Accelerating GMM-based patch priors for image restoration: Three ingredients for a 100× speed-up

Shibin Parameswaran, Charles-Alban Deledalle, Loïc Denis and Truong Q. Nguyen

Abstract—Image restoration methods aim to recover the underlying clean image from corrupted observations. The Expected Patch Log-likelihood (EPLL) algorithm is a powerful image restoration method that uses a Gaussian mixture model (GMM) prior on the patches of natural images. Although it is very effective for restoring images, its high runtime complexity makes EPLL ill-suited for most practical applications. In this paper, we propose three approximations to the original EPLL algorithm. The resulting algorithm, which we call the fast-EPLL (FEPLL), attains a dramatic speed-up of two orders of magnitude over EPLL while incurring a negligible drop in the restored image quality (less than 0.5 dB). We demonstrate the efficacy and versatility of our algorithm on a number of inverse problems such as denoising, deblurring, super-resolution, inpainting and desvgnetting. To the best of our knowledge, FEPLL is the first algorithm that can competitively restore a 512×512 pixel image in under 0.5s for all the degradations mentioned above without specialized code optimizations such as CPU parallelization or GPU implementation.

Index Terms—Image restoration, image patch, Gaussian mixture model, efficient algorithms

I. INTRODUCTION

Patch-based methods form a very popular and successful class of image restoration techniques. These methods process an image on a patch-by-patch basis where a patch is a small sub-image (e.g., of 8×8 pixels) that captures both geometric and textural information. Patch-based algorithms have been at the core of many state-of-the-art results obtained on various image restoration problems such as denoising, deblurring, super-resolution, defogging, or compression artifact removal to name a few. In image denoising, patch-based processing became popular after the success of the Non-Local Means algorithm [2]. Subsequently, continued research efforts have led to significant algorithmic advancements in this area [1], [8], [43], [11], [29], [40], [26]. Other inverse problems such as image super-resolution and image deblurring have also benefited from patch-based models [9], [16], [32], [37], [14], [27].

Among these various patch-based methods, the Expected Patch Log-Likelihood algorithm (EPLL) [43] deserves a special mention due to its restoration performance and versatility. The EPLL introduced an innovative application of Gaussian Mixture Models (GMMs) to capture the prior distribution of patches in natural images. Note that a similar idea was introduced concurrently in [40]. The success of this method is evident from the large number of recent works that extend the original EPLL formulation [19], [38], [28], [4], [31], [33], [20]. However, a persistent problem of EPLL-based algorithms is their high runtime complexity. For instance, it is orders of magnitude slower than the well-engineered BM3D image denoising algorithm [8]. However, extensions of BM3D that perform super-resolution [10] and other inverse problems [22] require fundamental algorithmic changes, making BM3D far less adaptable than EPLL. Other approaches that are as versatile as EPLL [36], [5], [21] either lack the algorithmic efficiency of BM3D or the restoration efficacy of EPLL.

Another class of techniques that arguably offers better runtime performance than EPLL-based methods (but not BM3D) are those based on deep learning. With the advancements in computational resources, researchers have attempted to solve some classical inverse problems using multi-layer perceptrons [3] and deep networks [6], [13], [23]. These methods achieve very good restoration performance, but are heavily dependent on the amount of training data available for each degradation scenario. Most of these methods learn filters that are suited to restore a specific noise level (denoising), blur (deblurring) or upsampling factor (super-resolution), which makes them less attractive to serve as generic image restoration solutions. More recently, Zhang et al. [42] demonstrated the use of deep residual networks for general denoising problems, single-image super-resolution and compression artifact removal. Unlike earlier deep learning efforts, their approach can restore images with different noise levels using a single model which is learned by training on image patches containing a range of degradations. Even in this case, the underlying deep learning model requires retraining whenever a new degradation scenario different from those considered during the learning stage is encountered.

More recently, [34] proposed training a single deep network to solve many inverse problems. This work uses an iterative scheme that alternatively enforces a good fit to the learned prior model of natural images (via a projection operator performed by a deep neural network) and a satisfying fidelity to the data (via the direct model). The projection operator takes the form of an auto-encoder, trained by an adversarial strategy, that processes 64×64 patches with 1024-dimensional latent space (i.e., 4 times smaller). This framework shows promise but, in its current form, requires a very large image database for training. In addition, almost all of the training and testing in the original publication are conducted on small 64×64 images. Another limitation, as noted by the authors, is that the

S. Parameswaran, C.-A. Deledalle and T. Nguyen are with the Department of Electrical and Computer Engineering, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093, e-mail: sparames@ucsd.edu, cdeledalle@eng.ucsd.edu and tqn001@eng.ucsd.edu.

C. Deledalle is also with IMB, CNRS, Univ. Bordeaux, Bordeaux INP, F-33405 Talence, France, e-mail: charles-alban.deledalle@math.u-bordeaux.fr.

L. Denis is with Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d’Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Etienne, France, e-mail: loic.denis@univ-st-etienne.fr.
regularization parameter that controls the contribution of prior is fixed during training stage. In other words, regularization parameter cannot be changed during test time which can be an issue in certain situations. In contrast, our approach needs much less data during training stage and, during test time, the regularization parameter can be tuned according to the signal-to-noise ratio of the image being addressed.

Moreover, it is much harder to gain insight into the actual model learned by a deep architecture compared to a GMM. For this reason, even with the advent of deep learning methods, flexible algorithms like EPLL that have a transparent formulation remain relevant for image restoration.

Recently, researchers have tried to improve the speed of EPLL by replacing the most time-consuming operation in the EPLL algorithm with a machine learning-based technique of their choice [39, 35]. These methods were successful in accelerating EPLL to an extent but did not consider tackling all of its bottlenecks. In contrast, this paper focuses on accelerating EPLL by proposing algorithmic approximations to all the prospective bottlenecks present in the original algorithm proposed by Zoran et al. [43]. To this end, we first provide a complete computational and runtime analysis of EPLL, present a new and efficient implementation of original EPLL algorithm and then finally propose innovative approximations that lead to a novel algorithm that is more than 100× faster compared to the efficiently implemented EPLL (and 350× faster than the runtime obtained by using the original implementation [43]).

Contributions: The main contributions of this work are the following. We introduce three strategies to accelerate patch-based image restoration algorithms that use a GMM prior. We show that, when used jointly, they lead to a speed-up of the EPLL algorithm by two orders of magnitude. Compared to the popular BM3D algorithm, which represents the current state-of-the-art in terms of speed among CPU-based implementations, the proposed algorithm is almost an order of magnitude faster. The three strategies introduced in this work are general enough to be applied individually or in any combination to accelerate other related algorithms. For example, the random subsampling strategy is a general technique that could be reused in any algorithm that considers overlapping patches to process images; the flat-tail spectrum approximation can accelerate any method that needs Gaussian log-likelihood or multiple Mahalanobis metric calculations; finally, the binary search tree for Gaussian matching can be included in any algorithm based on a GMM prior model and can be easily adapted for vector quantization techniques that use a dictionary.

For reproducibility purposes, we release our software on GitHub along with a few usage demonstrations (available at https://googl/xjqKUA).

