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 (dis-)similarity measure to the query shape. Their main weakness 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: Planar shape recognition, background model, number of false alarms, meaningful matches, level lines.

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 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 (dis-)similarity measure between features describing shapes) and decision. The first two stages have been widely addressed in the literature (see for instance [43] or [47] and references therein), and are 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,

[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 Vandœuvre-lès-Nancy Cedex, France.

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.

1.1 Shape extraction and representation

In computer vision, extraction of shape information from images dates back to Marr [27] but Attneave [6] as well as Wertheimer [45] 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). Geometrical shapes can then be modeled as simple closed curves. However, as pointed out by Kanisza [20], 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 curves corresponding to objects boundaries, but pieces of them. In this work we adopt this atomic shape representation; we call *shape element* any piece of 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 in this paper we do not address recognition of shapes as a whole, which can be performed by integrating the recognized shape elements, based on spatial coherence [9].

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 numerous curves arbitrarily close to a given curve [5]. On the other hand, 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, which we call a *code*. (Figure 1 illustrates this.)

1.2 Contribution to matching decision and related work

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 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 deci-

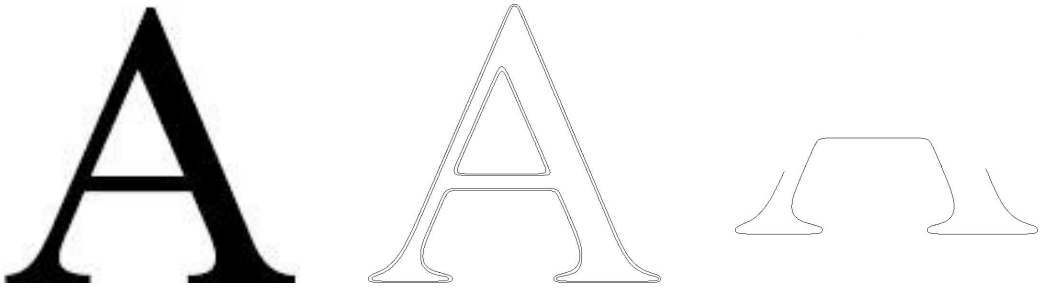


Figure 1: Some vocabulary. From left to right: a (synthetic) image, corresponding *level lines*, and a piece of level line (called *shape element*). The scope of this paper is to match shape elements. A so-called *code* is a list of K invariant descriptors (w.r.t. similarity or affine transformations) that represent a shape element. These invariant descriptors are called *features* and are defined in Section 3.

sion rule. We apply this methodology to the shape matching problem, but our scope is much wider since the principles that are used are general. To the best of our knowledge, a generic acceptance / rejection decision method for shape matching has not been proposed yet. In general, matches with a query shape are only ranked (for example along a distance [17], along some probability [40], or along some number of votes in hashing methods [46]).

Let us specify what we mean by "automatic decision rule" for shape matching. Assume we are looking in a shape database (usually extracted from an image or a set of images) for a query shape (element) $\mathcal{S}$. A distance between shapes is available, so that the smaller the distance is, the more similar the shapes are. The question is: what is the threshold value for that distance to ensure recognition? Given two shapes and an observed small distance 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 distance, 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 [14] for a comprehensive account), such a methodology is called *a contrario* decision. This detection framework has been recently applied to various situations, for example by Desolneux *et al.* to the detection of alignments [12] or contrasted edges [13], by Almansa *et al.* to the detection of vanishing points [2], by Stival and Moisan to stereo images [30], by Gousseau to the comparison of image "composition" [18] and by Cao to the detection of good continuations [8]. In computer vision, the first attempts to detect events in images against a random situation certainly date back to David Lowe's work on perceptual organization. In [25], 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 [34] 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 [19] 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 [19] inspired several works [1, 35]. Another approach is to simultaneously use a background and a shape model, as in [22], where performances of shape recognition algorithms are studied in a fairly general context. Following Huttenlocher and Grimson's work, Pennec [37] 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 restrictive assumptions (*a priori* shape model, uniform distribution of features), as in Huttenlocher and Grimson's work. As pointed out by Pennec, these limitations are nevertheless hardly verified in real cases.

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 [11] or [44]). 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.

Nevertheless, none of the preceding methods allow to fix the thresholds in an automated way. In the following, we intend to make such probabilistic methods reliable for the geometrical 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 and the database of observed shapes. 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 [32].

1.3 Organization of this paper

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 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.* [13] on boundaries extraction (improved in [10]) and by Lisani *et al.* [23, 24] on curve normalization, *shape elements* are extracted from images, then matching is performed and followed by the decision process. In Section 4, we present an experiment that shows the validity of the proposed model. It is verified that the methodology satisfies Helmholtz principle [14]: a meaningful match is a match that is not likely to occur in a context where noise overwhelms the information. Experimental results are discussed in Section 5. We conclude with Section 6.

2 An *a contrario* decision framework

The aim of this section is to present a general method to automatically fix an acceptance / rejection threshold for the recognition of shape elements, up to a given class of invariance. 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 of usual hypothesis testing theory.

2.1 Decision as hypothesis testing

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 $\mathcal{B}$. As mentioned in the previous section, we assume each shape element S to be represented by a *code*, that is a set of K features $x_1(S)$, $x_2(S)$, ..., $x_K(S)$, each of them belonging to a set E_i endowed with a dissimilarity measure d_i ($i \in \{1, \dots, K\}$). The "product dissimilarity measure" d is then defined over $E_1 \times E_2 \times \cdots \times E_K$, that is

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

The dissimilarity values are supposed to share the same range (and will be chosen accordingly), so that taking their maximum is sound. In what follows, these dissimilarities are named "distances", although they are not necessarily true metrics.

The distance function between shape elements being given, deciding whether a shape element matches another shape element consists in setting a threshold δ over this distance. Ideally, δ should

be set automatically, without any user tuning. We propose to use the hypothesis testing framework [15, 41] in order to replace the distance bound by a probability of false alarms bound.

2.1.1 Shape model *versus* background model

A database shape $\mathcal{S}'$ being observed, or more precisely the distance $d(\mathcal{S}, \mathcal{S}')$ between the query $\mathcal{S}$ and $\mathcal{S}'$ being observed, the problem is to decide whether $d(\mathcal{S}, \mathcal{S}')$ is small enough so that pairing $\mathcal{S}$ and $\mathcal{S}'$ makes sense, or whether the distance is quite large so that it does not make sense anymore. We are thus led to consider the two following alternative hypotheses:

- $\mathcal{H}_1$: "$d(\mathcal{S}, \mathcal{S}')$ is observed because of some causality", since we are interested in pairing shape elements that are instances of the same object.
- $\mathcal{H}_0$: "$d(\mathcal{S}, \mathcal{S}')$ is observed only *by chance*", for instance because the database contains many shape elements.

