Fast Bayesian JPEG Decompression and Denoising With Tight Frame Priors

Michal Šorel and Michal Bartoš

***Abstract*— JPEG decompression can be understood as an image reconstruction problem similar to denoising or deconvolution. Such problems can be solved within the Bayesian maximum a posteriori probability framework by iterative optimization algorithms. Prior knowledge about an image is usually described by the l_1 norm of its sparse domain representation. For many problems, if the sparse domain forms a tight frame, optimization by the alternating direction method of multipliers can be very efficient. However, for JPEG, such solution is not straightforward, e.g., due to quantization and subsampling of chrominance channels. Derivation of such solution is the main contribution of this paper. In addition, we show that a minor modification of the proposed algorithm solves simultaneously the problem of image denoising. In the experimental section, we analyze the behavior of the proposed decompression algorithm in a small number of iterations with an interesting conclusion that this mode outperforms full convergence. Example images demonstrate the visual quality of decompression and quantitative experiments compare the algorithm with other state-of-the-art methods.**

***Index Terms*— Image processing, image restoration, sparsity, JPEG, ADMM, denoising.**

I. Introduction

LOSSY compression of images using the JPEG standard [1] based on the quantization of discrete cosine transform (DCT) coefficients has become a standard way to store image data. Decompression specified in the JPEG standard was created mainly with speed in mind and typically results in artifacts along strong edges and a visually disturbing checkerboard pattern. However, decompression can be thought of as an image restoration problem. Indeed, since the adoption of the JPEG standard in 1992, the image processing community has worked on finding efficient and precise methods to restore the original data. Similarly to other specialized problems, the progress in JPEG restoration mostly reflects developments in general image restoration.

JPEG decompression/restoration techniques can be naturally categorized based on the Bayesian maximum a posteriori probability (MAP) framework. Even for various ad hoc artifact removal filters and other heuristic methods trying to reduce what is disturbing from a human perspective by suppressing the checkerboard pattern and smoothing along strong edges, we can usually find an intuitive explanation as an approximation of the MAP view.

The Bayesian MAP approach estimates the posterior probability of possible solutions and chooses the image with the highest probability. Following the Bayes formula, this posterior probability is proportional to the product of the likelihood, describing the error introduced during the compression process, and of an approximation of the prior probability. As a rule, instead of maximizing the probability, we equivalently minimize the negative log-probability, which transforms the product to the sum of the negative log-likelihood and a regularization function.

In JPEG restoration, Bayesian likelihood is considered in two forms representing two large groups of algorithms. The first group works with precise likelihood corresponding to the quantization constraint set (QCS) defined as an interval of DCT coefficients, rounding of which could have resulted in the integer coefficients stored in the JPEG file [2]–[6].

Instead of the QCS, the second group works with the distribution of quantization error approximated by a multivariate Gaussian function, which corresponds to a multivariate Gaussian function also in the spatial domain. Its variance is spatially varying with higher values along edges of JPEG blocks, and its covariance matrix has non-negligible off-diagonal elements [7]. Compared to the QCS, the Gaussian likelihood is differentiable and resulting function has no constraints, which simplifies optimization of the posterior probability and speeds up convergence [7]–[9].

In addition to the choice between the QCS and its Gaussian approximation, the quality of reconstruction depends mainly on the choice of image prior probability distribution represented by the corresponding regularization function. Early publications used smooth priors, for example the Huber function of spatial gradient in [8] or weighted quadratic function of gradient in [2]. Later methods incorporated non-differentiable sparse priors that provided state-of-the-art results for many other image restoration problems [10]. These include the total variation (TV) [3], [11], fields of experts (FoE) [9], total generalized variation (TGV) [4], [5], wavelets [12], non-local means [13] and sparse dictionaries [14], [15]. We should also mention algorithms that can be interpreted as working with DCT domain priors [16]. State-of-the-art algorithms usually build on the ideas of sparsity and non-local means

Manuscript received January 14, 2016; revised June 19, 2016, September 13, 2016, and November 4, 2016; accepted November 5, 2016. Date of publication November 10, 2016; date of current version December 2, 2016. This work was supported by the Czech Science Foundation under Grant GA16-13830S. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Abd-Krim Karim Seghouane.

The authors are with the Department of Image Processing, Institute of Information Theory and Automation, Czech Academy of Sciences, 182 08 Prague, Czech Republic (e-mail: sorel@utia.cas.cz; bartos@utia.cas.cz).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2627802

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2627802, IEEE Transactions on Image Processing

denoising [17]–[20] or neural networks [21], [22], achieving good results at the cost of longer run times.

The choice of the algorithm to optimize the MAP criterion depends on image priors. Early methods that used smooth priors applied gradient descent or for the QCS the projected gradient descent [8], [11]. Convexity of the QCS motivated the use of the projection on convex sets (POCS) method [23], [24]. As the POCS method works with no objective function, only constraint sets, these decompression methods replace smoothness priors by inequalities.

Bayesian estimation with the sparse priors based on l_1 norm found its use in many image processing, compressed sensing and machine learning applications. One difficulty with these priors is that resulting non-smooth functions are difficult to optimize by standard methods. This helped spreading of several first-order techniques for non-smooth optimization [25]–[28], which are relatively simple to implement and are fast enough to be practical. Probably the most popular are the alternating direction method of multipliers (ADMM) [29], also known as the split-Bregman method [26], and the accelerated Arrow-Hurwicz algorithm [30].

For JPEG decompression, the algorithm [30] applied on the QCS formulation with TV regularization was used in [3]. The main disadvantage of TV in this context is that it favors unnatural piecewise constant functions, which is exactly the character of JPEG artifacts on block boundaries. To alleviate this problem, in [4] and [5], the same authors proposed a modified regularization term - total generalized variation.

Based on our good experience with ADMM in other problems [31], this paper investigates the application of ADMM on the combination of the Gaussian approximation of the QCS and tight frame priors. In general, for image restoration problems with the sparse priors based on l_1 norm, ADMM can be very fast under the assumption that there is a closed-form non-iterative solution for the inverse of an operator corresponding to the regularized Hessian of image degradation (see Sec.~IV). The time complexity of this inverse depends on the properties of both the degradation and regularization functions. For deconvolution problems, fast inversion is possible in the Fourier domain for priors containing only convolutions, such as TV and FoE [32]. There is a group of degradations, where efficient inversion requires the linear operator used in the regularization function to form a tight frame [33]. Several examples of such degradations are derived for example in [27], including deconvolution, inpainting, and reconstruction from partial Fourier observations.

Tight frames sacrifice the orthonormality and linear independence of orthonormal bases while still enjoying the same efficient decomposition and reconstruction as orthonormal wavelet bases. In recent years, many tight frames have been proposed to more efficiently represent natural images, including the dual-tree complex wavelets (DT-CWT) [34], ridgelets [35], curvelets [36], bandlets [37], and shearlets [38]. Tight frames can be even learned from data [39].

A. Contributions

In this article we show that ADMM can be efficiently applied to solve an approximation of the MAP formulation of JPEG restoration with sparse priors forming a tight frame. To this end, we replace quantization noise by its Gaussian approximation and apply the Woodbury matrix inversion formula to express the inverse of the regularized Hessian of degradation operator, which has in this case a relatively complicated structure, consisting of the block-wise DCT, quantization and for chrominance channels also down-sampling. We derive how this inverse can be computed directly in the DCT domain.

