

Sparse Image Representation with Epitomes

Louise Benoît, Julien Mairal, Francis Bach and Jean Ponce

École Normale Supérieure, INRIA, Paris, France

1. Introduction

- **Goal.** Our goal is to explain a new formulation and an algorithm for learning epitomes and illustrate their use in image denoising.
- **What is an epitome ?** An epitome is a structured dictionary. It is itself a small image which summarizes the content of a larger one. For any patch in the large image, there is a similar one in the epitome.

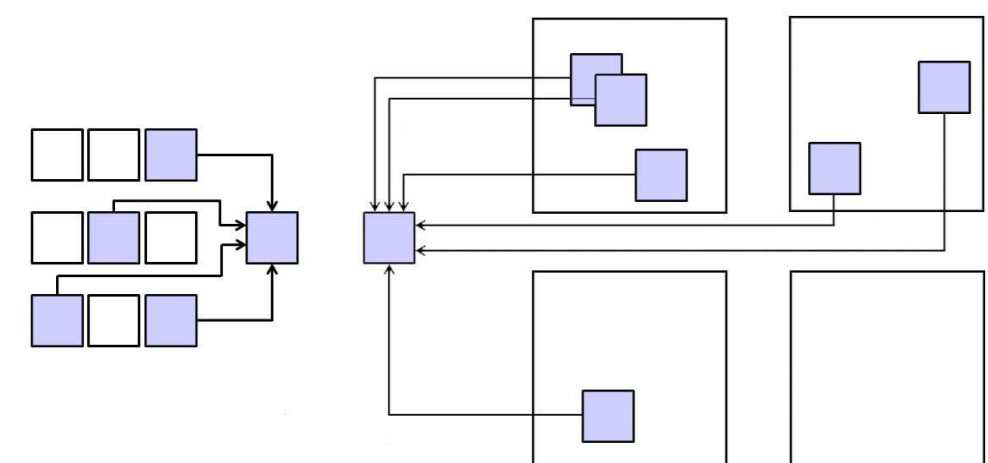


FIGURE : A “flat” dictionary (left) vs. a collection of 4 epitomes (right). The atoms are extracted from the epitomes and may overlap.

3. Algorithm

The optimization problem (3) is not jointly convex in (\mathbf{D}, \mathbf{A}) , but is convex with respect to each variable when the other one is fixed. We use a block coordinate descent scheme that alternates between the optimization of \mathbf{D} and \mathbf{A} .

- **Step 1.** Optimization of \mathbf{A} :
We use classical solvers for ℓ_1 decomposition problems (LARS algorithm).
- **Step 2.** Optimization of \mathbf{D} :
We use an accelerated projected gradient descent algorithm (FISTA algorithm):

$$f(\mathbf{D}) = \frac{1}{2} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda \sum_{j=1}^p \|\mathbf{d}_j\|_2 \|\boldsymbol{\alpha}^j\|_1,$$

We compute the gradient and project it orthogonally onto $\text{Im } \varphi$.

5. Denoising with epitomes

- **Method.** We test the performance of epitome learning on a denoising task. While following the assumption that the clean image can be approximated by a sparse linear combination of elements from the epitome, the method proceeds as follows :

- Learn the epitome on the noisy image,
- Approximate each noisy patch with a greedy algorithm (OMP) using the learned epitome:

$$\text{argmin}_{\mathbf{a}_i \in \mathbb{R}^p} \|\mathbf{a}_i\|_0 \text{ s.t. } \|\mathbf{y}_i - \mathbf{D}\mathbf{a}_i\|_2^2 \leq (\mathcal{C}\sigma^2),$$

- where $\mathbf{D}\mathbf{a}_i$ is a clean estimate of the patch,
- Average the clean estimates for each pixel.

