Matching Pursuit-Based Shape Representation and Recognition Using Scale-Space

François Mendels, Pierre Vandergheynst, Jean-Philippe Thiran

Ecole Polytechnique Federale de Lausanne, Institut de Traitement des Signaux ITS, STI,
1015 Lausanne, Switzerland

ABSTRACT: In this paper, we propose an analytical low-level representation of images, obtained by a decomposition process, namely the matching pursuit (MP) algorithm, as a new way of describing objects through a general continuous description using an affine invariant dictionary of basis function (BFs). This description is used to recognize multiple objects in images. In the learning phase, a template object is decomposed, and the extracted subset of BFs, called meta-atom, gives the description of the object. This description is then naturally extended into the linear scale-space using the definition of our BFs, and thus providing a more general representation of the object. We use this enhanced description as a predefined dictionary of the object to conduct an MP-based shape recognition task into the linear scale-space. The introduction of the scale-space approach improves the robustness of our method: we avoid local minima issues encountered when minimizing a nonconvex energy function. We show results for the detection of complex synthetic shapes, as well as real world (aerial and medical) images.   © 2007 Wiley Periodicals, Inc. Int J Imaging Syst Technol, 16, 162–180, 2006; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/ima.20078

Key words: matching pursuit; scale-space shape representation; shape recognition; sparse representation

I. INTRODUCTION

Shape modeling and recognition is one of the basic operations in image analysis, and represents a great deal of interest. These methods are used in many applications fields, such as face recognition, retrieval of images in databases (content-based image retrieval (CBIR)), satellite image analysis, or medical imaging.

Relevant regions, extracted from a previous segmentation process, have to be coded/represented in some way, so that a further analysis of their shape is possible. Numerous methods have been presented, but all of them are linked by the same need: the shape has to be coded in some way to be described efficiently. Styner summarizes it as "A shape description is said to be efficient if shapes are described by concise sets of parameters or features" (Styner and Gerig, 2000). Moreover, this description has to be unique: we want to get the same descriptor from the same object as a basis to perform a correct shape recognition task. These descriptions have to be in both cases invariant to translation, rotation, and scale, as these transformations do not change their intrinsic shape characterization. The shape recognition process is then realized, based on their description, usually in a one-to-one correspondence framework.

Many authors have presented different ways to produce a so-called shape descriptor or shape representation, depending on whether they give quantitative or qualitative measures (Pavlidis, 1980; Loncaric, 1998). The reader is referred to the works by Pavlidis (1980), Mehtre et al. (1997), Loncaric (1998), Veltkamp and Hagedoorn (1999), Kindratenko (2003), Safar and Shahabi (2003), and Zhang and Lu (2004) for further references on classical methods.

Different classifications of these methods were proposed according to general description criteria. Among all, Pavlidis (1980) has proposed three different classifications for the shape analysis methods (Pavlidis, 1980): the first one distinguishes between boundary-based (local) and region-based (global) methods. The second one takes into account the result of the analysis as numerical or not, bringing respectively scalar transform or spatial transform domain techniques. The third one relies on the information preserving property or not, of each methods. In this work, we are interested in methods related to spatial sparse representation models, with both local and global features.

Many of these shape description techniques can be seen as decomposition methods into fundamental elements/parts: many terms have been used to name them, such as shape primitives (Shapiro, 1980), atoms (Pavlidis, 1980), geons (Biederman, 1985; Ullman, 1989), codons (Hoffman and Richards, 1983), etc., depending of the features considered in each case. We use here *shape primitives* as a general naming convention for all these terms. This expression is normally dedicated to a structural model of shape (see, e.g., Shapiro, 1980). Shapiro (1980) defined them as follows:

> The primitives are the simple parts and intrusions of the shape which can be derived through the graph-theoretic clustering procedure.

Nevertheless, this generic term of *shape primitives* can be interpreted in other ways: let us look at the *chain code* method (Freeman, 1974), consisting in assigning a direction to each

segment composing a digitized contour, or its version with critical points (e.g. Freeman, 1978). We can extend the shape primitives concept and consider in this case the direction vectors adjoint to the shape variation in orientation as the shape primitives.

Of course, shape decomposition techniques, such as Fourier descriptors (Arbter et al., 1990; Kauppinen et al., 1995), iconic representation (Rao and Ballard, 1995a,b; Ben-Arie et al., 1996a,b), medial shape descriptors (Symmetric Axis Transform (SAT) (Blum, 1967; Blum and Nagel, 1978), and Spherical Harmonics functions (SPHARM (Styner and Gerig, 2000; Brechbühler et al., 1995; Gerig et al., 2001)), lead naturally to this vision of shape primitives, as each of them uses simple elements to describe a complex shape: this group of elements can be interpreted as the specific basis of the shape. Pavlidis (1980) made the following statement for global shape analysis:

> Among the boundary points find sets of points which are closely related. Such sets may be used to assign labels to corresponding parts of the object.

In Ullman (1989), Ullman joins Pavlidis's view in linking object description to the recognition process: he describes the decomposition approach in constituent parts as a natural basis for object recognition, where all objects are defined by a combination of these components. He pointed out basic components such as line, points, and corners as low-level representation opposed to more complex components, such as group of basic ones linked with a spatial relationship, as high-level representation.

The full shape is then described by the set of constituent fragments, overlapping each other or not: Ullman and Soloviev (1999) present an example of a face summarization with a description based on a set of overlapping parts of a face fragments training set, containing structures as eyes, mouth, border of the face, hairs, etc., taken at different level of resolution.

All these descriptors are linked by the same need: getting a concise description of a shape (Styner and Gerig, 2000), that may be useful for shape recognition purposes. Depending on the targeted recognition process, the shape descriptor benefits from invariance properties to transformations, such as rotation, translation, or scaling.

The image-coding domain has the same needs: representing a signal in a compact way, in order to get a sparse representation, used, e.g., to diminish transmission costs. Low-level representations are commonly used for this purpose to have a concise and unique, if possible, description of a signal. Mallat and Zhang (1993) define their role as

> Low level representations must also provide explicit information on very different properties, while giving simple cues to differentiate close patterns.

These representations reflects the signal structures to be analyzed, by isolating them when they are coherent with the reference basis. Linear expansions of waveforms selected from a redundant dictionary have been recently used to represent the signal, as they allow to model adaptive signal representations (Mallat and Zhang, 1993). Such expansions lead to sparse codes (Pece, 2002). Iterative greedy strategies were designed to obtain such descriptions by decomposing signals into an overcomplete set of basis functions (BFs). Many methods, such as the Method of Frames (MOF) (Daubechies, 1988), the Best Orthogonal Basis (BOB) (Coifman and Wickerhauser, 1992), the Matching Pursuit (MP) (Mallat and

Zhang, 1993; Bergeaud and Mallat, 1995), as some of its variations, the Orthogonal Matching Pursuit (OMP) (Pati et al., 1993; Gharavi-Alkhansari and Huang, 1998; Rebollo-Neira and Lowe, 2002), the Basis Pursuit (BP) (Chen et al., 1998), the Stochastic Matching Pursuit (Wang and Goblirsch, 1997), the Statistical Matching Pursuit (Wang et al., 1997), the weighted L^p Matching Pursuit (Donahue et al., 1996), the Multiresolution Matching Pursuit (Figueras i Ventura and Vandergheynst, 2002), the Inhibition method (Pece and Petkov, 2000), the High-Resolution Pursuit (HRP) (Jaggi et al., 1995, 1998, 1999), the Evolutionary Pursuit (Ferreira da Silva, 2003), were proposed to achieve these so-called *atomic decompositions*. The term *atom* is originated from designing BFs as atoms.

In this paper, an analytical low-level representation of images, obtained by a decomposition process, namely, the matching pursuit (MP) algorithm, is proposed as a new way of describing objects through a *general continuous description* using an overcomplete dictionary of basis functions, that is invariant to rotation, scale, and translation transformations. We choose to characterize the objects by their *boundaries*. Other decompositions methods may be applied as well; nevertheless, we choose MP, as it is the basis of a class of algorithms for linear approximations of functions.

Let us consider an overcomplete dictionary $\mathcal{D}$ of parametric BFs g_γ that ensures a possible perfect reconstruction of the object image. The set of parameters defining each function g_γ represents the shape vectors of the object.

The shape description method introduced has the following steps:

- From a template image of the object $\mathcal{O}$ to recognize, we operate a decomposition using MP with the redundant dictionary $\mathcal{D}$.
- We obtain a set of parameters indicating the position, scale, orientation, and amplitude of a subset of basis functions, extracted from $\mathcal{D}$, that best represents the template image $\mathcal{O}$. This group of atoms, that have a spatial relationship, is interpreted as a *meta-atom* g_γ^M. This gives a description with both local and global features.
- Using this template description, we generate a new redundant dictionary $\mathcal{D}^{\mathcal{M}}$, containing all variations of the previous determined meta-atom. We obtain a pseudo-affine invariant object dictionary by introducing a shearing parameter, in addition to a translation, isotropic dilation, and rotation parameters.
- We then perform the recognition process through an MP decomposition of the target image $\mathcal{T}$ (where we want to find the object) using this predefined dictionary $\mathcal{D}^{\mathcal{M}}$. Setting up a minimal error threshold, we find as many objects as $\mathcal{T}$ contains that are similar up to an affine transformation to the template object.
- We increase the robustness of the recognition process using a coarse-to-fine scale-space recognition approach where the implicit scale-space representation of the object is used.
- The implicit extension of the object dictionary into a linear scale-space is given, using the scale parameter σ as an isotropic dilation of our basis functions.

Figure 1 presents the general scheme for shape recognition using an MP-based object description.

The paper structure is as follows: in Section II, we briefly introduce sparse representation methods whose general idea is related to the method we introduce hereafter. A more detailed description of Matching Pursuit is given in Section III, where some important
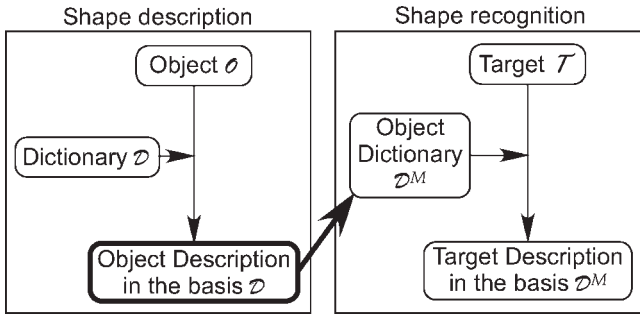
Figure 1. General scheme for shape description and recognition.

properties related to shape description purposes are presented. Section IV presents the shape descriptor, that uses anisotropic refinement (AR) atoms, and some examples are discussed in IV-B. Section IV-C discusses the possibility to choose another dictionary to describe the shape.

The natural extension of the model to a linear scale-space follows in Section V: this extension will find its complete justification in the next sections dealing with the recognition process.

Section VI first presents the object dictionary, based on the meta-atom extracted previously. A shearing parameter is added to obtain a pseudo-affine invariant dictionary. In section VII, we present the shape recognition process based on a MP process. The dictionary is extended into a linear scale-space by implicitly dilating the meta-atom (see section VIII-A). The shape recognition process is described at a given scale σ in the linear scale-space (see section VIII-B), and a multiscale scheme is proposed, based on a coarse-to-fine approach (see section VIII-C).

Section IX shows some results of shape recognition on different types of images, using different types of models, and also demonstrates the model capacity to contain more than one object into a single description.

II. RELATED APPROACHES

We are interested in methods related to spatial sparse representation models, with *both local and global features,* whose general idea is close to the iconic representation, first presented by Rao and Ballard (1995), later improved by Ben-Arie et al. (1996a,b) by reaching affine invariance property. In those works, the authors extract a localized affine invariant version of image patches using Gaussian kernels. The authors then use this shape signature in a one-to-one comparison framework to perform recognition.

Some authors have looked into local spatial representation to conduct recognition using local features: the banana wavelet decomposition scheme, introduced by Krüger and coworkers (1997) consists in extracting landmarks from images, in order to describe common features in the chosen set of images. The authors first produced a set of banana wavelets $B_{\alpha,c}: \mathbb{R}^2 \to \mathbb{C}$, by realizing the product of a curved Gaussian and a curved wave function as defined in Eq. (1):

$$B_{\alpha,c}(x,y) = \exp\left(-\frac{k^2}{2}\left(\left(\frac{x_c}{\sigma_x}\right)^2 + \left(\frac{y_c}{\sigma_y}\right)^2\right)\right)\exp(i\,kx_c)$$

$$\text{with}\begin{pmatrix} x_c \\ y_c \end{pmatrix} = C_c\left(\begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right),$$

$$\text{and } C_c\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x' - cy'^2 \\ y' \end{pmatrix}, \quad (1)$$

where α, c, $k \in \mathbb{R}$, C_c is the curving function, and k determines the frequency of the wavelet.

By varying the parameters α and c, they produce a so-called banana plant, with the real part of $B_{\alpha,c}$, shown in Figure 2.

By convolving the image with the banana plant, Krüger and Peters obtain the banana wavelet response for a given landmark. Finally, the landmarks joined with the best responses of each banana wavelet give a rough description of the surroundings boundaries of the object considered: this can be viewed as a set of atoms placed accurately along the most relevant edges.

Application-based attempts were drawn to look at special features such as nose localization in face recognition process, based on MP decomposition (Mallat and Zhang, 1993; Bergeaud and Mallat, 1995). In this face recognition application, Phillips (1994, 1998) introduces matching pursuit filters (MPFs) as an adapted wavelet expansion build from a training set of the object. This brings an adaptation to data, and to the pattern recognition problem to solve. Moreover, the provided description is able to capture both local and global features, as it consists in the sum of spatial features, represented here by the chosen set of wavelets.

Philips uses this description to describe a nose with Gabor wavelets, called *proto-nose*. By filtering the complete face image with the set of wavelets extracted, Phillips obtained the nose localization and performed face identification.

Matching pursuit filters have been used by Huang and Hsu (2000, 2001) to perform direct road sign recognition in a two-step approach: first, they perform a detection of a region of interest through template matching, and then the recognition part is conducted using MPF method based on a trained set of possible atoms combination to produce a template of a model, called *proto-sign,* highlighting its geometrical features.

Zhao and Nandhakumar (1998) combine a holistic template model with geometrical local features to perform face recognition. MPFs are used to give a set of geometrical features representing a training set of faces, while the linear discriminant analysis acts as a pattern classifier. The faces are then recognized based on a set of discriminant features, such as the nose and the eyes, identified with MPF parameters.

Other applications, such as voiced/unvoiced speech discrimination in noise, require the extraction of discriminant features to perform a classification process: Lobo and Loizou (2003) choose to represent the speech signal using Gabor functions, and conducts the decomposition of the signal using matching pursuit.

