Shape-based Invariant Texture Indexing

Gui-Song Xia, Julie Delon, Yann Gousseau

Shape-based Invariant Texture Indexing

Gui-Song Xia, Julie Delon, Yann Gousseau

Abstract This paper introduces a new texture analysis scheme, which is invariant to local geometric and radiometric changes. The proposed methodology relies on the topographic map of images, obtained from the connected components of level sets. This morphological tool, providing a multi-scale and contrast-invariant representation of images, is shown to be well suited to texture analysis. We first make use of invariant moments to extract geometrical information from the topographic map. This yields features that are invariant to local similarities or local affine transformations. These features are invariant to any local contrast change. We then relax this invariance by computing additional features that are invariant to local affine contrast changes and investigate the resulting analysis scheme by performing classification and retrieval experiments on three texture databases. The obtained experimental results outperform the current state of the art in locally invariant texture analysis.

Keywords: Topographic map, level lines, texture analysis, local invariance.

1 Introduction

Texture is widely considered as a fundamental ingredient of the structure of natural images. The analysis of texture, though, is a long standing and challenging problem in image processing and computer vision. Yves Meyer recently coined texture as "a subtle balance between repetition and innovation" (Meyer, 2007). Indeed, the repetitive nature of texture oriented some of

Telecom ParisTech, LTCI CNRS,
46, rue Barrault, 75013 Paris, France
Tel.: +33-1-45818139
Fax: +33-1-45817144

the very early research on automatic texture discrimination toward frequency or autocorrelation analysis, see e.g. (Kaizer, 1955; Chen, 1982). Next, in order to deal with local transitions as well as with the "innovative" part of textures, one has favored localized, Gabor or wavelet-like analysis, see e.g. (Jain and Farrokhnia, 1991). The ability of such mathematical tools to handle multi-scale structures has made them one of the most popular tool for analyzing textures. One limitation of such approaches, however, lies in their difficulty in efficiently representing the geometrical aspects of textures, such as sharp transitions and elongated contours. In order to overcome this difficulty, alternative wavelet-like approaches have been proposed to enable more efficient representations of structured textures, see e.g. (Peyré, to appear, 2009).

The Mathematical Morphology school has long ago (Haas et al, 1967; Serra, 1982) proposed a radically different multi-scale analysis tool for texture, the so-called granulometry. These are obtained from an image by applying elementary morphological operations with structuring elements of increasing sizes. Because such basic morphological operations operate on the level sets of images, the resulting analysis enables a direct handling of edges and shapes contained in textures. In this work, we show that a morphological and multi-scale decomposition of images, the topographic map as introduced by Caselles *et al.* (Caselles et al, 1999a), enables one to perform efficient texture analysis, while being invariant to *local* radiometric and geometrical changes.

Indeed, a challenging issue when analyzing texture is that texture surfaces are usually perceived under unknown viewing conditions. Except when dealing with a controlled image acquisition protocol, for instance in specific industrial applications, texture analysis methods should comply with some invariance requirements. The most basic ones are translation, scale and orienta-

tion invariances. It is also desirable to achieve invariance to some contrast changes, in order to deal with variable lighting conditions. Next, the requirement of invariance with respect to viewpoint changes for flat texture yields analysis that are invariant with respect to affine or projective transforms. Moreover, textures can live on non-flat surfaces, as it is the case for bark on a tree or for folded textiles. Such an example is shown in Figure 1, where two different samples of the same texture class (plaid) from the UIUC database (Lazebnik et al, 2005) are displayed. Several recent approaches to the analysis of such textures rely on the extraction of local features that are individually invariant to some geometric transforms, such as similarity or affine transforms (Lazebnik et al, 2003; Mellor et al, 2008). In contrast with previous works dealing with invariant 3D texture analysis, such *locally invariant* methods do not need any learning of the deformations (Varma and Zisserman, 2002; Leung and Malik, 2001) or explicit modeling (Wu and Chantler, 2003) of the 3D surfaces. In this paper, we show that a morphological analysis relying on connected components of level sets enables retrieval and classification of textures that equal or outperform the existing locally invariant approaches on several databases.
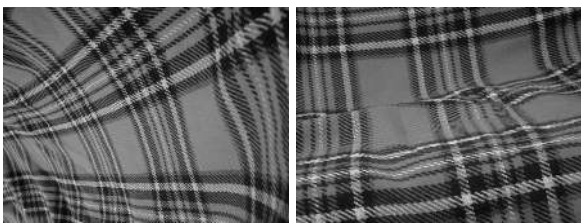

Fig. 1 Two samples of the same texture class from the UIUC database (Lazebnik et al, 2005). This texture lies on non-rigid surfaces implying complex deformations between the samples.

1.1 Previous and related work

This section briefly summarizes different directions that have been explored for the invariant analysis of texture images. Texture analysis has been a very active research field over the last four decades, and an exhaustive study of this field is of course beyond the scope of this paper. Some surveys and comparative studies of existing methods can be found in (Haralick, 1979; Tuceryan and Jain, 1993; Reed and du Buf, 1993; Randen and Husoy, 1999; Zhang and Tan, 2002), the last one being devoted to invariant texture analysis. In what follows, we first focus on classical approaches and the type of global invariances they allow. By global invariances, we mean invariances to global transforms of the image. We then summarize recent approaches to the analysis of texture that are invariant under local transforms of images. We focus on methods that are invariant *by design* and do not include in this short discussion methods that are invariant as the result of a learning process (Varma and Zisserman, 2002; Leung and Malik, 2001) or an explicit modeling of 3D textures surfaces (Wu and Chantler, 2003).

The use of co-occurrence matrices (Haralick et al, 1973; Davis et al, 1979) is still a popular approach, relying on non-parametric statistics at the pixel level. It is also worth noticing that this path along non-parametric statistics has been very fruitful for the purpose of texture synthesis (Efros and Leung, 1999). Rotation invariance can be achieved for such indexing methods by using polar coordinate systems, as detailed in (Davis, 1981). In a related direction, Pietikäinen *et al.* (Pietikäinen et al, 2000; Ojala et al, 2002) propose a rotation invariant local binary pattern (joint distribution of gray values on circular local neighborhoods) to describe texture images. Still at the pixel level, Kashyap and Khotanzad (Kashyap and Khotanzad, 1986) developed rotation invariant autoregressive models. Cohen *et al.* (Cohen et al, 1991), among others, have introduced rotation invariant Gaussian Markov random fields to model textures. However, the design of scale invariant Markov random field rapidly implies very involved computations, see e.g. (Gidas, 1989). Of course, pixel statistics can be averaged over different neighborhoods and make use of multi-resolution schemes, but these statistics are certainly not the easiest way to achieve scale or affine invariant analysis of textures.

A second popular and efficient way to analyze textures relies on filtering. Many works have focused on different filter bank families, different sub-band decompositions, and on the optimization of filters for texture feature separation, see e.g. (Simoncelli and Portilla, 1998; Randen and Husoy, 1999; Sandler and Lindenbaum, 2009). Many of these approaches enable translation invariance (by using over-complete representations), rotation and scale invariance, by using effective filter designs, see e.g. (Chen and Kundu, 1994; Do and Vetterli, 2002; Pun and Lee, 2003; Mellor et al, 2008). Some contrast invariance can also be achieved by normalizing responses to filters.

As already mentioned, an alternative approach to the analysis of textures has been proposed by the mathematical morphology school in the framework of granulometry. The idea is to characterize an image by the way it evolves under morphological operations such as opening or closing when the size of the structuring elements is increased (Serra, 1982; Maragos, 1989). These ideas have been successfully applied to the classifica-

tion of textures, see e.g. (Chen and Dougherty, 1994; Asano et al, 2000), as well as the related approach (Ayala and Domingo, 2001), making use of stochastic geometry. Several works rely on the theory of connected operators (Salembier and Serra, 1995) to compute granulometry without the need for structuring elements, see (Li et al, 1997; Fletcher and Evans, 2005), thus potentially enabling greater geometrical invariances. However, there are few works showing the benefit of the geometrical nature of morphological operators to achieve similarity or affine invariant texture classification, with the notable exception of (Urbach et al, 2007), where a shape-size pattern spectra is proposed as a way to classify images. In particular, it is shown that this spectra enables rotation-invariant classification of texture images. In (Hamdan and Larson, 2002), it is proposed to globally use the Earth Mover's Distance between topographic maps to perform scale invariant texture classification. To the best of our knowledge, no work has proposed the use of morphological attributes to achieve viewpoint invariant description of textures. Concerning radiometric invariant analysis of texture, the benefit of using contrast invariant morphological operators to recognize texture under various illumination conditions has not yet been demonstrated. Authors of (Hanbury et al, 2005) have developed an illumination invariant morphological scheme to index textures, but they achieve invariance thanks to histogram modification techniques and not by using the contrast invariant properties of morphological analysis.

Fractal geometry has also been used in the description of textures, see e.g. the early work (Peleg et al, 1984). Such approaches have been shown to enable globally invariant texture analysis. Recently, Xu *et al* (Xu et al, 2006) proposed the use of multifractal spectrum vectors to describe textures while achieving global invariance under bi-Lipschitz transforms, a general class of transforms which includes perspective transforms and smooth texture surface deformations.

Recently, several works proposed to use individually normalized local features in order to represent textures while being locally invariant to geometric or radiometric transforms, see (Lazebnik et al, 2005; Zhang et al, 2007; Varma and Garg, 2007; Mellor et al, 2008). In (Lazebnik et al, 2005) and (Zhang et al, 2007), a set of interest local affine regions are selected to build a sparse representation of textures relying on affine invariant descriptors. Textures are represented thanks to bag-of-features, a method that has been proved very efficient to recognize object categories, see e.g. (Li and Perona, 2005). In (Varma and Garg, 2007), textures are characterized statistically by the full joint PDF of their local fractal dimension and local fractal length, and this approach is shown to be discriminative and affine invariant. Very

recently, Mellor *et al.* (Mellor et al, 2008) have shown that similar local invariances can be obtained using a filter bank approach. These authors develop a new family of filters, enabling a texture analysis that is locally invariant to contrast changes and to similarities.

1.2 Our Contributions

As explained earlier in the introduction, the goal of this paper is to introduce a new method for texture analysis that in spirit is similar to morphological granulometries, while allowing a high degree of geometrical and radiometric invariances. The approach relies on the complete set of level lines of the image, the so-called topographic map, introduced by Caselles *et al.* (Caselles et al, 1999a). The *shapes* (that is, the interiors of the connected components of level lines) are the basic elements on which the proposed texture analysis is performed. We exhibit a set of simple statistics on these shapes, obtained using classical invariant shape moments. Therefore, and because each shape is individually normalized, the proposed texture indexing is invariant to local geometrical transforms, allowing for the recognition of non-rigid textures. Various experiments of texture classification and retrieval demonstrate the efficiency of the proposed analysis method on various databases.

The paper is organized as follows. First, in Section 2, we briefly recall the definition and elementary properties of the topographic map. Next, in Section 3 local features based on the topographic map are defined. In Section 4, the ability of these features to classify or retrieve texture is demonstrated on three databases: Brodatz's texture photo album (Brodatz, 1966), UIUC database (Lazebnik et al, 2005) and UMD database (Xu et al, 2009). Discussions, experiments involving unsupervised learning and exploring the scaling behavior of the method as well as an application to segmentation are also included in this section. A short version of this work has appeared in (Xia et al, 2008).

2 Topographic map

In this section, we recall the definition of the topographic map and its main properties. The topographic map has been suggested as an efficient way to represent images by Caselles *et al.* (Caselles et al, 1997, 1999a). It is made of the level lines, defined as the connected components of the topological boundaries of the level sets of the image. As we shall see, this map inherits a tree structure from the nesting properties of level sets and is an elegant way to completely represent the geometric

information of an image while remaining independent of the contrast.

The upper level sets of an image $u : \Omega \mapsto \mathbb{R}$ are defined as the sets

$$\chi_\lambda(u) = \{x \in \Omega; \ u(x) \geq \lambda\},$$

where $\lambda \in \mathbb{R}$. We can define in the same way the lower level sets $\chi^\lambda(u)$ of u by inverting the inequality. Remark that if φ is a strictly increasing contrast change, then

$$\chi_{\varphi(\lambda)}(\varphi(u)) = \chi_\lambda(u),$$

which means that the set of all upper level sets remains the same under increasing contrast changes. Moreover, the image is completely described by its upper level sets. Indeed, u can be reconstructed thanks to the following formula

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

Of course, the same property holds for lower level sets. Now, observe that these upper (lower) level sets constitute a decreasing (increasing) family. Indeed, if λ is greater than μ, then $\chi_\lambda(u)$ is included in $\chi_\mu(u)$ (and conversely $\chi^\lambda(u)$ contains $\chi^\mu(u)$). It follows that the connected components of upper level sets (respectively of the lower level sets) are naturally embedded in a tree structure. Several authors (Salembier and Serra, 1995; Caselles et al, 1997; Heijmans, 1999), have proposed to use these trees of connected components (one for the upper level sets, one for the lower level sets) as an efficient way to represent and manipulate images, thanks to their hierarchical structure and their robustness to *local* contrast changes. Observe that the *maximally stable extremal regions* (MSER) (Matas et al, 2002) detector in images also relies on connected component of level sets.

