



BUILDING INVARIANCE PROPERTIES FOR DICTIONARIES OF SAR IMAGE PATCHES

Sonia Tabti, Charles-Alban Deledalle, Loïc Denis, Florence Tupin

► **To cite this version:**

Sonia Tabti, Charles-Alban Deledalle, Loïc Denis, Florence Tupin. BUILDING INVARIANCE PROPERTIES FOR DICTIONARIES OF SAR IMAGE PATCHES. 2014. <hal-01001600v2>

HAL Id: hal-01001600

<https://hal.archives-ouvertes.fr/hal-01001600v2>

Submitted on 24 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BUILDING INVARIANCE PROPERTIES FOR DICTIONARIES OF SAR IMAGE PATCHES

Sonia Tabti¹, Charles-Alban Deledalle², Loïc Denis³, Florence Tupin¹

¹Institut Mines-Télécom, Télécom ParisTech CNRS-LTCl, Paris, France

²CNRS-Univ. Bordeaux, IMB, Talence, France

³CNRS-Univ. Saint-Etienne, Laboratoire Hubert Curien, Saint-Etienne, France

ABSTRACT

Adding invariance properties to a dictionary-based model is a convenient way to reach a high representation capacity while maintaining a compact structure. Compact dictionaries of patches are desirable because they ease semantic interpretation of their elements (atoms) and offer robust decompositions even under strong speckle fluctuations. This paper describes how patches of a dictionary can be matched to a speckled image by accounting for unknown shifts and affine radiometric changes. This procedure is used to build dictionaries of patches specific to SAR images. The dictionaries can then be used for denoising or classification purposes.

Index Terms— Patches, dictionary, shift-invariant, contrast-invariant, SAR images.

1. INTRODUCTION

Image *patches* are small rectangular regions, typically 3×3 to 11×11 pixels, that capture local geometry (e.g., edge, corner, smooth area, point-like source) and texture. Patch-based statistical modeling of natural images has encountered growing success in image processing the past decade, with applications such as denoising, compression and annotation. While non-local approaches rely on an auto-similarity principle (similar patches are likely to arise within the area surrounding a given patch), dictionary-based modeling decompose patches as a linear combination of (few) patches taken from a given collection (so-called dictionary of patches). Unlike fixed dictionaries based on Fourier or wavelets basis, learned dictionaries are suited to the specific content of a given image or class of images.

The dictionary learning procedure is usually expressed as the minimization of an energy that enforces the obtained dictionary to reconstruct well the image or its patches with only few atoms, i.e. sparsely. This idea was already at the heart of the pioneering work of Olshausen et al. in [1] and has more recently been popularized with the K-SVD algorithm [2]. Inspired by vectorial quantification methods such as K-means, K-SVD builds a dictionary iteratively in two steps: for each patch of the image a suitable linear combination of few atoms

is found (the sparse coding step), next the atoms are updated according to those coefficients (the dictionary update step).

Few researches considered the adaptation of dictionary learning methods for SAR (Synthetic Aperture Radar) images. Due to the multiplicative behavior of speckle, the standard K-SVD algorithm tends to learn dictionaries that essentially explain the brightest areas. For speckle reduction purposes, [3] suggested using a logarithmic transform of the data to map speckle noise to additive noise, and then applying the K-SVD procedure. To our knowledge, only [4] proposed a learning method that directly takes into account the specificity of non-additive noise. As a first step towards dictionary learning with non Gaussian noise, [4] suggested using only a vectorial quantification method, the k-medoids, in which each patch is only represented by one atom (1-sparse coding).

Although dictionary learning approaches can be very effective, they usually require a large dictionary to capture the diversity of patch structures found in an image. A notable reason is that all shifted versions of the structures must be encoded in the dictionary to be correctly modeled. If the dictionary incorporates in some way a shift-invariance, it can be made much more compact. In the MoTIF algorithm (Matching of Time Invariant Features) [5], the best shift parameter is computed over a training basis of patches by maximizing a square correlation. A shift invariant version of the Olshausen method was proposed in [6] and, more recently, the K-SVD algorithm has been extended in [7], [8]. Another proposed approach to implement shift-invariance is to use epitomes [9].

In this work, we design dictionaries for 1-sparse coding, as in [4]. Since atoms are not combined together, it is necessary to provide invariance properties to reach a sufficient expressive power. We show on numerical experiments that 1-sparse coding is a relevant choice for SAR images due to the strong noise level (Section 2). Next, as in [3], we consider the homomorphic transform, and we show how 1-sparse coding with shift and radiometric invariance can be implemented efficiently (Section 3). Finally, we apply our methodology for 1-sparse dictionary learning with shift and radiometric invariance (Section 4).