II. EXPECTED PATCH LOG-LIKELIHOOD (EPLL)

We consider the problem of estimating an image \( x \in \mathbb{R}^N \) (\( N \) is the number of pixels) from noisy linear observations \( y = Ax + w \), where \( A : \mathbb{R}^N \rightarrow \mathbb{R}^M \) is a linear operator and \( w \in \mathbb{R}^M \) is a noise component assumed to be white and Gaussian with variance \( \sigma^2 \). In a standard denoising problem, \( A \) is the identity matrix, but in more general settings, it can account for loss of information or blurring. Typical examples for operator \( A \) are: a low pass filter (for deconvolution), a masking operator (for inpainting), or a projection on a random subspace (for compressive sensing). To reduce noise and stabilize the inversion of \( A \), some prior information is used for the estimation of \( x \). The EPLL introduced by Zoran and Weiss [43] includes this prior information as a model for the distribution of patches found in natural images. Specifically, the EPLL defines the restored image as the maximum a posteriori estimate, corresponding to the following minimization problem:

\[
\arg\min_x \frac{P}{2\sigma^2} \|Ax - y\|^2 - \sum_{i \in I} \log p(P_i, x) \tag{1}
\]

where \( I = \{1, \ldots, N\} \) is the set of pixel indices, \( P_i : \mathbb{R}^N \rightarrow \mathbb{R}^P \) is the linear operator extracting a patch with \( P \) pixels centered at the pixel with location \( i \) (typically, \( P = 8 \times 8 \)), and \( p(.) \) is the a priori probability density function (i.e., the statistical model of noiseless patches in natural images). While the first term in eq. (1) ensures that \( Ax \) is close to the observations \( y \) (this term is the negative log-likelihood under the white Gaussian noise assumption), the second term regularizes the solution \( x \) by favoring an image such that all its patches fit the prior model of patches in natural images. The authors of [43] showed that this prior can be well approximated (upon removal of the DC component of each patch) using a zero-mean Gaussian Mixture Model (GMM) with \( K = 200 \) components, that reads for any patch \( z \in \mathbb{R}^P \), as

\[
p(z) = \sum_{k=1}^{K} w_k \frac{1}{(2\pi)^{P/2} |\Sigma_k|^{1/2}} \exp \left( -\frac{1}{2} z^t \Sigma_k^{-1} z \right), \tag{2}
\]

where the weights \( w_k \) (such that \( w_k > 0 \) and \( \sum_k w_k = 1 \)) and the covariance matrices \( \Sigma_k \in \mathbb{R}^{P \times P} \) are estimated using the Expectation-Maximization algorithm [12] on a dataset consisting of 2 million “clean” patches extracted from the training set of the Berkeley Segmentation (BSDS) dataset [30].

Half-quadratic splitting: Problem (1) is a large non-convex problem where \( A \) couples all unknown pixel values \( x \) and the patch prior is highly non-convex. A classical workaround, known as half-quadratic splitting [15, 25], is to introduce \( N \) auxiliary unknown vectors \( z_i \in \mathbb{R}^P \), and consider instead the penalized optimization problem that reads, for \( \beta > 0 \), as

\[
\arg\min_{x, z_1, \ldots, z_N} \frac{P}{2\sigma^2} \|Ax - y\|^2 + \frac{\beta}{2} \sum_{i \in I} \|P_i x - z_i\|^2 - \sum_{i \in I} \log p(z_i). \tag{3}
\]

When \( \beta \rightarrow \infty \), the problem (3) becomes equivalent to the original problem (1). In practice, an increasing sequence of \( \beta \) is considered, and an alternating optimization scheme is used:

\[
\hat{z}_i \leftarrow \arg\min_{z_i} \frac{\beta}{2} \|P_i \hat{x} - z_i\|^2 - \log p(z_i) \quad \text{for } i = 1, \ldots, N \tag{4}
\]

\[
\hat{x} \leftarrow \arg\min_{x} \frac{P}{2\sigma^2} \|Ax - y\|^2 + \frac{\beta}{2} \sum_{i \in I} \|P_i x - \hat{z}_i\|^2. \tag{5}
\]
TABLE I

COMPARISON OF THE EXECUTION TIME OF OUR IMPLEMENTATION OF EPLL WITH AND WITHOUT PROPOSED ACCELERATIONS. EXPERIMENT CONDUCTED ON 40 IMAGES EACH OF SIZE 481 × 321 IN DENOISING SETTING. PROFILING WAS CARRIED OUT USING MATLAB (R 2014 b) ON A PC WITH INTEL(R) CORE(TM) i7-4790K CPU @ 4.00GHz AND 16 GB RAM. EXECUTION TIMES ARE REPORTED AS AVERAGE NUMBER OF SECONDS PER IMAGE (S) AND PERCENTAGE OF THE TOTAL TIME (%).

<table>
<thead>
<tr>
<th>Step</th>
<th>Without accelerations</th>
<th>With the proposed accelerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Gaussian selection)</td>
<td>38.27s</td>
<td>55 %</td>
</tr>
<tr>
<td>(Patch estimation)</td>
<td>0.95s</td>
<td>2 %</td>
</tr>
<tr>
<td>(Patch extraction)</td>
<td>0.43s</td>
<td>1 %</td>
</tr>
<tr>
<td>(Patch reprojection)</td>
<td>0.15s</td>
<td>0 %</td>
</tr>
<tr>
<td>Others</td>
<td>0.52s</td>
<td>1 %</td>
</tr>
<tr>
<td>Total</td>
<td>40.32s</td>
<td>0.31s</td>
</tr>
</tbody>
</table>

Algorithm 1 The five steps of an EPLL iteration [43]

for all \(i \in I\)

\[
\begin{align*}
\hat{z}_i & \leftarrow \mathcal{P}_i \hat{x} \quad \text{(Patch extraction)} \\
\hat{k}_i^* & \leftarrow \arg \min_{1 \leq k \leq K} \log w_{k,i}^2 + \log \left| \Sigma_{k,i} + \frac{1}{P} \text{Id}_P \right| + \\
\hat{z}_i & \leftarrow \left( \Sigma_{k,i} + \frac{1}{P} \text{Id}_P \right)^{-1} \hat{z}_i \quad \text{(Gaussian selection)} \\
\hat{z}_i & \leftarrow \left( \Sigma_{k,i} + \frac{1}{P} \text{Id}_P \right)^{-1} \Sigma_{k,i} \hat{z}_i \quad \text{(Patch estimation)} \\
\hat{x}_i & \leftarrow \left( \sum_{i \in I} \mathcal{P}_i^T \mathcal{P}_i \right)^{-1} \sum_{i \in I} \mathcal{P}_i^T \hat{z}_i \quad \text{(Patch reprojection)} \\
\hat{x} & \leftarrow \left( \mathcal{A}' \mathcal{A} + \beta \sigma^2 \text{Id}_N \right)^{-1} \left( \mathcal{A}' \mathcal{y} + \beta \sigma^2 \mathcal{z} \right) \quad \text{(Image estimation)}
\end{align*}
\]  

(6)

where \(\mathcal{P}_i^T \mathcal{P}_i\) is a diagonal matrix whose \(i\)-th diagonal element corresponds to the number of patches overlapping the pixel of index \(i\). This number is a constant equal to \(P\) (assuming proper boundary conditions are used), which allows to split the computation into two steps Patch reprojection and Image estimation as shown in Alg. 1. Note that the step Patch reprojection is simply the average of all overlapping patches. In contrast, subproblem (4) cannot be obtained in closed form as it involves a term with the logarithm of a sum of exponentials. A practical solution proposed in [43] is to keep only the component \(k_i^*\) maximizing the likelihood for the given patch assuming it is a zero-mean Gaussian random vector with covariance matrix \(\Sigma_{k,i} + \frac{1}{P} \text{Id}_P\). With this approximation, the solution of (4) is also given by Wiener filtering, and the resulting algorithm iterates the steps described in Alg. 1. The authors of [43] found that using \(T = 5\) iterations, with the sequence \(\beta = \frac{1}{\sigma^2} \{1, 4, 8, 16, 32\}\), for the initialization \(\hat{x} = \mathcal{y}\), provides relevant solutions in denoising contexts for a wide range of noise levels \(\sigma^2\).

