of images with bounded variation are Jordan curves [17]. We consider it more easy to compute shape elements by locally encoding level lines.

- It is a hierarchical representation: since level sets are ordered by the inclusion relation (and so are their connected components), the topographic map may be embedded in a tree structure.
- Most important regarding the main subject of this chapter, object contours locally coincide with level lines very well. Basically, level lines are everywhere normal to the gradient as edges. Contrarily to local edges, level lines are accurate at occlusions. Whereas edges detectors usually fail near T-junctions (and additional treatments are necessary), there are several level lines at a junction (See Fig. 1).

Fig. 1. Level lines and T-junction. Depending on the grey level configuration between shapes and background, level lines may or may not follow the objects boundary. In any case, junctions appear where two level lines separate. Here, there are two kinds of level lines: the occluded circle and the shape composed of the union of the circle and the square. The square itself may be retrieved by difference.

However, level lines in textures are usually very complicated and are not always useful for shape recognition. Moreover, because of noise and interpolation, many level lines may follow roughly one and the same contour. Thus it is useful, for practical computational reasons, to select only the most meaningful level lines.

Remark that level lines are Jordan curves and, for continuous images, do not intersect. Moreover, for almost all level of a C^1 image, the interior of a closed level line is a simply connected set. Monasse and Guichard [32] call such a set a *Shape*, and to avoid any confusion, we will write *morphological shape*. Basically, morphological shapes are either connected components of level sets whose holes have been filled, or connected components of holes of level sets whose own holes have been filled. Actually, the situation is slightly more complicated because of open level lines that meet the image border, and that have to be closed one way or another. Nevertheless, Monasse and Guichard proved that these morphological shapes could be used to define a tree which is called the *tree of level lines*. Each node of the tree is a morphological shape, attached to a grey level, or equivalently the boundary of the morphological shape, which is a level line of the image. In Figure 2, we display an example of such a tree for a simple synthetic image. An algorithm called *Fast Level Set Transform* [32] allows to efficiently compute this tree. It may be extended to a bilinearly interpolated image, whose level lines suffer less from pixelization effects than those of a pixelwise constant image. In this case, images become continuous and have infinitely many level lines. Thus, there is a preliminary choice of quanti-

zation of these lines. If the original image was encoded on 8-bits, we simply choose a quantization step equal to 1, since we know that 256 grey levels give fair enough visual quality. Moreover, the selection procedure we describe below would not give more level lines with a thinner quantization step.

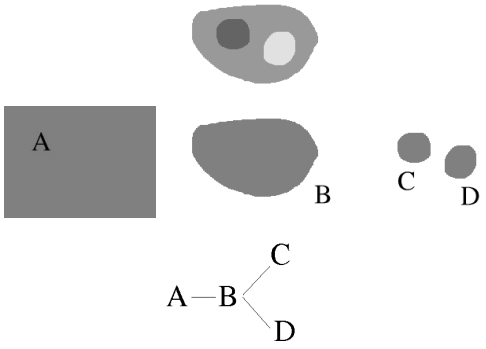


Fig. 2. Top : a synthetic image. Middle : its morphological shapes. Bottom : the corresponding tree.

Meaningful Boundaries

A very simple and efficient method to select the most meaningful level lines in the topographic map has been introduced by Desolneux, Moisan and Morel in [13]. Recent improvements have been achieved in [9] but, for the sake of simplicity, we only describe the original arguments which rely on an a contrario detection principle which will be detailed in Sect. 3.

Let $u : \mathbb{R}^2 \to \mathbb{R}$ be a differentiable grey level image. Assume that we have a measure of contrast. To simplify we take it here equal to the norm of the gradient. Assume that we know the distribution of the gradient of u, given by

$$H_c(\mu) = P(|Du| > \mu).$$

In practice, we shall take a finite differences approximation of the gradient. The empirical histogram is used to approximate H_c. That is, we assume that the gradient norm is distributed as the positive random variable X defined by

$$\forall \mu > 0, \quad P(X > \mu) = \frac{\#\{x \in \Gamma,\, |Du(x)| > \mu\}}{\#\{x \in \Gamma,\, |Du(x)| > 0\}}, \tag{2}$$

where the symbol $\#$ designs the cardinality of a set, Γ the finite sampling grid, $|Du|$ is computed by finite difference approximation.

Definition 1 ([13]). *Let E be a finite set of N_{ll} level lines of u. We say that a level line C is an ε-meaningful boundary if*

$$NFA(C) \equiv N_{ll} H_c(\min_{x \in C} |Du(x)|)^{l/2} < \varepsilon, \tag{3}$$

where l is the length of C. This number is called number of false alarms (NFA) of C.

This definition will be explained in details in Sect. 3.4, since its justification uses the *a contrario framework* that will be introduced in Sect. 3. Let us take it for granted for a while.

Maximal Boundaries

We know that all level lines are needed to perfectly reconstruct the image. Nevertheless, only a few of them suffice to describe most shape information. Because of interpolation and blur, level lines accumulate along edges, and even meaningful level lines still are locally redundant, as far as shapes are concerned. A very elegant way to eliminate some redundancy is to use the structure of the tree of level lines which simply contain the topological inclusion relation between level lines.

Definition 2 ([31]). *A monotone section of a tree of level lines is a part of a branch such that each node has a unique son and where grey level is monotone (no contrast reversal).*
A maximal monotone section is a monotone section which is not strictly included in another one.

Definition 3 ([13]). *We say that a meaningful boundary is maximal meaningful if it has a minimal NFA in a maximal monotone section of the tree of meaningful level lines.*

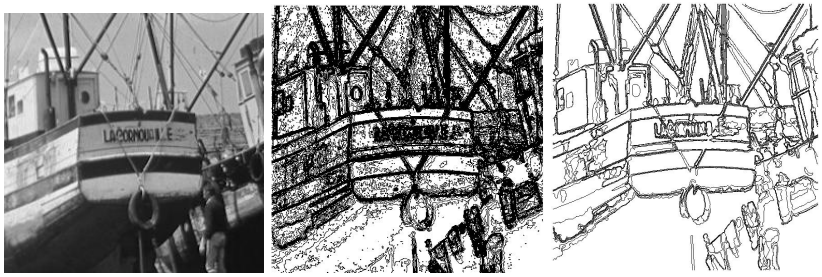
Remark that this definition makes sense since meaningful level lines still enjoy the same tree structure as level lines. In practice, meaningful level lines often represent less than 10% of the total number of level lines (most of which are actually very small and due to noise and texture). About one meaningful level line over 10 is usually a maximal one. Hence, about 99% of all level lines are eliminated.

Fig. 3 illustrates that the loss of information resulting from the use of meaningful level lines is negligible compared to the gain in information compactness. This reduction is crucial in order to speed up the shape matching stage that follows the encoding.

2.2 Level Lines Smoothing

Once meaningful level lines are extracted, we need to smooth them in order to eliminate noise and aliasing effects. The Geometric Affine Scale Space [2, 37] is fully convenient (since such a smoothing commutes with special affine transformations and since we are interested in affine invariance). It is given by the following motion by curvature

$$\frac{\partial x}{\partial t} = |\text{Curv}(x)|^{\frac{1}{3}} \mathbf{n}(x),$$



(a) (b) (c)

Fig. 3. Extraction of meaningful level lines. (a) original "La Cornouaille" image, (b) level lines, represented here with grey-level quantization step equal to 10 (there are 54790 level lines for a quantization step of 1, and they fill the whole image), (c) the 296 maximal meaningful level lines (there are 4342 meaningful level lines but with no real additional information).

where x is a point on a level line, $\mathrm{Curv}(x)$ the curvature and $\mathbf{n}(x)$ the normal to the curve, oriented towards concavity. We use a fast implementation by Moisan [29]. Logically, it would be interesting to use this equation in a true multiscale recognition procedure: each extracted shape should be described at several different scales. The price to pay is of course a higher numerical complexity. In this chapter, we only use this equation as a way to wipe out pixelization effect due to quantization, so that the invariance properties of the equation are not used to their full potential. The scale at which the smoothing is applied is fixed and given by the pixel size. Nevertheless, this is still very useful: the aim is to reduce the complexity of meaningful level lines by simplifying them. The final goal remains the same: to make the shape matching faster. Indeed, smoothing reduces the number of bitangents on level lines by eliminating those due to noise; consequently it also reduces the number of encoded shape elements, as it will become clear from what follows.

