



# A Bayesian Hyperprior Approach for Joint Image Denoising and Interpolation, with an Application to HDR Imaging

Cecilia Aguerrebere, Andrés Almansa, Julie Delon, Yann Gousseau, Pablo Musé

## ► To cite this version:

Cecilia Aguerrebere, Andrés Almansa, Julie Delon, Yann Gousseau, Pablo Musé. A Bayesian Hyperprior Approach for Joint Image Denoising and Interpolation, with an Application to HDR Imaging. IEEE Transactions on Computational Imaging, 2017, .

**HAL Id: hal-01107519**

**<https://hal.archives-ouvertes.fr/hal-01107519v5>**

Submitted on 15 May 2017

**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.

# A Bayesian Hyperprior Approach for Joint Image Denoising and Interpolation, with an Application to HDR Imaging

Cecilia Aguerrebere, Andrés Almansa, Julie Delon, Yann Gousseau and Pablo Musé

**Abstract**—Recently, impressive denoising results have been achieved by Bayesian approaches which assume Gaussian models for the image patches. This improvement in performance can be attributed to the use of per-patch models. Unfortunately such an approach is particularly unstable for most inverse problems beyond denoising. In this work, we propose the use of a hyperprior to model image patches, in order to stabilize the estimation procedure. There are two main advantages to the proposed restoration scheme: Firstly it is adapted to diagonal degradation matrices, and in particular to missing data problems (e.g. inpainting of missing pixels or zooming). Secondly it can deal with signal dependent noise models, particularly suited to digital cameras. As such, the scheme is especially adapted to computational photography. In order to illustrate this point, we provide an application to high dynamic range imaging from a single image taken with a modified sensor, which shows the effectiveness of the proposed scheme.

**Index Terms**—Non-local patch-based restoration, Bayesian restoration, Maximum a Posteriori, Gaussian Mixture Models, hyper-prior, conjugate distributions, high dynamic range imaging, single shot HDR, hierarchical models.

## I. INTRODUCTION

**D**IGITAL images are subject to a wide variety of degradations, which in most cases can be modeled as

$$\mathbf{Z} = \mathbf{D}\mathbf{C} + \mathbf{N}, \quad (1)$$

where  $\mathbf{Z}$  is the observation,  $\mathbf{D}$  is the degradation operator,  $\mathbf{C}$  is the underlying ground-truth image and  $\mathbf{N}$  is additive noise. Different settings of the degradation matrix  $\mathbf{D}$  model different problems such as zooming, deblurring or missing pixels. Different versions of the noise term  $\mathbf{N}$  include the classical additive Gaussian noise with constant variance or more complicated and realistic models such as signal dependent noise.

Due to the inherent ill-posedness of such inverse problems, standard approaches impose some prior on the image, in either variational or Bayesian approaches. Popular image models have been proposed through the total variation [1], wavelet decompositions [2] or the sparsity of image patches [3].

C. Aguerrebere is with the Department of Electrical and Computer Engineering, Duke University, Durham NC 27708, US (e-mail: cecilia.aguerrebere@duke.edu)

Y. Gousseau is with the LTCI, Télécom ParisTech, Université Paris-Saclay, 75013 Paris, France (e-mail: gousseau@telecom-paristech.fr).

A. Almansa and J. Delon are with MAP5 (CNRS UMR 8145), Université Paris Descartes, 75270 Paris Cedex 06 (e-mail: andres.almansa,julie.delon@parisdescartes.fr)

P. Musé is with the Department of Electrical Engineering, Universidad de la República, 11300 Montevideo, Uruguay (e-mail: pmuse@fing.edu.uy)

Buades et al. [4] introduced the use of patches and the self-similarity hypothesis to the denoising problem leading to a new era of patch-based image restoration techniques. A major step forward in fully exploiting the potential of patches was achieved by several state-of-the-art restoration methods with the introduction of patch prior models, in a Bayesian framework. Some methods are devoted to the denoising problem [5]–[8], while others propose a more general framework for the solution of image inverse problems [9], [10], including inpainting, deblurring and zooming. The work by Lebrun et al. [7], [11] presents a thorough analysis of several recent restoration methods, revealing their common roots and their relationship with the Bayesian approach.

Among the state-of-the-art restoration methods, two noticeable approaches are the patch-based Bayesian approach by Yu et al. [10], namely the piece-wise linear estimators (PLE), and the non-local Bayes (NLB) algorithm by Lebrun et al. [7]. PLE is a general framework for the solution of image inverse problems under Model (1), while NLB is a denoising method ( $\mathbf{D} = \mathbf{Id}$ ). Both methods use a Gaussian patch prior learnt from image patches through iterative procedures. In the case of PLE, patches are modeled according to a Gaussian Mixture Model (GMM), with a relatively small number of classes (19 in all their experiments), whose parameters are learnt from all image patches<sup>1</sup>. In the case of NLB, each patch is associated with a single Gaussian model, whose parameters are computed from similar patches chosen from a local neighbourhood, i.e., a search window centered at the patch. We refer hereafter to this kind of per-patch modelling as *local*.

Zoran and Weiss [9] (EPLL) follow a similar approach, but instead of iteratively updating the GMM from image patches, they use a larger number of classes that are fixed and learnt from a large database of natural image patches. Wang and Morel [8] claim that, in the case of denoising, it is better to have fewer models that are updated with the image patches (as in PLE) than having a large number of fixed models (as in EPLL).

All of the previous restoration approaches share a common Bayesian framework based on Gaussian patch priors. Relying on local priors [7], [8] has proven more accurate for the task of image denoising than relying on a mixture of a limited number of Gaussian models [9], [10]. In particular, NLB outperforms PLE for this task [12], mostly due to its local model estimation.

<sup>1</sup>Actually, the authors report the use of  $128 \times 128$  image sub-regions in their experiments, so we may consider PLE as a semi-local approach.

On the other hand, PLE yields state-of-the-art results in other applications such as interpolation of missing pixels (especially with high masking rates), deblurring and zooming.

As a consequence we are interested in taking advantage of a local patch modelling for more general inverse problems than denoising. The main difficulty lies in the estimation of the models, especially when the image degradations involve a high rate of missing pixels, in which case the estimation is seriously ill-posed.

In this work we propose to model image patches according to a Gaussian prior, whose parameters, the mean  $\mu$  and the covariance matrix  $\Sigma$ , will be estimated locally from similar patches. In order to tackle this problem, we include prior knowledge on the model parameters making use of a hyperprior, i.e. a probability distribution on the parameters of the prior. In Bayesian statistics,  $\mu$  and  $\Sigma$  are known as hyperparameters, while the prior on them is called a hyperprior. Such a framework is often called hierarchical Bayesian modelling [13]. Its application to inverse problems in imaging is not new. In particular, in the field of image restoration, this methodology was proposed by Molina et al. [14], [15], and was more recently applied to image unmixing problems [16] and to image deconvolution and the estimation of the point spread function of a camera [17]. However, to our knowledge, this is the first time that such a hierarchical Bayesian methodology is used to reduce ill-posedness in patch-based image restoration. In this context, the use of a hyperprior compensates for the patches missing information.

There are two main contributions of this work: First, as described above, we propose a robust framework enabling the use of Gaussian local priors on image patches for solving a useful family of restoration problems by drawing on a hierarchical Bayesian approach. The second advantage of the proposed framework is its ability to deal with signal dependent noise, therefore making it adapted to realistic digital photography applications.

Experiments on both synthetic and real data show that the approach is well suited to various problems involving a diagonal degradation operator. First, we show state-of-the-art results in image restoration problems such as denoising, zooming and interpolation of missing pixels. Then we consider the generation of high dynamic range (HDR) images from a single snapshot using spatially varying pixel exposures [18] and demonstrate that our approach significantly outperforms existing methods to deal with this inverse problem. It is worth mentioning that modified sensors enabling such approaches have been recently made available by Sony but are not yet fully exploited by available smartphones and digital cameras.

The article is organized as follows. Section II introduces the proposed approach while Section III presents the main implementation aspects. Supportive experiments are presented in Section IV. Section V is devoted to the application of the proposed framework to the HDR imaging problem. Last, conclusions are summarized in Section VI.

## II. HYPERPRIOR BAYESIAN ESTIMATOR

The proposed restoration method, called Hyperprior Bayesian Estimator (HBE), assumes a Gaussian prior for

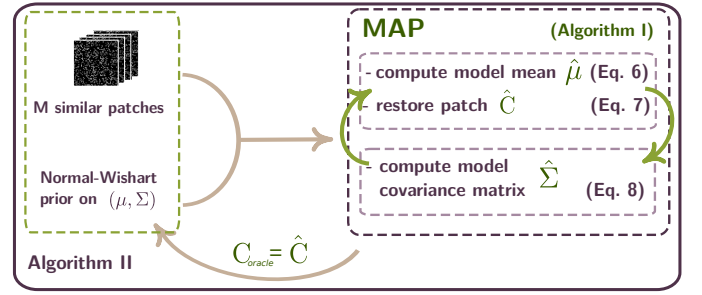


Fig. 1. Diagram of the proposed iterative approach.

image patches, with parameters  $\mu$  and  $\Sigma$ . A **joint maximum a posteriori** formulation is used to estimate both the image patches and the parameters  $\mu$  and  $\Sigma$ , thanks to a Bayesian hyperprior model on these parameters, stabilizing the local estimation of the Gaussian statistics. As a consequence, we can exploit the accuracy of local model estimation for general restoration problems, in particular with missing values (e.g. for interpolation or zooming). Figure 1 illustrates the proposed approach which is described in detail in the following.

### A. Patch degradation model

The observed image  $\mathbf{z}$  is decomposed into  $I$  overlapping patches  $\{\mathbf{z}_i\}_{i=1,\dots,I}$  of size  $\sqrt{n} \times \sqrt{n}$ . Each patch  $\mathbf{z}_i \in \mathbb{R}^{n \times 1}$  is considered to be a realization of the random variable  $\mathbf{Z}_i$  given by

$$\mathbf{Z}_i = \mathbf{D}_i \mathbf{C}_i + \mathbf{N}_i, \quad (2)$$

where  $\mathbf{D}_i \in \mathbb{R}^{n \times n}$  is a degradation operator,  $\mathbf{C}_i \in \mathbb{R}^{n \times 1}$  is the original patch we seek to estimate and  $\mathbf{N}_i \in \mathbb{R}^{n \times 1}$  is an additive noise term, modeled by a Gaussian distribution  $\mathbf{N}_i \sim \mathcal{N}(0, \Sigma_{N_i})$ . Therefore, the distribution of  $\mathbf{Z}_i$  given  $\mathbf{C}_i$  can be written as

$$\begin{aligned} p(\mathbf{Z}_i | \mathbf{C}_i) &\sim \mathcal{N}(\mathbf{D}_i \mathbf{C}_i, \Sigma_{N_i}) \\ &\propto |\Sigma_{N_i}^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{Z}_i - \mathbf{D}_i \mathbf{C}_i)^T \Sigma_{N_i}^{-1}(\mathbf{Z}_i - \mathbf{D}_i \mathbf{C}_i)\right). \end{aligned} \quad (3)$$

In this noise model, the matrix  $\Sigma_{N_i}$  is only assumed to be diagonal (the noise is uncorrelated). It can represent a constant variance, spatially variable variances or even variances dependent on the pixel value (to approximate Poisson noise).

This degradation model is deliberately generic. We will see in Section IV that keeping a broad noise model is essential to properly tackle the problem of HDR imaging from a single image. The model also includes the special case of multiplicative noise.

### B. Joint Maximum A Posteriori

We assume a Gaussian prior for each patch, with unknown mean  $\mu$  and covariance matrix  $\Sigma$ ,  $p(\mathbf{C}_i | \mu, \Sigma) \sim \mathcal{N}(\mu, \Sigma)$ . To simplify calculations we work with the precision matrix  $\Lambda = \Sigma^{-1}$ . As it is usual when considering hyperpriors, we assume that the parameters  $\mu$  and  $\Sigma$  follow a conjugate

distribution. In our case, that boils down to assuming a Normal-Wishart<sup>2</sup> prior for the couple  $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ ,

$$\begin{aligned} p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \mathcal{N}(\boldsymbol{\mu} | \boldsymbol{\mu}_0, (\kappa \boldsymbol{\Lambda})^{-1}) \mathcal{W}(\boldsymbol{\Lambda} | (\nu \boldsymbol{\Sigma}_0)^{-1}, \nu) \\ &\propto |\boldsymbol{\Lambda}|^{1/2} \exp\left(-\frac{\kappa}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \boldsymbol{\Lambda} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)\right) \\ &\quad |\boldsymbol{\Lambda}|^{(\nu-n-1)/2} \exp\left(-\frac{1}{2}\text{tr}(\nu \boldsymbol{\Sigma}_0 \boldsymbol{\Lambda})\right), \end{aligned} \quad (4)$$

with parameters  $\boldsymbol{\mu}_0$ ,  $\boldsymbol{\Sigma}_0$ , scale parameter  $\kappa > 0$  and  $\nu > n-1$  degrees of freedom.

Now, assume that we observe a group  $\{\mathbf{Z}_i\}_{i=1,\dots,M}$  of similar patches and that we want to recover the restored patches  $\{\mathbf{C}_i\}_{i=1,\dots,M}$ . If these unknown  $\{\mathbf{C}_i\}$  are independent<sup>3</sup> and follow the same Gaussian model, we can compute the joint maximum a posteriori

$$\begin{aligned} \arg \max_{\{\mathbf{C}_i\}, \boldsymbol{\mu}, \boldsymbol{\Lambda}} p(\{\mathbf{C}_i\}, \boldsymbol{\mu}, \boldsymbol{\Lambda} | \{\mathbf{Z}_i\}) &= \\ = p(\{\mathbf{Z}_i\} | \{\mathbf{C}_i\}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) p(\{\mathbf{C}_i\} | \boldsymbol{\mu}, \boldsymbol{\Lambda}) p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \\ = p(\{\mathbf{Z}_i\} | \{\mathbf{C}_i\}) p(\{\mathbf{C}_i\} | \boldsymbol{\mu}, \boldsymbol{\Lambda}) p(\boldsymbol{\mu}, \boldsymbol{\Lambda}). \end{aligned} \quad (5)$$

In this product, the first term is given by the noise model (3), the second one is the Gaussian prior on the set of patches  $\{\mathbf{C}_i\}$  and the third one is the hyperprior (4). In the last equality we omit the explicit dependence on  $\boldsymbol{\mu}$  and  $\boldsymbol{\Lambda}$  in  $p(\{\mathbf{Z}_i\} | \{\mathbf{C}_i\}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ , since these parameters are completely determined by the set  $\{\mathbf{C}_i\}$ .

### C. Optimality conditions

Computing the joint maximum a posteriori amounts to minimizing

$$\begin{aligned} f(\{\mathbf{C}_i\}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) &:= -\log p(\{\mathbf{C}_i\}, \boldsymbol{\mu}, \boldsymbol{\Lambda} | \{\mathbf{Z}_i\}) \\ &= \frac{1}{2}(\mathbf{Z}_i - \mathbf{D}_i \mathbf{C}_i)^T \boldsymbol{\Sigma}_{N_i}^{-1} (\mathbf{Z}_i - \mathbf{D}_i \mathbf{C}_i) \\ &\quad - \frac{\nu - n + M}{2} \log |\boldsymbol{\Lambda}| \\ &\quad + \frac{1}{2} \sum_{i=1}^M (\mathbf{C}_i - \boldsymbol{\mu})^T \boldsymbol{\Lambda} (\mathbf{C}_i - \boldsymbol{\mu}) \\ &\quad + \frac{\kappa}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \boldsymbol{\Lambda} (\boldsymbol{\mu} - \boldsymbol{\mu}_0) + \frac{1}{2} \text{trace}[\nu \boldsymbol{\Sigma}_0 \boldsymbol{\Lambda}], \end{aligned}$$

over the set  $\mathbb{R}^{nM} \times \mathbb{R}^n \times S_n^{++}(\mathbb{R})$ , with  $S_n^{++}(\mathbb{R})$  the set of real symmetric positive definite matrices of size  $n$ .