A. Complexity via eigenspace implementation

The algorithm summarized in the previous section may reveal cumbersome computations as it requires performing numerous matrix multiplications and inversions. Nevertheless, as the matrices \(\Sigma_k\) are known prior to any calculation, their eigendecomposition can be computed offline to improve the runtime. If we denote the eigendecomposition (obtained offline) of \(\Sigma_k = U_k S_k U_k^T\), such that \(U_k \in \mathbb{R}^{P \times P}\) is unitary and \(S_k\) is diagonal with positive diagonal elements ordered in decreasing order, steps Gaussian selection and Patch estimation can be expressed in the space of coefficients \(c\) as

\[
\begin{align*}
\tilde{c}_i^k & \leftarrow U_k^T \tilde{z}_i \quad \text{O}(NKP^2) \quad (7) \\
k_i^* & \leftarrow \arg \min_{1 \leq k \leq K} \sum_{i=1}^{N} \left( \log \nu_j + \frac{|\tilde{c}_i^k| \tilde{c}_i^k |}{\nu_j} \right) \\
\tilde{z}_i & \leftarrow U_k k_i^* \tilde{c}_i \quad \text{O}(NP^2) \quad (10)
\end{align*}
\]

where \(\tilde{c}_i^k\) denotes the \(j\)-th entry of vector \(\tilde{c}\), \(\nu_j = -2 \log w_j\), \(\nu_j = |S_k|_{jj} + \frac{1}{\nu_j}\), and \(\gamma_j = |S_k|_{jj} \nu_j^2\) with \(|S_k|_{jj}\) the \(j\)-th entry on the diagonal of matrix \(S_k\). The complexity of each operation is indicated on its right and corresponds to the number of operations per iteration of the alternate optimization scheme. Each steps Patch extraction and Patch reprojection share a complexity of \(O(NP)\). Finally, the complexity of step Image estimation depends on the transform \(\mathcal{A}\). In many scenarios of interest, \(\mathcal{A}' \mathcal{A}\) can be diagonalized using a fast transform and the inversion of \(\mathcal{A}' \mathcal{A} + \beta \sigma^2 \text{Id}_N\) can be performed efficiently in the transformed domain (since \(\text{Id}_N\) is diagonal in any orthonormal basis). For instance, it leads to \(O(N)\) operations for denoising or inpainting, and \(O(N \log N)\) for periodical deconvolutions or super-resolution problems, thanks to the fast Fourier transform (these are the settings we have adopted in this paper). If \(\mathcal{A}\) cannot be easily diagonalized, this step can be performed using conjugate gradient (CG) method, as done in [43], at a computational cost that depends on the number of CG iterations (i.e., on the conditioning of \(\mathcal{A}' \mathcal{A} + \beta \sigma^2 \text{Id}_N\)). In any case, as shown in the next section, this step has a complexity independent of \(P\) and \(K\) and is one of the faster operations in the image restoration problems considered in this paper.

B. Computation time analysis

In order to uncover the practical computational bottlenecks of EPLL, we have performed the following computational analysis. To identify clearly which part is time consuming, it is
important to make the algorithm implementation as optimal as possible. Therefore, we refrain from using the MATLAB code provided by the original authors [43] for speed comparisons. Instead, we use a MATLAB/C version of EPLL based on the eigenspace implementation described above, where some steps are written in C language and interfaced using mex functions. This version, which we refer to as EPLLc, provides results identical to the original implementation while being 2-3 times faster. The execution time of each step for a single run of EPLLc is reported in the second column of Table I. Reported times fit our complexity analysis and clearly indicate that the Step Gaussian selection causes significant bottleneck due to $O(NP^2K)$ complexity.

In the next section, we propose three independent modifications leading to an algorithm with a complexity of $O(NP\tilde{r} \log K/s^2)$ with two constants $1 \leq s^2 \leq P$ and $1 \leq \tilde{r} \leq P$ that control the accuracy of the approximations introduced. The algorithm, in practice, is more than 100 times faster as shown by its runtimes reported in the third column.

III. FAST EPLL: THE THREE KEY INGREDIENTS

We propose three accelerations based on (i) scanning only a (random) subset of the $N$ patches, (ii) reducing the number of mixture components matched, and (iii) projecting on a smaller subspace of the covariance eigenspace. We begin by describing this latter acceleration strategy in the following paragraph.

A. Speed-up via flat tail spectrum approximation

To avoid computing the $P$ coefficients of the vector $\tilde{c}^k_i$ in eq. (7), we rely on a flat-tail approximation. The $k$-th Gaussian model is said to have a flat tail if there exists a rank $r_k$ such that for any $j > r_k$, the eigenvalues are constant: $|S_k|_{j,j} = \lambda_k$. Denoting by $\tilde{U}_k \in \mathbb{R}^{P \times r_k}$ (resp. $\tilde{U}^*_k \in \mathbb{R}^{P \times r^*_k}$) the matrix formed by the $r_k$ first (resp. $r^*_k = P - r_k$ last) columns of $U_k$, we have $\tilde{U}_k^t (\tilde{U}_k^*)^t = \text{Id}_P - \tilde{U}_k \tilde{U}_k^t$. It follows

$$
\begin{align*}
(\Sigma_k + \frac{1}{\beta} \text{Id}_P)^{-1} &= \tilde{U}_k (\tilde{S}_k + \frac{1}{\beta} \text{Id}_{r_k})^{-1} (\tilde{U}_k^t + (\lambda_k + \frac{1}{\beta})^{-1} (\text{Id}_P - \tilde{U}_k \tilde{U}_k^t)), \\
(\Sigma_k + \frac{1}{\beta} \text{Id}_P)^{-1} \Sigma_k &= \tilde{U}_k (\tilde{S}_k + \frac{1}{\beta} \text{Id}_{r_k})^{-1} \tilde{S}_k \tilde{U}_k^t + \lambda_k (\lambda_k + \frac{1}{\beta})^{-1} (\text{Id}_P - \tilde{U}_k \tilde{U}_k^t),
\end{align*}
$$

where $\tilde{S}_k \in \mathbb{R}^{r_k \times r_k}$ is the diagonal matrix formed by the $r_k$ first rows and columns of $S_k$. Steps Gaussian selection and Patch estimation can thus be rewritten as

$$
\begin{align*}
\left\{ \hat{c}^k_i \leftarrow \tilde{U}_k^t \tilde{z}_i \right\}_{k=1..K} & \quad \mathcal{O}(NK\tilde{r}) \\
\left\{ k^*_i \leftarrow \arg\min_{1 \leq k \leq K} t_k^* + r^*_k \log \nu_P + \frac{||\tilde{z}_i||^2}{\nu_P} \\
& + \frac{\rho}{P} \sum_{j=1}^{r_k} \left( \log \nu_j^* + \frac{||\hat{c}^k_j||^2}{\nu_j^*} - \frac{||\hat{c}^k_j||^2}{\nu_j} \right) \right\}_{i=1..N} & \quad \mathcal{O}(NK^\tilde{r}) \\
\left\{ [\hat{c}_i]_j \leftarrow (\gamma^*_j - \gamma^*_P) [\hat{c}^k_i]_j \right\}_{j=1..r_k^*} & \quad \mathcal{O}(NP\tilde{r}) \\
\left\{ \tilde{z}_i \leftarrow \tilde{U}_k^t \hat{c}_i + \gamma^*_P \tilde{z}_i \right\}_{i=1..N} & \quad \mathcal{O}(NP\tilde{r})
\end{align*}
$$

where $\nu_P^* = \lambda_k + \frac{1}{\beta}$, $\gamma^*_P = \lambda_k / \nu_P^*$. As $||\tilde{z}_i||^2$ can be computed once for all $k$, the complexity of each step is divided by $P/\tilde{r}$, where $\tilde{r} = \frac{1}{K} \sum_{k=1}^{K} r_k$ is the average rank after which eigenvalues are considered constant.