Using the Gaussian approximation of quantization noise has several advantages. Likelihood becomes strongly convex, which in general improves convergence. Perhaps surprisingly, its use improves reconstruction quality both visually and in terms of the signal-to-noise ratio (SNR), as we show in our experiments. It also makes the result less sensitive to the number of iterations.

Asymptotical convergence properties of ADMM are well known. In contrast to most articles, we concentrate on its behavior in a small number of iterations, which is in practical applications more relevant. An important conclusion we draw is that a small number of iterations, such as five, can achieve a better ISNR (improvement in SNR) than much larger numbers. This also justifies the use of ADMM instead of accelerated primal-dual methods that in theory further improve convergence by preconditioning or utilizing the strong convexity of the likelihood function. Our experience is that for a small numbers of iterations ADMM is competitive and sometimes even faster than for example [30], which is not true for higher numbers. Since one iteration of ADMM in our case consists of only two DCTs and two frame transforms (in addition to a few element-wise operations), this makes MAP based iterative methods as practical as non-iterative methods.

In our experiments, we illustrate the behavior of the proposed algorithm for different priors and various stopping criteria, both in terms of SNR and visual impression. We compare the algorithm with similar methods based on the QCS model with the total variation [3] and total generalized variation [4], [5] priors, as well as the state-of-the-art method [20]. The latter gives better SNR but at the cost of much longer running time. Statistical experiments show the mean ISNR and its variance on a set of fifty images.

The Bayesian MAP approach can be extended to other image restoration problems involving JPEG compression, for example the resolution enhancement of compressed videos [16], [40]. Nevertheless, as a rule, their solution is complex and much more time-consuming than simple JPEG decompression. In this paper, we show that as an interesting side-effect of using the Gaussian approximation of quantization noise, a simple modification of the proposed decompression algorithm can be used to remove simultaneously image noise and compression artifacts.

The rest of the paper is organized as follows. Sec. II explains the MAP formulation of the problem of image restoration from JPEG data. Sec. III shortly explains the main optimization tool we use, the alternating direction method of multipliers. For grayscale images, the algorithm is derived in Sec. IV. Sec. V extends the algorithm to color images and discusses a modification that regularizes jointly in all channels. Sec. VI

TABLE I
NOTATION

x	unknown image we estimate
$\tilde{y}$	quantized DCT coefficients stored in JPEG file
y	observed JPEG image, $y = C^{-1}Q^{-1}\tilde{y}$
$\bar{y}$	constant part of the right-hand side of a linear system solved in each iteration of ADMM, $\bar{y} = C^T Q^T \tilde{y}$
$\bar{y}_c$	version of $\bar{y}$ used in chrominance channels
Q	diagonal operator of quantization
C	discrete cosine transform applied to each JPEG block
Φ	Tight frame used for sparsity-based l_1 regularization
D	down-sampling operator used for chrominance channels
q	vector of quantization coefficients, $Q = \mathrm{diag}(q)$
σ_q^2	variance of quantization noise, $\sigma_q^2 = \frac{1}{12}$
σ_g^2	noise variance for the extension to denoising
$a,\ d$	auxiliary variables in ADMM
τ	weight of regularization term
μ	ADMM parameter

shows how the proposed algorithm can be extended to the problem of denoising. Sec. VII demonstrates the performance of the proposed algorithm by both quantitative and qualitative experiments and results are summarized in Sec. VIII. To improve readability, we provide a complete list of used variables and operators in Tab. I.

II. Bayesian JPEG Restoration With Sparse Priors

We start by shortly recalling how the JPEG compression/decompression works and describing the MAP solution of the JPEG decompression problem. We assume that readers are familiar with the basic principles of JPEG compression [41], namely that it is based on the quantization of DCT (discrete cosine transform) coefficients, where the DCT is applied to small blocks of usually 8×8 pixels.

For grayscale images, the JPEG compression-decompression process can be described by a sequence of operators

$$y = C^{-1}Q^{-1}\left[QCx\right] \tag{1}$$

where x is an original image, y the decompressed image, Q and C the linear operators of quantization and DCT, respectively. The square brackets denote the operator of rounding, which can be thought of as a quantization noise. We can imagine that images x and y are vectors and operators Q and C matrices, even though the matrices will never be formed explicitly. C is a block diagonal matrix made up of the square matrices of DCT. C is orthogonal, because all the DCT sub-matrices are orthogonal. Q is a diagonal matrix corresponding to element-wise division by quantization coefficients from the quantization table stored in each JPEG file (64 values for 8×8 blocks), i.e. Q contains vectorized coefficients replicated for each block along diagonal. We will denote the vector of these coefficients as q, i.e. $Q = \mathrm{diag}(q)$. For color images, the image is first transformed into $Y'C_BC_R$ space and individual channels are stored separately. The luminance (brightness) is stored as described above but chrominance channels are often stored at smaller resolution, which complicates the degradation model. This is described in Sec. V.

Given an observation y, a model describing the probability distribution of possible observations $p(y|x)$ and a prior probability $p(x)$, the Bayesian MAP approach maximizes the posterior probability $p(x|y) \sim p(y|x)p(x)$. Since in our case of rounding to the nearest integer the likelihood $p(y|x)$ is uniform within the quantization interval $-0.5 < QCy - QCx \leq 0.5$. MAP estimation corresponds to the maximization of the prior probability $p(x)$ over all x satisfying this interval constraint, the QCS.

It is quite common in the MAP based image restoration that $p(x)$ is given by a sparsity inducing regularization $p(x) \sim \exp(-\tau \left\|\Phi^T x\right\|_1)$, where Φ^T is a linear analysis operator (transform to a sparse domain) such as the gradient, wavelets, or an overcomplete dictionary. We in addition constrain the operator to satisfy $\Phi\Phi^T = I$, which is called a normalized or Parseval frame [33]. Slightly more generally, we can define tight frames as satisfying $\Phi\Phi^T = tI$ (for a scalar constant t), which naturally arises for example from concatenating several orthogonal bases. Nevertheless, since tight frames can always be normalized, it is a common practice to derive algorithms for normalized frames but use the general term tight frames. This is the case of this paper too. In our experiments, we use mainly the DT-CWT tight frame [34] that represents well natural images and has a linear computational complexity. We also show results for the data specific tight frame [39]. Note that the well known constructions of overcomplete systems [35]–[37] describe well cartoon-like images but are not so good for natural images. For the TV and FoE priors in general $\Phi\Phi^T \neq I$ and therefore the procedure we derive cannot be used directly. In this case we can use less efficiently a similar procedure with ADMM replaced by its variant [27] or an alternative not requiring the tight frame condition [30], [42]–[45].

As mentioned in the introduction, an alternative to the QCS is the approximation of quantization error by a Gaussian function with variance σ_q^2

$$QCy = QCx + e, \quad e \sim N(0, I\sigma_q^2), \tag{2}$$

where $\sigma_q^2 = 1/12$ is the variance of the unit quantization noise [9]. Assuming that the errors introduced by rounding of DCT coefficients are independent, since C is orthogonal and Q diagonal, the covariance matrix of spatial domain noise is $E[C^{-1}Q^{-1}ee^TQ^{-T}C] = C^{-1}Q^{-1}E[ee^T]Q^{-T}C = C^{-1}Q^{-2}C\sigma_q^2$.

