An *a contrario* decision method for shape element recognition

Pablo MUSÉ[1], Frédéric SUR[1,4], Frédéric CAO[2], Yann GOUSSEAU[3], Jean-Michel MOREL[1]

Abstract: Shape recognition is the field of computer vision which addresses the problem of finding out whether a query shape lies or not in a shape database, up to a certain invariance. Most shape recognition methods simply sort shapes from the database along some similarity measure to the query shape. Their Achilles' heel is the decision stage, which should aim at giving a clear-cut answer to the question: "do these two shapes look alike?" In this article, the proposed solution consists in bounding the number of false correspondences of the query shape among the database shapes, ensuring that the obtained matches are not likely to occur "by chance". As an application, one can decide with a parameterless method whether any two digital images share some shapes or not.

Keywords: Shape recognition, *a contrario* decision, background model, Number of False Alarms, meaningful matches.

Contents

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

1 Introduction

Recognition is the ability to identify, based on prior knowledge. Visual recognition, in particular, is the process of finding correspondences between new elements and elements which have been previously seen, at least once, and live in our "world of images". In this work, we will focus on the problem of visual recognition based upon geometrical shape information. Shape recognition methods usually consist of three stages: feature extraction, matching (the core of this stage is the definition of a distance or dissimilarity measure between features describing shapes) and decision. The first two stages have been widely addressed in the literature (see for instance [38] or [42] and references therein), and will be discussed in Section 3. On the contrary, the decision problem for shape matching has been rarely studied, especially in a generic framework. Once two shapes are likely to match, how is it possible to come to a decision? The purpose of this article is not to propose a new shape recognition procedure, but to define statistical criteria leading to decide whether two shapes are alike or not.

In computer vision, extraction of shape information from images dates back to Marr [23], but Attneave [5], as well as Wertheimer [40] 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). Shapes can then be modeled as Jordan curves. However, as pointed out by Kanisza [18], in every day's vision most objects are partially hidden by other ones, and despite this occlusion phenomenon humans can still 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. The information regarding how shape elements are extracted from images is not necessary for the moment; we will just assume that the set of shape elements extracted from an image provides a suitable representation of its shape contents. Let us moreover point out that we do not address in this paper the recognition of shapes as a whole, which can be performed by integrating the recognized shape elements, based on spatial coherence [7].

When shapes are subject to weak perspective distortions, human perception is still able to recognize them. In order to be compared, shape representations should thus be invariant to these transformations. However in general, projective transformations have been shown not to behave well with regard to shape matching, because they permit to map a large class of curves arbitrarily close to a circle, and thus to map a finite number of curves arbitrarily close to a given curve [4]. Projective transformations can be locally approximated by affine transformations, and these approximations are particularly accurate under weak perspective distortions. Since shape elements are supposed to be quite local, an affine invariant representation of shape elements meets the geometric invariance requirement of shape representation. For a large class of applications, similarity invariance could even be enough. Consequently, a possible approach consists in representing each shape element $\mathcal{S}$ by a list of K affine or similarity invariant descriptors, $x_1(\mathcal{S}), \ldots, x_K(\mathcal{S})$, which we will call a *code*.

Having a shape representation which is consistent with the perceptual principles that guide recognition enables us to address the shape correspondence problem. Determining correspondences between shape elements not only means defining a notion of similarity between them, but also being able to decide whether the two shape elements are to be paired or not. The main goal of this paper is to propose a general framework that enables to reach that kind of decisions by introducing an automatic decision rule. We will apply this methodology to the shape matching problem, but the scope of this methodology is much wider since the principles that are used are general. Up to our knowledge, a generic acceptance / rejection decision method for shape matching has not yet been proposed. In general, matches with a query shape are, at most, only ranked (for example along a distance [14], along some probability [35], or along some number of votes in hashing methods [41]).

Let us specify what we mean by an "automatic decision rule" for shape matching. 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 extracted 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. Following a series of articles by Desolneux, Moisan, and Morel (see for example [10, 11, 12]), such a methodology is called *a contrario* decision. In computer vision, the first attempts to detect events in images *a contrario* to a random situation certainly date back to David Lowe's work on perceptual organization. In [22], Lowe studies whether a configuration of points shows some

intrinsic structure: "*[. . .] any relations which arise through some accident of viewpoint or position are of no use for recognition and will only confuse the interpretation process. This fact will provide the basic method for evaluating the usefulness of specific image relations – relations are useful only to the extent that they are unlikely to have arisen by accident*".

Several works on target detection follow the same principles. Olson and Huttenlocher [29] present a method for automatic target recognition under similarity invariance. Objects and images in which the objects are sought are encoded by oriented edges, and compared by using a relaxed Hausdorff distance. The authors give an estimate of the probability of a false alarm occurring over the entire image, which is used to take a decision. Let us quote the authors:"*One method by which we could use the estimate is to set the matching threshold such that the probability of a false alarm is below some predetermined probability. However, this can be problematic in very cluttered images since it can cause correct instances of targets that are sought to be missed.*" Grimson and Huttenlocher [16] propose to fix a threshold on the proportion of model features (edges) among image features (considered in the transformation space) upon which the detection is sure. Their main assumption is that the features are uniformly distributed; this "background model", according to the terminology we will soon define, governs random situations. This framework allows the authors to estimate the probability that a cluster in the feature space is due to the *"conspiracy of random"* in their words. Fixing a threshold on this probability gives sure detections: rare events are the most significant ones. The ideas developed in [16] inspired several works [1, 30]. Following Huttenlocher and Grimson's work, Pennec [32] presents a method to compute the intrinsic false alarm rate of commonly used methods such as Geometric Hashing and Generalized Hough Transform, by incorporating the uncertainty of measurements. The proposed computation relies on several limitative assumptions (*a priori* shape model, uniform distribution of features), as in Huttenlocher and Grimson's work. As pointed out by Pennec: "*[These limitations] are hardly ever verified in real cases. For instance in medical images of the head, extremal points are not uniformly distributed in the image, but more or less uniformly distributed on the surface of the brain and the skull [. . .]. A very interesting extension would be to compute the probability of false positives* online*, during the recognition itself. This would allow us to take into account the specific distribution of the model and scene features.*" This is precisely what we aim at.

Other examples illustrating the *a contrario* decision methodology can be found among the literature about detection of low resolution targets over a cluttered background (see for example [9] or [39]). A probabilistic model for the background over which the sought objects lie is first built, then objects are detected if they are not likely to be generated by the background.

In the following, we intend to make such probabilistic methods reliable for the shape correspondence problem, and we propose a method to automatically compute the right matching thresholds. Instead of defining a threshold distance for each query shape, we define a quantity (namely the Number of False Alarms) that can be thresholded independently of the query shape. This quantity can be interpreted as the expected number of random shapes at some given distance from a query shape. Even if thresholding this number naturally leads to threshold the matching distance, we show that we get an additional information about how likely the matching is, and therefore about how sure we are that the matching is correct.

A preliminary, less efficient version of the method presented here was also proposed in [27] and [28].

The plan of this article is as follows. In Section 2, we tackle the general problem of deciding whether two *shape elements*, represented by affine or similarity invariant *codes*, match or not. We introduce the notion of *meaningful match*. This concept enables to rank matches with a given shape element by a criterion which is given an accurate and handy meaning: the Number of False Alarms (NFA). However, contrarily to most existing methods, not only does the NFA enable us to rank candidate matches, but a detection threshold which adapts to the query shape and to the database is also derived from a uniform bound over this NFA. In Section 3 the presented decision methodology is specified for shape recognition in digital images; following works by Desolneux *et al.* [11] on boundaries extraction (improved in [8]) and by Lisani *et al.* [20, 21] on curve normalization, *shape elements* are extracted from images, then matching is performed and followed by the presented decision process. In Section 4, we present some experiments that show the validity of the proposed model. It is verified that the methodology satisfies Helmholtz principle [12]: 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 Section 5. We conclude in Section 6.