In practice, covariance matrices $\Sigma_k$ are not flat-tail but can be approximated by a flat-tail matrix by replacing the lowest eigenvalues by a constant $\lambda_k$. To obtain a small value of $\bar{r}$ (hence a large speed-up), we preserve a fixed proportion $\rho \in (0,1)$ of the total variability and replace the smallest eigenvalues accounting for the remaining $1-\rho$ fraction of the variability by their average (see Fig. 1): $r_k$ is the smallest integer such that $\text{Tr}(S_k) \geq P(1-\rho)$ Tr$(\bar{S}_k)$. Choosing $\rho = 0.95$ means that 5% of the variability, in the eigendirections associated to the smallest eigenvalues, is assumed to be evenly spread in these directions. In practice, the choice of $\rho = 0.95$ leads to an average rank of $\tilde{r} = 19.6$ (for $P = 8 \times 8$) for a small drop of PSNR as shown in Fig. 4. Among several other covariance approximations that we tested, for instance, the one consisting in keeping only the $r_k$ first directions, the flat tail approximation provided the best trade-off in terms of acceleration and restoration quality. The analyses showing the superiority of the proposed approximation over the more common approach of keeping only the first $r_k$ directions, and the effect of $\rho$ on image quality are included in the supplementary document.

B. Speed-up via a balanced search tree

As shown in Table I, the step Gaussian selection has a complexity of $O(NP^2K)$, reduced to $O(NPK\tilde{r})$ using the flat tail spectrum approximation. This step remains the biggest bottleneck since each query patch has to be compared to all the $K$ components of the GMM. To make this step even more efficient, we reduce its complexity using a balanced search
As described below, such a tree can be built offline by adopting a bottom-up strategy that repeatedly collapses the original GMM to models with fewer components, until the entire model is reduced to a single Gaussian model.

We progressively combine the GMM components from one level to form the level above, by clustering the \( K \) components into \( L < K \) clusters of similar ones, until the entire model is reduced to a single component. The similarity between two zero-mean Gaussian models with covariance \( \Sigma_1 \) and \( \Sigma_2 \) is measured by the Kullback-Leibler (KL) divergence

\[
KL(\Sigma_1, \Sigma_2) = \frac{1}{2} \text{Tr}(\Sigma_2^{-1}\Sigma_1 + \Sigma_1^{-1}\Sigma_2 - 2\text{Id}_P).
\]

Based on this divergence, at each level \( n \), we look for a partition \( \Omega^n \) of the \( K \) Gaussian components into \( L \) clusters (with about equal sizes) minimizing the following optimization problem

\[
\arg\min_{\Omega^n} \sum_{l=1}^{L} \sum_{k_1 \in \Omega^n_l} KL(\Sigma_{k_1}, \Sigma_{k_l}),
\]

such that \( \bigcup_{l=1}^{L} \Omega^n_l = [K] \) and \( \Omega^n_{l_1} \cap \Omega^n_{l_2} = \emptyset \), where \( \Omega^n_l \) is the \( l \)-th set of Gaussian components for the GMM at level \( n \). This clustering problem can be approximately solved using the genetic algorithm of [24] for the Multiple Traveling Salesmen Problem (MTSP). MTSP is a variation of the classical Traveling Salesmen Problem where several salesmen visit a unique set of cities and return to their origins, and each city is visited by exactly one salesman. This attempts to minimize the total distance traveled by all salesmen. Hence, it is similar to our original problem given in eq. (18) where the Gaussian components and the clusters correspond to \( K \) cities and \( L \) salesmen, respectively. Given the clustering at level \( n \), the new GMM at level \( n - 1 \) is obtained by combining the zero-mean Gaussian components such that, for all \( 1 \leq l \leq L \):

\[
w^n_l = \sum_{k \in \Omega^n_l} w^n_k \quad \text{and} \quad \Sigma^n_l = \frac{1}{w^n_l} \sum_{k \in \Omega^n_l} w^n_k \Sigma^n_k,
\]

where \( \Sigma^n_l \) and \( w^n_l \) are the corresponding covariance matrix and weight of the \( k \)-th Gaussian component at level \( n \).

MTSP compared to classical clustering approaches, is that this procedure can be adapted easily to enforce approximately equal sized clusters, simply by enforcing that each salesman visits at least \( 3 \) cities for the last level and \( 2 \) for the other ones.

We also experimented with other clustering strategies such as the hierarchical kmeans-like clustering in [17] and hierarchical agglomerative clustering. With no principled way to enforce even-sized clusters, these approaches, in general, lead to unbalanced trees (with comb structured branches) which result in large variations in computation times from one image to another. Although they all lead to similar denoising performances, we opted for MTSP based clustering to build our Gaussian tree in favor of obtaining a stable speed-up profile for our resulting algorithm. Please refer to the supplementary document for timing comparisons of MTSP vs. other tree building strategies.

In Fig. 2 we show that the tree obtained using MTSP-based clustering is almost a binary tree (left) and also display the types of patches it encodes along a given path (right). Such a balanced tree structure lets one avoid testing each patch against all \( K \) components. Instead, a patch is first compared to the two first nodes in level 1 of the tree, then the branch providing the smallest cost is followed and the operation is repeated at higher levels until a leaf has been reached. Using this balanced search tree reduces the complexity of step Gaussian selection to \( \mathcal{O}(NP \log K) \).

C. Speed-up via the restriction to a random subset of patches

The simplest and most effective proposed acceleration consists of subsampling the set \( \mathcal{I} \) of \( N \) patches to improve the complexity of the four most time-consuming steps, see Table I. One approach, followed by BM3D [8], consists of restricting the set \( \mathcal{I} \) to locations on a regular grid with spacing \( s \in [1, \sqrt{P}] \) pixels in both directions, leading to a reduction of complexity by a factor \( s^2 \). We refer to this approach as the regular patch subsampling. A direct consequence is that \( |\mathcal{I}| = N/s^2 \) and the complexity is divided by \( s^2 \). However, we observed that this strategy consistently creates blocky artifacts revealing the regularity of the extraction pattern. A random
sampling approach, called "jittering", used in the computer graphics community [7] is preferable to limit this effect. This procedure ensures that each pixel is covered by at least one patch. The location \((i_0, j_0)\) of a point of the grid undergoes a random perturbation, giving a new location \((i, j)\) such that

\[
i_0 - \frac{\sqrt{T-s}}{2} \leq i \leq i_0 + \frac{\sqrt{T-s}}{2}
\]

and

\[
j_0 - \frac{\sqrt{T-s}}{2} \leq j \leq j_0 + \frac{\sqrt{T-s}}{2},
\]

where \([ \cdot ]\) denotes the flooring operation. We found experimentally that independent and uniform perturbations offered the best performance in terms of PSNR and visual quality against all other tested strategies. In addition, we also resample these positions at each of the \(T\) iterations and add a (random) global shift to ensure that all pixels have the same expected number of patches covering them.

Figure 3 illustrates the difference between a regular grid and a jittered grid of period \(s = 6\) for patches of size \(P = 8 \times 8\). In both cases, all pixels are covered by at least one patch, but the stochastic version reveals an irregular pattern.

Nevertheless, when using random subsampling, a major bottleneck occurs when \(A^T A\) is not diagonal because the inversion involved in eq. (6) cannot be simplified as in Alg. 1. Using a conjugate gradient is a practical solution but will negate the reduction of complexity gained by using subsampling. To the best of our knowledge, this is the main reason why patch subsampling has not been utilized to speed up EPLL. Here, we follow a different path. We opt for approximating the solution of the original problem (involving all patches) rather than evaluating the solution of an approximate problem (involving random subsample of patches). More precisely, we speed up Alg. 1 by replacing the complete set of indices by the random subset of patches. In this case, step Patch reprojection consists of averaging only this subset of overlapping restored patches. This novel and nuanced idea avoids additional overhead and attains dramatic complexity improvements compared to the standard approach. Note that even in the case of some inverse problems, such as deblurring, super-resolution, inpainting and devignetting, this strategy can still be used in order to avoid conjugate gradient and maintain a large speed-up.