The MAP solution for this model is a convex problem

$$\arg\min_x \frac{1}{2\sigma_q^2}\left\|\tilde{y} - QCx\right\|^2 + \tau\left\|\Phi^T x\right\|_1, \tag{3}$$

where $\tilde{y} = QCy$ are the quantization coefficients stored in JPEG format. Equation (3) is a special case of what is also known as the analysis-based approach to sparsity restoration [46] with a special form of the degradation operator QC and observation in DCT domain $\tilde{y} = QCy$. The scalar parameter τ can be estimated from training data by fitting distribution

$$p(x) \propto \tau^N e^{-\tau\left\|\Phi^T x\right\|_1}, \tag{4}$$

where N is the dimension of x. The maximum likelihood estimate is straightforward by setting the derivative of (4) to zero, which gives $\tau = N/\left\|\Phi^T x\right\|_1$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2627802, IEEE Transactions on Image Processing

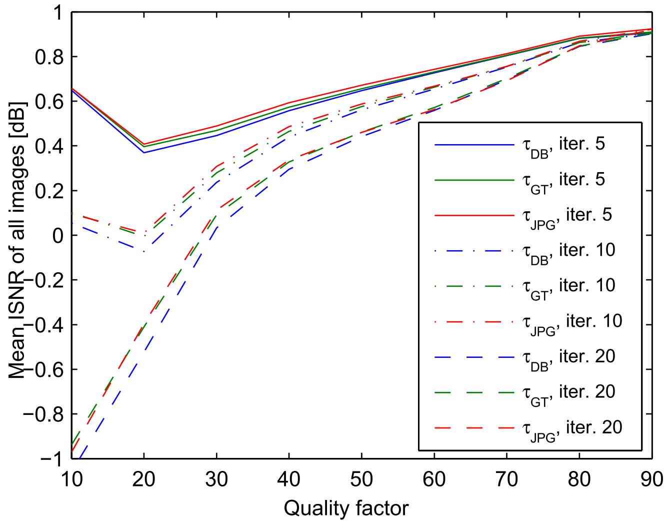


Fig. 1. Comparison of the mean ISNR (over 50 images) for combinations of three ML estimates of τ and three different numbers of iterations. τ_{DB} is trained from a database of images, τ_{GT} from ground truth and τ_{JPG} is estimated from input image. In all cases, 5 iterations achieve the best ISNR, 10 iterations is best visually, 20 iterations is visually close to full convergence.

III. Alternating Direction Method of Multipliers

The main optimization tool we use is the alternating direction method of multipliers (ADMM) [27], [29]. ADMM is a method to minimize the sum of two functions

$$\min_x f(x) + g(Gx), \tag{5}$$

where f and g are convex not necessarily differentiable functions and G a linear operator. ADMM consists of iteratively executing three update steps

$$x \leftarrow \arg\min_x f(x) + \frac{\mu}{2}\|Gx - a - d\|^2, \tag{6}$$

$$a \leftarrow \arg\min_a g(a) + \frac{\mu}{2}\|Gx - a - d\|^2, \tag{7}$$

$$d \leftarrow d - (Gx - a), \tag{8}$$

where scalar $\mu > 0$ is a parameter, a is an auxiliary variable representing a sparse domain counterpart of x and d a dual variable. Convergence is proved in [29]. Stopping criteria are discussed for example in [47], Section 3.3.1.

IV. Algorithm

Since (3) is obviously a special case of (5), its global minimum can be found by ADMM described in the previous section. Since (3) can be multiplied by σ_q^2 without changing its optimum, we can hide σ_q^2 in τ and without loss of generality assume $\sigma_q^2 = 1$. After this simplification, ADMM alternates solution of two convex problems

$$\arg\min_x \frac{1}{2}\|\tilde{y} - QCx\|^2 + \frac{\mu}{2}\|\Phi^T x - a - d\|^2 \tag{9}$$

and

$$\arg\min_a \tau\|a\|_1 + \frac{\mu}{2}\|\Phi^T x - a - d\|^2 \tag{10}$$

supplemented by a simple update of the dual variable. Equation (10) is a fast element-wise thresholding operation

$$a \leftarrow \operatorname{sgn}\left(\Phi^T x - d\right)\max\left(0; |\Phi^T x - d| - \frac{\tau}{2\mu}\right). \tag{11}$$

As a consequence, the critical point of the algorithm is a fast solution of (9), which can be expressed as a linear system

$$(C^T Q^T QC + \mu\Phi\Phi^T)x = C^T Q^T \tilde{y} + \mu\Phi(a + d). \tag{12}$$

The operator we invert can be interpreted as the regularized Hessian of image degradation.

Let us assume that Φ is a tight frame ($\Phi\Phi^T = I$). To further simplify notation, let us denote $\Phi(a+d)$ as z and $C^T Q^T \tilde{y} = C^T Q^T QCy$ as $\bar{y}$, which replaces (12) by

$$(C^T Q^T QC + \mu I)x = (\bar{y} + \mu z)\,. \tag{13}$$

Recall that DCT is orthogonal, i.e. $CC^T = C^T C = I$. Multiplying by C from left

$$(Q^T Q + \mu I)Cx = C\,(\bar{y} + \mu z). \tag{14}$$

Since Q is diagonal, $QQ^T + \mu I = \operatorname{diag}(q^2 + \mu)$ is also diagonal and has a trivial inverse. We get a non-iterative x-update

$$x \leftarrow C^T \operatorname{diag}\left(\frac{1}{q^2 + \mu}\right)C\,(\bar{y} + \mu\Phi(a + d))\,, \tag{15}$$

which only consists of element-wise operations and two DCTs.

V. Extension to Color and Joint Regularization

For decompression of color images, let us first consider three channels of $Y'C_bC_r$ space independently of each other. As a rule, JPEG images store its chrominance channels at half resolution, which gives a slightly modified problem

$$\arg\min_x \frac{1}{2}\|QCDy - QCDx\|^2 + \tau\left\|\Phi^T x\right\|_1 \tag{16}$$

where D is a down-sampling operator. In practice, there are two options for D. It is either a direct sampling at half resolution, i.e. taking every second pixel in each direction, or it is combined with a low-pass operator, which means that D computes the mean value of each square of size 2×2 pixels (in general we will consider averaging over $m \times n$ pixels). Denoting $D^T C^T Q^T QCy$ as $\bar{y}_c$, we get the x-update as

$$x \leftarrow (D^T C^T Q^T QCD + \mu I)^{-1}(\bar{y}_c + \mu z) \tag{17}$$

Again, our purpose is to compute this inverse directly in one step. Sherman-Morrison-Woodbury matrix inversion formula

$$(A + URV^T)^{-1} = A^{-1} - A^{-1}U(R^{-1} + V^T A^{-1}U)^{-1}V^T A^{-1} \tag{18}$$

after substitution $A = \mu I$, $U = D^T C^T Q^T$, $R = I$, $V^T = QCD$ transforms the inverse in (17) to

$$\frac{1}{\mu}\left(I - D^T C^T Q^T\left(QCDD^T C^T Q^T + \mu I\right)^{-1} QCD\right).$$

It can help only if $DD^T = kI$, because then $CkC^T = kI$. Indeed, it holds for both variants we consider. For simple

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2627802, IEEE Transactions on Image Processing