Now, the notion of level lines (topological boundaries of level sets) enables to merge both trees, which motivates further the use of the topographic map to represent images. Monasse and Guichard fully exploited this fact and, drawing on the notion of *shape*, developed an efficient way to compute this hierarchical representation of images (Monasse and Guichard, 2000a), called Fast Level Set Transform (FLST). A shape is defined as a connected component of an upper or lower level set, whose holes have been filled. A hole of a set A in an image is defined as a connected component of the complementary set of A that does not intersect the border of the image. It is shown in (Monasse and Guichard, 2000a) that the set of shapes of an image has a tree structure. Under some regularity assumption on the image, this tree is equivalent to the topographic map (that is the set of all level lines). For discrete images, the only

technicality needed in order to define the shapes is that two different notions of connectivity should be adopted for level sets : 8-connectivity for upper level sets and 4-connectivity for lower sets (the opposite convention could of course be adopted). For more precision and results on the topographic map, we refer to the recent monograph (Caselles and Monasse, 2008). For the experiments performed in this paper, we compute the topographic maps using the FLST code available in the free processing environment *Megawave2*[1]. For a recent alternative to the computation of the topographic map, see (Song, 2007). An example of the representation of a synthetic image by its topographic map is shown in Fig. 2.
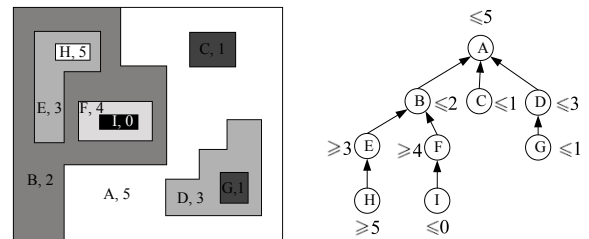

Fig. 2 Representation of an image by its topographic map (this example is taken from (Monasse, 2000)). Left: an original digital image, with gray levels from 0 to 5; Right: representation of the image by its tree of shapes, where $(A, B, \ldots, I)$ denote the corresponding shapes.

The topographic map has a natural scale-space structure, where the notion of scale corresponds to the areas of the shapes (Monasse and Guichard, 2000b). This is of course a first motivation to investigate its use for texture analysis. Moreover, because it is made of the level lines of the image, the topographic map permits to study textures at several scales without geometric degradation when going from fine to coarse scales. This is actually a very strong property of this scale-space representation. Contrarily to approaches using the linear scale space or linear filtering, it allows a faithful account of the geometry at all scales. Figure 3 illustrates this ability. This figure shows a needlework texture, in which the smallest scales represent the fine net of the needlework, while the large scales capture the boundaries of the flowers that are represented.

Next, the topographic map is invariant to any increasing contrast change. In fact, it is even invariant to any *local* contrast change as defined in (Caselles et al, 1999b). This property is of primary interest to define texture analysis schemes that are robust to illumination changes. At this point, it is important to add that
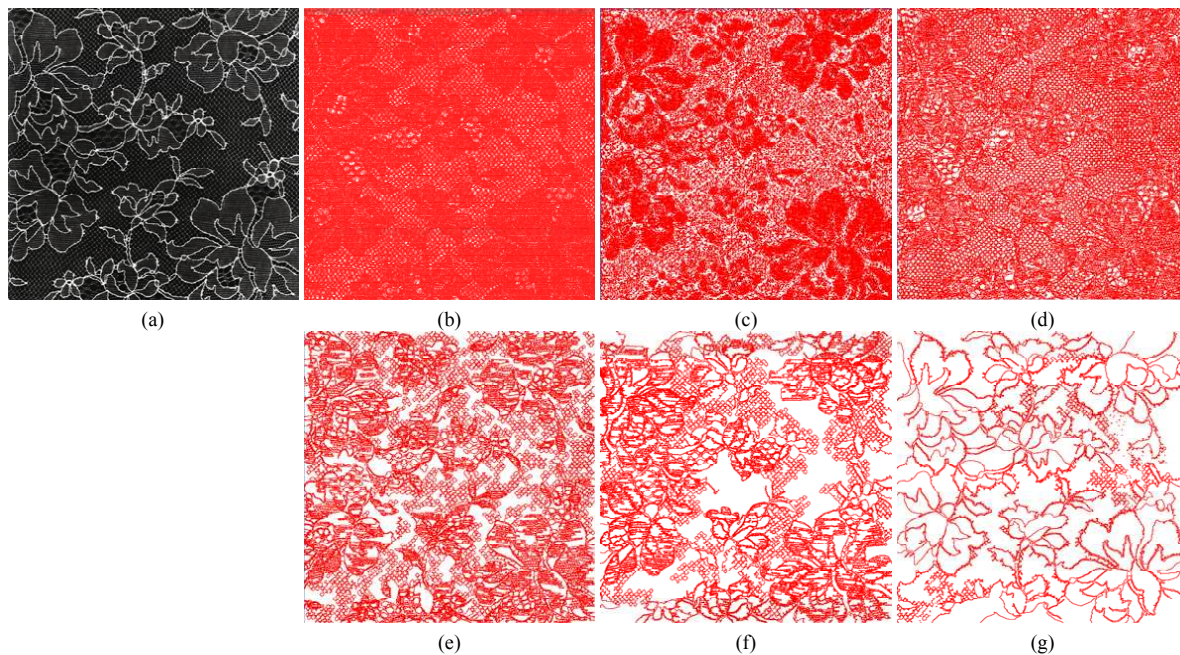
[1] http://www.cmla.ens-cachan.fr/Cmla/Megawave/

Fig. 3 Representation of a texture image by its topographic map. (a) original texture image D41 (of size 640×640) taken from Brodatz's photo album (Brodatz, 1966); (b) all shapes boundaries; (c)-(g) shape boundaries at different scales, respectively for shapes of areas in $[1, 10]$, in $[11, 125]$, in $[126, 625]$, in $[626, 3125]$, and in $[3126, 409600]$.

while individual lines may not be strictly invariant to illumination changes, *marginals* of the geometrical attributes of lines are, as will be demonstrated by the experimental section.

Last, the basic elements of the topographic map are shapes obtained from connected components of the level sets. Therefore, it provides a local representation of the image. As we shall see, this locality, combined with the fact that the topographic map is by nature a geometric representation of images, enables us to develop analysis schemes that are invariant to local geometrical distortions.

Now, it remains to show that the set of level lines contains pertinent information about the structure of textures. This fact is suggested in the original paper on the topographic map of images (Caselles et al, 1999a), where it is stated that "no matter how complicated the patterns of the level lines may be, they reflect the structure of the texture". A first attempt at using the topographic map to classify texture images has been proposed in (Hamdan and Larson, 2002). In the context of satellite imaging, scales computed from contrasted level lines have proven useful to discriminate between different textured areas (Luo et al, 2009). The use of level lines in the context of texture synthesis has also been investigated in (Gousseau, 2002). In the remaining of this work, we show the usefulness of level lines to index textures while being robust to viewpoints and illumination changes.

3 Invariant Texture Descriptors

The goal of this section is to define texture features that are both invariant to some geometric changes and discriminative enough. These features will be obtained from the shapes of the topographic map and it is therefore quite natural to consider the classical invariant shape moments, whose definition is recalled in this section. Observe that such shape moments are already used for image registration in (Monasse, 1999) and texture recognition in (Hamdan and Larson, 2002). However, it is well known that these moments rapidly loose robustness as their order increases, so that only a small number of these can be used to analyze real world textures. In order to enrich the proposed analysis, we take into account multi-scale shape dependencies on the topographic map. The resulting features are invariant to *any* local contrast change. Last, we suggest some contrast information that can be extracted from the shapes and will allow to improve the discriminative power of the proposed analysis scheme while still being invariant to local affine contrast changes.

3.1 Marginals of invariant moments

In this section, we first give a short reminder on the invariant moments that can be extracted from the inertia matrix of a shape, focusing on invariances to similarity and affine transforms. More information on this classi-

cal subject can be found e.g. in (Hu, 1962; Flusser and Suk, 1993; Liao and Pawlak, 1996; Zhang and Lu, 2004). Then, we show how this moments can be applied to shapes of the topographic map in order to perform locally invariant texture analysis.

3.1.1 Invariant moments reminder

For p, q integer values, the two-dimensional $(p+q)th$ order central moment $\mu_{pq}(s)$ of a shape $s \subset \mathbb{R}^2$ is defined as

$$\mu_{pq}(s) = \int \int_s (x - \overline{x})^p (y - \overline{y})^q \, dxdy, \tag{1}$$

where $(\overline{x}, \overline{y})$ is the center of mass of the shape, *i.e.*

$$\overline{x} = \frac{1}{\mu_{00}(s)} \int \int_s x \, dxdy, \text{ and } \overline{y} = \frac{1}{\mu_{00}(s)} \int \int_s y \, dxdy. \tag{2}$$

For the sake of simplicity, we will omit the variable s in the following and write μ_{pq} instead of $\mu_{pq}(s)$. Note that μ_{00} is the area of the shape and that all central moments μ_{pq} are invariant to **translations**.

In order to achieve invariance to **scale changes**, it is well known and easily shown that moments have to be normalized in the following way

$$\eta_{pq} = \mu_{pq}/\mu_{00}^{(p+q+2)/2}. \tag{3}$$

As a consequence, any function of the normalized moments η_{pq} is invariant to both scale changes and translations of the shape s. Now, the sensitivity to noise of these moments quickly increases as their order increases. We observed experimentally that moments of order bigger than two are not robust enough to faithfully account for texture characteristics, and we therefore limit the analysis to moments of order smaller than 2. Since $\eta_{00} = 1$ and $\eta_{01} = \eta_{10} = 0$, invariant features are all obtained from the normalized inertia matrix

$$C = \begin{pmatrix} \eta_{20} & \eta_{11} \\ \eta_{11} & \eta_{02} \end{pmatrix}. \tag{4}$$

In order to achieve **rotation** invariance, only two features remain, namely λ_1 and λ_2, the two eigenvalues of C, with $\lambda_1 \geq \lambda_2$. Observe that using these values boils down to fit to the shape an ellipse with semi-major axis $2\sqrt{\lambda_1}$ and semi-minor axis $2\sqrt{\lambda_2}$. Note also that from the seven similarity invariants proposed in the seminal work by Hu (Hu, 1962), the only ones of order two are $\lambda_1 + \lambda_2$ and $(\lambda_1 - \lambda_2)^2$. Now, any function of λ_1 and λ_2 would also be invariant to similarity. We chose to use

$$\epsilon = \lambda_2/\lambda_1, \tag{5}$$

and

$$\kappa = \frac{1}{4\pi\sqrt{\lambda_1 \lambda_2}}, \tag{6}$$

because these invariants have a clearer intuitive meaning and a simpler range than Hu's moments. The first one lies between 0 and 1 and describes the *elongation* or the flatness of the shape. It can be shown that the second one also lies between 0 and 1. This invariant can be seen as a measure of the *compactness* of the shape, which reaches its maximum at ellipses. Indeed, κ is a dimensionless ratio between the area of the shape (1 for a normalized shape) and the area of the best ellipse fitting the shape. Note that this invariant is more robust than a measure relying on the boundary of the shape, such as the isoperimetric ratio $\frac{4\pi}{p^2}$ (where p is the perimeter of the shape). Next, observe that κ (but not ϵ) is further invariant to **affine** transforms. In fact, κ^{-2} is the first affine invariant of Flusser *et al.*, defined in (Flusser and Suk, 1993).

3.1.2 Texture features from second order moments

As a first feature to represent textures, we simply compute the marginals over all shapes of the two features κ and ϵ. More precisely, for each of these two features, we compute a 1D-histogram by scanning all the shapes of the topographic map. The resulting 1D-histograms are invariant to *any local contrast change*, even decreasing ones. Now, it is well known that contrast inversion strongly affects the visual perception. For this reason, we restrict the invariance to *any local increasing contrast change* (Caselles et al, 1999b) by splitting each of the previous 1D-histograms in two histograms, one for shapes originating from upper level sets (bright shapes) and one for shapes originating from lower level sets (dark shapes). The concatenations of the bright and dark histograms are called respectively *elongation histogram* (EH) and *compactness histogram* (CpH).

Observe that since moments are individually normalized for each shape, the resulting features are invariant to local geometrical changes (similarity for EH and affinity for CpH). More precisely, applying a different geometrical transform on each shape does not affect the overall marginals of κ and ϵ. In particular, this should allow to recognize texture that have undergone non-rigid transforms.