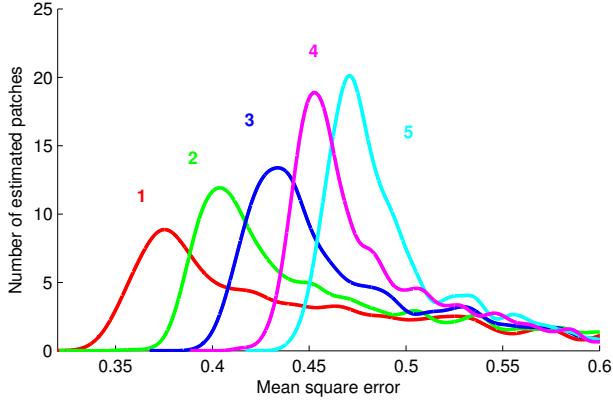


Fig. 1. Distributions of the mean squared error with respect to the sparsity level S . The MSE tends to increase with the number of atoms used in the linear combination for highly noisy patches.

2. APPROXIMATING NOISY SAR PATCHES WITH A SINGLE ATOM IS OPTIMAL IN THE MSE SENSE

While [4] motivates the 1-sparse coding assumption as a first step towards dictionary learning with non Gaussian noise, we draw here another motivation. To illustrate our point we conduct the following experiment. A dictionary of patches is learned with K-SVD on a log-transformed SAR image with very high signal-to-noise ratio (100 looks image). Then several patches are extracted from this SAR image and corrupted with multiplicative noise corresponding to single-look images. Log-transformed versions of these noisy patches are then sparse-coded with several sparsity degrees in order to evaluate which degree achieves the best approximation error (i.e., the smallest mean square error, MSE). The MSE distributions of the estimates obtained with different sparsity degrees are presented in Figure 1. We observe that sparsity 1 produces frequently the smallest MSE, and is optimal on average. Indeed, with such a high level of noise, higher sparsity levels lead to higher degrees of freedom and result in larger estimation variances. These experiments suggest that a single atom is enough to approximate noisy patches in SAR imaging.

3. CONTRAST AND SHIFT INVARIANT MATCHING OF A DICTIONARY ATOM

Dictionary learning procedures rely on some performance measure of the dictionary being learned, typically the approximation error of a given noisy image. Such a criterion is simple to evaluate in our context: since we consider 1-sparse codes, the performance of an atom is independent on other atoms. For speed and to reduce estimation variance, dictionaries should be as small as possible while being representative of the image content. To limit the size of dictionaries, a common idea is to let atoms represent a class of

patches that are identical up to some transformations. We propose here a methodology to match an atom up to a shift and affine radiometric change.

Let y be an observed patch corrupted by a white Gaussian noise with standard deviation σ (in the following, y will be extracted from a log-transformed SAR image). Given an atom a and the patch y , the approximation error when approximating patch y with atom a up to an affine radiometric change can be measured with the optimal sum of square difference (SSD) as described by the CI-GLR test (Contrast-Invariant Generalized Likelihood Ratio) in [10]:

$$\text{GLR}^{\text{CI}}(a, y) = \begin{cases} (1 - \mathcal{C}(y, a)^2) \frac{\|y - \bar{y}\mathbb{1}\|^2}{2\sigma^2} & \text{if } a \neq \bar{a}\mathbb{1} \text{ and } y \neq \bar{y}\mathbb{1} \\ \frac{\|y - \bar{y}\mathbb{1}\|^2}{2\sigma^2} & \text{otherwise,} \end{cases} \quad (1)$$

where $\mathcal{C}(y, a) = \left| \frac{\sum_{k=1}^P (y_k - \bar{y})(a_k - \bar{a})}{\sqrt{\sum_{k=1}^P (y_k - \bar{y})^2 \sum_{k=1}^P (a_k - \bar{a})^2}} \right|$ is the normalized correlation, and \bar{a} and \bar{y} are the empirical means of a and y , P is the size of each patch, y_k (resp. a_k) is the value of the patch y (resp. a) at location k , and $\mathbb{1}$ is a vector whose entries are all equal to 1. From Equation (1), the optimal atom is the one maximizing the correlation. This principle is indeed at the heart of several greedy approaches such as matching pursuit algorithms. However correlation itself is not a measure of fitness since the approximation error depends also on $\|y - \bar{y}\mathbb{1}\|^2/2\sigma^2$: a quantity related to the signal-to-noise ratio of the observation y . Hence, correlation indicates which atoms explain well a given patch, while (1) define which patches are well explained by a given atom.