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 2. Difference between the iteration with best ISNR (left), full convergence (middle) and early stopping rule (right) for TV-based method [3] and proposed combination of the Gaussian approximation of the QCS and DT-CWT prior. **Best viewed electronically.** (a) Original. (b) JPEG, quality = 50, 0 dB, 0.05 s. (c) State of the art [20]: 1.21 dB, 140 s. (d) TV [3], max. ISNR: 0.82 dB, 1.5 s, 7 it. (e) TV [3], converged: -0.35 dB, 178 s, 991 it. (f) TV [3], early stopping: 0.63 dB, 2.8 s, 15 it. (g) Proposed DT-CWT, max. ISNR: 0.79 dB, 0.4 s, 5 it. (h) Proposed DT-CWT, converged: 0.36 dB, 30 s, 500 it. (i) Proposed DT-CWT, early stopping: 0.68 dB, 0.7 s, 10 it.

down-sampling $DD^T = I$. In the variant with a low-pass filter averaging over $m \times n$ pixels $k = 1/(mn)$,[1] i.e. for the most common $2:1$ chroma subsampling [1] $k = 1/4$. We obtain

$$x \leftarrow \frac{1}{\mu}\left(I - D^T C^T Q^T \left(kQQ^T + \mu I\right)^{-1} QCD\right)(\bar{y}_c + \mu z)$$
$$= \frac{1}{\mu}\left(I - D^T C^T \operatorname{diag}\left(\frac{q^2}{kq^2 + \mu}\right)CD\right)(\bar{y}_c + \mu\Phi(a + d)) \tag{19}$$

[1]Proof is trivial for direct subsampling. Because D^T replicates one pixel to a square of $m \times n$ pixels and D picks one of them, we get $DD^T = I$. In the latter case, D^T replicates each pixel to $m \times n$ pixels multiplied by a scalar $k = 1/(mn)$. D computes an average of the same values, together giving $DD^T = kI$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2627802, IEEE Transactions on Image Processing

Algorithm 1 Proposed Algorithm for Decompression of Color Images. For Grayscale Images the Sum in Point 5) Disappears, in Other Respects it is a Special Case With $k = 1$ and $D = I$

1) For all three color channels $i = 1 \dots 3$ initialize $a_0^i = y^i$, $d_0^i = 0$ and pre-compute $\bar{y}_c^i = D_i^T C^T Q_i^T Q_i C D_i y^i$.
2) repeat
3) for $i = 1 \dots 3$
4) $x_{n+1}^i = \frac{1}{\mu}\left(I - D_i^T C^T \operatorname{diag}(\frac{q_i^2}{kq_i^2+\mu}) C D_i\right) \cdot \left(\bar{y}_c^i + \mu\Phi\left(a_n^i + d_n^i\right)\right)$
5) $a_{n+1}^i = \frac{\Phi^T x_{n+1}^i - d_n^i}{\sqrt{\sum_{j=1}^3 |\Phi^T x_{n+1}^j - d_n^j|^2}} \cdot \max\left(0; \sqrt{\sum_{j=1}^3 |\Phi^T x_{n+1}^j - d_n^j|^2} - \frac{\tau}{2\mu}\right)$
6) $d_{n+1}^i = d_n^i - \Phi^T x_{n+1}^i + a_{n+1}^i$
7) until stopping criterion is satisfied

which is for $D = I$ equivalent to (15) and compared to (15) adds one up-sampling and one down-sampling operation per iteration.

The procedure described above can be modified to utilize the correlation between color channels by regularizing jointly in all channels

$$\min_x \frac{1}{2}\sum_{i=1}^{3}\left\|Q_i C D_i y^i - Q_i C D_i x^i\right\|^2 + \tau\left\|\sqrt{\sum_{i=1}^{3}\left|\Phi^T x^i\right|^2}\right\|_1, \tag{20}$$

where the absolute value and square power in the regularization term are element-wise operations. Quantization tables and down-sampling operators are specific to each channel i. This approach is a special case of now widely used group sparsity but can be traced back to work on gradients of multi-valued images [48]. Applied to luminance-chrominance model as in this paper, joint regularization was used for example in [49].

Again, we can apply ADMM, which can be done independently for each channel except thresholding (10), solution of which becomes

$$a^i \leftarrow \frac{\Phi^T x^i - d^i}{\sqrt{\sum_{j=1}^3 |\Phi^T x^j - d^j|^2}} \cdot \max\left(0; \sqrt{\sum_{j=1}^{3} |\Phi^T x^j - d^j|^2} - \frac{\tau}{2\mu}\right). \tag{21}$$

For high quality factors, there is little difference between separate and joint regularization. The added value of joint regularization becomes apparent for highly compressed images, where strong quantization of down-sampled chrominance channels typically damages edges. As edge information is preserved in the luminance channel, it is transferred by joint regularization to chrominance channels. The resulting algorithm, including chrominance channels and joint regularization, is summarized in Alg. 1.

VI. Bayesian JPEG Denoising

An interesting consequence of approximating the quantization noise by a Gaussian distribution is a simple extension to denoising. More precisely, we consider the situation of an image degraded by a Gaussian noise $n \sim N(0, \sigma_g^2)$ and then compressed by the JPEG algorithm. The task is to recover the original image by maximizing the posterior probability using the same priors we use in the JPEG decompression.

For simplicity, we first show the derivation for grayscale images. The only difference with respect to (3) is introducing color noise to the observation $QC(x+n) = QCx + QCn$. DCT is orthogonal, so Cn is still white but the multiplication by Q generates a diagonal covariance matrix $E[QCnn^T C^T Q^T] = QCE[nn^T]C^T Q^T = \sigma_g^2 QQ^T = \sigma_g^2 \operatorname{diag}(q^2)$. Combined with the Gaussian approximation of quantization noise (noises are independent of each other) we end up with covariance matrix $\Sigma = \sigma_q^2 I + \sigma_g^2 \operatorname{diag}(q^2) = \operatorname{diag}(\sigma_q^2 + \sigma_g^2 q^2)$. Instead of (3) we solve

$$\arg\min_x \frac{1}{2}(\tilde{y} - QCx)^T \Sigma^{-1}(\tilde{y} - QCx) + \tau\left\|\Phi^T x\right\|_1,$$

which can be again solved by ADMM. The only difference is in the x-update, where instead of (9) we solve

$$\arg\min_x \frac{1}{2}(\tilde{y} - QCx)^T \Sigma^{-1}(\tilde{y} - QCx) + \frac{\mu}{2}\|\Phi^T x - a - d\|^2, \tag{22}$$

which is equivalent to

$$(C^T Q^T \Sigma^{-1} QC + \mu\Phi\Phi^T)x = C^T Q^T \Sigma^{-1}\tilde{y} + \mu\Phi(a + d). \tag{23}$$

Assuming again $\Phi\Phi^T = I$, denoting $\bar{y} = C^T Q^T \Sigma^{-1}\tilde{y}$, $z = \Phi(a + d)$ and multiplying from left by C gives

$$\left(Q^T \Sigma^{-1} Q + \mu I\right) Cx = C\left(\bar{y} + \mu z\right)$$

$$x \leftarrow C^T \operatorname{diag}\Big(\frac{1}{\frac{q^2}{\sigma_q^2 + \sigma_g^2 q^2} + \mu}\Big) C\left(\bar{y} + \mu\Phi(a + d)\right). \tag{24}$$

We can see that we obtained almost the same formula as (15), except the multiplication by the diagonal covariance matrix in the constant term on the right side and that weighting of quantization coefficients now depends on its frequency.