3.2 Dependencies in the topographic map

As explained in the previous section, requiring geometrical invariances and robustness restricts the number

of possible invariant moments to two. In order to define new features from the topographic map without going into complex geometrical descriptors relying e.g. on the boundary of shapes, it is natural to take shape dependencies into account. Indeed, invariant moment marginals as defined in the previous section do not reflect the relative positions or inclusions between shapes. Let us illustrate this point by a toy-example. Figure 4 shows two simple synthetic textures and their corresponding topographic maps. These two images share the same histograms EH and CpH, in spite of their structural differences.
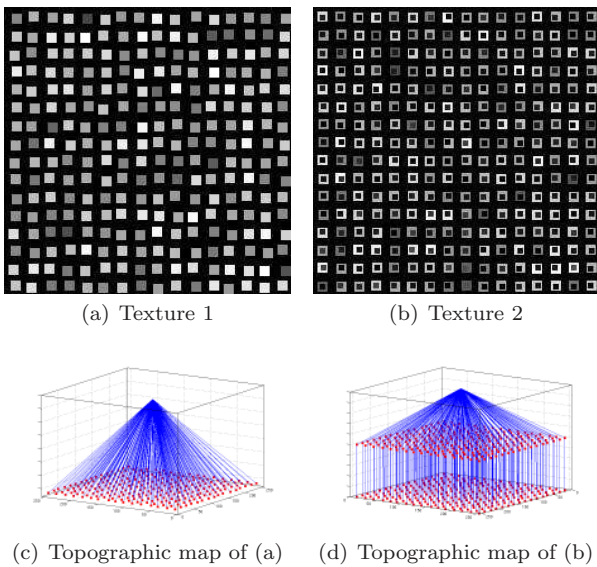

(a) Texture 1   (b) Texture 2

(c) Topographic map of (a)   (d) Topographic map of (b)

Fig. 4 Toy example: two synthetic textures and their corresponding topographic maps. Both images have the same shape marginals but different tree structures, as shown in (c) and (d).

We claim that the topographic map, because of its hierarchical structure, enables the extraction of shape dependency in an easy and intuitive way. In this work, we focus on children-parents relationships within the tree, although other relationships could be interesting.

Definition (*Ancestor family* $\mathcal{N}^M$) Let s be a shape of the image. Let s^m be the m-*th* cascaded ancestor of s, where m is an integer. That is, s^1 is the parent shape of s, s^2 the parent shape of s^1, etc. For $M \geq 1$, the Mth ancestor family of s is defined as $\mathcal{N}^M = \{s^m, 1 \leq m \leq M\}$.

Now, it is quite simple to extract affine invariant information from these ancestor families. Recall that $\mu_{00}(s)$ is the area of the shape s. An affine transformation $AX + b$ on s changes $\mu_{00}(s)$ into $det(A)\mu_{00}(s)$. As a consequence, if we define for any shape s

$$\alpha(s) = \frac{\mu_{00}(s)}{\langle \mu_{00}(s') \rangle_{s' \in \mathcal{N}^M}}, \tag{7}$$

where $\langle \cdot \rangle_{s' \in \mathcal{N}^M}$ is the mean operator on $\mathcal{N}^M$, then α is locally affine invariant, in the sense that for each shape s, $\alpha(s)$ is only sensitive to transformations applied to its M direct ancestors. Remark also that $0 < \alpha < 1$. Again, the distribution of α is represented by a 1D-histogram, split into dark and bright shapes. The corresponding feature is called *scale ratio histogram (SRH)*.

Remark Other features could be extracted from the ancestor family, built e.g. from elongation or compactness as defined in the previous section. However for the purpose of texture indexing, and in particular for the classification and retrieval tasks to be considered in the experimental section, we did not find them to be overly discriminative. These could however be useful for different tasks.

In what follows, we use two sets of texture features. The first one, called SI, is made of the features that are invariant to (local) similarity transforms, while the second one, called AI, is made of the (locally) affine invariant features. That is,

- $SI = \text{CpH+SRH+EH}$,
- $AI = \text{CpH+SRH}$,

where, as defined before, EH stands for elongation histogram, CpH for compactness histogram and SRH for scale ratio histogram. These are *geometric* features, in the sense that they are invariant to any (local) increasing contrast change. We believe that these descriptors illustrate the usefulness of the topographic map to analyze texture images, in particular allowing for relatively easy handling of invariances.

3.3 Contrast information

The previous geometric features are invariant to *any local increasing contrast change*, as defined in (Caselles et al, 1999a). This is a very strong invariance and we are not aware of any texture analysis scheme having this property. Now, we observed that this invariance is too strong to efficiently recognize many texture classes. In this section, we define contrast features that are invariant to *local affine contrast changes*. This is coherent with the contrast invariances considered in recent works to which we will compare our results, such as (Lazebnik et al, 2005; Mellor et al, 2008; Xu et al, 2006).

We choose to compute intensity histograms after local normalization by mean and variance on a neighborhood. Such photometric normalization approaches are relatively standard and have been used in local descriptors, see (Obdržálek and Matas, 2002; Schaffalitzky and Zisserman, 2001). Schaffalitzky *et. al* (Schaffalitzky and Zisserman, 2001) enable their texture descriptors to be invariant

to local affine illumination changes by normalizing the intensity of each point by the mean and standard deviation over a local adaptive neighborhood (a support region with detected adaptive scale). We follow a similar path, except that we rely on the topographic map to define local neighborhoods.

More precisely, at each pixel x, a normalized grey level value is computed as

$$\gamma(x) = \frac{u(x) - \text{mean}_{s(x)}(u)}{\sqrt{\text{var}_{s(x)}(u)}}, \qquad (8)$$

where $s(x)$ is the smallest shape of the topographic map containing x, $\text{mean}_{s(x)}(u)$ and $\text{var}_{s(x)}(u)$ are respectively the mean and the variance of u over $s(x)$. This results in a *contrast histogram (CtH)*, computed by scanning all pixels of u. Thanks to the adopted normalization, the resulting feature is invariant to local affine contrast changes, as the features in (Lazebnik et al, 2005; Mellor et al, 2008; Xu et al, 2006).

One particularity of the proposed normalization (8) is that the normalized value $\gamma(x)$ at x will generally be negative for shapes coming from an upper level set, and positive for shapes coming from a lower level set (this property is not systematic but very often satisfied on natural images).

Observe that this last feature, CtH, is not invariant to *local* similarity (or affine) transforms. Indeed, contrast histograms are computed on a pixel by pixel basis which breaks the geometrical invariances we add preserved so far. Now, we observed that this feature is very robust to geometrical distortions of the textures, even in some extreme cases, as will be demonstrated by the experimental section.

4 Experiments

In this part, we first explain in Section 4.1 how to compare texture images using the features introduced in the previous section. We then investigate in Section 4.2 the performances of the resulting comparison scheme by confronting it with state-of-the-art texture descriptors. More precisely, we follow the experimental protocols presented in (Lazebnik et al, 2005) and reproduced in (Mellor et al, 2008). These protocols consist of retrieval and classification tasks. In order to meet the standards of the current literature in texture indexing, these experiments are performed on three different databases, namely the classical Brodatz database, the UIUC database (Lazebnik et al, 2005) and the more recent UMD database (Xu et al, 2009). On these three databases, the descriptors introduced in this paper show similar or better results than the descriptors presented in (Lazebnik et al,

2005; Mellor et al, 2008; Xu et al, 2006). For the sake of completeness, all the results of our retrieval experiments are available at the Internet address (Xia, 2009). We also investigate the scaling properties of the proposed scheme on the reunion of the three aforementioned databases.

After these comparative experiments, we show in Section 4.3 that the proposed texture indexing scheme can strongly benefit from an unsupervised learning procedure. We show how the recognition performances are enhanced through the use of manifold learning. Then, Section 4.4 is devoted to a discussion on invariance to resolution changes (illustrated by experiments on our own high resolution texture database) as well as on the trade-off between invariance and discriminative power. Last, it is suggested in Section 4.6 that the features introduced in this paper enable one to segment images made of spatially varying textures regions.

For all experiments of this section, histograms EH, CpH and SRH are computed over 25 bins for bright shapes and 25 bins for dark shapes. Histogram CtH is computed over 50 bins. The value of M used to compute SRH is set to $M = 3$.

4.1 Descriptors comparison

Two texture samples u and v are compared through the distribution of features, simply by comparing the corresponding histograms. We choose to compare histograms through Jeffrey divergence, a symmetric modification of the Kullback-Leibler (K-L) divergence.

For two discrete distributions $P = (p_1, \ldots p_N)$ and $Q = (q_1, \ldots q_N)$, the Jeffrey divergence between P and Q is defined as

$$\mathcal{D}(P, Q) = \sum_{i=1}^{N} (p_i \log \frac{p_i}{m_i} + q_i \log \frac{q_i}{m_i}) \qquad (9)$$

where $m_i = \frac{p_i + q_i}{2}$.

In our tests, probabilistic measures of similarities such as Jeffrey divergence or χ^2-divergence (used by (Mellor et al, 2008)) yield better results than L^p-distances (e.g. Manhattan, $p = 1$, or Euclidean, $p = 2$). Using one-dimensional Earth mover's distance between histograms yields consequently poorer results, probably due to a larger variability in the relative weight of bins than in their positions.

We denote by $\mathcal{D}_k(u, v)$ the Jeffrey divergence between the kth histograms of the descriptors of u and v (in this paper $k \in \{1, \ldots 3\}$ if we use the descriptor EH+CtH and $k \in \{1, \ldots 4\}$ if we use SI+CtH). The fi-

nal distance between u and v can be computed as a weighted sum of the distances $\mathcal{D}_k(u, v)$,

$$\mathcal{D}(u, v) = \frac{\sum_{k=1}^{K} \omega_k \mathcal{D}_k(u, v)}{\sum_{k=1}^{K} \omega_k} \tag{10}$$

where ω_k is the weight assigned to the kth feature. For the sake of simplicity, in the following experiments the weights ω_k have been chosen as equal. These weights could have been adapted by learning their respective discriminative power on a training data set (see e.g. (Zhang et al, 2007)).

4.2 Comparative evaluations

4.2.1 Experimental protocols

As explained before, we reproduce exactly the retrieval and classification experiments described in the papers of Lazebnik *et al.* (Lazebnik et al, 2005), Mellor *et al.* (Mellor et al, 2008) and Xu *et al.* (Xu et al, 2009).

Recall that the approach of Lazebnik *et al.* relies on local descriptors. These descriptors are computed on a sparse set of affine invariant regions of interest. This kind of approach is popular in computer vision and known to be very efficient for object recognition. In the work of Lazebnik *et al.*, the best results are obtained with the combination of two region detectors (Harris and Laplacian) and two local descriptors (spin images and RIFT descriptors). The corresponding texture description, which is denoted by (H+L)(S+R), is locally invariant to affine transformations and locally robust to affine contrast changes. The approach of Mellor *et al.* relies on histograms of several invariant combinations of linear filters. This description is locally invariant to similarities and globally invariant to contrast changes. Finally, the method developed by Xu *et al.* is based on a multifractal description of textures. Their description is invariant under many viewpoint changes and non-rigid deformations, as well as local affine contrast changes.

In order to compare the performances of the descriptors we introduced with the best results provided by these papers, experiments are performed on three different databases: the Brodatz database, the UIUC database (Lazebnik et al, 2005) and UMD database (Xu et al, 2009). It is worth noticing that the corresponding results should be taken cautiously and not directly compared with other retrieval or classification experiments which do not follow exactly the same experimental protocols.

The **retrieval experiment** consists in using one sample of the database as a query and retrieving the

N_r most similar samples. The average number of correctly retrieved samples (generally called *recall*) when the query spans the whole database is drawn as a function of N_r.

For the **classification experiment**, N_t samples are extracted from each class and used as a training set. Each remaining sample in the database is then affected to the class which contains the nearest training sample. For each value N_t, an average classification rate is computed by using randomly selected training sets, in order to eliminate the dependence of the results on some particular sets.

4.2.2 Databases

The tree different databases used for the comparison tasks are now briefly described.

- **Brodatz Dataset**: The Brodatz's photo album (Brodatz, 1966) is a well known benchmark database used to evaluate texture recognition algorithms. Although it lacks some interclass variations, Lazebnik *et al.* (Lazebnik et al, 2005) point out that this database is a challenging platform for testing the discriminative power of texture descriptors, thanks to its *variety of scales and geometric patterns*. This database contains 111 different texture images. Following the protocols of (Lazebnik et al, 2005; Mellor et al, 2008), we divide each of these images into 9 non overlapping samples of resolution 215×215. As a result, the complete dataset is composed of 111 texture classes, each one being represented by 9 samples (all in all, 999 samples).
- **UIUC Database**: This texture database (Lazebnik et al, 2005) contains 25 texture classes, each one being composed of 40 samples of size 640×480 (*i.e.* 1000 samples altogether). Inside each class, the samples are subject to drastic viewpoint changes, contrast changes or even non-rigid deformations.
- **UMD Database**: This database, introduced by Xu *et al* (Xu et al, 2009) in order to test globally projective invariant features, is composed of 25 different textures classes, each one being represented by 40 samples (1000 samples altogether). These samples show strong viewpoint and scale changes, and significant contrast differences. A significant proportion of this database is made of textures consisting in the repetition of objects. The resolution of these images is 1280×960.

4.2.3 Performances on Brodatz

Figure 5 shows the retrieval and classification results obtained with the different indexing schemes on the Brodatz database.