2 An *a contrario* decision framework

The aim of this section is to present a method to fix an acceptance/rejection threshold for the recognition of *shape elements*, up to a given class of invariance. Each shape element is described by a *code* (a list of invariant features belonging to some feature space) as mentioned in the previous section. 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 the 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 [10] or contrasted edges [11], by Almansa *et al.* to the detection of vanishing points [2], by Stival and Moisan to stereo images [25] , by Gousseau to the comparison of image "composition" [15] and by Cao to the detection of good continuations [6]. The main advantage of this technique is that the only parameter which controls the detection is the Number of False Alarms, that will be defined in Section 2.3. We present here the *a contrario* decision methodology in terms of hypothesis testing in order to link the number of false alarms first defined by Desolneux *et al.* and the probability of false alarms introduced in usual hypothesis testing theory.

2.1 Shape model *versus* background model

Let us precisely define the shape element representation and give 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 each shape element S to be represented by a set of K features $x_1(S)$, $x_2(S)$, ..., $x_K(S)$, each of them belonging to a metric space (E_i, d_i) ($i \in \{1, \ldots, K\}$). We define a distance between shape elements as the product distance over $E_1 \times E_2 \times \cdots \times E_K$, that is

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

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 $\mathcal{S}$. Having such a model would imply an extra knowledge (for instance some "expert" should have first built up the models). 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* $\mathcal{S}$.

We are therefore 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 the shape elements belong to some probability space $(\Omega, \mathcal{A}, \overline{\Pr})$ (which will not be explicitly precised) such that the following definition is valid.

Definition 1 *We call* background model *any random model* $(\Omega, \mathcal{A}, \overline{\Pr})$ *such that the following assumption holds:*

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

For every $i \in \{1, \ldots, K\}$, the probability $\overline{\Pr}(\mathcal{S}' \in \Omega, \text{ s.t. } d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \leq \delta)$ is denoted by $P_i(\mathcal{S}, \delta)$. In the remainder of this article, the empirical frequency

$$\frac{1}{N} \cdot \#\{\mathcal{S}' \in \mathcal{B}, d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \leq \delta\}$$

(where $\#\cdot$ denotes the cardinality of any finite set and N is the cardinality of the database $\mathcal{B}$) is taken as an estimator of $P_i(\mathcal{S}, \delta)$ for every δ. That is, we use in practice $P_i(\mathcal{S}, \delta) = \frac{1}{N} \cdot \#\{\mathcal{S}' \in \mathcal{B}, d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) \leq \delta\}$.

2.2 An *a contrario* decision methodology

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. We propose to use the hypothesis testing framework [13, 36] in order to replace the distance bound by a probability of false alarms bound.

A shape $\mathcal{S}'$ being observed, the hypothesis we are interested in is $\mathcal{H}_0$: "$\mathcal{S}'$ has been generated by the shape model of $\mathcal{S}$". However, handling this hypothesis with our assumption (no available shape model for $\mathcal{S}$) is simply impossible. We are therefore led to concentrate on the alternative hypothesis $\mathcal{H}_1$: "$\mathcal{S}'$ has been generated by the background model". For each δ, the set of the shape elements is split into two subsets $\Omega_0(\delta)$ and $\Omega_1(\delta)$, respectively made of the shape elements whose distance to $\mathcal{S}$ is lower than δ (and for which hypothesis $\mathcal{H}_0$ is accepted), and of those for which the distance to $\mathcal{S}$ is larger than δ (for which hypothesis $\mathcal{H}_0$ is rejected).

Definition 2 *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 $\mathcal{H}_0$ is accepted ($\mathcal{S}'$ is near $\mathcal{S}$ because of some causality). In this case, $\mathcal{S}'$ is classified in $\Omega_0(\delta)$.*
- *Otherwise, $\mathcal{H}_0$ is rejected and the alternative hypothesis $\mathcal{H}_1$ is accepted ($\mathcal{S}'$ is near $\mathcal{S}$ casually). In this case, $\mathcal{S}'$ is classified in $\Omega_1(\delta)$.*

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

- The *probability of non-detection* or *probability of a miss* (associated with type I error) $\alpha' = \Pr(\Omega_1(\delta)|\mathcal{H}_0)$;
- The *probability of false alarms* (associated with type II error) $\alpha = \Pr(\Omega_0(\delta)|\mathcal{H}_1)$;

provided $\Pr(\cdot|\mathcal{H}_0)$ (resp. $\Pr(\cdot|\mathcal{H}_1)$) is the likelihood of $\mathcal{H}_0$ (resp. $\mathcal{H}_1$) over Ω.

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 widely used techniques for doing this are the *likelihood ratio test* and the *Bayesian test*. However, the practical limits of this theoretical framework are obvious. They indeed need that one knows the likelihood of both the hypothesis $\mathcal{H}_0$ and the counter-hypothesis $\mathcal{H}_1$, 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 $\mathcal{H}_0$). Moreover, the Bayesian approach needs prior information, which remains either spoilt by arbitrariness, or is strongly related to a specific problem for which supplementary information is provided.

2.2.1 Estimating the probability of false alarms

Let us summarize the situation. We are not able to compute the probability of non-detection $\Pr(\Omega_1(\delta)|\mathcal{H}_0)$. On the other hand, a straightforward computation provides the value of the probability of false alarms of the statistical test $\mathcal{T}_\delta(\mathcal{S})$, denoted by $\text{PFA}(\mathcal{S},\delta) := \Pr(\Omega_0(\delta)|\mathcal{H}_1)$. Since $\mathcal{S}' \in \Omega_0(\delta)$ if and only if $d(\mathcal{S},\mathcal{S}') \leq \delta$, it follows that

$$\begin{aligned}\text{PFA}(\mathcal{S},\delta) &= \Pr\left(\mathcal{S}' \in \Omega \text{ s.t. } d(\mathcal{S},\mathcal{S}') \leq \delta | \mathcal{H}_1\right) \\ &= \Pr\left(\mathcal{S}' \in \Omega \text{ s.t. } \max_{i\in\{1,\ldots,K\}} d_i(x_i(\mathcal{S}),x_i(\mathcal{S}')) \leq \delta \mid \mathcal{H}_1\right).\end{aligned}$$

Now, by the definition of $\mathcal{H}_1$, $\Pr(\cdot|\mathcal{H}_1) = \overline{\Pr}(\cdot)$, thus

$$\text{PFA}(\mathcal{S},\delta) = \overline{\Pr}\left(\mathcal{S}' \in \Omega \text{ s.t. } \max_{i\in\{1,\ldots,K\}} d_i(x_i(\mathcal{S}),x_i(\mathcal{S}')) \leq \delta\right).$$

Assumption **(A)** then yields

$$\begin{aligned}\text{PFA}(\mathcal{S},\delta) &= \prod_{i\in\{1,\ldots,K\}} \Pr\left(\mathcal{S}' \in \Omega, \text{ s.t. } d_i(x_i(\mathcal{S}),x_i(\mathcal{S}')) \leq \delta\right) \\ &= \prod_{i\in\{1,\ldots,K\}} P_i(\mathcal{S},\delta). \qquad (1)\end{aligned}$$

Therefore, we have just proved the following proposition.