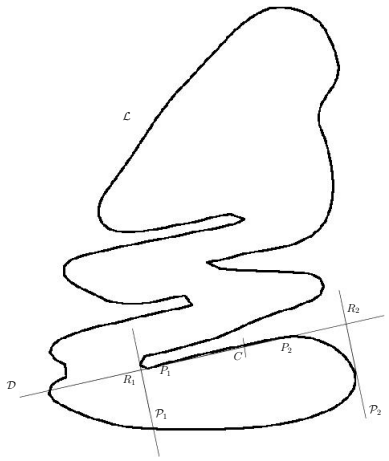
2.3 Semi-local Normalization and Encoding

The last stage of the invariant shape encoding algorithm is semi-local normalization and encoding. Roughly speaking, in order to build invariant representations (up to either similarity or affine transformations), we define local frames for each level line, based on robust directions (tangent lines at flat parts, or bitangent lines). Such a representation is obtained by uniformly sampling a piece of curve in this normalized frame. The following section is devoted to an improvement of Lisani's algorithm.

The proposed semi-local normalization of level lines or, more generally speaking, of Jordan curves is based on robust directions. These directions are given by bitangent lines, or by tangent lines at flat parts (a flat part is a portion of a curve which is everywhere unexpectedly close to the segment joining its endpoints, relatively to an adequate background model [33, 40]).

We now detail the procedures used to achieve similarity invariance for semi-local normalization and encoding of Jordan curves. In what follows we consider direct

Euclidean parameterization for level lines. We treat the similarity invariant case, and refer the reader to [25] for the affine invariant encoding.



In order to represent a level line $\mathcal{L}$, for each flat part, and for each couple of points on which the same straight line is tangent to the curve, do:

a) Let P_1 and P_2 be either the tangency points when dealing with bitangency, or the endpoints for the detected segment when dealing with flat parts. Consider the tangent line $\mathcal{D}$ to these points;
b) Starting backward from P_1, call $\mathcal{P}_1$ the previous tangent to $\mathcal{L}$, orthogonal to $\mathcal{D}$. Starting forward from P_2, call $\mathcal{P}_2$ the next tangent to $\mathcal{L}$, orthogonal to $\mathcal{D}$;
c) Find the intersection points between $\mathcal{P}_1$ and $\mathcal{D}$, and between $\mathcal{P}_2$ and $\mathcal{D}$. Call them R_1 and R_2, respectively;
d) Store the *normalized* coordinates of M equi-distributed points over an arc on $\mathcal{L}$ of normalized length F, centered at C, the intersection point of $\mathcal{L}$ with the perpendicular bisector of $[R_1 R_2]$. By "normalized coordinates" we mean coordinates in the similarity invariant frame defined by points R_1, R_2 mapped to $(-\frac{1}{2}, 0), (\frac{1}{2}, 0)$, respectively.

Fig. 4. Similarity invariant semi-local encoding. On the left, an illustration based on a flat part.

The procedure is illustrated and detailed in Fig. 4. Two implementation parameters, F and M, are involved in this normalization procedure. The value of F determines the normalized length of the shape elements, and is to be chosen having in mind the following trade-off: if F is too large, shape elements will not deal well with occlusions, while if it is too small, shape elements will not be discriminatory enough. One therefore faces a classical dilemma in shape analysis: locality versus globality of shape representations. The choice of M is less critical from the shape representation viewpoint, since it is just a precision parameter. Its value is to be chosen as a compromise between accuracy of the shape element representation and computational load.

On Fig. 5 we show several normalized shape elements extracted from a single line, taking $F = 5$ and $M = 45$. Notice that the representation is quite redundant. While the representation is certainly not optimal because of redundancy, it increases the possibility of finding common shape elements when corresponding shapes are present in images, even if they are degraded or subject to partial occlusions.

All experiments to be presented in Sect. 5 concerning matching based on this semi-local encoding (or the affine invariant procedure detailed in [25]) were carried out using $F = 5$ and $M = 45$, since it seems to be a good compromise solution. We observed that in general these parameters can be fixed once and for all, and do not need to be tuned by the user. Let us notice that some curves cannot be coded with $F = 5$: when their length is too small with respect to the length of the segment line $[R_1 R_2]$, the resulting shape element would overlap itself.

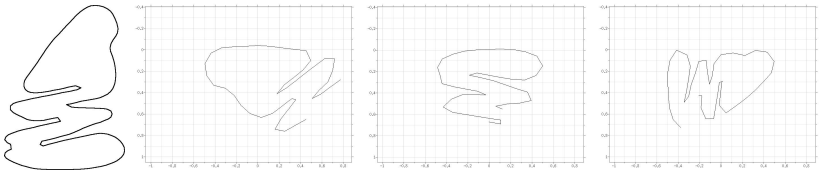


Fig. 5. Example of semi-local similarity invariant encoding. The line on the left generates 19 shape elements ($F = 5$, $M = 45$). Twelve of them are based on bitangent lines, the other ones are based on flat parts. The representation is quite redundant. Here are displayed three normalized shape elements, one deriving from bitangent lines, and two from a flat part.

3 An A Contrario Decision Framework

By applying the procedures of the previous section, similarity invariant *shape elements* are extracted from images. (Affine invariant shape elements may be extracted as well, see [25].) These are the basic objects to be recognized. Generally speaking, the recognition problem is hard since sorting the shape elements along a similarity measure (such as a distance) to a query shape element is not sufficient; we must answer by *yes* or *no* the question "does that shape element look like the query shape element"? In that case, the problem consists in automatically setting a threshold δ over the similarity measure and in giving a confidence level to this decision. This is precisely the aim of the proposed methodology. We shall first build up an empirical statistical model of the shape elements database. The relevant matches will be detected a contrario as rare events for this *background model*. This detection framework has been recently applied by Desolneux et al. to the detection of alignments [12] or contrasted edges [13], by Almansa et al. to the detection of vanishing points [1], by Stival and Moisan to stereo images [30], by Gousseau to the comparison of image "composition" [19] and by Cao to the detection of good continuations [7]. The main advantage of this technique is that the only parameter which controls the detection is the Number of False Alarms, that has already been introduced for level lines selection in Sect. 2 and that will be defined for shape matching in Sect. 3.2.

3.1 Shape model versus background model

Let us first introduce some notations. Our aim is to compare a given query shape element $\mathcal{S}$ with the N shape elements of a database $\mathcal{B}$. We assume S to be represented by a code, that is a set of K features $x_1(S), x_2(S), \ldots, x_K(S)$, each of them belonging to a set E_i endowed with a dissimilarity measure d_i. We then define the product dissimilarity measure on $E_1 \times E_2 \times \cdots \times E_K$ by

$$d(\mathcal{S}, \mathcal{S}') = \max_{i \in \{1,\ldots,K\}} d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')).$$

Observe that in order for this definition to be sound, the d_is are supposed to have the same range, a property that will be satisfied by the dissimilarity measures to be considered in this chapter. In what follows, we will call this dissimilarity measures distances, although they are not necessarily metrics.

We assume no other information but the observed set of features, and we are interested in shape elements which are close to the query shape element $\mathcal{S}$ because their generation shares some common cause with the generation of $\mathcal{S}$. But what is the underlying common cause? We probably do not know, and this is the point. Indeed, directly addressing this problem is not possible, unless we have the exact model of $\mathcal{S}$ at hand. Such a model would imply an extra knowledge (for instance some expert should have first built it up). We are therefore unable to compute the probability that a shape element is near $\mathcal{S}$ *because it has been generated by the shape model of* $\mathcal{S}$.

Consequently, we are led to wonder whether a database shape element is near the query $\mathcal{S}$ "just by chance", and to detect correspondences as unexpected coincidences. In order to address this latest point, we have to build up a *background model*: a model to compute the probability that a shape element from the database is near $\mathcal{S}$ *by chance*. We assume that shape elements are defined on some probability space $(\Omega, \mathcal{A}, \Pr)$. A background model, at fixed $\mathcal{S}$, is defined as follows.

