Rotation and scale invariant shape representation and recognition using Matching Pursuit

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

Using a low-level representation of images, like matching pursuit, we introduce a new way of describing objects through a general description using a translation, rotation, and isotropic scale invariant dictionary of basis functions.

This description is then used as a predefined dictionary of the object to conduct a shape recognition task. We show some promising results for both parts of description and detection with simple shapes.

1. Introduction

Many ways exist in image processing for segmenting shapes and performing shape recognition tasks. Some of them use a statistical approach (e.g. moments methods [7]), nodes representations (e.g. with splines [10]). Here, we develop an approach which is close to the iconic representation [2], improved by Ben-Arie [1].

We introduce a new shape descriptor, using matching pursuit (MP) algorithm [3][9] as a decomposition process. Let us consider a redundant dictionary $\mathcal{D}$ of parametric basis functions that ensures a possible perfect reconstruction of the object image. These function parameters represent the shape vectors of the object.

The method we introduce has the following steps :

- from a template image of the object 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, deviation and amplitude of the basis functions that best represent the template image $\mathcal{O}$.
- using this template description, we generate a new dictionary $\mathcal{O}_n$ by varying the parameters of the previous extracted subset of basis functions, while keeping their intrinsic relationship.
- we then perform a MP decomposition of the target image $\mathcal{T}$ (where we want to find the object) using this predefined dictionary $\mathcal{O}_n$. Setting up a minimal error threshold, we find as many objects as $\mathcal{T}$ contains that are covariant with the template object.

The paper structure is as follows: in section 2, we will introduce briefly MP process and the important points that showed up here. Section 3 will describe the new shape descriptor we introduce, and the following shape recognition task will be described in section 4.

2. Describing images with Matching Pursuit

The matching pursuit algorithm, first introduced for monodimensional signals by Mallat & Zhang [9], is an iterative greedy [4] process that decomposes a function f in a Hilbert space $\mathcal{H}$, using a redundant dictionary $\mathcal{D} = \{g_\gamma\}_{\gamma\in\Gamma}$ of functions g_{γ_i} usually called *atoms*.

The goal of this section is not to describe MP in a complete manner, but rather to point out the basic concepts and the interesting properties of MP for the purpose of shape description. We refer the reader e.g. to [3][8][9][11] for further details.

Each step n of the algorithm consists in a projection of an atom $g_{\gamma_{n-1}}$ on f where the residue of n^{th} order $R^n f$ has to be minimized. Thus we have, with $R^0 f = f$,

$$R^1 f = f - \langle f, g_{\gamma_0}\rangle g_{\gamma_0}, \tag{1}$$

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

When $R^{i+1} f$ is minimized for a given g_{γ_i}, the projection between the previous residue and the actual atom $\langle R^i f, g_{\gamma_i}\rangle$ 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 $N \to \infty$ [9]. This describes the decomposition process.

From eq.2, we easily deduce the reconstruction process which corresponds to the MP invertibility property (e.g. [5]). We are able to reconstruct exactly f if $N \rightarrow \infty$:

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

By looking at the energy conservation property of MP [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.

3. Object/shape description using MP

While describing the MP process, and following the iconic representation idea of analyzing shapes [2], we already pointed out the framework for decomposing a shape in its principal *'shape vectors'*, which are our atoms.

Here, we make the choice of describing the template image of the object by characterizing its boundaries. The second choice concerns the dictionary taken to analyze the image: it has a large influence on the number of atoms needed to describe the shape accurately. Vandergheynst & Frossard [11] 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 [3].

Following this idea, we will use here these anisotropic refinement atoms [11]:

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

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

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 deviation.

The following parameters completely defined g_{γ_k} :

$$g_{\gamma_k} : \begin{bmatrix} p_{x_k} & p_{y_k} & \sigma_{x_k} & \sigma_{y_k} & \theta_k \end{bmatrix}. \quad (5)$$

This set of basis functions is invariant under translation, rotation and isotropic scaling (see eq.8 to 10 in [11] for details). Moreover, the general form of the atoms seems to fit well with the boundary behavior: they can reproduce the abrupt change in grayscale of a step edge (e.g. Fig.1 (b) to (e)).

Our object $\mathcal{O}$ is described as a reconstruction of a given subset of basis functions g_{γ_k} given by the MP process. Thus we have:

$$\mathcal{O} \cong \sum_{k=0}^{K} \langle R^k f, g_{\gamma_k} \rangle g_{\gamma_k}, \quad (6)$$

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 object.