Proposition 1 *The probability of false alarms of the statistical test $\mathcal{T}_\delta(\mathcal{S})$ is*

$$\textit{PFA}(\mathcal{S},\delta) = \prod_{i\in\{1,\ldots,K\}} P_i(\mathcal{S},\delta).$$

We recall that, in numerical experiments, we will use the estimator for $P_i(\mathcal{S},\delta)$ ($i \in \{1,\ldots,K\}$)

$$P_i(\mathcal{S},\delta) = \frac{1}{N} \cdot \#\{\mathcal{S}' \in \mathcal{B}, d_i(x_i(\mathcal{S}'),x_i(\mathcal{S})) \leq \delta\}.$$

2.2.2 Deriving an *a contrario* decision rule

Now, the next step is to bound the PFA. Indeed, the probability of false alarms $\text{PFA}(\mathcal{S}, \delta)$ being non-decreasing with δ, an upper bound p on this quantity immediately provides an upper bound δ^* over the distances, namely

$$\delta^*(p) = \max\{\delta > 0, \text{PFA}(\mathcal{S}, \delta) < p\}.$$

Consequently, if the test is to accept $\mathcal{H}_0$ if the observed distance is below $\delta^*(p)$, and to reject this hypothesis otherwise, then the associated probability of false alarms is bounded by p. This rule is said to be an *a contrario* decision since we accept the null hypothesis as soon as the alternative hypothesis is not likely to be valid (*i.e.* the probability of false alarms of the associated statistical test is very low). Applied here to the shape recognition problem, 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". Notice that, according to this decision, all we are saying is that, under the background model, such a coincidence is so astonishing that there must be a better explanation than randomness. We are by no means asserting that this better explanation is "matched shape elements correspond to instances of the same object", though this might be the cause, among other possibilities. Experiments (see Section 5) indeed show matched shapes that are actually alike, but do not correspond to the same object.

2.3 A detection terminology

2.3.1 Number of False Alarms

The *a contrario* decision that has just been introduced consists in fixing a threshold over the probability of false alarms rather than over the distance between shape elements. Since a probability has little meaning *per se*, we now introduce the *number of false alarms*. Let us recall that we are interested in a situation in which a query shape element is compared to shape elements from a database of size N.

Definition 3 *The* Number of False Alarms *of the shape element $\mathcal{S}$ at a distance d is*

$$\begin{aligned} \textit{NFA}(\mathcal{S}, d) &:= N \cdot \textit{PFA}(\mathcal{S}, d) \\ &= N \cdot \prod_{i \in \{1,\dots,K\}} P_i(\mathcal{S}, d). \end{aligned}$$

Since the latest product of probabilities is the probability of false alarms when testing if the database shape elements are at a distance lower than d to $\mathcal{S}$ (*cf* Equation (1)), the number of false alarms can be seen as the average number of false alarms that are expected when we test whether the distance from each shape element in the database to $\mathcal{S}$ is below d.
Remark: Since each P_i is empirically estimated over the database $\mathcal{B}$, the NFA also depends on $\mathcal{B}$.

Definition 4 *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 $d(\mathcal{S}, \mathcal{S}')$:*

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

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: 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.

2.3.2 Meaningful matches

Instead of directly bounding the probability of false alarms in order to deduce a distance threshold (as explained in Section 2.2.2), we bound the number of false alarms, since this quantity that has an interpretation in terms of expected frequencies.

Definition 5 *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:*

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

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 $\text{NFA}(\mathcal{S}, d) := N \cdot \prod_{i \in \{1,\ldots,K\}} P_i(\mathcal{S}, d)$ is pseudo-invertible with respect to d. That is, there exists a unique positive real number $\delta^*(\varepsilon/N)$ (also depending on the query shape $\mathcal{S}$) such that

$$\delta^*(\varepsilon/N) := \max\{\delta > 0, \text{PFA}(\mathcal{S}, \delta) \leq \varepsilon/N\}.$$

The proposition that follows is then straightforward.

Proposition 2 *A shape element $\mathcal{S}'$ is an ε-meaningful match of the query shape element $\mathcal{S}$ if and only if*

$$d(\mathcal{S}, \mathcal{S}') \leq \delta^*(\varepsilon/N).$$

The ε-meaningful matches of $\mathcal{S}$ are then those shape elements for which the distance to $\mathcal{S}$ is below $\delta^*(\varepsilon/N)$ (the probability of false alarms of the associated test is consequently less than ε/N). We therefore expect on the average less than ε false alarms among all ε-meaningful matches over the N tested shape elements. This methodology does not enable to estimate the number of ε-meaningful matches. However, if all shape elements in the database were generated by the background model, then hypothesis $\mathcal{H}_0$ should never be accepted, all ε-meaningful detections should thus be considered as false alarms. The following proposition makes this claim more formal.

Proposition 3 *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 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 $\mathcal{S}'_j$ actually 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 $E(R) = \sum_{j=1}^{N} E(\chi_j)$. Using Proposition 2, it follows that

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

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

$$E(\chi_j) = \Pr(d(\mathcal{S}, \mathcal{S}'_j) \leq \delta^*(\varepsilon/N) | \mathcal{H}_1) = \text{PFA}(\mathcal{S}, \delta^*(\varepsilon/N)).$$

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

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

2.3.3 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, the probabilities are in fact empirical frequencies estimated over the database. 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 $d \leq d'$,

$$P_i(\mathcal{S}_1, d) \leq P_i(\mathcal{S}_2, d).$$

This yields $\delta^*_{\mathcal{S}_2} \leq \delta^*_{\mathcal{S}_1}$ (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"

$$P_i^1(\mathcal{S}, d) \leq P_i^2(\mathcal{S}, d),$$

where P_i^1 and P_i^2 are respectively estimated over $\mathcal{B}_1$ and $\mathcal{B}_2$. This yields $\delta_2^*(\mathcal{S}) \leq \delta_1^*(\mathcal{S})$.

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.

2.3.4 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 been already recognized before somehow or other.

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, and not only a single shape element to a database. This corresponds to the experiments of Section 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 6 *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 (Proposition 3) still holds.

2.4 Building statistically independent features

Now, why is it so important to consider independent features (*cf* **(A)**)? The reason is the following one: using independent features is a way to beat the *curse of dimensionality* [17]. 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 [22] 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,

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

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 that the number of false alarms between two similar shapes can be as low as 10^{-10}. This means that we need to observe a database 10^{10} times larger in order that a meaningful match at the same distance ought to be a false alarm.

To sum up, in our framework, 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.

2) Features are mutually statistically independent (more precisely speaking, distances between features are independent).

3) Their number is as large as possible.

The first requirement means that the features describe shapes well, the second one is imposed in order to design the background model, and the third requirement is needed in order to reach low numbers of false alarms. 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 falls.

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 pieces of Jordan curves (the *shape elements*). Shape elements are normalized before comparison in order to meet the geometric invariance requirement of recognition (see Section 3.2); therefore we will more specifically deal with *normalized shape elements*, and extract independent features from them (Section 3.3).