Experiments conducted on our validation dataset show that this strategy used with \(s = 6\) leads to an acceleration of about \(36 \times\) with less than a 0.2dB drop in PSNR. In comparison, for a similar drop of PSNR, the regular patch subsampling can only achieve a \(9 \times\) acceleration with \(s = 3\) (plots included in the supplementary).

D. Performance analysis

Figure 4 shows the image restoration performance and speed-up obtained when the three ingredients are applied separately or in combination. The results are averaged over 40 images from the test set of BSDS dataset [30] that is set aside for validation purposes. The speed-up is calculated with respect to the EPLLc implementation which is labeled “original” in Fig. 4. Among the three ingredients, random subsampling or jittering (labeled “subsampling”) leads to the largest speed-up (32\(\times\)), while the usage of the search tree provides more than 7\(\times\) faster processing. The average speed-up obtained when combining all three ingredients is around 179\(\times\) on our validation set consisting of images of size \(481 \times 321\), for an average drop of PSNR less than 0.5dB. Note that most of the runtime gain is achieved by combining binary search tree and random subsampling (166\(\times\)), while the flat tail acceleration leads to a modest acceleration of less than 10\%. Nevertheless, we include it in our approach because it introduces another parameter (\(\rho\)) that addresses the bottleneck presented by the dimensionality (\(P^2\)) in a unique way. Adjusting \(\rho\) lets a practitioner choose an appropriate operating point suitable for their need in the speed-vs-quality trade-off space without losing too much in terms of quality.

IV. RELATED METHODS

To the best of our knowledge, there are only two other approaches [39], [35] that have attempted to accelerate EPLL. Unlike our approach, these methods focus on accelerating only one of the steps of EPLL namely the Gaussian selection step. Both use machine learning techniques to reduce its runtime.

In [39], the authors use a binary decision tree to approximate the mapping \(\tilde{z}_i \mapsto k_i^*\) performed in step Gaussian selection. At each node \(k\) and level \(n\) of the tree, the patch \(\tilde{z}_i\) is confronted with a linear separator in order to decide if the recursion should continue on the left or right child given by

\[
\langle a_k^*, \tilde{z}_i \rangle + b_k^* \geq 0
\]

where \((a_k^*, b_k^*)\) are the parameters of the hyperplane for the \(k\)-th node at level \(n\). These separators are trained offline on all pairs of \((\tilde{z}_i, k_i^*)\) obtained after the first iteration of EPLL for a given \(\beta\) and noise level \(\sigma\). Once a leaf has been reached, its index provides a first estimate for the index \(k_i^*\). To reduce errors due to large variations among the neighboring pixels, this method further employs a Markov random fields on the resulting map of Gaussian components which runs in \(O(NK)\) complexity.
Hence, their overall approach reduces the complexity of step Gaussian selection from $O(NKP^2)$ to $O(N(PD+K))$, where $D=12$ is the depth of the learned decision tree.

In [35], the authors approximate the Gaussian selection step, by using a gating (feed-forward) network with one hidden layer

$$\tilde{z}_i \mapsto \nu^k + \sum_{j=1}^{Q} \left( \log \nu_j^k + \omega_j^k (\tilde{d}_i)_{j}^2 \right) \text{ with } \tilde{d}_i = V^t \tilde{z}_i \quad (22)$$

where $Q$ is the size of the hidden layer. The matrix $V \in \mathbb{R}^{P \times Q}$ encodes the weights of the first layer, $\omega^k$ corresponds to the weights of the hidden layer and they are learned discriminatively to approximate the exact posterior probability:

$$\tilde{z}_i \mapsto \nu^k + \sum_{j=1}^{P} \left( \log \nu_j^k + \frac{(c_j^k)^2}{\nu_j^k} \right) \text{ with } c_j^k = U_k^t \tilde{z}_i \quad (23)$$

that we encounter in eq. (7) and (8). Theoretically, a new network will need to be trained for each type of degradations, noise levels and choices of $\beta$ (recall that $\nu_j^k = (S_k)_{jj} + \frac{1}{2\beta}$). However, the findings of [35] indicate that applying a network learned on clean patches with $\frac{1}{\beta} = 0$ is effective regardless of the type of degradation or the value of $\beta$. Their main advantage can be highlighted by comparing eq. (22) and (23) where complexity is reduced from $O(NKP^2)$ to $O(NQ(K + P))$. The authors utilize this benefit by choosing $Q = 100$.

Unlike these two approaches, our method does not try to learn the Gaussian selection rule directly (which depends on both the noise level through $1/\beta$ and the prior model through the GMM). Instead, we simply define a hierarchical organization of the covariance matrices $\Sigma_k$. In other words, while the two other approaches try to infer the posterior probabilities (or directly the maximum a posteriori), our approach provides an approximation to the prior model. During runtime, this approximation of the prior is used in the posterior for the Gaussian selection task. Please note that the value of $\beta$ does not play a role in determining the prior. This allows us to use the same search tree independently of the noise level, degradations, etc. Given that the main advantage of EPLL is that the same model can be used for any type of degradations, it is important that this property remains true for the accelerated version. Last but not least, the training of our search tree takes a few minutes while the training steps for the above mentioned approach take from several hours to a few days [39].

Apart from methods that accelerate EPLL, [20] and [19] are two works that use strategies sharing similarities to our proposed flat-tail and binary search tree approximations, respectively. In [20], the authors use a flat-tail GMM to model distribution of noisy patches. Since the GMM is learned directly on noisy patches, the constant value of the tail corresponds to the noise variance. This allows to improve inference when retrieving the intrinsic dimension of each cluster. This is different from our approach which uses GMM priors learned on clean patch and provides flat-tail approximation using mean of least significant coefficients. In [19], the authors introduce a general data structure called covariance tree (CovTree). A CovTree is constructed by first building a binary space partition tree on patches (data points) and patches in each node are then modeled using a Gaussian distribution. While the resulting tree of Gaussians do share some similarities with our binary search tree, the learning process is very different. The covariance tree is built directly on data points, our approach is applied on an already learned GMM. In addition, unlike CovTree, our proposed strategy can handle any number of components (not just powers of two) and also encodes all parameters of mixture components including the mixing weights.

In the next section, we show that our proposed accelerations produce restoration results with comparable quality to competing methods while requiring a smaller amount of time.

V. Numerical experiments

In this section, we present the results obtained on various image restoration tasks. Our experiments were conducted on standard images of size $512 \times 512$ such as Barbara, Boat, Couple, Fingerprint, Lena, Mandrill and on 60 test images of size $481\times321$ from the Berkeley Segmentation Dataset (BSDS) [30] (the original BSDS test set contains 100 images, the other 40 was used for validation purposes while setting parameters $\rho$ and $s$). For denoising, we compare the performance of our fast EPLL (FEPLL) to the original EPLL algorithm [43] and BM3D [8]. For the original EPLL, we have included timing results given by our own MATLAB/C implementation (EPLLc) and the MATLAB implementation provided by the authors (EPLLm). We also compare our restoration performance and runtime against other fast restoration methods introduced to achieve competitive trade-off between runtime efficiency and image quality. These methods include RoG [35] (a method accelerating EPLL based on feedforward networks described in Sec. IV), and CSF [36] (a fast restoration technique using random field-based architecture).

For deblurring experiments, we additionally compare with field-of-experts (FoE)-based non-blind deconvolution [5] denoted as iPiano. We contacted the corresponding author of [39] and got confirmation that the implementation of their algorithm (briefly described in IV) is not publicly available. Due to certain missing technical details, we were unable to reimplement it faithfully. However, the results reported in [39] indicate that their algorithm performs in par with BM3D in terms of both PSNR and time. Hence, BM3D results can be used as a faithful proxy for the expected performance of Wang et al.’s algorithm [39].

