Image Bit-depth Enhancement via Maximum-A-Posteriori Estimation of AC Signal

Pengfei Wan, Gene Cheung, *Senior Member, IEEE,* Dinei Florencio, *Fellow, IEEE,* Cha Zhang, *Senior Member, IEEE,* and Oscar C. Au, *Fellow, IEEE*

***Abstract*—When images at low bit-depth are rendered at high bit-depth displays, missing least significant bits need to be estimated. We study the image bit-depth enhancement problem: estimating an original image from its quantized version from a minimum mean squared error (MMSE) perspective. We first argue that a graph-signal smoothness prior—one defined on a graph embedding the image structure—is an appropriate prior for the bit-depth enhancement problem. We next show that solving for the MMSE solution directly is in general too computationally expensive to be practical. We then propose an efficient approximation strategy. Specifically, we first estimate the AC component of the desired signal in a maximum a posteriori (MAP) formulation, efficiently computed via convex programming. We then compute the DC component with an MMSE criterion in closed form given the computed AC component. Experiments show that our proposed two-step approach has improved performance over conventional bit-depth enhancement schemes in both objective and subjective comparisons.**

***Index Terms*—Bit-depth enhancement, graph signal processing**

I. Introduction

It is undeniable that there exists an insatiable human desire to create bigger and more realistic visual displays. In terms of spatial resolution (number of pixels per image), television has evolved from VGA (640×480) to HD (1280×720), and soon to 4K and 8K ultra HD (3840×2160 and 7680×4320 respectively). In terms of bit-depth (number of bits per pixel), *high dynamic range* (HDR) technologies [1] have promised 10 or more bits per pixel—as opposed to conventional 8 bits per pixel—for finer-grained quantization of real pixel values to discrete levels.

However, though display technologies have continued to improve, the bulk of legacy image and video contents were captured and recorded using older imaging devices, often in lower spatial resolution and shallower bit-depth than what modern displays are capable of rendering. Super-resolution [2] addresses the first problem of increasing the spatial resolution of an image. In contrast, we address the second problem of bit-depth enhancement. Bit-depth enhancement is equivalent to the problem of signal reconstruction from its per-pixel quantized version. Thus bit-depth enhancement can also be called the *pixel domain de-quantization* problem.

P. Wan is with Meitu, Inc. Email: wpf@meitu.com.

G. Cheung is with National Institute of Informatics, 2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan. Email: cheung@nii.ac.jp.

D. Florencio and C. Zhang are with Microsoft Research, One Microsoft Way, Redmond, WA 98052 USA. Email: {dinei, chazhang}@microsoft.com.

This work is partially supported by Microsoft Research CORE program.

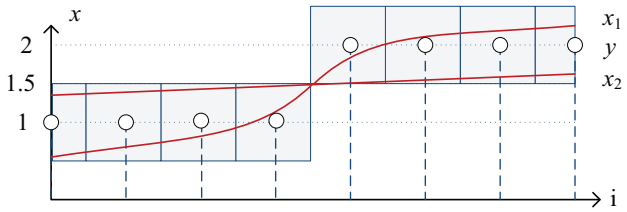


Fig. 1: Examples of quantized and reconstructed 1D signals.

One major visual artifact caused by shallow bit-depth is false contouring, also called posterization artifact. Informally it is the unnatural appearance of discrete color bands due to coarse quantization into finite bits, while a high bit-depth equivalent image patch would appear as smooth transition from one color to another. Thus there is a need to suitably increase the bit-depth of legacy image content.

Existing techniques for bit-depth enhancement in the literature [3–7] exploit the notion of *smoothness*. The key observation is that true image signals tend to be smooth, and thus given an observed *low bit-depth* (LBD) signal, applying a smoothing operator across consecutive quantization levels would likely result in a better quality reconstructed signal. As an illustration, we see in Fig. 1 an 8-sample one-dimensional (1D) signal $\mathbf{y}$ quantized to integer values 1 and 2. If a smoothing operator is applied, it can result in $\mathbf{x}_1$ or $\mathbf{x}_2$, depending on the amount and type of smoothing applied. While the notion of smoothness is intuitive, defining an appropriate mathematical definition and applying it optimally for *high bit-depth* (HBD) signal reconstruction is nontrivial.

In this paper, leveraging on recent *graph signal processing* (GSP) techniques for images [8–18], we first formally define "smoothness" via a signal-dependent graph Laplacian. Specifically, a *graph Fourier transform* (GFT) computed from a defined graph can decompose a graph-signal (*e.g.*, a pixel patch) into graph frequency components, and a graph-signal is deemed smooth if it contains mainly low graph frequencies. Further, unlike spectral decomposition based on fixed transforms like discrete cosine transform (DCT), image gradients can be embedded as edge weights into the graph, so that discontinuities in natural images will nonetheless be interpreted as low graph frequencies, reducing the chance of over-smoothing.