3 From images to normalized shape elements to independent features

3.1 Representing shapes by level lines

In this section we discuss how the proposed methodology for decision making can be used in a realistic shape recognition system. An algorithm extracting pieces of Jordan curves corresponding to invariant local representations of shapes in images was proposed by Lisani *et al.* [20, 21]. It proceeds with the following steps:

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

Let us detail and argue each of these steps. Consider the set of level lines in an image (*i.e.* the boundaries of the connected components of its level sets). This representation has several advantages. Although it is not invariant under *scene illumination* changes (in this case the image itself is changed and any descriptor hardly remains invariant), it is invariant under *contrast* changes. The mathematical morphology school has claimed that all shape information is contained in level lines, and this is certainly correct, in the sense that we can reconstruct the whole image from its level lines. Moreover, the boundaries of the objects lying in the image are well represented by the union of some pieces of level lines. Thus, level lines can be viewed as concatenations of pieces of boundaries of objects and therefore encode all shape information.

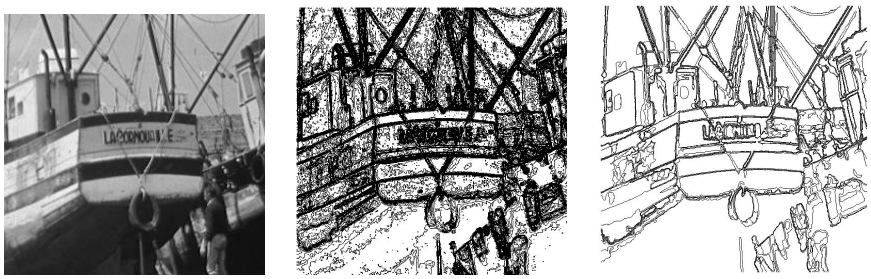
Nevertheless, the representation provided by level lines is highly redundant, and may also contain useless information. That is why, Desolneux *et al.* [11] proposed a method to extract *meaningful* level lines from images, which was later improved by Cao *et al.* [8]. Experimentally, these lines proved to locally coincide with boundaries of perceptually significant objects in images. The algorithm needs no parameter tuning, since parameters are automatically set based on statistical arguments derived from perceptual principles. Meaningful level lines are not contrast invariant, since their detection depends on the contrast distribution in the image. However, it turns out that they are invariant with respect to globally affine contrast changes.

Figure 1 illustrates that the loss of information implied by 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. Otherwise, the presented methodology could not be realistic for applications such as image retrieval from databases.

Once meaningful level lines are extracted, we need to smooth them in order to eliminate noise and aliasing effects. The Geometric Affine Scale Space [3, 34] is fully convenient (since such a smoothing commutes with special affine transformations and since we are interested in affine invariance):

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

where x is a point on a level line, $\text{Curv}(x)$ the curvature and $\vec{n}(x)$ the normal to the curve, oriented towards concavity. We use a fast implementation by Moisan [24]. The scale at which the smoothing is applied is fixed and given by the pixel size. We fix the smoothing scale in order to wipe out details of size one pixel on the curves which are commonly extracted. The aim is to reduce the complexity of meaningful level lines by simplifying them. The final goal remains



(a) (b) (c)

Figure 1: Extraction of meaningful level lines. (a) original "La Cornouaille" image, (b) level lines, represented here with grey-level quantization step equal to 10 (54790 level lines), (c) meaningful level lines (296 detections).

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.

The last stage of the invariant shape encoding algorithm is 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 pieces, or bitangent lines). Such a representation is obtained by uniformly sampling a piece of curve in this normalized frame.

The conjunction of these three stages was first introduced by Lisani *et al.* [20, 21]; the third stage is also based on the seminal work of Lamdan *et al.* [19], followed by Rothwell's work on invariant indexing [33] and more recently by Orrite *et al.* [31]. The following section is devoted to an improvement of Lisani's algorithm.

3.2 Semi-local normalization and encoding

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 pieces (a flat zone is a portion of a curve which is everywhere unexpectedly close to the segment joining its endpoints, relatively to an adequate background model [26, 37]). While bitangency is an affine invariant property, it is not the case for flat pieces. However, two arguments stand for its consideration. The first one is that, under reasonable zoom factors, flat pieces are preserved. The second argument is that inflexion points, which are conserved by affine transformations, are most of the time surrounded by a flat piece, which is by consequent also conserved by affine transformations. If it is not the case, the tangent at the inflexion point will not be a robust direction. In that sense, tangent at flat pieces can also be considered as robust versions of tangents at inflexion points (which Lisani's original algorithm use, together with bitangent lines and a non-robust version of flat pieces).

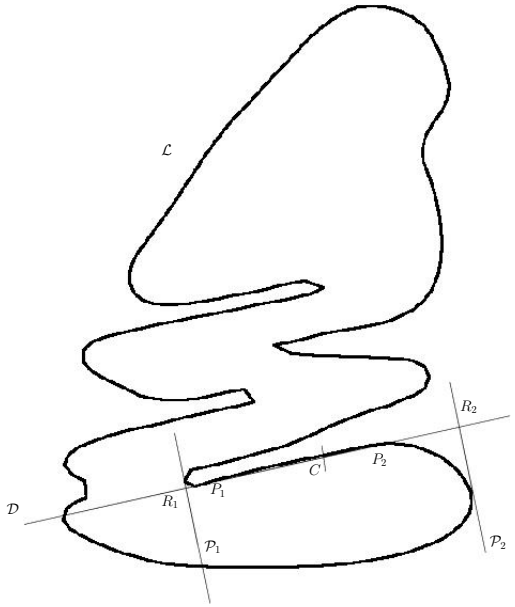
We now detail the procedures used to achieve similarity and affine invariance for semi-local normalization / encoding of Jordan curves. In what follows we consider direct Euclidean parameterization for level lines.

3.2.1 Similarity invariant normalization and encoding

The procedure is illustrated and detailed in Figure 2. Two implementation parameters, F and N, 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 be well adapted to deal 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 N 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 Figure 3 we show several normalized shape elements extracted from a single line, taking $F = 5$ and $N = 45$. 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 Section 5 concerning matching based on this semi-local encoding were carried out using $F = 5$ and $N = 45$, 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



In order to represent a level line $\mathcal{L}$, for each flat piece, 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 pieces. 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 N 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 $[R_1R_2]$. 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.

Figure 2: Similarity invariant semi-local encoding. On the left, an illustration based on a bitangent line.

be coded with $F = 5$: when their length is too small with respect to the length of the segment line $[R_1R_2]$, the resulting shape element would overlap itself.

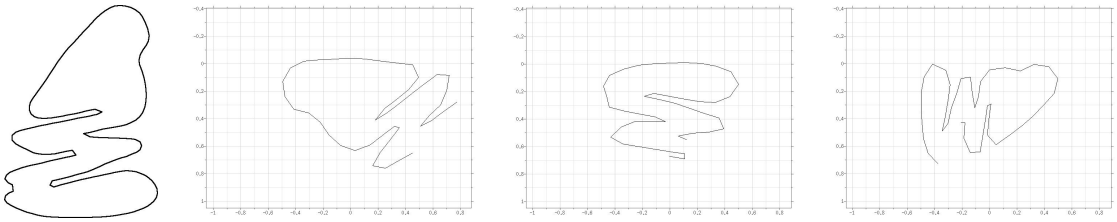


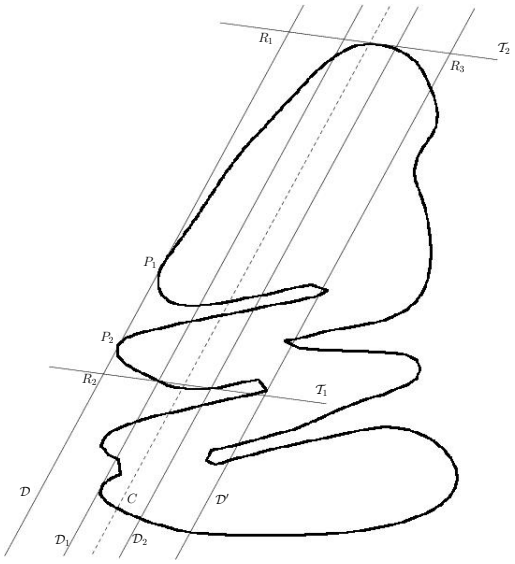
Figure 3: Example of semi-local similarity invariant encoding. The line on the left generates 19 shape elements ($F = 5$, $N = 45$). Twelve of them are based on bitangent lines, the other ones are based on flat pieces. The representation is quite redundant. Here are displayed three normalized shape elements, two deriving from bitangent lines, and one from a flat piece.