The function  $f$  is biconvex respectively in the variables  $(\{\mathbf{C}_i\}, \boldsymbol{\mu})$  and  $\boldsymbol{\Lambda}$ . To minimize this energy for a given set of hyper-parameters  $(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$ , we will use an alternating convex minimization scheme. At each iteration,  $f$  is first minimized with respect to  $(\{\mathbf{C}_i\}, \boldsymbol{\mu})$  with  $\boldsymbol{\Lambda}$  fixed, then viceversa.

Differentiating  $f$  with respect to each variable, we get explicit optimality equations for the minimization scheme. The proofs of the following propositions are straightforward and available in the supplementary material.

**Proposition 1.** Assume that  $\boldsymbol{\Lambda}$  is fixed and that the covariance  $\boldsymbol{\Sigma}_{N_i}$  does not depend on the  $\{\mathbf{C}_i\}$ . The function  $(\{\mathbf{C}_i\}, \boldsymbol{\mu}) \mapsto$

$f(\{\mathbf{C}_i\}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$  is convex on  $\mathbb{R}^{n(M+1)}$  and attains its minimum at  $(\{\hat{\mathbf{C}}_i\}, \hat{\boldsymbol{\mu}})$ , given by

$$\hat{\boldsymbol{\mu}} = \left( \kappa \text{Id} + \sum_{i=1}^M \mathbf{A}_i \mathbf{D}_i \right)^{-1} \left( \sum_{i=1}^M \mathbf{A}_i \mathbf{Z}_i + \kappa \boldsymbol{\mu}_0 \right). \quad (6)$$

$$\hat{\mathbf{C}}_i = \mathbf{A}_i (\mathbf{Z}_i - \mathbf{D}_i \hat{\boldsymbol{\mu}}) + \hat{\boldsymbol{\mu}}, \quad \forall i \in \{1, \dots, M\} \quad (7)$$

with  $\mathbf{A}_i = \boldsymbol{\Lambda}^{-1} \mathbf{D}_i^T (\mathbf{D}_i \boldsymbol{\Lambda}^{-1} \mathbf{D}_i^T + \boldsymbol{\Sigma}_{N_i})^{-1}$ .

**Proposition 2.** Assume that the variables  $(\{\mathbf{C}_i\}, \boldsymbol{\mu})$  are fixed. The function  $\boldsymbol{\Lambda} \rightarrow f(\{\mathbf{C}_i\}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$  is convex on  $S_n^{++}(\mathbb{R})$  and attains its minimum at  $\hat{\boldsymbol{\Lambda}}$  such that

$$\hat{\boldsymbol{\Lambda}}^{-1} = \frac{\nu \boldsymbol{\Sigma}_0 + \kappa (\boldsymbol{\mu} - \boldsymbol{\mu}_0)(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T + \sum_{i=1}^M (\mathbf{C}_i - \boldsymbol{\mu})(\mathbf{C}_i - \boldsymbol{\mu})^T}{\nu + M - n}. \quad (8)$$

The expression of  $\hat{\mathbf{C}}_i$  in (7) is obtained under the hypothesis that the noise covariance matrix  $\boldsymbol{\Sigma}_{N_i}$  does not depend on  $\mathbf{C}_i$ . Under the somewhat weaker hypothesis that the noise  $N_i$  and the signal  $\mathbf{C}_i$  are uncorrelated, this estimator is also the affine estimator  $\tilde{\mathbf{C}}_i$  that minimizes the Bayes risk  $\mathbb{E}[(\tilde{\mathbf{C}}_i - \mathbf{C}_i)^2]$  (c.f. supplementary material). The uncorrelatedness of  $N_i$  and  $\mathbf{C}_i$  is a reasonable hypothesis in practice. This includes various noise models, such as  $N_i = f(C_i)\varepsilon_i$  with  $\varepsilon_i$  independent of  $C_i$ , which approximates CMOS and CCD raw data noise [20].

From (6), we find that the MAP estimator of  $\boldsymbol{\mu}$  is a weighted average of two terms: the mean estimated from the similar restored patches and the prior  $\boldsymbol{\mu}_0$ . The parameter  $\kappa$  controls the confidence level we have on the prior  $\boldsymbol{\mu}_0$ . With the same idea, we observe that the MAP estimator for  $\boldsymbol{\Lambda}$  is a combination of the prior  $\boldsymbol{\Lambda}_0$  on  $\boldsymbol{\Lambda}$ , the covariance imposed by  $\boldsymbol{\mu}$  and the covariance matrix estimated from the patches  $\mathbf{C}_i$ .

### D. Alternating convex minimization of $f$

The previous propositions imply that we can derive an explicit alternating convex minimization scheme for  $f$ , presented in Algorithm 1. Starting with a given value of  $\boldsymbol{\Lambda}$ , at each step,  $\boldsymbol{\mu}^l$  and  $\mathbf{C}^l$  are computed according to Equations (6) and (7), then  $\boldsymbol{\Lambda}^l$  is updated according to (8). We show in Appendix A

---

#### Algorithm 1: Alternating convex minimization for $f$

---

**Input:**  $\mathbf{Z}$ ,  $\mathbf{D}$ ,  $\boldsymbol{\mu}_0$ ,  $\boldsymbol{\Sigma}_0$ ,  $\kappa$ ,  $\nu$

**Output:**  $\{\hat{\mathbf{C}}_i\}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Lambda}}$

1 **Initialization:** Set  $\boldsymbol{\Lambda}^0 = \boldsymbol{\Sigma}_0^{-1}$

2 **for**  $l = 1$  to  $\text{maxIts}$  **do**

3     Compute  $(\mathbf{C}^l, \boldsymbol{\mu}^l) = \arg \min_{(\mathbf{C}, \boldsymbol{\mu})} f(\mathbf{C}, \boldsymbol{\mu}, \boldsymbol{\Lambda}^{l-1})$  by equations (6) and (7)

4     Compute  $\boldsymbol{\Lambda}^l = \arg \min_{\boldsymbol{\Lambda}} f(\mathbf{C}^l, \boldsymbol{\mu}^l, \boldsymbol{\Lambda})$  by Eq. (8)

5 **end**

6  $\{\hat{\mathbf{C}}_i = \mathbf{C}_i^{\text{maxIts}}\}, \hat{\boldsymbol{\mu}} = \boldsymbol{\mu}^{\text{maxIts}}, \hat{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}^{\text{maxIts}}$

---

the following convergence result for the previous algorithm. The proof adapts the arguments in [21] to our particular case.

**Proposition 3.** The sequence  $f(\{\mathbf{C}_i^l\}, \boldsymbol{\mu}^l, \boldsymbol{\Lambda}^l)$  converges monotonically when  $l \rightarrow +\infty$ . The sequence  $\{\{\mathbf{C}_i^l\}, \boldsymbol{\mu}^l, \boldsymbol{\Lambda}^l\}$  generated by the alternate minimization scheme has at least

<sup>2</sup>The Normal-Wishart distribution is the conjugate prior of a multivariate normal distribution with unknown mean and covariance matrix.  $\mathcal{W}$  denotes the Wishart distribution [19].

<sup>3</sup>We rely on the classical independence assumption made in the patch-based literature, even if it is wrong in case patches overlap.



one accumulation point. The set of its accumulation points forms a connected and compact set of partial optima and stationary points of  $f$ , and they all have the same function value.

In practice, we observe in our experiments that the algorithm always converges after a few iterations.

#### E. Full restoration algorithm

The full restoration algorithm used in our experiments is summarized in Algorithm 2 and illustrated by Figure 1. It alternates between two stages: the minimization of  $f$  using Algorithm 1, and the estimation of the hyper-parameters  $\mu_0, \Sigma_0$ . In order to estimate these parameters, we rely on an *oracle image* computed by aggregation of all the patches estimated on the first stage (details are provided in Section III-D).

---

#### Algorithm 2: HBE algorithm.

---

**Input:**  $\mathbf{Z}, \mathbf{D}, \mu_0, \Sigma_0, \kappa, \nu$  (see details in Section III-C)

**Output:**  $\tilde{\mathbf{C}}$

- 1 Decompose  $\mathbf{Z}$  and  $\mathbf{D}$  into overlapping patches.
  - 2 **Initialization:** Compute first oracle image  $\mathbf{C}_{\text{oracle}}$  (see details in Section III-D)
  - 3 **for**  $it = 1$  to  $\text{maxIts}_2$  **do**
  - 4     **for all patches not yet restored do**
  - 5         Find patches similar ( $L^2$  distance) to the current  $\mathbf{z}_i$  in  $\mathbf{C}_{\text{oracle}}$  (see details in Section III-A).
  - 6         Compute  $\mu_0$  and  $\Sigma_0$  from  $\mathbf{C}_{\text{oracle}}$  (see details in Section III-C).
  - 7         Compute  $(\{\hat{\mathbf{C}}_i\}, \hat{\mu}, \hat{\Sigma})$  following Algorithm 1.
  - 8     **end**
  - 9     Perform aggregation to restore the image.
  - 10    Set  $\mathbf{C}_{\text{oracle}} = \tilde{\mathbf{C}}$ .
  - 11 **end**
- 

### III. IMPLEMENTATION DETAILS

#### A. Search for similar patches

The similar patches are all patches within a search window centered at the current patch, whose  $L_2$  distance to the central patch is less than a given threshold. This threshold is given by a tolerance parameter  $\varepsilon$  times the distance to the nearest neighbour (the most similar one). In all our experiments, the search window was set to size  $25 \times 25$  (with a patch size of  $8 \times 8$ ) and  $\varepsilon = 1.5$ . The patch comparison is performed in an oracle image (i.e. the result of the previous iteration), so all pixels are known. However, it may be useful to assign different confidence levels to the known pixels and to those originally missing and then restored. For all the experimental results presented in Section IV, the distance between patches  $\mathbf{c}_p$  and  $\mathbf{c}_q$  in the oracle image  $\mathbf{C}_{\text{oracle}}$  is computed as

$$d(p, q) = \frac{\sum_{j=1}^N (\mathbf{c}_p^j - \mathbf{c}_q^j)^2 \omega_{p,q}^j}{\sum_{j=1}^N \omega_{p,q}^j}, \quad (9)$$

where  $j$  indexes the pixels in the patch,  $\omega_{p,q}^j = 1$  if  $\mathbf{D}_p^j = \mathbf{D}_q^j = 1$  (known pixel) and  $\omega_{p,q}^j = 0.01$  otherwise

(originally missing then restored pixel) [22]. With this formulation, known pixels are assigned a much higher priority than unknown ones. Variations on these weights could be explored.

#### B. Optional speed-up by Collaborative Filtering

The proposed method computes one Gaussian model per image patch according to Equations (6) and (8). In order to reduce the computational cost, we can rely on the collaborative filtering idea previously introduced for patch-based denoising techniques [7], [23]. Based on the hypothesis that similar patches share the same model, we assign the same model to all patches in the set of similar patches (as defined in Section III-A).

The number of similar patches jointly restored depends on the image and the tolerance parameter  $\varepsilon$ , but it is often much smaller than what would result from the patch clustering performed by methods that use global GMMs such as PLE or EPLL. Performance degradation is observed in practice when using a very large tolerance parameter ( $\varepsilon > 3$ ), showing that mixing more patches than needed is detrimental. The collaborative filtering strategy helps accelerating the algorithm up to a certain point, but a trade-off with performance needs to be considered.

#### C. Parameter choices

The four parameters of the Normal-Wishart distribution:  $\kappa$ ,  $\nu$ , the prior mean  $\mu_0$  and the prior covariance matrix  $\Sigma_0$ , must be set in order to compute  $\mu$  and  $\Sigma$ .

a) *Choice of  $\kappa$  and  $\nu$ :* The computation of  $\mu$  according to (6) combines the mean  $\sum_{i=1}^M \mathbf{A}_i \mathbf{Z}_i$  estimated from the similar patches and the prior mean  $\mu_0$ . The parameter  $\kappa$  is related to the degree of confidence we have on the prior  $\mu_0$ . Hence, its value should be a trade-off between the confidence we have in the prior accuracy vs. the one we have in the information provided by the similar patches. The latter improves when both  $M$  (i.e. the number of similar patches) and  $P = \text{trace}(\mathbf{D}_i)$  (i.e. the number of known pixels in the current patch) increase. These intuitive insights suggest the following rule to set  $\kappa$ :

$$\kappa = M\alpha, \quad \alpha = \begin{cases} \alpha_L & \text{if } P \text{ and } M > \text{threshold} \\ \alpha_H & \text{otherwise.} \end{cases} \quad (10)$$

A similar reasoning leads to the same rule for  $\nu$ ,

$$\nu = M\alpha + n \quad (11)$$

where the addition of  $n$  ensures the condition  $\nu > n - 1$  required by the Normal-Wishart prior to be verified.

This rule is used to obtain the experimental results presented in Section IV, and proved to be a consistently good choice despite its simplicity. However, setting these parameters in a more general setting is not a trivial task and should be the subject of further study. In particular we could explore a more continuous dependence of  $\alpha$  on  $P$ ,  $M$ , and possibly a third term  $Q = \sum_{i=1}^n S_{ii}$  where  $S = \sum_{j=1}^M \mathbf{\Lambda}^{-1} \mathbf{D}_j \mathbf{\Lambda}_j^* \mathbf{D}_j$ . This term estimates to what an extent similar patches cover the missing pixels in the current patch.

b) *Setting of  $\mu_0$  and  $\Sigma_0$* : Assuming an oracle image  $\mathbf{C}_{\text{oracle}}$  is available (see details in Section II-E),  $\mu_0$  and  $\Sigma_0$  can be computed using the classical MLE estimators from a set of similar patches  $(\tilde{\mathbf{c}}_1, \dots, \tilde{\mathbf{c}}_M)$  taken from  $\mathbf{C}_{\text{oracle}}$

$$\mu_0 = \frac{1}{M} \sum_{j=1}^M \tilde{\mathbf{c}}_j, \quad \Sigma_0 = \frac{1}{M-1} \sum_{j=1}^M (\tilde{\mathbf{c}}_j - \mu_0)(\tilde{\mathbf{c}}_j - \mu_0)^T. \quad (12)$$

This is the same approach followed by Lebrun et al. [7] to estimate the patch model parameters in the case of denoising.

#### D. Initialization

A good initialization is crucial since we aim at solving a non-convex problem through an iterative procedure. To initialize the proposed algorithm we follow the approach proposed by Yu et al. [10] (described in detail in Appendix A in the supplementary material). They propose to initialize the PLE algorithm by learning the  $K$  GMM covariance matrices from synthetic images of edges with different orientations as well as from the DCT basis to represent isotropic patterns. As they state, in dictionary learning, the most prominent atoms represent local edges which are useful to represent and restore contours. Hence, this initialization helps to correctly restore corrupted patches even in quite extreme cases. The oracle of the first iteration of the proposed approach is the output of the first iteration of the PLE algorithm.