A classical test for choosing between $\mathcal{H}_0$ and $\mathcal{H}_1$ is to compare the distance $d(\mathcal{S}, \mathcal{S}')$ with some predetermined value δ and to decide that $\mathcal{H}_1$ holds whenever $d(\mathcal{S}, \mathcal{S}') \leq \delta$. Otherwise, $\mathcal{H}_1$ is rejected and the null hypothesis $\mathcal{H}_0$ is accepted. Let us call $\mathcal{T}_\delta(\mathcal{S})$ this (statistical) test. The quality of a statistical test is measured by the probability of taking wrong decisions. Two kinds of errors are possible: reject $\mathcal{H}_1$ for an observation $\mathcal{S}'$ for which $\mathcal{H}_1$ is actually true (non-detection), and accept $\mathcal{H}_1$ for $\mathcal{S}'$ although $\mathcal{H}_1$ is false (false alarm). A probability measure can be associated to each type of error:

- The *probability of false alarms* $\alpha = \Pr(d(\mathcal{S}, \mathcal{S}') \leq \delta | \mathcal{H}_0)$;
- The *probability of non-detection* or *probability of a miss* $\alpha' = \Pr(d(\mathcal{S}, \mathcal{S}') > \delta | \mathcal{H}_1)$;

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

However, we assume no other information but the observed set of features. We are therefore unable to estimate $\Pr(\cdot|\mathcal{H}_1)$, which would need 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, or a database made of various instances of the shape of interest could be provided, enabling a shape model estimation as in [48]). As a consequence, any classical method such as *likelihood ratio test* and the *Bayesian test* is simply unusable here. 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.

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 $\Pr(\cdot|\mathcal{H}_0)$. Since no common causality is assumed under hypothesis $\mathcal{H}_0$, it is natural to impose that the distances between shape features are independent. More precisely, we assume that the shape elements belong

to some probability space $(\Omega, \mathcal{A}, \overline{\Pr})$ such that $\Pr(\cdot|\mathcal{H}_0) = \overline{\Pr}$ and 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* $\Sigma \mapsto d_i(x_i(\mathcal{S}), x_i(\Sigma))$ *(*$i \in \{1, \ldots, K\}$*) from* Ω *to* $\mathbb{R}^+$ *are mutually statistically independent.*

In the remainder of this article, shape elements with respect to which probabilities are computed are denoted by Σ, while fixed shapes are denoted by $\mathcal{S}$, $\mathcal{S}'$... For every $i \in \{1, \ldots, K\}$, the probability $\overline{\Pr}(d_i(x_i(\mathcal{S}), x_i(\Sigma)) \leq \delta)$ is denoted by $P_i(\mathcal{S}, \delta)$.

2.1.2 Estimating the probability of false alarms

Although we are not able to compute the probability of non-detection $\Pr(d(\mathcal{S}, \Sigma) > \delta|\mathcal{H}_1)$, a straightforward computation provides the value of the probability of false alarms of the statistical test $\mathcal{T}_\delta(\mathcal{S})$, denoted by $\text{PFA}(\mathcal{S}, \delta) := \Pr(d(\mathcal{S}, \Sigma) \leq \delta|\mathcal{H}_0)$. By the definition of d and of the null hypothesis:

$$\begin{aligned}\text{PFA}(\mathcal{S}, \delta) &= \Pr\left(\max_{i\in\{1,\ldots,K\}} d_i(x_i(\mathcal{S}), x_i(\Sigma)) \leq \delta \mid \mathcal{H}_0\right). \\ &= \overline{\Pr}\left(\max_{i\in\{1,\ldots,K\}} d_i(x_i(\mathcal{S}), x_i(\Sigma)) \leq \delta\right).\end{aligned}$$

Assumption **(A)** then yields

$$\text{PFA}(\mathcal{S}, \delta) = \prod_{i\in\{1,\ldots,K\}} \overline{\Pr}\left(d_i(x_i(\mathcal{S}), x_i(\Sigma)) \leq \delta\right).$$

Thus, we have just proved the following proposition.

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

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

2.1.3 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

$$\delta^*(p) = \max\{\delta > 0, \text{PFA}(\mathcal{S}, \delta) < p\}. \tag{2}$$

Consequently, if the test is to accept $\mathcal{H}_1$ 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 hypothesis of interest as soon as the null 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 that 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".

2.2 Automatic setting of the distance threshold

2.2.1 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 2 *The* Number of False Alarms *of the shape element $\mathcal{S}$ at a distance δ is*

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

By a slight abuse of notation, we also define the number of false alarms of the query shape element $\mathcal{S}$ and a database shape element $\mathcal{S}'$ as 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 can be seen as the average number of detections that are expected when we test whether the distance from each shape element in the database to $\mathcal{S}$ is below d (first formulation), or the average number of database shapes which are "false alarms" and whose distance to $\mathcal{S}$ is lower than $d(\mathcal{S}, \mathcal{S}')$ (second formulation).
Remark: Let us moreover notice that the arguments of $\text{NFA}(\mathcal{S}, \mathcal{S}')$ (seen as a two variables function) do not play a symmetric role.

2.2.2 Meaningful matches

Instead of directly bounding the probability of false alarms in order to deduce a distance threshold (as explained in Section 2.1.3), we bound the number of false alarms, since this quantity has an interpretation in terms of expected frequencies.

Definition 3 *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:*

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

A straightforward consequence is that, with Equation (2), a shape element $\mathcal{S}'$ is an ε-meaningful match of the query shape element $\mathcal{S}$ if and only if

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

The ε-meaningful matches of $\mathcal{S}$ are then those shape elements for which the distance to $\mathcal{S}$ is below $\delta^* \left(\frac{\varepsilon}{N} \right)$, and the probability of false alarms of the associated test is consequently less than $\frac{\varepsilon}{N}$. 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 actual number of ε-meaningful matches. However, if all shape elements in the database were generated by the background model, then hypothesis $\mathcal{H}_1$ should never be accepted, all ε-meaningful detections should thus be considered as false alarms. The following proposition makes this claim more formal, and is the keystone of what we propose.

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 Σ_j $(1 \leq j \leq N)$ denote the shape elements in the database, and χ_j the indicator function of the event e_j: "Σ_j is an ε-meaningful match of the query $\mathcal{S}$ " (*i.e.* its value is 1 if Σ_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 $\mathcal{S}$.
Linearity of expectation implies that the expectation of R is $E(R) = \sum_{j=1}^{N} E(\chi_j)$. By the remark above about δ^*, and since shape elements from the database are assumed to satisfy the assumptions of the background model, it follows that

$$E(\chi_j) = \Pr(\Sigma_j \text{ is an } \varepsilon\text{-meaningful match of } \mathcal{S} | \mathcal{H}_0) = \Pr\left(d(\mathcal{S}, \Sigma_j) \leq \delta^* \left(\frac{\varepsilon}{N} \right) \mid \mathcal{H}_0 \right).$$

By the definition of PFA, $E(\chi_j) = \text{PFA}\left(\mathcal{S}, \delta^* \left(\frac{\varepsilon}{N} \right)\right)$, thus $E(R) = \sum_{j=1}^{N} \text{PFA}\left(\mathcal{S}, \delta^* \left(\frac{\varepsilon}{N} \right)\right)$. Hence, by the 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 us to compute $E(R)$. Since dependencies between events e_j are unknown, we are not able to estimate the probability law of R.