Definition 4. *We call* background model *any random model $\mathcal{S}'$ for which the following holds true:*

(A) *The random variables $d_i(x_i(\mathcal{S}), x_i(\mathcal{S}'))$ ($i \in \{1, \ldots, K\}$) from Ω to $\mathbb{R}^+$ are mutually independent.*

From now on, at fixed $\mathcal{S}$, and for every $i \in \{1, \ldots, K\}$, we denote

$$P_i(\mathcal{S}, \delta) := \Pr(d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \leq \delta).$$

3.2 A Detection Terminology

Number of False Alarms

A distance function between shape elements being given, deciding whether a shape element matches another shape element consists in setting a threshold δ over the distances. Ideally, δ should be set automatically, without any user tuning. In order to do so, we will rely on the following quantities.

Definition 5. *The* Number of False Alarms *of the shape element $\mathcal{S}$ at a distance d is*

$$NFA(\mathcal{S}, d) := N \prod_{i \in \{1,\dots,K\}} P_i(\mathcal{S}, d).$$

We will see later in Sect. 3.3 that this number can be seen as the average number of false alarms that are expected in a statistical test framework, when we test whether the distance from each shape element in the database to $\mathcal{S}$ is below d.

Definition 6. *The number of false alarms of the query shape element $\mathcal{S}$ and a database shape element $\mathcal{S}'$ is the number of false alarms of $\mathcal{S}$ at a distance $d(\mathcal{S}, \mathcal{S}')$:*

$$NFA(\mathcal{S}, \mathcal{S}') := NFA(\mathcal{S}, d(\mathcal{S}, \mathcal{S}')).$$

The *number of false alarms between $\mathcal{S}$ and $\mathcal{S}'$* corresponds to the expected number of database shapes which are "false alarms" and whose distance to $\mathcal{S}$ is lower than $d(\mathcal{S}, \mathcal{S}')$.

Remark 1. For the sake of simplicity, the same notation is used for both the preceding definitions of the number of false alarms. Let us moreover notice that the arguments of this latest NFA (seen as a two variables function) do not play a symmetric role.

Meaningful Matches

Next, we decide which shapes of the database match the query shape $\mathcal{S}$ by bounding the number of false alarms.

Definition 7. *A shape element $\mathcal{S}'$ is an* ε-meaningful match *of the query shape element $\mathcal{S}$ if their number of false alarms is bounded by ε:*

$$NFA(\mathcal{S}, \mathcal{S}') \leq \varepsilon.$$

Notice that since the functions $P_i(\mathcal{S}, d) : d \mapsto \Pr(y \in E_i \text{ s.t. } d_i(x_i(\mathcal{S}), y) \leq d)$ are non-decreasing, the function $d \mapsto \text{NFA}(\mathcal{S}, d)$ is pseudo-invertible. That is, there exists a unique positive real number $\delta(\varepsilon)$ (depending on the query shape $\mathcal{S}$) such that

$$\delta(\varepsilon) := \max\{\delta > 0, \text{NFA}(\mathcal{S}, \delta) \leq \varepsilon\}.$$

The proposition that follows is then straightforward.

Proposition 1. *A shape element $\mathcal{S}'$ is an ε-meaningful match of the query shape element $\mathcal{S}$ if and only if*

$$d(\mathcal{S}, \mathcal{S}') \leq \delta(\varepsilon).$$

The decision rule we propose thus amounts, for a fixed $\mathcal{S}$, to compare $d(\mathcal{S}, \mathcal{S}')$ to the bound $\delta(\varepsilon)$. The justification behind this rule is that the expectation of the number of shapes that match $\mathcal{S}$ "by chance" is then bounded by ε. The following proposition makes this claim more formal.

Proposition 2. *Under the assumption that the database shape elements are identically distributed following the background model, the expectation of the number of ε-meaningful matches is less than ε.*

Proof. Let $\mathcal{S}'_j$ ($1 \leq j \leq N$) denote the shape elements in the database, and χ_j be the indicator function of the event e_j: "$\mathcal{S}'_j$ is an ε-meaningful match of the query $\mathcal{S}$" (i.e. its value is 1 if $\mathcal{S}'_j$ is an ε-meaningful match of $\mathcal{S}$, and 0 otherwise). Let $R = \sum_{j=1}^{N} \chi_j$ be the random variable representing the number of shapes ε-meaningfully matching $\mathcal{S}$.
The expectation of R is $\mathbb{E}(R) = \sum_{j=1}^{N} \mathbb{E}(\chi_j)$. Using Prop. 1, it follows that

$$\mathbb{E}(\chi_j) = \Pr(\mathcal{S}'_j \text{ is an } \varepsilon\text{-meaningful match of } \mathcal{S}) = \Pr(d(\mathcal{S}, \mathcal{S}'_j) \leq \delta(\varepsilon)).$$

Since shape elements from the database are assumed to satisfy the assumptions of the background model, one has

$$\mathbb{E}(\chi_j) = \prod_{i=1}^{K} P_i(\mathcal{S}, \delta(\varepsilon)) = \frac{1}{N}\text{NFA}(\mathcal{S}, \delta(\varepsilon)).$$

Linearity of expectation implies $\mathbb{E}(R) = \frac{1}{N} \sum_{j=1}^{N} \text{NFA}(\mathcal{S}, \delta(\varepsilon))$. Hence, by definition of δ, this yields $\mathbb{E}(R) \leq \sum_{j=1}^{N} \varepsilon \cdot N^{-1}$; therefore $\mathbb{E}(R) \leq \varepsilon$. □

The key point is that the linearity of the expectation allows to compute $\mathbb{E}(R)$. Since dependencies between events e_j are unknown, we are not able to estimate the probability law of R.

Let us now summarize. A reference shape $\mathcal{S}$ being given, we seek its ε-meaningful matches, which by Prop. 1 boils down to a bound on distances. For each matching shape $\mathcal{S}'$, the number $NFA(\mathcal{S}, \mathcal{S}')$ quantifies the quality of the match, and Prop. 2 gives a handy meaning to the number ε (we will always use $\varepsilon = 1$ in subsequent experiments).

Recognition Threshold is Relative to the Context

Notice that the empirical probabilities take into account the 'rareness' or 'commonness' of a possible match; indeed the threshold δ is less restrictive in the first case and stricter in the other one. If a query shape $\mathcal{S}_1$ is rarer than another one $\mathcal{S}_2$, then the database contains more shapes close to $\mathcal{S}_2$ than shapes close to $\mathcal{S}_1$, below a certain fixed distance d'. Now, probabilities will be estimated through empirical frequencies over the database (see Sect. 3.6). As a consequence, if a query shape $\mathcal{S}_1$ is rarer than another one $\mathcal{S}_2$, then we have, for $i \in \{1, \ldots, K\}$ and $d \leq d'$,

$$\hat{P}_i(\mathcal{S}_1, d) \leq \hat{P}_i(\mathcal{S}_2, d).$$

This yields (with obvious notations) $\delta_{\mathcal{S}_2} \leq \delta_{\mathcal{S}_1}$ (provided both quantities are below d'), i.e. the rarer the sought shape, the higher the recognition threshold.

Another formulation of the same property is that if a given query shape is rarer among the shapes out of a database $\mathcal{B}_1$ than among the shapes out of a database $\mathcal{B}_2$, then we get for every $i \in \{1, \ldots, K\}$ and for d "small enough"

$$\hat{P}_i^1(\mathcal{S}, d) \leq \hat{P}_i^2(\mathcal{S}, d),$$

where $\hat{P}_i^1$ and $\hat{P}_i^2$ are respectively estimated over $\mathcal{B}_1$ and $\mathcal{B}_2$. This yields $\delta_2 \leq \delta_1$. This latest point is in fact one of the key points of the proposed methodology, and should be the cornerstone of every shape recognition system. Suppose that we seek a character, let us say a 'a' among different characters from an ordinary scanned text. Then the recognition threshold (under which a character matches the sought 'a') should be larger than the one obtained when searching the same 'a' among other 'a' of various slightly different fonts. Indeed, the sought shape would be relatively much rarer in the latter case than in the former. The conclusion is that the distance threshold proposed by our algorithm auto-adapts to the relative "rareness" of the query shape among the database shapes. The "rarer" the query shape, the more permissive the corresponding distance threshold, and conversely.

Observe also that the number of false alarms, and therefore the confidence of a recognition, depends upon the size of the searched database. This is counterintuitive, but only at first sight. Indeed, when the size of the searched database grows, the probability that a piece of shape be created "just by chance" grows too. Let us take a classical example. Images of vegetation can lead humans and computer vision algorithm to hallucinate faces. The explanation is simple ; such textured images create lots of casual spatial arrangements. Some of them can look like a searched shape. The larger the database, the likelier such false alarms.