#### E. Computational complexity

With the original per-patch strategy, the complexity of the algorithm is given by step 3 in Algorithm 1:  $[(4n^3 + n^3/3)M + n^3/3]$ , so the total complexity is  $[(4n^3 + n^3/3)M + n^3/3] \times \max Its \times \max Its_2 \times T$  (where  $T$  = total number of patches to be restored and assuming the Cholesky factorization is used for matrix inversion). The collaborative filtering strategy reduces this value by a factor that depends on the number of groups of similar patches, which depends on the image contents and the distance tolerance parameter  $\varepsilon$ . The main difference with the PLE algorithm complexity  $((3n^3 + n^3/3) \times its_{PLE} \times T)$  is a factor given by the number of groups defined by the collaborative filtering approach and the ratio between  $its_{PLE}$  and  $\max Its \times \max Its_2$ . As mentioned by Yu et al. [10], computational complexity can be further reduced in the case of binary masks by removing the zero rows and inverting a matrix of size  $n^2/S \times n^2/S$  instead of  $n^2 \times n^2$  where  $S$  is the masking ratio. Moreover, the proposed algorithm can be run in parallel in different image subregions thus allowing for even further acceleration in multiple-core architectures. The complexity comparison to NLB needs to be made in the case where the degradation is additive noise with constant variance (translation invariant degradation), which is the task performed by NLB. In that case, the complexity of the proposed approach (without considering collaborative filtering nor parallelization, which are both done also in NLB), is  $11n^3/3 \times \max Its \times \max Its_2 \times T$  whereas that of NLB is  $2 \times (4n^3/3)$ .

## IV. IMAGE RESTORATION EXPERIMENTS

In this section we illustrate the ability of the proposed method to solve several image inverse problems. Both synthetic (i.e., where we have added the degradation artificially) and real data (i.e., issued from a real acquisition process) are used. The considered problems are: interpolation, combined interpolation and denoising, denoising, and zooming. The reported values of peak signal-to-noise ratio ( $PSNR = 20 \log_{10}(255/\sqrt{MSE})$ ) are averaged over 10 realizations for each experiment (variance is below 0.1 for interpolation and below 0.05 for combined interpolation and denoising and for denoising only). Similar results are obtained with the structural similarity index (SSIM) which is included in the supplementary material (Appendix B).

#### A. Synthetic degradation

a) *Interpolation*: Random masks with 50%, 70% and 90% missing pixels are applied to the tested ground-truth images. The interpolation performance of the proposed method is compared to that of PLE [10], EPLL [9] and E-PLE [25] using a patch size of  $8 \times 8$  for all methods. PLE parameters are set as indicated in [10] ( $\sigma = 3$ ,  $\varepsilon = 30$ ,  $K = 19$ ). We used the EPLL code provided by the authors [24] with default parameters and the E-PLE code available in [25] with the parameters set as specified in the companion demo. The parameters for the proposed method are set to  $\alpha_H = 1$ ,  $\alpha_L = 0.5$  ( $\alpha_H$  and  $\alpha_L$  define the values for  $\kappa$  and  $\nu$ , see Section III-C). The PSNR results are shown in Table I. Figure 2 shows some extracts of the obtained results, the PSNR values for the extracts and the corresponding difference images with respect to the ground-truth. The proposed method gives sharper results than the other considered methods. This is specially noticeable on the reconstruction of the texture of the fabric of Barbara's trousers shown in the first row of Figure 2 or on the strips that appear through the car's window shown in the second row of the same figure.

b) *Combined interpolation and denoising*: For this experiment, the ground-truth images are corrupted with additive Gaussian noise with variance 10, and a random mask with 70% and 90% of missing pixels. The parameters for all methods are set as in the previous interpolation-only experiment. Table I summarizes the PSNR values obtained by each method. Figure 3 shows some extracts of the obtained results, the PSNR values for the extracts and the corresponding difference images with respect to the ground-truth. Once again, the results show that the proposed approach outperforms the others. Fine structures, such as the mast and the ropes of the ship, as well as textures, as in Barbara's headscarf, are much better preserved.

c) *Denoising*: For the denoising task the proposed approach should perform very similarly to the state-of-the-art denoising algorithm NLB [7]. The following experiments are conducted in order to verify this.

The ground-truth images are corrupted with additive Gaussian noise with variance  $\sigma^2 = 10, 30, 50, 80$ . The code provided by the authors [26] automatically sets the NLB parameters from the input  $\sigma^2$  and the patch size, in this case

Interpolation - PSNR (dB)													Interpolation & Denoising - PSNR (dB)							
50%				70%				90%					70%				90%			
	HBE	PLE	EPLL	E-PLE	HBE	PLE	EPLL	E-PLE	HBE	PLE	EPLL	E-PLE	HBE	PLE	EPLL	E-PLE	HBE	PLE	EPLL	E-PLE
barbara	<b>39.11</b>	36.93	32.99	35.43	<b>34.69</b>	32.50	27.96	28.77	<b>24.86</b>	23.62	23.30	23.26	<b>33.34</b>	31.99	27.63	27.75	<b>24.57</b>	23.53	23.27	23.20
boat	<b>34.92</b>	34.32	34.21	33.59	<b>31.37</b>	30.74	30.38	30.26	<b>25.96</b>	25.35	24.72	25.43	<b>30.61</b>	30.41	30.15	29.54	<b>25.78</b>	25.45	24.71	25.47
traffic	30.17	30.12	<b>30.19</b>	28.86	<b>27.27</b>	27.12	27.13	26.64	<b>22.84</b>	22.34	21.85	22.27	26.99	26.98	<b>27.05</b>	26.35	<b>22.87</b>	22.43	22.21	22.35
Denoising - PSNR (dB)													Zooming - PSNR (dB)							
$\sigma^2$	10				30				50				80				$\times 2$			
	HBE	NLB	EPLL	HBE	NLB	EPLL	HBE	NLB	EPLL	HBE	NLB	EPLL	HBE	PLE	EPLL	E-PLE	Lanczos			
barbara	<b>41.26</b>	41.20	40.56	<b>38.40</b>	38.26	37.32	<b>37.13</b>	36.94	35.84	<b>35.96</b>	35.73	34.51	<b>38.17</b>	37.11	31.34	36.51	28.01			
boat	<b>40.05</b>	39.99	39.47	36.71	<b>36.76</b>	36.34	35.41	<b>35.46</b>	35.13	34.30	<b>34.33</b>	34.12	<b>32.35</b>	31.96	31.95	32.08	29.60			
traffic	40.73	<b>40.74</b>	40.55	<b>37.03</b>	36.99	36.86	<b>35.32</b>	35.26	35.20	<b>33.78</b>	33.70	33.72	25.05	24.78	<b>25.17</b>	24.91	21.89			

TABLE I

RESULTS OF THE INTERPOLATION, COMBINED INTERPOLATION AND DENOISING, DENOISING AND ZOOMING TESTS DESCRIBED IN SECTION IV-A. PATCH SIZE OF  $8 \times 8$  FOR ALL METHODS IN ALL TESTS. PARAMETER SETTING FOR INTERPOLATION, COMBINED INTERPOLATION AND DENOISING, AND ZOOMING, HBE:  $\alpha_H = 1$ ,  $\alpha_L = 0.5$ , PLE:  $\sigma = 3$ ,  $\varepsilon = 30$ ,  $K = 19$  [10], EPLL: DEFAULT PARAMETERS [24], E-PLE: PARAMETERS SET AS SPECIFIED IN [25]. PARAMETER SETTING FOR DENOISING, HBE:  $\alpha_H = \alpha_L = 100$ , NLB: CODE PROVIDED BY THE AUTHORS [26] AUTOMATICALLY SETS PARAMETERS FROM INPUT  $\sigma^2$ , EPLL: DEFAULT PARAMETERS FOR THE DENOISING EXAMPLE [24]

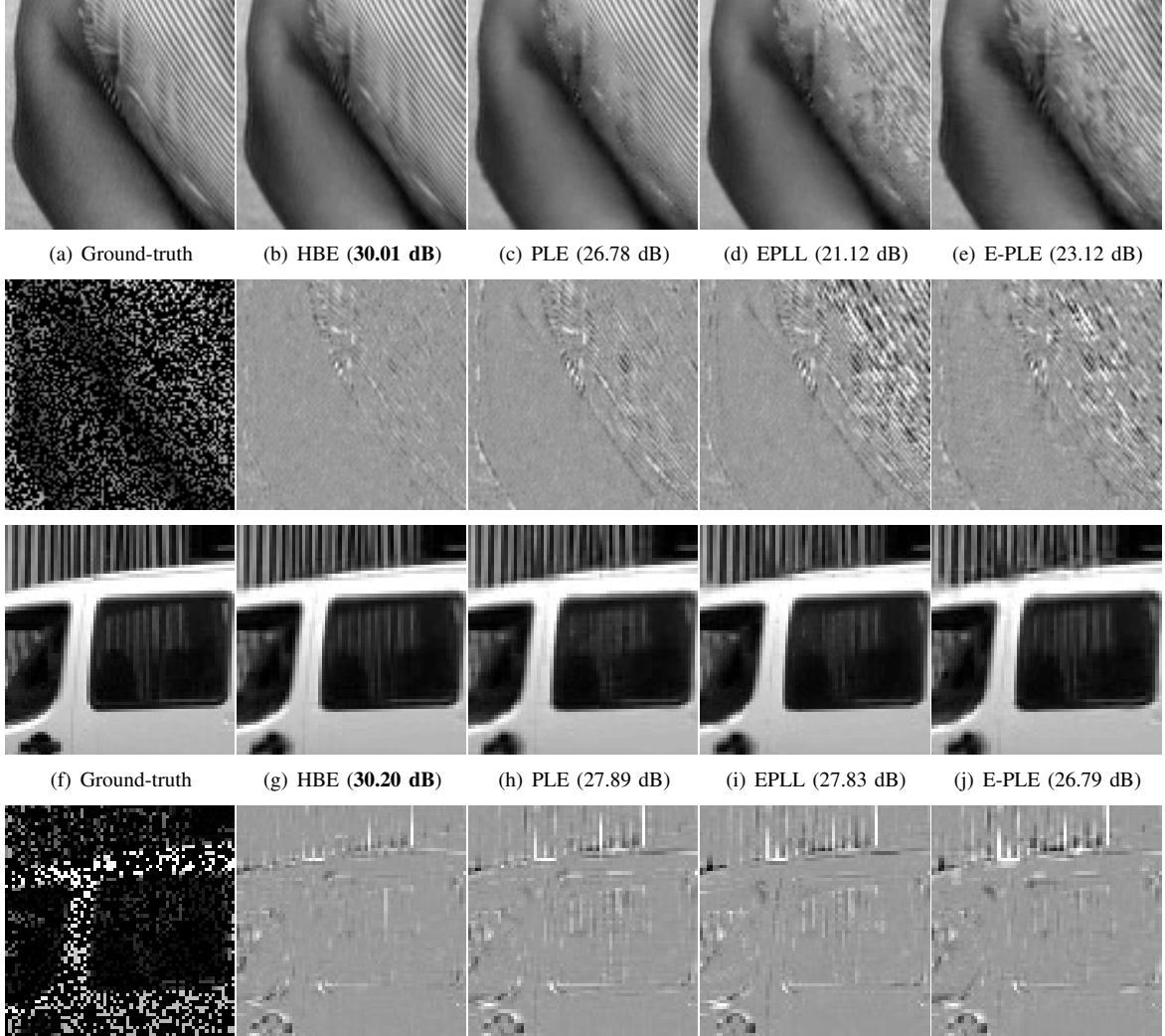


Fig. 2. Synthetic data. Interpolation with 70% of randomly missing pixels. Left to right: (first row) Ground-truth (extract of barbara), result by HBE, PLE, EPLL, E-PLE. (second row) input image, difference with respect to the ground-truth of each of the corresponding results. (third and fourth row) Idem for an extract of the traffic image. See Table I for the PSNR results for the complete images. Please see the digital copy for better details reproduction.

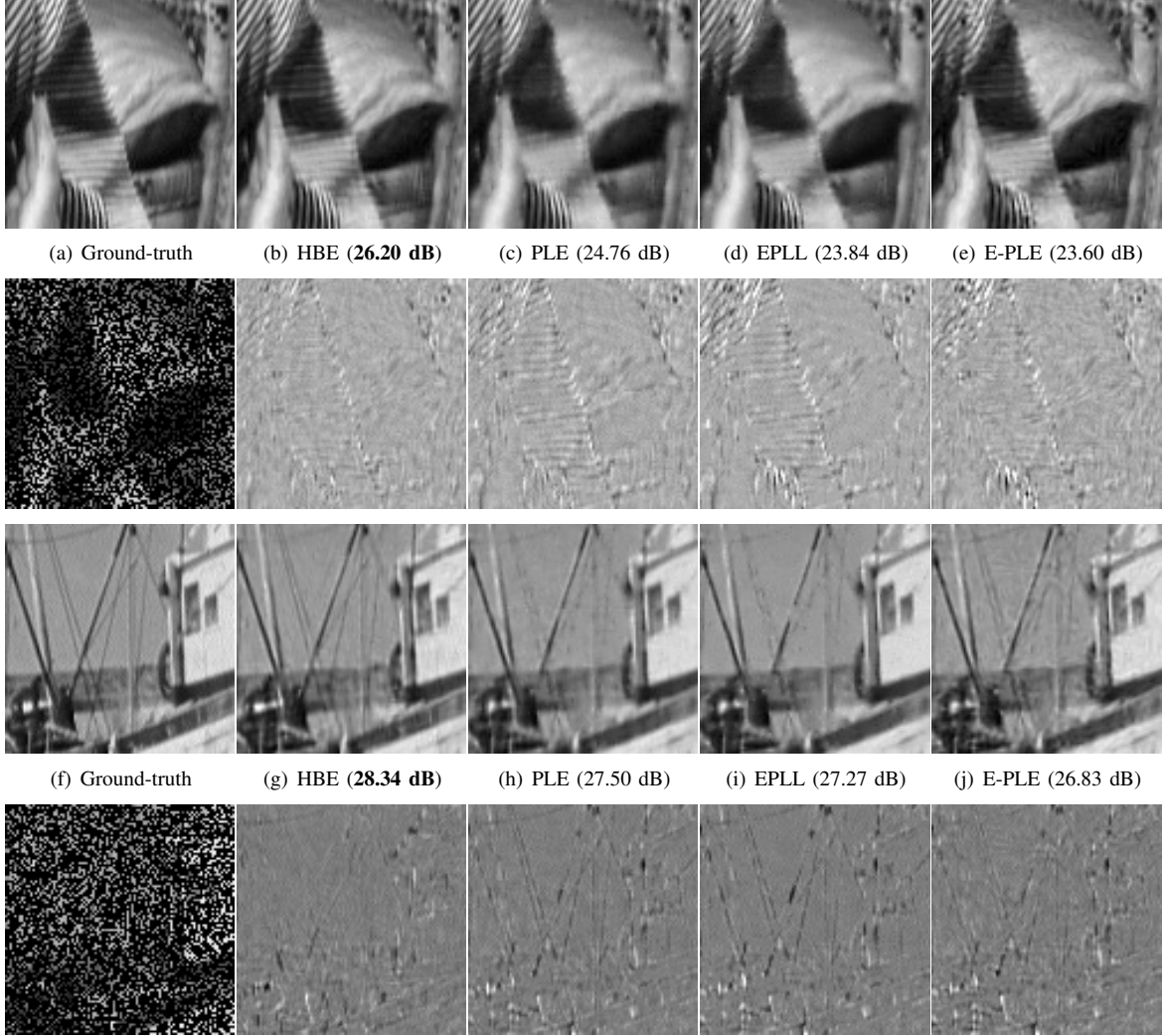


Fig. 3. **Synthetic data. Combined interpolation and denoising with 70% of randomly missing pixels and additive Gaussian noise ( $\sigma^2 = 10$ ).** **Left to right:** (first row) Ground-truth (extract of barbara), result by HBE, PLE, EPLL, E-PLE. (second row) input image, difference with respect to the ground-truth of each of the corresponding results. (third and fourth row) Idem for an extract of the boat image. See Table I for the PSNR results for the complete images. Please see the digital copy for better details reproduction.