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

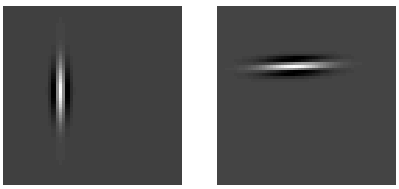
$$\mathcal{O} : \begin{bmatrix} C_{g_{\gamma_0}} & p_{x_0} & p_{y_0} & \sigma_{x_0} & \sigma_{y_0} & \theta_0 \\ & & \vdots & & & \\ C_{g_{\gamma_i}} & p_{x_i} & p_{y_i} & \sigma_{x_i} & \sigma_{y_i} & \theta_i \\ & & \vdots & & & \\ C_{g_{\gamma_K}} & p_{x_K} & p_{y_K} & \sigma_{x_K} & \sigma_{y_K} & \theta_K \end{bmatrix}. \quad (7)$$

We use 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 [5][6]. So we will obtain a suboptimal solution at each step, but the overall convergence of MP ensures to get a complete description of our object.

3.1. Template Example

Fig.1 presents the resulting first four atoms of the decomposition of the model of a square shape (64×64 image) using $\mathcal{D}$. This confirm our intuitive idea of removing the biggest energy part of the boundaries, here contained in each side of the square.

(a) Model (b) 1^{st} atom (c) 2^{nd} atom



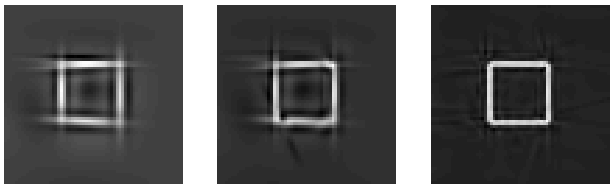
(d) 3^{rd} atom (e) 4^{th} atom

Figure 1. Original model and first 4 anisotropic refinement atoms

By looking 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 precision of the shape description, which relates the sensitivity of our model, will depend on the number of atoms we consider.

3.2. Object dictionary

We can reproduce the analyzed shape for any translation $[p'_x, p'_y]$, any isotropic scaling $\sigma'_x = \sigma'_y$, and any deviation



(a) 10 atoms (b) 20 atoms (c) 100 atoms

Figure 2. Reconstruction using 10, 20 and 100 anisotropic refinement atoms

θ' just by modifying directly the matrix coefficients as :

$$\begin{bmatrix} C'_{g_{\gamma_i}} \\ p'_{x_i} \\ p'_{y_i} \\ \sigma'_{x_i} \\ \sigma'_{y_i} \\ \theta'_i \end{bmatrix} = \begin{bmatrix} C_{g_{\gamma_i}} \\ \sigma_{x_i}(\cos(\theta_i)p_{x_i} + \sin(\theta_i)p_{y_i}) + p'_x \\ \sigma_{y_i}(-\sin(\theta_i)p_{x_i} + \cos(\theta_i)p_{y_i}) + p'_y \\ \sigma_{x_i} \cdot \sigma'_x \\ \sigma_{y_i} \cdot \sigma'_y \\ \theta_i - \theta' \end{bmatrix}, \quad (8)$$

with $i = 0...K$.

We call the modified shape $O' = \mathcal{O}_{[p'_x,p'_y,\sigma'_x,\sigma'_y,\theta']}$, so we can express $\mathcal{O}$ as $O_{origin} = \mathcal{O}_{[0,0,1,1,0]}$.

4. Shape recognition using MP description

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 $\mathcal{O}_n = \{O_n = \mathcal{O}_{[p'_{x_n},p'_{y_n},\sigma'_{x_n},\sigma'_{y_n},\theta'_n]}\}$ giving $\mathcal{O}$ for any translation, isotropic dilation, and rotation. So we have :

$$R^1\mathcal{T} = \mathcal{T} - \langle \mathcal{T}, O_1 \rangle O_1, \quad (9)$$

with $O_1 = \mathcal{O}_{[p'_{x_1},p'_{y_1},\sigma'_{x_1},\sigma'_{y_1},\theta'_1]}$.

Following MP principles, the best match (here O_1 for the 1^{st} step) will give the object location in $\mathcal{T}$. In order to avoid false match (e.g. in the case of partial object that could be seen as a stable local minimum by the genetic algorithm), we setup a minimal error threshold.

In the following examples, this threshold will be fixed to 90% of the ideal case (exact match between the object and the found solution): as we are dealing with non-perfect description of the object (finite number of g_{γ_k}), we cannot achieve the ideal case.

In the case of multiple solutions, we will re-iterate the MP algorithm using $\mathcal{O}_n$, until no match are found considering the minimal error we choose.

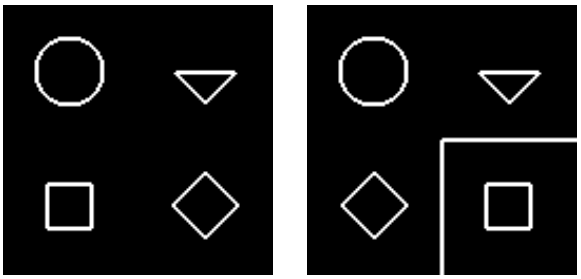
We obtain for the step N:

$$R^N\mathcal{T} = \mathcal{T} - \sum_{n=0}^{N-1} \langle R^n\mathcal{T}, O_n \rangle O_n, \quad (10)$$

where each O_n is a solution for the template $\mathcal{O}$.

