Affine invariant Matching Pursuit-based shape representation and recognition using scale-space

François Mendels, Pierre Vandergheynst, and Jean-Philippe Thiran
Signal Processing Institute, Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
{francois.mendels, pierre.vandergheynst, jp.thiran}@epfl.ch

Abstract — In this paper, we propose an analytical low-level representation of images, obtained by a decomposition process, here the matching pursuit (MP) algorithm, as a new way of describing objects through a general continuous description using an affine invariant dictionary of basis functions. This description is used to recognize objects in images. In the learning phase, a template object is decomposed, and the extracted subset of basis functions, called meta-atom, gives the description of our object. We then extend naturally this description into the linear scale-space using the definition of our basis functions, and thus bringing a more general representation of our object.

We use this enhanced description as a predefined dictionary of the object to conduct an MP-based shape recognition (MPSR) task into the linear scale-space. The introduction of the scale-space approach improves the robustness of our method, and permits to avoid local minima problems encountered when minimizing a non-convex energy function. We show results for the detection of complex synthetic shapes, as well as natural (aerial and medical) images.

1 Introduction

Many ways exist in image processing for performing shape analysis and recognition tasks. Some uses a so-called "shape descriptor" (SD) or shape representation, depending if they give quantitative or qualitative measures. These methods have to be in both cases invariant to translation, rotation and scale, as these transformations do not change the intrinsic shape characterization of an object. The shape recognition process is then realized, based on its description, usually in a one-to-one correspondence framework. Different classifications were proposed [1, 2] : one distinguishes between local SD, such as nodes representations (e.g. with splines [3]), or template matching-like approaches (e.g.[4]), global SD, which can be based on a statistical approach (e.g. moments methods [5]), and methods linking both views (e.g. medial axis transform [6]).

In this work, we are interested in methods related to spatial sparse representation models, with both local and global features. We develop an approach whose general idea is close to the iconic representation (IC), first presented by Rao and Ballard in [7], lately improved by Ben-Arie et al. [8, 9] by reaching affine invariance property. In those works, the authors extract a localized affine invariant version of image patches using Gaussian kernels, and realize a one-to-one correspondence using Multi-Dimensional Indexing (MDI) in the spectral domain to conduct the recognition process.

Other authors have looked into local spatial representation to conduct recognition using local features: the banana wavelet decomposition scheme, introduced by [10, 11] consist in extracting landmarks from images, in order to describe common features in the chosen set of images. Finally, the landmarks obtained 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 matching pursuit (MP) decomposition [12, 13]. In this face recognition application, Phillips introduces matching pursuit filters as an adapted wavelet expansion build from a training set of the object [14]. 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.

Matching pursuit filters have been used by Huang and Hsu to perform direct road sign recognition [15, 16] 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 matching pursuit filter method based on a trained set of possible atoms combination.

Following the idea of describing an object with a set of vectors that have a spatial relationship, we introduce in this paper a new shape descriptor, using the matching pursuit algorithm as a decomposition process, and its natural extension to a linear scale-space.

The MP-based approach allows us to address the multiple target recognition and localization problem, and to consider both local and global features.

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

The method we introduce 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 $\mathcal{D}$. Therefore, 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 represent the template image $\mathcal{O}$. This group of atoms is interpreted as a meta-atom g_γ^M.

- 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 an affine invariant object dictionary by introducing a shearing parameter, in addition to a translation, isotropic dilation, and rotation parameters.

- we create an implicit extension of our object into a linear scale-space, using the scale parameter as an isotropic dilation of our basis functions.

- we then perform a 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 our object is used.

The paper structure is as follows: in Section 2, we introduce briefly MP and some important properties which present interest in both shape description and recognition processes. Section 3 presents the shape descriptor, and its extension to a linear scale-space follows in Section 4. The shape recognition task is then introduced in Section 5, and conducted into a linear scale-space. Section 6 presents results of the MP-based shape recognition process conducted on synthetic and natural (aerial and medical) images.

2 Describing images with Matching Pursuit

The goal of this section is to point out the basic concepts and the interesting properties of the Matching Pursuit (MP) process for the purpose of shape description, rather than describing it and its properties completely. We refer e.g. to [12, 13, 17, 18] for more details.

The matching pursuit algorithm, first introduced for monodimensional signals by Mallat and Zhang [13], is an iterative greedy process [19] 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_i} are basis functions (BF) of unit norm usually called *atoms* [13], 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 residue of n^{th} order $R^n f$, when projecting orthogonally $g_{\gamma_{n-1}}$ on $R^{n-1}f$. Thus we have, with $R^0 f = f$,

$$R^n f = R^{n-1} f - \langle R^{n-1} f, g_{\gamma_{n-1}} \rangle g_{\gamma_{n-1}}, \tag{1}$$

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

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:

$$R^N f = f - \sum_{n=0}^{N-1} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n}, \tag{2}$$

where $R^N f \to 0$ when $N \to \infty$ [20]. This corresponds to the decomposition process.

From Eq.(2), we easily deduce the reconstruction process which corresponds to the MP invertibility property [21]. We are able to reconstruct exactly f if $N \to \infty$:

$$f = \sum_{n=0}^{\infty} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n}. \tag{3}$$

As we use a dictionary composed by functions of unit norm, each decomposition step gives an atom $g_{\gamma_{n-1}}$ that is orthogonal to the corresponding residue $R^n f$ [12]. Thus, following Eq.(2), we obtain the energy conservation property of MP [12]:

$$\|f\|^2 = \sum_{n=0}^{N-1} |\langle R^n f, g_{\gamma_n}\rangle|^2 + \|R^N f\|^2. \tag{4}$$

By looking at Eq.(2), 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.

3 Object/shape description using MP

While describing the MP process, and following the iconic representation idea of analyzing shapes [8, 9], we already pointed out the framework for decomposing a shape in its principal *'shape vectors'*, which are our atoms. Note that 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 Basis Pursuit (BP) [22], can be used as well.

As previously said, Phillips has also introduced in [14] the notion of describing a model, called proto-object, using only a few 2D directional wavelets with Matching Pursuit Filters (MPF), 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.