The situation is analogous for chrominance channels regardless of whether we use separate or joint regularization. The noise covariance becomes $\Sigma = \operatorname{diag}(\sigma_q^2 + k\sigma_g^2 q^2)$, the right hand side changes to $\bar{y}_c = D^T C^T Q^T \Sigma^{-1}\tilde{y}$ and (17) becomes

$$x \leftarrow (D^T C^T Q^T \Sigma^{-1} QCD + \mu I)^{-1}(\bar{y}_c + \mu z). \tag{25}$$

In the inversion formula (18) we add $R = \Sigma^{-1}$, which in the x-update (19) changes only the diagonal term $\operatorname{diag}(\frac{q^2}{kq^2+\mu})$ giving

$$x \leftarrow \frac{1}{\mu}\left(I - D^T C^T \operatorname{diag}\Big(\frac{q^2}{kq_f^2 + \mu\left(\sigma_q^2 + k\sigma_g^2 q^2\right)}\Big) CD\right) \cdot \left(D^T C^T \operatorname{diag}\Big(\frac{q}{\sigma_q^2 + k\sigma_g^2 q^2}\Big)\tilde{y} + \mu\Phi\left(a + d\right)\right). \tag{26}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2627802, IEEE Transactions on Image Processing

TABLE II

QUANTITATIVE COMPARISON OF EXPERIMENT IN FIG. 5, JPEG QUALITY = 30. ISNR/PSNR VALUES ARE IN dB. ALL VALUES ARE FOR Y' CHANNEL

Image	Butterfly			Flower			Bike			Parrot			Time [s] /
Method	ISNR	PSNR	SSIM	ISNR	PSNR	SSIM	ISNR	PSNR	SSIM	ISNR	PSNR	SSIM	iterations
Standard JPEG		28.93	0.906		30.67	0.894		27.91	0.895		32.99	0.916	˜0 / 0
DT-CWT	0.71	29.65	0.938	0.36	31.03	0.906	0.54	28.45	0.909	0.48	33.47	0.926	˜2 / 10
Learned frame [39]	0.99	29.92	0.935	0.84	31.51	0.915	0.99	28.90	0.918	0.83	33.82	0.929	˜30 / 20
TGV[5]	1.64	30.57	0.944	0.25	30.92	0.900	0.29	28.20	0.900	0.17	33.15	0.921	˜7 / 50
Liu et al. [20]	2.37	31.30	0.951	1.21	31.88	0.916	1.89	29.80	0.927	1.19	34.18	0.929	˜220 / NA

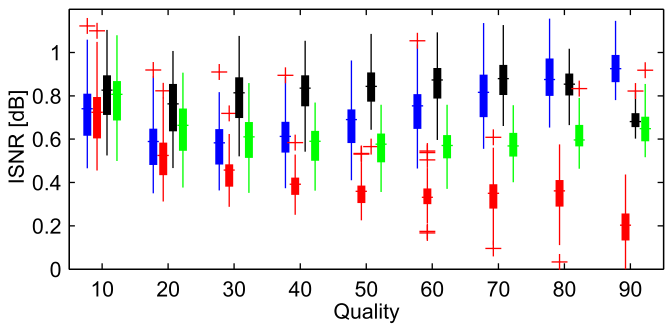


Fig. 3. Statistical comparison of the maximum achievable ISNR on a set of 50 images for the algorithm with the QCS or proposed Gaussian approximation and two different priors. The order of the variants in each group representing the same quality from left to right: Gaussian approx. with DT-CWT, QCS with DT-CWT, Gaussian approx. with learned frame [39], QCS with learned frame [39].

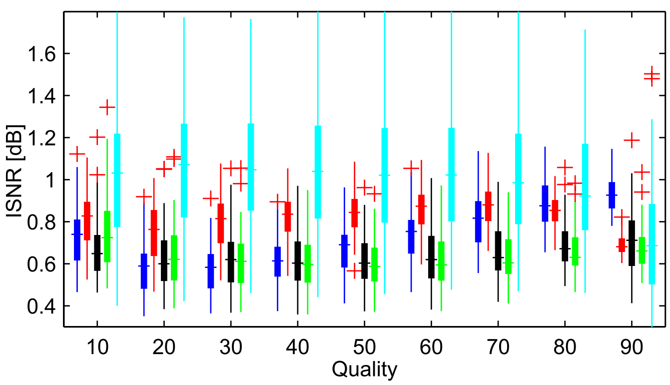


Fig. 4. Statistical comparison of the maximum achievable ISNR on a set of 50 images for different JPEG decompression algorithms. The order of the methods in each group representing the same quality is from left to right: Proposed with DT-CWT, proposed with learned frame [39], QCS with TV priors [3], QCS with TGV priors [5], state-of-the-art method [20].

VII. EXPERIMENTS

In this section, we experimentally demonstrate our contributions. Statistical experiments are carried out on a database of fifty outdoor images pre-processed to contain almost no noise. This was achieved by taking RAW images by a full-frame SLR camera under good lighting conditions and decreasing image resolution by a factor of eight in each direction. Since each pixel of the resulting image was computed as an average of 64 original pixels, noise standard deviation decreased eight times. In addition, we use also images taken from [20]. In all statistical experiments (Figs. 3, 4 and 8) the improvement is given in terms of the ISNR computed only on the luminance channel in the iteration with the best ISNR. In our opinion it is a reasonable option how to show the power of models we compare independently of various stopping rules. We used a variant of the box-and-whisker plot, showing the median value, and first and third quartiles. Wherever in the experiments a variant is denoted as "proposed", it is meant that it uses the Gaussian approximation of the QCS and ADMM as described in the paper. For several example images, we give the values of the ISNR, PSNR and SSIM [50].

The proposed methods were implemented in Matlab without any parallelizations. For comparison we used the implementation of [3], [5], and [20] available on authors' web pages, all of them also in Matlab. For time measurements, we modified [3], [5] to use the DCT implementation from Phil Salee's Matlab JPEG toolbox, which is much faster than the original implementations. JPEG data were also read using this toolbox.

A. JPEG Decompression

We start by demonstrating the fact that the Gaussian approximation of the QCS condition does not harm the reconstruction and even improve results. Fig. 3 shows for two different priors (DT-CWT and the tight frame learned from patches collected from the image database using [39]) that the ISNR is significantly better using the Gaussian approximation than using the QCS. It also shows that the learned frame is slightly better than DT-CWT, which comes at the cost of longer running time. Note that the QCS has no additional parameters and the regularization parameter τ for the case of Gaussian approximation was set by the ML estimate from the image database. The combination of the QCS with the DT-CWT or the learned frame was optimized also using ADMM (Sec. III) with an indicator function instead of the least squares in (3). We describe this algorithm (similar to [3] and [5]) in [6].