Equation (1) provides a criterion to evaluate how a noisy patch y matches an atom a up to an affine change of contrast. To evaluate the overall performance of an atom a , we then need to be able to quickly match it against all the patches y that could be extracted from the whole image \mathbf{y} . In [9], the authors proposed an algorithm to extract quickly an atom from a large image (called epitome) that correlates best with a given patch. We employ here a similar strategy to compute for all patches of a noisy image \mathbf{y} the 1-sparse approximation error obtained with a given atom. The map associating to all pixel locations the approximation error (1) between a and \mathbf{y} can be computed by:

$$\text{SSD}_{\text{map}}^{\text{CI}}(a, \mathbf{y}) = \frac{1}{2\sigma^2} \times \begin{cases} \mathbf{y}^2 \star \iota - P(\mathbf{y} \star \iota/P)^2 - \frac{(\mathbf{y} \star (a - \bar{a}\mathbb{1}))^2}{\|a - \bar{a}\mathbb{1}\|^2} & \text{if } a \neq \bar{a}\mathbb{1} \\ \mathbf{y}^2 \star \iota - P(\mathbf{y} \star \iota/P)^2 & \text{otherwise,} \end{cases} \quad (2)$$

where CI stands for contrast invariance and ι is the support of a patch (typically a $P \times P$ squared box window). The approximation of a noisy image with a given dictionary can be computed efficiently using Equation (2) since each convolution can be computed separately in the Fourier domain. $\text{SSD}_{\text{map}}^{\text{CI}}(a, \mathbf{y})$ can thus be evaluated for a given atom a against all patches of an image with a complexity

$O(N \log N)$ where N is the image size. In comparison, applying (1) after extracting all patches would be of complexity $O(NP)$.

Beyond contrast change invariance, we wish to implement a shift invariance. Rather than requiring that all patches from the noisy image be approximated with the dictionary (i.e., that each pixel from the image be represented several times), in each patch that includes the pixel, we require that each pixel be well approximated at least once. The approximation error is then the smallest error among all patches including the pixel. The map giving at each pixel the optimal approximation error between a given atom and all surrounding patches can readily be obtained by gray-level erosion (in the mathematical morphology sense) of the map given by equation (2):

$$\text{SSD}_{\text{map}}^{\text{CI+SI}}(a, \mathbf{y}) = \text{erode}(\text{SSD}_{\text{map}}^{\text{CI}}(a, \mathbf{y}), \iota) \quad (3)$$

where SI stands for shift-invariance and $\text{erode}(\cdot, \iota)$ is the erosion operator with structural element ι .

Figure 2 illustrates the efficiency of our shift invariant procedure. In fact, in this toy example, we first computed an initial $\text{SSD}_{\text{map}}^{\text{CI+SI}}$ with a constant atom. Then we selected an atom (in this experiment: a portion of an edge) to prove that the shift invariance is verified by the criterion of equation (2) (i.e. the whole edge is explained).

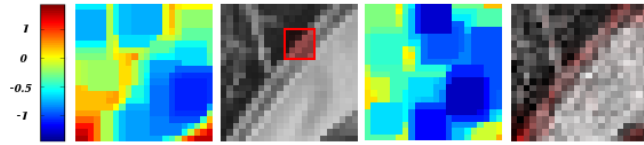


Fig. 2. From left to right: the initial $\text{SSD}_{\text{map}}^{\text{CI+SI}}$ obtained with the constant atom, the selected atom (indicated by the red box); the updated $\text{SSD}_{\text{map}}^{\text{CI+SI}}$ using this atom given by the minimum values of the two maps. Blue indicates small values (i.e., good approximation).

4. LEARNING A DICTIONARY OF SAR IMAGE PATCHES

Based on the shift and contrast invariant matching criterion $\text{SSD}_{\text{map}}^{\text{CI+SI}}$ derived in previous paragraph, we design a dictionary learning procedure. The aim of this dictionary is to provide a compact summary of a noisy image \mathbf{y} . The atoms of this dictionary are extracted from a high signal-to-noise ratio¹ reference SAR image \mathbf{y}^* . Both \mathbf{y} and \mathbf{y}^* are log-transformed in order to turn speckle fluctuations into an additive term with constant variance.