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 [23]. 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. Vandergheynst and Frossard [18] have proven that anisotropic refinement atoms are more suitable to describe images and boundaries than e.g. the original Gabor wavelet dictionary initially proposed in [12]. Other families of atoms have been investigated as well, like e.g. banana wavelets (product of a curved Gaussian and a curved wave function) [10] or the combination of gaussian function with a triangular function [24], and show good results for the targeted applications (resp. face description and coding the displaced frame difference (dfd)). Authors interested in the sparse representation of a signal have also produced different dictionaries: for example, Donoho and Huo 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 [25]. Recently, Peotta et al. 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 [26].

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 [18] :

$$g_{\gamma_k}(x,y) = (4x^2 - 2)e^{-(x^2+y^2)}, \tag{5}$$

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

where $(\tilde{x},\tilde{y})$ 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.

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

$$\gamma_k \equiv \begin{bmatrix} p_{x_k} & p_{y_k} & \sigma_{x_k} & \sigma_{y_k} & \theta_k \end{bmatrix}. \tag{6}$$

In the proposed method, we describe our object o as a given subset of basis functions g_{γ_k} chosen during the 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. We now consider this subset of functions as a meta-atom g_γ^M, which we normalize. Thus we have :

$$o : g_\gamma^M = \frac{\sum_{k=0}^{K} \langle R^k f, g_{\gamma_k}\rangle g_{\gamma_k}}{\| \sum_{k=0}^{K} \langle R^k f, g_{\gamma_k}\rangle g_{\gamma_k} \|}, \tag{7}$$

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.

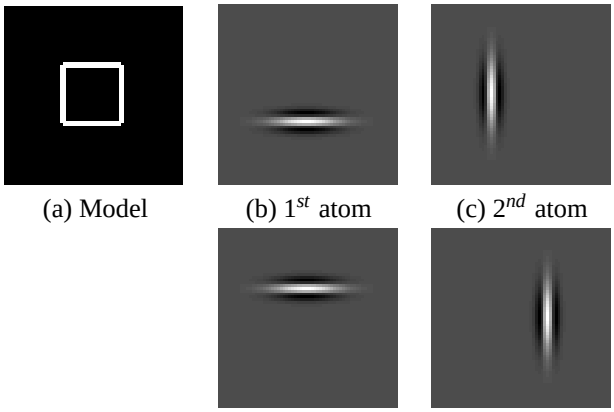
So o can be represented as a set of those vectors :

$$g_\gamma^M : \begin{bmatrix} C_{g_{\gamma_0}} & p_{x_0} & p_{y_0} & \sigma_{x_0} & \sigma_{y_0} & \theta_0 \\ & & \vdots & & & \\ C_{g_{\gamma_i}} & p_{x_i} & p_{y_i} & \sigma_{x_i} & \sigma_{y_i} & \theta_i \\ & & \vdots & & & \\ C_{g_{\gamma_K}} & p_{x_K} & p_{y_K} & \sigma_{x_K} & \sigma_{y_K} & \theta_K \end{bmatrix}. \tag{8}$$

Phillips, Hsu and Huang both use a full search procedure to obtain their atoms [14, 16], 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 [21, 27]. So we obtain a suboptimal solution at each step, but the overall convergence of MP in the reconstruction process (see Eq.(3)) ensures to get a complete description of our object.

3.1 Template Example

Fig.1(b) to (e) shows 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 illustrates that they accurately model straight edges.



(a) Model (b) 1^{st} atom (c) 2^{nd} atom (d) 3^{rd} atom (e) 4^{th} atom

Fig. 1. Original model and first 4 anisotropic refinement atoms

If we look at the reconstruction in Fig.2, 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 : Fig.2(d) to (f) shows 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.

3.2 Object dictionary

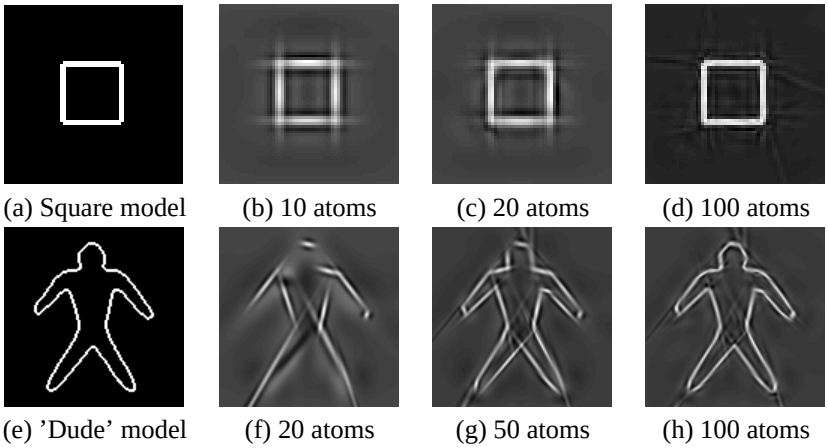
The extracted subset of basis functions g_γ^M, i.e. our object representation o, is invariant with respect to similitude transformations, due to the atom properties (see [18] for details). Here, we increase our range of variability by introducing a shearing parameter, thus leading to affine invariance. We obtain a new redundant dictionary $\mathcal{D}^{\mathcal{M}} = \{g_\gamma^M\}_{\gamma\in\Gamma'}$, where $\Gamma' = [p'_x, p'_y, \sigma'_x, \sigma'_y, \theta', S_m]$ is the set of possible indexes.

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:

$$g_{\gamma_k}(x', y') = (4x'^2 - 2)e^{-(x'^2+y'^2)}. \tag{9}$$

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

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



(a) Square model (b) 10 atoms (c) 20 atoms (d) 100 atoms

(e) 'Dude' model (f) 20 atoms (g) 50 atoms (h) 100 atoms

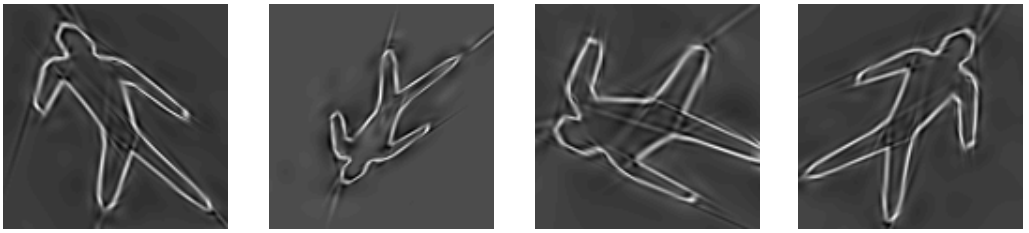
Fig. 2. Variations in the reconstruction using different number of atoms: (a) to (d) square model and its reconstructions, (e) to (h) 'dude' model and its reconstruction