2.2.3 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 is, the "more certain" the ε-meaningful detections are. Of course, the same claim is true when considering distances: the lower the distance threshold δ, the more certain the corresponding matches, but considering the NFA quantifies this confidence level.

2.3 How to attain very small number of false alarms?

Let us notice that the proposed *a contrario* decision methodology (that is thresholding the probability of false alarms of a statistical test in order to derive a distance threshold) would be fully consistent without the independence assumption (*cf* **(A)**), which in particular does not intervene in the proof of the central Proposition 2. Now, why is it so important to consider several independent features instead of a single, possibly highly multidimensional, feature? The reason is the following one: using independent features is a way to beat the *curse of dimensionality* [7]. 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 [25] 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.*"

Consider the following heuristic argument. Assume that the probabilities $P_i(\mathcal{S}, d)$ are estimated by empirical frequencies on a set of N shape elements, and that the database in which the query shape $\mathcal{S}$ is sought has also cardinality N. Then the lowest attainable probability is $1/N$. Consequently, if the background model is built on $K = 1$ feature, then the lowest attainable number of false alarms would be $N \cdot 1/N = 1$. 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 $K > 1$ independent features, then the lowest reachable number of false alarms would be $N \cdot 1/N^K = N^{1-K}$, which can now be much less than 1.

In practice we extract $K = 6$ independent features and 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, for the shape recognition task to be reliable in our framework, shape features have to meet the three following requirements:

1) Features provide a complete description: two shapes with the same features are identical;

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 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 curves corresponding to invariant local representations of shapes in images was proposed by Lisani *et al.* [23, 24]. It proceeds with the following steps:

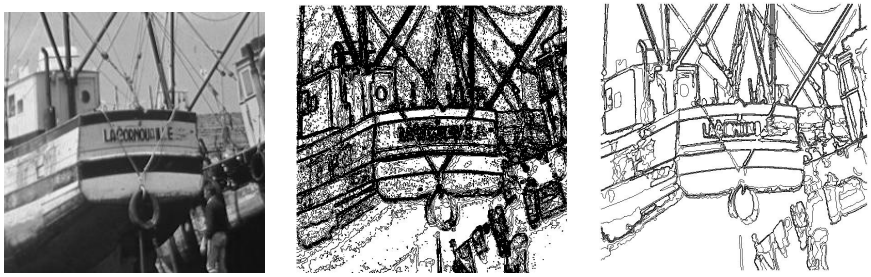
1. Extraction of meaningful level lines.
2. Affine invariant smoothing of the extracted level lines.
3. Local encoding of pieces of level lines after affine or similarity normalization.

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.* [13] proposed a method to extract *meaningful* level

lines from images, which was later improved by Cao *et al.* [10]. 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 2 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.



(a) (b) (c)

Figure 2: 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).

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, 39] is fully suitable (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 [29]. 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 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 parts, 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.* [23, 24]; the third stage is also based on the seminal work of Lamdan *et al.* [21], which was followed by Rothwell's work on invariant indexing [38]. For a more recent application of similar ideas, see Orrite *et al.* [36]. The following section is devoted to an improvement of Lisani's algorithm.

3.2 Semi-local normalization and encoding

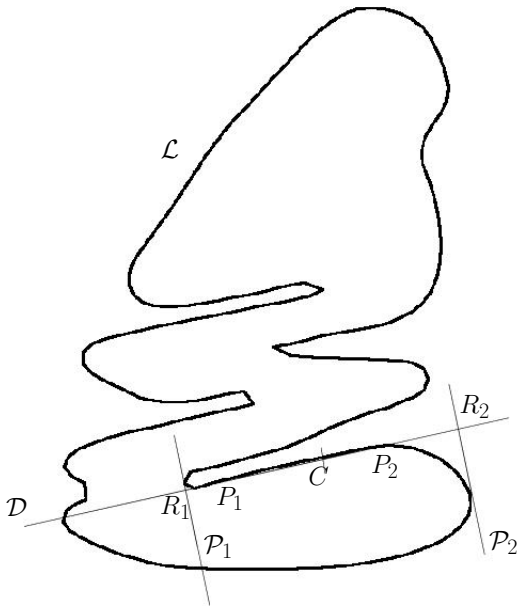
In a founding paper [16], Fischler and Bolles argue that any curve partitioning technique must satisfy two general principles: stability of the description and a complete, concise, and complexity limited explanation. Inspired from these considerations, the proposed semi-local normalization of level lines or, more generally speaking, of curves is based on a local encoding of pieces derived from robust directions. The following curve representation can be considered as a robust alternative to the Curvature Primal Sketch by Asada and Brady [4], in which the *locations* of significant curvature changes are used, although they may suffer from instability. We thus use *directions* which are given by bitangent lines and 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 [31, 42]). While bitangency is an affine invariant property, it is not the case for flat parts. However, two arguments stand for its consideration. The first one is that, under reasonable zoom factors, flat parts are preserved. The second argument is that inflexion points, which are conserved by affine transformations, are most of the time surrounded by a flat part, which is by consequent also conserved by affine transformations. If it is not the case, the tangent at the inflexion point is not a robust direction. In that sense, tangent at flat parts 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 parts).

We now detail the procedures used to achieve similarity and affine invariance for semi-local normalization / encoding of 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 3. 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 are not well adapted to deal with occlusions, while if it is too small, shape elements are not discriminatory enough. One therefore faces a classical dilemma in shape analysis: local *versus* global nature 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 Figure 4 we show several normalized shape elements extracted from a single line, with $F = 5$ 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



Given a level line $\mathcal{L}$, for each flat part, or for each bitangent line, do the following:

a) Call P_1 the first tangency point and P_2 the other one (for flat parts, P_1 and P_2 are the endpoints of the detected flat segment). Consider the tangent line $\mathcal{D}$ containing these points;

b) Call $\mathcal{P}_1$ the first tangent line to $\mathcal{L}$ which is orthogonal to $\mathcal{D}$, starting from P_1 in the negative direction. Call $\mathcal{P}_2$ the first tangent line to $\mathcal{L}$ which is orthogonal to $\mathcal{D}$, starting from P_2 in the positive direction;

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 length $F \cdot \|R_1R_2\|$, centered at C, the intersection point of $\mathcal{L}$ with the perpendicular bisector of $[R_1R_2]$ (the first intersection starting from P_1). By "normalized coordinates", one has to understand 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 3: Similarity invariant semi-local encoding. On the left, an illustration based on a flat part.

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 $M = 45$, 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 remark that a few 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.