A similar experiment in Fig. 4 compares two variants of the proposed approach (Gaussian approximation of the QCS, the DT-CWT and learned frame [39] as priors) with QCS-based algorithms (TV [3], TGV [5]) and the state-of-the-art algorithm [20]. As mentioned earlier in all cases we chose the iteration with the best ISNR but similar differences would be observed using the early stopping rule proposed in [3] and [5]. We can see that the learned frame is slightly better than DT-CWT except very high quality factors, Gaussian approximation overcomes the QCS (except qualities 20 and 30, where DT-CWT is slightly worse than TV and TGV) and [20] gives the best ISNR at all qualities except 90. On the other hand, it comes at the cost of much higher running time (see next experiment Fig. 5 and Tab. II). Surprisingly, in terms

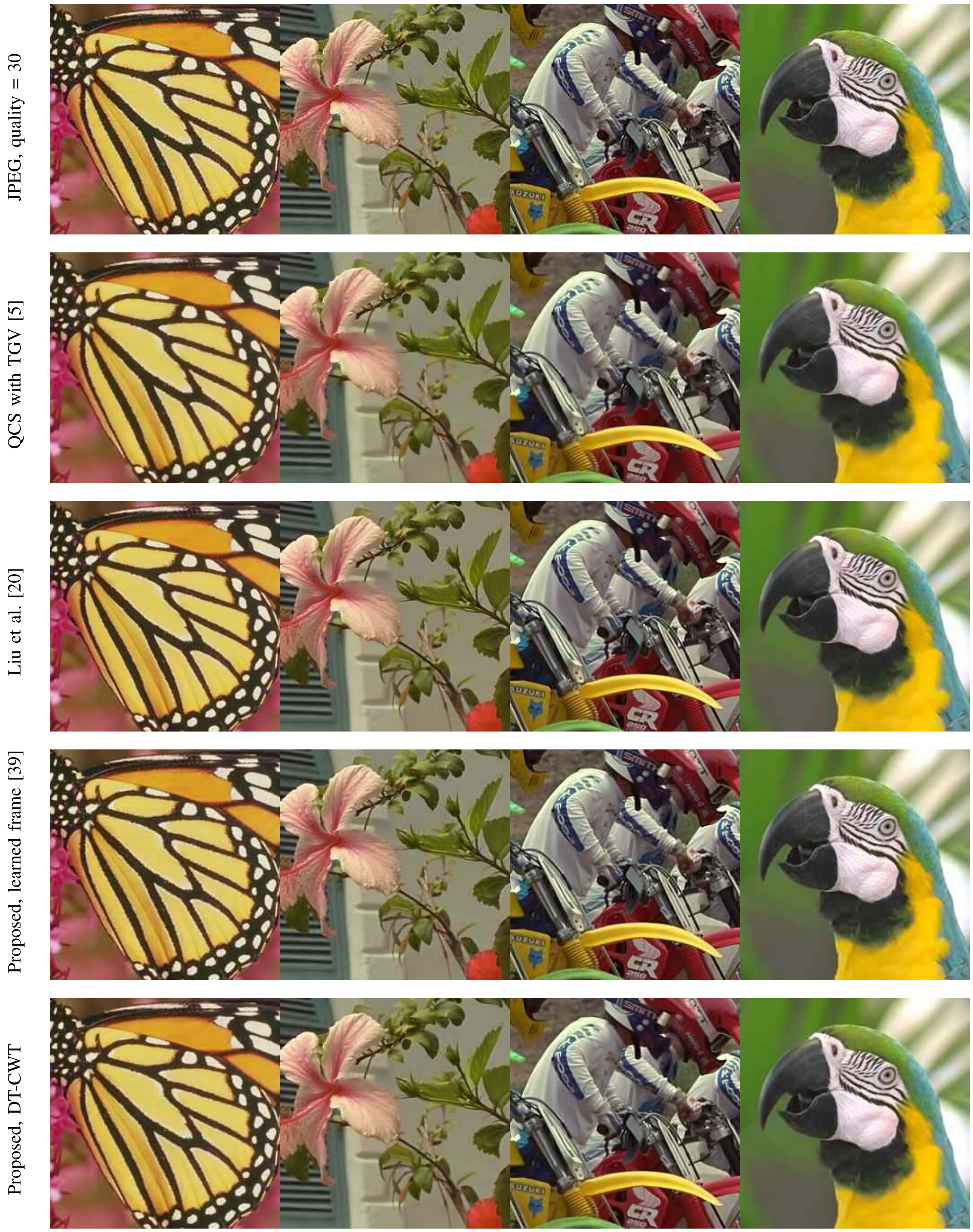


Fig. 5. Visual comparison of the proposed approach using DT-CWT or learned priors [39] with TGV [5] and Liu et al. [20]. The ISNR, PSNR and SSIM values are summarized in Table II. **Best viewed electronically.**

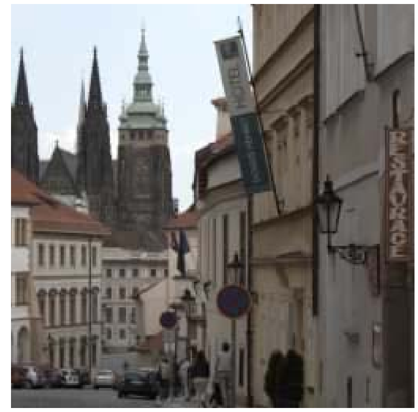
of ISNR in Fig. 4, the results of TGV-based method [5] can be slightly worse than those of the TV-based [3]. This holds especially for higher qualities and small number of iterations, where we get the optimal SNR. For full convergence it would be the other way round (see also Fig. 2(e)). Results are always better than standard JPEG decompression with improvement from 0.4 to 1.1 dB (1.8 dB for [20]). Even though this improvement does not look large, it is important to realize that this occurs mainly along edges that constitute only a small proportion of total image area.

Analogously to other image reconstruction problems, there is a trade-off between quality of reconstruction and time requirements. One of motivations for our research was our experience that elegant formulations with l_1 priors give in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2627802, IEEE Transactions on Image Processing

(a) Standard JPEG decompression, quality=50, SNR (dB) - $Y' : 21.09, C_b : 5.67, C_r : 5.95$, PSNR (dB) - $Y' : 33.0, C_b : 39.3, C_r : 39.7$, SSIM - $Y' : 0.932, C_b : 0.922, C_r : 0.935$

(b) DT-CWT without joint regularization, ISNR (dB) - $Y' : 0.71, C_b : 0.35, C_r : 0.41$, PSNR (dB) - $Y' : 33.7, C_b : 39.6, C_r : 40.2$, SSIM - $Y' : 0.945, C_b : 0.927, C_r : 0.941$



(c) DT-CWT with joint regularization, ISNR (dB) - $Y' : 0.71, C_b : 0.51, C_r : 0.54$, PSNR (dB) - $Y' : 33.7, C_b : 39.8, C_r : 40.3$, SSIM - $Y' : 0.945, C_b : 0.929, C_r : 0.942$

(d) C_b-channel of standard JPEG, quality=50

(e) C_b-channel without joint regularization

(f) C_b-channel with joint regularization

Fig. 6. Comparison of the proposed approach (DT-CWT prior) with (c, f) and without (b, e) joint regularization of channels shown on RGB (top) and C_b (bottom) channels. **Best viewed electronically.**

practice results almost as good as much more complicated models. Fig. 5 and Tab. II shows this trade-off for our case of JPEG decompression. While the state-of-the-art method [20] gives the best SNR and preserves more details than the rest, we can see that visually both the QCS with TGV [5] and the proposed algorithm give results that are hard to distinguish from [20] at a fraction of time. Similarly, it is hard to tell visually whether the more time consuming data specific tight frame [39] is better than TGV or DT-CWT, although it wins in numbers.

Behavior of reconstruction algorithms can be strikingly different depending on the chosen stopping criterion. This is shown in Fig. 2, where we compared the results for convergence stopped in the iteration with the best SNR, early stopping rule (for [3] it is based on the value of dual gap, our algorithm was simply stopped after 10 iterations) and full convergence. As a rule, full convergence makes the result inferior both visually and in numbers. We can see that the DT-CWT prior is more natural than TV in the sense that the result has about the same character after a small number of iterations as for full convergence. For [3] and our algorithm, the best SNR is achieved around 5 iterations and visually best results in 10-15 iterations.