Next, armed with our defined smoothness prior for graph-signals, we formulate an optimization problem for reconstructed signal $\hat{\mathbf{x}}$ that minimizes the average mean squared error (MSE) (or, equivalently, the expected distortion) given

quantized signal $\mathbf{y}$. Observing that the optimization is difficult to solve directly, we next demonstrate that the most probable signal via *maximum a posteriori* (MAP) estimation, given observed $\mathbf{y}$ and the smoothness prior, can lead to large expected distortion, and thus applying MAP directly is a poor proxy for the original optimization objective. However, we argue that with a simple twist—computing the most probable AC component $\hat{\mathbf{x}}_A$ of the reconstructed signal first via MAP and then a distortion-minimizing DC component $\hat{x}_D$ subsequently—MAP can still be a useful and efficient tool that returns good approximate solutions to the original MMSE problem. Experiments demonstrate that our proposed two-step method produces HBD images that are better than competing schemes in both objective and subjective comparisons.

The outline of the paper is as follows. We first overview related work in Section II. We then formulate and analyze the bit-depth enhancement problem in Section III. Proposed image bit-depth enhancement and the corresponding optimization procedure are elaborated in Section IV and Section V respectively. Finally, we discuss experimental results and draw conclusions in Section VI and VII, respectively.

II. Related Work

Bit-depth enhancement is different from the inverse tone mapping problem in HDR imaging [19, 20]. Distortions in inverse tone mapping are typically caused by unknown non-linear tone mapping operator or the over-saturation of camera sensors, while distortion in bit-precision enhancement problem is introduced by scalar quantization. Thus the desired output of bit-precision enhancement does not hallucinate lost details for improved subjective quality, but estimates the original HBD image from the coarsely quantized LBD image.

Coarse quantization may cause false contours which degrade the visual quality of the image. Besides naïve algorithms such as zero-padding and bit-replication [21], previous works on false contour removal and bit-precision enhancement [4, 5, 7, 22] are typically smoothing schemes using filtering or spatial interpolation, which do not optimize an objective metric such as mean squared error (MSE). [23] presented a bit-depth enhancement method by modeling the error of intra-pixel prediction, but its objective is to minimize the classification risk, rather than to minimize the MSE relative to the ground-truth. Also note that in our problem setup the input image has already been quantized, so dithering-like approaches [24–26] that change the signal before quantization are not applicable.

An example of filtering-based image bit-depth enhancement method is predictive cancellation [7]. The basic idea is to predict the quantization error of the HBD image given the LBD image. The LBD image is first low-pass filtered, and the quantization error of the low-passed image approximates that of the input LBD image in low frequencies. Thus by subtracting the approximated quantization error, the contours in low frequency image regions can be removed. The problem of this method is that the low-pass filter might blur non-contour edges, and it is hard for the filter kernel to adapt to false contours in different sizes.

Examples of interpolation-based bit-depth enhancement method are [5, 27]. Two distance maps, up map and down map, are generated to measure the contour region progression. A variable called step ratio is calculated from the distance maps showing the percentage of progression of the contour region. Finally, the estimated HBD signal is calculated by linear interpolation based on the step ratio. With the help of detected skeletons, [5] obtained good bit-depth enhancement results even for textureless regions that are local maximum and minimum. But the simple linear interpolation does not guarantee good reconstruction performances as natural images are 2D signals that typically have signal-dependent structures and non-linear transitions.

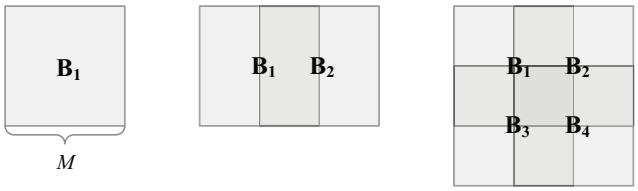


Fig. 2: Illustration of block arrangement. We use $M \times M$ half-overlapped blocks to reduce artefacts due to block-based processing. From left to right: one block, two overlapped blocks, four overlapped blocks. Note that the boundary pixels of a block are the inner pixels of other blocks.

Graph signal processing (GSP) is the study of signals that live on structured data kernels described by graphs [8]. GSP tools can also be applied to traditional signals such as images that live on regular 2D grids [9–18, 28] or point cloud structures [29], where the idea is to embed signal gradients into the graph before signal processing. Similarly, to reconstruct a HBD signal we first compute edge weights of a graph based on the signal gradients (deduced from the observed input LBD image) and then define signal smoothness via the graph Laplacian, so that a signal with enhanced bit-precision can be reconstructed without over-smoothing natural image gradients.