3.2.2 Affine invariant normalization/encoding

The procedure is illustrated in Figure 4. As we did for the similarity invariant normalization, implementation parameters were fixed once and for all to $F = 5$ and $N = 45$. Figure 5 shows several shape elements extracted from a single line for this choice of parameters. The encoding is in fact less redundant than for the similarity encoding procedure. This is due to the fact that the construction of affine invariant local frames imposes more constraints on the curve than the one for similarity invariant frames.

3.3 From normalized shape elements to independent features

In this section, we explain the procedure we apply to extract features from these shape elements. We empirically found that the best trade-off achieving simultaneously the three feature requirements that we pointed out in Section 3.1 is the following (see Figure 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 $C_1, C_2, \ldots, C_5$ (each of those C_i being discretized with 9 points). These features ought to be independent; nevertheless, $C_1, \ldots, C_5$ 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 Section 2.1 are made of these six 'generic' shape features $C_1, \ldots, C_6$. Using the notations



In order to derive an affine invariant representation of a level line $\mathcal{L}$, for each flat piece, 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 pieces. Consider the tangent line $\mathcal{D}$ to these points;

b) Starting forward from P_2, find the next tangent to $\mathcal{L}$ which is parallel to $\mathcal{D}$. Call it $\mathcal{D}'$;

c) Consider the straight lines which are parallel to $\mathcal{D}$ and lay at $1/3$ and $2/3$ of distance from $\mathcal{D}$ to $\mathcal{D}'$. Call them $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively;

d) Starting forward from P_2, find the next intersection points between $\mathcal{L}$ and $\mathcal{D}_1$, and $\mathcal{L}$ and $\mathcal{D}_2$. Consider the straight line $\mathcal{T}_1$ defined by these two points;

e) Starting backward from P_1, find the previous tangent to $\mathcal{L}$ parallel to $\mathcal{T}_1$, and call it $\mathcal{T}_2$;

f) Define points R_1, R_2, and R_3 as the intersections between $\mathcal{D}$ and $\mathcal{T}_2$, $\mathcal{D}$ and $\mathcal{T}_1$, and $\mathcal{D}'$ and $\mathcal{T}_2$, respectively;

g) Points R_1, R_2, R_3 define an affine basis. The affine normalization is fixed by mapping $\{R_1, R_2, R_3\}$ into $\{(0,0),(1,0),(0,1)\}$ if $\{R_1, R_2, R_3\}$ is a direct frame, and into $\{(0,0),(1,0),(0,-1)\}$ if not;

h) Encoding: consider the intersection point between $\mathcal{L}$ and the straight line equidistant from $\mathcal{D}$ and $\mathcal{D}'$ (the first one starting from P_2). Call it C. Normalize the portion of $\mathcal{L}$ having normalized length F/2 at both sides of C. Store N equi-distributed points over the normalized piece of curve.

Figure 4: Affine invariant semi-local encoding. The encoded shape element is based on the bitangent line $\mathcal{D}$.

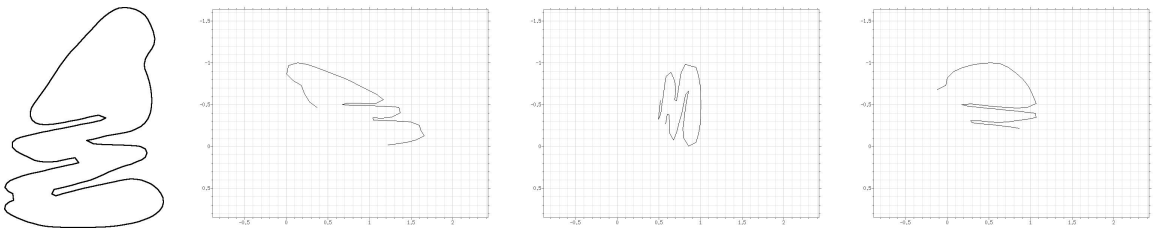


Figure 5: Example of semi-local affine invariant encoding. The line on the left generates 7 shape elements ($F = 5$, $N = 45$); three of them are represented here.

introduced in the previous sections, we have $x_i(\mathcal{S}) = C_i$ $(i \in \{1, \ldots, 6\})$. For every $i \in \{1, \ldots, 5\}$, $E_i = (\mathbb{R}^2)^9$, $E_6 = (\mathbb{R}^2)^6$, and the distances d_i between them are L^∞-distances.

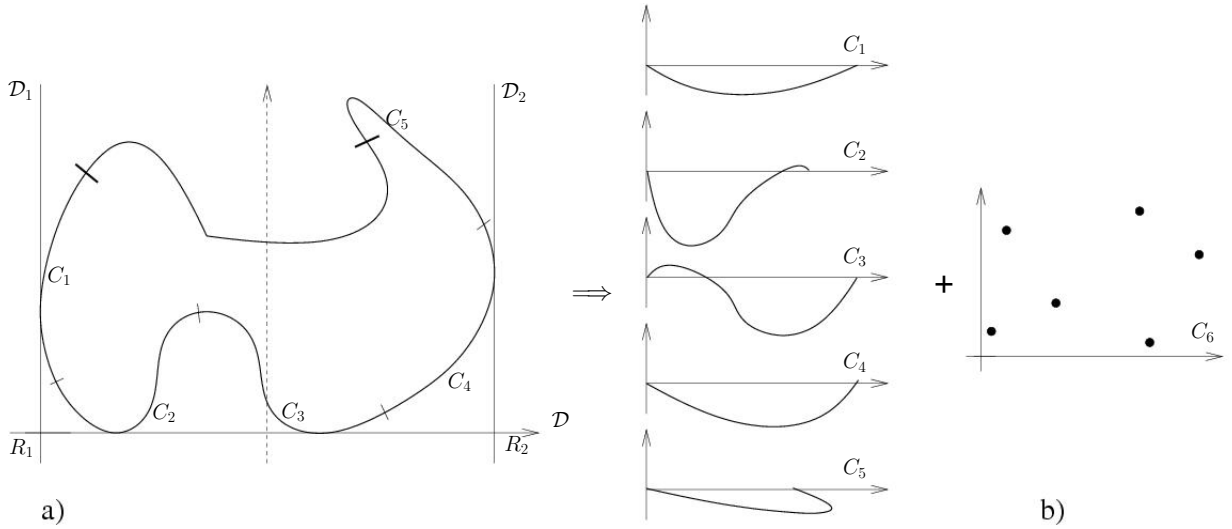


Figure 6: Building independent features. Example of a similarity-invariant encoding (see Section 3). Sketch a): original shape as a Jordan curve in a normalized frame based on a bitangent line. Both ends of the considered Jordan curve piece are marked with bold lines: this representation is split into 5 pieces C_1, C_2, C_3, C_4, and C_5. Sketch b): each one of them is normalized, and a sixth feature C_6 made of the endpoints of these pieces is also built.