$$\begin{cases} C'_{g_{\gamma_k}} & = & C_{g_{\gamma_k}} \\ \begin{bmatrix} p'_{x_k} \\ p'_{y_k} \end{bmatrix} & = & \begin{bmatrix} 1 & S_m \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma'_x & 0 \\ 0 & \sigma'_y \end{bmatrix} \begin{bmatrix} \cos(\theta') & sin(\theta') \\ -sin(\theta') & \cos(\theta') \end{bmatrix} \begin{bmatrix} p_{x_k} \\ p_{y_k} \end{bmatrix} + \begin{bmatrix} p'_x \\ p'_y \end{bmatrix} \\ \sigma'_{x_k} & = & \sigma_{x_k} \cdot \sigma'_x \\ \sigma'_{y_k} & = & \sigma_{y_k} \cdot \sigma'_y \\ \theta'_k & = & \theta_k - \theta' \end{cases} \tag{10}$$

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

We call the modified shape $g_\gamma^{M'} = g_{\gamma\ [p'_x,p'_y,\sigma'_x,\sigma'_y,\theta',S_m]}^{M}$, so we can express the original model O as $g_{\gamma\ origin}^{M} = g_{\gamma\ [0,0,1,1,0,0]}^{M}$.

We show in Fig.3 some examples of the "dude" model subject to given affine variations.



$[0,0,1,1,0,0.5]$ $[0,0,0.7,0.7,70,-0.5]$ $[0,0,1.1,1.1,40,0.5]$ $[0,0,1,1,-10,-0.25]$

Fig. 3. Examples of affine variations for the model "dude" using the following parameters $[p'_x, p'_y, \sigma'_x, \sigma'_y, \theta' * \pi/72, S_m]$

Variability of the MP-based object description As we will see in Section 6 with some examples, this shape description presents also a great range of variability in modelling different categories of objects. Thus, we can consider objects by characterizing not only outer boundaries, but also inner ones. This model also enables to include structural details of the object (e.g. the texture in the spatial domain, its orientation with respect to the object boundaries), and introduce them in the recognition process. We can also give a description for an object composed of different regions, these regions being connected or not. Moreover, this model does not restrict us to describe uniquely binary template objects : grayscale template objects may be modelled as well (see Fig.18(b) in Section 6.3). This implies that we are able to give a stronger weight to parts of the model arbitrarily chosen (see Fig.15(c) in Section 6.2 where more importance is given to the front of the boat).

This also addresses a common problem encountered in shape description, where the object is described by a non-closed curve: considering the boundaries as a sum of patches (our atoms) resolves implicitly this question.

Moreover, the direct extension of the sketch description given to describe the object leads to the possibility of representing a group of objects, while maintaining their spatial relation (e.g. see Fig.18(c) in Section 6.3).

4 MP-based shape representation in a linear scale-space

Using the MP-based shape model given in Eq.(7), we extend now our description to a linear scale-space, obtained by convolving our object with a Gaussian function G_σ, namely the normalized Gaussian kernel [28], given for a bidimensional signal by :

$$G_\sigma(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{\left[-\frac{x^2+y^2}{2\sigma^2}\right]}, (x,y) \in \mathbb{R}^2, \sigma \in \mathbb{R}^+, \sigma \neq 0. \tag{11}$$

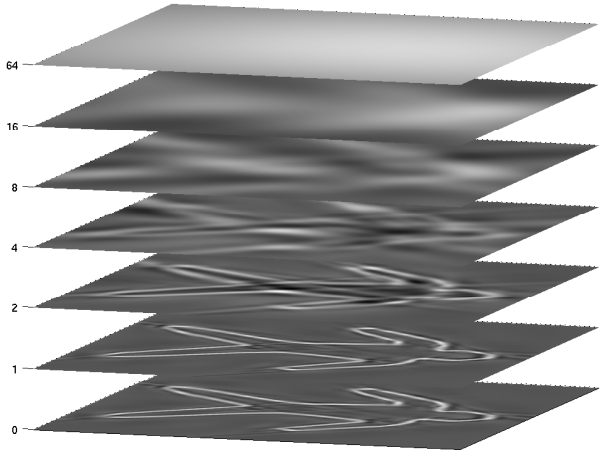
As our atoms satisfy the linear heat equation, their dilated version is given by the product of each scale parameter with σ for all atoms ($\sigma_{x_k,\sigma} = \sigma_{x_k}.\sigma$ and $\sigma_{y_k,\sigma} = \sigma_{y_k}.\sigma$). Thus the explicit convolution is not required to build a scale-space representation of our object.

By following this property, our object o modelled by g_γ^M can be expressed at scale σ as :

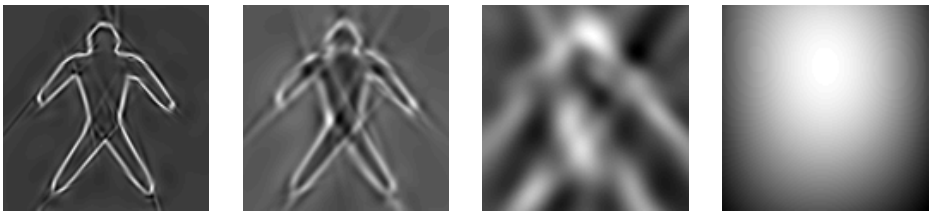
$$[g_\gamma^M]_\sigma = \sum_{k=0}^{K} \langle R^k f, g_{\gamma_k} \rangle g_{\gamma_k} * G_\sigma = \sum_{k=0}^{K} C_k g_{\gamma_k,\sigma}, \tag{12}$$

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

(a) Dude object and its description with 100 atoms



(b) Pyramidal representation of the scale-space generated with meta-atom dilation



$\sigma = 0$ $\sigma = 2$ $\sigma = 8$ $\sigma = 64$

(c) Planes of the "dude" scale-space with 100 atoms

Fig. 4. Linear Scale-space of the "dude" object o with $\sigma = 0..64$

Fig.4 illustrates the construction of a linear scale of our object "dude". The model at original scale is first shown (see (a)) and then its respective scale pyramid (see (b-c)), construct following Eq.(12). 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.