In Runkle et al. (1999), matching pursuit have also been used for target characterization purpose in synthetic aperture radar images, as wave-based features, extracted from multiple views with different orientations. The target discrimination step is then realized based on a Hidden Markov Model (Bharadwaj et al., 1999, 2001; Carin et al., 1999; Runkle et al., 1999; Runkle and Carin, 1999).
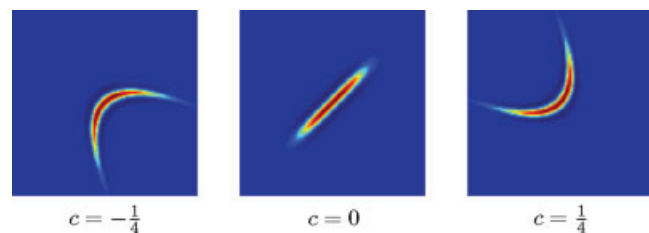

Figure 2. Banana plant for $c \in [-0.25, 0.25]$ and $\alpha = \pi/4$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

In this section, we have seen that matching pursuit algorithm, or more generally, sparse representation methods based on a linear decomposition process can be useful for extracting characteristics of a signal and subsequently for using them as features in a recognition process. We now focus in the next section on the MP algorithm itself.

III. DESCRIBING IMAGES WITH MATCHING PURSUIT

The goal of this section is to point out the basic concepts and the interesting properties of the MP process for the purpose of shape description, rather than describing it and its properties completely. We refer the reader to, e.g., Mallat and Zhang (1993), Bergeaud and Mallat (1995), Mallat (1998), Frossard (2000), and Vandergheynst and Frossard (2001) for more details.

The matching pursuit algorithm, first introduced for monodimensional signals by Mallat and Zhang (1993), is an iterative greedy process (Davis et al., 1997) that decomposes a function f in a Hilbert space $\mathcal{H}$, using a redundant dictionary $\mathcal{D} = \{g_\gamma\}_{\gamma \in \Gamma}$, where g_γ, are continuous BFs of unit norm usually called *atoms* (Mallat and Zhang, 1993), and Γ is the set of possible indexes. We will consider here bidimensional functions f.

Each step n of the algorithm consists in finding the atom $g_{\gamma_{n-1}}$ that minimizes the residues of nth order $R^n f$, when projecting orthogonally $g_{\gamma_{n-1}}$ on $R^{n-1}f$. Thus we have, with $R^0 f = f$,

$$R^n f = R^{n-1}f - \langle R^{n-1}f, g_{\gamma_{n-1}}\rangle g_{\gamma_{n-1}}, \qquad (2)$$

where $\langle \cdot, \cdot \rangle$ is the scalar product, and

$$|\langle R^{n-1}f, g_{\gamma_{n-1}}\rangle| = \sup_{\gamma \in \Gamma} |\langle R^{n-1}f, g_\gamma\rangle|. \qquad (3)$$

When $R^n f$ is minimized for a given $g_{\gamma_{n-1}}$, the projection between the previous residue and the actual atom $\langle R^{n-1}f, g_{\gamma_{n-1}}\rangle$ is maximized. Iteratively, we obtain for N atoms

$$R^N f = f - \sum_{n=0}^{N-1} \langle R^n f, g_{\gamma_n}\rangle g_{\gamma_n}, \qquad (4)$$

where $R^N f \to 0$ when $N \to \infty$ (L.K., 1987). This corresponds to the decomposition process.

From Eq. (4), we easily deduce the reconstruction process which corresponds to the MP invertibility property (Figueras i Ventura and Vandergheynst, 2002). Perfect reconstruction of f is possible if $N \to \infty$:

$$f = \sum_{n=0}^{\infty} \langle R^n f, g_{\gamma_n}\rangle g_{\gamma_n}. \qquad (5)$$

We can easily deduce f from Eq. (4) for a finite number of atoms N:

$$f = \sum_{n=0}^{N-1} \langle R^n f, g_{\gamma_n}\rangle g_{\gamma_n} + R^N f. \qquad (6)$$

When neglecting the residue $R^N f$ that tends to zero when N is big, we obtain an approximation $\hat{f}$ of f:

$$\hat{f} = \sum_{n=0}^{N-1} \langle R^n f, g_{\gamma_n}\rangle g_{\gamma_n}. \qquad (7)$$

The effects of this incomplete set of atoms can be seen in Figure 3, wherein we reconstruct Mona Lisa's face using a finite number of atoms. The result is refined as we consider more atoms, but note that the general structure of the image appears at very early stages when few atoms are considered. The hair structure can already be identified in Figure 3b, and the face components, such as mouth and eyes, appear in (c).
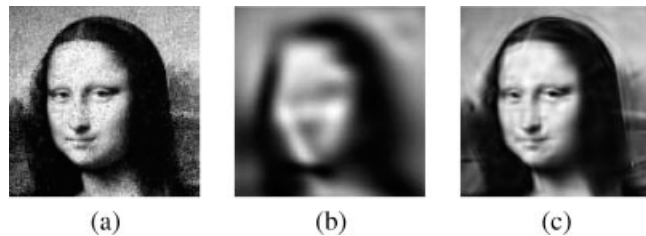

Figure 3. Face reconstruction of the Mona Lisa from a matching pursuit decomposition process using anisotropic refinement (AR) atoms (Vandergheynst and Frossard, 2001): (a) original Mona Lisa's face, (b) reconstruction with 20 AR atoms, and (c) reconstruction with 200 atoms.

It is also interesting to note the sketchy behavior of the process: the original image contains some noise, and the reconstruction is built up like a painting with strokes overlapping each other, to get to the final result. This follows Mallat's (1993) statement:

A matching pursuit decomposition provides an interpretation of the signal structures.

If we get back to the first step of the decomposition from Eq. (2), the function f is composed of the contribution of the first atom extracted g_{γ_0}, and the residue $R^1 f$ such that

$$f = \langle f, g_{\gamma_0}\rangle g_{\gamma_0} + R^1 f. \qquad (8)$$

From this equation, we can see that the residue $R^1 f$ is orthogonal to g_{γ_0} (Bergeaud and Mallat, 1995), and as g_{γ_0} is of unit norm, this leads to

$$\|f\|^2 = |\langle f, g_{\gamma_0}\rangle|^2 + \|R^1 f\|^2. \qquad (9)$$

Thus, following Eq. (6), we obtain the energy conservation property of MP (Mallat and Zhang, 1993; Bergeaud and Mallat, 1995):

$$\|f\|^2 = \sum_{n=0}^{N-1} |\langle R^n f, g_{\gamma_n}\rangle|^2 + \|R^N f\|^2. \qquad (10)$$

By looking at Eqs. (4) and (3), we can give a more intuitive look at this process: at each step, we choose the atom that will remove the biggest energy of the image. Note that we did not make any assumption on the shape of the atoms g_γ, apart from setting their norm to one, for energy conservation purposes.

Recent developments in the MP algorithms family were brought by Pece and Petkov (2000): they introduce a new fast atomic decomposition, the inhibition method, which is related to matching pursuit but allows updating of more than one coding coefficient per iteration, thus bringing a speed up in the decomposition process. The updated coefficients correspond to mutually orthogonal elements of the dictionary, and so avoiding a double contribution of the atoms in the function representation.

IV. OBJECT/SHAPE DESCRIPTION USING MP

A. Object Model. While describing the MP process, and following the iconic representation idea of analyzing shapes (Ben-Arie et al., 1996a,b), we already pointed out the framework for decomposing a shape in its principal "*shape vectors,*" which are our atoms. Our shape model follows the idea of having a patches-like structures to describe a shape, such as in Ullman and Basri (1991),

where they represent each view of an object by a linear combination of prototypical views.

This is also somewhat close to the textons concept: they are defined by Zhu et al. (2002) as *a mini-template that consists of a number of bases at some geometric and photometric configurations,* i.e., they are a clustering of basis functions that have a spatial relationship.

As we said in Section I, this first decomposition step can be achieved in many different ways: MP algorithm represents a possibility to achieve it, but other similar approaches linked to the sparse representation domain, such as e.g., BP (Chen et al., 1998), the inhibition method presented briefly previously (Pece and Petkov, 2000). However, e.g., BP is used to extract an exact representation of a signal: in our case, we will see that this condition is not automatically required. A controlled error may be admissible.

Phillips (1998) has also introduced the notion of describing a model, called proto-object, using only a few 2D directional wavelets with MPFs, as he did not required his description to be very accurate, but rather to fit an average description extracted from a subset of image models (see Section II).

On the contrary, we make here the choice of describing the template image of the object by characterizing its boundaries, following ideas that we introduced in Mendels et al. (2002).

Obviously, the choice of the dictionary to analyze the image has a big influence on the number of atoms needed to describe the shape accurately. This choice is intimately linked to the following question: *What is an optimal representation for an image given a library of templates functions and how to select it?* (Geiger et al., 1999).

Vandergheynst and Frossard (2001) have proven that AR atoms are more suitable to describe images and boundaries than, e.g., the original Gabor wavelet dictionary initially proposed in Bergeaud and Mallat (1995). Other families of atoms have been investigated as well, like, e.g., banana wavelets (see II) (Krüger and Peters, 1997) or the combination of Gaussian function with a triangular function (Moschetti et al., 2002), and show good results for the targeted applications (respectively face description and coding the displaced frame difference). Authors interested in the sparse representation of a signal have also produced different dictionaries: for example, Donoho and Huo (2001) introduced the beamlet dictionary as a dyadically organized library of line segments at a range of

locations, orientations, and scales, thus enabling a multiscale analysis approach. Recently, Peotta et al. (2003) have also introduced another curved function, incorporating a bending parameter, that shows significant improvement in very low bit rate image coding applications, using sparse representation of the signal.

Considering that the overall shape we are looking for corresponds to a high variation in the gradient (step edge), and as their description remains quite simple, we will use here the anisotropic refinement atoms (Vandergheynst and Frossard, 2001):

$$g_{\gamma_k}(x, y) = (4x^2 - 2)e^{-(x^2 + y^2)}, \qquad (11)$$

$$\text{with} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos(\theta_k) & \sin(\theta_k) \\ -\sin(\theta_k) & \cos(\theta_k) \end{bmatrix} \begin{bmatrix} (\bar{x} - p_{x_k})/\sigma_{x_k} \\ (\bar{y} - p_{y_k})/\sigma_{y_k} \end{bmatrix},$$

where $(\bar{x}, \bar{y})$ are the original pixel coordinates, $[p_{x_k}, p_{y_k}]$ are the horizontal and vertical translation, $[\sigma_{x_k}, \sigma_{y_k}]$ the horizontal and vertical scaling factors, and θ_k the orientation.

An example of an anisotropic refinement atom is shown in Figure 4.

The following set of parameters completely define g_{γ_k}:

$$\gamma_k \equiv [p_{x_k} \quad p_{y_k} \quad \sigma_{x_k} \quad \sigma_{y_k} \quad \theta_k]. \qquad (12)$$

In the proposed method, we describe the object $\mathcal{O}$ as a *given subset of basis functions* g_{γ_k} chosen during the greedy MP process: these are the first atoms extracted, each one of them containing at each step the biggest energy part of the model to fit. Thus, we obtain a compact representation of the object. We now consider this subset of functions as a *meta-atom* g_γ^M, which we normalize. Thus we have

$$\mathcal{O} : g_\gamma^M = \frac{\sum_{k=0}^K \langle R^k f, g_{\gamma_k} \rangle g_{\gamma_k}}{\| \sum_{k=0}^K \langle R^k f, g_{\gamma_k} \rangle g_{\gamma_k} \|}, \qquad (13)$$

where K is the number of atoms chosen for the description, $C_{g_{\gamma_k}} = \langle R^k f, g_{\gamma_k} \rangle$ is the coefficient factor, and f is the image of the template object.
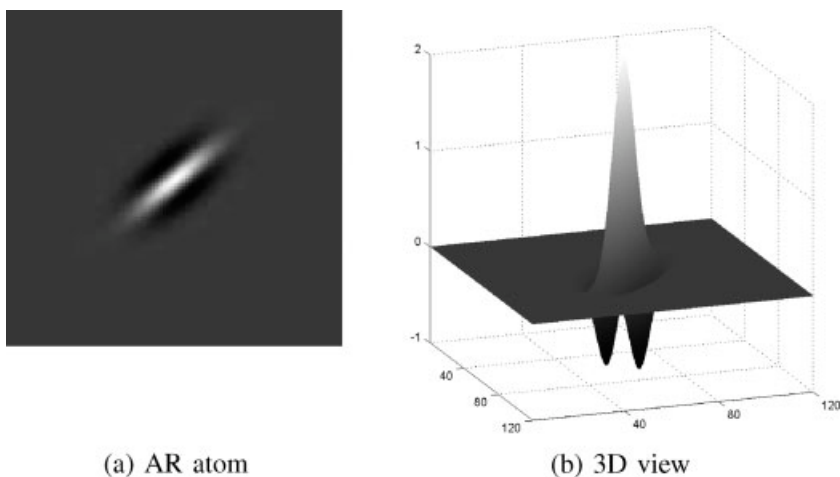

(a) AR atom

(b) 3D view

Figure 4. Anisotropic refinement atom example, with its 3D view.
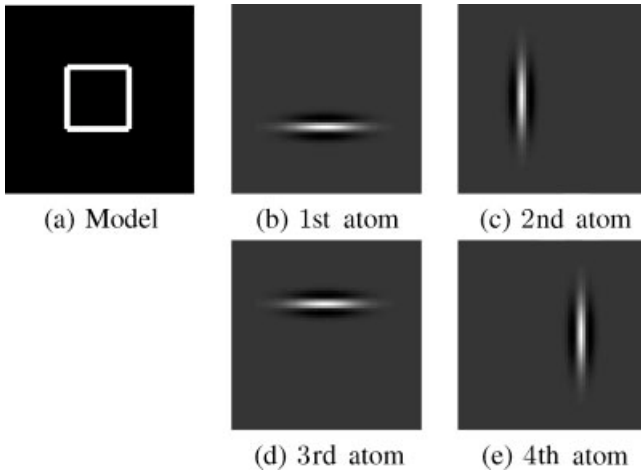
Figure 5. Original model and first four anisotropic refinement atoms.

So $\mathcal{O}$ can be represented as a set of those vectors:

$$g_\gamma^M : \begin{bmatrix} C_{g_{\gamma_0}} & p_{x_0} & p_{y_0} & \sigma_{x_0} & \sigma_{y_0} & \theta_0 \\ & & \vdots & & & \\ C_{g_{\gamma_i}} & p_{x_i} & p_{y_i} & \sigma_{x_i} & \sigma_{y_i} & \theta_i \\ & & \vdots & & & \\ C_{g_{\gamma_K}} & p_{x_K} & p_{y_K} & \sigma_{x_K} & \sigma_{y_K} & \theta_K \end{bmatrix}. \qquad (14)$$

Because of the nature of the initial anisotropic refinement dictionary $\mathcal{D}$, the description of the template object is invariant to similitude transformations, i.e., rotation, translation, and isotropic scaling. These properties follow the general statement given for shape descriptors in Section I: an object has the same shape when submitted to rotation, translation, and isotropic scaling, thus its description should remain the same. Of course, in the case of recognition and localization process, these invariance properties, such as, e.g., translation invariance, may not be needed.