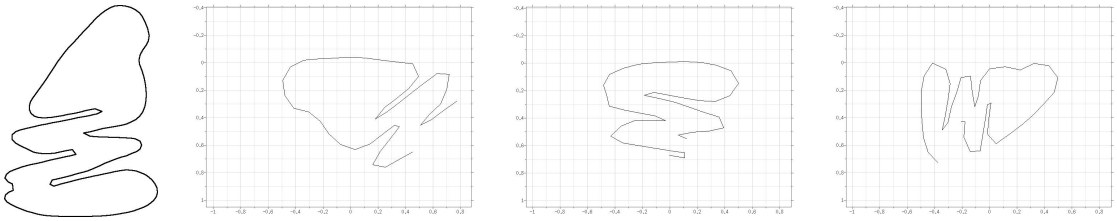


Figure 4: Example of semi-local similarity invariant encoding. The line on the left generates 19 shape elements ($F = 5$, $M = 45$). Twelve of them are based on bitangent lines, the other ones are based on flat parts. The representation is quite redundant. Here are displayed three normalized shape elements, two deriving from bitangent lines, and one from a flat part.

3.2.2 Affine invariant normalization/encoding

The procedure is illustrated in Figure 5. As we did for the similarity invariant normalization, implementation parameters were fixed once and for all to $F = 5$ and $M = 45$. Figure 6 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.

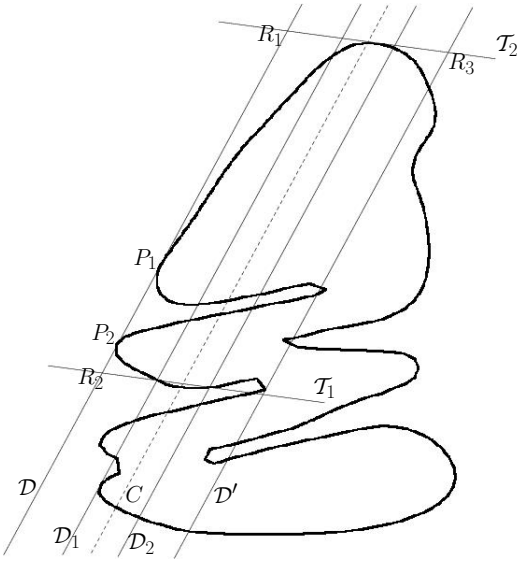
3.3 From normalized shape elements to independent features

In this section, we explain the procedure we apply to extract independent 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 7 for an illustration). Each piece of 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,

The affine invariant representation of a level line $\mathcal{L}$ is computed by applying the following procedure for each flat part or bitangent of $\mathcal{L}$:

a) Call P_1 the first tangency point and P_2 the other one (for flat parts, P_1 and P_2 are the endpoints of the detected flat segment). Consider the tangent line $\mathcal{D}$ to these point;

b) Starting 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 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 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 M equi-distributed points over the normalized piece of curve.

Figure 5: Affine invariant semi-local encoding. The encoded shape element is based on the bitangent line $\mathcal{D}$.

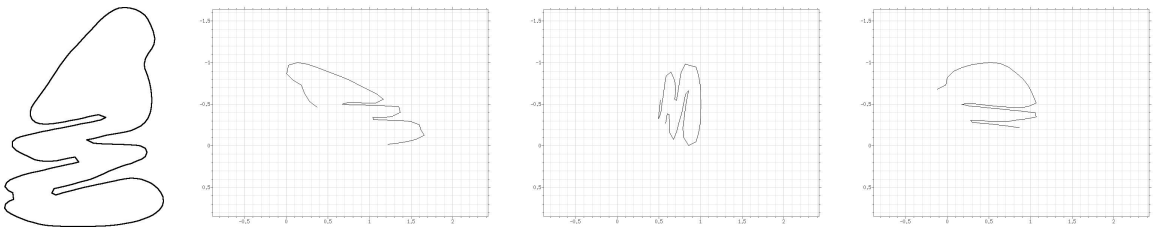


Figure 6: Example of semi-local affine invariant encoding. The line on the left generates 7 shape elements ($F = 5$, $M = 45$); three of them are represented here.

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 introduced in the previous sections, we have $x_i(\mathcal{S}) = C_i$ ($i \in \{1, \ldots, 6\}$), and for every $i \in \{1, \ldots, 5\}$, $E_i = (\mathbb{R}^2)^9$, $E_6 = (\mathbb{R}^2)^6$. On the one hand, this six features encoding turns out to give independent enough feature (see Section 4), and on the other hand, it enables to get very low NFAs (see Section 5), which means very reliable detections.

Some care must be taken concerning the distances d_i. For example, naively choosing the L^∞-distance for d_i would introduce a bias when computing the product distance d as the maximum over the d_i's. The range of the L^∞-distance with respect to the global feature C_6 is indeed not the same as the range of the other ones. While numerous normalizations are possible, the following dissimilarity measures are built based on the L^∞-distance: for every $i \in \{1, \ldots, 6\}$

$$d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) = \Pr\left(\|x_i(\mathcal{S}) - x(\Sigma)\|_\infty \leq \|x_i(\mathcal{S}) - x_i(\mathcal{S}')\|_\infty\right).$$

Let us point out that the d_i's are not symmetric but are defined with respect to a fixed, predetermined feature, which in fact corresponds to the fixed, predetermined query shape.

This particular normalization choice leads to the following property. Assuming that the K distribution functions $F_i : \delta \mapsto \Pr\left(\|x_i(\mathcal{S}) - x_i(\Sigma)\|_\infty \leq \delta\right)$ are invertible, one has with notations of Definition 2:

$$NFA(\mathcal{S}, \mathcal{S}') = N \cdot \left(\max_{i \in \{1,\ldots,K\}} d_i(x_i(\mathcal{S}), x_i(\mathcal{S}'))\right)^K. \tag{3}$$

Proving this equality simply consists in rewriting the NFA. Let $\delta = \max_{i \in \{1,\ldots,K\}} d_i(x_i(\mathcal{S}), x_i(\mathcal{S}'))$. By Definition 2: $\text{NFA}(\mathcal{S}, \mathcal{S}') = N \cdot \prod_{i \in \{1,\ldots,K\}} P_i(\mathcal{S}, \delta)$. Now, for every $i \in \{1, \ldots, K\}$, the definition of the P_i's implies that $P_i(\mathcal{S}, \delta) = \overline{\Pr}(d_i(x_i(\mathcal{S}), x_i(\Sigma)) \leq \delta) = \overline{\Pr}(F_i(\|x_i(\mathcal{S}) - x_i(\Sigma)\|_\infty) \leq \delta)$. By the invertibility of the F_i's, $P_i(\mathcal{S}, \delta) = \delta$, yielding the result.

In numerical experiments, the NFA is computed with Equation (3), while the d_i's are estimated through the empirical frequencies:

$$d_i(x_i(\mathcal{S}), x_i(\mathcal{S}')) = \frac{1}{N} \cdot \#\Big\{\mathcal{S}'' \in \mathcal{B}, \|x_i(\mathcal{S}) - x_i(\mathcal{S}'')\|_\infty \leq \|x_i(\mathcal{S}) - x_i(\mathcal{S}')\|_\infty\Big\},$$

where $\#\cdot$ denotes the cardinality of any finite set and N is the cardinality of the database $\mathcal{B}$.