4.1 *Residual error in scale-space reconstruction*

Fig.5 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.

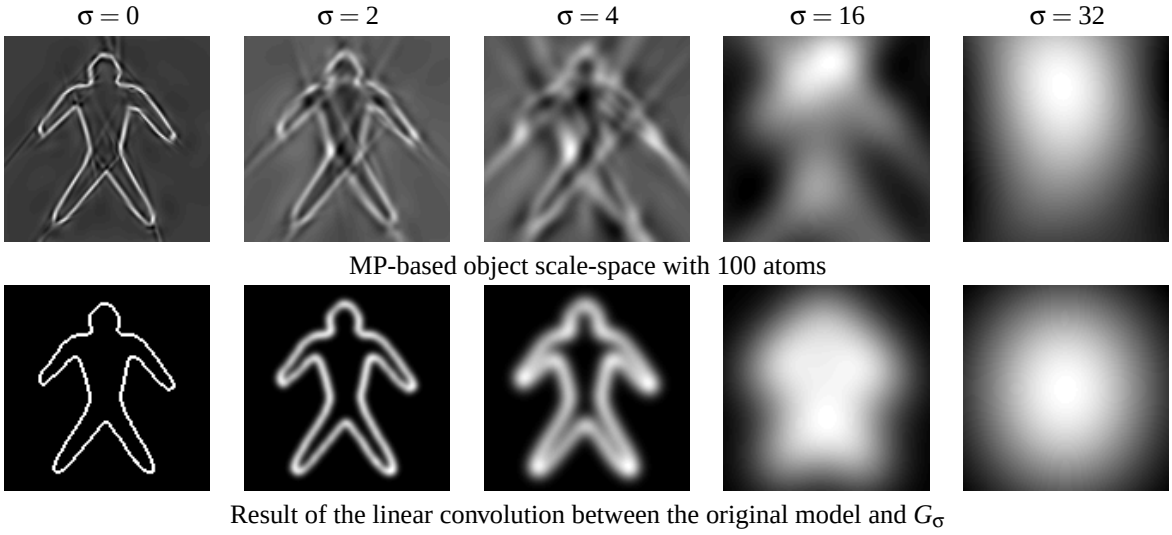


Fig. 5. Pyramidal scale-space of the MP-based object and its linear scale-space obtained by convolution

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 we need an infinite number of atoms to get a perfect reconstruction (see Eq.(3)). This confirms that we fit the model and its dilated version 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 influence over all scales. We have to minimize at step i the following system for each residue $R^i f_{\sigma_k}$ with $k = 0...K$:

$$\left\{\begin{array}{lcl} R^i f & = & f - \sum_{n=0}^{i-1} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} \\ R^i f_{\sigma_1} & = & f * G_{\sigma_1} - \sum_{n=0}^{i-1} \langle R^n f_{\sigma_1}, g_{\gamma_n,\sigma_1} \rangle g_{\gamma_n,\sigma_1} \\ R^i f_{\sigma_2} & = & f * G_{\sigma_2} - \sum_{n=0}^{i-1} \langle R^n f_{\sigma_2}, g_{\gamma_n,\sigma_2} \rangle g_{\gamma_n,\sigma_2} \\ & \vdots & \\ R^i f_{\sigma_K} & = & f * G_{\sigma_K} - \sum_{n=0}^{i-1} \langle R^n f_{\sigma_K}, g_{\gamma_n,\sigma_K} \rangle g_{\gamma_n,\sigma_K} \end{array}\right. \tag{13}$$

Fig.6 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 Fig.5, but it requires more atoms as well : the reconstruction in Fig.6 needed *n times the number of scale considered* compared to n atoms in Fig.5.

Moreover, by considering higher and lower scales in the minimization process, the choice of each atom may lead to induce perturbation between the scales, as we use the same atom dilated with respect to its scale, and that boundaries moves in a linear scale-space. This could lead to consider a non-linear 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.7).

But, as we will see later in Section 6, 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.

5 MP-based shape recognition using a linear scale-space

Let us recall the shape recognition process using an MP-based shape description (MPSR).

The recognition task follows directly the idea of describing the object 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_γ^M given by $\mathcal{D}^{\mathcal{M}} = \{[g_\gamma^M]_n = [g_\gamma^M]_{[p'_{x_n}, p'_{y_n}, \sigma'_{x_n}, \sigma'_{y_n}, \theta'_n, S_{m_n}]}\}$ for any translation

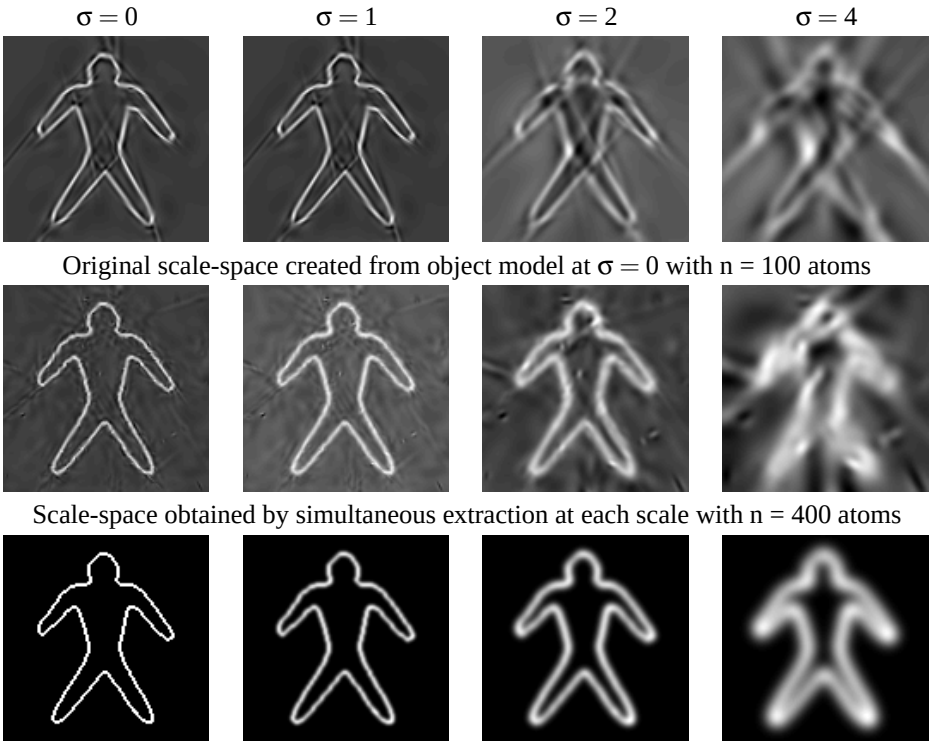


Fig. 6. Pyramidal scale-space of the MP-based object with 2 different solutions and its linear scale-space obtained by convolution