Another possibility that we have investigated is to use a principal component analysis (PCA) [28]. 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. The presented independent features extraction is much more reliable and provides much better experimental results (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 $\text{PFA}(\mathcal{S}, \delta)$ that a shape element could fall just by chance at a distance lower than δ to a query shape $\mathcal{S}$ is correct under the independence assumption **(A)** on the distances between features. Of course, the degree of trust that we are able to give to the associated Number of False Alarms $\text{NFA}(\mathcal{S}, \delta)$ (Definitions 3 and 6) strongly depends on the validity of this independence assumption. The expected number of false alarms among all ε-meaningful matches with the query shape should be lower than ε. Nevertheless, we are not able to separate false alarms and real matches: we only observe detections. Now, Helmholtz principle [12] states that no detection in "noise" (which has to be precised) should be considered as relevant. All ε-meaningful matches in the noise should thus be considered as false alarms: in such a noise situation there should be on the average about ε many of them. The following experiments test this claim. We show that in this situation the NFA is a pretty good prediction of the number of detections. The independence assumption is valid enough, so that the claim according to which there is on average at most ε false alarms among ε-meaningful matches still holds.

As a first experiment we check the detection thresholds on a very simple model: we consider as database and query some random walks with independent increments (instead of the normalized shape elements of Section 3). Although in this case the shape elements do not come from Jordan curves, the background model is ensured to be true, in the sense that the considered shape elements fit perfectly the independence assumption.

Table 1 shows that the Number of False Alarms is very accurately predicted for various database sizes: the number of detections with an NFA lower than ε is about ε indeed.

value of ε:	0.01	0.1	1	10	100	1,000	10,000
100,000 shape elements	0	0	2.3	15.2	122.2	1,075.5	9,872.2
50,000 shape elements	0.2	0.3	1.5	11.9	106.1	1,001.1	9,789.5
10,000 shape elements	0	0	1.2	12.5	108.4	985.0	–

Table 1: Random walks. Average (over 10 samples) number of 1-meaningful detections *versus* ε. First row: database of 100,000 shape elements. Second row: database of 50,000 shape elements. Third row: database of 10,000 shape elements.

Of course, modeling shape elements with random walks is not realistic. 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 Section 3) which introduces some structural similarities (for example, shape elements coming from bitangent points show mostly common structures). In order to quantify the "amount of dependency" due to these two aspects, we have led the following additional experiment.

Let us consider databases made of shape elements extracted from pieces of level lines in white noise images. Table 2 shows that the number of detections is not as precisely predicted as in the preceding experiment, at least for small values of ε. Nevertheless, the order of magnitude is still correct, and does not depend on the size of the database. These properties are sufficient for setting the Number of False Alarms threshold based on Helmholtz principle. Following this method, a match is supposed to be highly relevant if it cannot happen in white noise images. According to Table 2, matches with an NFA lower than 0.1 are ensured to be very unlikely in white noise images. If we want to ensure a strong confidence in the detected matches, we are thus led to consider 0.1-meaningful matches in realistic experiments.

value of ε:	0.01	0.1	1	10	100	1,000	10,000
104,722 shape elements	0.3	1.5	6.5	31.5	173.9	1,264.4	9,803.1
47,033 shape elements	0.1	0.3	3.7	20.2	125.4	976.3	9,854.2
10,784 shape elements	0	0.2	2.6	14.8	107.6	973.3	–

Table 2: Normalized pieces of white noise level lines. Average (over 10 samples) number of 1-meaningful detections *versus* ε. First row: database of 104,722 shape elements. Second row: database of 47,033 shape elements. Third row: database of 10,784 shape elements.

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 Section 3. 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 Definition 6) are highlightened along the following experiments. More experiments can be seen in [26] and [37].

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 ("de-normalized" 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. 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). False matches have generally an NFA larger than 10^{-1}. If we are concerned with very sure detections, we simply set the NFA threshold to 10^{-1}.

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 Figure 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 Proposition 3, 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".

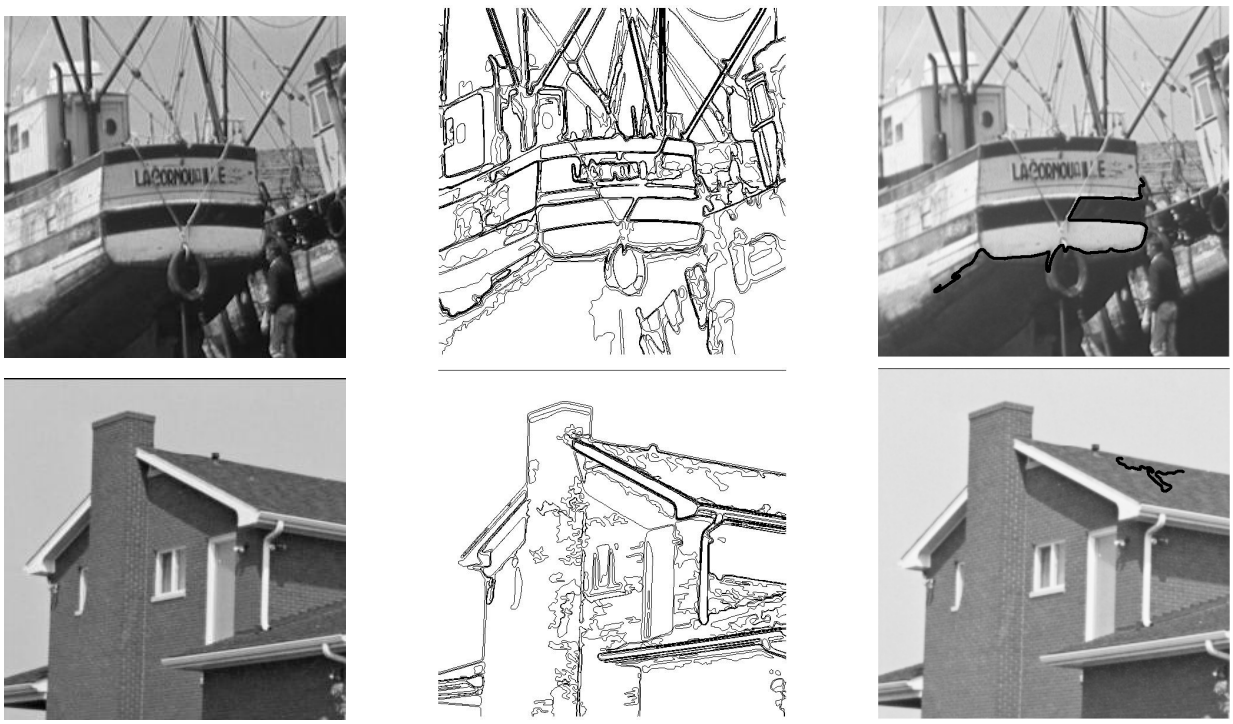


Figure 7: Two unrelated images. Original images (left), and meaningful level lines (middle). The 846 normalized shape elements from the top image are searched among the 281 normalized shape elements from the bottom image. Only one 1-meaningful match is detected (right). Its NFA is 0.2, which is very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

5.2 Perspective distortion

It is not surprising that the affine method performs better than the similarity method, when dealing with images related through an affine transformation. In this second experiment, we show that, as expected, the affine method also performs better (with regards to the lowest reachable NFA) than the similarity method, when applied to real images related through moderately weak perspective transformations. The two images considered in this experiment (which we call "Hitchcock experiment") are two different snapshots of the same scene. They are shown in Figure 8, with their corresponding level lines.