Why an A Contrario Decision?

The advantages of the a contrario decision based on the NFA compared to the direct setting of a distance threshold between shape elements are obvious. On the one hand, thresholding the NFA is much more handy than thresholding the distance. Indeed, we simply put $\varepsilon = 1$ and allow at most one false alarm among meaningful matches (we simply refer to 1-meaningful matches as "meaningful matches"), or $\varepsilon = 10^{-1}$ if we want to impose a higher confidence in the obtained matches. The detection threshold ε is set uniformly whatever the query shape element and the database may be: the resulting distance threshold adapts automatically according to them as explained in the preceding section. On the other hand, the lower ε, the "surer" the ε-meaningful detections are. Of course, the same claim is true when considering distances: the lower the distance threshold δ, the surer the corresponding matches, but considering the NFA quantifies this confidence level. Moreover, computing the NFA does not need any shape model. This is a major advantage of the proposed method, since having a shape model means that the query shape has already been recognized before somehow or other. We will see in Sect. 3.3 how this point relates to the control of false positives in a classical hypothesis testing framework.

Comparing Two Databases

Let us end up with the definition of the number of false alarms when comparing all shape elements in a database to all shape elements in another database. This corresponds to the experiments of Sect. 5 where the shape contents of two images are compared. When searching the shapes belonging to a database $\mathcal{B}_1$, made of N_1 shape elements, among the N_2 shape elements belonging to a database $\mathcal{B}_2$, we define:

Definition 8. *The* Number of False Alarms *of a shape $\mathcal{S}$ (belonging to $\mathcal{B}_1$) at a distance d is*

$$NFA(\mathcal{S}, d) = N_1 \cdot N_2 \cdot \Pr\left(\mathcal{S}', \max_{i\in\{1\dots K\}} d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \leq d\right).$$

The probabilities (depending on the searched shape $\mathcal{S}$) are estimated as before, as a product of K empirical estimates over the database $\mathcal{B}_2$ among which the query shapes are sought. For each shape in $\mathcal{B}_1$ we also define ε-meaningful matches. The claim up to which we shall expect on the average ε false alarms among the ε-meaningful matches over all $N_1 \cdot N_2$ tested pairs of shapes (Prop. 2) still holds.

3.3 A Contrario Decision as Hypothesis Testing

In this section, we show how the proposed methodology can be interpreted in a statistical testing framework [16, 39]. A shape $\mathcal{S}'$ being observed, the hypothesis we are interested in is $\mathcal{H}_1$: "$\mathcal{S}'$ has been generated by the shape model of $\mathcal{S}$". However, as explained before, handling this hypothesis with our assumption (no available shape model for $\mathcal{S}$) is impossible. We are therefore led to concentrate on an alternative hypothesis $\mathcal{H}_0$: "$\mathcal{S}'$ follows the background model". We consider a test relying on the distance between shapes.

Definition 9. *A query shape element $\mathcal{S}$ being given, the statistical test $\mathcal{T}_\delta(\mathcal{S})$ is defined by:*

- *if a database shape element $\mathcal{S}'$ is such that $d(\mathcal{S}, \mathcal{S}') < \delta$ then hypothesis $\mathcal{H}_1$ is accepted ($\mathcal{S}'$ is near $\mathcal{S}$ because of some causality).*
- *Otherwise, $\mathcal{H}_1$ is rejected and the null hypothesis $\mathcal{H}_0$ is accepted ($\mathcal{S}'$ is near $\mathcal{S}$ casually).*

The quality of a statistical test is measured by the probability of taking wrong decisions : reject $\mathcal{H}_0$ for $\mathcal{S}$ although $\mathcal{H}_0$ is valid (type I error, false positive) or reject $\mathcal{H}_1$ for an observation $\mathcal{S}$ for which $\mathcal{H}_1$ is actually true (type II error, mis-detection). A probability measure can be associated to each type of error.

- The *probability of false alarms* (associated with type I error) $\alpha = \Pr(d(\mathcal{S}, \mathcal{S}') < \delta | \mathcal{H}_0)$.
- The *probability of non-detection* or *probability of a miss* (associated with type II error) $\alpha' = \Pr(d(\mathcal{S}, \mathcal{S}') \geq \delta | \mathcal{H}_1)$.

It is clear that the lower α and α', the better the test, but it is also clear that α and α' cannot be independently optimized. The problem is to find a trade-off between these two probabilities. Two well known approaches to this problem are *Neyman-Pearson Theory* and *Bayesian tests*. However, the practical limits of this theoretical framework are obvious. They indeed need that one knows the likelihood of both $\mathcal{H}_0$ and $\mathcal{H}_1$, which is in general unrealistic if the aim is to recognize an unspecified query shape (a generative model is indeed needed for the query shape $\mathcal{S}$ to compute the likelihood of a shape $\mathcal{S}'$ under hypothesis $\mathcal{H}_1$). Moreover, the Bayesian approach needs prior information, which is either arbitrary or is strongly related to a specific problem for which supplementary information is provided.

Let us summarize the situation. We are not able to compute the probability of non-detection $\Pr(d(\mathcal{S},\mathcal{S}') \geq \delta|\mathcal{H}_1)$. On the other hand, a straightforward computation provides the value of the probability of false alarms of the statistical test $\mathcal{T}_\delta(\mathcal{S})$, that is, $\Pr(d(\mathcal{S},\mathcal{S}') < \delta|\mathcal{H}_0)$. Indeed, by definition of d,

$$\Pr(d(\mathcal{S},\mathcal{S}') < \delta|\mathcal{H}_0) = \Pr\left(\max_{i\in\{1,\dots,K\}} d_i(x_i(\mathcal{S}),x_i(\mathcal{S}')) \leq \delta \mid \mathcal{H}_0\right).$$

Now, by definition of $\mathcal{H}_0$, the independence assumption **(A)** holds true so that

$$\begin{aligned}\Pr(d(\mathcal{S},\mathcal{S}') < \delta|\mathcal{H}_0) &= \prod_{i\in\{1,\dots,K\}} Pr\left(d_i(x_i(\mathcal{S}),x_i(\mathcal{S}')) \leq \delta\right)\\ &= \prod_{i\in\{1,\dots,K\}} P_i(\mathcal{S},\delta). \qquad (4)\end{aligned}$$

It is therefore straightforward (by definition of $\delta(\varepsilon)$) that the statistical test $\mathcal{T}_{\delta(\varepsilon)}(\mathcal{S})$ has a probability of false alarm bounded by ε/N:

$$\Pr(d(\mathcal{S},\mathcal{S}') < \delta|\mathcal{H}_0) \leq \varepsilon/N.$$

The a contrario decision rule therefore consists in accepting hypothesis $\mathcal{H}_1$ when the null hypothesis $\mathcal{H}_0$ is unlikely, this likeliness being quantified by ε. Recall also that Prop. 2 shows that the average number of false alarms when testing a shape against all shapes in the database is bounded by ε, therefore giving a clear meaning to this bound. In short, we accept the hypothesis "a database shape element $\mathcal{S}'$ matches the query shape element $\mathcal{S}$ " as soon as it is not likely that $\mathcal{S}'$ is near $\mathcal{S}$ "by chance".

Several earlier works conceive the shape recognition problem in the same spirit, being based on the computation of a probability of false alarms. This computation can be achieved by following several approaches. All of them are of course based on the background modeling. For instance, Grimson and Huttenlocher [20] estimate the probability that some features of the sought shapes are retrieved just thanks to the so-called "conspiracy of random", by assuming that features are uniformly distributed. A more accurate approach consists in building a tighter background model. Examples can be found among the literature about detection of low resolution targets over a cluttered background (see for example [11]). Such approaches are derived from classical signal processing methods (where the noise is modeled as a Gaussian

process, thus enabling to exactly compute the probability of false alarms and derive the detection threshold, see e.g. [4]). Specific background models can also be built, depending on the considered problem. For instance, Amit *et al.* [3] address a shape classification problem in this perspective. Another approach is to simultaneously use a background and a shape model, as done by Lindenbaum in [23] where performances of shape recognition algorithms are studied in a fairly general context. Let us also name the Statistical Parametric Mapping which is a popular method for analysis of brain imaging data sequences: by integrating spatial dependences, large deviations with respect to the SPM are attributed to the cognitive process of interest (see [18] for an introduction).