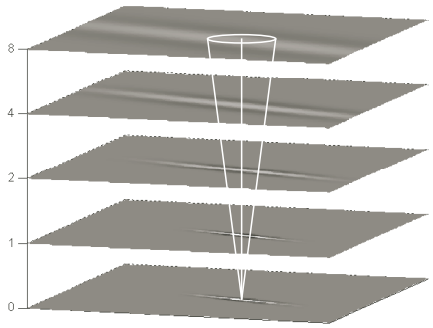


Fig. 7. Atom dilated over the scales ($\sigma = 0..8$) and the inverse pyramid showing possible variations in the position of the center that could be introduced

$[p'_{x_n}, p'_{y_n}]$, isotropic dilation $[\sigma'_{x_n} = \sigma'_{y_n}]$, rotation θ'_n , and shearing S_{m_n}. For the 1^{st} step, we have:

$$R^1\mathcal{T} = \mathcal{T} - \langle \mathcal{T}, [g_\gamma^M]_1 \rangle [g_\gamma^M]_1, \tag{14}$$

with $[g_\gamma^M]_1 = [g_\gamma^M]_{[p'_{x_1}, p'_{y_1}, \sigma'_{x_1}, \sigma'_{y_1}, \theta'_1, S_{m_1}]}$.

Following MP principles, the best match (here $[g_\gamma^M]_1$) will give the object location in $\mathcal{T}$.

In case of multiple solutions, we re-iterate the MP algorithm using $[g_\gamma^M]_n$, until no match is found considering a minimal error threshold fixing the tolerance of the difference between the object detected and the model (here 90% of the ideal case).

We obtain for the step N:

$$R^N\mathcal{T} = \mathcal{T} - \sum_{n=0}^{N-1} \langle R^n\mathcal{T}, [g_\gamma^M]_n \rangle [g_\gamma^M]_n, \tag{15}$$

where each $[g_\gamma^M]$ is a solution of the minimization problem (see Eq.(15)) for the template $\mathcal{O}$.

5.1 *Square Template Example*

Using the square template model previously introduced and described (see Fig.1-2), we perform the shape recognition task on the following image (see Fig.8(a)).

(a) Target 1 $\mathcal{T}_1$ (b) First solution $[g^M_{\gamma_k}]_1$ (c) second solution $[g^M_{\gamma_k}]_2$ (d) Final solution

Fig. 8. Target 1 and the 2 objects extracted (128×128 image)

We have two solutions for the square, one identical to the model (see Fig.1(a)), and the other one rotated by 45^o. For the minimization process (see Eq.(15)), the parameters of the Genetic algorithm (GA) are set to 27 chromosomes and 50 generations for each step of MP decomposition, and will remain the same for all the following experiments.

5.2 *Extension to a linear scale-space*

The tolerance threshold introduced permits to consider solutions that have local variations, but it may be to too sensitive, or we may extract partial solutions that correspond to a local minima of the energy function.

Our model handles affine variations (see Fig.3), but bigger perturbations (see Fig.9) may lead to false solutions.

Fig. 9. Local variations in ”dude” model

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 conduct to an incomplete solution: Fig.10(c) shows a false match where one leg of the model is captured correctly.

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

Fig. 10. Target image and the false match extracted

In order to increase the robustness of our process, we will use the linear scale-space extension of our model $[g^M_\gamma]_\sigma$, and thus realize the Matching Pursuit-based Shape Recognition (MPSR) task into a linear scale-space of our target image $\mathcal{T}$.

5.2.1 *MP decomposition at scale* σ

We now describe how to realize a MP decomposition at a scale bigger than the original scale ($\sigma = 0$).

The decomposition process described in Eq.(14) for the original scale of the object, gives at the first step the object $[g_\gamma^M]_1$ where $[g_\gamma^M]_1 = \sum_{k=0}^{K} C_{k_1} g_{\gamma_{k_1}}$.

Performing this MP decomposition at a different scale $\sigma > 0$ implies that our target image $\mathcal{T}$ becomes $\mathcal{T} * G_\sigma$, and that our object $[g_\gamma^M]_n$ becomes $[g_\gamma^M]_{n,\sigma}$.

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

$$R^1\mathcal{T}_\sigma = \mathcal{T} * G_\sigma - \langle \mathcal{T} * g_\sigma, \sum_{k=0}^{K} C_{k_1} g_{\gamma_{k_1},\sigma} \rangle \sum_{k=0}^{K} C_{k_1} g_{\gamma_{k_1},\sigma}. \tag{16}$$

We obtain the following objects present in the target image by re-iterating the MP decomposition at scale σ. For the second object $[g_\gamma^M]_2$, we have to minimize $R^2\mathcal{T}_\sigma$:

$$R^2\mathcal{T}_\sigma = R^1\mathcal{T}_\sigma - \langle R^1\mathcal{T}_\sigma, \sum_{k=0}^{K} C_{k_2} g_{\gamma_{k_2},\sigma} \rangle \sum_{k=0}^{K} C_{k_2} g_{\gamma_{k_2},\sigma}. \tag{17}$$

If N objects are present in $\mathcal{T}$, we obtain for the N^{th} object:

$$R^N\mathcal{T}_\sigma = R^{N-1}\mathcal{T}_\sigma - \langle R^{N-1}\mathcal{T}_\sigma, \sum_{k=0}^{K} C_{k_N} g_{\gamma_{k_N},\sigma} \rangle \sum_{k=0}^{K} C_{k_N} g_{\gamma_{k_N},\sigma}. \tag{18}$$

5.2.2 Performing shape recognition through multi-scale MP

As we are able to decompose an image $\mathcal{T}$ into our object basis at any given scale σ, we now drive through the scale to realize the multi-scale recognition task. As shown in Fig.11, the extraction of a possible object solution begins at a coarser scale (here represented by σ_2), and each solution is then propagated iteratively to finer scales (σ_1, σ_0), until we reach the object scale ($\sigma = 0$).