$8 \times 8$ . For this experiment, there are no unknown pixels to interpolate (the mask  $D$  is the identity matrix).

The results of both methods are very similar if HBE is initialized with the output of the first step of NLB [7] (instead of using the initialization described in Section II-E) and the parameters  $\kappa$  and  $\nu$  are large enough. In this case,  $\mu_0$  and  $\Sigma_0$  are prioritized in equations (6) and (8) and both algorithms are almost the same. That is what we observe in practice with  $\alpha_H = \alpha_L = 100$ , as demonstrated in the results summarized in Table I. The denoising performance of HBE is degraded for small  $\kappa$  and  $\nu$  values. This is due to the fact that  $\mu_0$  and  $\Sigma_0$ , as well as  $\mu$  and  $\Sigma$  in NLB, are computed from an oracle image resulting from the first restoration step. This restoration includes not only the denoising of each patch, but also an aggregation step that greatly improves the final result. Therefore, the contribution of the first term of (6) to the computation of  $\hat{\mu}$  degrades the result compared to that of using  $\mu_0$  only (i.e. using a large  $\kappa$ ).

*d) Zooming:* In order to evaluate the zooming capacity of the proposed approach, ground-truth images are downsampled by a factor 2 (no anti-aliasing filter is used) and the zooming is compared to the ground-truth. The results are compared with PLE, EPLL, E-PLE and Lanczos interpolation. Table I summarizes the obtained PSNR values. Figure 4 shows extracts from the obtained results, the PSNR values for the extracts and the corresponding difference images with respect to the ground-truth. Again, HBE yields a sharper reconstruction than the other methods.

### B. Real data

For this experiment, we use raw images captured with a Canon 400D camera (set to ISO 400 and exposure time 1/160 seconds). The main noise sources for CMOS sensors are: the Poisson photon shot noise, which can be approximated by a Gaussian distribution with equal mean and variance; the thermally generated readout noise, which is modeled as an

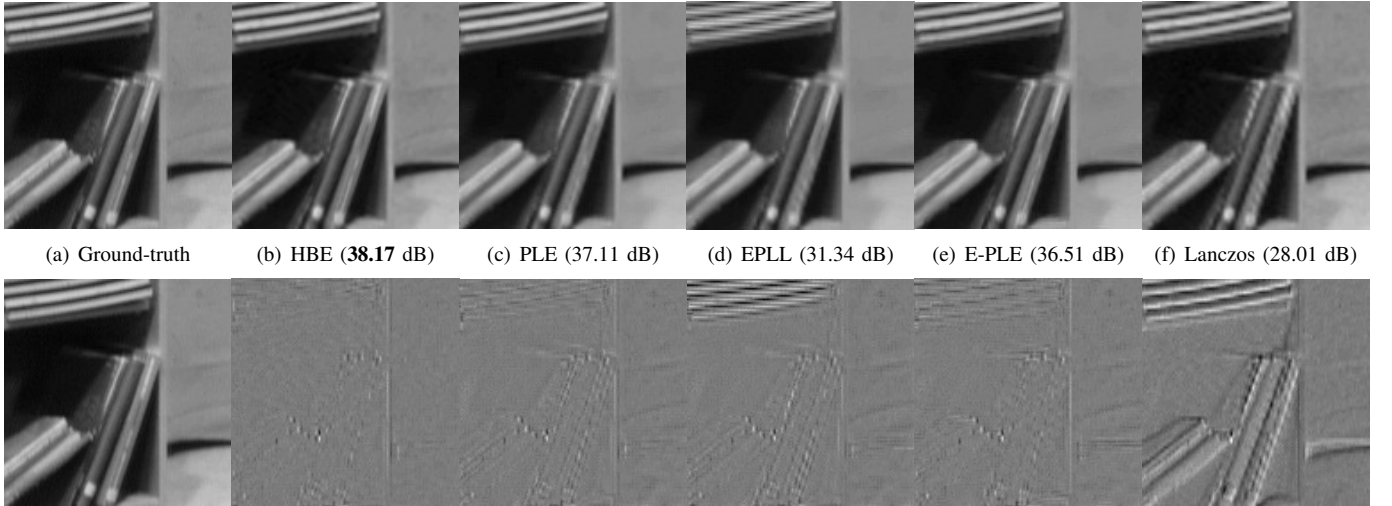


Fig. 4. **Synthetic data. Zooming  $\times 2$ .** Left to right: (first row) Ground-truth high resolution image (extract of barbara). Result by HBE, PLE, EPLL, E-PLE, lanczos interpolation. (second row) Input low-resolution image, difference with respect to the ground-truth of each of the corresponding results. Please see the digital copy for better details reproduction.



Fig. 5. **Left. Real data.** JPEG version of the raw image used in the experiments presented in Section IV-B. The boxes show the extracts displayed in Figure 6. **Right. Synthetic data.** Ground-truth images used in the experiments presented in Section IV-A. The green boxes indicate the extracts used for the zooming experiments.

additive Gaussian distributed noise and the spatially varying gain given by the photo response non uniformity (PRNU) [20], [27]. We thus consider the following noise model for the non saturated raw pixel value  $Z(p)$  at position  $p$

$$Z(p) \sim \mathcal{N}(g a_p \tau C(p) + \mu_R, g^2 a_p \tau C(p) + \sigma_R^2), \quad (13)$$

where  $g$  is the camera gain,  $a_p$  models the PRNU factor,  $\tau$  is the exposure time,  $C(p)$  is the irradiance reaching pixel  $p$ ,  $\mu_R$  and  $\sigma_R^2$  are the readout noise mean and variance. The camera parameters have to be estimated by a calibration procedure [20]. The noise covariance matrix  $\Sigma_N$  is thus diagonal with entries that depend on the pixel value  $(\Sigma_N)_p = g^2 a_p \tau C(p) + \sigma_R^2$ .

In order to evaluate the interpolation capacity of the proposed approach, we consider the pixels of the green channel only (i.e. 50% of the pixels in the RGGB Bayer pattern) and interpolate the missing values. We compare the results to those obtained using an adaptation of PLE to images degraded with noise with variable variance (PLEV) [28]. The results for the EPLL and E-PLE methods are not presented here since these methods are not suited for this kind of noise. Figure 6 shows extracts of the obtained results (see Figure 5 for a JPEG version of the raw image showing the location

of the extracts). As it was already observed in the synthetic data experiments, fine details and edges are better preserved. Compare for example the reconstruction of the balcony edges and the wall structure in the first row of Figure 6, as well as the structure of the roof and the railing in the second row of the same image.

### C. Discussion

In all the previous experiments, the results obtained with HBE outperform or are very close to those obtained by the other evaluated methods. Visually, details are better reconstructed and images are sharper.

We interpret this as the inability of a fixed set of patch classes (19 for PLE) to accurately represent patches that seldom appear in the image, such as edges or textures (as in Barbara's trouser). The fact that methods such as PLE are actually semi-local (classes are estimated on  $128 \times 128$  regions [10]) does not solve this issue. On the contrary, a local model estimation as the one performed by HBE correctly handles those cases.

The performance difference is much more remarkable for the higher masking rates. In such cases, the robustness of the estimation is crucial. Indeed the proposed method performs the estimation from similar patches in a local window. The hypothesis of self-similarity being reinforced by considering local neighbourhoods, such a strategy restricts the possible models, therefore making the estimation more robust. Furthermore, the local model estimation, previously shown to be successful at describing patches [7], gives a better reconstruction even when a very large part of the patch is missing.

EPLL uses more mixture components (200 components are learnt from  $2 \times 10^6$  patches of natural images [9]) in its GMM model than PLE. It was observed in [8] that this strategy is less efficient than PLE for the denoising task. In this work, we also observe that the proposed approach outperforms EPLL, not only in denoising, but also in inpainting and zooming.



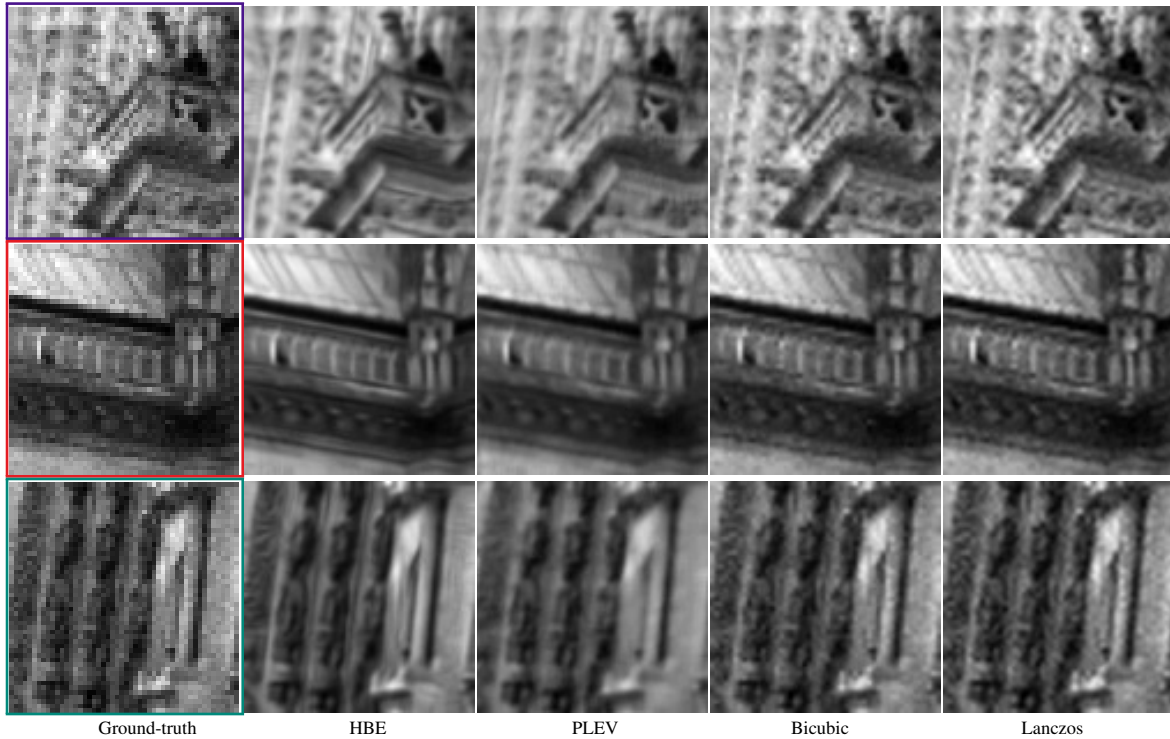


Fig. 6. **Real data. Zooming  $\times 2$ .** Interpolation of the green channel of a raw image (RGGGB). **Left to right:** Input low-resolution image, result by HBE, PLEV [28], bicubic and lanczos interpolation.

However, here it is harder to tell if the improvement is due to the local model estimation performed from similar patches or to the different restoration strategies followed by these methods.

## V. HIGH DYNAMIC RANGE IMAGING FROM A SINGLE SNAPSHOT

In this section, we apply the proposed framework to generate HDR images from a single shot. HDR imaging aims at reproducing an extended dynamic range of luminosity compared to what can be captured using a standard digital camera, which is often not enough to produce an accurate representation of real scenes. In the case of a static scene and a static camera, the combination of multiple images with different exposure levels is a simple and efficient solution [27], [29], [30]. However, several problems arise when either the camera or the elements in the scene move [31], [32].

An alternative to the HDR from multiple frames is to use a single image with spatially varying pixel exposures (SVE) [18]. An optical mask with spatially varying transmittance is placed adjacent to a conventional image sensor, thus controlling the amount of light that reaches each pixel (see Figure 7) [18], [33], [34].

The greatest advantage of this acquisition method is that it avoids the need for image alignment and motion estimation. Another advantage is that the saturated pixels are not organized in large regions. Indeed, some recent multi-image methods tackle motion problems by taking a reference image and then by estimating motion or reconstructing the image relative to this reference [31], [35]. A problem encountered

by these approaches is the need to inpaint very large saturated and underexposed regions in the reference frame. The SVE acquisition strategy avoids this problem since, in general, all scene regions are sampled by at least one of the exposures.

Taking advantage of the ability of the proposed framework to simultaneously estimate missing pixels and denoise well-exposed ones, we propose a novel approach to generate HDR images from a single shot acquired with spatially varying pixel exposures. The proposed approach shows significant improvements over existing methods.

### A. Spatially varying exposure acquisition model

An optical mask with spatially varying transmittance is placed adjacent to a conventional image sensor to give different exposure levels to the pixels. This optical mask does not change the acquisition process of the sensor. Hence, the noise model (13) can be adapted to the SVE acquisition by including the per-pixel SVE gain  $o_p$ <sup>4</sup>:

$$Z(p) \sim \mathcal{N}(go_p a_p \tau C(p) + \mu_R, g^2 o_p a_p \tau C(p) + \sigma_R^2). \quad (14)$$

In the approach proposed by Nayar and Mitsunaga [18], the varying exposures follow a regular pattern. Motivated by the aliasing problems of regular sampling patterns, Schöberl et al. [36] propose to use spatially varying exposures on a non-regular pattern. Figure 7 shows examples of both acquisition patterns. This observation led us to choose the non-regular pattern in the proposed approach.

<sup>4</sup>Some noise sources not modeled here, such as blooming, might have an impact in the SVE acquisition strategy and should be considered in a more accurate image modeling.

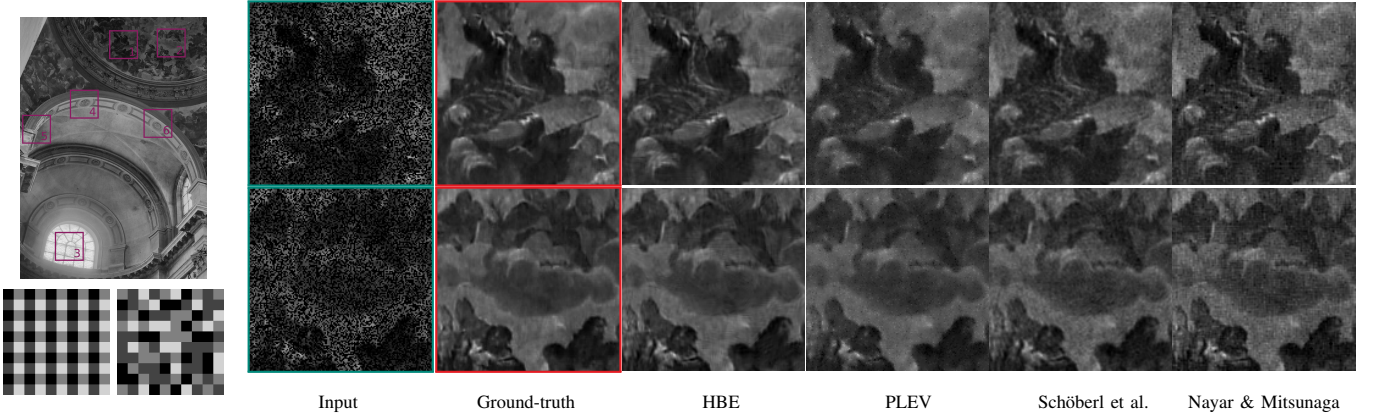


Fig. 7. **Synthetic data.** Left: (top) Tone mapped version of the ground-truth image used for the experiments in Section V-C1. (bottom) Regular (left) and non-regular (right) optical masks for an example of 4 different filters. Right: Results for extracts 1 and 6. From left to right: Input image with random pattern, ground-truth, results by HBE, PLEV [28], Schöberl et al. [36], Nayar and Mitsunaga [18]. 50% missing pixels (for both random and regular pattern). See PSNR values for these extracts in Table II. Please see the digital copy for better details reproduction.