To explicitly illustrate the quality vs. runtime tradeoff of FEPLL, we include results obtained using a slightly slower version of FEPLL referred to as FEPLL’, that does not use the balanced search tree and uses a flat tail spectrum approximation with $\rho = 0.98$. Please note that FEPLL’ is not meant to be better or worse than FEPLL, it is just another version running at a different PSNR/time tradeoff which allows us to compare our algorithm to others operating in different playing fields.

Finally, to illustrate the versatility of FEPLL, we also include results for other inverse problems such as denoising, super-resolution, and inpainting. Additional results can be found in the supplementary.

Parameter settings: In our experiments, we use patches of size $P = 8 \times 8$, and the GMM provided by Zoran et
TABLE II
PSNR, SSIM AND EXECUTION TIME ON THE BSDS TEST SET (AVERAGE ON 60 IMAGES OF SIZE 481×321), AND ON SIX STANDARD IMAGES (EACH OF SIZE 512×512 AND RESULTS AVERAGED OVER 10 INDEPENDENT NOISE REALIZATIONS) FOR THE PROPOSED FEPLL AND FEPLL', EPLL (WITH TIMING GIVEN FOR BOTH EPLLm [43] AND OUR EPLLc), BM3D [8], CSF [36], RoG [35] AND DnCNN [41] WITH 3 DIFFERENT LEVELS OF NOISE.

<table>
<thead>
<tr>
<th>σ</th>
<th>Algo.</th>
<th>Berkeley</th>
<th>Barbara</th>
<th>Boat</th>
<th>Couple</th>
<th>Fingerprint</th>
<th>Lena</th>
<th>Mandrill</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>RoG</td>
<td>37.1 / 0.959</td>
<td>36.7 / 0.954</td>
<td>36.5 / 0.920</td>
<td>37.2 / 0.946</td>
<td>36.4 / 0.987</td>
<td>38.3 / 0.939</td>
<td>35.1 / 0.956</td>
</tr>
<tr>
<td></td>
<td>EPLL</td>
<td>37.3 / 0.963</td>
<td>37.6 / 0.962</td>
<td>36.8 / 0.933</td>
<td>37.3 / 0.950</td>
<td>36.4 / 0.987</td>
<td>38.6 / 0.944</td>
<td>35.2 / 0.960</td>
</tr>
<tr>
<td></td>
<td>DnCNN</td>
<td>5.9 / 0.915</td>
<td>5.5 / 0.970</td>
<td>4.4 / 0.878</td>
<td>3.5 / 0.893</td>
<td>3.2 / 0.968</td>
<td>6.3 / 0.970</td>
<td>3.3 / 0.899</td>
</tr>
<tr>
<td>20</td>
<td>RoG</td>
<td>29.4 / 0.928</td>
<td>28.4 / 0.838</td>
<td>30.5 / 0.815</td>
<td>30.3 / 0.827</td>
<td>28.3 / 0.922</td>
<td>32.5 / 0.865</td>
<td>26.5 / 0.794</td>
</tr>
<tr>
<td></td>
<td>BM3D</td>
<td>29.4 / 0.824</td>
<td>31.7 / 0.904</td>
<td>30.8 / 0.824</td>
<td>30.7 / 0.842</td>
<td>28.8 / 0.928</td>
<td>33.0 / 0.876</td>
<td>26.6 / 0.794</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>29.0 / 0.805</td>
<td>28.3 / 0.821</td>
<td>30.2 / 0.802</td>
<td>29.9 / 0.812</td>
<td>28.0 / 0.916</td>
<td>31.8 / 0.837</td>
<td>26.1 / 0.778</td>
</tr>
<tr>
<td></td>
<td>DnCNN</td>
<td>30.0 / 0.847</td>
<td>31.0 / 0.896</td>
<td>31.1 / 0.831</td>
<td>31.0 / 0.848</td>
<td>28.8 / 0.930</td>
<td>33.4 / 0.883</td>
<td>27.0 / 0.815</td>
</tr>
<tr>
<td>60</td>
<td>RoG</td>
<td>24.6 / 0.623</td>
<td>23.3 / 0.626</td>
<td>25.6 / 0.654</td>
<td>25.2 / 0.640</td>
<td>22.4 / 0.722</td>
<td>27.4 / 0.742</td>
<td>21.4 / 0.668</td>
</tr>
<tr>
<td></td>
<td>BM3D</td>
<td>24.6 / 0.637</td>
<td>26.3 / 0.757</td>
<td>25.9 / 0.671</td>
<td>25.6 / 0.666</td>
<td>23.8 / 0.801</td>
<td>28.2 / 0.778</td>
<td>21.7 / 0.500</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>22.0 / 0.489</td>
<td>21.4 / 0.490</td>
<td>22.7 / 0.502</td>
<td>22.5 / 0.505</td>
<td>22.1 / 0.727</td>
<td>23.4 / 0.511</td>
<td>20.2 / 0.466</td>
</tr>
<tr>
<td></td>
<td>DnCNN</td>
<td>25.3 / 0.665</td>
<td>24.9 / 0.717</td>
<td>26.4 / 0.686</td>
<td>26.0 / 0.683</td>
<td>23.1 / 0.781</td>
<td>28.5 / 0.790</td>
<td>22.1 / 0.555</td>
</tr>
</tbody>
</table>

Time (in seconds)

<table>
<thead>
<tr>
<th>σ</th>
<th>Algo.</th>
<th>Berkeley</th>
<th>Barbara</th>
<th>Boat</th>
<th>Couple</th>
<th>Fingerprint</th>
<th>Lena</th>
<th>Mandrill</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>RoG</td>
<td>0.27</td>
<td>0.38</td>
<td>0.38</td>
<td>0.37</td>
<td>0.36</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>EPLL</td>
<td>0.96</td>
<td>1.28</td>
<td>1.28</td>
<td>1.28</td>
<td>1.26</td>
<td>1.28</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td>EPLLc</td>
<td>43.82</td>
<td>71.14</td>
<td>71.24</td>
<td>71.31</td>
<td>71.28</td>
<td>71.23</td>
<td>71.31</td>
</tr>
<tr>
<td>20</td>
<td>RoG</td>
<td>1.16</td>
<td>1.92</td>
<td>1.92</td>
<td>1.93</td>
<td>1.91</td>
<td>1.93</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>BM3D</td>
<td>1.60</td>
<td>2.32</td>
<td>2.68</td>
<td>2.59</td>
<td>2.17</td>
<td>2.61</td>
<td>2.64</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>0.87</td>
<td>1.09</td>
<td>1.09</td>
<td>1.13</td>
<td>1.09</td>
<td>1.09</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>CSFp</td>
<td>0.38</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>DnCNN</td>
<td>2.09</td>
<td>3.57</td>
<td>3.55</td>
<td>3.60</td>
<td>3.60</td>
<td>3.60</td>
<td>3.61</td>
</tr>
<tr>
<td>60</td>
<td>RoG</td>
<td>0.28</td>
<td>0.63</td>
<td>0.63</td>
<td>0.64</td>
<td>0.63</td>
<td>0.64</td>
<td>0.64</td>
</tr>
</tbody>
</table>

(a) Reference x  
(b) Noisy image y  
(c) BM3D result  
(d) EPLLc result  
(e) FEPLL result  

Fig. 5. Illustration of a denoising problem with noise standard deviation σ = 20. Part of: (a) the original image (b) its noisy version (c-e) denoised results of competitive methods with PSNR/SSIM and time (inset).

al. [43] with K = 200 components. The 200-components GMM is progressively collapsed into smaller GMMs with K = 64, 32, 16, 8, 4, 2 and 1, and then all Gaussians of the tree are modified offline by flat-tail approximations with ρ = 0.95. The final estimate for the restored image is obtained after 5 iterations of our algorithm with β set to λσ−2 [1, 4, 8, 16, 32] where λ = min{N−1||A′A||F,||A||2, 250σ2}. For denoising, where A is identity, λ = 1 which boils down to the setting used by Zoran et al. [43]. For inverse problems, we found that the initialization ẑ = (A′A + 0.2σ2/λ∇)−1A′y, with ∇ the image Laplacian, provides relevant solutions whatever the linear operator A and the noise level σ2. While the authors of [43] do not provide any further direction for setting β and the initialization in general inverse problems, our proposed setting leads to competitive solutions irrespective of A and σ2. For BM3D [8], EPLLm [43], RoG [35], CSF [36] and iPiano [5] we use the implementations provided by the original authors and use the default parameters prescribed by them.