The regularization parameter τ can be set in various ways. It can be found experimentally and fixed for each quality level or can be estimated as the value with maximum likelihood (see end of Sec. II). The ML estimate can be obtained either from an image database or simply from the input image. Fig. 1 compares these two options with the estimate from ground truth. We can see the ISNR does not depend much on which option we choose. The best results are achieved for a small number of iterations and this number is important mainly for high compression ratios. In our experiments, we estimate τ from the image database.

The last experiment in this section compared the joint regularization described in Sec. V with reconstruction of each channel independently (Fig. 6). Although the differences in color images appear mostly invisible because of the reduced sensitivity to colors of human eye, they are present on small color objects as e.g. the traffic sign and colors are in general faded out due to the smoothing in chrominance channels. This is further supported by the images of the C_b channel, where e.g. the tower was reconstructed based on the information from luminance channel. Interestingly, joint regularization does not affect the PSNR and SSIM values of the luminance channel.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2627802, IEEE Transactions on Image Processing

(a) Standard JPEG decompression, SNR (dB) - Y' : $19.45, C_b : 5.69, C_r : 5.77$, PSNR (dB) - Y' : $31.4, C_b : 39.3, C_r : 39.6$, SSIM - Y' : $0.853, C_b : 0.920, C_r : 0.931$

(b) State-of-the-art JPEG restoration [20], ISNR (dB) - Y' : $3.04, C_b : 0.24, C_r : 0.41$, PSNR (dB) - Y' : $34.4, C_b : 39.5, C_r : 40.0$, SSIM - Y' : $0.943, C_b : 0.925, C_r : 0.937$

(c) Proposed combined JPEG + denoising, ISNR (dB) - Y' : $2.43, C_b : 0.70, C_r : 1.03$, PSNR (dB) - Y' : $33.8, C_b : 40.0, C_r : 40.6$, SSIM - Y' : $0.941, C_b : 0.930, C_r : 0.944$

Fig. 7. Comparison of the proposed combined JPEG-denoising approach with DT-CWT priors and 10 iterations (c) with standard JPEG decompression (a) and state-of-the-art JPEG decompression method [20] (b). Noise standard deviation was 10, quality factor 80. **Best viewed electronically.**

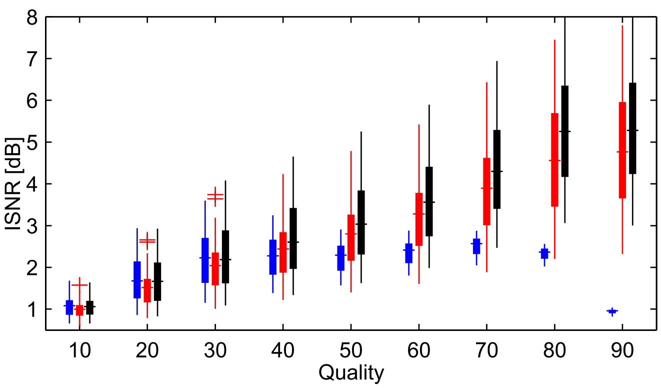


Fig. 8. Statistical comparison (maximum achievable ISNR on a set of 50 images) of combined JPEG-denoising approach with JPEG decompression alone and standard l_2-denoising, all of them with the same DT-CWT priors. Original images were distorted by additive Gaussian noise with $\sigma_g = 10$ and compressed at nine different quality levels. The order of the methods in each group representing the same quality is from left to right: Proposed JPEG decompression without considering noise, l_2-denoising, and proposed combined JPEG-denoising approach.

B. Combined JPEG Decompression and Denoising

In Sec. VI we derived a simple extension of the proposed JPEG decompression algorithm to combination with image denoising. Given that image noise is partially suppressed by JPEG compression, the question arises, whether such model is necessary. Fig. 7 shows an example, where the answer is affirmative. The image was degraded by Gaussian noise ($\sigma =$ 10 out of 256 levels) and compressed with quality factor 80. Standard JPEG decompression contains visible artifacts, which are not removed even by the state-of-the-art algorithm [20] set manually to achieve the best possible result. Our approach, using known noise standard deviation, removes the artifacts satisfactorily. Fig. 8 demonstrates benefits of the combined approach in comparison with JPEG decompression alone and standard MAP-based l_2-denoising, all of them with the same DT-CWT priors. Results are intuitive. If image noise is much stronger than quantization, which in our example happens for quality factors higher than 60, standard l_2-denoising works well, even though slightly worse than the combined approach. On the other hand, if image noise is much smaller than quantization noise, it is basically removed by JPEG compression and it is sufficient to use JPEG decompression without denoising modification. Benefits of combined approach are mostly visible in situations where image noise is of about the same strength as quantization.

VIII. Conclusion

In this paper, we derived a fast solution of the problem of JPEG decompression based on the MAP formulation with sparse priors by ADMM. The main contribution is the observation that using a Gaussian approximation of the quantization noise and a tight frame in the sparse prior based on l_1-norm allows for fast computation of the inverse critical for the ADMM method. Derived formulas allow to solve the formulation with the Gaussian approximation of the QCS by other proximal techniques to further speed up convergence [30], [42]–[45].

We showed that the Gaussian approximation of the QCS gives a better SNR than using the QCS. This counter-intuitive fact probably results from the partial inadequacy of sparse priors preferring smooth functions in our situation, where high frequencies are damaged by JPEG compression. Gaussian approximation favors solutions closer to original JPEG decompression, which prevents the algorithm to make result too smooth. As an example of tight frame priors we used the dual-tree complex wavelets and learned frames [39] and demonstrated that if they are coupled with the Gaussian approximation of the QCS, they are superior to the TV-based [3] and TGV-based [4], [5] methods in terms of SNR. On the other hand, differences are not large and even the simple TV prior can give very good reconstructions if stopped early enough. We also investigated the trade-off between the time and quality of reconstruction compared to the state-of-the-art method [20]. While [20] restores slightly more details, results are sometimes hard to distinguish from two orders of magnitude faster methods such as the method proposed in this paper. Also using more time-consuming priors such as [39] is justified only when computational aspects are less important.

Finally, we showed that thanks to the approximation of the likelihood by a Gaussian distribution, the proposed decompression algorithm can be naturally extended to solve simultaneously the denoising problem in basically the same time as the original algorithm. Benefits of combined approach are mostly visible for noise of about the same strength as quantization.

A simplified version of the proposed algorithm was used in the image forensics tool [51]. Matlab code of several variants of the algorithm described in this paper is available at http://zoi.utia.cas.cz/jpegrestoration.

REFERENCES

[1] "JPEG file interchange format (JFIF)," Ecma Int., Geneva, Switzerland, Tech. Rep. ECMA TR/98, 2009.

[2] Y. Yang, N. P. Galatsanos, and A. K. Katsaggelos, "Projection-based spatially adaptive reconstruction of block-transform compressed images," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 896–908, Jul. 1995.

[3] K. Bredies and M. Holler, "A total variation–based jpeg decompression model," *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 366–393, 2012.

[4] K. Bredies and M. Holler, "A TGV-based framework for variational image decompression, zooming, and reconstruction. Part I: Analytics," *SIAM J. Imag. Sci.*, vol. 8, no. 4, pp. 2814–2850, 2015.