First, T patches in \mathbf{y}^* are extracted at random locations. They define a set \mathcal{C} of candidate elements that will be considered for inclusion in the final dictionary \mathcal{D} . The total number

¹in practice, such an image can be obtained by multi-looking a very high resolution image (e.g., 100 looks from decimeter to meter resolution).

of elements in the dictionary, K , is set manually. An initial dictionary is built from the first K elements of the set of candidate patches \mathcal{C} .

We define the map of patch approximation errors of image \mathbf{y} with dictionary \mathcal{D} for each pixel i as:

$$\text{SSD}_{\text{map}}^{\text{CI+SI}}(\mathcal{D}, \mathbf{y})_i = \min_{k_i} \text{SSD}_{\text{map}}^{\text{CI+SI}}(a_{k_i}, \mathbf{y})_i, \quad (4)$$

The dictionary is updated by considering all candidate patches in \mathcal{C} in turn and replacing atoms of the dictionary each time the average approximation error can be reduced. When considering replacing patch a_k from the dictionary with patch c_t from the candidates set \mathcal{C} , a map of local gains is computed:

$$\text{SSD}_{\text{map}}^{\text{CI+SI}}(\mathcal{D}, \mathbf{y})_i - \text{SSD}_{\text{map}}^{\text{CI+SI}}(\mathcal{D} \cup \{c_t\} \setminus \{a_k\}, \mathbf{y})_i. \quad (5)$$

Candidate patch c_t makes its way into current dictionary \mathcal{C} if an atom a_k can be found such that the average gain is positive, i.e., the approximation error can be decreased by replacing a_k with c_t . The process is then repeated for c_{t+1} .

The computation of the gain can be made fast by keeping track of the best two approximation errors and approximating atoms with indices $k_i^{1\text{st}}, k_i^{2\text{nd}}$ for each pixel i of the image \mathbf{y} . Computation of the gain defined in equation (5) then requires only the computation of a map $\text{SSD}_{\text{map}}^{\text{CI+SI}}(c_t, \mathbf{y})$ using (3). Three cases can occur:

- The atom a_k to be replaced is not the optimum at pixel i (i.e., $k \neq k_i^{1\text{st}}$). In this case either c_t improves at pixel i or either it does not affect the reconstruction error. The gain at pixel i is then given by:

$$\max(\text{SSD}_{\text{map}}^{\text{CI+SI}}(\mathcal{D}, \mathbf{y})_i - \text{SSD}_{\text{map}}^{\text{CI+SI}}(c_t, \mathbf{y})_i, 0),$$

- Otherwise $a_{k_i^{1\text{st}}}$ is going to be replaced. Suppose the candidate c_t is better than the second best atom $a_{k_i^{2\text{nd}}}$ in terms of reconstruction error, so the gain at pixel i is the difference:

$$\text{SSD}_{\text{map}}^{\text{CI+SI}}(\mathcal{D}, \mathbf{y})_i - \text{SSD}_{\text{map}}^{\text{CI+SI}}(c_t, \mathbf{y})_i,$$

- Otherwise $a_{k_i^{1\text{st}}}$ is going to be replaced and the candidate c_t is worse than $a_{k_i^{2\text{nd}}}$, so the gain at pixel i is:

$$\text{SSD}_{\text{map}}^{\text{CI+SI}}(\mathcal{D}, \mathbf{y})_i - \text{SSD}_{\text{map}}^{\text{CI+SI}}(a_{k_i^{2\text{nd}}}, \mathbf{y})_i.$$

Overall the complexity is of order $O(TN \log N)$ independently of the number of atoms K .

An example of dictionary produced by our proposed approach is presented on figure 3. A K-SVD dictionary learned on the same noisy image is also presented. They are learned using the same 1-look SAR image (cf. fig. 4) and both are composed of 64 atoms of size 8×8 . One can observe that our dictionary summarizes well the structures of the SAR image (the backscattering lines for example). The K-SVD dictionary seems to explain well edges but contains too many homogeneous atoms which makes the information provided by these atoms too redundant.

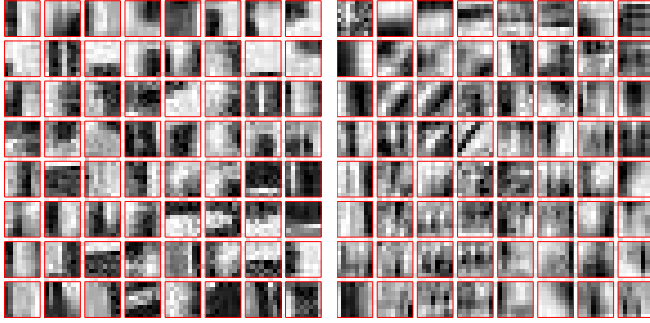


Fig. 3. From left to right: our dictionary and the K-SVD dictionary, both of them composed of 64 atoms of size 8×8 and learned from the images in figure 4.