For the affine semi-local invariant method, 1150 and 853 shape elements were extracted from the target image and from the database image, respectively. The number of 1-meaningful matches detected was 16. These 16 matched shape elements are shown, superimposed to images, in Figure 9. No false matches were detected, and all matches have their NFA below 0.1. The best match, shown in Figure 10, reaches NFA $= 6.5\,10^{-11}$. This value is remarkably low, considering that ideal perfect matches in this experiment would have a number of false alarms of $1150 \times 853/853^6 = 2.5\,10^{-12}$ (when the empirical distributions of distances to target codes are learned using only the considered database image, as we do here).

In Figure 11 we display the meaningful matches which were detected using the similarity semi-local invariant recognition method. In this case, $2,033$ and $1,463$ shape elements were extracted from the target image and from the database image, respectively. As pointed out in Section 3.2.2, the similarity method leads to extract more shape elements than the affine method, that is why more 1-meaningful matches (26) are detected in this case. The meaningful matches for the similarity method are shown in Figure 11. The lowest NFA reached with the similarity method is $3.8\,10^{-8}$, and corresponds to the shape elements presented in Figure 12. In Figures 11(b) and 11(c), we present, respectively, the shape elements matching at $\varepsilon < 0.1$, and those for which the NFA is between 0.1 and 1. Notice that none of the 10^{-1}-meaningful matches are false matches, and that the corresponding shape elements are in general much more local than the shape elements matching in Figure 11(c). Indeed, the more global the shape elements, the less accurate the similarity approximation of the underlying transformation, which is in fact a projective transformation. As a consequence,

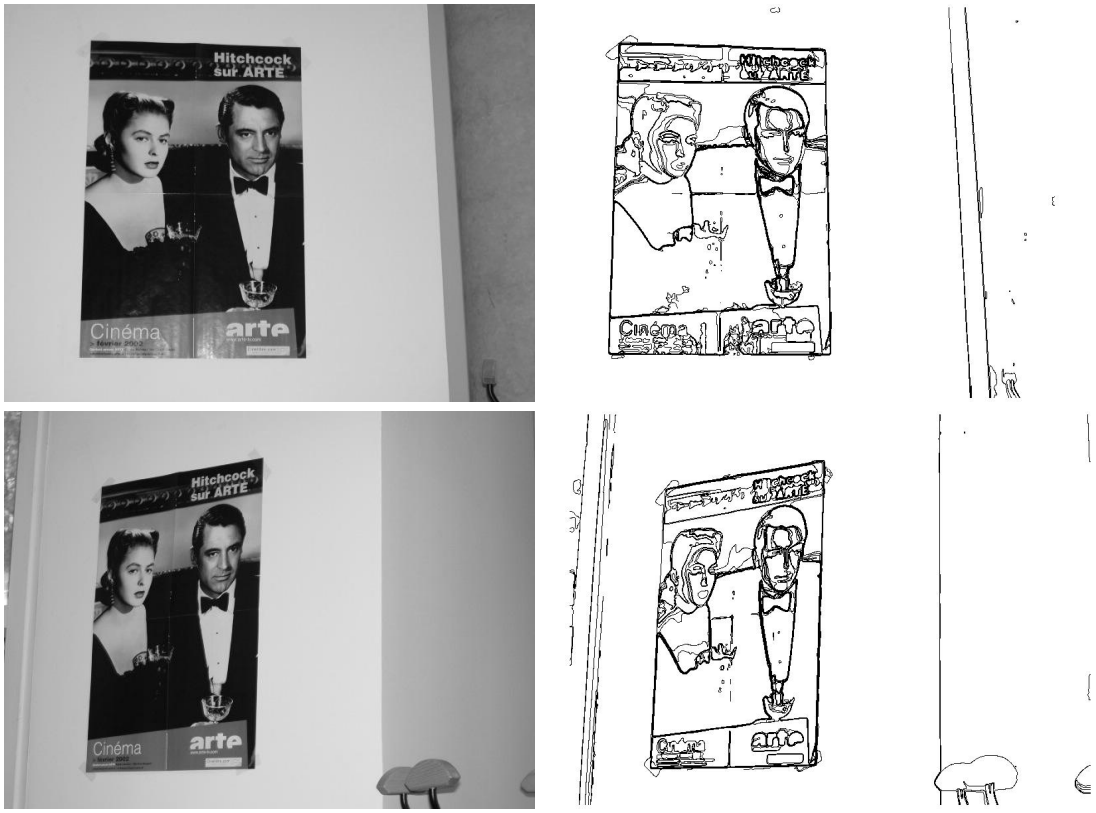


Figure 8: Hitchcock experiment: original images (corresponding to two different snapshots of the same scene) and their corresponding level lines to be encoded. The image on top is considered as "target" image. In the target image, 307 meaningful level lines are detected, and 266 meaningful level lines are detected in the database image.

the lowest NFA reached with the similarity method ($3.8\,10^{-8}$) is larger than the lowest NFA reached with the affine method (which is $6.5\,10^{-11}$). In fact, NFAs are much lower for the affine method than for the similarity method, as the weak projective transformation applied here is better locally approximated by affine transformations than by similarity transformations.

Two false matches, for which the NFA is larger than 0.1, can be seen in Figure 11(c). In Figure 13 we show the shape elements of these false matches, as well as the superimposed normalized shape elements represented in the normalized frame. Let us notice that the distance threshold corresponding to such NFA is large enough so that no "good matches" are missed. Although the proposed method does not quantify the probability of non-detection (denoted by α' in Section 2.2), experimental evidence shows that this probability is very low.

Figure 9: Affine invariant semi-local recognition method: the 16 meaningful matches between shape elements. No false matches were detected (in the sense that all meaningful matches correspond to the same "piece of object"), and all detections show an NFA below 0.1. The lowest NFA is $6.5\,10^{-11}$.

Figure 10: Affine invariant semi-local recognition method: the match showing the lowest NFA ($6.5\,10^{-11}$).

5.3 Dealing with partial occlusions and contrast changes

The following experiment consists in comparing the codes extracted from two views of Velasquez' painting *Las Meninas* (see Figure 14). The codes extracted from the query image ($11,332$ codes) are searched among the codes extracted from the database image ($12,833$ codes). Shape elements are here normalized with respect to similarity transformations. Note that the target image is a photography which was taken in the museum: visitors' heads hide a part of the painting.

Figure 15 shows on the left the set of pieces of level lines in the target image that match a piece of level line in the database image with a corresponding Number of False Alarms less than 1 (meaningful matches), and on the right the set of shape elements from the database image that correspond to at least one shape element in the query image. The algorithm identifies 55 meaningful matches. Only 5 false matches can be seen among them. They all have an NFA between 1 and 10^{-1}. In fact, 36 matches show an NFA lower than 10^{-1}.

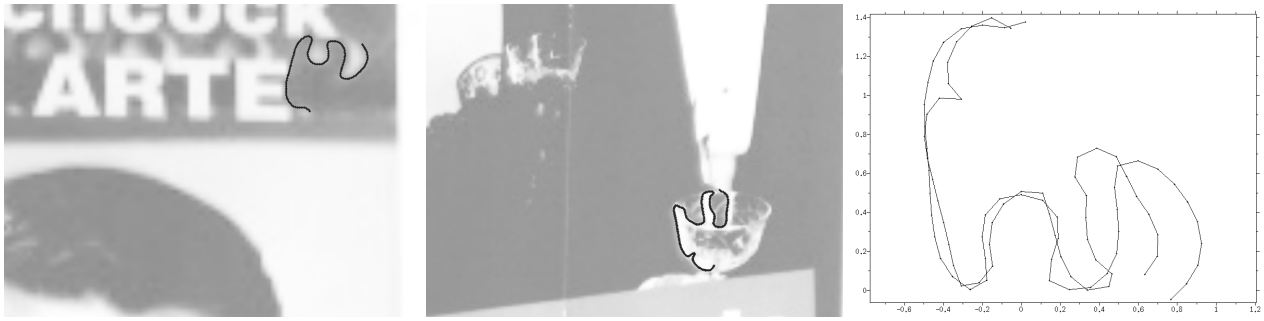
(a) *All* 26 *matches having an NFA below* 1.