In the retrieval experiment, shown on Figure 5 (a), the number of retrieved samples N_r takes values from 8 to 50. Since each class contains 9 samples, a perfect indexing method should reach an average recall of 100% for $N_r = 8$. For this number of retrieved samples, the affine invariant descriptor AI+CtH reaches 77.33%, while the similarity invariant descriptor SI+CtH reaches 80.44%. These results slightly outperform those of Lazebnik's affine invariant texture descriptor (H+L)(R+S) (76.97% recall) and Mellor's similarity invariant texture descriptors (77.65% recall). This trend remains valid when N_r increases. It should be remarked that in order to obtain such results on Brodatz, Lazebnik *et al.* add a shape channel to their description, and lose thereby their invariance to local affine changes.

Following (Lazebnik et al, 2005; Mellor et al, 2008), classification rates are estimated by averaging the results on randomly selected training sets. When the number of training samples is 3 for each class, the average classification rate reaches 88.31% for AI+CtH and 90.66% for SI+CtH. For the same level of invariance, these results are equivalent to those reported by Lazebnik *et al.* (88.15%) and Mellor *et al.* (89.71% for their similarity invariant descriptor) with the same protocol.

Now, as observed in (Mellor et al, 2008), some images of the original Brodatz database represent the same texture at different scales. Nevertheless, these images are considered as different textures by the experimental protocol, which penalizes invariant indexing schemes. In the same way, we should keep in mind that texture samples are created by cutting each texture of the Brodatz database into pieces. As a consequence, the resulting dataset lacks of viewpoint and scale changes. Consequently, a well chosen non-invariant indexing scheme should naturally provide better results on this database. In order to check this statement and for the sake of completeness, we tried to add some non-invariant features to our invariant descriptors. For this purpose, we added to the SI+CtH descriptor the histogram of shapes areas and the histogram of shapes orientations (the orientation being defined as the direction of the principal eigenvector of the inertia matrix (4)). The corresponding retrieval and classification results are shown in Figures 5 (a) and (b). Observe that, as it could be expected, all the results are clearly improved by adding these features.

4.2.4 Comparisons on UIUC Dataset

Figures 6 (a) and (b) show the retrieval and classification results of the AI+CtH and SI+CtH descriptors on the UIUC database. For the same level of invariance, these results are better than those reported in (Lazebnik et al, 2005) and (Mellor et al, 2008).

Let us observe that we were able to obtain better results than those reported in Figure 6 by weighting the contribution of each shape in the descriptors by a power of its area. This trick allows to give more weight to large shapes than to small ones, and hence to take more into account the geometrical aspect of textures. Now, using this trick on the Brodatz database yields a decrease of performances. Therefore, and since we did not find an automatic way to tune this weighting, we chose not to develop this possibility in the present study.

It is also interesting to note that local similarity invariance is enough to correctly retrieve texture classes with strong viewpoint variations. This property is illustrated by Figure 7, which shows the 39 first samples retrieved by SI+CtH when the query is the sample T15_01. This descriptor retrieves 38 samples of the class perfectly, despite the strong viewpoint changes between different samples. This is due both to the fact that three out of four features of SI+CtH are locally affine invariant, as well as to the fact that, as demonstrated by the experiments in Mellor *et al.*, invariance to local similarity already enables a good handling of viewpoints changes. In fact, local similarity invariance yields better results than local affinity invariance on this database, as will be further discussed in Section 4.4.2.

Another specific retrieval result is shown on Figure 8 for the texture class T25 of the UIUC database. This class, which represents a plaid under different viewpoints, contains many distortions and non-rigid deformations. Nevertheless, the SI+CtH descriptor retrieves the samples of this class quite well (the average retrieval rate on the whole class reaches 65.26% for 39 retrieved samples). It is also worth noting that 6 out of the 8 errors (highlighted in red on Figure 8) come from the same class T03. The retrieval of these samples is false but consistent. An example of a texture yielding a bad retrieval rate is shown in Figure 9. The corresponding texture class exhibits both blur and a very strong variability.

For classification of the UIUC database, the descriptors AI+CtH and SI+CtH also show better performances than the methods of Lazebnik *et al.* (Lazebnik et al, 2005) and Mellor *et al.* (Mellor et al, 2008). More precisely, the classification rate reached by AI+CtH is 66.56% and the one reached by SI+CtH is 70.69% when only one sample is used. These numbers should be compared
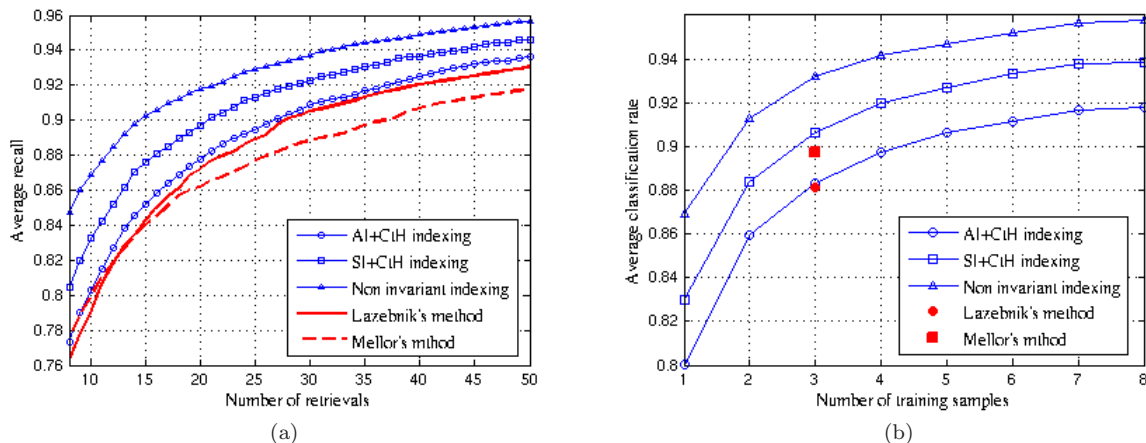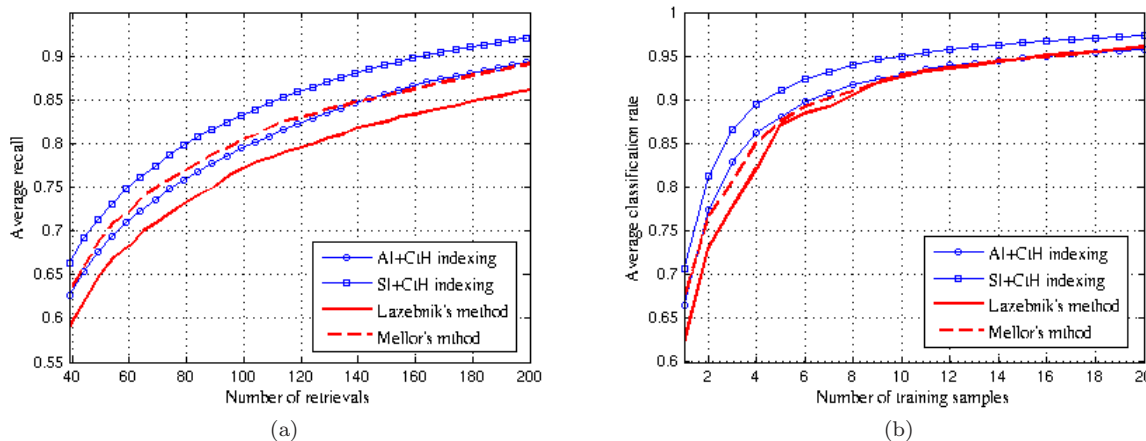
Fig. 5 Average retrieval (a) and classification (b) performances of different texture indexing schemes on the Brodatz dataset. The blue curves correspond to the performances of the descriptors SI+CtH and AI+CtH (recall that SI stands for similarity invariant local features, AI stands for affine invariant local features and CtH for locally affine invariant contrast histogram; all these features are described in Section 3.2), while the red curves show the performances of (Lazebnik et al, 2005) and (Mellor et al, 2008). The performance of a non-invariant indexing scheme is also shown for the sake of completeness.

Fig. 6 Average retrieval (a) and classification (b) performances of different texture indexing schemes on the UIUC database. The blue curves correspond to the performances of the descriptors SI+CtH and AI+CtH (recall that SI stands for similarity invariant local features, AI stands for affine invariant local features and CtH for locally affine invariant contrast histogram; all these features are described in Section 3.2), while the red curves show the performances of (Lazebnik et al, 2005) and (Mellor et al, 2008).

to the rates of 62.15% and 67.10% achieved respectively in (Lazebnik et al, 2005) and (Mellor et al, 2008). An interesting point is that the performances of our descriptors decrease on texture classes containing blur. The descriptors provided in the work of Lazebnik *et al.* (Lazebnik et al, 2005) appear to be more robust to blur and perform better on these specific classes. It is worth noticing that similar findings have been reported on region detectors by Mikolajczyk *et al.* (Mikolajczyk et al, 2005), who observed that MSER (Matas et al, 2002), a local region detector based on level sets, is more sensitive to blur than other region detectors, e.g. Harris-affine and Hessian-affine regions. This may be due to the use of the linear scale space in the process of key-

points extraction and scale computation for such descriptors.

4.2.5 Comparisons on UMD Database

Using the same strategy as before, Figure 10 shows the retrieval and classification performances of the descriptors AI+CtH and SI+CtH, along with the results obtained by the method of Xu (Xu et al, 2006), as well as those obtained on this database with the method of Lazebnik (Lazebnik et al, 2005) as reported in (Xu et al, 2006). Observe that our indexing scheme is particularly well adapted to this database. Indeed, the curves of Figure 10 show that both SI+CtH and AI+CtH descriptors
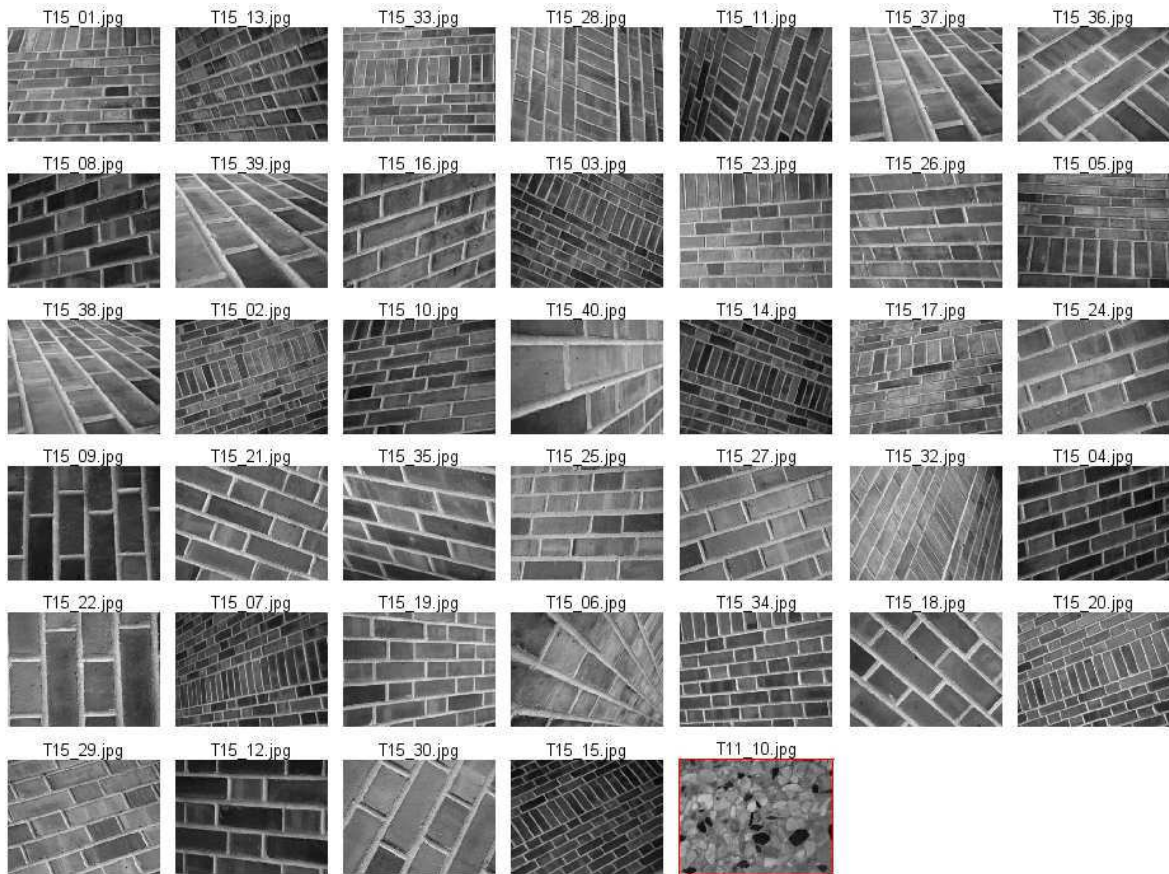
Fig. 7 One of the best retrieval results on the UIUC database, obtained on the texture class T15 using the SI+CtH descriptor (*Similarity Invariant features + Contrast Histogram*). The query image is in first position and the 39 most similar samples follow, ordered according to their matching scores. It is worth noticing that no learning is involved in these experiments. Retrieval results for all texture samples are available at the address (Xia, 2009).

perform significantly better than other methods. This may be due to the fact that this representation relies on geometry and is thereby well adapted to highly resolved and structured textures. Figure 11 shows two specific retrieval results, an almost perfect result on a texture made of apple stacks, as well as a result on a texture made of bamboos, for which the retrieval rate is roughly the one we get on the whole database. The AI+CtH and SI+CtH descriptors deal quite well with large scale and illumination changes on the fruit texture. Concerning the bamboos texture, one observes that textures RT21 and RT20 (corn leaves) are visually very similar and relatively hard to discriminate.