Phillips, and Hsu and Huang (1998, 2001) use a full search procedure to obtain their atoms, as they need a rough description with few atoms. We prefer using genetic algorithms (GA) in our implementation for speeding up the search process, but this implies that we choose the best atom up to an insurance interval (Figueras i Ventura and Vandergheynst, 2001, 2002). So we obtain a suboptimal solution at each step, but the overall convergence of MP in the reconstruction process (see Eq. (5)) ensures to get a complete description of the object. At each step of the GA, we extract a set of possible solutions and test them against the current best solution found. Details on the implementation and procedure of the GA are given in Figueras i Ventura and Vandergheynst (2001). Note that the numerical complexity of the implementation can be reduced because of the structure of the dictionary and the structure of the atoms that are well localized; thus, calculation is performed only in the areas containing the biggest energetic part of each atom composing the meta-atom.

As the decomposition process removes at each step the biggest energy part of the residue, this implies naturally that the subset of the chosen basis functions is unique: consequently, the *object representation is unique*. The optimization algorithm selected to do the matching pursuit has an influence on this uniqueness property: in the case of genetic algorithms, the uniqueness is not guaranteed as we choose atoms up to a confidence interval. The uniqueness is guaranteed in the case of a full search procedure, that may be computationally intensive. Nevertheless, the choice of genetic algorithms remains consistent as the overall convergence is guaranteed by the MP process.

B. Template Example. Figures 5b–5e show the example of the first four atoms extracted to model a square template (64×64 image) using a dictionary $\mathcal{D}$ of anisotropic refinement atoms, and illustrate that they accurately model straight edges. As all square boundaries have an identical strength, the order of the extraction of the first four atoms (see Figs. 5b–5e) is not unique and depends on the optimization algorithm chosen, here the GA. Note that this does not change the overall description of the object.

If we look at the reconstruction in Figure 6, we clearly see that the error between the reconstruction and the model gets lower as we consider more atoms. So the sensitivity of our shape model will depend on the number of atoms we consider, linked with the complexity of the object to represent: Figures 6d–6f show that the "dude" model contains more variations than the square one. Consequently, we will require a higher number of atoms to get a description that reaches a minimal threshold of resemblance we choose. This threshold is set depending on the energetic part of the residue: e.g., we can state that we reach a good shape description when the residue contains less than 5% of the total energy of the model. Other adaptive criteria may of course be applied to set this threshold of resemblance.

As we do not give any restriction concerning possible symmetries that the template may present, the description obtained may not respect these symmetries. This is a direct consequence of the atom shape.

Note that adding more atoms when we have already reached the threshold of resemblance is not necessary. Moreover, we will introduce "nonfeature" elements, i.e., elements that are not here to model the shape, but to correct small deformations introduced in previous steps (Jaggi et al., 1995). In Figure 7, we can observe this effect when zooming on, e.g., the right arm: the reconstruction with 1000 atoms is better than the one with 100 atoms, i.e., the boundaries are
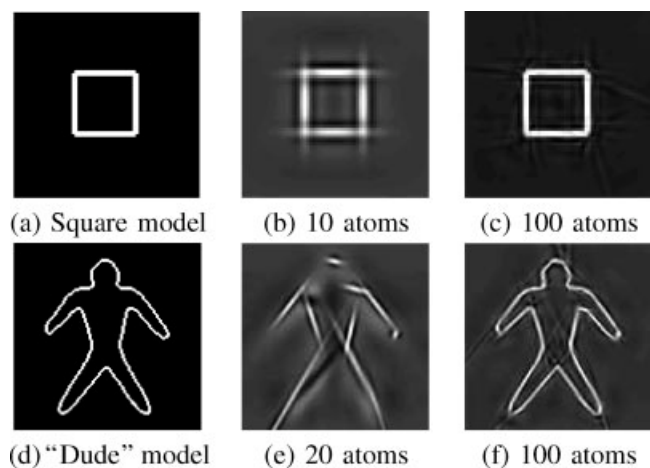

Figure 6. Variations in the reconstruction using different number of atoms: (a)–(c) square model and its reconstructions; (d)–(f) dude model and its reconstruction.
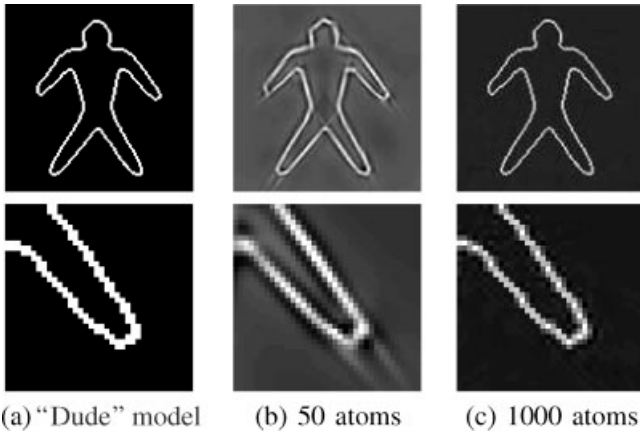
(a) "Dude" model   (b) 50 atoms   (c) 1000 atoms

Figure 7. Reconstruction artifacts with raising the number of atoms considered: (a) dude model and its reconstruction with (b) 50 atoms and (c) 1000 atoms; the first line gives the complete model, and the second line is a zoom on the right arm.

refined but also surrounded by a small local noise, that can be described as ringing artifacts.

Jaggi et al. (1995, 1998, 1999) propose a framework called High Resolution Pursuit to overcome such drawbacks, by changing the function that measures the correlation between the atom and the image, i.e., they do not minimize the L^2 norm of the residual, or measuring the atom contribution in terms of scalar products. The contribution is measured using the sum of a weighted scalar product between the image and the atom given at different scales.

C. Curved Atoms? We have chosen here anisotropic refinement atoms that are not curved, but the question "what are the best atoms?" remains. As we have shown in Figure 6 a shape (dude model) with curved parts in its boundaries, we can think that a curved atom will behave in a better way than a noncurved (straight) one. We have chosen here to describe the 'dude' model using curved basis functions $\text{atom}_{(t_x,t_y,\theta,r,\sigma_x,\sigma_y)}$ of unit norm from Peotta et al. (2003) described hereafter in Eqs. 15–17, as shown in Figure 8:

$$\text{atom}_{(t_x,t_y,\theta,r,\sigma_x,\sigma_y)}(x,y) = \mathcal{T}_{(t_x,t_y)}\mathcal{R}_\theta \mathcal{B}_r \mathcal{S}_{(s_x,s_y)} g_p(x,y), \quad (15)$$

where

$$g_p(x,y) = x\, e^{-\left(\frac{|x|}{\sigma_x}+y^2\right)}, \quad (16)$$

and

$$\mathcal{B}_r g_p(x,y)$$
$$= \begin{cases} g_p\left(r - \sqrt{(x-r)^2 + y^2},\, r\,\arctan\left(\frac{y}{r-x}\right)\right), & \text{if } x < r \\ g_p\left(r - |x|, x - r + \frac{r\pi}{2}\right), & \text{if } x \geq r \end{cases}$$
$$(17)$$

with r radius of the bending function, and the transformation $\mathcal{T}_{(t_x,t_y)}$ is a translation of (t_x, t_y), $\mathcal{R}_\theta$ is a rotation of θ, and $\mathcal{S}_{(s_x,s_y)}$ is an anisotropic scaling of (s_x,s_y) respectively in the x- and y-direction.

The artifacts due to the atom shape are present in both cases, even though they are not of the same nature and they seem to be diminished in the case of curved atoms for a perfect model. The

conclusion we can draw from this comparison is that a representation with noncurved atoms needs more elements to model the shape in presence of a high number of curved parts. Nevertheless, as the description remains simple and as the anisotropic refinement atoms fit our needs, we do not need in a first place to extend our overcomplete dictionary to curved functions, or to consider both dictionaries at the same time.

Moreover, in this case, it is necessary to ensure that straight parts will be modeled with straight atoms and not with a sum of curved ones. A good compromise may be to keep straight atoms for the beginning of the model and curved ones for small refinements.

V. MP-BASED SHAPE REPRESENTATION AND SCALE-SPACE

A. Extending the MP-Based Model in a Linear Scale-Space. In order to produce a more general description, we now extend the MP-based object representation in a linear scale-space: this will be very useful in the next sections for recognition purposes, if we want to be less sensitive to small deformations that can occur in presence of noise, or just when considering an object that is close to the model but not identical.

Using the MP-based shape model given in Eq. (13), we extend naturally the description in a linear scale-space, obtained by convolving the object with a Gaussian function G_σ, namely the normalized Gaussian kernel (Lindeberg, 1994), given for a bidimensional signal by

$$G_\sigma(x,y,\sigma) = \frac{1}{2\pi\sigma^2}\, e^{\left[-\frac{x^2+y^2}{2\sigma^2}\right]}, \quad (x,y) \in \mathbb{R}^2, \quad \sigma \in \mathbb{R}^+, \quad \sigma \neq 0.$$
$$(18)$$

As our atoms satisfy the linear heat equation (they result from the combination of a Gaussian function in the first direction and the second derivative of a Gaussian in the direction normal to the first one (Figueras i Ventura and Vandergheynst, 2001, 2002)), their dilated version is given by the product of each scale parameter with σ for all atoms ($\sigma_{x_k,\sigma} = \sigma_{x_k}\sigma$ and $\sigma_{y_k,\sigma} = \sigma_{y_k}\sigma$). Thus the explicit

(a) 20 atoms   (b) 50 atoms   (c) 100 atoms
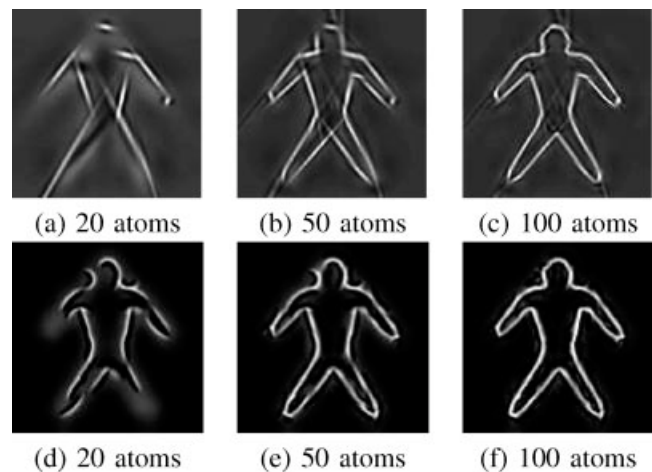
(d) 20 atoms   (e) 50 atoms   (f) 100 atoms

Figure 8. Variations in the reconstruction using different number of atoms using straight anisotropic refinement atoms and curved atoms (model courtesy of Ph.Jost): (a) original model, (a)–(c) reconstructions with straight anisotropic refinement atoms, and (d)–(f) reconstructions with curved atoms.

(a) Dude object and its description with 100 atoms
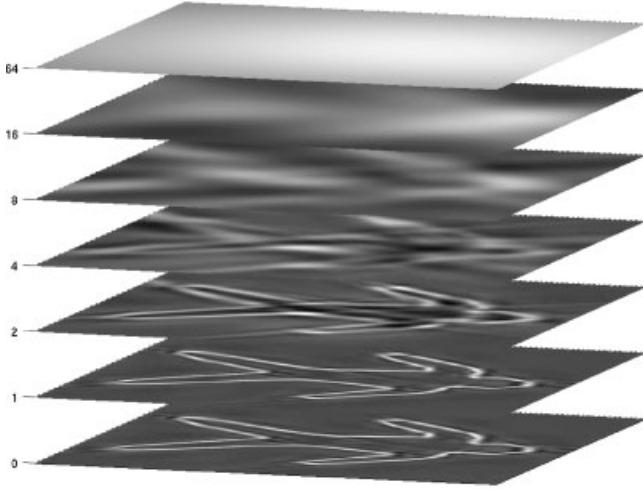
(b) Pyramidal representation of the scale-space
generated with meta-atom dilation

Original model $\qquad$ $\sigma = 2$ $\qquad$ $\sigma = 64$
(c) Planes of the dude scale-space with 100 atoms

Figure 9. Linear scale-space of the 'dude' object $\mathcal{O}$ with $\sigma = 1..64$, where level 0 represents the original model.

convolution is *not required* to build a scale-space representation of the object. Note that this dilation does not imply a modification of the atom location.

By following this property, the object $\mathcal{O}$ modeled by g_γ^M can be expressed at scale σ as

$$[g_\gamma^M]_\sigma = \sum_{k=0}^{K} \langle R^k f, g_{\gamma_k} \rangle g_{\gamma_k} * G_\sigma = \sum_{k=0}^{K} C_{g_{\gamma_k}} g_{\gamma_k, \sigma}, \qquad (19)$$

with $C_{g_{\gamma_k}} = \langle R^k f, g_{\gamma_k} \rangle$, and $g_{\gamma_k, \sigma} = g_{[p_{x_k}, p_{y_k}, \sigma_{x_k} \sigma, \sigma_{y_k} \sigma, \theta_k]}$.

Figure 9 illustrates the construction of a linear scale of the object "dude". The model at original scale is first shown (see (a)) and then its respective scale pyramid (see (b–c)), construct following Eq. (19). We clearly see here the classical scale-space implications: the loss of small details as scale increases, and the variations in the object boundaries locations.

B. Residual Error in Scale-Space Reconstruction. Figure 10 shows the difference between the reconstruction based on the MP description, through the scales, and the linear convolution of our original model with the normalized Gaussian kernel.

The MP-based representation in the scale-space follows what is obtained via the classical convolution method, but the reconstruction suffers from the incomplete form of the representation, as an infinite number of atoms is required to get a perfect reconstruction (see Eq. (5)). This confirms that we fit the model and its dilated ver-

sion up to an insurance interval, and the influence of each atoms is more visible in the scale-space extension.

In order to reduce the residual error of reconstruction, we perform the MP-decomposition over all scales that are considered at the same time: we have to choose each atom while maximizing its
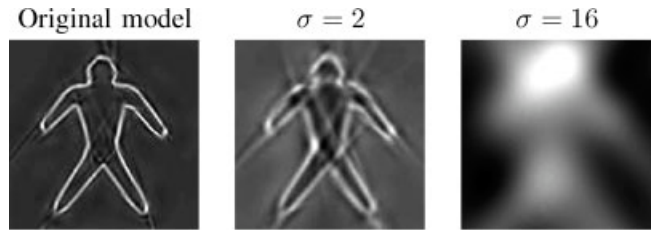
Original model $\qquad$ $\sigma = 2$ $\qquad$ $\sigma = 16$

MP-based object scale-space with 100 atoms

Result of the linear convolution between
the original model and G_σ

Figure 10. Pyramidal scale-space of the MP-based object and its linear scale-space obtained by convolution.