### B. Hyperprior Bayesian Estimator for Single Shot High Dynamic Range Imaging

1) *Problem statement:* In order to reconstruct the dynamic range of the scene we need to solve an inverse problem. We want to estimate the irradiance image  $C$  from the SVE image  $Z$ , knowing the exposure levels of the optical mask and the camera parameters. For this purpose we map the raw pixel values to the irradiance domain  $Y$  with

$$Y(p) = \frac{Z(p) - \mu_R}{g o_p a_p \tau}. \quad (15)$$

We take into account the effect of saturation and under-exposure by introducing the exposure degradation matrix  $D$ , a diagonal matrix given by

$$(D)_p = \begin{cases} 1 & \text{if } \mu_R < Z(p) < z_{sat}, \\ 0 & \text{otherwise,} \end{cases} \quad (16)$$

with  $z_{sat}$  equal to the pixel saturation value. From (14) and (16),  $Y(p)$  can be modeled as

$$Y(p)|(D)_p \sim \mathcal{N}\left((D)_p C(p), \frac{g^2 o_p a_p \tau (D)_p C(p) + \sigma_R^2}{(g o_p a_p \tau)^2}\right). \quad (17)$$

Notice that (17) is the distribution of  $Y(p)$  for a given exposure degradation factor  $(D)_p$ , since  $(D)_p$  is itself a random variable that depends on  $Z(p)$ . The exposure degradation factor must be included in (17) since the variance of the over or under exposed pixels no longer depends on the irradiance  $C(p)$  but is only due to the readout noise  $\sigma_R^2$ . From (17) we have

$$Y = DC + N, \quad (18)$$

where  $N$  is zero-mean Gaussian noise with diagonal covariance matrix  $\Sigma_N$  given by

$$(\Sigma_N)_j = \frac{g^2 o_p a_p \tau (D)_p C(p) + \sigma_R^2}{(g o_p a_p \tau)^2}. \quad (19)$$

Then the problem of irradiance estimation can be stated as retrieving  $C$  from  $Y$ , which implies denoising the well-exposed pixel values ( $(D)_p = 1$ ) and estimating the unknown ones ( $(D)_p = 0$ ).

2) *Proposed solution:* From (18), image  $Y$  is under the hypothesis of the HBE framework introduced in Section II. The proposed HDR imaging algorithm consists of the following steps: 1) generate  $D$  from  $Z$  according to (16), 2) obtain  $Y$  from  $Z$  according to (15), 3) apply the HBE approach to  $Y$  with the given  $D$  and  $\Sigma_N$ .

### C. Experiments

The proposed reconstruction method was thoroughly tested in several synthetic and real data examples. A brief summary of the results is presented in this section.

1) *Synthetic data:* Sample images are generated according to Model (18) using the HDR image in Figure 7 as the ground-truth. Both a random and a regular pattern with four equiprobable exposure levels  $o = \{1, 8, 64, 512\}$  are simulated. The exposure time is set to  $\tau = 1/200$  seconds and the camera parameters are those of a Canon 7D camera set to ISO 200 ( $g = 0.87$ ,  $\sigma_R^2 = 30$ ,  $\mu_R = 2048$ ,  $v_{sat} = 15000$ ) [27].

Figure 7 shows extracts of the results obtained by the proposed method, by PLEV [28] (basically an adaptation of PLE to the same single image framework) and by Schöberl et al. [36] for the random pattern and by Nayar et Mitsunaga [18] using the regular pattern. The percentage of unknown pixels in the considered extracts is 50% (it is nearly the same for both the regular and non-regular pattern). Table II shows the PSNR values obtained in each extract marked in Figure 7. The proposed method manages to correctly reconstruct the irradiance on the unknown pixels. Moreover, its denoising performance is much better than that of Schöberl et al. and Nayar and Mitsunaga, and still sharper than PLEV.

2) *Real data:* The feasibility of the SVE random pattern has been shown in [34] and that of the SVE regular pattern in [33]. Nevertheless, these acquisition systems are still not available for general usage.<sup>5</sup> However, as stated in Section V-A, the only

<sup>5</sup>While writing the last version of this article the authors got aware of Sony's latest sensor IMX378. This sensor has a special mode called SME-HDR, which is a variation of the SVE acquisition principle. Whereas this sensor was adopted by the Google Pixel smartphone in 2016, the special SME-HDR mode is never activated by the Google Pixel phone, according to experts from the company DxO, and we found no way to activate it and access the raw image.



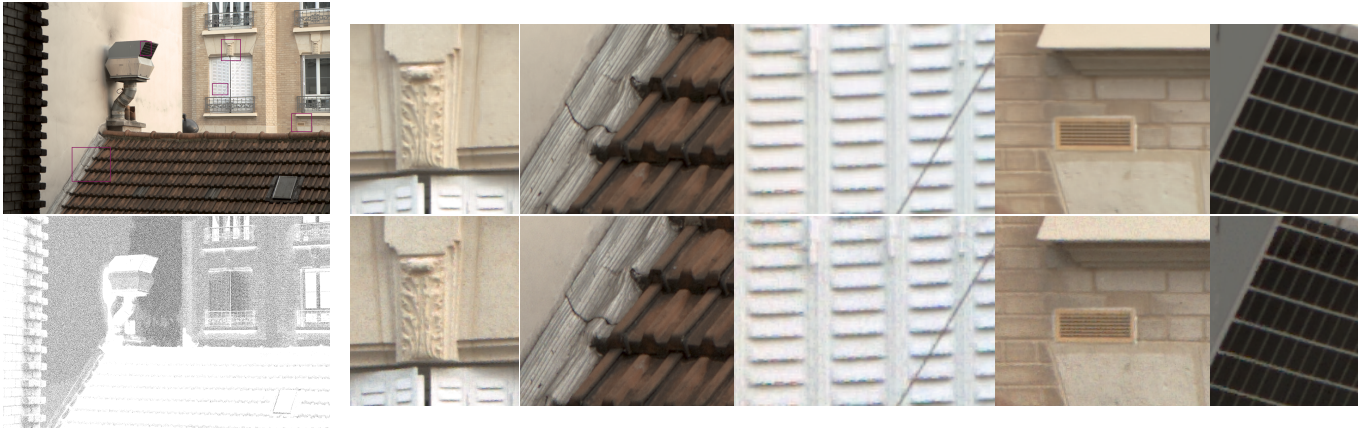


Fig. 8. **Real data.** **Left:** Tone mapped version of the HDR image obtained by the proposed approach and its corresponding mask of unknown (black) and well-exposed (white) pixels. **Right:** Comparison of the results obtained by the proposed approach (first row) and PLEV (second row) in the extracts indicated in the top image. Please see the digital copy for better details reproduction.

	PSNR (dB)					
	1 (Fig. 7)	2 (Fig. 7)	3	4	5	6
HBE	<b>33.08</b>	<b>33.87</b>	22.95	<b>35.10</b>	<b>36.80</b>	<b>35.66</b>
PLEV	29.65	30.82	22.77	33.99	36.42	34.73
Schöberl et al.	30.38	31.16	21.39	30.04	32.84	31.02
Nayar and Mitsunaga	29.39	30.10	<b>23.24</b>	25.83	30.26	26.90

TABLE II  
PSNR VALUES FOR THE EXTRACTS SHOWN IN FIGURE 7.

variation between the classical and the SVE acquisition is the optical filter. Hence, the noise at a pixel  $p$  captured using SVE with an optical gain factor  $o_p$  and exposure time  $\tau/o_p$  and a pixel captured with a classical camera using exposure time  $\tau$  should be very close. We take advantage of this fact in order to evaluate the reconstruction performance of the proposed approach using real data.

For this purpose, we generate an SVE image  $\mathbf{z}_{sve}$  from four raw images  $\{\mathbf{z}_{raw}^i\}_{i=1,\dots,4}$  acquired with different exposure times. The four different exposure times simulate four different filters of the SVE optical mask. The value at position  $(x, y)$  in  $\mathbf{z}_{sve}$  is chosen at random among the four available values at that position  $\{\mathbf{z}_{raw}^i(x, y)\}_{i=1,\dots,4}$ . Notice that the Bayer pattern is kept on  $\mathbf{z}_{sve}$  by construction. The images  $\{\mathbf{z}_{raw}^i\}_{i=1,\dots,4}$  are acquired using a remotely controlled camera and a tripod so as to be perfectly aligned. This protocol does not allow us to take scenes with moving objects. Let us emphasize, however, that using a real SVE device, this, as well as the treatment of moving camera, would be a non-issue.

Figures 8 and 9 show the results obtained from two real scenes, together with the masks of well-exposed (white) and unknown (black) pixels (the SVE raw images are included in Appendix B in the supplementary material). Recall that among the unknown pixels, some of them are saturated and some of them are under exposed. Square patches of size 6 and 8 were used for the examples in Figure 9 and Figure 8 respectively. Demosaicing [37] and tone mapping [38] are used for displaying purposes.

We compare the results to those obtained by PLEV [28]. A comparison against the methods by Nayar and Mitsunaga

and Schöberl et al. is not presented since they do not specify how to treat raw images with a Bayer pattern. The proposed method manages to correctly reconstruct the unknown pixels even in extreme conditions where more than 70% of the pixels are missing, as for example the last extract in Figure 9.

These examples show the suitability of the proposed approach to reconstruct the irradiance information in both very dark and bright regions simultaneously. See for instance the example in Figure 9, where the dark interior of the building (which can be seen through the windows) and the highly illuminated part of another building are both correctly reconstructed (see the electronic version of the article for better visualization).

## VI. CONCLUSIONS

In this work we have presented a novel image restoration framework. It has the benefits of local patch characterization (that was key to the success of NLB as a state of the art denoising method), but manages to extend its use to more general restoration problems where the linear degradation operator is diagonal, by combining local estimation with Bayesian restoration based on hyperpriors. This includes problems such as zooming, inpainting and interpolation. In this way, all these restoration problems are set under the same framework. It does not include image deblurring or deconvolution, since the degradation operator is no longer diagonal. Correctly addressing deconvolution with large kernels with patch-based approaches and Gaussian prior models is a major challenge that will be the subject of future work.

We have presented a large series of experiments both on synthetic and real data that confirm the robustness of the proposed strategy based on hyperpriors. These experiments show that for a wide range of image restoration problems HBE outperforms several state-of-the-art restoration methods.

This work opens several perspectives. The first one concerns the relevance of the Gaussian patch model and its relation to the underlying image patches manifold. If this linear approximation has proven successful for image restoration, its full relevance in other areas remains to be explored, especially in all domains requiring to compare image patches. Another

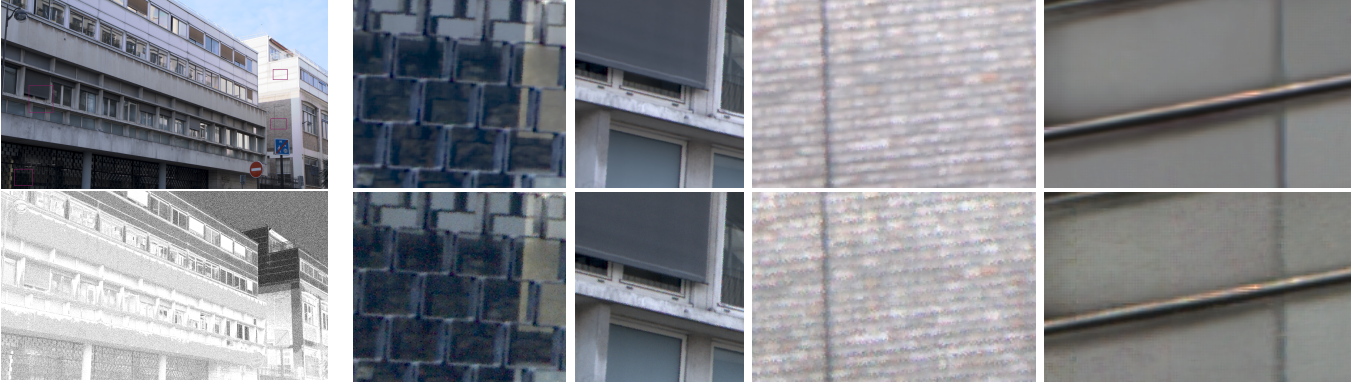


Fig. 9. **Real data. Left:** Tone mapped version of the HDR image obtained by the proposed approach and its corresponding mask of unknown (black) and well-exposed (white) pixels. **Right:** Comparison of the results obtained by the proposed approach (first row) and PLEV (second row) in the extracts indicated in the top image. Please see the digital copy for better details reproduction.

important related question is the one of the estimation of the degradation model in images jointly degraded by noise, missing pixels, blur, etc. Restoration approaches generally rely on the precise knowledge of this model and of its parameters. In practice however, we often deal with images for which the acquisition process is unknown, and that have possibly been affected by post-treatments. In such cases, blind restoration remains an unsolved challenge.

Finally, we have presented a novel application of the proposed general framework to the generation of HDR images from a single variable exposure (SVE) snapshot. The SVE acquisition strategy allows the creation of HDR images from a single shot without the drawbacks of multi-image approaches, such as the need for global alignment and motion estimation to avoid ghosting problems. The proposed method manages to simultaneously denoise and reconstruct the missing pixels, even in the presence of (possibly complex) motions, improving the results obtained by existing methods. Examples with real data acquired in very similar conditions to those of the SVE acquisition show the capabilities of the proposed approach.

## APPENDIX A

### ALTERNATE MINIMIZATION SCHEME CONVERGENCE

We study in the following the convergence of the alternate minimization algorithm 1. To show the main convergence result, we need the following lemma

**Lemma 1.** *The function  $f$  is coercive on  $\mathbb{R}^{n(M+1)} \times S_n^{++}(\mathbb{R})$ .*

*Proof.* We need to show that

$$\lim_{\|(\{C_i\}, \mu, \Lambda)\| \rightarrow +\infty} f(\{C_i\}, \mu, \Lambda) = +\infty.$$

Now,  $\|(\{C_i\}, \mu, \Lambda)\| \rightarrow +\infty$  if and only if  $\|C_i\| \rightarrow +\infty$  or  $\|\mu\| \rightarrow +\infty$  or  $\|\Lambda\| \rightarrow +\infty$ .