(b) 12 *matches show an NFA below* 0.1.

(c) 14 *matches show an NFA between* 0.1 *and* 1.

Figure 11: Similarity invariant semi-local recognition method: meaningful matches between shape elements. Among the 26 matches having an NFA below 1, 12 are 10^{-1}-meaningful. False matches (two) can only be seen in (c), and their NFA is above 0.1.

Figure 12: Similarity invariant semi-local recognition method: the match showing the lowest NFA ($3.8\,10^{-8}$).



(a) *False match,* $NFA = 0.64$

(b) *False match,* $NFA = 0.68$

Figure 13: Similarity semi-local invariant method: the two false matches. Their NFA are larger than 0.1. We could expect such an NFA since the curves show local variations, but their global aspect is the same, as it can be noticed when looking closer at the corresponding normalized shape elements.

Figure 14: 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.

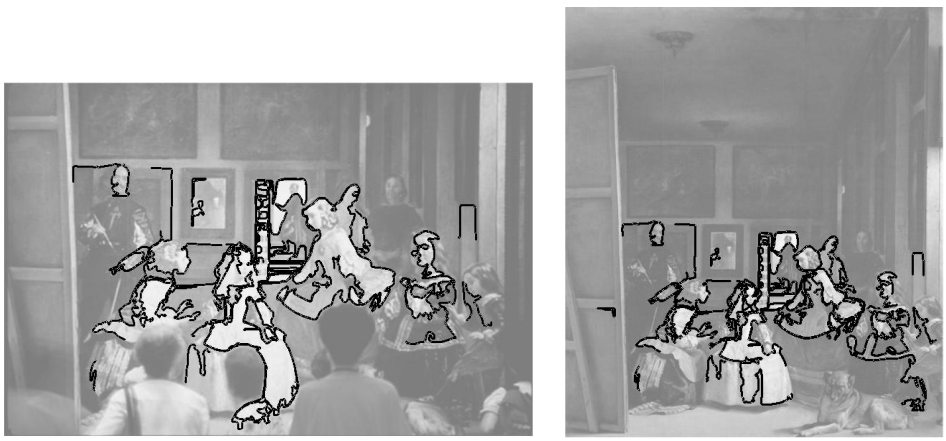


Figure 15: Las Meninas. The 55 meaningful matches. Half of them has an NFA lower than 10^{-3}. The best match has an NFA equal to 4.10^{-14}. To each bold piece of level line on the right corresponds a bold piece of level line on the left.

6 Conclusion and perspectives

In this article, 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 article 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 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 [7]. As noticed on preliminary results [26, 37], combining the spatial information furnished by matched shape elements strongly reinforces the recognition confidence of the method.

Acknowledgements: 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).

References

[1] A.A. Adjeroh and M.C. Lee. An occupancy model for image retrieval and similarity evaluation. *IEEE Transactions on Image Processing*, 9(1):120–131, 2000.

[2] 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.

[3] 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.

[4] K. Åström. Fundamental limitations on projective invariants of planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):77–81, 1995.

[5] F. Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183–193, 1954.

[6] 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.

[7] 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.

[8] F. Cao, P. Musé, and F. Sur. Extracting meaningful curves from images. *Journal of Mathematical Imaging and Vision*, 2004. To appear.

[9] 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.

[10] A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful alignments. *International Journal of Computer Vision*, 40(1):7–23, 2000.

[11] A. Desolneux, L. Moisan, and J.-M. Morel. Edge detection by Helmholtz principle. *Journal of Mathematical Imaging and Vision*, 14(3):271–284, 2001.

[12] A. Desolneux, L. Moisan, and J.-M. Morel. *A theory of digital image analysis*. 2004. Book in preparation.

[13] P.A. Devijver and J. Kittler. *Pattern recognition - A statistical approach*. Prentice Hall, 1982.

[14] P. Frosini and C. Landi. Size functions and formal series. *Applicable Algebra in Engineering, Communication and Computing*, 12:327–349, 2001.

[15] Y. Gousseau. Comparaison de la composition de deux images, et application à la recherche automatique. In *proceedings of GRETSI 2003*, Paris, France, 2003.

[16] 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.

[17] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2001.

[18] G. Kanizsa. *La Grammaire du Voir*. Diderot, 1996. Original title: *Grammatica del vedere*. French translation from Italian.

[19] 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.

[20] J.L. Lisani. *Shape Based Automatic Images Comparison*. PhD thesis, Université Paris 9 Dauphine, France, 2001.

[21] 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.

[22] D.G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publisher, 1985.

[23] D. Marr. *Vision*. Freeman Publishers, 1982.

[24] L. Moisan. Affine plane curve evolution: A fully consistent scheme. *IEEE Transactions on Image Processing*, 7(3):411–420, 1998.

[25] 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.

[26] P. Musé. *On the definition and recognition of planar shapes in digital images*. PhD thesis, École Normale Supérieure de Cachan, 2004.

[27] P. Musé, F. Sur, F. Cao, and Y. Gousseau. Unsupervised thresholds for shape matching. In *Proceedings of IEEE International Conference on Image Processing*, Barcelona, Spain, 2003.

[28] P. Musé, F. Sur, and J.-M. Morel. Sur les seuils de reconnaissance des formes. *Traitement du Signal*, 20(3):279–294, 2003.

[29] C. Olson and D.P. Huttenlocher. Automatic target recognition by matching oriented edge pixels. *IEEE Transactions on Image Processing*, 6(12):103–113, 1997.

[30] C.F. Olson. Improving the generalized Hough transform through imperfect grouping. *Image and Vision Computing*, 16(9-10):627–634, 1998.

[31] 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.

[32] X. Pennec. Toward a generic framework for recognition based on uncertain geometric features. *Videre: Journal of Computer Vision Research*, 1(2):58–87, 1998.

[33] C.A. Rothwell. *Object Recognition Through Invariant Indexing*. Oxford Science Publications, 1995.

[34] G. Sapiro and A. Tannenbaum. Affine invariant scale-space. *International Journal of Computer Vision*, 11(1):25–44, 1993.

[35] C. Schmid. A structured probabilistic model for recognition. In *Proceedings of Conference on Computer Vision and Pattern Recognition*, volume 2, pages 485–490, Fort Collins, Colorado, USA, 1999.

[36] S.D. Silvey. *Statistical Inference*. Chapman and Hall, 1975.

[37] F. Sur. *A contrario decision for shape recognition*. PhD thesis, Université Paris Dauphine, 2004.

[38] 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.

[39] G.H. Watson and S.K. Watson. Detection of unusual events in intermittent non-gaussian images using multiresolution background models. *Optical Engineering*, 35(11):3159–3171, 1996.

[40] 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.

[41] H.J. Wolfson and I. Rigoutsos. Geometric hashing: an overview. *IEEE Computational Science & Engineering*, 4(4):10–21, 1997.

[42] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.