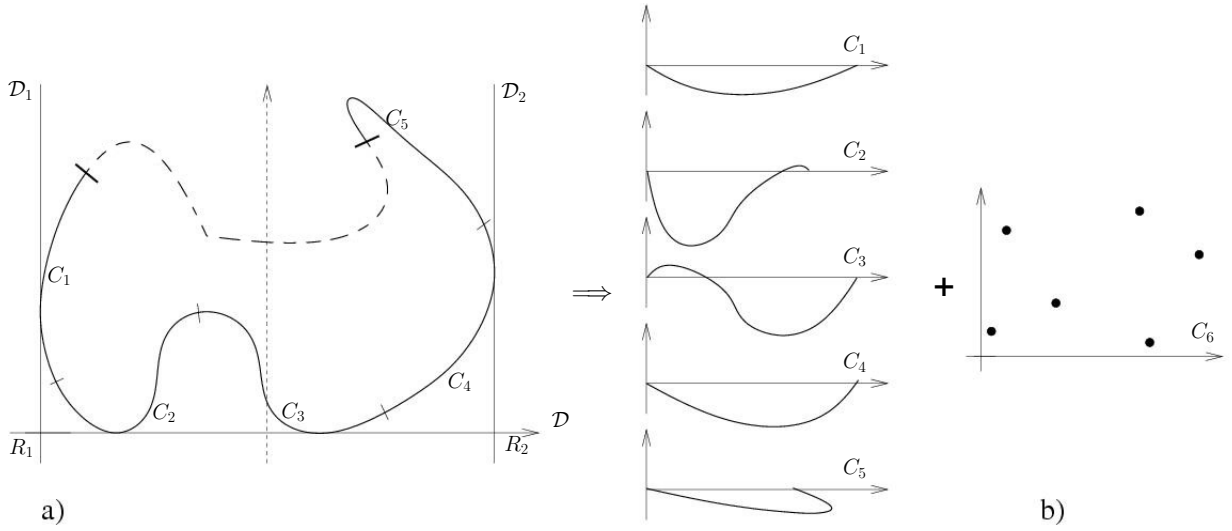


Figure 7: Building independent features. Example of a similarity-invariant encoding (see Section 3.2.1). Sketch a): original shape as a curve in a normalized frame based on a bitangent line. Both ends of the considered curve piece are marked with bold lines: this representation is split into 5 pieces 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.

Remark: Another possibility to get independent or at least uncorrelated features from the shape elements, would be to first apply a whitening technique. However, such techniques need extra assumption on the relation between the observations, that are not likely to be sound when dealing with shapes. For example, we have investigated a principal component analysis (PCA) based model [33]: results are not as good as they should be. Indeed, PCA would be correct provided the feature space is linear, which is clearly not true for the space of shapes. The presented independent features extraction is much more reliable and provides much better experimental results.

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 $\text{NFA}(\mathcal{S}, \delta)$ (Definition 2) 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 [14] states that no detection in "noise" or "among random curves" (in a sense which is to be made clear) should be considered as relevant. All ε-meaningful matches in the noise should thus be considered as false alarms: in such a situation there should be on the average about ε many of them.

If random curves are modeled as random walks with independent increments, and those random

curves taken as database and query (instead of the normalized shape elements of Section 3), then it is clear that the independence assumption holds. As a consequence, the claim according to which there is on average at most ε false alarms (detections, here) among ε-meaningful matches still holds.

Nevertheless, modeling shape elements with random walks is not realistic. On the one hand, shape elements correspond to pieces of level lines, 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 experiment, where random curves are modeled as level lines from white noise images.

Let us consider databases made of normalized shape elements extracted from pieces of level lines in white noise images. Table 1 shows that, although a slight drift can be observed, the order of magnitude of the number of detections is still correct, whatever the size of the database. This property is 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 1, matches with an NFA lower than 0.1 are ensured to be very unlikely in white noise images. If we want to ensure a very 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	$1,000$	$10,000$
$100,000$ 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	$9,997.85$

Table 1: Normalized pieces of level lines from white noise images. Average number over $1,000$ queries of ε-meaningful detections *versus* ε. Three databases of different sizes were tested. The number of ε-meaningful detections is still about ε.

5 Experiments and discussion

In this section, 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 semi-locally encoded. Since the problem of interest is no longer to compare a single shape with a database, but two databases, the NFA definition has to be adapted: the Number of False Alarms of a shape $\mathcal{S}$ (belonging to $\mathcal{B}_1$) at a distance d is

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

For each shape in $\mathcal{B}_1$ we define ε-meaningful matches as in Definition 3. By this way, the claim according 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 2) still holds.

In the following experiments, 1-meaningful matches are highlighted. 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 elements in some sense) are shown here for the sake of clarity. More experiments can be seen in [31] and [42].

What we call "false matches" along this section 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). Experimentally, false matches generally have 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 Recognition threshold is relative to the context

Before considering real images comparison, let us notice that the empirical probabilities take into account the rarity 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 this yields that the distance threshold is larger when estimated with respect to $\mathcal{B}_1$ than to $\mathcal{B}_2$.

The conclusion is that the distance threshold proposed by our algorithm auto-adapts to the relative "rarity" of the query shape among the database shapes. The "rarer" the query shape is, the more permissive the corresponding distance threshold are, and conversely.

5.2 Dealing with partial occlusions and contrast changes

The first experiment consists in comparing the codes extracted from two views of Velázquez' painting *Las Meninas* (see Figure 8). The codes extracted from the query image ($13,851$ codes) are searched for among the codes extracted from the database image ($10,351$ codes). Shape elements

are here normalized with respect to similarity transformations. Note that the query image is a photograph which was taken in the museum: visitors' heads hide a part of the painting.

Figure 9 shows on the left the set of pieces of level lines in the query 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 80 meaningful matches. Only a few false matches can be seen among them. They all have an NFA between 1 and 10^{-1}. In fact, 62 matches show an NFA lower than 10^{-1}. One can notice that some parts of the painting are not retrieved in the database image. In particular, the character standing on the right in the background, or the body of the character on the left are not retrieved. Looking carefully at the level lines, one can realize that these characters are not correctly described by a set of level lines. This situation is quite frequent when dealing with paintings, since contours are usually only suggested and not harshly marked. The picture frame in the background is not retrieved either, although the corresponding level lines are correctly extracted. This is due to the fact that the normalization procedure applied here cannot deal with curves that are too short with respect to the "normalized length" F (*cf* comments at the end of Section 3.2.1).