3.4 Meaningful Boundaries and A Contrario Framework

Let us now interpret the definition of meaningful boundaries that we gave in Sect. 2 in the a contrario decision framework. The implicit definition of contours contained in Def. 1 is that the norm of the gradient should be large everywhere along an edge. Again, we consider it hard to model the dependence of the gradient values along a true edge and prefer to take the decision by contradicting an independence hypothesis.

Assume that X is a real random variable described by the inverse distribution function $H(\mu) = \Pr(X \geq \mu)$. Assume that u is a random image such that the values $|Du|$ are independent with the same law as X. Let now E be a set of random curves (C_i) in u such that $\#E$ (the cardinality of E) is independent of each C_i. For each i, we note $\mu_i = \min_{x \in C_i} |Du(x)|$. We also assume that we can choose L_i independent points on C_i (points that are afar at least by Nyquist's distance, a property which in particular bounds L_i from above). We can think of the C_i as random walks with independent increments but since we choose a finite number of samples on each curve, the law of the C_i does not really matter. We assume that L_i is independent from the pixels crossed by C_i.
By mimicking Def. 1, we say that C_i is ε-meaningful if

$$NFA(C_i) = \#E \cdot H(\mu_i)^{L_i} < \varepsilon.$$

Proposition 3. *The expected number of ε-meaningful curves in a random set E of random curves is smaller than ε.*

Proof. Let us denote by X_i the binary random variable equal to 1 if C_i is meaningful and to 0 else. Let also $N = \#E$. Let us denote by $\mathbb{E}(X)$ the expectation of a random variable X in the a contrario model. We then have

$$\mathbb{E}\left(\sum_{i=1}^{N} X_i\right) = \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{N} X_i | N\right)\right).$$

We have assumed that N is independent from the curves. Thus, conditionally to $N = n$, the law of $\sum_{i=1}^{N} X_i$ is the law of $\sum_{i=1}^{n} Y_i$, where Y_i is a binary variable equal to 1 if $nH(\mu_i)^{L_i} < \varepsilon$ and 0 else. By linearity of expectation,

$$\mathbb{E}\left(\sum_{i=1}^{N} X_i | N = n\right) = \mathbb{E}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \mathbb{E}(Y_i).$$

Since Y_i is a Bernoulli variable, $\mathbb{E}(Y_i) = \Pr(Y_i = 1) = \Pr(nH(\mu_i)^{L_i} < \varepsilon) = \sum_{l=0}^{\infty} \Pr(nH(\mu_i)^{L_i} < \varepsilon | L_i = l) P(L_i = l)$. Again, we have assumed that L_i is independent of the gradient distribution in the image. Thus conditionally to $L_i = l$, the law of $nH(\mu_i)^{L_i}$ is the law of $nH(\mu_i)^l$. Let us finally denote by $(\alpha_1, \cdots, \alpha_l)$ the l (independent) values of $|Du|$ along C_i. We have

$$\begin{aligned}
\Pr\left(nH(\mu_i)^l < \varepsilon\right) &= \Pr\left(H(\min_{1\le k\le l} \alpha_k) < \left(\frac{\varepsilon}{n}\right)^{1/l}\right) \\
&= \Pr\left(\max_{1\le k\le l} H(\alpha_k) < \left(\frac{\varepsilon}{n}\right)^{1/l}\right) \text{ since } H \text{ is nonincreasing} \\
&= \prod_{k=1}^{l} \Pr\left(H(\alpha_k) < \left(\frac{\varepsilon}{n}\right)^{1/l}\right) \text{ by independence} \\
&\le \frac{\varepsilon}{n},
\end{aligned}$$

since if H is the inverse distribution function of X, $\Pr(H(X) < t) \le t$. The last term in the above inequalities does not depend upon l, thus

$$\sum_{l=0}^{\infty} \Pr(nH(\mu_i)^{L_i} < \varepsilon | L_i = l) \Pr(L_i = l) \le \frac{\varepsilon}{n} \sum_{l=0}^{\infty} \Pr(L_i = l) = \frac{\varepsilon}{n}.$$

Hence,

$$\mathbb{E}\left(\sum_{i=1}^{N} X_i | N = n\right) \le \varepsilon.$$

This finally implies $\mathbb{E}\left(\sum_{i=1}^{N} X_i\right) \le \varepsilon$, which exactly means that the expected number of meaningful curves is less than ε. □

3.5 Building Statistically Independent Features

We now return to the matching problem. Why is it so important to consider independent features (cf. Assump. **(A)** in Def. 4)? The reason is the following one: using independent features is a way to beat the *curse of dimensionality* [6]. By combining a few independent features, we can easily reach very low numbers of false alarms without needing huge databases to estimate the probability of false alarms. In his pioneering work, D. Lowe [26] presents this same viewpoint for visual recognition: "*Due to limits in the accuracy of image measurements (and possibly also the lack of precise relations in the natural world) the simple relations that have been described often fail to generate the very low probabilities of accidental occurrence that would make them strong sources of evidence for recognition. However, these useful unambiguous results can often arise as a result of combining tentatively-formed relations*

to create new compound relations that have much lower probabilities of accidental occurrence".

Let us give a numerical example. If the considered database is made of N shape elements, the lowest value reachable by each empirical probability,

$$\hat{P}_i(\mathcal{S}, d) = \frac{1}{N} \cdot \#\left\{\mathcal{S}' \in \mathcal{B}, d_i(x_i(\mathcal{S}'), x_i(\mathcal{S})) \leq d\right\},$$

is at least $1/N$. Consequently, if the background model is built on $K = 1$ feature, and the database is made of $N = 1000$ shapes, then the lowest reachable number of false alarms would be $1000 \cdot 1/1000 = 1$. This means that even if two shape elements $\mathcal{S}$ and $\mathcal{S}'$ are almost identical, based on the NFA we cannot ensure that this match is not casual. Indeed, an NFA equal to 1 means that, on the average, one of the shape elements in the database can match $\mathcal{S}$ by chance. Assume now that the background model is built on $K = 6$ features (and still $N = 1000$), then the lowest reachable number of false alarms would be $1000 \cdot 1/1000^6 = 10^{-15}$.

In practice, we observe number of false alarms between similar shapes as low as 10^{-10}. This means that such matches would still be meaningful in a database 10^{10} times larger.

To summarize, in our framework and in order to be reliable for the shape recognition task, shape features have to meet the three following requirements:

1) Features provide a complete description: two shapes with the same features are alike (so that shapes are accurately described).
2) Distances between features are mutually independent (so that we may design the background model).
3) Their number is as large as possible (so that we may reach low NFAs).

Finding features that meet these three requirements together is a hard problem. Indeed, there must be enough features in order that the first requirement is valid, but not too many otherwise the second requirement fails.

The decision framework we have been describing so far is actually completely general, in the sense that it can be applied to find correspondences between any kind of structures for which K statistically independent features can be extracted. In the following section, we concentrate on the problem of extracting independent features from the normalized shape elements defined in Sect. 2.

3.6 From Normalized Shape Elements to Independent Features

In this section, we explain the procedure we apply to extract independent features from normalized shape elements. We empirically found that the best trade-off achieving simultaneously the three feature requirements pointed out in Sect. 3.5 is the following (see Fig. 6 for an illustration). Each piece of Jordan curve C is split into five subpieces of equal length. Each one of these pieces is normalized by mapping the chord between its first and last points on the horizontal axis, the first point

being at the origin: the resulting "normalized small pieces of curve" are five features C_1, C_2, ... , C_5 (each of those C_i being discretized with 9 points). These features ought to be independent; nevertheless, C_1, ... , C_5 being given, it is impossible to reconstruct the shape they come from. For the sake of completeness a sixth global feature C_6 is therefore made of the endpoints of the five previous pieces, in the normalized frame. For each piece of level line, the shape features introduced in Sect. 3.1 are made of these six 'generic' shape features $C_1, \dots, C_6$. Using the notations introduced in the previous sections, we have $x_i(\mathcal{S}) = C_i$ $(i \in \{1, \dots, 6\})$. For every $i \in \{1, \dots, 5\}$ and $E_i = (\mathbb{R}^2)^9$, $E_6 = (\mathbb{R}^2)^6$.