Original model $\sigma = 1$ $\sigma = 2$

Original scale-space created from object model
with n = 100 atoms

Scale-space obtained by simultaneous extraction
at each scale with n = 400 atoms

Result of the linear convolution between
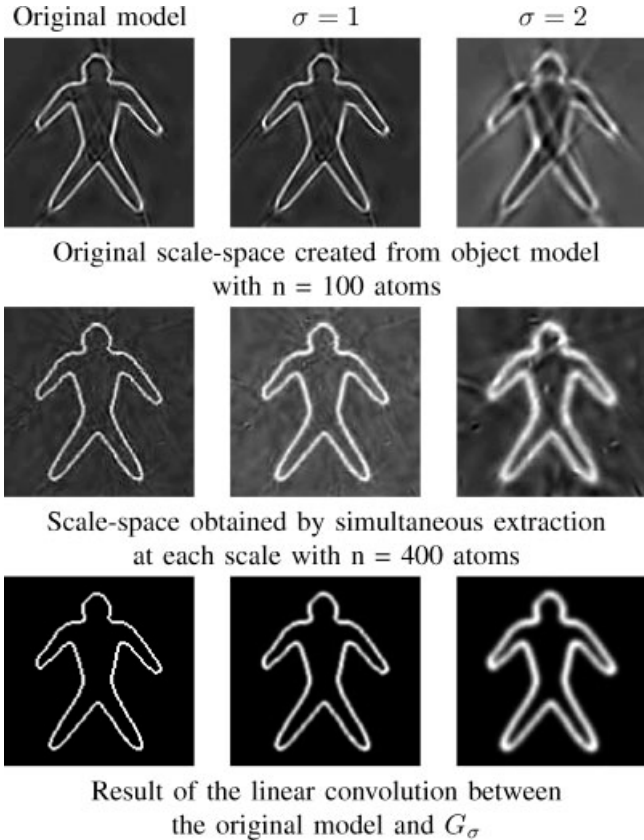the original model and G_σ

Figure 11. Pyramidal scale-space of the MP-based object with two different solutions and its linear scale-space obtained by convolution.

influence over all scales. We have to minimize at step i the following system for each residue $R^i f_{\sigma_k}$ with $k = 0..K$:

$$
\begin{cases}
R^i f &= f - \sum_{n=0}^{i-1} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n}, \\
R^i f_{\sigma_1} &= f * G_{\sigma_1} - \sum_{n=0}^{i-1} \langle R^n f_{\sigma_1}, g_{\gamma_n,\sigma_1} \rangle g_{\gamma_n,\sigma_1}, \\
R^i f_{\sigma_2} &= f * G_{\sigma_2} - \sum_{n=0}^{i-1} \langle R^n f_{\sigma_2}, g_{\gamma_n,\sigma_2} \rangle g_{\gamma_n,\sigma_2}, \\
&\vdots \\
R^i f_{\sigma_K} &= f * G_{\sigma_K} - \sum_{n=0}^{i-1} \langle R^n f_{\sigma_K}, g_{\gamma_n,\sigma_K} \rangle g_{\gamma_n,\sigma_K}.
\end{cases} \tag{20}
$$

Figure 11 shows the results of this approach. A qualitative visual estimation illustrates that this representation is more coherent with respect to the convolution process at higher scales than the one presented in Figure 10, but it requires more atoms as well: the reconstruction in Figure 11 needed *n times the number of scale considered,* compared to *n* atoms in Figure 10.

Moreover, by considering higher and lower scales in the minimization process, the choice of each atom may induce perturbation between the scales, as we use the same atom dilated with respect to its scale, and that boundaries move in a linear scale-space. This could lead to consider a nonlinear scale-space extension, or to consider atoms with centers $[p'_{x_k} p'_{y_k}]$ contained in an inverse pyramid, taking into account the scale factor σ (see Fig. 12).

But, as we will see later in section IX, it appears that the residual error induced by the model will be compensated in the recognition

process by the scale-space coarse-to-fine extraction approach, and so, a finer model, requiring more atoms, is not needed in this case.

VI. OBJECT DICTIONARY

The extracted subset of basis functions g_γ^M, i.e., the representation of the object $\mathcal{O}$, is invariant with respect to similitude transformations, because of the atom properties ((Vandergheynst and Frossard, 2001; Frossard, 2000; Peotta et al., 2003); and section IV-A for details). Here, we increase our range of variability by introducing a shearing parameter, thus leading to pseudo-affine invariance, as we consider isotropic dilations to maintain the coherence of the object model. We obtain a new redundant dictionary $\mathcal{D}^M = \{[g_\gamma^M]_{\gamma'}\}_{\gamma' \in \Gamma'}$, where $\Gamma' = [p'_x, p'_y, \sigma'_x, \sigma'_y, \theta', S_m]$ is the set of possible indexes, that represents the possible variations of the meta-atom.

The affine transformation of the basis functions g_{γ_k} gives, for any translation $[p'_x, p'_y]$, any isotropic scaling $\sigma'_x = \sigma'_y$, any orientation θ', and any shearing S_m:

$$
g_{\gamma_k, \gamma'}(x', y') = (4x'^2 - 2)e^{-(x'^2 + y'^2)}, \tag{21}
$$

with $\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta'_k) & \sin(\theta'_k) \\ -\sin(\theta'_k) & \cos(\theta'_k) \end{bmatrix} \begin{bmatrix} 1 & -S_m \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (\bar{x} - p'_{x_k})/\sigma'_{x_k} \\ (\bar{y} - p'_{y_k})/\sigma'_{y_k} \end{bmatrix},$

where $(\tilde{x}, \tilde{y})$ are the original pixel coordinates, and the new parameters of the atoms $[p'_x p'_{y_k} \sigma'_{x_k} \sigma'_{x_k} \theta'_k]$, $k = 0..K$ submitted to this affine transformation are obtained using the following relations:

$$
\begin{cases}
C'_{g_{\gamma_k}} = C'_{g_{\gamma_k}}, \\
\begin{bmatrix} p'_{x_k} \\ p'_{y_k} \end{bmatrix} = \begin{bmatrix} 1 & S_m \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma'_x & 0 \\ 0 & \sigma'_y \end{bmatrix} \begin{bmatrix} \cos(\theta') & \sin(\theta') \\ -\sin(\theta') & \cos(\theta') \end{bmatrix} \begin{bmatrix} p_{x_k} \\ p_{y_k} \end{bmatrix} + \begin{bmatrix} p'_x \\ p'_y \end{bmatrix}, \\
\sigma'_{x_k} = \sigma_{x_k} \sigma'_x, \\
\sigma'_{y_k} = \sigma_{y_k} \sigma'_y, \\
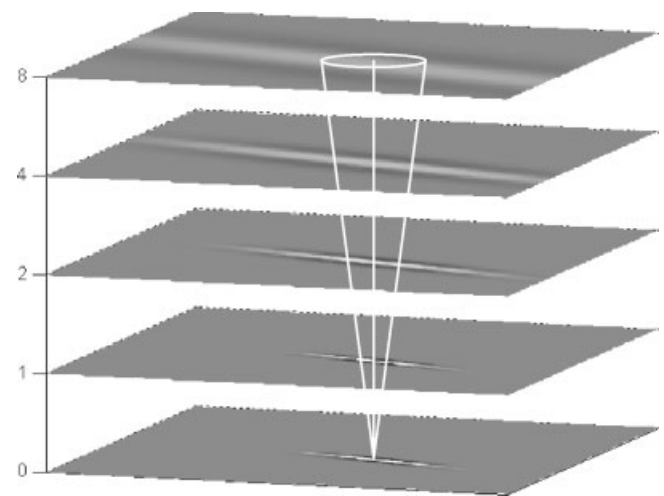\theta'_k = \theta_k - \theta'.
\end{cases} \tag{22}
$$

Figure 12. Atom dilated over the scales ($\sigma = 0..8$) and the inverse pyramid showing possible variations in the position of the center that could be introduced.

Figure 13. Examples of affine variations for the model 'dude' using the following parameters $[p'_x = 0, p'_y = 0, \sigma'_x, \sigma'_y, \theta', S_m]$.

The modified shape is then fully described by the subset of atoms with respect to its general affine transformation γ', given by $[p'_x, p'_y, \sigma'_x, \sigma'_y, \theta', S_m]$:

$$\gamma' \equiv [p'_x \ p'_y \ \sigma'_x \ \sigma'_y \ \theta' \ S_m] \quad (23)$$

We call the modified shape $[g^M_\gamma]_{\gamma'} = g^M_{\gamma \ [p'_x, p'_y, \sigma'_x, \sigma'_y, \theta', S_m]}$. The modified object $\mathcal{O}'$ submitted to the affine transformation γ' can be represented as a set of those vectors:

$$[g^M_\gamma]_{\gamma'} : \begin{bmatrix} C'_{g_{\gamma_0}} & p'_{x_0} & p'_{y_0} & \sigma'_{x_0} & \sigma'_{y_0} & \theta'_0 \\ & & \vdots & & & \\ C'_{g_{\gamma_i}} & p'_{x_i} & p'_{y_i} & \sigma'_{x_i} & \sigma'_{y_i} & \theta'_i \\ & & \vdots & & & \\ C'_{g_{\gamma_K}} & p'_{x_K} & p'_{y_K} & \sigma'_{x_K} & \sigma'_{y_K} & \theta'_K \end{bmatrix}. \quad (24)$$

So we can express the original model $\mathcal{O}$ as $[g^M_\gamma]_{\text{origin}} = g^M_{\gamma \ [0,0,1,1,0,0]}$.

We show in Figure 13 some examples of the "dude" model subject to given affine variations.

Figure 14 shows one column of the previous figure with $\theta' = 0$, and $S_m = 0.5$, to see the variations of the model when the scales $\sigma'_x = \sigma'_y$ vary from $[0.7, 1.1]$.

Note that the spatial relationship between the atoms that compose the meta-atom evolves in function of σ'_x and σ'_y: for $\sigma'_x = \sigma'_y$, the isotropic dilation does not modify the object as when submitted to a linear scale-space, as we have seen in the previous sections.

VII. MP-BASED SHAPE RECOGNITION

Shape recognition methods are closely related to the nature of the shape descriptor chosen. They are usually based on a one-to-one comparison framework, using a tree-based approach or a metric-based approach (Veltkamp and Hagedoorn, 1999; Zhang and Lu, 2004). In this paper, we aim to realize a shape recognition process based on an energetic measure as a similarity measure between two objects, that extracts and localizes as many objects similar to the chosen template, as present in the target image, with respect to invariance properties.

We join the views of Pavlidis (1980), van Tonder (2000), and Ullman (1989, 1991, 1999), by extending the concept of object description into its constituent parts to the MP-based description where the atoms are considered as shape fragments. We now extend this concept to shape recognition by considering the object itself as a constituent part of the image: the assumption we make relies on describing the target image in the template shape basis.

We use again the Matching Pursuit algorithm as a decomposition process for the recognition process.

A. Performing Shape Recognition Using a Matching Pursuit Algorithm. The recognition task follows directly the idea of describing the object $\mathcal{O}$ using a redundant dictionary: in order to find the object in the target image $\mathcal{T}$, we will now decompose this image using the predefined dictionary composed by all affine variations of g^M_γ given by $\mathcal{D}^M = \{[g^M_\gamma]_{\gamma'}\}_{\gamma' \in \Gamma'} = \{[g^M_\gamma]_{\gamma'_{a,n}}$ with $\gamma'_{a,n} \equiv [p'_{x_{a,n}}, p'_{y_{a,n}}, \sigma'_{x_{a,n}} \sigma'_{y_{a,n}}, \theta'_{a,n}, S_{m_{a,n}}]\}$ for any translation $[p'_{x_{a,n}}, p'_{y_{a,n}}]$, isotropic dilation $\sigma'_{x_{a,n}} = \sigma'_{y_{a,n}}$, rotation $\theta'_{a,n}$, and shearing $S_{m_{a,n}}$, where the index stand for the current affine transformation, and n to design the nth meta-atom.

Figure 15 presents a graphical explanation of the different set of indices that compose the affine variations of the object, where the simulated template is a square.

For the first decomposition step, we want to minimize the residue $R^1 \mathcal{T}$ given by

$$R^1 \mathcal{T} = \mathcal{T} - \langle \mathcal{T}, [g^M_\gamma]_{\gamma'_{a,0}} \rangle [g^M_\gamma]_{\gamma'_{a,0}}, \quad (25)$$

with $[g^M_\gamma]_{\gamma'_{a,0}} = [g^M_\gamma]_{[p'_{x_{a,0}}, p'_{y_{a,0}}, \sigma'_{x_{a,0}} \sigma'_{y_{a,0}}, \theta'_{a,0}, S_{m_{a,0}}]}$.

Following MP principles, the best match (here $[g^M_\gamma]_{\gamma'_{a,0}}$) will give the first object location in $\mathcal{T}$.