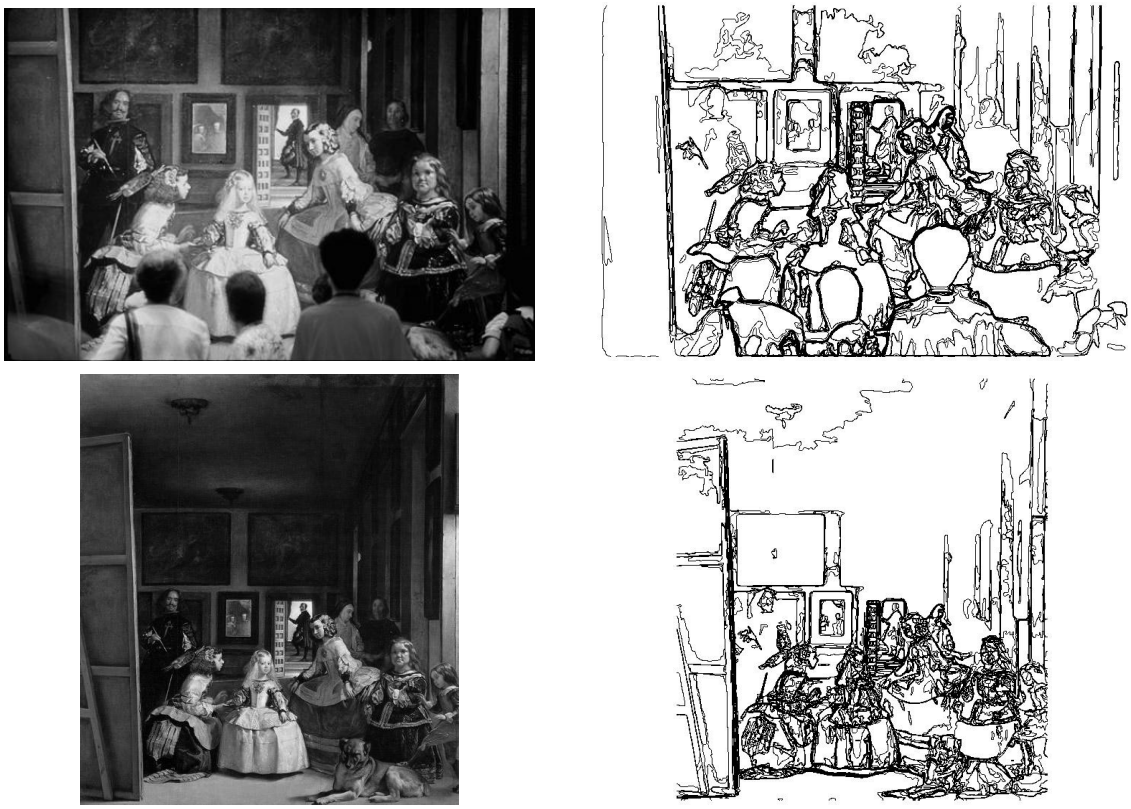


Figure 8: Las Meninas original images (on the left) and meaningful level lines (on the right). Top: query image and its level lines. Bottom: database image and its level lines. Note the contrast change and the partial occlusion between the two views. The codes from the query image are sought among the codes from the database image. Normalization is here with respect to similarity transformations.

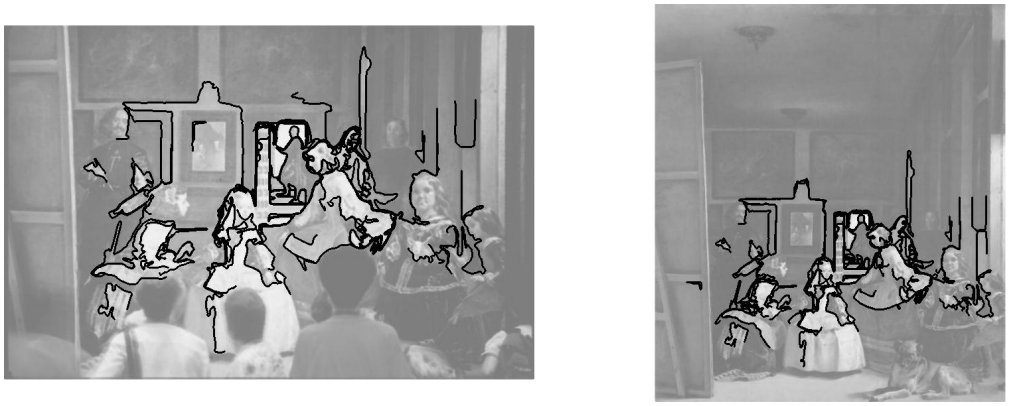


Figure 9: Las Meninas. The 80 1-meaningful matches. Half of them has an NFA lower than $5\ 10^{-4}$. The best match has an NFA equal to $8.5\ 10^{-14}$. To each bold piece of level line on the right corresponds a bold piece of level line on the left.

5.3 Multiple occurrences of the sought shape

The second experiment consists in detecting the "Coca-Cola" logo in the left image in Figure 10. All 1-meaningful matches are displayed in the right image. There are four occurrences of the logo in the image, two of them being strongly occluded, and one of them being very distorted by projection on the can and perspective effect). In spite of these obstacles, many parts of the original logo are retrieved. All false detections show NFA larger than 0.1 (see caption), except for one (see Figure 11), while the majority of the true ones have NFA between 10^{-5} and 10^{-10}. This experiment was led by using the semi-local similarity invariant encoding. Similar results were obtained with the semi-local affine invariant encoding.

5.4 A rejection test

In this experiment, the "Coca-Cola" logo is sought in a database of 150 images, that *do not* contain the logo. The number of tests is about $60 \cdot 10^6$. Less than 10 matches between shape elements were retrieved, all having an NFA larger than 0.1. Moreover, the false matches mostly correspond to characters and exhibit some visual similarity with parts of the logo. Again, the NFA is a good approximation of the number of casual matches that can be expected. As announced by Proposition 2, one should expect at most about one meaningful match. Although the order of magnitude is still correct, a bias is introduced by the fact that some shapes share similar parts because they are produced by a common process which is *not* the background model, such as e.g. human handwriting.

5.5 Similarity invariant matching

Images from this section and the following one come from K. Mikolajczyk's web page[1]. This experiment shows the invariance of the procedure under similarity transformations. The encoding

[1] http://lear.inrialpes.fr/people/Mikolajczyk/Database/index.html

Figure 10: Coca-Cola experiment. Left: image in which the Coca-Cola logo is sought. Right: 1-meaningful matches. 441 shape elements were extracted from the logo, and $30,866$ from the image on the left. There are 105 1-meaningful matches, eight of them are false detections (that is detections that do not correspond to shapes matching between a part of the logo and the corresponding part of a logo in the image). Seven of them have an NFA of $0.95, 0.88, 0.73, 0.60, 0.43, 0.35, 0.13, 0.10$, while one of them has an NFA of $8.7\ 10^{-4}$ (see Figure 11). The best match has $NFA = 6\ 10^{-10}$.