Two conclusions arise after the comparison of the descriptors proposed in this paper with the approaches of (Lazebnik et al, 2005; Mellor et al, 2008; Xu et al, 2006) on three different texture databases. First, both AI+CtH and SI+CtH are efficient for texture retrieval and classification. These descriptors show robust and consistent results on all three datasets, outperforming state of the art approaches. Second, similarity invariant

descriptors always perform better than affine invariant descriptors on all three databases. This aspect will be discussed in the last part of the section.

It is also worth noting that the texture features that we introduced are relatively compact in size. More precisely, each texture sample is represented by 4 histograms of 50 bins each, *i.e.* 200 values altogether. This size is comparable to that of Xu's descriptors (Xu et al, 2006), which use 78 values for each texture sample. In comparison, Lazebnik *et al.* (Lazebnik et al, 2005) use between 1200 and 4000 values for each sample (40 clusters of 32 or 100-dimensional descriptors), while Mellor *et al.* (Mellor et al, 2008) represent each sample by a histogram of 4096 bins.

4.3 Improving the performances through geodesic distances

In Section 4.2, we compared the texture indexing scheme proposed in this paper with three recent sets of invariant descriptors. We therefore followed a prescribed
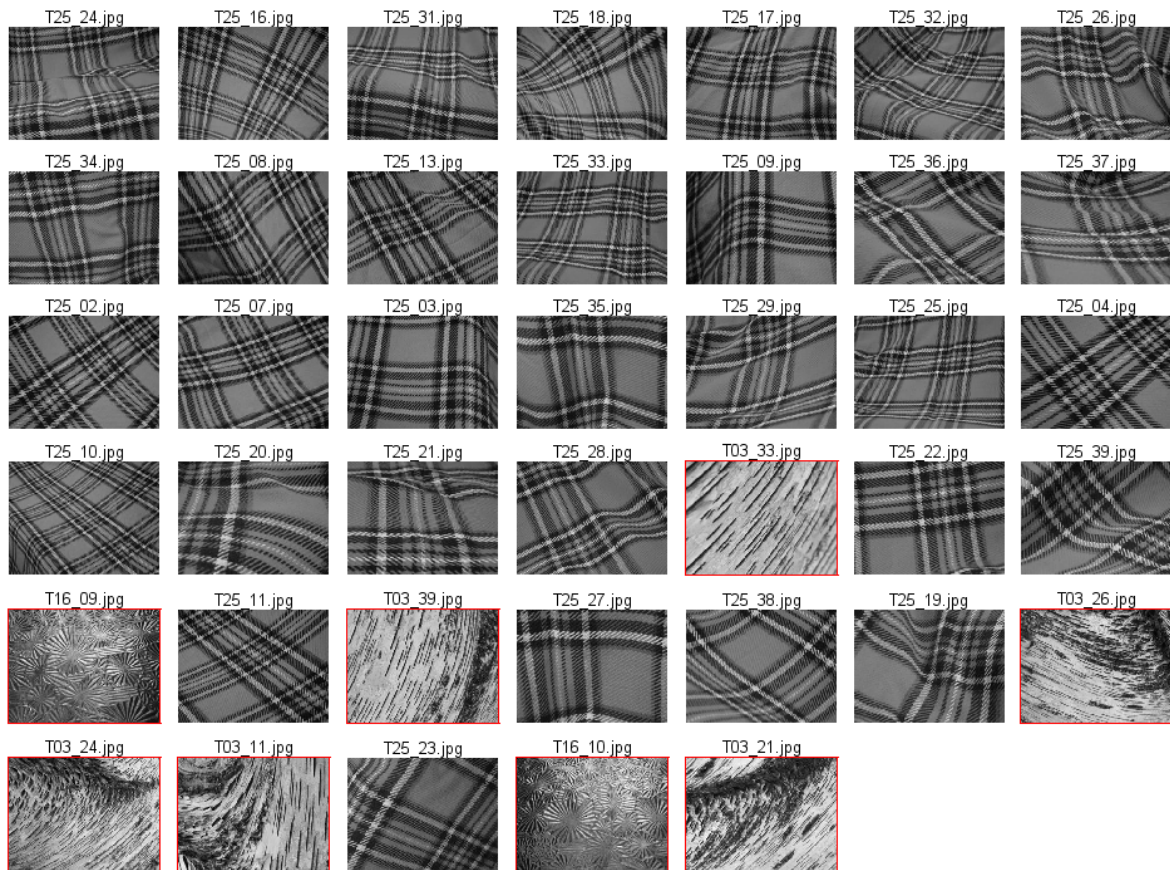
Fig. 8 Retrieval result obtained on the texture class T25 of the UIUC database with the descriptor SI+CtH (*Similarity Invariant features + Contrast Histogram*). The query image is in first position and the 39 most similar samples follow, ordered according to their matching scores. Retrieval errors are indicated in red. It is worth noticing that no learning is involved in these experiments. Retrieval results for all texture samples are available at the address (Xia, 2009).

protocol for classification relying on nearest neighbors. Usually, however, classification performances can benefit from some more powerful classifiers such as support vector machine (SVM) or from some classification schemes such as Adaboost. For example, Zhang *et al.* (Zhang et al, 2007) confirmed that using an SVM classifier instead of the nearest-neighbor classifier improves the classification performance of a bag-of-features representation.

In this section, we propose a very simple unsupervised way to improve recognition performances. Textures are considered as points lying on some intrinsic manifold obtained by the classical Isomap algorithm (Tenenbaum et al, 2000). Distances between these points are then computed as geodesics on the manifold. Since we are not concerned here with dimension reduction, we do not need to compute the manifold explicitly and compute distances between textures as follows. We first compute the pairwise distance matrix using Jeffrey divergence as in the previous sections. We then construct the k-nearest neighbor graph of the points (two vertices

are connected if one is among the k-nearest neighbors of the other). The geodesic distance between two points is then the shortest path between them on the graph. This shortest path may for instance be computed using Floyd algorithm (Floyd, 1962). Replacing Jeffrey divergence by this geodesic distance yields a very clear improvement of the retrieval and classification performances, especially for the retrieval task, as can be observed in Figure 12. In these experiments, a value of $k = 10$ has been used for the number of neighbors.

4.4 On invariance and discriminative power

4.4.1 Invariance to resolution changes

It was shown in section 3 that descriptors SI and AI are invariant to, respectively, local similarities and local affine transforms. In particular, the invariance to scale changes was ensured by the use of normalized moments computed on the topographic map, which do not change under a perfect, theoretical scale change. How-
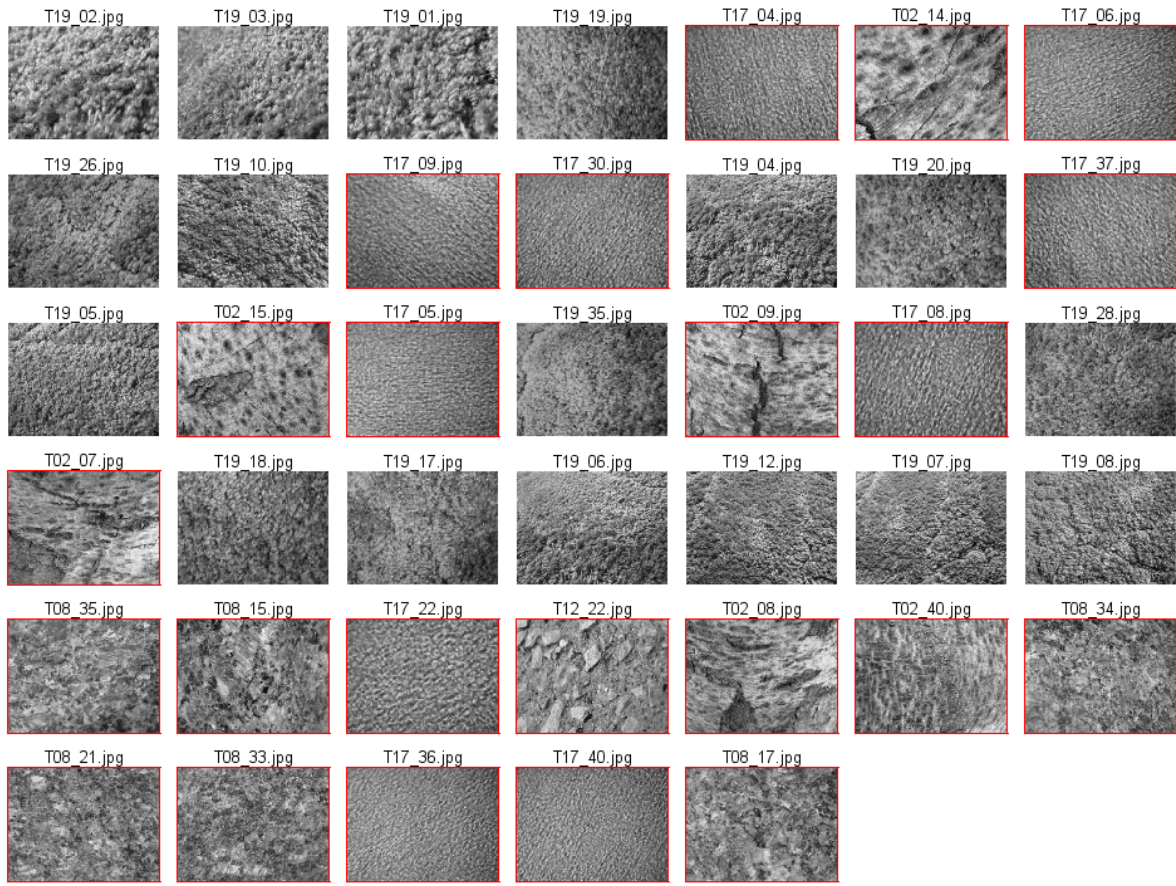
Fig. 9 A "bad" retrieval result obtained on the UIUC database with the descriptor SI+CtH (*Similarity Invariant features + Contrast Histogram*). The query image is in first position and the 39 most similar samples follow, ordered according to their matching scores. This result corresponds to the class T19. The corresponding texture class exhibits both blur and a very strong variability. Observe also that one half of the retrieval errors (indicated in red) are from the texture class T17, which at some scales looks similar to the class T19.


(a)