In order to find all the solutions present in the target image, we re-iterate the MP algorithm to find the next solution designed by $[g^M_\gamma]_{\gamma'_{a,n-1}}$ for the $(n-1)$th object, until no match is found considering a minimal error threshold α fixing the tolerance of the difference between the object detected and the model. As we have a meta-atom of unit norm, the process stops when, for the step n, the energetic contribution of the solution found is lower than the maximal (ideal) contribution of a meta-atom in the target image times the threshold:

$$\left| \langle R^{n-1} \mathcal{T}, [g^M_\gamma]_{\gamma'_{a,n-1}} \rangle \right| < \alpha |\langle \mathcal{T}, [g^M_\gamma] \rangle|. \quad (26)$$

Iteratively, we obtain the Nth object from

$$R^N \mathcal{T} = R^{N-1} \mathcal{T} - \langle R^{N-1} \mathcal{T}, [g^M_\gamma]_{\gamma'_{a,N-1}} \rangle [g^M_\gamma]_{\gamma'_{a,N-1}} \quad (27)$$

The residue $R^N \mathcal{T}$ after the extraction of each object is given by the substraction of the sum of the objects contributions to the target image:
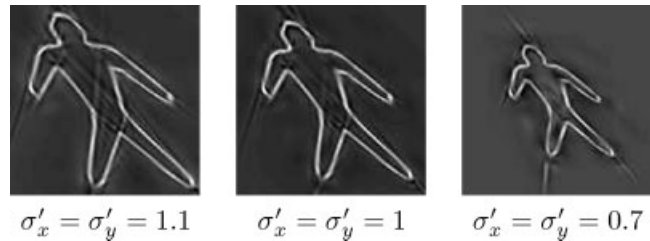

Figure 14. Detailed example of one affine variation for the model 'dude' using the following parameters $[p'_x = 0, p'_y = 0, \sigma'_x = \sigma'_y = [0.7, 1, 1.1], \theta' = 0, S_m = 0.5]$.
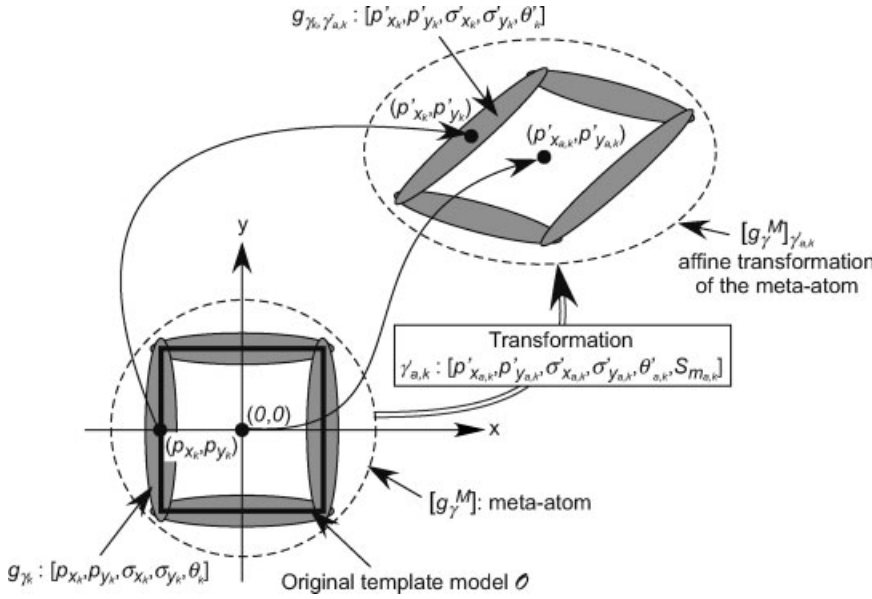
Figure 15. Examples of affine variations for the square model described with the meta-atom g_γ^M that becomes $[g_\gamma^M]_{\gamma'_{a,k}}$ when submitted to an affine transformation $\gamma'_{a,k} = [p'_{x_{a,k}}, p'_{y_{a,k}}, \sigma'_{x_{a,k}}, \sigma'_{y_{a,k}}, \theta'_{a,k}, S_{m_{a,k}}]$: the atom g_{γ_k} defined by $[p_{x_k}, p_{y_k}, \sigma_{x_k}, \sigma_{yk}, \theta_k]$ becomes $g_{\gamma_k, \gamma'_{a,k}}$ defined by $[p'_{x_k}, p'_{y_k}, \sigma'_{x_k}, \sigma'_{y_k}, \theta'_k]$.

$$R^N\mathcal{T} = \mathcal{T} - \sum_{n=0}^{N-1} \langle R^n\mathcal{T}, [g_\gamma^M]_{\gamma'_{a,n}} \rangle [g_\gamma^M]_{\gamma'_{a,n}}, \qquad (28)$$

where each $[g_\gamma^M]$ is a solution of the minimization problem (see Eq. (27)) for the template $\mathcal{O}$. Following the shape description step, we also use here genetic algorithms for the minimization process described in Eqs. (25) and (27).

B. Square Template Example. Using the square template model previously introduced and described (see Figs. 5 and 6), we perform the shape recognition task on the target image presented in Figure 16a, that contains two squares, a circle, and a triangle. Figure 16 shows the complete decomposition process. The first row corresponds to the first decomposition step using the Matching Pursuit algorithm, with a dictionary containing all affine variations of the square template model: the residue $R^1\mathcal{T}_1$ is obtained by subtracting
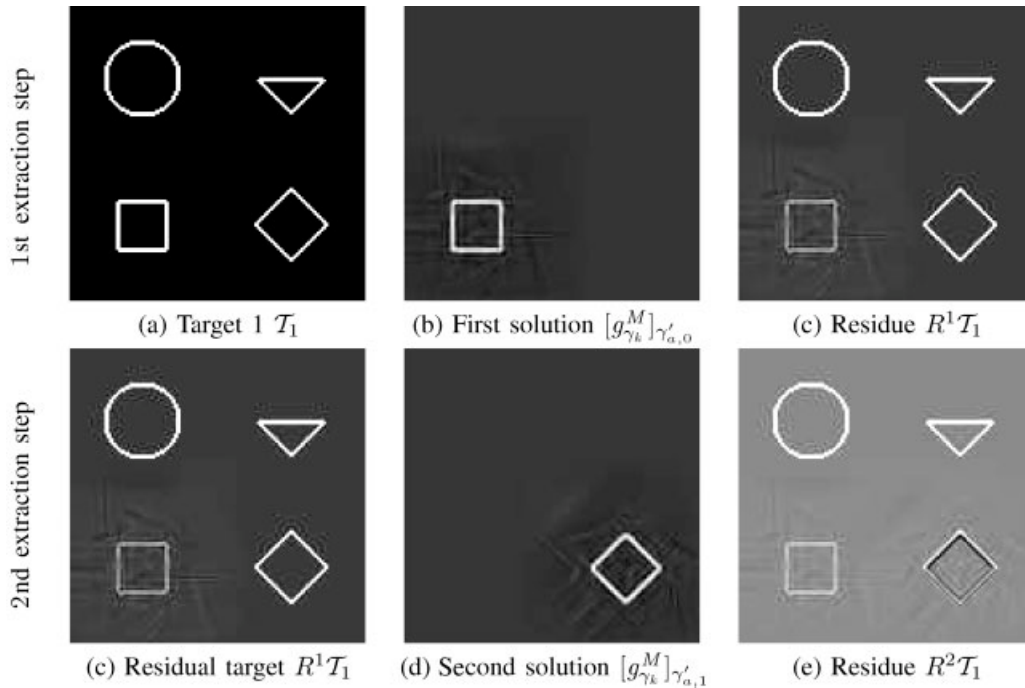

Figure 16. Example of a synthetic Target 1 and the MP-based shape recognition process: the first row shows the first step of the extraction (from (a)–(c)), and the second row shows the second square extraction (128 × 128 image).

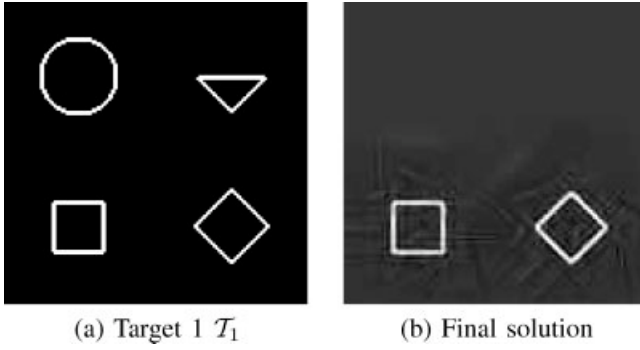(a) Target 1 $\mathcal{T}_1$ (b) Final solution

Figure 17. Target 1 and the 2 objects extracted corresponding to the square template.

the contribution of the first square found, shown in Figure 16b, from the target image $\mathcal{T}_1$. The second square is found in the residual part $R^1\mathcal{T}_1$ in the second extraction step, shown in Figures 16c–16e. As the next step produces a solution whose contribution is lower than the threshold of resemblance, we stop the MP decomposition at this point. The minimal error threshold giving the tolerance of the difference between the object detected and the model is fixed here to 90% of the ideal case, i.e., the object found is identical to the template object.

We have two solutions for the square, one identical to the model (see Fig. 16b), and the other one rotated by $45°$ (see Fig. 16d). The final solution of the extraction is given by their sum (see Fig. 17b).

For the minimization process (see Eq. (27)), the parameters of the GA (Figueras i Ventura and Vandergheynst, 2001) are set to 27 chromosomes and 50 generations for each step of MP decomposition, and will remain the same for all the following experiments.

Note that for a one-to-one object comparison application, i.e., the target image contains only one object, such as in CBIR applications (see, e.g., Wang et al., 2004) or classical face recognition process (see, e.g., Zhao and Nandhakumar, 1998), measuring the distance between two set of indexes to give a direct comparison on the indexes is possible, as the shape description is unique (see section IV). Indeed, by using a full search optimization, we would obtain the same set of indexes for both model and target in case of identity between them.

C. Towards an Extension to a Linear Scale-Space.
The tolerance threshold introduced permits to consider solutions that have local variations, but it may be too sensitive, or we may extract partial solutions that correspond to a local minima of the energy function.

Our model handles affine variations (see Figs. 13 and 14), but bigger perturbations (see Fig. 18) may lead to false solutions.

As we consider an energy function that is not necessarily convex, experience shows that we may be sensitive to local minima, even if GA leads theoretically to a global minimum. This may con-

Figure 18. Local variations in 'dude' model.

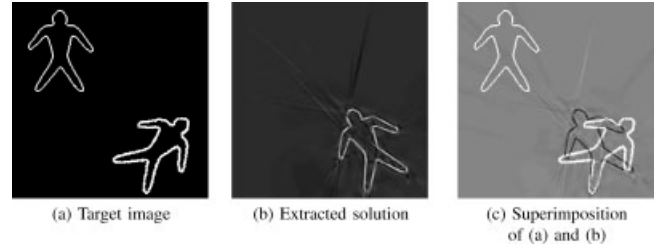(a) Target image   (b) Extracted solution   (c) Superimposition of (a) and (b)

Figure 19. Target image and the false match extracted.

duct to an incomplete solution: Figure 19c shows a false match where one leg of the model is captured correctly.

In order to increase the robustness of our process, we will use the linear scale-space extension of our model $[g_\gamma^M]_{\gamma',\sigma}$, and thus realize the Matching Pursuit-based Shape Recognition (MPSR) task into a linear scale-space of our target image $\mathcal{T}$.

VIII. MP-BASED SHAPE RECOGNITION USING A LINEAR SCALE-SPACE

A. Object Dictionary Extension in the Linear Scale-Space.
We need to extend our redundant dictionary $\mathcal{D}^M = \{[g_\gamma^M]_{\gamma'}\}_{\gamma'\in\Gamma'}$, in the linear scale-space, where the scale parameter is given by σ, with $\sigma > 0$.

Following Eq. (19), the scale-space extension of the meta-atom at scale σ gives the following parameters for each atom defined in Eq. (21):

$$
\begin{cases}
C'_{g_{\gamma_k}} = C'_{g_{\gamma_k}} \\
\begin{bmatrix} p'_{x_k} \\ p'_{y_k} \end{bmatrix} = \begin{bmatrix} 1 & S_m \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma'_x & 0 \\ 0 & \sigma'_y \end{bmatrix} \begin{bmatrix} \cos(\theta') & \sin(\theta') \\ -\sin(\theta') & \cos(\theta') \end{bmatrix} \begin{bmatrix} p_{x_k} \\ p_{y_k} \end{bmatrix} + \begin{bmatrix} p'_x \\ p'_y \end{bmatrix} \\
\sigma'_{x_k} = \sigma_{x_k}\, \sigma'_x\, \sigma \\
\sigma'_{y_k} = \sigma_{y_k}\, \sigma'_y\, \sigma \\
\theta'_k = \theta_k - \theta'
\end{cases}
$$
(29)

where $[C_{g_\gamma}, p_{x_k}, p_{y_k}, \sigma_{x_k}, \sigma_{y_k}, \theta_k]$ are the initial atoms parameters for $k = 0..K$, and $[p'_x, p'_y, \sigma'_x, \sigma'_y, \theta', S_m]$ are the parameters of the affine transformation γ'. We can observe that the scale parameter σ, corresponding to a local isotropic dilation, does not influence the positions $[p'_{x_k}, p'_{y_k}]$ but only the variances $[\sigma'_{x_k}, \sigma'_{y_k}]$ of each atom.

The object submitted to an affine transformation in a scale-space is then given by

$$
\mathcal{O}' : [g_\gamma^M]_{\gamma',\sigma} = \sum_{k=0}^K C'_{g_{\gamma_k}} g_{\gamma_k,\gamma'}\; G_\sigma = \sum_{k=0}^K C_{g_{\gamma_k}} g_{\gamma_k,\gamma',\sigma}, \tag{30}
$$

Figure 20 shows an affine variation of the dude model with $[p'_x, p'_y, \sigma'_x, \sigma'_y, \theta', S_m] = [0, 0, 0.8, 0.8, \pi/4, 0.5]$ and its associated scale-space for $\sigma = 1..4$.

Therefore, we obtain an extended overcomplete dictionary $\mathcal{D}^M_\sigma = \{[g_\gamma^M]_{\gamma',\sigma}\}_{\gamma'\in\Gamma',\sigma\in\mathbf{R}^+}$, where σ is the set of possible indexes in the linear scale-space, with $\sigma > 0$.

B. MP Decomposition at Scale σ.
We now describe how to realize an MP decomposition at a scale bigger than the original scale ($\sigma > 0$).

$$R^N\mathcal{T}_\sigma = R^{N-1}\mathcal{T}_\sigma - \langle R^{N-1}\mathcal{T}_\sigma, [g_\gamma^M]_{\gamma'_{a,N-1},\sigma}\rangle [g_\gamma^M]_{\gamma'_{a,N-1},\sigma}. \quad (33)$$
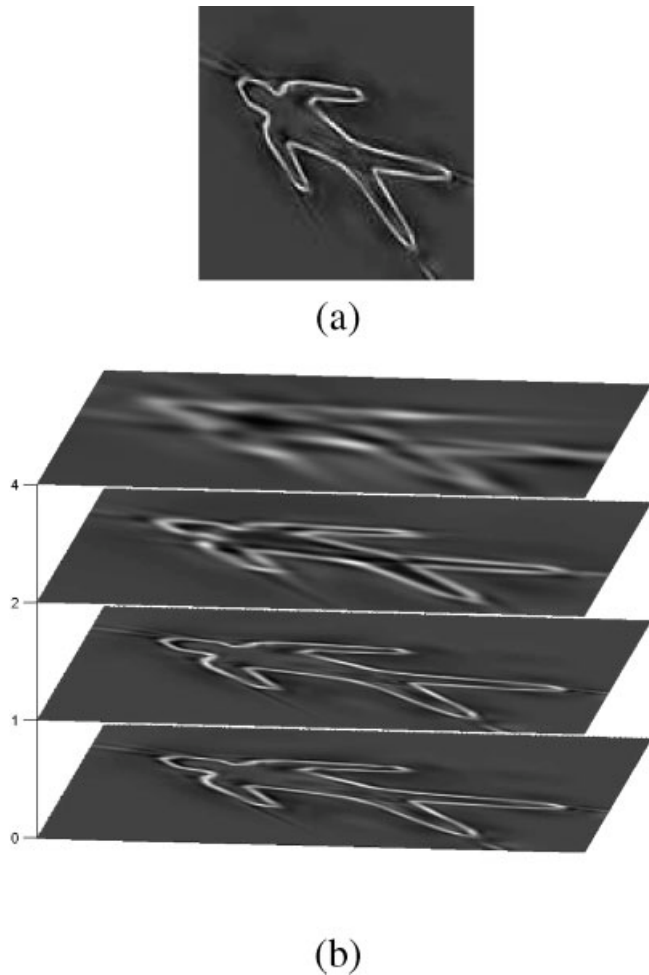

Figure 20. Example of affine variations for the model 'dude', and its associated scale-space for $\sigma = 1..4$: (a) and (b) show respectively an affine variation of the model with $[p'_x, p'_y, \sigma'_x, \sigma'_y, \theta', S_m] = [0, 0, 0.8, 0.8, \pi/4, 0.5]$, and its associated scale-space, where level 0 represents the affine variation of the original dude model.