Figure 11: Coca-Cola experiment. One of the false detections has an NFA of $8.7\ 10^{-4}$. One can see that it corresponds to shape elements that are actually "casually" similar, for which a quite low NFA is normal.

procedure is exactly the one described in Section 3. There are 74 matches between normalized shape elements (see Figure 12). Some of them are false and correspond to the top of the grass on the foreground. The problem is actually numeric: since normalized shape elements are all represented by the same number of points (in this case 45), very long and oscillating curves can be too coarsely sampled. Nonetheless, the NFA of these matches is between 0.1 and 1. The lowest NFA is $6.\,10^{-10}$. On Figure 13 we show the registration that can be computed from the matching shape elements. This computation first requires a grouping of the shape elements, which is also based on an *a contrario* approach and is detailed in a general framework in [9]. The left image is the superposition of the two registered images. The right image shows all common pieces of (meaningful) level lines that are almost superposed after registration[2]. This procedure allows to recover a complete geometrical description of common parts between two views, although all pieces of level lines have not been directly matched.

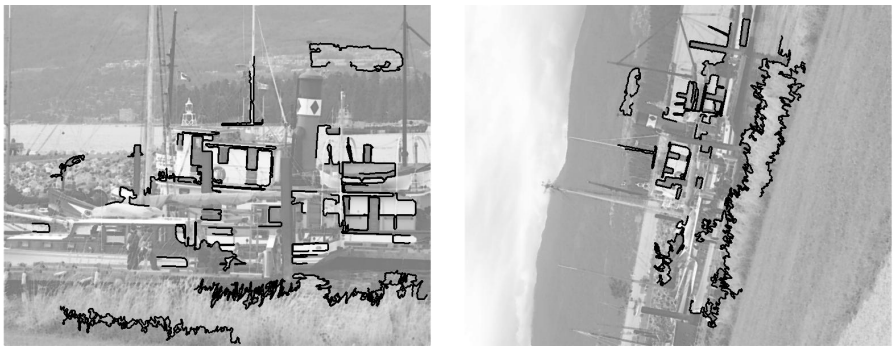


Figure 12: Similarity invariant matching of two photographs. The left image (resp. the right image) leads to 1,983 level lines (resp. 1,149) then to 35,058 encoded shape elements (resp. 23,326) Despite a strong scaling and rotation, there are many meaningful matches on the main shapes. There are also some false matches on the very curvy lines corresponding to the top of the grass. Actually the shape elements corresponding to these level lines are too roughly sampled and do not represent the shape very well. Nevertheless all their NFA is between 0.1 and 1.

5.6 Affine invariant matching and comparison with distance matching

The same matching procedure can be applied for perspective transformation. For planar shapes, the transformation is an homography. In this experiment, projective invariance is replaced by the weaker affine invariance (described in Section 3.2.2). This is sound since shape elements are local. Figure 14 shows the 1-meaningful matches. Due to affine invariance, there are some casual matches between pieces of curves which have an elliptic shape. Figure 15 shows the 0.01-meaningful matches. All the casual matches have disappeared.

Figure 16 shows the registration of the two images through an homography. This homography is computed by a regression based on points from the 1-meaningful local matches. The grouping

[2]Consider two pieces of level lines C_1 and C_2 in the registered image, parameterized by their arc length. Assume that their length is equal and equal to l (here $l = 40$). Then, if $|C_1(s) - C_2(s)| < \delta$ for all $0 \leq s \leq l$, then plot C_2. We take $\delta = 4$ in this experiment.

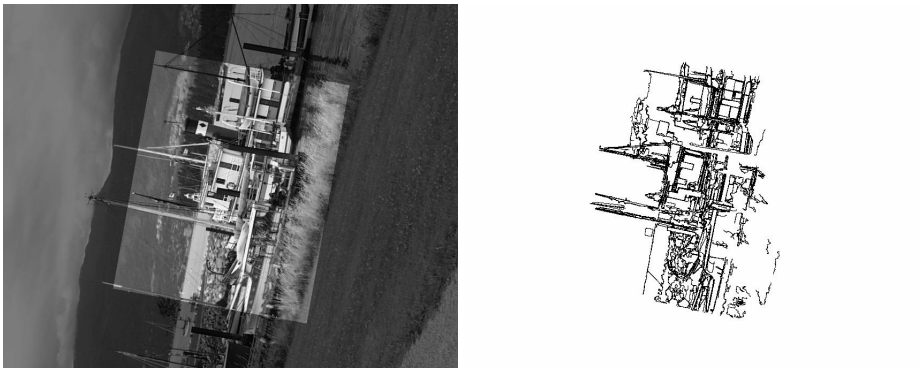


Figure 13: Registration of the two images. Each match uniquely defines a similarity. By using a grouping procedure described in [9], these similarity can be grouped together. In this experiment, a single group is detected and can be used to compute a mean similarity. The two images are then registered. The left image shows the superposition after registration. The right image shows the common pieces of level lines. (See text.)

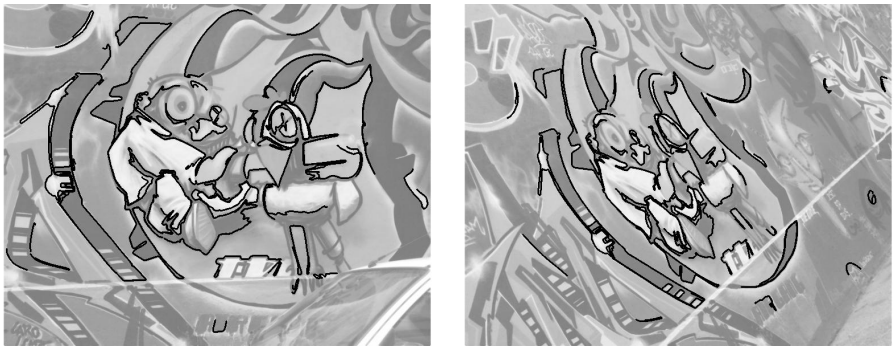


Figure 14: Affine invariant matching of two photographs. The left image (resp. the right image) leads to 792 level lines (resp. 995) then to 14,137 (resp. 12,924) encoded shape elements. There are 625 1-meaningful matches between affine invariant shape elements. The lowest NFA is 10^{-14}.

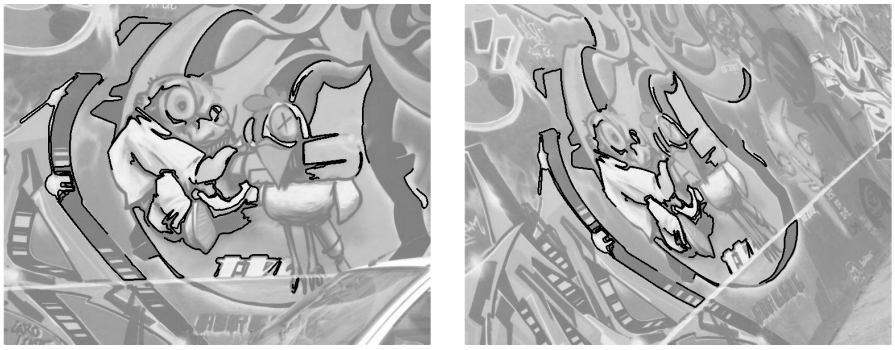


Figure 15: Affine invariant matching. There are 340 0.01-meaningful matches. Let us notice that the threshold on the NFA has a limited influence on the number of (appropriate) detections: dividing the threshold by 100 does not lead to dividing the number of detections by 100.

procedure of [9] is applied before this regression. This automatic procedure definitely eliminates all casual matches that can be still noticed on Fig. 15. The superposition of the images shows that the homography is very accurately estimated.