The matrix  $\Lambda$  being positive-definite, the terms  $\frac{1}{2} \sum_{i=1}^M (C_i - \mu)^T \Lambda (C_i - \mu)$  and  $\frac{\kappa}{2} (\mu - \mu_0)^T \Lambda (\mu - \mu_0)$  are both positive. Thus

$$\begin{aligned} f(\{C_i\}, \mu, \Lambda) &\geq -\frac{\nu - n + M}{2} \log |\Lambda| \\ &+ \frac{1}{2} \text{trace}[\nu \Sigma_0 \Lambda]. \end{aligned}$$

Now, this function of  $\Lambda$  is convex and coercive on  $S_n^{++}(\mathbb{R})$ , which implies that  $f(\{C_i\}, \mu, \Lambda) \rightarrow +\infty$  as soon as  $\|\Lambda\| \rightarrow +\infty$ . It also follows that the previous function of  $\Lambda$  has a global minimum that we denote by  $m_\Lambda$ . We can now write

$$\begin{aligned} f(\{C_i\}, \mu, \Lambda) &\geq m_\Lambda + \frac{1}{2} \sum_{i=1}^M (C_i - \mu)^T \Lambda (C_i - \mu) \\ &+ \frac{\kappa}{2} (\mu - \mu_0)^T \Lambda (\mu - \mu_0) \end{aligned}$$

and this function of  $(\{C_i\}, \mu)$  clearly tends towards  $+\infty$  as soon as  $\|C_i\| \rightarrow +\infty$  or  $\|\mu\| \rightarrow +\infty$ .  $\square$

We now show the main convergence result for our alternate minimization algorithm. The proof adapts the arguments in [21] to our case.

**Proposition 3.** *The sequence  $f(\{C_i^l\}, \mu^l, \Lambda^l)$  converges monotonically when  $l \rightarrow +\infty$ . The sequence  $\{\{C_i^l\}, \mu_l, \Lambda^l\}$  generated by the alternate minimization scheme has at least one accumulation point. The set of its accumulation points forms a connected and compact set of partial optima and stationary points of  $f$ , all having the same function value.*

*Proof.* The sequence  $f(\{C_i^l\}, \mu^l, \Lambda^l)$  obviously decreases at each step by construction. The coercivity and continuity of  $f$  imply that this sequence is also bounded from below, and thus converges. The convergence of  $f(\{C_i^l\}, \mu^l, \Lambda^l)$  implies that the sequence  $\{\{C_i^l\}, \mu_l, \Lambda^l\}$  is bounded. It follows that it has at least one accumulation point  $(\{C_i^*\}, \mu^*, \Lambda^*)$  and that there exists a strictly increasing sequence  $(l_k)_{k \in \mathbb{N}}$  of integers such that  $\{\{C_i^{l_k}\}, \mu^{l_k}, \Lambda^{l_k}\}_{k \in \mathbb{N}}$  converges to  $\{\{C_i^*\}, \mu^*, \Lambda^*\}$ .

Now, we can show that such an accumulation point is a partial optimum of  $f$ , i.e. that  $f(\{C_i^*\}, \mu^*, \cdot)$  attains its minimum at  $\Lambda^*$  and  $f(\cdot, \cdot, \Lambda^*)$  attains its minimum at  $(\{C_i^*\}, \mu^*)$ . By construction,

$$f(\{C^{l_k}\}, \mu^{l_k}, \Lambda^{l_k}) \leq f(\{C^{l_k}\}, \mu^{l_k}, \Lambda), \quad \forall \Lambda \in S_n^{++}(\mathbb{R})$$

which implies by continuity of  $f$  that

$$f(\{C^*\}, \mu^*, \Lambda^*) = \arg \min_{\Lambda \in S_n^{++}(\mathbb{R})} f(\{C^*\}, \mu^*, \Lambda). \quad (20)$$

Let us denote  $G(\{\mathbf{C}\}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = (\{\mathbf{C}'\}, \boldsymbol{\mu}', \boldsymbol{\Lambda}')$  with

$$\begin{aligned} (\{\mathbf{C}'\}, \boldsymbol{\mu}') &= \arg \min_{(\{\mathbf{C}\}, \boldsymbol{\mu})} f(\{\mathbf{C}\}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \\ \boldsymbol{\Lambda}' &= \arg \min_{\boldsymbol{\Lambda}} f(\{\mathbf{C}'\}, \boldsymbol{\mu}', \boldsymbol{\Lambda}). \end{aligned}$$

The alternate minimization scheme consists in updating  $G$  at each iteration. From Equations (6), (7), and (8), we see that  $G$  is explicit and continuous. Since  $\{l_k\}_{k \in \mathbb{N}}$  is strictly increasing, for each  $k \in \mathbb{N}$ ,  $l_k \geq l_{k-1} + 1$ . The sequence  $\{f(\{\mathbf{C}_i^l\}, \boldsymbol{\mu}^l, \boldsymbol{\Lambda}^l)\}_{l \in \mathbb{N}}$  decreases, so

$$\begin{aligned} f(G(\{\mathbf{C}^{l_{k-1}}\}, \boldsymbol{\mu}^{l_{k-1}}, \boldsymbol{\Lambda}^{l_{k-1}})) &= f(\{\mathbf{C}^{l_{k-1}+1}\}, \boldsymbol{\mu}^{l_{k-1}+1}, \boldsymbol{\Lambda}^{l_{k-1}+1}) \\ &\geq f(\{\mathbf{C}^{l_k}\}, \boldsymbol{\mu}^{l_k}, \boldsymbol{\Lambda}^{l_k}) \\ &\geq f(G(\{\mathbf{C}^{l_k}\}, \boldsymbol{\mu}^{l_k}, \boldsymbol{\Lambda}^{l_k})). \end{aligned}$$

Therefore, as  $k \rightarrow +\infty$ , since  $G$  is continuous, it follows that

$$f(G(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*)) = f(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*).$$

Now, writing  $(\{\mathbf{C}^{**}\}, \boldsymbol{\mu}^{**}, \boldsymbol{\Lambda}^{**}) = G(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*)$ , we get

$$\begin{aligned} f(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*) &\geq \arg \min_{(\mathbf{C}, \boldsymbol{\mu})} f(\{\mathbf{C}\}, \boldsymbol{\mu}, \boldsymbol{\Lambda}^*) = f(\{\mathbf{C}^{**}\}, \boldsymbol{\mu}^{**}, \boldsymbol{\Lambda}^*) \\ &\geq \arg \min_{\boldsymbol{\Lambda}} f(\{\mathbf{C}^{**}\}, \boldsymbol{\mu}^{**}, \boldsymbol{\Lambda}) = f(\{\mathbf{C}^{**}\}, \boldsymbol{\mu}^{**}, \boldsymbol{\Lambda}^{**}). \end{aligned}$$

We can conclude that all these terms are equal and in particular

$$f(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*) = f(\{\mathbf{C}^{**}\}, \boldsymbol{\mu}^{**}, \boldsymbol{\Lambda}^*) = \arg \min_{(\mathbf{C}, \boldsymbol{\mu})} f(\{\mathbf{C}\}, \boldsymbol{\mu}, \boldsymbol{\Lambda}^*) \quad (21)$$

From (20) and (21) we deduce that the accumulation point  $(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*)$  is a partial optimum of  $f$  and since  $f$  is differentiable, it is also a stationary point of  $f$ . Moreover, since  $f(\cdot, \cdot, \boldsymbol{\Lambda}^*)$  is strictly convex and has a unique minimum, it follows from (21) that  $(\{\mathbf{C}^{**}\}, \boldsymbol{\mu}^{**}) = (\{\mathbf{C}^*\}, \boldsymbol{\mu}^*)$ . As a consequence,  $\boldsymbol{\Lambda}^{**} = \arg \min_{\boldsymbol{\Lambda}} f(\{\mathbf{C}^{**}\}, \boldsymbol{\mu}^{**}, \boldsymbol{\Lambda}) = \arg \min_{\boldsymbol{\Lambda}} f(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}) = \boldsymbol{\Lambda}^*$ . Therefore, the accumulation point  $(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*)$  (and actually any accumulation point of the sequence) is also a fixed point of function  $G$ .

We have shown that accumulation points of the sequence  $\{\{\mathbf{C}^l\}, \boldsymbol{\mu}^l, \boldsymbol{\Lambda}^l\}$  are partial optima of  $f$  and fixed points of the function  $G$ . The set of accumulation points is obviously compact. Let us show that it is also a connected set. First, observe that whatever the norm  $\|\cdot\|$ , the sequence  $\|(\{\mathbf{C}^{l+1}\}, \boldsymbol{\mu}^{l+1}, \boldsymbol{\Lambda}^{l+1}) - (\{\mathbf{C}^l\}, \boldsymbol{\mu}^l, \boldsymbol{\Lambda}^l)\|$  converges to 0 when  $l \rightarrow \infty$ . If it was not the case, it would be possible to extract a subsequence  $(\{\mathbf{C}^{l_k}\}, \boldsymbol{\mu}^{l_k}, \boldsymbol{\Lambda}^{l_k})$  converging to an accumulation point  $(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*)$  while  $(\{\mathbf{C}^{l_k+1}\}, \boldsymbol{\mu}^{l_k+1}, \boldsymbol{\Lambda}^{l_k+1})$  converges to a different accumulation point  $(\{\mathbf{C}'\}, \boldsymbol{\mu}', \boldsymbol{\Lambda}')$ , but we know that it is impossible since  $(\{\mathbf{C}^{l_k+1}\}, \boldsymbol{\mu}^{l_k+1}, \boldsymbol{\Lambda}^{l_k+1}) = G(\{\mathbf{C}^{l_k}\}, \boldsymbol{\mu}^{l_k}, \boldsymbol{\Lambda}^{l_k})$  would also tend toward  $(\{\mathbf{C}^*\}, \boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*)$ . The sequence  $(\{\mathbf{C}^l\}, \boldsymbol{\mu}^l, \boldsymbol{\Lambda}^l)$  being bounded and such that  $\|(\{\mathbf{C}^{l+1}\}, \boldsymbol{\mu}^{l+1}, \boldsymbol{\Lambda}^{l+1}) - (\{\mathbf{C}^l\}, \boldsymbol{\mu}^l, \boldsymbol{\Lambda}^l)\|$  converges to 0, the set of its accumulation points is connected (see [39]). The fact that all accumulation points have the same function value is obvious since the sequence  $\{f(\{\mathbf{C}^l\}, \boldsymbol{\mu}^l, \boldsymbol{\Lambda}^l)\}_{l \in \mathbb{N}}$  decreases.  $\square$

## ACKNOWLEDGEMENT

We would like to thank the reviewers for their thorough and fruitful contributions, as well as Mila Nikolova and Alasdair Newson for their insightful comments and the authors of [24]–[26] for kindly providing their code. This work has been partially funded by the French Research Agency (ANR) under grant nro ANR-14-CE27-001 (MIRIAM), and by the “Investissement d’avenir” project, reference ANR-11-LABX-0056-LMH.

## REFERENCES

- [1] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.
- [2] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image denoising using scale mixtures of gaussians in the wavelet domain,” *IEEE Transactions on Image processing*, vol. 12, no. 11, pp. 1338–1351, 2003.
- [3] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [4] A. Buades, B. Coll, and J. M. Morel, “A Review of Image Denoising Algorithms, with a New One,” *SIAM Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.
- [5] S. Lyu and E. Simoncelli, “Modeling Multiscale Subbands of Photographic Images with Fields of Gaussian Scale Mixtures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 693–706, 2009.
- [6] P. Chatterjee and P. Milanfar, “Patch-Based Near-Optimal Image Denoising,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1635–1649, 2012.
- [7] M. Lebrun, A. Buades, and J. Morel, “A nonlocal bayesian image denoising algorithm,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1665–1688, 2013.
- [8] Y.-Q. Wang and J.-M. Morel, “SURE Guided Gaussian Mixture Image Denoising,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 999–1034, 2013.
- [9] D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration,” in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 479–486.
- [10] G. Yu, G. Sapiro, and S. Mallat, “Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity,” *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, 2012.
- [11] M. Lebrun, M. Colom, A. Buades, and J. Morel, “Secrets of image denoising cuisine,” *Acta Numerica*, vol. 21, no. 1, pp. 475–576, 2012.
- [12] Y.-Q. Wang, “The Implementation of SURE Guided Piecewise Linear Image Denoising,” *Image Processing On Line*, vol. 3, pp. 43–67, 2013.
- [13] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*. Chapman & Hall/CRC Boca Raton, FL, USA, 2014, vol. 2.
- [14] R. Molina, “On the hierarchical bayesian approach to image restoration: applications to astronomical images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 11, pp. 1122–1128, 1994.
- [15] R. Molina, A. K. Katsaggelos, and J. Mateos, “Bayesian and regularization methods for hyperparameter estimation in image restoration,” *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 231–246, 1999.
- [16] N. Dobigeon, J.-Y. Tourneret, and C.-I. Chang, “Semi-supervised linear spectral unmixing using a hierarchical bayesian model for hyperspectral imagery,” *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 2684–2695, 2008.
- [17] F. Orieux, J. Giovannelli, and T. Rodet, “Bayesian estimation of regularization and point spread function parameters for wiener–hunt deconvolution,” *J. Opt. Soc. Am. A*, vol. 27, no. 7, pp. 1593–1607, Jul 2010.
- [18] S. Nayar and T. Mitsunaga, “High Dynamic Range Imaging: Spatially Varying Pixel Exposures,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, Jun 2000, pp. 472–479.
- [19] H. Raiffa and R. Schlaifer, *Applied statistical decision theory*. Division of Research, Graduate School of Business Administration, Harvard University Boston, 1961.



- [20] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, “Study of the digital camera acquisition process and statistical modeling of the sensor raw data,” *Technical report hal-00733538v1*, 2012.
- [21] J. Gorski, F. Pfeuffer, and K. Klamroth, “Biconvex sets and optimization with biconvex functions: A survey and extensions,” *Mathematical Methods of Operations Research*, vol. 66, no. 3, pp. 373–407, 2007.
- [22] P. Arias, V. Caselles, and G. Facciolo, “Analysis of a Variational Framework for Exemplar-Based Image Inpainting,” *SIAM MMS*, vol. 10, no. 2, pp. 473–514, jan 2012.
- [23] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3d transform-domain collaborative filtering,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, 2007.
- [24] D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration,” <http://people.csail.mit.edu/danielzoran/epllcode.zip>, accessed: 29/09/2014.
- [25] Y.-Q. Wang, “E-PL: an Algorithm for Image Inpainting,” *Image Processing On Line*, vol. 2013, pp. 271–285, 2013.
- [26] M. Lebrun, A. Buades, and J.-M. Morel, “Implementation of the “Non-Local Bayes” (NL-Bayes) Image Denoising Algorithm,” *Image Processing On Line*, vol. 3, pp. 1–42, 2013.
- [27] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, “Best Algorithms for HDR Image Generation. A Study of Performance Bounds,” *SIAM Journal on Imaging Sciences*, vol. 7, no. 1, pp. 1–34, 2014.
- [28] C. Aguerrebere, A. Almansa, J. Delon, Y. Gousseau, and P. Musé, “Single shot high dynamic range imaging using piecewise linear estimators,” in *Proceedings of IEEE International Conference on Computational Photography (ICCP)*, 2014.
- [29] P. E. Debevec and J. Malik, “Recovering High Dynamic Range Radiance Maps from Photographs,” in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH, 1997, pp. 369–378. [Online]. Available: <http://dx.doi.org/10.1145/258734.258884>
- [30] M. Granados, B. Afdin, M. Wand, C. Theobalt, H. P. Seidel, and H. P. A. Lensch, “Optimal HDR reconstruction with linear digital cameras,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 215–222.
- [31] C. Aguerrebere, J. Delon, Y. Gousseau, and P. Musé, “Simultaneous HDR image reconstruction and denoising for dynamic scenes,” in *Proceedings of IEEE International Conference on Computational Photography (ICCP)*, 2013, pp. 1–11.
- [32] D. Sidibé, W. Puech, and O. Strauss, “Ghost detection and removal in high dynamic range images,” in *Proceedings of the European Signal Processing Conference*, 2009.
- [33] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, “Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum,” *IEEE Transactions on Image Processing*, vol. 99, Mar 2010.
- [34] M. Schöberl, A. Belz, A. Nowak, J. Seiler, A. Kaup, and S. Foessel, “Building a high dynamic range video sensor with spatially nonregular optical filtering,” in *Proceedings of Proc. SPIE 8499, Applications of Digital Image Processing XXXV*, vol. 8499, 2012, pp. 84 990C–84 990C–11.
- [35] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, “Robust patch-based HDR reconstruction of dynamic scenes,” *ACM Transactions on Graphics*, vol. 31, no. 6, pp. 203:1–203:11, 2012.
- [36] M. Schöberl, A. Belz, J. Seiler, S. Foessel, and A. Kaup, “High dynamic range video by spatially non-regular optical filtering,” in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2012, pp. 2757–2760.
- [37] J. Hamilton and J. Adams, “Adaptive color plan interpolation in single sensor color electronic camera,” US Patent 5,629,734, 1997.
- [38] R. Mantiuk, S. Daly, and L. Kerofsky, “Display adaptive tone mapping,” *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 68:1–68:10, 2008.
- [39] A. Ostrowski, *Solution of equations and systems of equations*. Academic Press New York, 1960, vol. 2.