Denoising: Table II shows the quantitative performances of FEPLL on the denoising task compared to EPLLm [43],
EPLLc (our own MATLAB/C implementation), RoG [35] BM3D [8], CSF [36] and DnCNN [42]. We evaluate the algorithms under low-, mid- and high-noise settings by using Gaussian noise of variance $5^2$, $20^2$ and $60^2$, respectively. The result labeled “Berkeley” is an average over 60 images from the BSDS testing set [30]. All numbers are averages obtained on 10 independent noise realizations. Figures 6 provide graphical representations of these performances in terms of PSNR/SSIM versus computation time for the BSDS images for the noise variance setting $\sigma^2 = 20^2$. In this figure, we have also included a recent approach based on weighted nuclear norm minimization (WNNM) [18]. On average, FEPPL results are 0.5dB below regular EPLL and BM3D; however, FEPPL is approximately 7 times faster than BM3D, 170-200 times faster than EPLLc and over 350 times faster than EPLlm. FEPPL outperforms the faster CSF algorithm in terms of both PSNR and time. In this case, FEPPL is even faster than the GPU accelerated version of CSF (CSF$_{gpu}$). Our approach is 4 times faster than RoG with a PSNR drop of 0.1-0.3dB. Nevertheless, if we slow down FEPPL to FEPPL’, we can easily neutralize this quality deficit while still being faster than RoG. While DnCNN offers better results, DnCNN (CPU) is about $10 \times$ slower than FEPPL (also CPU). Our FEPPL (CPU) is also slightly faster than DnCNN (GPU). WNNM offers comparable results to FEPPL’ but is about $500 \times$ slower. Note that these accelerations are obtained purely based on the approximations and no parallel processing is used. Also, in most cases, a loss of 0.5dB may not affect the visual quality of the image. To illustrate this, we show a sample image denoised by BM3D, EPLL and FEPPL in Fig. 5.

Deblurring: Table III shows the performance of FEPLL when used for deblurring as compared to RoG [35], iPiano [5] and CSF [36]. Once again, FEPPL uses Alg. 1 with operator $A$ in the Image estimation step defined by $Ax = F^{-1}[F[h] \odot F[x]]$, where $F$ is the fast Fourier transform and $F^{-1}$ its inverse, $\odot$ indicates element-wise product and $h$ is the blur kernel. For these experiments, we use the blur kernel provided by Chen et al. [5] along with their algorithm implementation. The results under the label “Berkeley” are averaged over 60 images from the BSDS test dataset [30]. The results labeled “Classic” is averaged over the six standard images (Barbara, Boat, Couple, Fingerprint, Lena and Mandrill). FEPPL consistently outperforms its efficient competitors both in terms of quality and runtime. Although the GPU version of CSF is faster, the restoration quality obtained by CSF is 2-3dB lower than FEPPL. The proposed algorithm outperforms RoG by 1-1.8dB while running 3 and 5 times faster on “Berkeley” and “Classic” datasets, respectively.

The superior qualitative performance of FEPPL is demonstrated in Fig. 7. For brevity, we only include the deblurring results obtained from the top competitors of FEPPL algorithm in terms of both quality and runtime. As observed, FEPPL provides the best quality vs. runtime efficiency trade-off. In contrast, a deblurring procedure using the regular EPLL is around 350 times slower than FEPPL with the original implementation [43]. Specifically, on the sample image shown in Fig. 7, EPLL provides a qualitatively similar result (not shown in the figure) with a PSNR of 32.7 dB and SSIM of 0.922 in 142 seconds.

Other inverse problems: Unlike BM3D, EPLL and FEPPL are more versatile and handle a wide range of inverse problems without any change in formulation. In Fig. 8, we show the results obtained by FEPPL on problems such as (a) devignetting, which involves a progressive loss of intensity, (b) super-resolution and (c) inpainting. To show the robustness of our method, the input images of size $481 \times 321$ were degraded with zero-mean Gaussian noise with $\sigma = 2$. All of the restoration results were obtained within/under 0.4 seconds and with the same set of parameters explained above (cf. Parameter settings).

### Table III

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>PSNR/SSIM</th>
<th>Time (s)</th>
<th>PSNR/SSIM</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPiano</td>
<td>29.5 / .824</td>
<td>29.53</td>
<td>29.9 / .848</td>
<td>59.10</td>
</tr>
<tr>
<td>CSF$_{pw}$</td>
<td>30.2 / .875</td>
<td>0.50 (0.14$^*$)</td>
<td>30.5 / .870</td>
<td>0.47 (0.14$^*$)</td>
</tr>
<tr>
<td>RoG</td>
<td>31.3 / .897</td>
<td>1.19</td>
<td>31.8 / .915</td>
<td>2.07</td>
</tr>
<tr>
<td>FEPPL</td>
<td>33.1 / .928</td>
<td>0.40</td>
<td>32.8 / .931</td>
<td>0.46</td>
</tr>
<tr>
<td>FEPPL’</td>
<td>33.2 / .930</td>
<td>1.01</td>
<td>33.0 / .933</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Fig. 6. (a) PSNR and (b) SSIM versus time for different restoration methods in a denoising problem with noise standard deviation $\sigma = 20$. PSNR, SSIM and time are averaged on the 60 BSDS test images, each of size $481 \times 321$. Optimal methods tend to be in the top-left corner.
In this paper, we accelerate EPLL by a factor greater than 100 with negligible loss of image quality (less than 0.5dB). This is achieved by combining three independent strategies: a flat tail approximation, matching via a balanced search tree, and stochastic patch sampling. We show that the proposed accelerations are effective in denoising and deblurring problems, as well as in other inverse problems such as super-resolution and denoising. An important distinction of the proposed accelerations is their genericity: the accelerated EPLL prior can be applied to many restoration tasks and various signal-to-noise ratios, in contrast to existing accelerations based on learning techniques applied to specific conditions (such as image size, noise level, blur kernel, etc.) and that require an expensive re-training to address a different problem.

Since the speed-up is achieved solely by reducing the algorithmic complexity, we believe that further inclusion of accelerations based on parallelization and/or GPU implementations will allow for real-time video processing. Moreover, the acceleration techniques introduced in this work are general strategies that can be used to speed up other image restoration and/or related machine learning algorithms. For reproducibility purposes, the code of FEPLL is made available on GitHub\(^1\).

VI. CONCLUSION

REFERENCES

Fig. 8. FEPLL on various inverse problems. All inputs contain Gaussian noise with $\sigma = 2$. Top row: (a) the observation in a devignetting problem, (b) the bi-cubic interpolation and the actual low-resolution size image (inset) in a $\times 3$ super-resolution problem and (c) the observation in an inpainting problem with 50% of missing pixels shown in red. Bottom row: respective FEPLL results all obtained in less than 0.4s.


[38] J. Sulam and M. Elad. Expected patch log likelihood with a sparse prior.
Supplementary for

"Accelerating GMM-based patch priors for image restoration: Three ingredients for a 100× speed-up"

Shibin Parameswaran∗, Charles-Alban Deledalle∗†, Loïc Denis‡ and Truong Q. Nguyen∗.