It now remains to define similarity measures d_i. As mentioned earlier, since the distance d between shape elements is defined as the maximum over the d_i's, we must choose distances having the same range. This will for example not be the case with L^∞-distances (the range of the L^∞ distance between the features C^6 is clearly not the same as for the other features). We choose the following normalization for the d_i's:

$$d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) = \Pr(\mathcal{S}'' \in \Omega \text{ s.t. } ||x_i(\mathcal{S}) - x_i(\mathcal{S}'')||_\infty \leq ||x_i(\mathcal{S}) - x_i(\mathcal{S}')||_\infty).$$

Note that the d_i's are not symmetrical. This normalization yields the following result, whose proof is left as an exercise.

Proposition 4. *Assume that, for $i = 1, \dots, K$, the distribution functions $\delta \mapsto \Pr(||x_i - x||_\infty \leq \delta)$ are invertible, then*

$$NFA(\mathcal{S}, \mathcal{S}') = N \left(\max_i d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \right)^K. \tag{5}$$

In practice, NFA's are computed thanks to formula (5), the d_i's being computed using empirical frequencies:

$$d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) = \frac{1}{N} \#\{\mathcal{S}'' \in \mathcal{B} \text{ s. t. } ||x_i(\mathcal{S}) - x_i(\mathcal{S}'')||_\infty \leq ||x_i(\mathcal{S}) - x_i(\mathcal{S}')||_\infty\},$$

where as before N is the cardinality of the database.

Another possibility that we have investigated is to use a principal component analysis (PCA) [34]. Although PCA does not provide independent features but uncorrelated ones, the computation of the number of false alarms appears to be still valid. However, results are not as good as they should be. PCA indeed suffers from an inherent drawback: it is correct under the strong assumption that the feature space is linear. This is clearly not true for the space of shapes. We experimentally observed that the presented independent features extraction is much more reliable (in the sense that meaningful matches actually mostly correspond to shape elements coming from the same "object").

4 Testing the Background Model

The computation of the probability that a shape element fall by chance at a distance lower than δ to a query shape $\mathcal{S}$ is correct under the independence assumption **(A)**

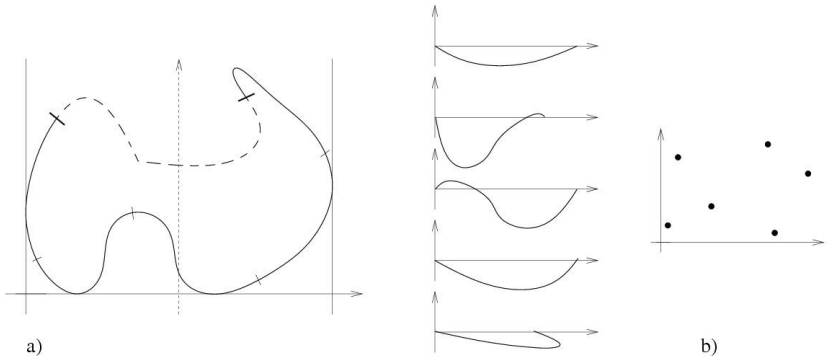


Fig. 6. Building independent features. Sketch a): example of a Jordan curve in a normalized frame based on a bitangent line. The bold part corresponds to a shape element; it is split into 5 pieces C_1, C_2, C_3, C_4, and C_5. Sketch b): each piece is normalized and a sixth feature C_6 made of the endpoints of these pieces is added.

on the distances between features. Of course, the accuracy of the Number of False Alarms NFA$(\mathcal{S}, \delta)$ (Defs. 5 and 8) strongly relies on this independence assumption. In most experiments, we are not able to objectively separate false alarms and correct matches, so that we cannot check whether the NFA is effectively bounded by ε. Now, Helmholtz principle [15] states that no detection in "noise" (the definition of which has to be precised) should be considered as relevant. All ε-meaningful matches in the noise should thus be considered as false alarms and therefore in such a situation there should on the average be about ε of them.

In this section we test the proposed procedure on shape element features that are extracted from a white noise image. Observe that in this case the background model (independence of features) is not necessarily true. On the one hand, shape elements correspond to pieces of Jordan curves, and consequently are constrained not to self-intersect. On the other hand, shape element features derive from a normalization procedure (as explained in Sect. 2) which introduces some structural similarities (for example, shape elements coming from bitangent points show mostly common structures). Table 1 quantifies the "amount of dependency" due to these two aspects. We can see that the observed number of matches in white noise images is indeed very near to ε and does not depend on the database size.

Table 1. Normalized pieces of white noise level lines. Average (over 1000 queries) number of ε-meaningful detections versus ε, for databases of various sizes.

value of ε:	0.01	0.1	1	10	100	1,000	10,000
100,000 shape elements	0.09	0.77	3.38	19.98	134.71	1,073.23	9,777.80
50,000 shape elements	0.07	0.45	2.45	17.19	123.07	1,038.41	9,771.81
10,000 shape elements	0.08	0.31	2.1	13.41	107.18	980.43	9,997.85

5 Experiments

In this chapter, we present several experiments that illustrate the a contrario decision methodology applied to the normalization of level lines explained in Sect. 2. A "query image" and a "database image" being given, meaningful level lines from each of them are encoded. Then, 1-meaningful matches (in the sense of Def. 8) are highlightened along the following experiments. More experiments can be seen in [33] and [40].

Although images and pieces of level lines superimposed to images are shown, the reader should keep in mind that the decision rule actually only deals with *normalized shape elements*. However, the results for the corresponding pieces of level lines ("denormalized" shape element in some sense) are shown here for the sake of clarity.

What we call "false matches" along the following sections are in fact meaningful matches that do not correspond to the same "object" (in the broadest sense). Only an a posteriori examination of the meaningful matches enables to distinguish them from matches which are semantically correct. As explained in the previous sections, we actually only detect matches that are not likely to occur by chance, or more precisely speaking, matches that are not expected to be generated more than once by the background model (by fixing the NFA threshold to 1). We experimentally observed that most of the time, false matches have an NFA larger than 10^{-1}. If we are concerned with very sure detections, we simply set the NFA threshold to 10^{-1}.

All the following experiments have been performed with the same values for the parameters of the encoding stage ($M = 45$ and $F = 5$, see Section 2.3) and by thresholding the NFAs to 1 (both for the extraction of meaningful lines and the matching).

5.1 Two Unrelated Images

The aim of this experiment is to check the main property of the proposed method, namely that the Number of False Alarms is an estimation of the expected number of matches that are due to chance. On Fig. 7 one can see two different images (results are representative of what is obtained when considering other images). The similarity invariant normalized shape elements of the meaningful level lines from the first one are searched among the normalized shape elements from the second one. Only one 1-meaningful match is retrieved (i.e. the NFA of this match is below 1). As announced by Prop. 2, one should at most expect about one meaningful match. Although the method does not distinguish between good and false matches, the NFA gives a good estimate on "how good a match is".

5.2 Perspective Distortion

This second experiment illustrates the proposed matching method in the presence of weak perspective distortions. The target image is a photography of Picasso's painting "Les Demoiselles d'Avignon" in a museum and from an angle, whereas the database

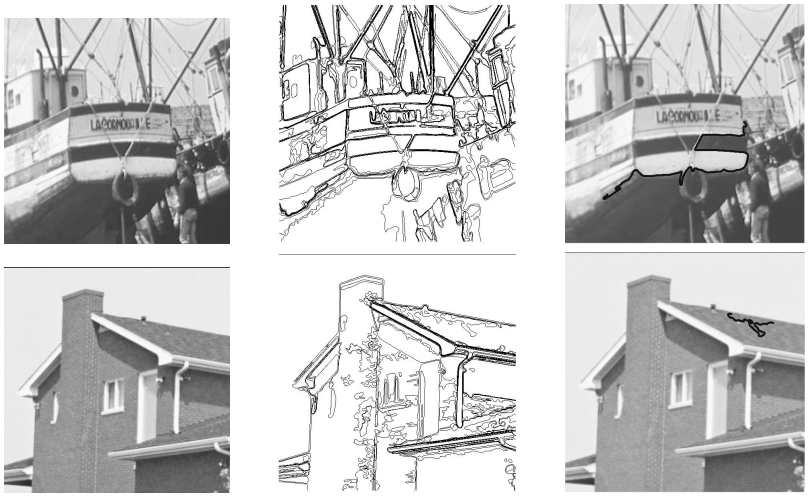


Fig. 7. Two unrelated images. Original images (left), and meaningful level lines (middle). The 846 normalized shape elements from the top image are searched among the 281 normalized shape elements from the bottom image. Only one 1-meaningful match is detected (right). Its NFA is 0.2, which is very near to 1. This match actually corresponds to pieces of level lines that look coarsely alike "by chance".

image is a poster of the same painting. They are shown in Fig. 8, together with the corresponding maximal meaningful boundaries.