- **Results.**



σ	IE	E	ISD	Epit	K-SVD	BM3D	LSSC
10	34.83	34.67	34.71	28.83	34.76	35.24	35.32
15	32.95	32.79	32.84	28.92	32.87	33.43	33.50
20	31.55	31.41	31.36	28.55	31.52	32.15	32.18
25	30.41	30.29	29.99	28.12	30.42	31.15	31.11
50	26.57	26.52	25.91	25.21	26.66	27.69	27.87
mean	31.26	31.14	30.96	27.93	31.25	31.93	32.00

2. Approach

The classical dictionary learning formulation is given by:

$$\min_{\mathbf{D} \in \mathcal{D}, \mathbf{A} \in \mathbb{R}^{p \times n}} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \lambda \|\mathbf{a}_i\|_1 \right].$$

We denote by \mathbf{E} in \mathbb{R}^m an epitome of size $\sqrt{m} \times \sqrt{m}$. We introduce a linear operator φ that extracts all overlapping patches from the epitome \mathbf{E} , and rearranges them into the columns of the “flat” dictionary \mathbf{D} .

We introduce an equivalent formulation adapted to epitome learning, which can be expressed as follows:

$$\min_{\substack{\mathbf{D} \in \text{Im } \varphi, n \\ \mathbf{A} \in \mathbb{R}^{p \times n}}} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \lambda \sum_{j=1}^p \|\mathbf{d}_j\|_2 |\alpha_{j,i}| \right].$$

Motivations

- This formulation can easily be adapted to other types of structured dictionaries by changing φ .
- We can use the dictionary learning algorithms to learn the epitome.

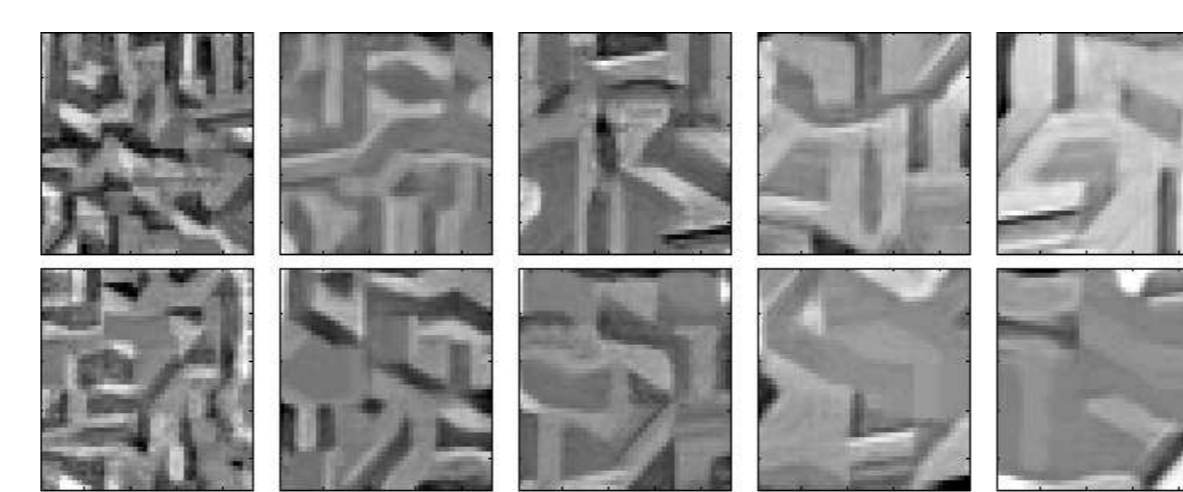
4. Experiments

Judging the visual aspect of the epitome is debatable, but we can nevertheless compare them.