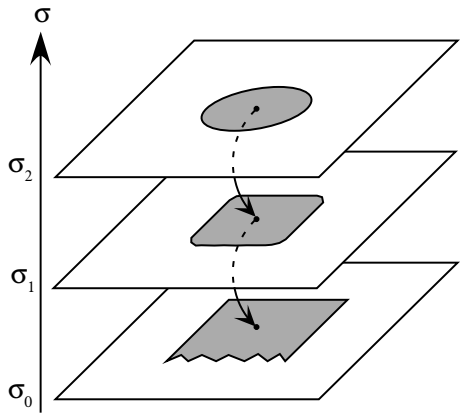


Fig. 11. Scale-space recognition of the object $\mathcal{O}$ in a target $\mathcal{T}$

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.

6 Results

6.1 Synthetic images

Using the target $\mathcal{T}$ presented in Fig.10(a), we conduct the multi-scale approach building each tree following each object. Fig.12 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 the scale, up to the original object scale.

Fig.13 shows the final result of the extraction, obtained by summing each step of the decomposition. Both objects are correctly extracted.

In Fig.14, we realize another detection test on a synthetic image build with two objects presenting bigger variations in the dude model (see Fig.9) and other different objects, such as dog and fish shapes. We show here

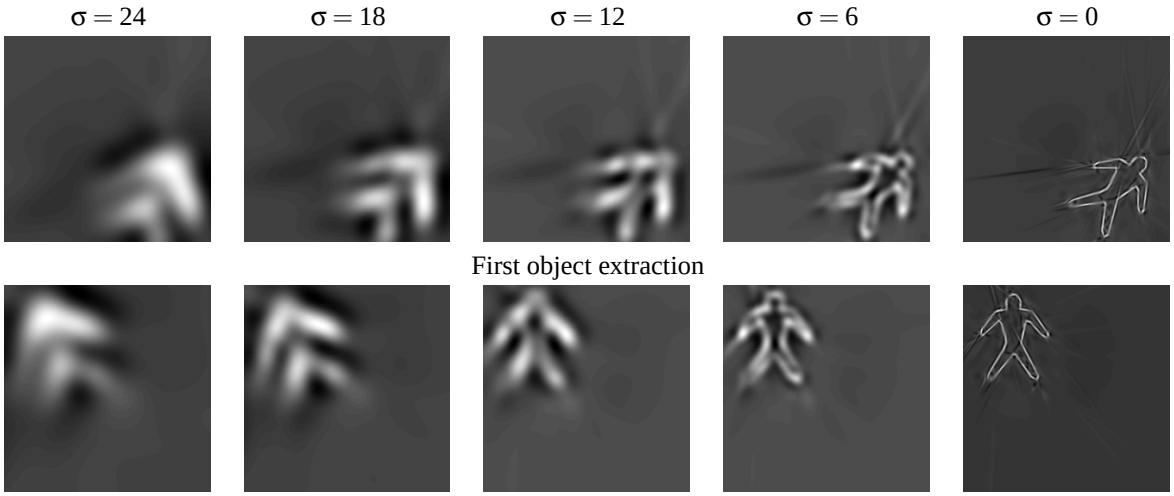


Fig. 12. First and second pyramidal object extraction from $\mathcal{T}$

(a) Original Target (b) Objects extracted at $\sigma = 0$

Fig. 13. Original target $\mathcal{T}$ and the result of the extraction

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). Both objects corresponding to our model are correctly found, with respect to their local deformations, which are not matched as we do not allow non-rigid deformation in our model. Matched parts are shown in white (see Fig.14(c)).

(a) Original Target (b) Objects extracted at $\sigma = 0$ (c) superimposition of the solution and the original target

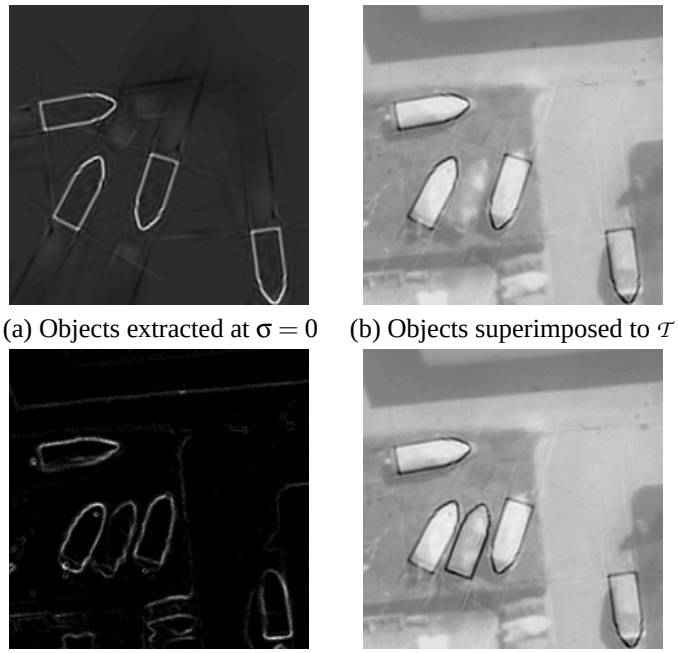
Fig. 14. Original target $\mathcal{T}$, the result of the extraction and the superimposition of (a) and (b) where the corresponding parts are shown in white (c)

6.2 Natural images

The following tests are conducted on an aerial image containing several boats (see Fig.15(a)). The model to be extracted is a boat (see Fig.15(c)) modelled by our MP description (see Eq.(7)).