August 8, 2018

This document gives additional results and analysis of FEPLL algorithm described on “Accelerating GMM-based patch priors for image restoration: Three ingredients for a 100× speed-up”.

1 Additional analysis

Flat tail approximation based acceleration First, we demonstrate the superiority of the proposed flat tail approximation over the commonly used approach of ignoring the least significant eigendirections. In Supplemental Figure 1, we show that the naïve approach of zeroing out (or ignoring) coefficients of the least significant eigendirections is inferior to the proposed approach of replacing these coefficients by the mean. Our approach can provide reasonably good performance even for very low values of $\rho$. For example, when $\rho = 0.20$ our approach leads to a PSNR of 27.5 compared to 25.1 obtained by the zeroing out strategy.

In a second experiment, we analyze the influence of the parameter $\rho \in (0, 1]$ used in the flat tail approximation. Supplemental Figure 2 shows the curve of PSNR as a function of speed-up for varying values of $\rho$. This experiment is repeated twice either with or without enabling the other two accelerations. Visual results highlight that as $\rho$ decreases to zero, residual noise starts appearing around salient structures. From these experiments, the choice of $\rho = 0.95$ leads to a good trade-off in terms of speed and visual quality for a drop of PSNR lower than 0.2dB.

Search tree acceleration In the second supplementary experiment, we analyze the influence of the clustering method used to successively collapse the GMM model at a given level into a smaller model at a lower level. We compare the proposed approach based on Multiple Traveling Salesmen Problem [2] (MTSP) using the symmetric KL divergence as a semi-metric to a hierarchical GMM clustering algorithm similar to K-means based on KL divergence as proposed in [1] and the standard hierarchical agglomerated clustering (HAC) using symmetric KL divergence. Supplemental Figure 3 shows that MTSP leads to a well-balanced tree, with a height of 7 (almost a binary tree except for the last level due to 200 mixture components). In comparison, using [1]’s K-Means inspired algorithm provides a tree of height 7 but is not balanced and HAC leads to a tree of height 59 with comb structured branches. Please note that a shorter tree is preferred for a faster Gaussian selection step. The tree built using the proposed MTSP strategy not only leads to better PSNR/SSIM, but also provides a more stable computation time for all images of a fixed size (irrespective of content) due to its balanced structure. This is confirmed in Supplemental Figure 4 that displays box-plots

∗Department of Electrical and Computer Engineering, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093, e-mail: sparames@ucsd.edu, cdeledalle@eng.ucsd.edu and tqn001@eng.ucsd.edu.
†IMB, CNRS, Univ. Bordeaux, Bordeaux INP, F-33405 Talence, France, e-mail: charles-alban.deledalle@math.u-bordeaux.fr
‡Univ Lyon, UJM-Saint-Étienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Étienne, France, e-mail: loic.denis@univ-st-etienne.fr.
Supplemental Figure 1: Comparison between the naïve approach of zeroing out (ignoring) coefficients of the least significant eigendirections versus the proposed flat-tail approximation approach of replacing the coefficients with mean. Performance is compared in terms of PSNR obtained while denoising with different values of $\rho$ (percent energy captured) used for thresholding singular values.

Supplemental Figure 2: Top: PSNR as a function of speed-up for varying values of the proportion $\rho$ used in the flat tail approximation when (a) no other accelerations are used and (b) the other two accelerations are also used. The parameter offering the best speed-up for a drop of at most $-0.2$dB is indicated. Bottom: (c) a noisy image with $\sigma=20$ and (d-g) FEPLL results obtained for increasing value of $\rho$ (includes the other two accelerations as well).

obtained from computation time statistics based on five independent runs of the algorithm on 40 different images of the BSDS dataset (all images have the same size). In contrast to our MTSP based strategy, the trees built using [1]’s method and HAC strategy lead to computation times that vary drastically depending on the image content.

Stochastic patch sub-sampling The last analysis focuses on the influence of the period $1 \leq s \leq \sqrt{P}$ used in the stochastic patch sub-sampling. In addition, we also compare against regular sub-sampling. Supplemental Figure 5 shows the curve of PSNR as a function of speed-up for varying values of $s$ where $P = 8$. The experiment has been performed in both cases where either the other two accelerations are disabled or enabled. Visual results highlight that as $s$ increases to $\sqrt{P}$, blocky
artifacts start appearing especially when using a regular patch sub-sampling. In comparison, the stochastic sub-sampling leads to competing results even for $s = \sqrt{P}$, which corresponds to a regular grid that is globally shifted by a random shift at each iteration of FEPLL. These experiments show that the choice of $s = 6$ leads to good trade-off in terms of speed and visual quality for a PSNR drop of less than 0.2dB.

2 Additional results

Denoising results As the first set of additional results, we provide the visual results obtained on five of the standard images: Barbara, Boat, Couple, Fingerprint, Lena and Mandrill, in the setting $\sigma = 20$. Results are displayed in Supplemental Figure 6.

Deconvolution results As the second set additional results, we show deconvolution results for five of testing images from the BSDS dataset [3] with a Gaussian blur setting of 3 pixels (standard deviation 3) and noise level $\sigma = 2$. Results are displayed in Supplemental Figure 7.

Super-resolution results As the last set of additional results, super-resolution results are given for five of the testing images from the BSDS dataset [3] for super-resolution by a factor of 3 with a noise level of $\sigma = 2$. Note that the sub-sampling operator also includes a small Gaussian blur of 0.5 pixels as well as a Kaiser window apodization for anti-aliasing. See Supplemental Figure 8.

References


Supplemental Figure 3: Influence of clustering techniques on the search tree and the results. (left) Multiple Traveling Salesmen Problem [2] (MTSP), (center) Hierarchical K-means like clustering proposed in [1] and (right) Hierarchical agglomerated clustering (HAC). (a-c) The corresponding trees. (d-i) The corresponding results on two denoising problems with $\sigma = 20$ involving two different images of size $481 \times 321$. PSNR/SSIM and time are indicated for each result.

Supplemental Figure 4: Timing scatter obtained over 5 runs for 40 images using Gaussian trees built using different clustering methods we tested. “MTSP” refers to the Multiple Traveling Salesman (proposed), “K-Means” is the method introduced by [1] and “HAC” is a simple Hierarchical agglomerative clustering approach.
Supplemental Figure 5: PSNR as a function of speed-up for varying values of (top) the patch extraction expected period $s$ when (a) no other accelerations are used and (b) the other two accelerations are also used. Stochastic and regular sub-sampling are compared. The parameter offering the best speed-up within a drop of at most $-0.2\text{dB}$ is indicated. (c) A reference image, (h) its noisy version with $\sigma = 20$ and results obtained for increasing value of $s$ (including the other two accelerations as well) for (d-g) the stochastic version and (i-l) the regular version.
Supplemental Figure 6: Denoising results where $\sigma = 20$ obtained on five standard images, from top to bottom: Barbara, Boat, Couple, Fingerprint, Lena and Mandrill.
**Supplemental Figure 7:** Illustration of a deblurring problem of a Gaussian convolution of width 3 pixels with a noise of standard deviation $\sigma = 2$. (a) The original high-resolution (HR) image $x$. (b) The low-resolution (LR) image $y = Ax + w$ and its bicubic interpolation. (c) The super-resolution result $\hat{x}$ obtained by our Fast EPLL.
Supplemental Figure 8: Illustration of a super-resolution by a factor $\times 3$ with a noise of standard deviation $\sigma = 2$. (a) The original high-resolution (HR) image $x$. (b) The low-resolution (LR) image $y = Ax + w$ and its bicubic interpolation. (c) The super-resolution result $\hat{x}$ obtained by our Fast EPLL.