Associate position.



**Cecilia Aguerrebere** received the B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from Universidad de la República, Uruguay, in 2006, 2011 and 2014 respectively, the M.Sc. degree in applied mathematics from ENS Cachan, France, in 2011, and the Ph.D. degree in Signal and Image Processing from Télécom ParisTech, France, in 2014 (joint Ph.D program with Universidad de la República, Uruguay). From August 2015 she is with the Electrical and Computer Engineering Department, Duke University, where she holds a Postdoctoral Research

**Andrés Almansa** received his HDR, Ph.D. and M.Sc./Engineering degrees in Applied Mathematics and Computer Science from Université Paris-Descartes, ENS Cachan (France) and Universidad de la República (Uruguay), respectively, where he is an Associate Professor since 2004. His current interests as a CNRS Research Scientist at Telecom ParisTech include image restoration and analysis, subpixel stereovision and applications to earth observation, high quality digital photography and film restoration.

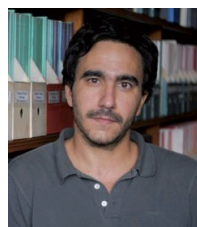


ANR project FREEDOM between 2007 and 2011.

**Julie Delon** studied mathematics from the École Normale Supérieure de Cachan and received the Ph. D. degree from ENS Cachan, France, in 2004 and the HDR degree from ENS Cachan in 2011. Between 2005 and 2013, she was a CNRS Researcher with Télécom ParisTech, Paris, France. Since 2013, she is a Professor of applied mathematics with Paris-Descartes University. Her current research interests include mono and multi-image restoration, optimal transport, and stochastic approaches in computer vision. She was coordinator of the young researcher



**Yann Gousseau** received the engineering degree from the École Centrale de Paris, France, in 1995, and the Ph.D. degree in applied mathematics from the University of Paris-Dauphine in 2000. He is currently a professor at Télécom ParisTech. His research interests include the mathematical modeling of natural images and textures, stochastic geometry, image analysis, computer vision and image processing.



**Pablo Musé** received his Electrical Engineering degree from Universidad de la República, Uruguay, in 1999, his M.Sc. degree in Mathematics, Vision and Learning and his Ph.D. in Applied Mathematics from ENS de Cachan, France, in 2001 and 2004 respectively. From 2005 to 2006 he was with Cognitech, Inc., Pasadena, CA, USA, where he worked on computer vision and image processing applications. In 2006 and 2007, he was a Postdoctoral Scholar with the Seismological Laboratory, California Institute of Technology, Pasadena, working on remote sensing using optical imaging, radar and GPS networks. Since 2008, he has been with the Division of Electrical Engineering, School of Engineering, Universidad de la República, where he is currently a Full Professor of signal processing.

# A Bayesian Hyperprior Approach for Joint Image Denoising and Interpolation, with an Application to HDR Imaging

## Supplementary Material

Cecilia Aguerrebere, Andrés Almansa, Julie Delon, Yann Gousseau and Pablo Musé

### APPENDIX A MAXIMUM A POSTERIORI

In the following, we derive the optimality equations for the function

$$\begin{aligned} f(\{\mathbf{C}_i\}, \mu, \mathbf{\Lambda}) &:= -\log p(\{\mathbf{C}_i\}, \mu, \mathbf{\Lambda} \mid \{\mathbf{Z}_i\}) \\ &= \frac{1}{2}(\mathbf{Z}_i - \mathbf{D}_i \mathbf{C}_i)^T \mathbf{\Sigma}_{N_i}^{-1} (\mathbf{Z}_i - \mathbf{D}_i \mathbf{C}_i) \\ &\quad - \frac{\nu - n + M}{2} \log |\mathbf{\Lambda}| + \frac{1}{2} \sum_{i=1}^M (\mathbf{C}_i - \mu)^T \mathbf{\Lambda} (\mathbf{C}_i - \mu) \\ &\quad + \frac{\kappa}{2} (\mu - \mu_0)^T \mathbf{\Lambda} (\mu - \mu_0) + \frac{1}{2} \text{trace}[\nu \mathbf{\Sigma}_0 \mathbf{\Lambda}], \end{aligned}$$

over the set  $\mathbb{R}^{nM} \times \mathbb{R}^n \times S_n^{++}(\mathbb{R})$ , with  $S_n^{++}(\mathbb{R})$  the set of real symmetric positive definite matrices of size  $n$ .

#### A. Optimality equations

**Proposition 1.** Assume that  $\mathbf{\Lambda}$  is fixed and that  $\mathbf{\Sigma}_{N_i}$  and  $\mathbf{C}_i$  are independent. Then the function  $(\{\mathbf{C}_i\}, \mu) \mapsto f(\{\mathbf{C}_i\}, \mu, \mathbf{\Lambda})$  is convex on  $\mathbb{R}^{n(M+1)}$  and has a unique minimizer  $(\{\hat{\mathbf{C}}_i\}, \hat{\mu})$ , given by

$$\hat{\mu} = \left( \kappa \text{Id} + \sum_{i=1}^M A_i \mathbf{D}_i \right)^{-1} \left( \sum_{i=1}^M A_i \mathbf{Z}_i + \kappa \mu_0 \right). \quad (1)$$

$$\hat{\mathbf{C}}_i = A_i (\mathbf{Z}_i - \mathbf{D}_i \hat{\mu}) + \hat{\mu}, \quad \forall i \in \{1, \dots, M\} \quad (2)$$

with  $A_i = \mathbf{\Lambda}^{-1} \mathbf{D}_i^T (\mathbf{D}_i \mathbf{\Lambda}^{-1} \mathbf{D}_i^T + \mathbf{\Sigma}_{N_i})^{-1}$ .

*Proof.* Showing that  $f$  is convex in  $(\{\mathbf{C}_i\}, \mu)$  is a simple exercise, left to the reader. It is also easy to show that the minimum of the quadratic function  $f$  is unique. Indeed the expression of this unique solution involves the inversion of a positive definite matrix. Differentiating  $f$  with respect to each  $\mathbf{C}_i$  yields

$$\frac{\partial f(\{\mathbf{C}_i\}, \mu, \mathbf{\Lambda} \mid \{\mathbf{Z}_i\})}{\partial \mathbf{C}_i} = -\mathbf{D}_i^T \mathbf{\Sigma}_{N_i}^{-1} (\mathbf{Z}_i - \mathbf{D}_i \mathbf{C}_i) + \mathbf{\Lambda} (\mathbf{C}_i - \mu). \quad (3)$$

Then, equating to zero, the explicit solution must satisfy

$$\begin{aligned} \mathbf{C}_i &= \mathbf{\Lambda}^{-1} \mathbf{D}_i^T (\mathbf{D}_i \mathbf{\Lambda}^{-1} \mathbf{D}_i^T + \mathbf{\Sigma}_{N_i})^{-1} (\mathbf{Z}_i - \mathbf{D}_i \mu) + \mu \\ &= A_i (\mathbf{Z}_i - \mathbf{D}_i \mu) + \mu. \end{aligned} \quad (4)$$

We recognize here a Wiener-like estimator for each  $\mathbf{C}_i$ .

Differentiating with respect to  $\mu$  yields

$$\frac{f(\{\mathbf{C}_i\}, \mu, \mathbf{\Lambda} \mid \{\mathbf{Z}_i\})}{\partial \mu} = \mathbf{\Lambda} \sum_{i=1}^M (\mu - \mathbf{C}_i) + \kappa \mathbf{\Lambda} (\mu - \mu_0), \quad (5)$$

which is zero if and only if (assuming  $\mathbf{\Lambda}$  is invertible)

$$\mu = \frac{M \bar{C} + \kappa \mu_0}{M + \kappa}, \quad \text{with } \bar{C} = \frac{1}{M} \sum_{i=1}^M \mathbf{C}_i. \quad (6)$$

Replacing the  $\mathbf{C}_i$ 's in the previous expression of  $\mu$  by their the expression as minimizers (Eq. (4)) leads to

$$\mu = \frac{\sum_{i=1}^M A_i (\mathbf{Z}_i - \mathbf{D}_i \mu) + M \mu + \kappa \mu_0}{M + \kappa},$$

Finally, grouping all the terms in  $\mu$  on the left hand side of the equation, we obtain the expected explicit solution

$$\mu = (\kappa \text{Id} + \sum_{i=1}^M A_i \mathbf{D}_i)^{-1} \left( \sum_{i=1}^M A_i \mathbf{Z}_i + \kappa \mu_0 \right). \quad (7)$$

□

**Proposition 2.** Assume that the variables  $(\{\mathbf{C}_i\}, \mu)$  are fixed. The function  $\mathbf{\Lambda} \mapsto f(\{\mathbf{C}_i\}, \mu, \mathbf{\Lambda})$  is convex on  $S_n^{++}(\mathbb{R})$  and its unique minimizer is

$$\hat{\mathbf{\Lambda}} = \left( \frac{\nu \mathbf{\Sigma}_0 + \kappa (\mu - \mu_0)(\mu - \mu_0)^T + \sum_{i=1}^M (\mathbf{C}_i - \mu)(\mathbf{C}_i - \mu)^T}{\nu + M - n} \right)^{-1}. \quad (8)$$

*Proof.* When  $(\{\mathbf{C}_i\}, \mu)$  is fixed, the function  $f$  is convex in  $\mathbf{\Lambda}$ . Indeed,  $f$  is a sum of linear terms in  $\mathbf{\Lambda}$  and a term in  $-\log |\mathbf{\Lambda}|$  which is also convex [1]. In order to compute the partial derivative with respect to  $\mathbf{\Lambda}$  (see for instance [2] for these derivations), observe that

$$\begin{aligned} f(\{\mathbf{C}_i\}, \mu, \mathbf{\Lambda}) &= -\frac{\nu - n + M}{2} \log |\mathbf{\Lambda}| \\ &\quad + \frac{1}{2} \sum_{i=1}^M (\mathbf{C}_i - \mu)^T \mathbf{\Lambda} (\mathbf{C}_i - \mu) \\ &\quad + \frac{\kappa}{2} (\mu - \mu_0)^T \mathbf{\Lambda} (\mu - \mu_0) \\ &\quad + \frac{1}{2} \text{trace}[\nu \mathbf{\Sigma}_0 \mathbf{\Lambda}]. \end{aligned} \quad (9)$$

The derivative of the first term is ( $\mathbf{\Lambda}$  is symmetric)

$$-\frac{\nu - n + M}{2} (\mathbf{\Lambda}^{-1})^T = -\frac{\nu - n + M}{2} \mathbf{\Lambda}^{-1}.$$

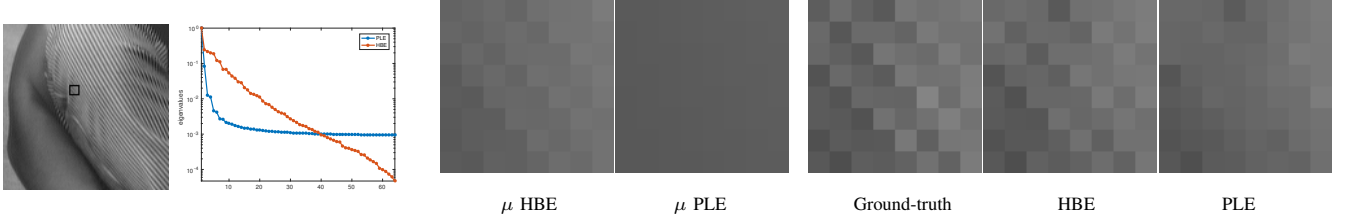


Fig. 1. **Left to right:** Chosen patch indicated by the black box. Decay of the eigenvalues of the patch model covariance matrix  $\Sigma$ . Patch model means  $\mu$  by each method. Ground-truth and restored patches by each method.

The derivative of the second and third terms are

$$\frac{1}{2} \sum_{i=1}^M (\mathbf{C}_i - \mu)(\mathbf{C}_i - \mu)^T$$

and

$$\frac{\kappa}{2} (\mu - \mu_0)(\mu - \mu_0)^T.$$

Finally, the derivative of the fourth term is

$$\frac{1}{2} \nu \Sigma_0^T = \frac{1}{2} \nu \Sigma_0.$$

Then it follows that

$$\begin{aligned} \frac{\partial f(\{\mathbf{C}_i\}, \mu, \Lambda \mid \{\mathbf{Z}_i\})}{\partial \Lambda} &= -\frac{\nu - n + M}{2} \Lambda^{-1} \\ &+ \frac{1}{2} \sum_{i=1}^M (\mathbf{C}_i - \mu)(\mathbf{C}_i - \mu)^T \\ &+ \frac{\kappa}{2} (\mu - \mu_0)(\mu - \mu_0)^T \\ &+ \frac{1}{2} \nu \Sigma_0. \end{aligned}$$

Equating to zero, this yields the unique minimizer

$$\Lambda^{-1} = \frac{\nu \Sigma_0 + \kappa (\mu - \mu_0)(\mu - \mu_0)^T + \sum_{i=1}^M (\mathbf{C}_i - \mu)(\mathbf{C}_i - \mu)^T}{\nu + M - n}. \quad (10)$$

□

### B. Affine risk minimizer

**Proposition 3.** Assume that the noise has zero mean and is not correlated to the signal  $\mathbf{C}_i$ . Then, the affine estimator  $\tilde{\mathbf{C}}_i$  that minimizes the Bayes risk  $\mathbb{E}[(\tilde{\mathbf{C}}_i - \mathbf{C}_i)^2]$  is given by

$$\tilde{\mathbf{C}}_i = \Lambda^{-1} \mathbf{D}_i^T (\mathbf{D}_i \Lambda^{-1} \mathbf{D}_i^T + \Sigma_{\mathbf{N}_i})^{-1} (\mathbf{Z}_i - \mathbf{D}_i \mu) + \mu. \quad (11)$$

*Proof.* Let us first consider the case  $\mu = 0$ . If we consider linear estimators only, we look for the matrix  $\tilde{\mathbf{W}}$  that verifies

$$\tilde{\mathbf{W}} = \arg \min_{\mathbf{W}} \mathbb{E}[(\mathbf{W} \mathbf{Z}_i - \mathbf{C}_i)^2]. \quad (12)$$