Using the affine invariant encoding (detailed in [25]), 1727 and 1595 shape elements were extracted respectively from the target image and from the database image. The number of 1-meaningful matches detected is 12. These 12 matched shape elements are shown, superimposed to images, in Fig. 9. Only one false match is detected, with an NFA of 0.12. The best match has an NFA of $2\,10^{-8}$ and corresponds to the face in the upper right part of the painting. Observe that ideal perfect matches in this experiment would have a number of false alarms of $1727 \times 1595/1595^6 = 1.7\,10^{-13}$ (when the empirical distributions of distances to target codes are learned using the considered database image only, as we do here).

5.3 Logo Recognition

In this experiment, we apply the method with a small logo image as target, and an image containing the logo as database. We use the similarity invariant encoding. The number of target codes is 80, whereas the number of database codes is 8866. As we can see on Figure 10, there are no false matches, even though the database image is complex and relatively cluttered. Recall also that, as illustrated in the experiment of Section 5.1, the method not only enables to locate the logo, but also to decide whether the logo is present or not in an image or a collection of images. We display a similar experiment in Figure 11, where pieces of a street nameplate are sought for.

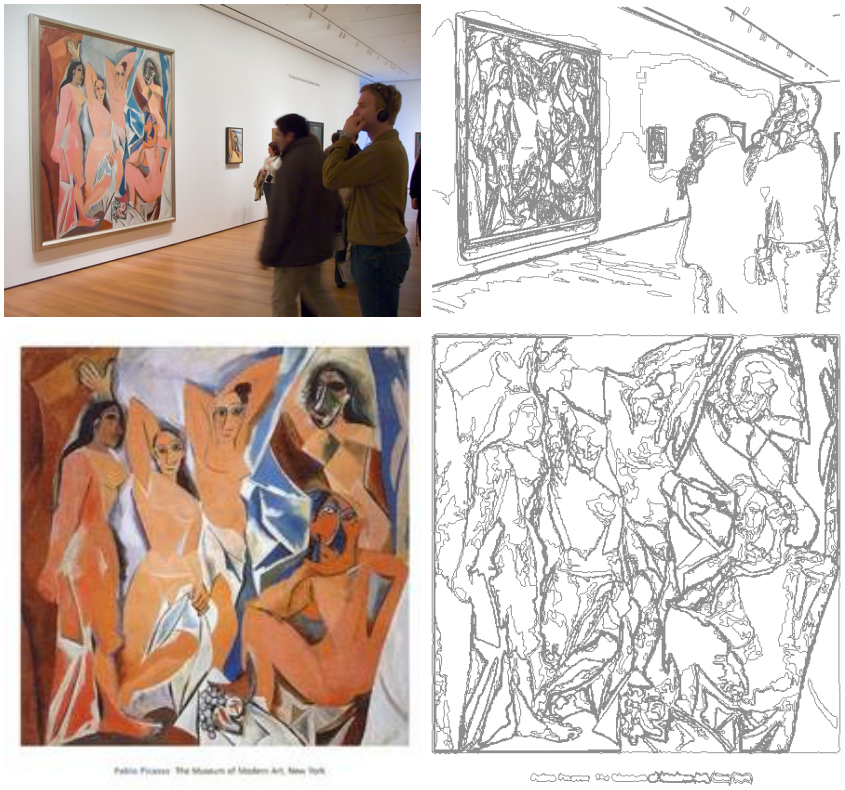


Fig. 8. Les Demoiselles d'Avignon: original images and their corresponding maximal meaningful boundaries to be encoded. The image on top is considered as "target" image. In the target image, 889 level lines are detected, and 212 level lines are detected in the database image.

In this experiment, there are five false matches, and they all have an NFA between 0.1 and 1.

5.4 Dealing with Partial Occlusions and Contrast Changes

The last experiment consists in comparing the codes extracted from two views of Velázquez' painting *Las Meninas* (see Fig. 12). The codes extracted from the query image (11, 332 codes) are searched for among the codes extracted from the database image (12, 833 codes). Shape elements are here normalized with respect to similarity transforms. Note that the target image is a photography which was taken in the museum: visitors' heads hide a part of the painting.

Fig. 9. Affine invariant semi-local recognition method: the 12 meaningful matches between shape elements. Only one false match is detected (other matches correspond to the same "piece of object"), with an NFA of 0.12. The lowest NFA is $2\,10^{-8}$ and corresponds to the contour of the face in the upper right part of the painting.

Fig. 10. The logo on the left is searched for in the image on the right, using a similarity invariant encoding. There are 9 matches, and none of them is false, in the sense that they correspond to the same pieces of the logo. The smallest NFA has a value of $4.4\,10^{-12}$.

Fig. 11. Pieces of the street nameplate on the left are sought for in the right image, using a similarity invariant encoding. The two plates comes from different locations in the street. There are 15 matches. Five of them are false but they all have an NFA between 10^{-1} and 1.

Fig. 13 shows all 55 1-meaningful matches. Only 5 false matches can be seen among them and they all have an NFA between 1 and 10^{-1}. In fact, 36 matches show an NFA lower than 10^{-1}.

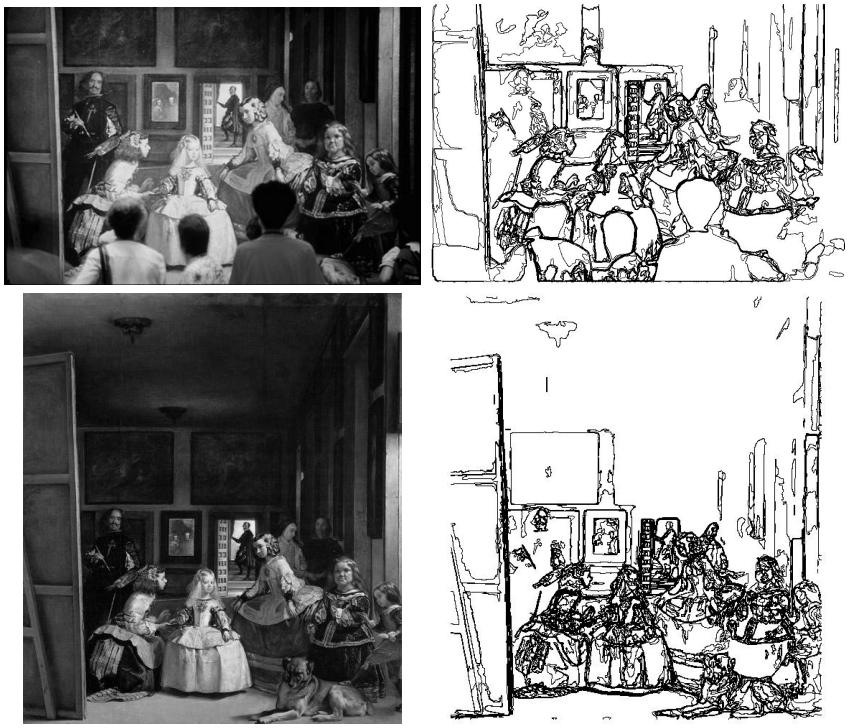


Fig. 12. Las Meninas original images (on the left) and meaningful level lines (on the right). Top: query image and its level lines. Bottom: database image and its level lines. The codes from the query image are sought among the codes from the database image. Normalization is here with respect to similarity transformations.