(b)
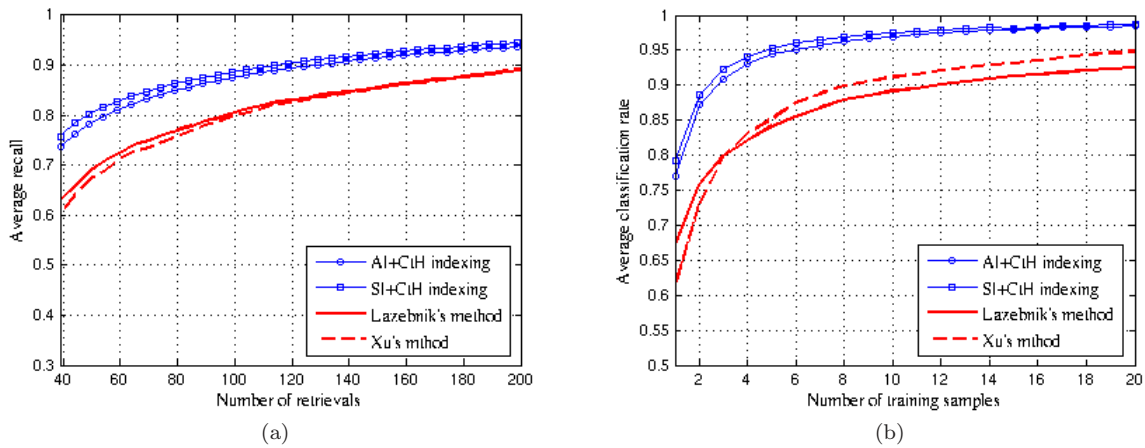
Fig. 10 Average retrieval (a) and classification (b) performances of different texture indexing schemes on UMD database (Xu et al, 2009). The blue curves correspond to the performances of the descriptors SI+CtH and AI+CtH (recall that SI stands for similarity invariant local features, AI stands for affine invariant local features and CtH for locally affine invariant contrast histogram; all these features are described in Section 3.2), while the red curves show the performances of (Xu et al, 2009) on this database, as well as those using the method from (Lazebnik et al, 2005) as reported by (Xu et al, 2009).
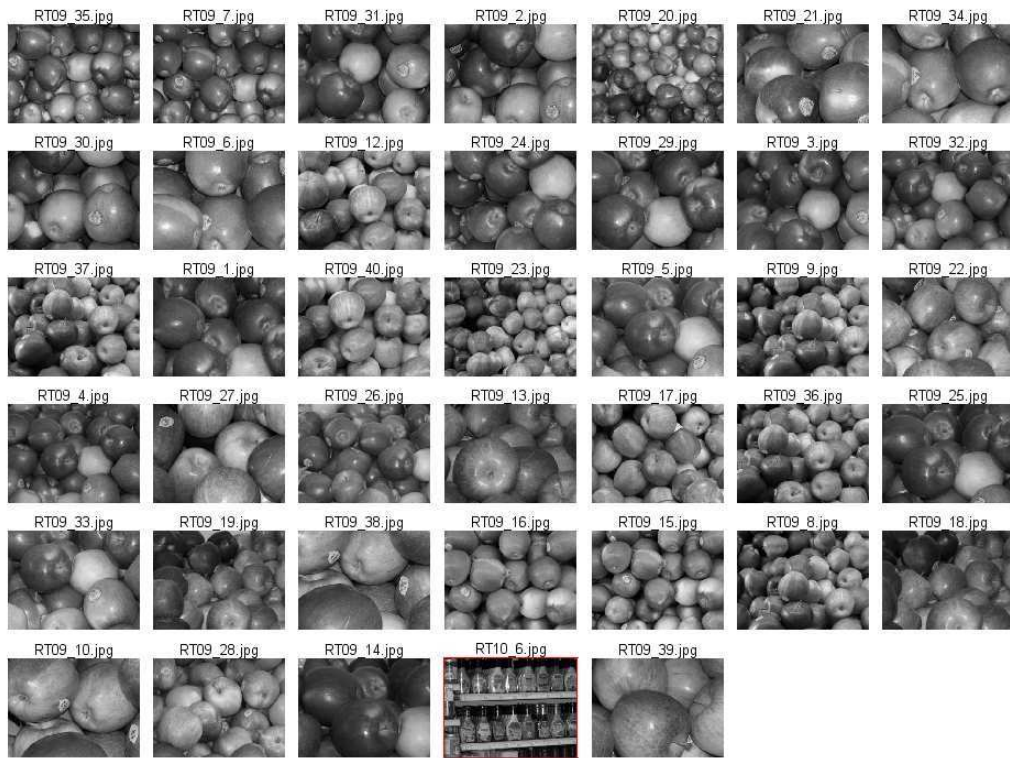
(a)
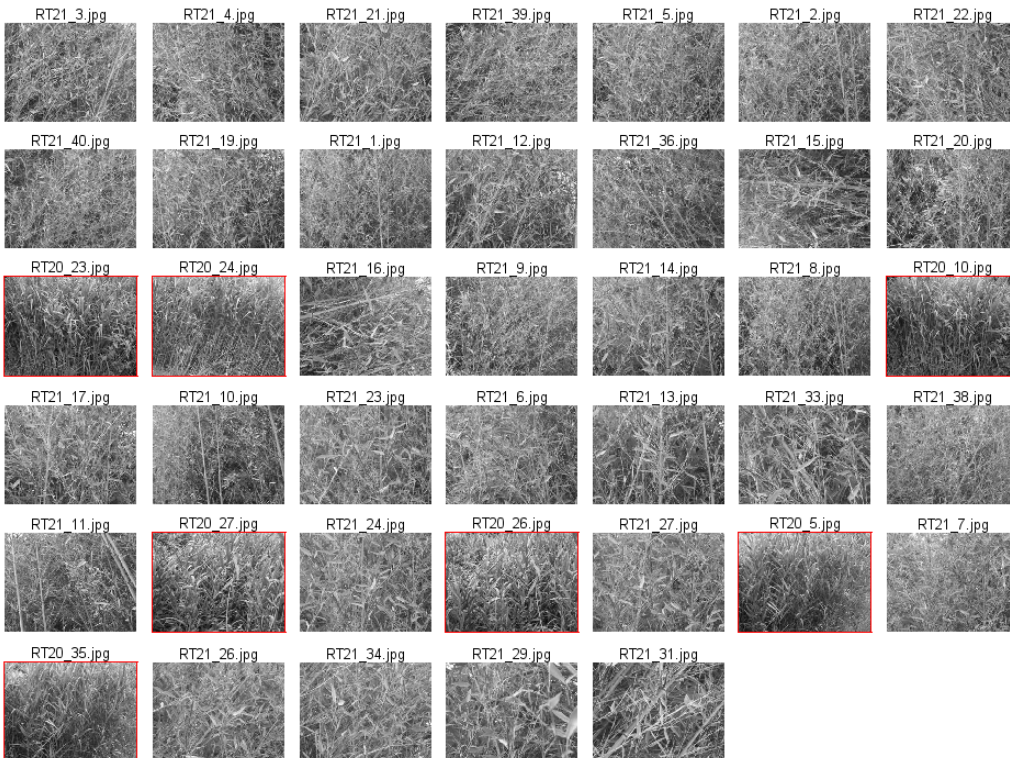
(b)

Fig. 11 Two retrieval results, respectively on (a) class RT9 and (b) class RT21 of UMD database, using the descriptor SI+CtH (*Similarity Invariant features + Contrast Histogram*). The query image is in first position and the 39 most similar samples are ordered according to their matching scores. Both examples correspond to non-planar textures. Observe that all errors for the class RT21 (bamboos) come from the class RT20 (corn leaves), which is visually quite similar to RT21. Again, no learning is involved in these experiments. Retrieval results for all texture samples are available at the address (Xia, 2009).
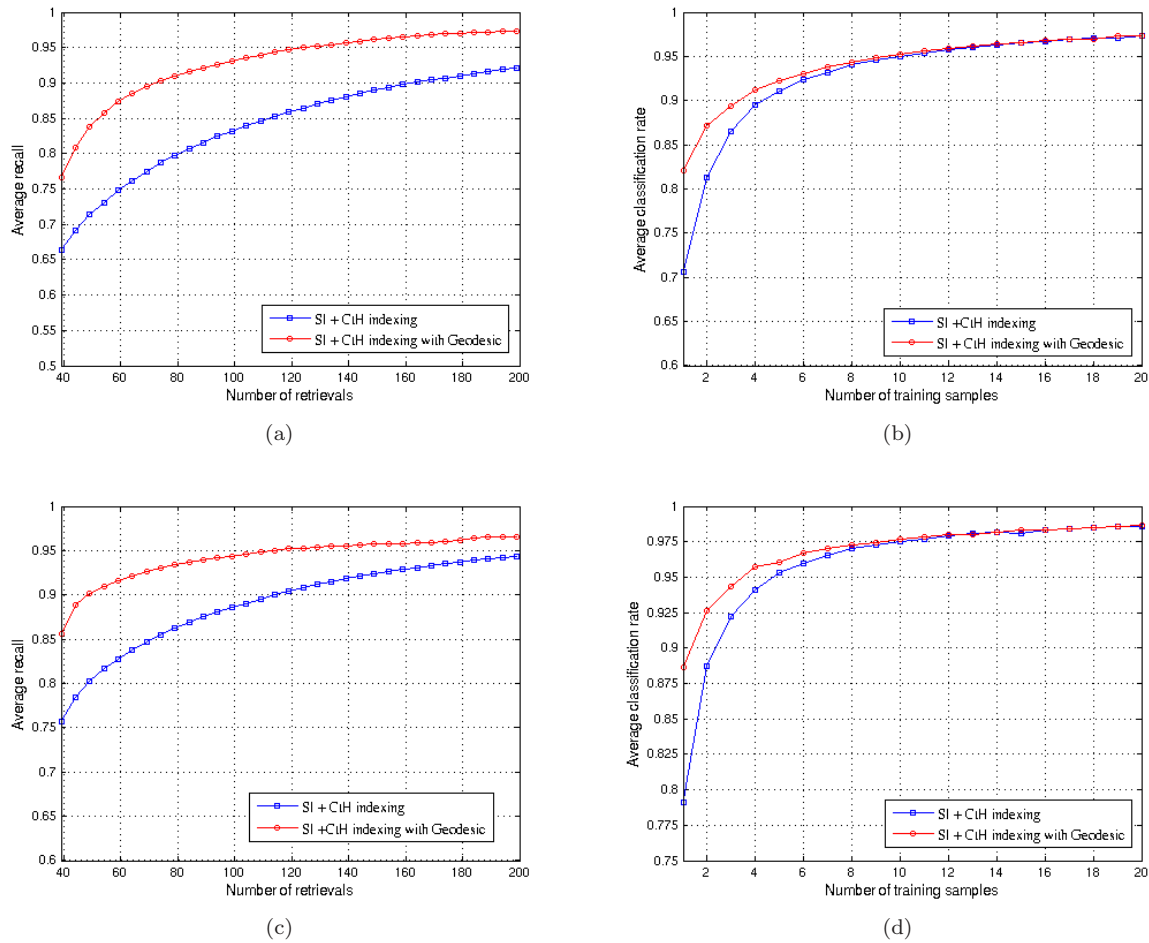
(a)

(b)

(c)

(d)

Fig. 12 Improving performances through geodesic distances Average retrieval (a) and classification (b) performances of the descriptor SI+CtH (*Similarity Invariant features + Contrast Histogram*) on UIUC database, with (red curve) and without (blue curve) geodesic distances. Figures (c) and (d) : same layout for the UMD database.

ever, in practice, scale changes on images often imply resolution changes. These changes can affect texture indexing methods, as investigated in (Luo et al, 2008). Such transformations involve blur, which affects the topographic map of images. In order to check the robustness of the descriptors to such changes, we set up the following experiment. Starting from 20 highly resolved texture images (see Figure 13), we build a database of 20 texture classes. In each class, the samples are generated by zooming each original texture image by a factor t, using bilinear interpolation. Here t takes its values among T as follows,

$$T = \{0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.3,$$
$$0.35, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9\}.$$

As a consequence, the whole database contains 20 classes of 16 samples, *i.e.* 320 texture samples. The size of the original images being 3072×2040, the smallest image size is 384×255.

Figure 14 shows the histograms SRH, CpH, EH and CtH of the 15-th texture shown in Figure 13 (pebble beach) for different zoom factors t. Observe that the curves coincide as long as the zoom factor remains larger than 0.5 (blue curves). When this factor decreases, the histograms move away from the original ones (for $t = 1$) but remain close to it. Similar behaviors are observed on other textures. This proves empirically the robustness of these features to real resolution changes with a zoom factor larger than .125.

In order to test the discriminative power of these features within the framework of resolution changes, we perform a simple retrieval experiment on this multiresolution database. For each zoom value t in T, and each texture class i, let M_t^i be the subset of the class made of the images having a resolution larger than t. A sample of resolution t and class i being given, its retrieval rate is defined as the proportion of well retrieved samples in M_t^i. As usual, the final retrieval rate $r(t)$ is the mean of
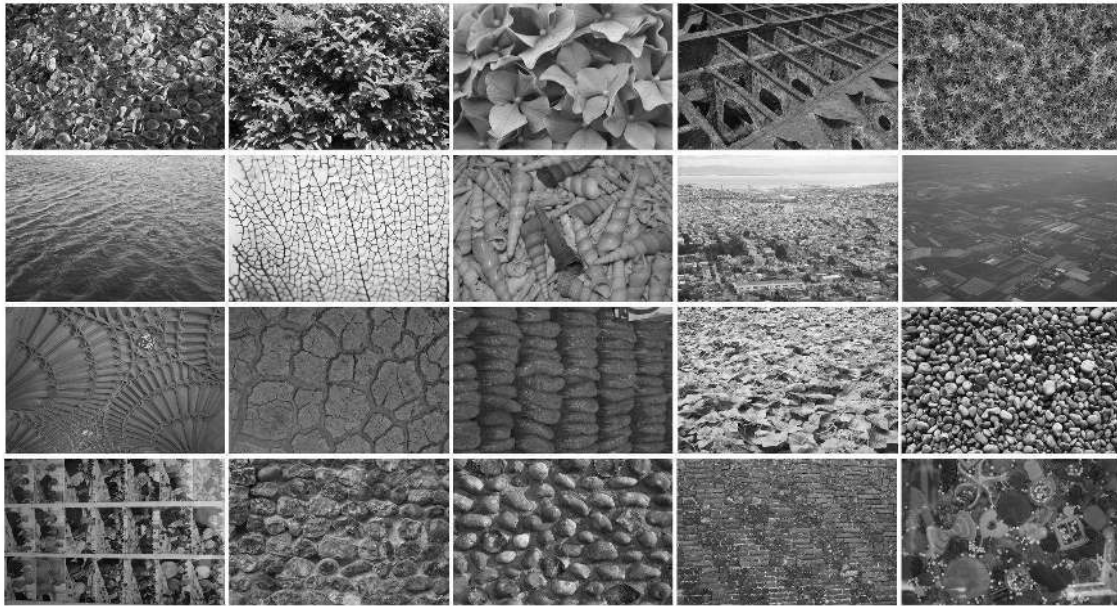
Fig. 13 Set of 3072×2040 texture images used to compute a multiresolution database. For each image, 15 samples are created by sub-sampling the original image with a zoom factor t taking its value in the set $T = \{0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

the retrieval rates over all samples of resolution t. Figure 15 shows the curves of $r(t)$ when t varies from 0.125 to 1 and when using different texture descriptors. Observe that up to a scale factor of 4, the retrieval results are perfect for SI+CtH.
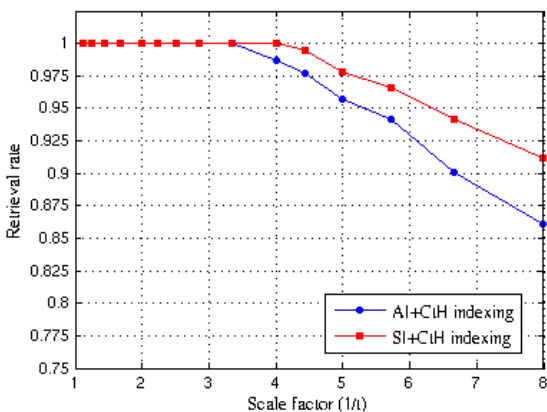

Fig. 15 Average retrieval performances of the descriptors SI+CtH and AI+CtH (recall that SI stands for similarity invariant local features, AI stands for affine invariant local features and CtH for locally affine invariant contrast histogram; all these features are described in Section 3.2) on the multiresolution database presented in section 4.4.1.