(a) Original Target $\mathcal{T}$ (b) gradient of $\mathcal{T}$ (c) Boat model at $\sigma = 0$

Fig. 15. Original target $\mathcal{T}$ and the boat model used for the extraction process



(a) Objects extracted at $\sigma = 0$ (b) Objects superimposed to $\mathcal{T}$

(c) Enhanced gradient of $\mathcal{T}$ (d) Objects superimposed to $\mathcal{T}$

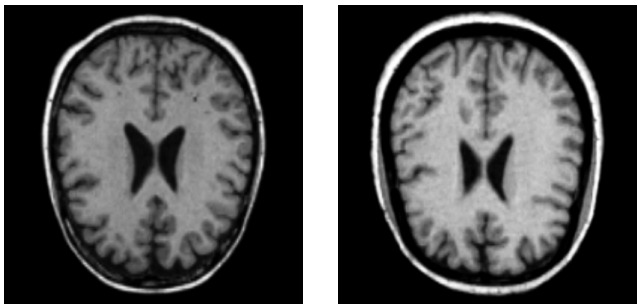
Fig. 16. Original boat target $\mathcal{T}$ and the result of the extraction, using normalized gradient (a-b), and enhanced gradient (c-d)

The detected objects are presented in Fig.16. The boat in the middle of the group of 3 boats is not found due to the lack of strength of its boundaries (see Fig.15(b)). Correct segmentation is realized in most cases, even though 2 solutions are validated with a wrong orientation : this shows a possible problem that occurs when the object searched presents strong symmetries. By enhancing the gradient (see Fig.16(c)), we realize the detection of the boat we did not found using the normalized gradient (see Fig.16(d)).

6.3 Medical images

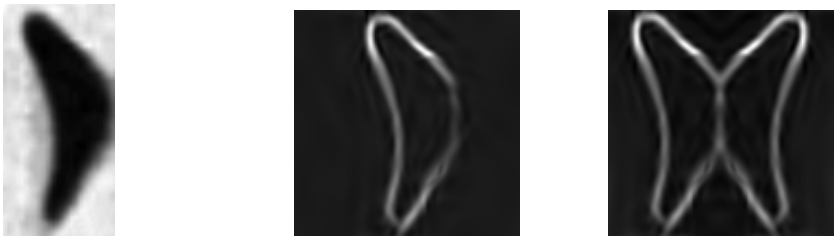
Here, we consider two axial slices of a Magnetic Resonance (MR) image of the brain (see Fig.17). We extract the left ventricle (LV) from the axial cut 1 (see Fig.18(a)), and then construct a model with $n = 100$ atoms (see Fig.18(b)) based on the gradient of the LV. We generate a new dictionary for the 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 $2.n$ atoms, while maintaining their intrinsic spatial relationship (see Fig.18(c)). This relationship is inversely linked to the scale of the ventricles : as the ventricles have a bigger scale, the distance between them reduces.

The extraction is conducted for both ventricles simultaneously in the axial cut 2 (see Fig.19). The solution, highlighted in white, is shown twice: firstly, we superimpose it to the axial cut (see Fig.19(b)), and secondly, we show in Fig.19(c) the zoom of the solution, thus presenting its correspondence with the gradient of the axial cut 2



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

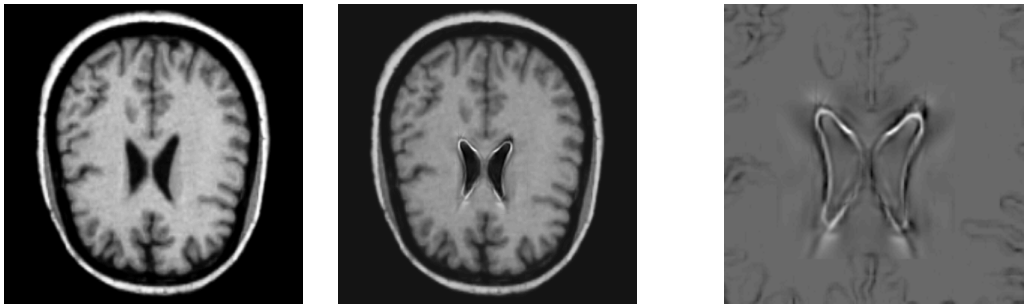
Fig. 17. Original axial cut for the model of the left ventricle (LV) of the human brain and the targeted axial cut for ventricle extraction



(a) LV extracted from Axial cut 1 (b) LV modelled at $\sigma = 0$, $n = 100$ atoms (c) both ventricles at $\sigma = 0$

Fig. 18. Original model O for the left ventricle (LV) of the human brain and its extension to both ventricles

(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.



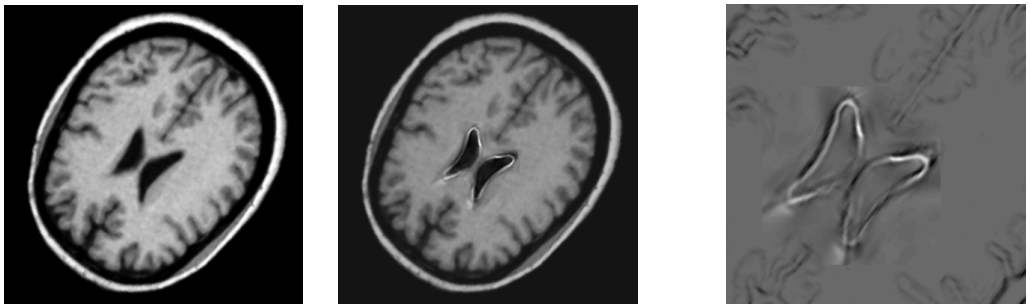
(a) Axial cut (b) superimposition of the solution and (a) (c) zoom of the superimposition of the solution and the gradient of (b)

Fig. 19. Axial cut of a MR brain image and extraction of both ventricles

Fig.20 shows a ventricles extraction for the axial cut 2 submitted to an arbitrary rotation, thus confirming the robustness of our method to changes in orientation. The solution is presented in the same way as before, and highlighted in white. We note also here the small shift on the location of the RV. If we want to achieve a better match, we have to extract both ventricle separately: in this case, 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.21): the solution is shown in white in both Fig.21(b) and (c). These two examples confirms the ability to represent simultaneously multiple objects (see Section 3.2) using this model, as well as its accuracy in different conditions.