6 Conclusion and Perspectives

In this chapter, we considered shape elements as pieces of long and contrasted enough level lines. This definition naturally comes from an analysis of the requirements that shape recognition meets, namely robustness to "small" contrast changes, robustness to occlusions, and concentration of the information along contours (i.e. regions where grey level changes abruptly). The purpose of this chapter is to propose a method to compute the Number of False Alarms of a match between some shape elements, up to a given class of invariance. Computing this quantity is useful

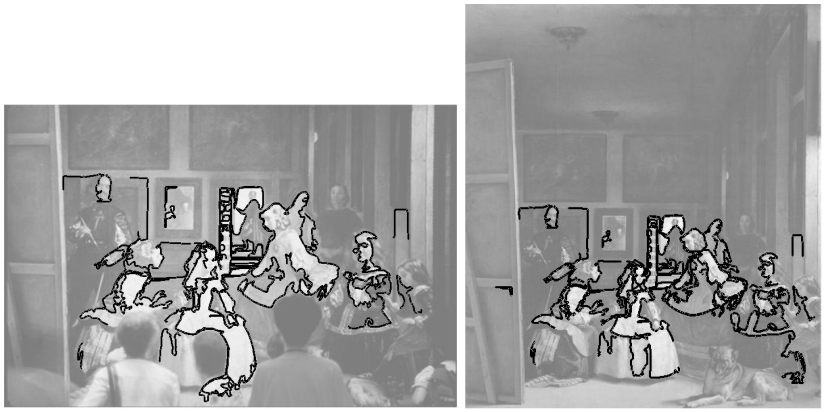


Fig. 13. Las Meninas. The 55 meaningful matches. Half of them have an NFA lower than 10^{-3}. The best match has an NFA equal to 4.10^{-14}. To each bold piece of level line on the right corresponds a bold piece of level line on the left.

because it leads to an acceptance/rejection threshold for partial shape matching. The proposed decision rule is to keep in consideration the matches with an NFA lower than 1 (or 10^{-1} if we are concerned with "surer" detections). This automatically yields a distance threshold that depends on both the database and the query.

Of course, dealing only with pieces of level lines is not enough to decide whether an object is present or not in a given image. Nevertheless, object edges coincide well with pieces of level lines, so that it is worth taking them into account. A further step should thus combine the matches, by taking account of their spatial coherence. Indeed, as we can see in the experiments we have presented, false matches (i.e. matches that do not actually correspond to the same "object") are not distributed over the images in a conspicuous way, unlike "good" matches. Each pair of matching shape elements leads to a unique transformation between images, which can be represented as a pattern in a transformation space. Hence, spatially coherent meaningful matches correspond to clusters in the transformation space, and their detection can then be formulated as a clustering problem. To achieve this task, we have developed an unsupervised clustering algorithm, still based on an a contrario model [8]. As noticed on preliminary results [33, 40], combining the spatial information furnished by matched shape elements strongly reinforces the recognition confidence of the method.

Acknowledgement. This work was supported by the Office of Naval Research under grant N00014-97-1-0839, by the Centre National d'Études Spatiales, and by the Réseau National de Recherche en Télécommunications (projet ISII). Algorithms were implemented using the MegaWave2 free software.

References

1. A. Almansa, A. Desolneux, and S. Vamech. Vanishing point detection without any a priori information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):502–507, 2003.
2. L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations of image processing: Multiscale analysis and P.D.E. *Archive for Rational Mechanics and Analysis*, 16(9):200–257, 1993.
3. Y. Amit, D. Geman, and X. Fan. A coarse-to-fine strategy for multiclass shape detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12):1606–1621, 2004.
4. N. Arnaud, F. Cavalier, M. Davier, and P. Hello. Detection of gravitational wave bursts by interferometric detectors. *Physical review D*, 59(8):082002–1 – 082002–9, 1999.
5. F. Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183–193, 1954.
6. R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.
7. F. Cao. Application of the Gestalt principles to the detection of good continuations and corners in image level lines. *Computing and Visualisation in Science*, 7(1):3–13, 2004.
8. F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur. An *a contrario* approach to clustering and validity assessment. Preprint CMLA No 2004-13.
9. F. Cao, P. Musé, and F. Sur. Extracting meaningful curves from images. *Journal of Mathematical Imaging and Vision*, 2004. To appear.
10. V. Caselles, B. Coll, and J.-M. Morel. Topographic maps and local contrast changes in natural images. *International Journal of Computer Vision*, 33(1):5–27, 1999.
11. P.B. Chapple, D.C. Bertilone, R.S. Caprari, and G.N. Newsam. Stochastic model-based processing for detection of small targets in non-gaussian natural imagery. *IEEE Transactions on Image Processing*, 10(4):554–564, 2001.
12. A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful alignments. *International Journal of Computer Vision*, 40(1):7–23, 2000.
13. A. Desolneux, L. Moisan, and J.-M. Morel. Edge detection by Helmholtz principle. *Journal of Mathematical Imaging and Vision*, 14(3):271–284, 2001.
14. A. Desolneux, L. Moisan, and J.-M. Morel. A grouping principle and four applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):508–513, 2003.
15. A. Desolneux, L. Moisan, and J.-M. Morel. *A theory of digital image analysis*. 2004. Book in preparation.
16. P.A. Devijver and J. Kittler. *Pattern recognition - A statistical approach*. Prentice Hall, 1982.
17. L.C. Evans and R.Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, 1992.
18. K.J. Friston. Introduction: Experimental design and statistical parametric mapping. In R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, K.J. Friston, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny, editors, *Human Brain Function*. Academic Press, 2nd edition, 2003.
19. Y. Gousseau. Comparaison de la composition de deux images, et application à la recherche automatique. In *proceedings of GRETSI 2003*, Paris, France, 2003.
20. W.E.L. Grimson and D.P. Huttenlocher. On the verification of hypothesized matches in model-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(12):1201–1213, 1991.

21. G. Kanizsa. *La Grammaire du Voir*. Diderot, 1996. Original title: *Grammatica del vedere*. French translation from Italian.
22. Y. Lamdan, J.T. Schwartz, and H.J. Wolfson. Object recognition by affine invariant matching. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 335–344, Ann Arbor, Michigan, U.S.A., 1988.
23. M. Lindenbaum. An integrated model for evaluating the amount of data required for reliable recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1251–1264, 1997.
24. J.L. Lisani. *Shape Based Automatic Images Comparison*. PhD thesis, Université Paris 9 Dauphine, France, 2001.
25. J.L. Lisani, L. Moisan, P. Monasse, and J.-M. Morel. On the theory of planar shape. *SIAM Multiscale Modeling and Simulation*, 1(1):1–24, 2003.
26. D.G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publisher, 1985.
27. D. Marr. *Vision*. Freeman Publishers, 1982.
28. G. Matheron. *Random Sets and Integral Geometry*. John Wiley and Sons, 1975.
29. L. Moisan. Affine plane curve evolution: A fully consistent scheme. *IEEE Transactions on Image Processing*, 7(3):411–420, 1998.
30. L. Moisan and B. Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal on Computer Vision*, 57(3):201–218, 2004.
31. P. Monasse. *Représentation morphologique d'images numériques et application au recalage, Morphological Representation of Digital Images and Application to Registration*. PhD thesis, Université Paris 9 Dauphine, France, 2000.
32. P. Monasse and F. Guichard. Fast computation of a contrast invariant image representation. *IEEE Transactions on Image Processing*, 9(5):860–872, 2000.
33. P. Musé. *On the definition and recognition of planar shapes in digital images*. PhD thesis, École Normale Supérieure de Cachan, 2004.
34. P. Musé, F. Sur, and J.-M. Morel. Sur les seuils de reconnaissance des formes. *Traitement du Signal*, 20(3):279–294, 2003.
35. C. Orrite, S. Blecua, and J.E. Herrero. Shape matching of partially occluded curves invariant under projective transformation. *Computer Vision and Image Understanding*, 93(1):34–64, 2004.
36. C.A. Rothwell. *Object Recognition Through Invariant Indexing*. Oxford Science Publications, 1995.
37. G. Sapiro and A. Tannenbaum. Affine invariant scale-space. *International Journal of Computer Vision*, 11(1):25–44, 1993.
38. J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
39. S.D. Silvey. *Statistical Inference*. Chapman and Hall, 1975.
40. F. Sur. *A contrario decision for shape recognition*. PhD thesis, Université Paris Dauphine, 2004.
41. R. Veltkamp and M. Hagedoorn. State-of-the-art in shape matching. In M.S. Lew, editor, *Principles of Visual Information Retrieval*, volume 19. Springer Verlag, 2001.
42. M. Wertheimer. Untersuchungen zur Lehre der Gestalt, II. *Psychologische Forschung*, (4):301–350, 1923. Translation published as Laws of Organization in Perceptual Forms, in Ellis, W. (1938). A source book of Gestalt psychology (pp. 71-88). Routledge & Kegan Paul.
43. D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.