Hence,  $\tilde{\mathbf{W}}$  must verify

$$\mathbb{E}[(\tilde{\mathbf{W}} \mathbf{Z}_i - \mathbf{C}_i) \mathbf{Z}_i^T] = 0, \quad (13)$$

and we have

$$\tilde{\mathbf{W}} = \mathbb{E}[\mathbf{C}_i \mathbf{Z}_i^T] (\mathbb{E}[\mathbf{Z}_i \mathbf{Z}_i^T])^{-1}. \quad (14)$$

Since the noise  $N_i$  has zero mean and is not correlated to the signal  $\mathbf{C}_i$ , the element  $(p, q)$  of matrix  $\mathbb{E}[\mathbf{C}_i \mathbf{Z}_i^T]$  is given by

$$\mathbb{E}[\mathbf{C}_i \mathbf{Z}_i^T]_{p,q} = \mathbb{E}[\mathbf{C}_i (\mathbf{D}_i \mathbf{C}_i + \mathbf{N}_i)^T]_{p,q} \quad (15)$$

$$= \mathbb{E}[\mathbf{C}_i^p (\mathbf{D}_i \mathbf{C}_i)_q + \mathbf{C}_i^p \mathbf{N}_i^q] \quad (16)$$

$$= (\Lambda^{-1} \mathbf{D}_i^T)_{p,q}. \quad (17)$$

Also, the element  $(p, q)$  of matrix  $\mathbb{E}[\mathbf{Z}_i \mathbf{Z}_i^T]$  is given by

$$\mathbb{E}[\mathbf{Z}_i \mathbf{Z}_i^T]_{p,q} = \mathbb{E}[(\mathbf{D}_i \mathbf{C}_i + \mathbf{N}_i)(\mathbf{D}_i \mathbf{C}_i + \mathbf{N}_i)^T]_{p,q} \quad (18)$$

$$= \mathbb{E}[(\mathbf{D}_i \mathbf{C}_i)_p (\mathbf{D}_i \mathbf{C}_i)_q^T + (\mathbf{D}_i \mathbf{C}_i)_p (\mathbf{N}_i)_q^T] \quad (19)$$

$$+ (\mathbf{N}_i)_p (\mathbf{D}_i \mathbf{C}_i)_q + (\mathbf{N}_i)_p (\mathbf{N}_i)_q^T] \quad (20)$$

$$= (\mathbf{D}_i \Lambda^{-1} \mathbf{D}_i^T)_{p,q} + (\Sigma_{\mathbf{N}_i})_{p,q}. \quad (21)$$

Hence we have,

$$\tilde{\mathbf{W}} = \Lambda^{-1} \mathbf{D}_i^T (\mathbf{D}_i \Lambda^{-1} \mathbf{D}_i^T + \Sigma_{\mathbf{N}_i})^{-1}. \quad (22)$$

In the general case where  $\mu \neq 0$ , we can always consider the centered version of the patches  $(\mathbf{Z}_i - \mathbf{D}_i \mu)$  and apply the previous result. Therefore, the estimator of  $\mathbf{C}_i$  that minimizes the risk function  $\mathbb{E}[(\tilde{\mathbf{C}} - \mathbf{C}_i)^2]$  among all affine estimators is given by (11). □

## APPENDIX B INITIALIZATION

A good initialization is crucial since we aim at solving a non-convex problem through an iterative procedure. Yu et al. [3] propose to initialize the PLE algorithm by learning the  $K$  GMM covariance matrices from synthetic images of edges with different orientations as well as the DCT basis to represent isotropic patterns. As they state, in dictionary learning, the most prominent atoms represent local edges which are useful at representing and restoring contours. Hence, this initialization helps to correctly restore corrupted patches even in quite extreme cases.

Each covariance matrix  $\Sigma_k$ ,  $k = 1, \dots, K - 1$ , corresponds to one of  $K - 1$  orientations, uniformly sampled from directions 0 to  $\pi$ . For a given orientation  $\theta$ , a synthetic black-and-white image is generated and patches that touch the contour at different positions are randomly sampled from it. A covariance matrix is then computed from the sampled patches. The first eigenvector of the covariance matrix, which is almost constant, is replaced by a constant vector. This allows a class  $k$  of a given orientation to restore patches having different means. Up to a certain gray level difference, dark or bright edges with the same orientation are correctly represented by the same class. A Gram-Schmidt orthogonalization is computed

on the other eigenvectors to ensure the orthogonality of the basis. The eigenvalues of all bases are initialized with the same values, chosen to have a fast decay. Lastly, the DCT basis is added to represent isotropic image patterns, making a total of  $K$  classes. The mean of each class  $\mu_k$ ,  $k = 1, \dots, K$ , is initialized to zero. The authors claim that they have found in practice that  $K = 19$  classes (i.e. 18 orientations, 10 degrees apart) give a correct reconstruction and are a good compromise between performance and complexity for a patch size of  $8 \times 8$ . The fact that the algorithm is applied in regions of size  $128 \times 128$ , and therefore localized, also explains why this a priori small number of classes can be suitable to describe all image patches.

To initialize the proposed algorithm, we follow the approach by Yu et al. [3] and compute the  $K$  covariance matrices  $\Sigma_k$  as previously described. Then, each image patch  $\mathbf{z}_i$  is reconstructed under each class  $k = 1, \dots, K$  as

$$\tilde{\mathbf{C}}_i^k = \Sigma_k \mathbf{D}_i^T (\mathbf{D}_i \Sigma_k \mathbf{D}_i^T + \Sigma_{\mathbf{N}_i})^{-1} (\mathbf{z}_i - \mathbf{D}_i \mu_k) + \mu_k. \quad (23)$$

The best suited class  $\tilde{k}_i$  is chosen as the one maximizing the posterior probability of the patch  $p(\mathbf{C}_i | \mathbf{z}_i, \mu_k, \Sigma_k)$  over  $k$  assuming  $\mathbf{C} = \tilde{\mathbf{C}}_i^k$ :

$$\begin{aligned} \tilde{k}_i &= \arg \max_k \log p(\mathbf{C} | \mathbf{z}_i, \mu_k, \Sigma_k) \\ &= \arg \min_k \left( (\mathbf{z}_i - \mathbf{D}_i \tilde{\mathbf{C}}_i^k)^T \Sigma_{\mathbf{N}}^{-1} (\mathbf{z}_i - \mathbf{D}_i \tilde{\mathbf{C}}_i^k) \right. \\ &\quad \left. + (\tilde{\mathbf{C}}_i^k - \mu_k)^T \Sigma_k^{-1} (\tilde{\mathbf{C}}_i^k - \mu_k) + \ln |\Sigma_k| \right). \end{aligned} \quad (24)$$

The first oracle is thus created by aggregating the estimations of all patches corresponding to the chosen classes. In other words, the oracle of the first iteration of the proposed approach is the output of the first iteration of the PLE algorithm. Algorithm 1 summarizes the initialization steps.

---

**Algorithm 1:** Summary of the initialization procedure.

---

**Input:**  $\mathbf{Z}$ ,  $\mathbf{D}$ ,  $K$

**Output:**  $\mathbf{C}_{oracle}$

- 1 Decompose  $\mathbf{Z}$  and  $\mathbf{D}$  into overlapping patches.
  - 2 Compute the covariance matrix of the  $K$  classes from synthetic images of edges plus the DCT.
  - 3 Project all patches into the  $K$  classes using (23) and chose the best class using (24).
  - 4 Compute the first oracle  $\mathbf{C}_{oracle}$  aggregating the estimations of all patches for the chosen class.
- 

## APPENDIX C

### EXTRA EXPERIMENTAL RESULTS

#### A. Synthetic Degradation

In this section we include further results showing the performance of the proposed approach. Table I shows the structural similarity index (SSIM) obtained by each method under the different degradation models.

#### B. Real Data

Figures 2 and 3 show the SVE raw images used to create the HDR examples shown in Section V.C, figures 8 and 9 respectively.

## APPENDIX D

### COMPARISON TO THE PLE RESTORATION

Despite close, HBE and PLE have fundamental differences that make the obtained models considerably different. The two main differences are the computation of the model parameters  $\mu$  and  $\Sigma$  and the group patches involved in this computation. PLE alternates between the patch restoration and the maximum-likelihood estimation of the model parameters, whereas in the case of HBE, it uses a joint maximum a posteriori formulation to estimate both the image patches and the parameters  $\mu$  and  $\Sigma$ , thanks to a Bayesian hyperprior model on these parameters. In the case of PLE, the included patches are all those belonging to the most probable cluster, while in the case of HBE are all patches similar to the current one in a local search window (often much fewer than in the case of PLE).

Both algorithms share the same initialization. Figure 1 shows the final model after subsequent iterations of each approach, represented by the decay of the covariance matrix eigenvalues and the corresponding means, for the patch indicated by the black box in the same figure. Figure 1 also shows the final patch reconstructions obtained by each method, clearly showing the performance difference already presumed by the difference in the patch models.

Figure 4 shows examples of the 18 patches used by HBE to compute the model and the cluster map produced by PLE, which indicates the cluster (among the possible 19) assigned to each pixel. In this example, the chosen patch belongs to the cluster represented by the blue color (indicated in white in the binary image next to the cluster map). As it can be seen in Figure 4 the PLE model includes much more varied patches (all those with value 0 in the cluster image). PLE mixes constant patches from the arm, giving a much smoother result.

## REFERENCES

- [1] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [2] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Technical University of Denmark, Tech. Rep., 2012.
- [3] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, 2012.
- [4] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," <http://people.csail.mit.edu/danielzoran/eplcode.zip>, accessed: 29/09/2014.
- [5] Y.-Q. Wang, "E-PL: an Algorithm for Image Inpainting," *Image Processing On Line*, vol. 2013, pp. 271–285, 2013.
- [6] M. Lebrun, A. Buades, and J.-M. Morel, "Implementation of the "Non-Local Bayes" (NL-Bayes) Image Denoising Algorithm," *Image Processing On Line*, vol. 3, pp. 1–42, 2013.





Fig. 2. SVE raw image used in the HDR generation example in Section V.C Figure 8



Fig. 3. SVE raw image used in the HDR generation example in Section V.C Figure 9

Interpolation - PSNR (dB)													Interpolation & Denoising - PSNR (dB)							
50%					70%				90%				70%				90%			
	HBE	PLE	EPLL	E-PLE	HBE	PLE	EPLL	E-PLE	HBE	PLE	EPLL	E-PLE	HBE	PLE	EPLL	E-PLE	HBE	PLE	EPLL	E-PLE
barbara	<b>0.92</b>	0.90	0.87	0.88	<b>0.85</b>	0.81	0.73	0.76	<b>0.58</b>	0.48	0.43	0.46	<b>0.78</b>	0.76	0.67	0.68	<b>0.53</b>	0.46	0.41	0.44
boat	<b>0.87</b>	0.85	0.84	0.83	<b>0.75</b>	0.72	0.71	0.70	<b>0.50</b>	0.46	0.40	0.43	<b>0.70</b>	0.68	0.67	0.64	<b>0.47</b>	0.44	0.38	0.44
traffic	<b>0.86</b>	0.85	0.85	0.82	<b>0.76</b>	0.73	0.72	0.73	<b>0.52</b>	0.48	0.43	0.47	<b>0.67</b>	<b>0.67</b>	0.66	0.65	<b>0.47</b>	0.44	0.41	0.44
Denoising - PSNR (dB)													Zooming - PSNR (dB)							
$\sigma^2$	10			30			50			80			$\times 2$							
	HBE	NLB	EPLL	HBE	NLB	EPLL	HBE	NLB	EPLL	HBE	NLB	EPLL	HBE	PLE	EPLL	E-PLE	Lanczos			
barbara	<b>0.89</b>	0.88	0.88	<b>0.83</b>	0.81	0.81	<b>0.80</b>	0.78	0.78	<b>0.77</b>	0.75	0.76	<b>0.81</b>	0.80	0.77	0.79	0.72			
boat	<b>0.91</b>	0.90	0.89	<b>0.82</b>	0.80	0.78	<b>0.78</b>	0.75	0.73	<b>0.73</b>	0.70	0.69	<b>0.75</b>	0.70	0.68	0.69	0.66			
traffic	<b>0.87</b>	0.86	<b>0.87</b>	<b>0.82</b>	0.80	0.81	<b>0.79</b>	0.77	<b>0.79</b>	0.76	0.75	<b>0.77</b>	<b>0.84</b>	0.81	0.83	0.82	0.74			

TABLE I

RESULTS OF THE INTERPOLATION, COMBINED INTERPOLATION AND DENOISING, DENOISING AND ZOOMING TESTS DESCRIBED IN SECTION IV-A. PATCH SIZE OF  $8 \times 8$  FOR ALL METHODS IN ALL TESTS. PARAMETER SETTING FOR INTERPOLATION, COMBINED INTERPOLATION AND DENOISING, AND ZOOMING, HBE:  $\alpha_H = 1$ ,  $\alpha_L = 0.5$ , PLE:  $\sigma = 3$ ,  $\varepsilon = 30$ ,  $K = 19$  [3], EPLL: DEFAULT PARAMETERS [4], E-PLE: PARAMETERS SET AS SPECIFIED IN [5]. PARAMETER SETTING FOR DENOISING, HBE:  $\alpha_H = \alpha_L = 100$ , NLB: CODE PROVIDED BY THE AUTHORS [6] AUTOMATICALLY SETS PARAMETERS FROM INPUT  $\sigma^2$ , EPLL: DEFAULT PARAMETERS FOR THE DENOISING EXAMPLE [4]

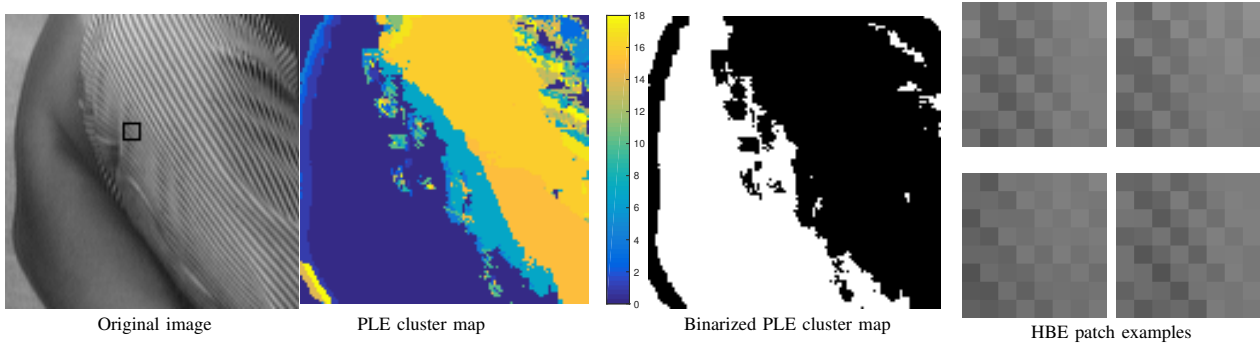


Fig. 4. **Left:** Original image with considered patch indicated by the black box, cluster map obtained by PLE for the corresponding image and binarized version of the cluster map. Each color of the PLE cluster map indicates the cluster, among the possible 19, assigned to each pixel. The binary image shows in white all pixels assigned to the same cluster as the considered patch. As it can be seen, pretty varied patches are being combined. **Right:** Four examples of the patches used by HBE for the restoration of the considered patch.