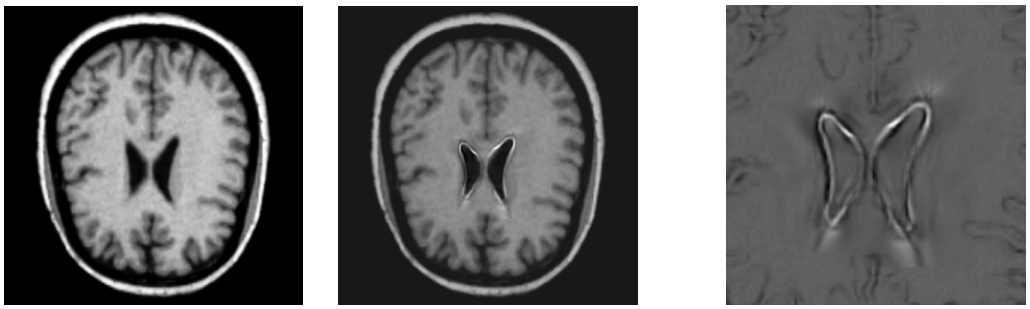
7 Conclusions

We have proposed a new shape descriptor based on 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



(a) Axial cut rotated (b) superimposition of the solution (white) and (a) (c) zoom of the superimposition of the solution (black) and the gradient of (b)

Fig. 20. Axial cut of a MR brain image and extraction of both ventricles



(a) Axial cut (b) superimposition of the solution (white) and (a) (c) zoom of the superimposition of the solution (black) and the gradient of (b)

Fig. 21. Axial cut of a MR brain image and result of the separated extraction of left and right ventricles

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).

Due to 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 lead us to 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 our 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 natural (aerial and medical) images, to demonstrate its stability in more complex conditions.

References

[1] T. Pavlidis. Algorithms for shape analysis of contours and waveforms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2.4:301–312, July 1980.

[2] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31.8:983–1001, August 1998.

[3] Fabian W. Meier, Guido M. Schuster, and Aggelos K. Katsaggelos. A mathematical model for shape coding with B-splines. *Signal Processing: Image Communication*, 15(7-8):685–701, May 2000.

[4] K.R. Rao and J. Ben-Arie. Multiple template matching using the expansion filter. *IEEE Trans. on Circuits and Systems for Video Technology*, 4.5:490–503, October 1994.

[5] C. Kan and M.D. Srinath. Invariant character recognition with Zernike and orthogonal Fourier-Mellin moments. *Pattern Recognition*, 35:143–154, 2002.

[6] H. Blum and R. Nagel. Shape description using weighted symmetric axis features. *Pattern Recognition*, 10.3:167–180, 1978.

[7] R.P.N. Rao and D.H. Ballard. An active vision architecture based on iconic representations. *Artificial Intelligence*, 78.1-2:461–505, October 1995.

[8] Jezekiel Ben-Arie, K.R. Rao, and Z. Wang. Affine invariant shape representation and recognition using gaussian kernels and multi-dimensional indexing. In *IEEE ICASSP*, volume 6, pages 3470–3473, May 1996.

[9] J. Ben-Arie, K.R. Rao, and Z. Wang. Iconic representation and recognition using affine-invariant spectral signatures. In *ARPA Image Understanding Workshop*, Feb. 1996.

[10] N. Krüger and G. Peters. Object recognition with banana wavelets. In *Proc. European Symposium on Artificial Neural Networks (ESANN'1997)*, pages 61–66, April 1997.

[11] G. Peters and N. Krüger. Learning object representations by clustering banana wavelet responses. In *Proceedings of the 1st International Workshop on Statistical Techniques in Pattern Recognition, (STIPR97)*, pages 113–118, June 1997.

[12] F. Bergeaud and S. Mallat. Matching pursuit of images. In *Proc. IEEE International Conference on Image Processing ICIP'95*, volume 1, pages 53–56, October 1995.

[13] S. Mallat and Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 41(12):3397–3415, December 1993.

[14] P.J. Phillips. Matching pursuit filters applied to face identification. *IEEE Trans. on Image Processing*, 7.8: 1150–1164, August 1998.

[15] C. L. Huang and S. H. Hsu. Road sign interpretation using matching pursuit method. In *Proc. IEEE International Conference on Pattern Recognition (ICPR)*, volume 1, pages 329–333, September 2000.

[16] S. H. Hsu and C. L. Huang. Road sign detection and recognition using matching pursuit method. *Image and Vision Computing*, 19.3:119–129, February 2001.

[17] S. Mallat. *A wavelet tour of signal processing*. Academic Press, USA, 1998.

[18] P. Vandergheynst and P. Frossard. Efficient image representation by anisotropic refinement in matching pursuit. In *Proc. of IEEE ICASSP'01*, volume 3, pages 1757–1760, Salt Lake City UT, May 2001.

[19] Geoffrey Davis, Stéphane Mallat, and Marco Avellaneda. Adaptive greedy approximations. *Constructive Approximations*, 1997. Springer-Verlag NY.

[20] Jones L.K. On a conjecture of huber concerning the convergence of projection pursuit regression. *The Annals of Statistics*, 15(2):880–882, 1987.

[21] Rosa M. Figueras i Ventura and Pierre Vandergheynst. Evolutionary multiresolution matching pursuit and its relations with the human visual system. In *Proceedings of EUSIPCO 2002*, volume 2, pages 395–398, Toulouse, France, September 2002.

[22] S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20.1:33–61, 1998.

[23] F. Mendels, P. Vandergheynst, and J.-P. Thiran. Rotation and scale invariant shape representation and recognition using matching pursuit. In *Proc. IEEE Int. Conf. on Pattern Recognition (ICPR) 2002*, pages 326–329, Qubec, August, 11-15 2002.

[24] F. Moschetti, Granai L., P. Vandergheynst, and P. Frossard. New dictionary and fast atom searching method for matching pursuit representation of displaced frame difference. In *IEEE Proc. International Conference on Image Processing (ICIP2002)*, volume 3, pages 685–688, 2002.

[25] D.L. Donoho and X. Huo. Beamlets and multiscale image analysis. Technical Report 2001-08, Stanford University, Department of Statistics, April 2001.

[26] L. Peotta, L. Granai, and P. Vandergheynst. Very low bit rate image coding using redundant dictionaries. In *Proceedings of SPIE 48th Annual Meeting*, August 2003.

[27] Rosa M. Figueras i Ventura and Pierre Vandergheynst. Matching pursuit through genetic algorithms. Technical Report 01.02, ITS-EPFL, 2001.

[28] T. Lindeberg. Scale-space theory : A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21.2:225–270, 1994.