The decomposition process described in Eq. (25) for the original scale of the object gives at the first step the object $[g_\gamma^M]_{\gamma'_{a,0}}$ where $[g_\gamma^M]'_{\gamma_{a,0}} = \Sigma_{k=0}^K C_{g_{\gamma_k}} g_{\gamma_k \gamma'_{a,0}}$.

Performing this MP decomposition at a different scale $\sigma > 0$ implies that our target image $\mathcal{T}$ becomes $\mathcal{T} G_\sigma$, and that the object $[g_\gamma^M]_{\gamma'_{a,n}}$ becomes $[g_\gamma^M]_{\gamma_{a,n},\sigma'}$, following Eq. (30).

The first MP decomposition step at a fixed scale $\sigma > 0$ becomes

$$R^1\mathcal{T}_\sigma = \mathcal{T} G_\sigma - \langle \mathcal{T} G_\sigma, [g_\gamma^M]_{\gamma'_{a,0},\sigma}\rangle [g_\gamma^M]_{\gamma'_{a,0},\sigma}. \quad (31)$$

We obtain the following objects present in the target image by reiterating the MP decomposition at scale σ. For the second object $[g_\gamma^M]_{\gamma'_{a,1}}$, we have to minimize $R^2\mathcal{T}_\sigma$:

$$R^2\mathcal{T}_\sigma = R^1\mathcal{T}_\sigma - \langle R^1\mathcal{T}_\sigma, [g_\gamma^M]_{\gamma'_{a,1},\sigma}\rangle [g_\gamma^M]_{\gamma'_{a,1},\sigma}. \quad (32)$$

If N objects are present in $\mathcal{T}$, we obtain for the Nth object $[g_\gamma^M]_{\gamma'_{a,N-1}}$ by minimizing the residue $R^N\mathcal{T}_\sigma$:

C. Performing Shape Recognition Through Multiscale MP. Multiscale strategies in image processing have been widely used to increase the robustness of the process considered, such as, e.g., shape recognition: they have been jointly used, e.g., with Markov Random Fields (Zhang and Ma, 2000), with a tree-based comparison (Demirci et al., 2003), or with a curvature scale-space descriptor (Mokhtarian, 1995). These three approaches have in common that they consider coarse-to-fine process. This is of course one of the two possibilities the scale-space approach offers, the other one being a stack approach.

As we are able to decompose an image $\mathcal{T}$ into the object basis at any given scale σ (see Eqs. (31)–(33)), we now drive through the scale to realize the multiscale recognition task, in order to overcome local minima problems such as, e.g., the result shown in Figure 19. As shown in Figure 21, the extraction of a possible object solution begins at a coarser scale (here represented by σ_2), and each solution obtained at a coarser scale is then propagated iteratively to finer scales (σ_1, σ_0) to act as an initialization step, until we reach the object scale ($\sigma = 0$). We minimize for each object $[g_\gamma^M]_{\gamma'_{a,n}}$ at each scale σ_k with $k = N..0$:

$$R^{n+1}\mathcal{T}_{\sigma_k} = R^n\mathcal{T}_{\sigma_k} - \langle R^n\mathcal{T}_{\sigma_k} : [g_\gamma^M]_{\gamma'_{a,n},\sigma_k}\rangle [g_\gamma^M]_{\gamma'_{a,n},\sigma_k}, \quad (34)$$

with $\sigma_N > \cdots > \sigma_k > \cdots \geq \sigma_0$.

This process leads to more stable solutions as we propagate an approximate solution (the dilated object) through the scales. And thus, it is less sensitive to small variations of the object, as they have less influence at higher scales.

We show an example of this approach, using the target $\mathcal{T}$ presented in Figure 19a to extract both dude models, given in Figures 6d–6f: we conduct the multiscale approach building each tree following each object. Figure 22 shows the pyramid scale that conducts to respectively the first "dude" object, and the second "dude" object. We can see that the direction of each object is refined through scale, up to the original object scale.
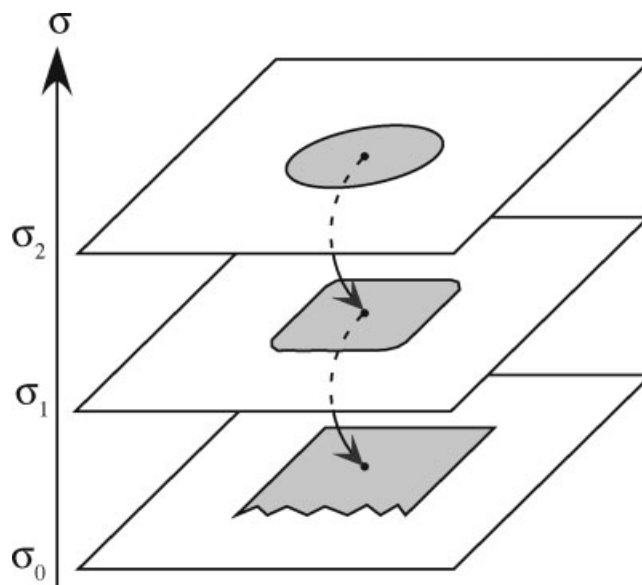

Figure 21. Scale-space recognition of the object $\mathcal{O}$ in a target $\mathcal{T}$.
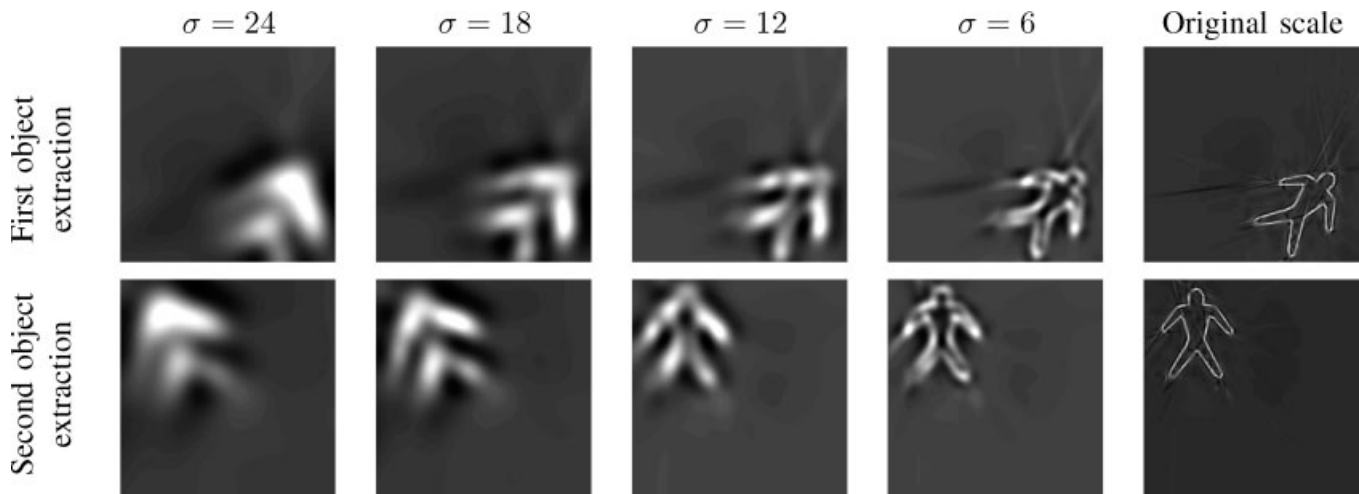
Figure 22. First and second pyramidal object extraction from $\mathcal{T}$.

Figure 23 shows the final result of the extraction, obtained by summing each step of the decomposition. Both objects are correctly extracted.

In this synthetic example, both objects have the same scale, and so the coarser scale chosen here is the same for both objects. But this question of the best coarser scale selection is closely related to this kind of coarse-to-fine extraction, as if we choose a scale bigger than the optimal one, we might miss one object at this scale, and consequently not propagate this solution to finer scales. Strategies for automatic scale selection have been investigated by, e.g., Lindeberg (1996). This represents a possibility to determine the ideal scale to begin the process, but is difficult to set up as the basic scale of the object varies. Another approach could be to perform first the extraction for the initial solutions we obtain and to refine afterwards in the considered scale if new solutions to our minimization problem exist. Combining these considerations could lead us to consider the second possibility for a multiscale approach, i.e., realizing the extraction in a stack scale-space. In this case, we would extract each object in all planes at the same time.

We will now show some results of this approach with different kinds of images, and shape models.

IX. RESULTS

A. Introduction. In this section, we will perform shape recognition process on different kind of images, using different shape models, namely,
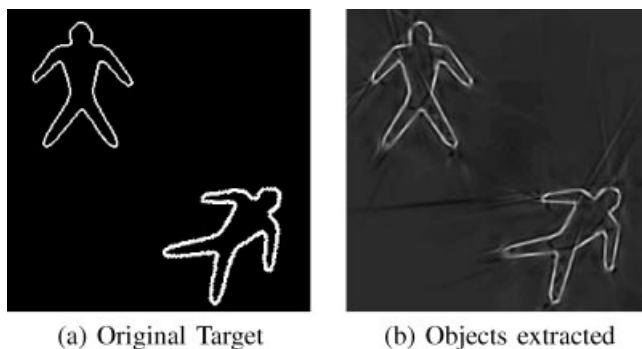
- two kind of tests on synthetic images using the 'dude' model previously used, where salt and pepper noise will be added afterwards;
- two kind of tests on natural aerial images, to detect boats parked side by side; with a model of boat that gives a stronger weight to the front part (grayscale model), where noise is added afterwards; and
- a test on recognition of brain ventricles in brain magnetic resonance (MR) image: we model from a left ventricle extract from a first brain a pair of ventricles (group of objects) to be matched in a second brain.
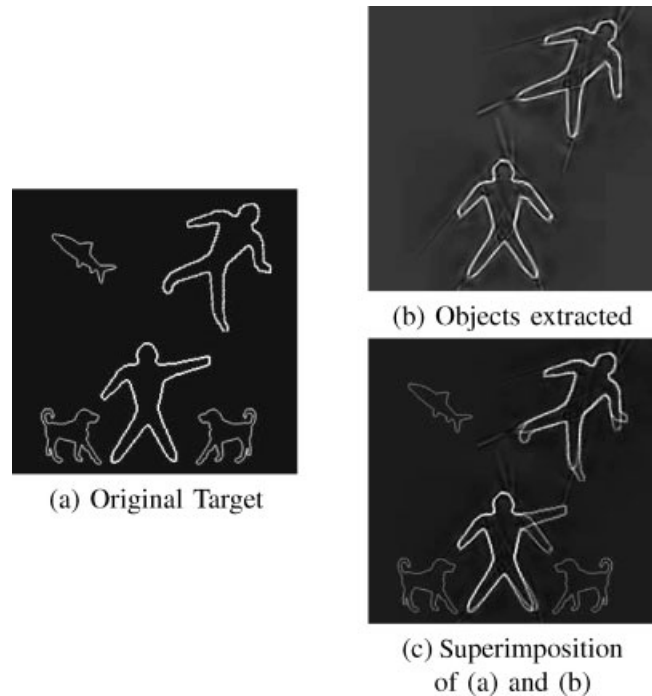

(a) Original Target

(b) Objects extracted

Figure 23. Original target $\mathcal{T}$ and the result of the extraction.

(a) Original Target

(b) Objects extracted

(c) Superimposition of (a) and (b)

Figure 24. Original target $\mathcal{T}$, the result of the extraction and the superimposition of (a) and (b) where the corresponding parts are shown in white (c).
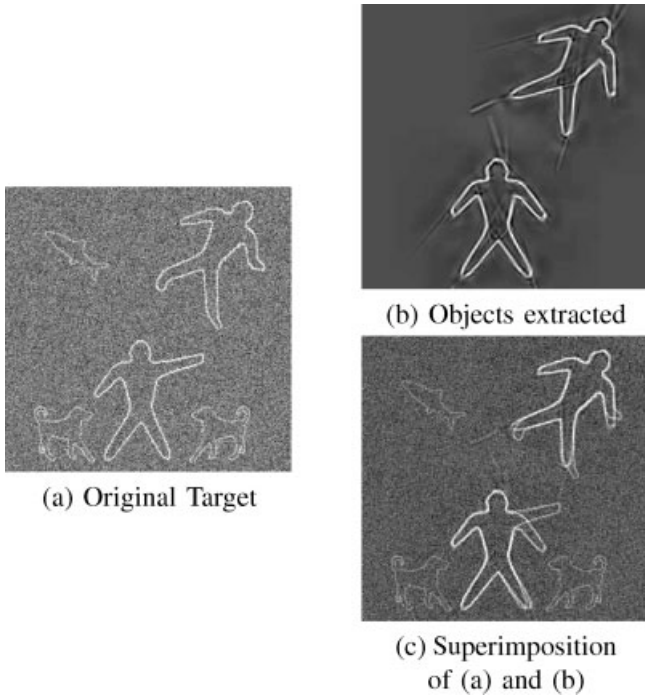
(a) Original Target

(b) Objects extracted

(c) Superimposition
of (a) and (b)

Figure 25. Frames of the recognition process of the 'dude' model in the target T with white noise added: (b) the result of the extraction and (c) the superimposition of (a) and (b).

B. Synthetic Images. In Figure 24, we realize a detection test on a synthetic image build with two objects presenting bigger variations in the dude model (see Fig. 18) and other different objects,
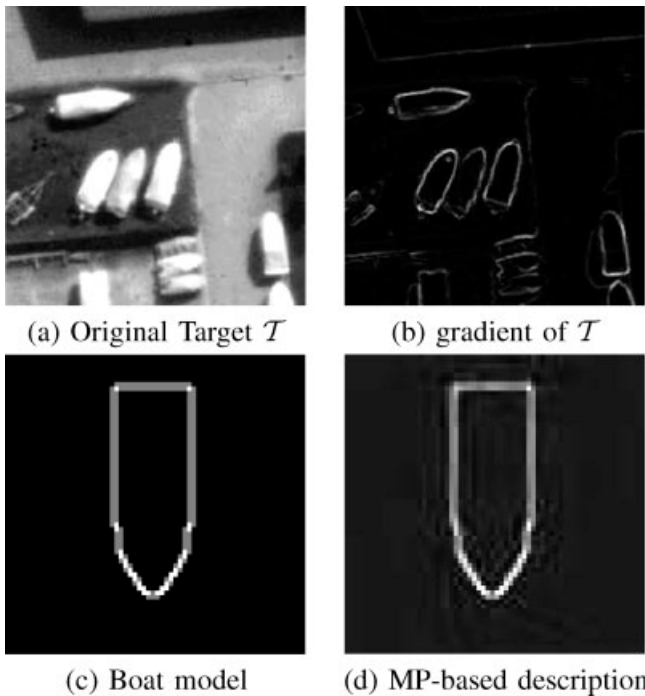
such as dog and fish shapes. Note that the second variation of the dude model that has an arm going up, has also a difference in the right leg: from now on, we denote it as "dude no. 2," and the other one as "dude no. 1."