5. CONCLUSION

In this paper, we have introduced a new methodology to design dictionaries of patches suited to SAR images with shift and radiometric invariances. As a first contribution, we have provided a measure of performance to evaluate the quality of a dictionary in approximating a log transformed SAR image. The second contribution was to show that this approximation error is computed with an $O(N \log N)$ complexity. This low complexity makes it so tractable to be used as an objective criterion for a dictionary learning technique, as shown with a proposed algorithm based on a simple replacement strategy.

The quality of the obtained results demonstrates the potential of the approach. Future work will focus on designing more sophisticated learning approaches based on our shift and contrast invariant criterion and apply them to denoising and clustering.

6. REFERENCES

- [1] B.A. Olshausen et al., “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” *Nature*, 1996.
- [2] M. Aharon, M. Elad, and A.M. Bruckstein, “The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation,” *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [3] S. Foucher, “SAR image filtering via learned dictionaries and sparse representations,” in *IGARSS. IEEE*, 2008, pp. 229–232.
- [4] P. Chainais, “Towards dictionary learning from images with non gaussian noise,” in *Workshop on Machine Learning for Signal Processing. IEEE*, 2012, pp. 1–6.
- [5] S. Lesage, P. Jost, P. Vandergheynst and R. Gribonval, “MoTIF: an efficient algorithm for learning translation

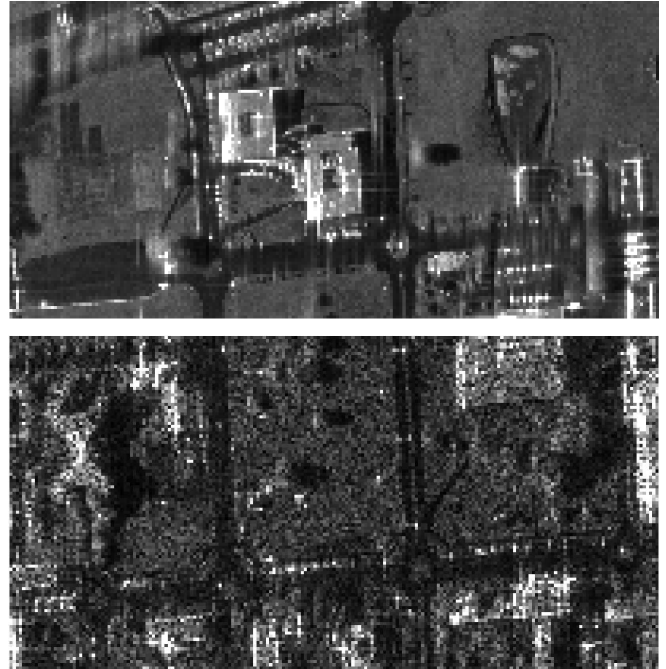


Fig. 4. On top: the noiseless reference image \mathbf{y}^* (Copyright CNES) in which candidate patches are extracted; on the bottom: the 1-look image \mathbf{y} (Copyright CNES) that is to be explained by the constructed dictionary. Our method selects K candidate atoms from \mathbf{y}^* based on the average reconstruction error of \mathbf{y} . The resulting dictionary is given in figure 3.

invariant dictionaries,” in *ICASSP. IEEE*, 2006, vol. 5, p. V.

- [6] W. Hashimoto and K. Kurata, “Properties of basis functions generated by shift invariant sparse representations of natural images,” *Biological Cybernetics*, vol. 83, no. 2, pp. 111–118, 2000.
- [7] J.J. Thiagarajan, K.N. Ramamurthy, and A. Spanias, “Shift-invariant sparse representation of images using learned dictionaries,” in *Workshop on Machine Learning for Signal Processing. IEEE*, 2008, pp. 145–150.
- [8] B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot, and P. Vandergheynst, “Shift-invariant dictionary learning for sparse representations: extending K-SVD,” in *EU-SIPCO*, 2008.
- [9] L. Benoit, J. Mairal, F. Bach, and J. Ponce, “Sparse image representation with epitomes,” in *CVPR. IEEE*, 2011, pp. 2913–2920.
- [10] C.-A. Deledalle, L. Denis, and F. Tupin, “Template matching with noisy patches: A contrast-invariant GLR test,” *EUSIPCO*, 2013.