4.4.2 Local invariance vs discriminative power

Following the experiments of section 4.2, the question of the level of invariance required to index a particular database arises naturally. We saw on Brodatz that removing invariance to scale and orientation greatly improved the results, which seems to be coherent with the fact that this database does not present many geometric distortions. Of course, the best level of invariance depends on the database. On UIUC and UMD databases, all descriptors invariant to local similarity changes show significantly better results than locally affine invariant descriptors, which confirms the results presented in (Mellor et al, 2008) and (Zhang et al, 2007). Moreover, we observe that the advantage of similarity invariance on affine invariance remains true if we restrict ourselves to textures containing strong distortions. This can be surprising since these two databases contain classes with strong non-rigid deformations. We could theoretically expect that local affine invariance, or even local projective invariance would be needed to index such classes correctly (recall that UMD database, for instance, has been built on purpose to test projective invariant descriptors). The fact that features that are only invariant to local similarities show the best results despite these variations can only be explained by a better discriminative power. In other words, there is a natural trade-off between the level of invariance of a texture description and the discriminative power of this description.
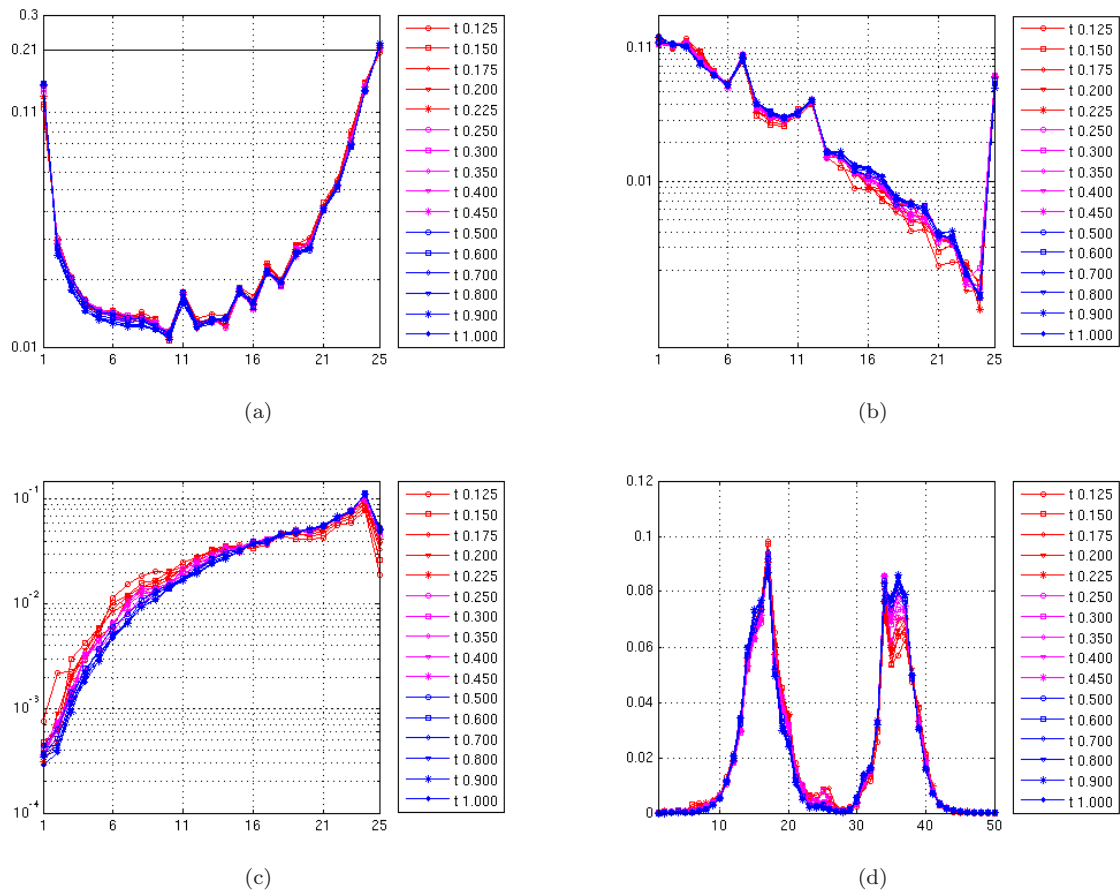
Fig. 14 Histograms of the pebble beach texture, the 15th texture image shown in Fig. 13: (a) Scale-ratio Histogram (SRH), (b) Elongation Histogram (EH), (c) Compactness Histogram (CpH) and (d) Contrast Histogram (CtH), for different zoom factors t.

Observe that the question of the best level of invariance needed for indexing is also addressed in (Zhang et al, 2007; Varma and Ray, 2007), where learning is used to estimate the optimal weights of the different descriptors.

These remarks also lead to question the need for further invariance in texture indexing. The previous observations suggest that achieving invariance to local similarities may be enough to account for viewpoint variations or non-rigid deformations. Furthermore, to the best of our knowledge, there exists no texture database in the literature on which complete local affine invariance is needed (in the sense that it yields better results than weaker invariances). Without such a database, it seems vain to try to develop features with more sophisticated invariances.

4.5 Scaling behavior of the analysis scheme

In this section, we briefly investigate how the proposed texture analysis scheme behaves when the numbers of texture classes and samples are increased. For this purpose, we simply build up a single database from the three texture databases considered so far (UIUC, UMD and Brodatz) therefore reaching 161 classes and 2999 samples.

We repeat the retrieval and classification experiments described in Section 4.2.1 for each sample of the whole Brodatz+UIUC+UMD database. Observe that in this configuration, classification and retrieval become noticeably more difficult : for example, for each query of the UIUC database, only 39 samples among 2999 belongs to the same class, instead of 39 among 1000 in the experiments of Section 4.2.1. Figure 16 shows the retrieval and classification rates averaged over all Brodatz samples, while Figure 17 shows the retrieval and classification rates averaged respectively over all UIUC samples, or over all UMD samples. Observe that although the performances decrease when the numbers of samples and classes increase, the proposed texture analysis scheme scales very well. In particular, it performs better on this large combined database of 2999 samples than Lazebnik's descriptors on the single UIUC

database (1000 samples), and its performances are comparable to those of Xu's descriptors on the single UMD database (1000 samples).

4.6 A segmentation experiment

Considering that the texture features introduced in Section 3 are efficient for classification tasks, it is natural to investigate their ability to segment texture images. Observe that the topographic map has a scale-space structure in which no regularization of the geometry is involved. This property makes it particularly interesting in the context of image segmentation. In this paragraph, we present a simple segmentation experiment, in which each pixel x is described by the features of $s(x)$, the smallest shape of the topographic map containing x. Five features are used: the contrast information of $s(x)$, defined in Section 3.3, the scale ratio of $s(x)$ defined in Section 3.2, the orientation of $s(x)$, as well as its elongation and compactness, both defined in Section 3.1. A recent active contour model (Houhou et al, 2008) is then applied to the resulting vectorial image. Two examples of the resulting segmentation scheme are displayed in Figure 18. The first one is composed of two different textures, which have been radiometrically corrected in order to share the same global mean and standard deviation. The second one is a photograph of a corn field. Because of the perspective, the corn texture is present at different scales. The segmentation results obtained are shown on the right column of the figure. Although this approach show promising results, it is important to notice that those results highly depend on the parameters of the method (mostly the regularization parameter λ in the energy and the initial contour), as it is usual with active contour models. These results could certainly benefit from recent developments in global minimization for active contour models such as those of (Bresson et al, 2007).

5 Conclusion

In this paper, it is shown that the topographic map is an efficient and intuitive tool to analyze texture images. Geometrical features are computed from the level sets of images, enabling state-of-the-art retrieval and classification results on challenging databases. In particular, this shows that morphological, granulometry-like indexing methods can deal with complex, potentially highly resolved texture images, even in the case of nonrigid transforms. To the best of our knowledge, such invariant analysis were only reported in the literature
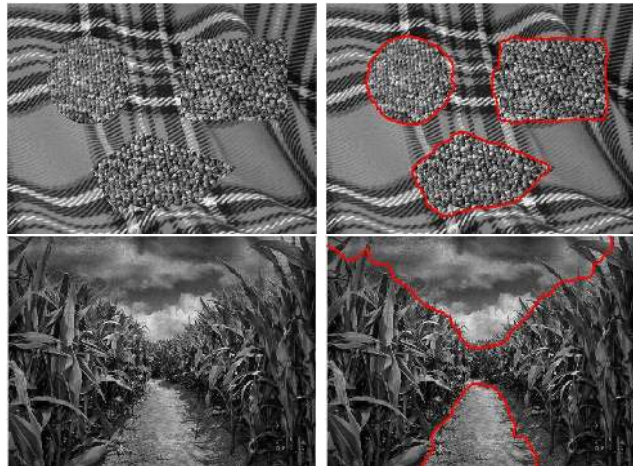

Fig. 18 Top: Image composed of two textures (taken from the UIUC database), radiometrically normalized to share the same mean and standard derivation. On the right, segmentation result using the active contour model described in (Houhou et al, 2008) and different features described in this paper. **Bottom**: same experiment with a photograph of a corn field.

using wavelet-like features, local descriptors or pixel-based features.

This work opens several perspectives. First, the hierarchical structure of the topographic map is only partially accounted for in the present work. It is of interest to further investigate the descriptive power of statistics on the tree of level lines, making use of specific neighborhoods and higher dependencies in the tree, possibly using probabilistic graphical models. One difficulty is to achieve this while preserving radiometric and geometric invariances. Next, and going beyond local contrast invariances, one could study the behavior of level line statistics under illumination changes in greater details. We show in this paper that lines statistics yield efficient retrieval results on databases with varying illumination conditions. The next step could be either to explicitly model level lines variations or to investigate the ability of the topographic map to learn the effects of illumination changes using databases such as CUReT (Dana et al, 1999). Other possible applications of the proposed framework include the registration of non-rigid objects, shape from texture or material recognition. Another possible extension is the design of locally invariant morphological filters, that could be designed by pruning the topographic map depending on features values.
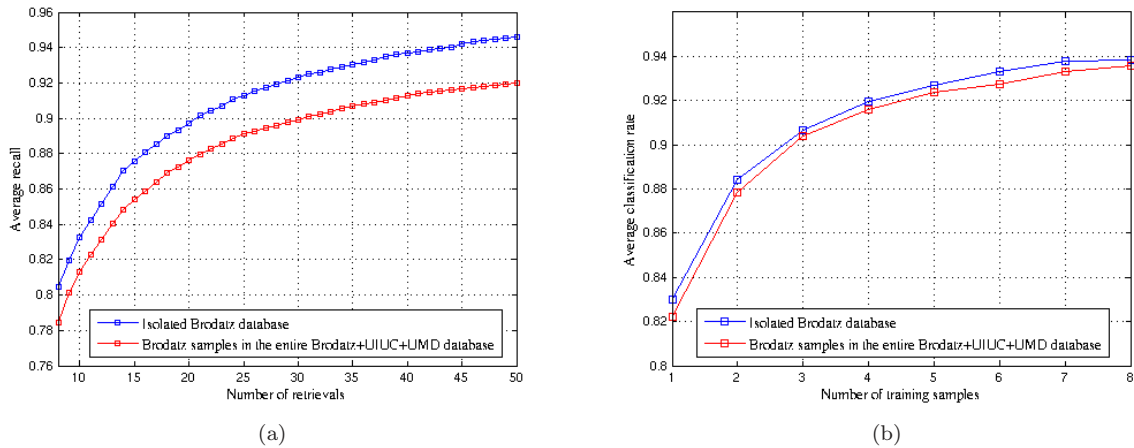
6 Acknowledgement

Fig. 16 In red, average retrieval (a) and classification (b) performances of Brodatz samples in the entire Brodatz+UIUC+UMD database, using SI+CtH (recall that SI stands for similarity invariant local features and CtH for locally affine invariant contrast histogram; all these features are described in Section 3.2). In blue, for comparison, average retrieval (a) and classification (b) performances of the same samples among the isolated Brodatz database.
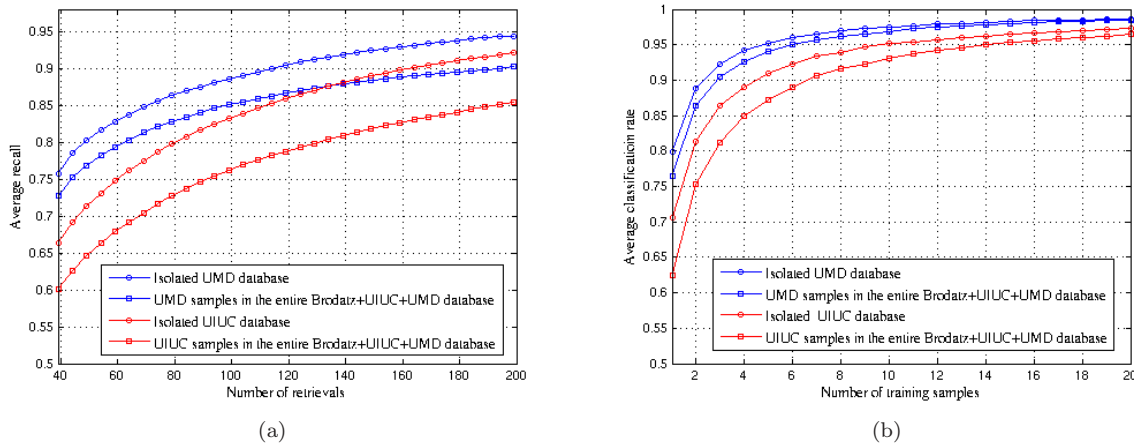

Fig. 17 In red, average retrieval (a) and classification (b) performances of UIUC (resp. UMD) samples in the entire Brodatz+UIUC+UMD database, using SI+CtH (recall that SI stands for similarity invariant local features and CtH for locally affine invariant contrast histogram; all these features are described in Section 3.2). In blue, for comparison, average retrieval (a) and classification (b) performances of the same samples among the isolated UIUC (resp. UMD) database.