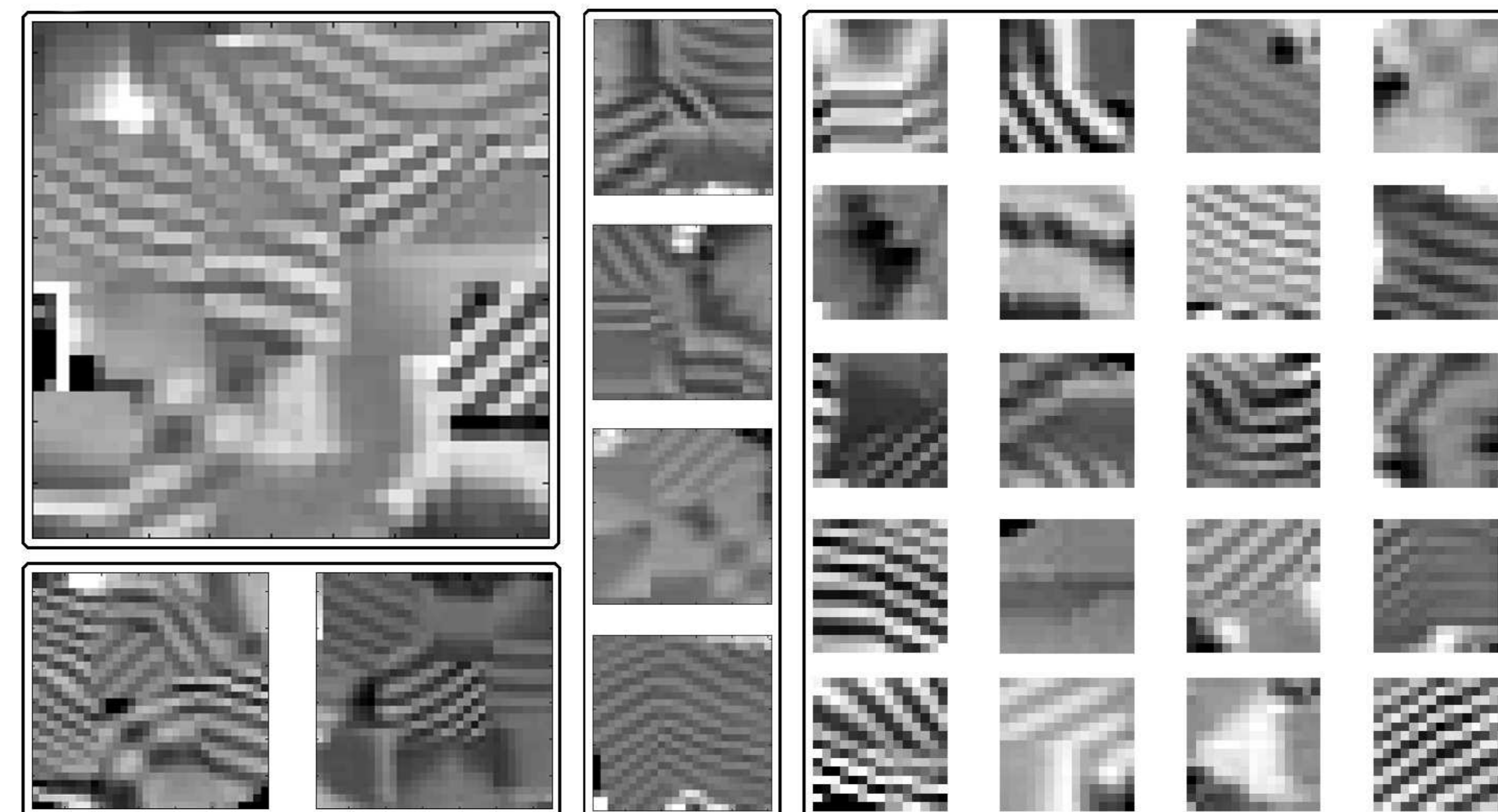
- Influence of the initialization :



- Influence of the size of the patches :



- Influence of the number of epitomes :



6. Conclusion

Future Work

- Extend the framework to video,
- Apply to other image processing tasks : inpainting, etc.,
- Learn image features for classification or recognition tasks,
- Encode other invariant properties through different mapping functions φ .

Acknowledgments. This work was partly supported by the European Community under the ERC grants “VideoWorld” and “Sierra”.

Main references

- M. Aharon and M. Elad. Sparse and redundant modeling of image content using an image-signature-dictionary. *SIAM Journal on Imaging Sciences*, 1(3):228–247, July 2008.
- N. Jovic, B. Frey, and A. Kannan. Epitomic analysis of appearance and shape. In *Proc. ICCV*, 2003.
- J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11:19–60, 2010.