Figure 16: Registration of the two images. The grouping method of [9] detects a single group of coherent affine transforms. A mean homography is then used to register the two images. One can see that the registration is very accurately estimated. Common pieces of level lines are shown on the right (algorithm described in Sec. 5.5.)

On Figure 17 some results are presented that can be achieved with Lowe's software[3] based on SIFT features [26]. Our purpose here is not to compare the proposed methodology with other ones, but to illustrate what makes it interesting for shape or, possibly, region descriptor matching. A simple matching algorithm is used for the SIFT features: a descriptor from the first image is matched with its nearest neighbor in the second one, provided the ratio of the distances between the nearest and the second nearest neighbor is below some threshold (this latest condition should reinforce the confidence, making this matching algorithm more stable than simply thresholding the distances). The lower the threshold is, the more reliable the matches should be. Despite this, one can observe that false matches are mixed with good matches, even with a decreased threshold (see [28] where this analysis is systematically led with several descriptors). This is not surprising since distances (or ratio of distances) cannot be uniformly thresholded: some statistically rare sought descriptors would deserve a relaxed threshold, while more care should be taken with common descriptors. In order to get rid of these false matches mixed with good matches, a postprocessing based on the Generalized Hough Transform is used (as e.g. in [26]). However, voting thresholds and bins sizes are also touchy parameters that introduce either false positives or mis-detections.

The presented *a contrario* matching procedure precisely takes these statistical properties into account and automatically produces an adaptive distance that depends on the sought shape element. On the contrary, not only is a uniform distance threshold very hard to set, but it is also not suitable. Experiments show indeed that it highly depends on the images that are compared. On the other hand, the NFA threshold can safely be fixed to 1 in most situations.

[3] http://www.cs.ubc.ca/~lowe/keypoints/

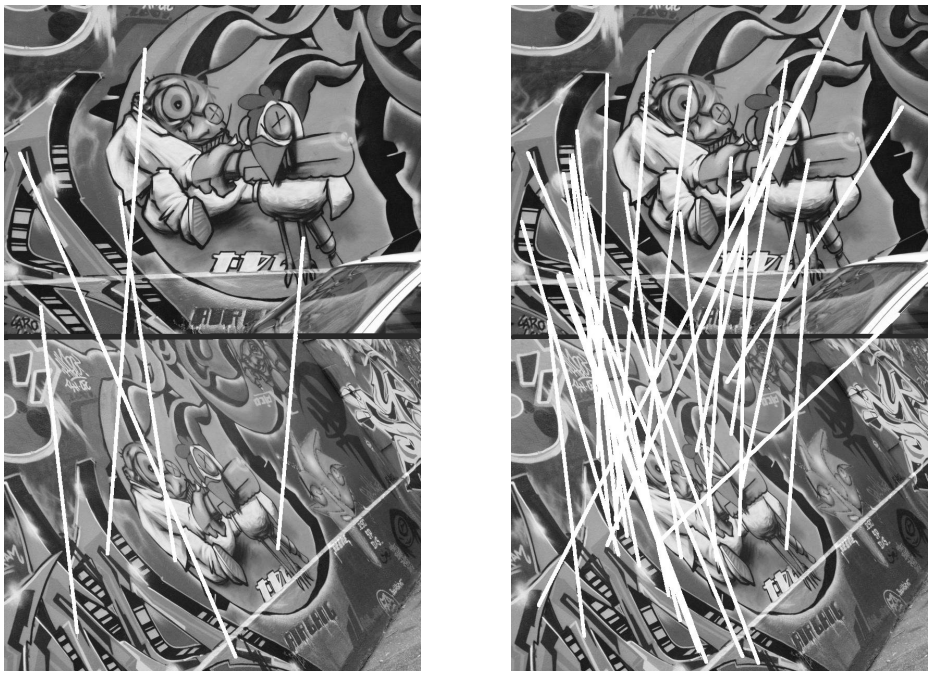


Figure 17: Lowe's SIFT descriptors [26] and a nearest neighbors-like matching strategy. Two matching experiments are shown here, with different values of the matching parameter (ratio of the distances between the two nearest neighbors below 0.5 for the left image and 0.6 for the right one). The matched descriptors are linked with a white straight line. While setting the matching parameter is hard and highly depends on the considered images, one can see that some parts of the images are missed, and that some false detections pass the harder test as well as good ones (2 of the 5 matches are false on the left, and 10 of the 29 on the right). Such results are not due to the descriptors quality, but to the matching strategy by itself.

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 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. This automatically yields a distance threshold that depends on both the database and the query. Moreover, the framework is general and could be adapted for other shape representations (for instance Lowe's SIFT descriptors).

As one can see from the experiments, very reliable detections (those with an NFA roughly lower than 10^{-5}) always correspond to correspondences between parts of the same object. However, if an object shows symmetries, or repetitive parts, one can not distinguish between them. A further step should thus combine the matches, by taking account of their spatial coherence, or relative positions. 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 [9]. Moreover, this clustering stage also automatically wipes out the few false matches with NFA lower than 1 (*i.e.* meaningful matches that do not actually correspond to the same "object"). These matches are actually not distributed over the images in a conspicuous way, unlike "good" matches. As shown in [9], combining the spatial information distribution of matched shape elements strongly reinforces the recognition confidence of the method.

Acknowledgments: 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 developed within MegaWave 2 free software. We thank the anonymous reviewers for their fruitful comments.

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] H. Asada and M. Brady. The curvature primal sketch. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):2–14, 1986.

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

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

[7] R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[8] 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.

[9] F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur. A unified framework for detecting groups and application to shape recognition. Technical Report 5766, INRIA, 2005. Submitted.

[10] F. Cao, P. Musé, and F. Sur. Extracting meaningful curves from images. *Journal of Mathematical Imaging and Vision*, 22(2-3):159–181, 2004.

[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. *Computational Gestalt Theory*. Lecture Notes in Mathematics, Springer Verlag, 2005. To appear.

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

[16] M.A. Fischler and R.C. Bolles. Perceptual organization and curve partitioning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):100–105, 1986.

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

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

[19] 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.

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

[21] 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.

[22] 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.

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

[24] 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.

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

[26] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

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

[28] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. To appear in IEEE Pattern Analysis and Machine Intelligence, 2005.

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

[32] 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.

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

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

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

[36] 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.

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

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

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

[40] 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.

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

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

[43] 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.

[44] 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.

[45] 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.

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

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

[48] S.C. Zhu. Embedding Gestalt laws in Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1170–1187, 1999.