References

Asano A, Miyagawa M, Fujio M (2000) Texture modelling by optimal gray scale structuring elements using morphological pattern spectrum. In: Proc. International Conference on Pattern Recognition, pp III: 475–478

Ayala G, Domingo J (2001) Spatial size distributions: Applications to shape and texture analysis. IEEE Trans Pattern Analysis and Machine Intelligence 23(12):1430–1442

Bresson X, Esedoglu S, Vandergheynst P, Thiran JP, Osher S (2007) Fast global minimization of the active contour/snake model. Journal of Mathematical Imaging and Vision 28(2):151–167

Brodatz P (1966) Textures: A Photographic Album for Artists and Designers. Dover, New York

Caselles V, Monasse P (2008) Geometric Description of Topographic Maps and Applications to Image Processing. Lecture Notes in Mathematics, Springer, to Appear

Caselles V, Coll B, Morel J (1997) Scale space versus topographic map for natural images. In: Proc. Inter-

national Conference on Scale-Space Theory in Computer Vision, vol 11, pp 29–49

Caselles V, Coll B, Morel JM (1999a) Topographic maps and local contrast changes in natural images. International Journal of Computer Vision 33(1):5–27

Caselles V, Lisani JL, Morel JM, Sapiro G (1999b) Shape preserving local histogram modification. IEEE Trans Image Processing 8(2):220–230

Chen CH (1982) A study of texture classification using spectral features. In: Proc. International Conference on Pattern Recognition

Chen J, Kundu A (1994) Rotation and gray scale transform invariant texture identification using wavelet decomposition and hidden markov model. IEEE Trans Pattern Analysis and Machine Intelligence 16(2):208–214

Chen Y, Dougherty E (1994) Gray-scale morphological granulometric texture classification. Optical Eng 33(8):2713–2722

Cohen FS, Fan Z, Patel MA (1991) Classification of rotated and scaled textured images using gaussian markov random field models. IEEE Trans Pattern Analysis and Machine Intelligence 13(2):192–202

Dana K, Van-Ginneken B, Nayar S, Koenderink J (1999) Reflectance and Texture of Real World Surfaces. ACM Transactions on Graphics (TOG) 18(1):1–34

Davis L (1981) Polarograms: a new tool for image texture analysis. Pattern Recognition 13(3):219–223

Davis L, Johns S, Aggarwal J (1979) Texture analysis using generalized co-occurrence matrices. IEEE Trans Pattern Analysis and Machine Intelligence 1(3):251–259

Do MN, Vetterli M (2002) Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models. IEEE Trans Multimedia 4(4):517–527

Efros AA, Leung TK (1999) Texture synthesis by non-parametric sampling. In: Proc. IEEE International Conference on Computer Vision, Corfu, Greece, pp 1033–1038

Fletcher N, Evans A (2005) Texture segmentation using area morphology local granulometries. In: Proc. International Symposium on Mathematical Morphology, vol 30, pp 367–376

Floyd RW (1962) Algorithm 97: Shortest path. Commun ACM 5(6):345, DOI http://doi.acm.org/10.1145/367766.368168

Flusser J, Suk T (1993) Pattern recognition by affine moment invariants. Pattern Recognition 26:167–174

Gidas B (1989) A renormalization group approach to image processing problems. IEEE Trans Pattern Analysis and Machine Intelligence 11(2):164–180

Gousseau Y (2002) Texture synthesis through level sets. In: Proc. International Workshop on Texture Analysis and Synthesis, Copenhagen, pp 53–57

Haas A, Matheron G, Serra J (1967) Morphologie mathématique et granulométries en place. Annales des Mines 11:736–753

Hamdan HM, Larson LM (2002) Texture classification through level lines. In: Proc. International Conference on Image Processing, pp 937–940

Hanbury A, Kandaswamy U, Adjeroh DA (2005) Illumination-invariant morphological texture classification. In: Proc. International Symposium on Mathematical Morphology, vol 30, pp 377–386

Haralick RM (1979) Statistical and structural approaches to texture. Proceedings of the IEEE 67(5):786–804

Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Trans Systems, Man and Cybernetics SMC-3(6):610–621

Heijmans HJAM (1999) Connected morphological operators for binary images. Computer Vision and Image Understanding 73(99-120)

Houhou N, Thiran J, Bresson X (2008) Fast texture segmentation model based on the shape operator and active contour. In: Proc. Computer Vision and Pattern Recognition

Hu MK (1962) Visual pattern recognition by moment invariants. IRE Trans Information Theory 8:179–187

Jain AK, Farrokhnia F (1991) Unsupervised texture segmentation using gabor filters. Pattern Recognition 24(12):1167–1186

Kaizer H (1955) A quantification of textures on aerial photographs. Tech. Rep. 121, Boston University

Kashyap RL, Khotanzad A (1986) A model-based method for rotation invariant texture classification. IEEE Trans Pattern Analysis and Machine Intelligence 8(4):472–481

Lazebnik S, Schmid C, Ponce J (2003) Affine-invariant local descriptors and neighborhood statistics for texture recognition. In: Proc. International Conference on Computer Vision, pp 649–655

Lazebnik S, Schmid C, Ponce J (2005) A sparse texture representation using local affine regions. IEEE Trans Pattern Analysis and Machine Intelligence 27(8):1265–1278

Leung T, Malik J (2001) Representing and recognizing the visual appearance of materials using three-dimensional textons. International Journal of Computer Vision 43(1):29–44

Li FF, Perona P (2005) A bayesian hierarchical model for learning natural scene categories. In: Proc. Computer Vision and Pattern Recognition, pp II: 524–531

Li W, Haese Coat V, Ronsin J (1997) Residues of morphological filtering by reconstruction for texture classification. Pattern Recognition 30(7):1081–1093

Liao S, Pawlak M (1996) On image-analysis by moments. IEEE Trans Pattern Analysis and Machine Intelligence 18(3):254–266

Luo B, Aujol JF, Gousseau Y, Ladjal S (2008) Indexing of satellite images with different resolutions by wavelet features. IEEE Trans on Image Processing 17(8):1465–1472

Luo B, Aujol JF, Gousseau Y (2009) Local scale measure from the topographic map and application to remote sensing images. SIAM Multiscale Modeling and Simulation 8(1):1–29

Maragos P (1989) Pattern spectrum and multiscale shape representation. IEEE Trans Pattern Analysis and Machine Intelligence 11(7):701–716

Matas J, Chum O, Martin U, Pajdla T (2002) Robust wide baseline stereo from maximally stable extremal regions. In: Proc. British Machine Vision Conference, London, vol 1, pp 384–393

Mellor M, Hong BW, Brady M (2008) Locally rotation, contrast, and scale invariant descriptors for texture analysis. IEEE Trans Pattern Analysis and Machine Intelligence 30(1):52–61

Meyer Y (2007) Workshop : An interdisciplinary approach to Textures and Natural Images Processing, Institut Henri Poincaré, Paris

Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T, Van Gool L (2005) A comparison of affine region detectors. International Journal of Computer Vision 65(1-2):43–72

Monasse P (1999) Contrast invariant registration of images. In: Proc. International Conference on Acoustics, Speech and Signal Processing, pp 3221–3224

Monasse P (2000) Morphological representation of digital images and application to registration. PhD thesis, Paris-Dauphine University

Monasse P, Guichard F (2000a) Fast computation of a contrast invariant image representation. IEEE Trans Image Processing 9(5):860–872

Monasse P, Guichard F (2000b) Scale-space from a level lines tree. Journal of Visual Communication and Image Representation 11(2):224–236

Obdrzálek S, Matas J (2002) Object recognition using local affine frames on distinguished regions. In: Proc. British Machine Vision Conference, pp 113–122

Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Analysis and Machine Intelligence 24(7):971–987

Peleg S, Naor J, Hartley R, Avnir D (1984) Multiple resolution texture analysis and classification.

IEEE Trans Pattern Analysis and Machine Intelligence 6(4):518–523

Peyré G (to appear, 2009) Texture synthesis with grouplets. IEEE Trans Pattern Analysis and Machine Intelligence

Pietikäinen M, Ojala T, Xu Z (2000) Rotation-invariant texture classification using feature distributions. Pattern Recognition 33(1):43–52

Pun CM, Lee MC (2003) Log-polar wavelet energy signatures for rotation and scale invariant texture classification. IEEE Trans Pattern Anal Mach Intell 25(5):590–603

Randen T, Husoy JH (1999) Filtering for texture classification: A comparative study. IEEE Trans Pattern Analysis and Machine Intelligence 21(4):291–310

Reed TR, du Buf JMH (1993) A review of recent texture segmentation and feature extraction techniques. Computer Vision, Graphics and Image Processing: Image Understanding 57(3):359–372

Salembier P, Serra J (1995) Flat zones filtering, connected operators, and filters by reconstruction. IEEE Trans Image Processing 4(8):1153–1160

Sandler R, Lindenbaum M (2009) Optimizing gabor filter design for texture edge detection and classification. International Journal of Computer Vision 84(3):308–324

Schaffalitzky F, Zisserman A (2001) Viewpoint invariant texture matching and wide baseline stereo. In: Proc. International Conference on Computer Vision, vol 2, pp 636–643

Serra J (1982) Image Analysis and Mathematical Morphology. Academic Press

Simoncelli EP, Portilla J (1998) Texture characterization via joint statistics of wavelet coefficient magnitudes. In: Proc. International Conference on Image Processing, pp 4–7

Song Y (2007) A topdown algorithm for computation of level line trees. IEEE Trans Image Processing 16(8):2107–2116

Tenenbaum JB, Silva V, Langford JC (2000) A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290(5500):2319–2323

Tuceryan M, Jain AK (1993) Texture analysis. Handbook of pattern recognition and computer vision pp 235–276

Urbach ER, Roerdink JBTM, Wilkinson MHF (2007) Connected shape-size pattern spectra for rotation and scale-invariant classification of gray-scale images. IEEE Trans Pattern Analysis and Machine Intelligence 29(2):272–285

Varma M, Garg R (2007) Locally invariant fractal features for statistical texture classification. In: Proc. International Conference on Computer Vision, pp 1–

8

Varma M, Ray D (2007) Learning the discriminative power-invariance trade-off. In: Proc. International Conference on Computer Vision, pp 14–21

Varma M, Zisserman A (2002) Classifying images of materials: achieving viewpoint and illumination independence. In: Proc. European Conference on Computer Vision, pp III: 255–271

Wu J, Chantler M (2003) Combining gradient and albedo data for rotation invariant classification of 3d surface texture. In: Proc. International Conference on Computer Vision, pp 848–855

Xia GS (2009) Invariant texture analysis web page. URL http://www.tsi.enst.fr/~xia/texture.html

Xia GS, Delon J, Gousseau Y (2008) Locally invariant texture analysis from the topographic map. In: Proc. International Conference on Pattern Recognition

Xu Y, Ji H, Fermuller C (2006) A projective invariant for textures. In: Proc. Computer Vision and Pattern Recognition, pp 1932–1939

Xu Y, Ji H, Fermuller C (2009) Viewpoint invariant texture description using fractal analysis. International Journal of Computer Vision 83(1):85–100

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

Zhang J, Tan T (2002) Brief review of invariant texture analysis methods. Pattern Recognition 35(3):735–747

Zhang J, Marszalek M, Lazebnik S, Schmid C (2007) Local features and kernels for classification of texture and object categories: A comprehensive study. International Journal of Computer Vision 73(2):213–238