We show here the increase in stability due to the scale-space introduction in the shape recognition task, in order to avoid false detection in adverse conditions (presence of local minima) such as in Figure 19, where the recognition process leads to a false solution. Note that the model shown in Figure 24a is more complex than the one presented in Figure 23.

Figure 24 presents the recognition process result for the multi-scale method described in Section VIII-C, which is denominated as the "*coarse-to-fine approach.*"

In Figure 24c, both objects corresponding to our model are correctly found, with respect to their local deformations, which are not matched as we do not allow nonrigid deformation in our model. Matched parts are shown in white.

In Figure 25, we show the recognition process conducted on the same target model of Figure 24a, with white noise added (SNR = 5 dB (Mallat, 1998)). We observe that the coarse-to-fine approach
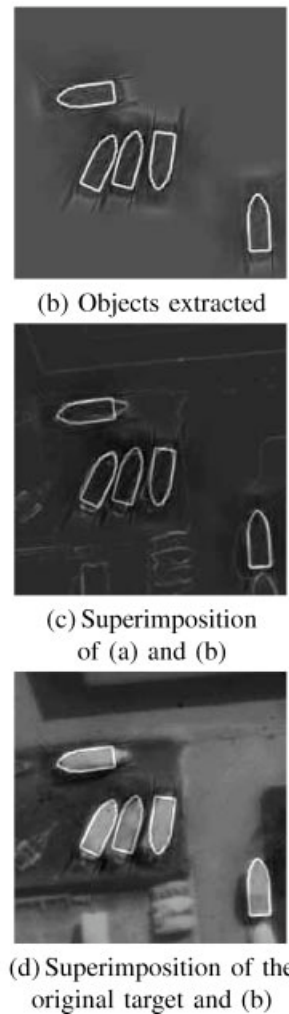
Coarse-to-fine approach

(b) Objects extracted

(a) Original Target T

(b) gradient of T

(c) Boat model

(d) MP-based description

Figure 26. (a) Original target T, (b) its gradient, (c) the boat model used for the extraction process, and (d) its MP-based representation with 100 atoms.

(a) Original target gradient

(c) Superimposition
of (a) and (b)

(d) Superimposition of the
original target and (b)

Figure 27. Original target T, the result of the extraction and the superimposition of (a) and (b) where the corresponding parts are shown in white (c).

remains robust to the presence of noise in the image, as both objects are correctly recognized.

C. Real World Images. In order to assess the stability of the process in real world images, the following recognition tests are conducted on an aerial image containing several boats (see Fig. 26a). The model to be extracted is a boat (see Fig. 26c) modeled by our MP description with 100 atoms (see Eq. (13), shown here in Figure 26d. Note that we use a grayscale model of the boat as we want to give a stronger weight to the front of the boat.

The detected objects are presented in Figure 27. Globally, we can say that the five biggest boats are found, even though some solutions (two for the case in Fig. 27b) are validated with a wrong orientation: this shows a possible problem that occurs when the object searched presents strong symmetries.

This could be considered as a wrong match. However, in this case, we decided to validate it as the solution found is in agreement with the minimal threshold of resemblance we chose. As we fix a threshold of resemblance, we do not give a matching percentage as
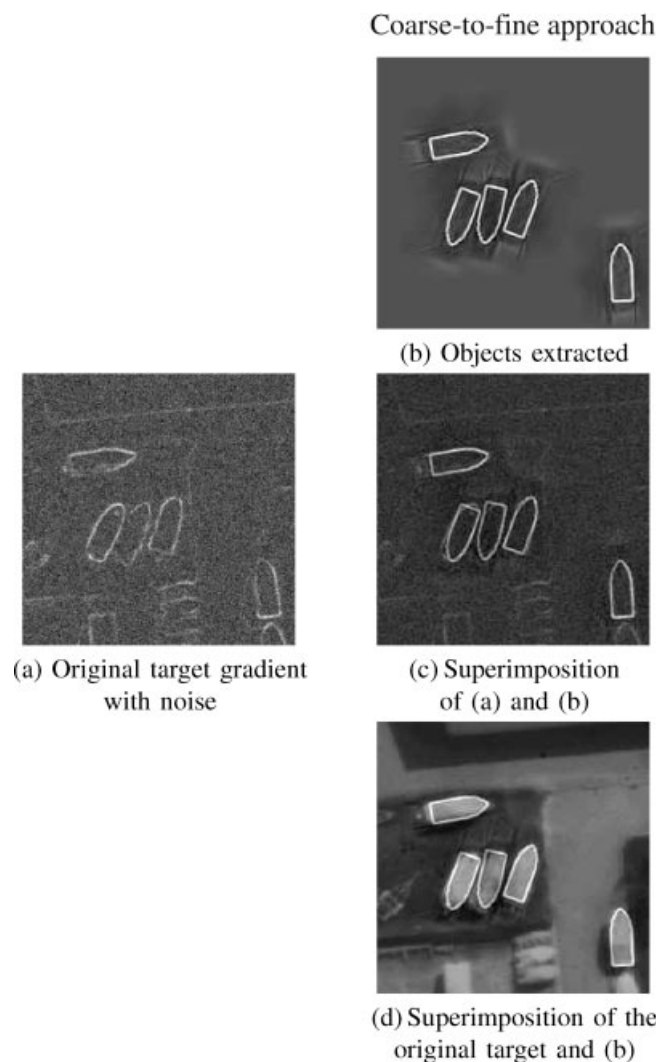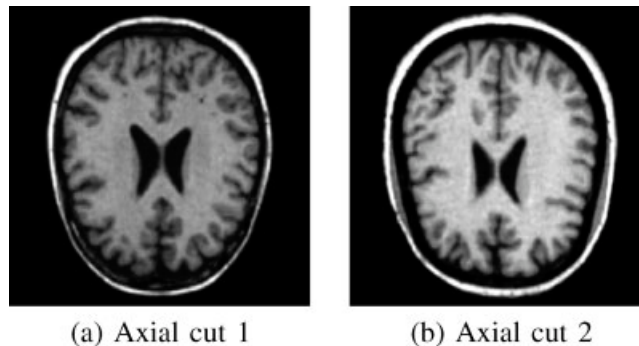

(a) Axial cut 1 (b) Axial cut 2

Figure 29. Original axial cut for the model of the left ventricle (LV) of the human brain and the targeted axial cut for ventricle extraction.

it is usually done in the pattern recognition domain: we just validate the solutions that verify Eq. (26) (see section VII-A).

The presence of high gradient level in the back of the boats due, e.g., to the presence of a motor may also cause problems: as we give a higher weight to the front of the boats, the process may stay stuck into a local minimum.

This process overcomes the problem of having close solutions, such as the group of three boats in the middle of the target image. However, small boats at the bottom of the image, whose gradients are low, are not detected, as their contribution is too low when compared with the threshold of resemblance we fix.

In Figure 28a, we introduce white noise directly on the gradient of the target image (SNR = 9 dB). The scale-space approach behaves correctly: all the five boats are found (see Fig. 28b) for the objects), even if the orientation problem is still present in two solutions of boats. As the noise added covers the gradient signatures of the small boats, the process is unable to extract them. For more clarity, Figure 28d presents the target image without noise, with the boat solutions found superimposed.

Because of the integration of the scale-space, the white noise present in Figure 28a is averaged through the scales: its influence diminishes with increasing σ. Therefore, the shape signatures at higher scales are less disturbed than the ones at lower ones, and so helps guiding the extraction process at lower scales.

D. Medical Images. In Figure 29, we consider two axial slices of a MR image of the brain: we want to recognize both ventricles in the second axial cut (see Fig. 29b) based on a dictionary of ventricle shapes created from another different brain (see Fig. 29a). In this example, we realize the tests with the coarse-to-fine approach, as our interest here is more in demonstrating the variability of the shape model when considering a group of separate objects, and its integration in the MP-based shape recognition process.

We extract the left ventricle (LV) from the axial cut 1 (see Fig. 30a), and then construct a model with $n = 100$ atoms (see Fig. 30b) based on the gradient of the LV.

We generate a new dictionary for the model of the right ventricle (RV), using an axial symmetry, achieved by shifting p_{x_k} indexes and changing the atoms orientations ($\theta'_k = -\theta_k$). By concatenating both LV and RV descriptions, we obtain a model for a pair of identical ventricles: this one is fully described using $2n$ atoms, while maintaining their intrinsic spatial relationship (see Fig. 30c). This relationship is inversely linked to the scale of the ventricles: as the

Coarse-to-fine approach

(b) Objects extracted

(a) Original target gradient with noise

(c) Superimposition of (a) and (b)

(d) Superimposition of the original target and (b)

Figure 28. Original target $\mathcal{T}$ with noisc added, the result of the extraction and the superimposition of (a) and (b) where the corresponding parts are shown in white (c).

(a) LV extracted from Axial cut 1   (b) LV modelled with $n = 100$ atoms   (c) Both ventricles
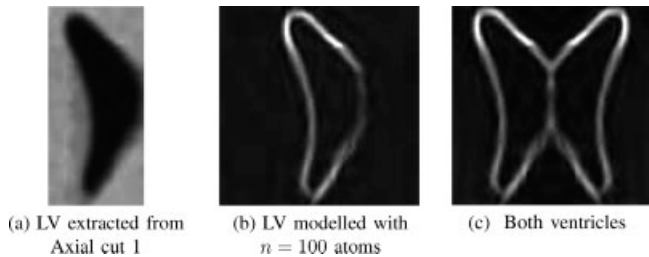
Figure 30. Original model $\mathcal{O}$ for the left ventricle (LV) of the human brain and its extension to both ventricles.

ventricles have a bigger scale, the distance between them reduces. This relationship is illustrated in Figure 31: in (a), the ventricles have a scaling factor σ_{d_1}, and are at a distance d_1, and in (b), the ventricles have a scaling factor σ_{d_2}, and are at a distance d_2. We propose the following relation linking the distance d_i between the ventricles and their scaling factor σ_{d_i}, with $i = 1,2$:

$$\boldsymbol{d_i} = \frac{A}{\sigma_{d_i}}, \tag{35}$$

where A is a positive constant, $A \in \mathbb{R}_*^+$.

As shown in Figure 31, having $\sigma_{d_1} > \sigma_{d_2}$ implies that $d_1 < d_2$.
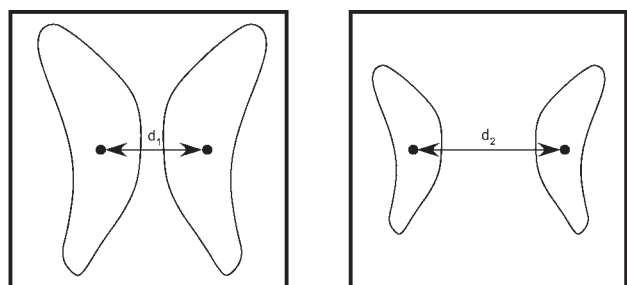
The extraction is then conducted for both ventricles simultaneously in the axial cut 2 (see Fig. 32).

The solution, highlighted in white, is shown twice: we first superimpose it to the axial cut (see Fig. 32b), and second, we show in Figure 32c the zoom of the solution, thus presenting its correspondence with the gradient of the axial cut 2 (in black).

Both ventricles are correctly localized, even though the location of the RV presents a small shift with respect to the gradient. This is due to the application of the same scaling factor on the two ventricles, as they are contained in an unique model, but we can easily observe that the size of ventricles differs. In these conditions, the minimization process brings the best trade-off between the different size of the ventricles, maximize the contribution of the left ventricle, and rotate the block of ventricles to get the best match possible on the right one.

If we want to achieve a better match, we have to extract both ventricles separately: we note that the scaling factor is not identical in both cases, confirming the shape difference of the ventricles.

The result of the extraction obtained by this way is now correct for both ventricles (see Fig. 33): the solution is shown in white in both Figures 33b and 33c.

(a) Axial cut 2   (b) Solution superimposed on (a)

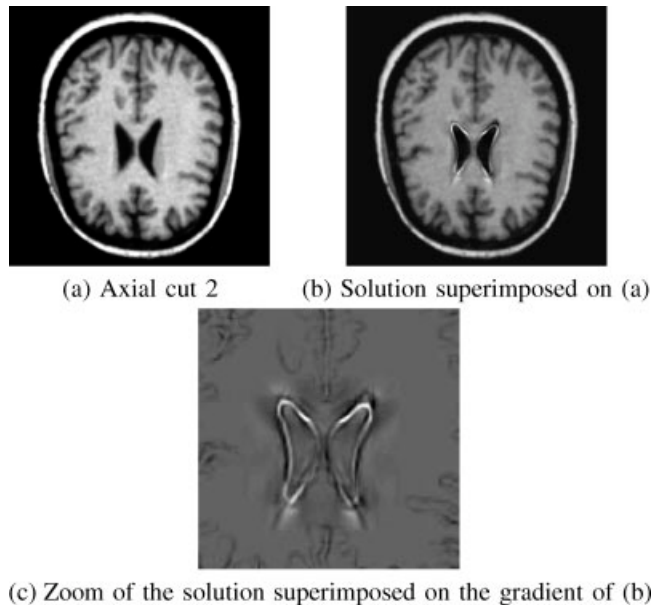(c) Zoom of the solution superimposed on the gradient of (b)

Figure 32. Axial cut of a MR brain image and extraction of both ventricles.

Through the example shown in Figure 32, we demonstrate the ability of the MP-based representation to describe simultaneously multiple objects using a single model, as well as its accuracy in different conditions. Even if the extraction of the ventricles step by step, i.e., left ventricle then right ventricle, gives the optimal solution, we can clearly see the advantage of considering a group of objects that have a variable spatial relationship: this enables to consider models that need rigid parts, that can have an "internal movement" driven by a predefined set of rules (here just based on the distance between the ventricles).