[5] K. Bredies and M. Holler, "A TGV-based framework for variational image decompression, zooming, and reconstruction. Part II: Numerics," *SIAM J. Imag. Sci.*, vol. 8, no. 4, pp. 2851–2886, 2015.

[6] M. Sorel and M. Bartos, "Efficient JPEG decompression by the alternating direction method of multipliers," in *Proc. Int. Conf. Pattern Recognit.*, Dec. 2016.

[7] M. A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 27–38, Jan. 2005.

[8] R. L. Stevenson, "Reduction of coding artifacts in transform image coding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5. Apr. 1993, pp. 401–404.

[9] D. Sun and W.-K. Cham, "Postprocessing of low bit-rate block DCT coded images based on a fields of experts prior," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2743–2751, Nov. 2007.

[10] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. San Diego, CA, USA: Academic, 2008.

[11] F. Alter, S. Durand, and J. Froment, "Adapted total variation for artifact free decompression of JPEG images," *J. Math. Imag. Vis.*, vol. 23, no. 2, pp. 199–211, 2005.

[12] Z. Xiong, M. T. Orchard, and Y.-Q. Zhang, "A deblocking algorithm for JPEG compressed images using overcomplete wavelet representations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 433–437, Apr. 1997.

[13] C. Wang, J. Zhou, and S. Liu, "Adaptive non-local means filter for image deblocking," *Signal Process., Image Commun.*, vol. 28, no. 5, pp. 522–530, 2013.

[14] C. Jung, L. Jiao, H. Qi, and T. Sun, "Image deblocking via sparse representation," *Signal Process., Image Commun.*, vol. 27, no. 6, pp. 663–677, 2012.

[15] H. Chang, M. K. Ng, and T. Zeng, "Reducing artifacts in JPEG decompression via a learned dictionary," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 718–728, Feb. 2014.

[16] S. C. Park, M. G. Kang, C. A. Segall, and A. K. Katsaggelos, "Spatially adaptive high-resolution image reconstruction of DCT-based compressed images," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 573–585, Apr. 2004.

[17] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, "Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4613–4626, Dec. 2013.

[18] J. Zhang, S. Ma, Y. Zhang, and W. Gao, "Image deblocking using group-based sparse representation and quantization constraint prior," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 306–310.

[19] Y. Kwon, K. I. Kim, J. Tompkin, J. H. Kim, and C. Theobalt, "Efficient learning of image super-resolution and compression artifact removal with semi-local Gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1792–1805, Sep. 2015.

[20] X. Liu, X. Wu, J. Zhou, and D. Zhao, "Data-driven soft decoding of compressed images in dual transform-pixel domain," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1649–1659, Apr. 2016.

[21] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 576–584.

[22] Z. Wang, S. Chang, D. Liu, Q. Ling, and T. S. Huang, "D3: Deep dualdomain based fast restoration of jpeg-compressed images," in *Proc. IEEE CVPR*, Jun. 2016, pp. 2764–2772.

[23] A. W. C. Liew, H. Yan, and N.-F. Law, "POCS-based blocking artifacts suppression using a smoothness constraint set with explicit region modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 795–800, Jun. 2005.

[24] T. Kartalov, Z. A. Ivanovski, L. Panovski, and L. J. Karam, "An adaptive POCS algorithm for compression artifacts removal," in *Proc. 9th Int. Symp. Signal Process. Appl. (ISSPA)*, Feb. 2007, pp. 1–4.

[25] J.-F. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," *Multiscale Model. Simul.*, vol. 8, no. 2, pp. 337–369, 2010.

[26] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

[27] M. V. Afonso, J.-M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[28] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[29] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1, pp. 293–318, Jun. 1992.

[30] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[31] M. Šorel and F. Šroubek, "Fast convolutional sparse coding using matrix inversion lemma," *Digit. Signal Process.*, vol. 55, pp. 44–51, Aug. 2016.

[32] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2774–2781.

[33] P. G. Casazza, G. Kutyniok, and F. Philipp, "Introduction to finite frame theory," in *Finite Frames*. Boston, MA, USA: Springer, 2013, pp. 1–53.

[34] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Appl. Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, May 2001.

[35] E. J. Candes, "Ridgelets: Estimating with ridge functions," *Ann. Statist.*, vol. 31, no. 5, pp. 1561–1599, 2003.

[36] E. J. Candes and D. L. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," *Ann. Statist.*, vol. 30, no. 3, pp. 784–842, 2002.

[37] E. L. Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 423–438, Apr. 2005.

[38] P. Kittipoom, G. Kutyniok, and W.-Q. Lim, "Construction of compactly supported shearlet frames," *Constructive Approx.*, vol. 35, no. 1, pp. 21–72, 2012.

[39] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, "Data-driven tight frame construction and image denoising," *Appl. Comput. Harmon. Anal.*, vol. 37, no. 1, pp. 89–105, 2014.

[40] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 37–48, May 2003.

[41] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*, 1st ed. Norwell, MA, USA: Academic, 1992.

[42] N. Komodakis and J. C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, Nov. 2015.

[43] K. Bredies and H. Sun, "Preconditioned alternating direction method of multipliers for the minimization of quadratic plus non-smooth convex functionals," SpezialForschungBereich F 32, Graz, Austria, Tech. Rep. 006-2015, 2015.

[44] Y. Ouyang, Y. Chen, G. Lan, and E. Pasiliao, Jr., "An accelerated linearized alternating direction method of multipliers," *SIAM J. Imag. Sci.*, vol. 8, no. 1, pp. 644–681, 2015.

[45] K. Bredies and H. Sun. (Apr. 2016). "Accelerated douglas-rachford methods for the solution of convex-concave saddle-point problems." [Online]. Available: https://arxiv.org/abs/1604.06282

[46] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, p. 947, 2007.

[47] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[48] S. D. Zenzo, "A note on the gradient of a multi-image," *Comput. Vis., Graph., Image Process.*, vol. 33, no. 1, pp. 116–125, 1986.

[49] T. F. Chan, S. H. Kang, and J. Shen, "Total variation denoising and enhancement of color images based on the CB and HSV color models," *J. Vis. Commun. Image Represent.*, vol. 12, no. 4, pp. 422–435, 2001.

[50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[51] J. Kamenicky *et al.*, "PIZZARO: Forensic analysis and restoration of image and video data," *Forensic Sci. Int.*, vol. 264, pp. 153–166, Jul. 2016.

Michal Šorel (M'10) received the M.Sc. and Ph.D. degrees in computer science from Charles University, Prague, Czech Republic, in 1999 and 2007, respectively. From 2012 to 2013, he was with Heriot-Watt University in Edinburgh, Scotland, and the University of Bern, Switzerland. He is currently a Research Fellow with the Institute of Information Theory and Automation, Czech Academy of Sciences, Prague. His research interests lie on the interface of computational mathematics, machine learning, and image processing. His research work currently concentrates on development of efficient image restoration algorithms.

Michal Bartoš received the M.Sc. degree (Hons.) in biomedical and ecological engineering and the Ph.D. degree in biomedical electronics and biocybernetics from the Brno University of Technology, Czech Republic, in 2007 and 2015, respectively. He is currently a Post-Doctoral Researcher with the Institute of Information Theory and Automation, Czech Academy of Sciences, Prague. His research interests are in image and signal processing mostly in biomedical applications, including perfusion estimation using magnetic resonance imaging, ultrasound imaging, and image registration.