4.1. Experimental results

Using the template model previously introduced and described (see Fig.1-2), we will now perform the shape recognition task on the two following images (see Fig.3).



(a) Target 1 $\mathcal{T}_1$ (b) Target 2 $\mathcal{T}_2$

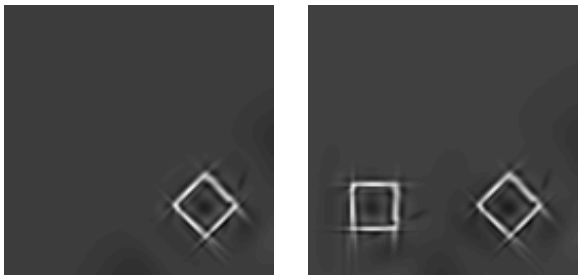
Figure 3. Target 1 and 2 (128×128 image) examples

As we can see, in the two target images, we have two solutions for the square, one identical to the model, and the other one rotated by 45^o. Parameters of the GA are set to 27 chromosomes and 50 generations for each step of MP decomposition. The result of the first step of the MP algorithm (see eq.9) for 10 and 20 atoms in the description of the object shape is shown in Fig.4.

(a) 10 atoms (b) 20 atoms

Figure 4. First solution for the square in $\mathcal{T}_1$, using a different number of atoms in the description of the object shape

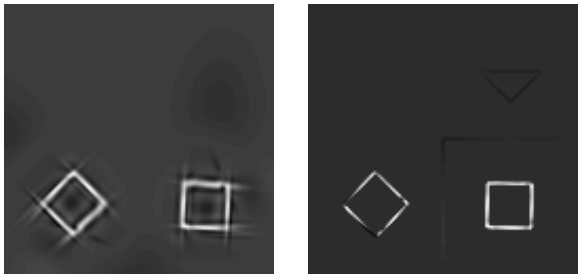
We clearly see here that, as long as the shape is sufficiently described, the number of atoms doesn't influence the result we get, even though the resolution of the object improves with the number of atoms. Although the other rotated square could have been found as a first solution, this doesn't change the final result (see Fig.5): the second square is detected at the second iteration of the algorithm (see eq.10). The detection works fine also for the $\mathcal{T}_2$ case



(a) 2^{nd} solution (b) Final result

Figure 5. Second solution, and final result for the localization of the square in $\mathcal{T}_1$, with 20 atoms

(see Fig.6(a)), even if we obtain approximate borders due to the nature of the uncomplete description of the object. In order to avoid such inaccuracy in the boundaries definition, one could easily recombine the initial target with the found solution, by, e.g., doing a point-to-point multiplication (see Fig.6(b) for $\mathcal{T}_2$).



(a) Final result (b) Combination

Figure 6. Final result with 20 atoms for the localization of the square in target 2 and recombination with the initial $\mathcal{T}_2$ image

In the combination figure, we can see that the rotated solution for the square does not fit completely the initial target. This leads to two remarks :

- the simple recombination may be not sufficient in this case: the setup of a threshold for the minimal acceptable error induces a small variance in the final location of the shape.
- one could also use this result as a starting point for more high-level segmentations procedures.

5. Conclusions

We have introduced a new shape descriptor based on the use of matching pursuit as a shape analysis tool. We first decompose the shape in its principal atoms, and then used this description as a new dictionary for shape recognition task. Due to the nature of the initial anisotropic refinement dictionary D, we are invariant under translation, rotation and isotropic scaling for the description and for the detection of the object.

We also showed accurate promising results for this method, with simple shapes. This method has now to be tested on natural images, to assess its stability in more complex conditions.

References

[1] J. Ben-Arie, K. Rao, and Z. Wang. Affine invariant shape representation and recognition using gaussian kernels and multi-dimensional indexing. *IEEE ICASSP*, 6:3470–3473, May 1996.

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

[3] F. Bergeaud and S. Mallat. Matching pursuit of images. *Proc. IEEE ICIP*, 1:53–56, Oct. 1995.

[4] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive Approximations*, 1997. Springer-Verlag NY.

[5] R. M. Figueras i Ventura and P. Vandergheynst. Evolutionary multiresolution matching pursuit and its relations with the human visual system. *submitted to EUSIPCO 2002*. Toulouse.

[6] R. M. Figueras i Ventura and P. Vandergheynst. Matching pursuit through genetic algorithms. *LTS-EPFL tech. report*.

[7] C. Kan and M. Srinath. Invariant character recognition with zernike and orthogonal fourier-mellin moments. *Pattern Recognition*, 35:143–154, 2002.

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

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

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

[11] P. Vandergheynst and P. Frossard. Efficient image representation by anisotropic refinement in matching pursuit. *Proc. of IEEE ICASSP'01*, 3:1757–1760, May 2001.