(a) Axial cut 2   (b) Solution superimposed on (a)

(c) Zoom of the solution superimposed on the gradient of (b)

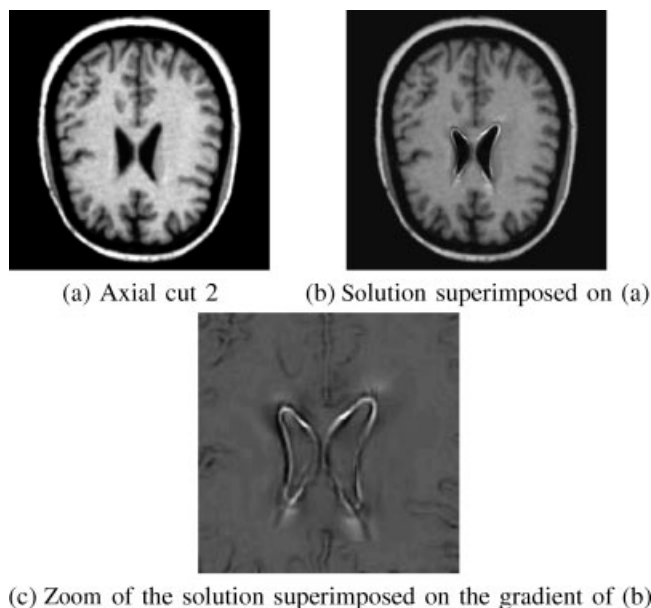**Figure 33.** Axial cut of a MR brain image and result of the separated extraction of left and right ventricles.

(a) Pair of ventricles at σ_{d_1} (b) Pair of ventricles at σ_{d_2}

Figure 31. Variations of the size and distance between both ventricles, given with different scaling factors $\sigma_x' = \sigma_y' = \sigma_{d_{1,2}}$.

To improve this example, we can think of adding one degree of freedom in the variable spatial relationship of the two ventricles, e.g., allowing a rotation of one ventricle with a center placed at the middle of the segment indicated in Figure 31, but this also increases the complexity of the process.

X. CONCLUSIONS

We have proposed a new shape descriptor on the basis of the use of matching pursuit as a shape analysis tool, and extended it by introducing a scale-space approach and an affine invariant dictionary. We first decompose the shape in its principal "shape vectors" (atoms), and then used this description, called meta-atom g_γ^M, as a new dictionary for shape recognition task (MPSR).

Because of the nature of the initial anisotropic refinement dictionary $\mathcal{D}$, the description and the detection of the object are invariant to similitude transformations. Introducing a shearing parameter leads us to pseudo-affine invariance for the new dictionary $\mathcal{D}^\mathcal{M} = \{g_\gamma^M\}$ composed by all variations of the meta-atom. Moreover, the atom definition allows to easily generalize the object definition to a scale-space representation, and gives an analytical definition of the object in the linear scale-space. We use it in the MPSR task, in a coarse-to-fine approach, and improve the robustness of our extraction process.

We showed accurate results for this method, with complex synthetic shapes. This method has also been assessed on real world (aerial and medical) images, to demonstrate its stability in more complex conditions, as well as its capability to handle multiple separated objects with a single model, and/or give a different weight to parts of the model.

REFERENCES

K. Arbter, W. Snyder, H. Burkhardt, and G. Hirzinger, Application of affine-invariant Fourier descriptors to recognition of 3-D objects, IEEE PAMI 12 (1990), 640–647.

J. Ben-Arie, K. Rao, and Z. Wang, Affine invariant shape representation and recognition using Gaussian kernels and multidimensional indexing, in IEEE ICASSP 6 (1996a), 3470–3473.

J. Ben-Arie, K. Rao, and Z. Wang, Iconic representation and recognition using affine-invariant spectral signatures, ARPA Image Understanding Workshop, Feb. 1996b.

F. Bergeaud and S. Mallat, Matching pursuit of images, Proc IEEE Int Conf Image Process, ICIP'95, 1 (1995), 53–56.

P. Bharadwaj, P. Runkle, and L. Carin, Target identification with wave-based matched pursuits and hidden Markov models, IEEE Trans Antennas Propag 47 (1999), 1543–1554.

P. Bharadwaj, P. Runkle, L. Carin, J. Berrie, and J. Hughes, Multiaspect classification of airborne targets via physics-based hmms and matching pursuits, IEEE Trans Aero Electron Syst 37 (2001), 595–606.

I. Biederman, Human image understanding: Recent research and theory, Comput Vis Graph Image Process 32 (1985), 29–73.

H. Blum, "A transformation for extracting new descriptors of shape," MIT Press models for the perception of speech and visual form, 1967, pp. 362–380.

H. Blum and R. Nagel, Shape description using weighted symmetric axis features, Pattern Recogn 10 (1978), 167–180.

C. Brechbühler, G. Gerig, and O. Kübler, Parametrization of closed surfaces for 3D shape description, Comput Vis Image Understand 61 (1995), 154–170.

L. Carin, G. Ybarra, P. Bharadwaj, and P. Runkle, Physics-based classification of targets in sar imagery using subaperture sequences, IEEE Proc Int Conf Acoust Speech Signal Process (ICASSP) 6 (1999), 3341–3344.

S. Chen, D. Donoho, and M. Saunders, Atomic decomposition by basis pursuit, SIAM J Sci Comput 20 (1998), 33–61.

R. Coifman and M. Wickerhauser, Entropy-based algorithms for best-basis selection, IEEE Trans Inform Theory 38 (1992), 713–718.

I. Daubechies, Time-frequency localization operators: A geometric phase space approach, IEEE Trans Inform Theory 34 (1988), 605–612.

G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," Journal of Constructive Approximation 13 (1997), 57–98.

M. Donahue, D. Geiger, R. Hummel, and T.-L. Liu, "Sparse representations for image decomposition with occlusions, IEEE Proc Int Conf Comput Vis Pattern Recogn (CVPR), 1996, pp. 7–12.

D. Donoho and X. Huo, Beamlets and multiscale image analysis, Tech. Rep. 2001-08, Department of Statistics, Stanford University, April 2001.

M. F. Demirci, A. Shokoufandeh, Y. Keselman, S. Dickinson, and L. Bretzner, Many-to-many matching of scale-space feature hierarchies using metric embedding, Scale-Space 2003, LNCS 2695, June 2003, pp. 17–32.

A. Ferreira da Silva, Atomic decomposition with evolutionary pursuit, Digit Signal Process 13 (2003), 317–337.

R. M. Figueras i Ventura and P. Vandergheynst, Matching pursuit through genetic algorithms, Tech. Rep. 01.02, ITS-EPFL, 2001.

R. M. Figueras i Ventura and P. Vandergheynst, Evolutionary multiresolution matching pursuit and its relations with the human visual system, Proc EUSIPCO 2002, Vol. 2, Toulouse, France, September 2002, pp. 395–398.

H. Freeman, Computer processing of line drawing images, Comput Surv 6 (1974), 57–98.

H. Freeman, Shape description via the use of critical points, Pattern Recogn 10 (1978), 159–166.

P. Frossard, Robust and multiresolution video delivery: From h.26x to matching pursuit based technologies, Ph.D. Dissertation, Electrical Department, Signal Processing Institute (ITS), Swiss Federal Institute of Technology (EPFL), 2000.

D. Geiger, T.-L. Liu, and M. Donahue, Sparse representations for image decompositions, Int J Comput Vis (IJCV) 33 (1999), 139–156.

G. Gerig, M. Styner, D. Jones, D. Weinberger, and J. Lieberman, Shape analysis of brain ventricles using SPHARM, IEEE Proc MMBIA 2001, 2001, pp. 171–178.

M. Gharavi-Alkhansari and T. Huang, A fast orthogonal matching pursuit algorithm, IEEE Proc Int Conf Acoust Speech Signal Process (ICASSP), 3 (1998), 1389–1392.

D. Hoffman and W. Richards, Parts of recognition, Tech. Rep. AIM-732, 1983.

S.H. Hsu and C.L. Huang, Road sign detection and recognition using matching pursuit method, Image Vis Comput 19 (2001), 119–129.

C.L. Huang and S.H. Hsu, Road sign interpretation using matching pursuit method, Proc IEEE Int Conf Pattern Recogn (ICPR) 1 (2000), 329–333.

L.K. Jones, On a conjecture of huber concerning the convergence of projection pursuit regression, Ann Stat 15 (1987), 880–882.

S. Jaggi, W. Karl, S. Mallat, and A. Willsky, High resolution pursuit for feature extraction, Appl Comput Harmonic Anal 5 (1998), 428–449.

S. Jaggi, W. Karl, S. Mallat, and A. Willsky, Silhouette recognition using high-resolution pursuit, Pattern Recogn 32 (1999), 753–771.

S. Jaggi, A. Willsky, W. Karl, and S. Mallat, Multiscale geometrical feature extraction and object recognition with wavelets and morphology, IEEE Proc Int Conf Image Process (ICIP) 3 (1995), 372–375.

H. Kauppinen, T. Seppanen, and M. Pietikainen, An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification, IEEE Trans Pattern Anal Mach Intell 17 (1995), 201–207.

V. Kindratenko, On using functions to describe the shape, J Math Imaging Vis 18 (2003), 225–245.

N. Krüger and G. Peters, Object recognition with banana wavelets, Proc Eur Symp Artificial Neural Networks (ESANN 1997), April 1997, pp. 61–66.

T. Lindeberg, Scale-space theory: A basic tool for analysing structures at different scales, J Appl Stat 21 (1994), 225–270.

T. Lindeberg, Edge detection and ridge detection with automatic scale selection, IEEE Proc Comput Vis Pattern Recogn, June 1996, pp. 465–470.

A. Lobo and P. Loizou, Voiced/unvoiced speech discrimination in noise using gabor atomic decomposition, IEEE Proc Int Conf Acoust Speech Signal Process (ICASSP) 1 (2003), 820–823.

S. Loncaric, A survey of shape analysis techniques, Pattern Recogn 31 (1998), 983–1001.

S. Mallat and Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Trans Signal Process 41 (1993), 3397–3415.

S. Mallat, A wavelet tour of signal processing, 2nd ed., Academic Press, San Diego, 1998.

B. Mehtre, M. Kankanhalli, and W. Lee, Shape measures for content based image retrieval: A comparison, Inform Process Manag 33 (1997), 319–337.

F. Mendels, P. Vandergheynst, and J.-P. Thiran, Rotation and scale invariant shape representation and recognition using matching pursuit, Proc IEEE Int Conf Pattern Recogn (ICPR), Quebec, 11–15 August, 2002, Vol. 4, pp. 326–329.

F. Mokhtarian, Silhouette-based isolated object recognition through curvature scale space, IEEE Trans Pattern Anal Mach Intell 17 (1995), 539–544.

F. Moschetti, L. Granai, P. Vandergheynst, and P. Frossard, New dictionary and fast atom searching method for matching pursuit representation of displaced frame difference, IEEE Proc Int Conf Image Process (ICIP2002) 3 (2002), 685–688.

Y. Pati, R. Rezaiifar, and P. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, IEEE Proc Asilomar Conf Signals Syst Comput 1 (1993), 40–44.

T. Pavlidis, Algorithms for shape analysis of contours and waveforms, IEEE Trans Pattern Anal Mach Intell 2 (1980), 301–312.

A. Pece, The problem of sparse image coding, J Math Imaging Vis 17 (2002), 89–108.

A. Pece and N. Petkov, Fast atomic decomposition by the inhibition method, IEEE Proc Int Conf Pattern Recogn (ICPR) 3 (2000), 211–214.

L. Peotta, L. Granai, and P. Vandergheynst, Very low bit rate image coding using redundant dictionaries, Proc SPIE 48th Annu Meet, San Diego, August 2003.

G. Peters and N. Krüger, Learning object representations by clustering banana wavelet responses, Proc 1st Int Workshop Stat Tech Pattern Recogn, (STIPR97), June 1997, pp. 113–118.

P. Phillips, Matching pursuit filter design, in IEEE Proc Int Conf Pattern Recogn (ICPR) 3 (1994), 57–61.

P. Phillips, Matching pursuit filters applied to face identification, IEEE Trans Image Process 7 (1998), 1150–1164.

R. Rao and D. Ballard, An active vision architecture based on iconic representations, Artif Intell 78(1/2) (1995), 461–505.

R. Rao and D. Ballard, Object indexing using an iconic sparse distributed memory, IEEE Int Conf Comput Vis, June 1995, pp. 24–31.

L. Rebollo-Neira and D. Lowe, Optimized orthogonal matching pursuit approach, IEEE Signal Process Lett 9 (2002), 137–140.

P. Runkle and L. Carin, Multiaspect target identification with wave-based matching pursuits and continuous hidden Markov models, IEEE Int Conf Acoust Speech Signal Process (ICASSP) 4 (1999), 2115–2118.

P. Runkle, L. Carin, L. Couchman, T. Yoder, and J. Bucaro, Multiaspect target identification with wave-based matched pursuits and continuous hidden Markov models, IEEE Trans Pattern Anal Mach Intell 21 (1999), 1371–1378.

M. Safar and C. Shahabi, Shape analysis and retrieval of multimedia objects, Kluwer, Academic Publishers, Portland, 2003, Multimedia systems and applications series, Vol. 23.

L. Shapiro, A structural model of shape, IEEE Trans Pattern Anal Mach Intell 2 (1980), 111–126.

M. Styner and G. Gerig, Hybrid boundary-medial shape description for biologically variable shapes, IEEE Proc MMBIA 2000, June 2000, pp. 235–242.

S. Ullman, Aligning pictorial descriptions: An approach to object recognition, Cognition 32 (1989), 193–254.

S. Ullman and R. Basri, Recognition by linear combinations of models, IEEE Trans Pattern Anal Mach Intell 13 (1991), 992–1006.

S. Ullman and S. Soloviev, Computation of pattern invariance in brain-like structures, Neural Networks 12 (1999), 1021–1036.

P. Vandergheynst and P. Frossard, Efficient image representation by anisotropic refinement in matching pursuit, Proc IEEE ICASSP'01, Salt Lake City, UT, May, 2001, Vol. 3, pp. 1757–1760.

G. van Tonder and Y. Ejima, The patchwork engine: Image segmentation from shape symmetries, Neural Networks 13 (2000), 291–303.

R. Veltkamp and M. Hagedoorn, State-of-the-art in shape matching, Tech. Rep Uu-CS-1999-27, Utrecht University, The Netherlands, 1999.

K. Wang and D. Goblirsch, Extracting dynamic features using the stochastic matching pursuit algorithm for speech event detection, IEEE Proc Workshop Automat Speech Recogn Understand, December 1997, pp. 132–139.

K. Wang, C.-H. Lee, and B.-H. Juang, Selective feature extraction via signal decomposition, IEEE Signal Process Lett 4 (1997), 8–11.

T. Wang, Y. Rui, and J.-G. Sun, Constraint based region matching for image retrieval, Int J Comput Vis (IJCV) 56(1/3) (2004), 37–45.

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

Y. Zhang and S. Ma, Unsupervised segmentation of noisy image in a multiscale framework, IEEE Proc Int Conf Signal Process (ICSP) 2 (2000), 905–909.

W. Zhao and N. Nandhakumar, Linear discriminant analysis of mpf for face recognition, IEEE Int Conf Pattern Recogn 1 (1998), 185–188.

S.-C. Zhu, C. Guo, Y. Wu, and Y. Wang, What are textons? Proc Eur Conf Comput Vis (ECCV) 4 (2002), 793–807.