III. Problem Formulation

We first overview the image bit-depth enhancement problem. We then define the problem objective in terms of MMSE. We need an appropriate signal prior to regularize our inverse problem; we argue that the graph-signal smoothness prior is suitable for the bit-depth enhancement problem. We conclude this section by showing that solving the formulated MMSE problem directly is difficult—numerical evaluation is computationally complex, while solving MAP as a proxy can lead to bad MSE performance.

A. Block-based Bit-depth Enhancement

Image bit-depth enhancement problem is an inverse imaging problem. We interpret an image signal $\mathbf{x}^o$ to be an observation of an N-dimensional random vector $\mathbf{x}$, whose *probability density function* (PDF) is described mathematically by a signal prior (to be formally defined). Without loss of generality, we assume that the signal $\mathbf{x}^o$ is normalized such that each pixel i has value $x_i^o \in [0, 1]$. To represent each pixel value in finite bits, each x_i^o is separately scalar-quantized to value $y_i = (\text{floor}(x_i^o/Q) + 0.5)\, Q$ [30] in vector $\mathbf{y}$ using a known

scalar quantization step size $Q = 1/2^b$ with bit-depth b.[1] At the receiver, only $\mathbf{y}$ and Q are known, which translate to a *per-pixel quantization constraint* for the original signal: x_i^o must lie within the *quantization bin* $\left[y_i - \frac{Q}{2}, y_i + \frac{Q}{2}\right)$, $\forall i$. Given the set of N quantization constraints and a pre-defined signal prior, the problem is to find the "best" estimate $\hat{\mathbf{x}}$ of the original image signal $\mathbf{x}^o$.

For complexity considerations, we divide an image into overlapping pixel blocks for block-based processing. The estimated HBD pixel values are updated through processing of blocks in a raster-scan order. Specifically, we use $M \times M$ square blocks, each overlapping adjacent blocks on the same row or column by $\frac{M}{2}$ pixels (M is an even integer); see Fig. 2 for an illustration. Thus by image signal, we mean a vectorized image block with length $N = M^2$.

B. MMSE Objective for Bit-depth Enhancement

Our objective is to find an estimate $\hat{\mathbf{x}}^{\text{MMSE}}$ with the minimum mean squared error (MMSE):

$$\hat{\mathbf{x}}^{\text{MMSE}} = \arg\min_{\hat{\mathbf{x}}} \int \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \; f(\mathbf{x} \,|\, \mathbf{y}) \; d\mathbf{x} \tag{1}$$

where $f(\mathbf{x} \,|\, \mathbf{y})$ is the posterior PDF of original signal $\mathbf{x}$ given observation $\mathbf{y}$. To derive $\hat{\mathbf{x}}^{\text{MMSE}}$, we set the derivative of (1) with respect to $\hat{\mathbf{x}}$ to zero, leading to:

$$\int (\hat{\mathbf{x}} - \mathbf{x}) \; f(\mathbf{x} \,|\, \mathbf{y}) \; d\mathbf{x} = 0 \tag{2}$$

which implies $\hat{\mathbf{x}} \int f(\mathbf{x} \,|\, \mathbf{y}) \; d\mathbf{x} = \int \mathbf{x} f(\mathbf{x} \,|\, \mathbf{y}) \; d\mathbf{x}$. By definition $\int \; f(\mathbf{x} \,|\, \mathbf{y}) \; d\mathbf{x} = 1$, so the MMSE solution $\hat{\mathbf{x}}^{\text{MMSE}}$ is also the expectation of $\mathbf{x}$ given $\mathbf{y}$:

$$\hat{\mathbf{x}}^{\text{MMSE}} = \int \mathbf{x} f(\mathbf{x} \,|\, \mathbf{y}) \; d\mathbf{x} = \mathbf{E}(\mathbf{x}|\mathbf{y}) \tag{3}$$

To compute $\hat{\mathbf{x}}^{\text{MMSE}}$ in (3) we need to compute the posterior $f(\mathbf{x} \,|\, \mathbf{y})$. By Bayes' theorem [31], we know that $f(\mathbf{x} \,|\, \mathbf{y}) \propto f(\mathbf{x}) f(\mathbf{y} \,|\, \mathbf{x})$, where $f(\mathbf{x})$ and $f(\mathbf{y} \,|\, \mathbf{x})$ are the prior and likelihood respectively. In our bit-depth enhancement problem, the likelihood takes a simple form due to the nature of quantization:

$$f(\mathbf{y}|\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \mathbb{F}(\mathbf{y}) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $\mathbb{F}(\mathbf{y})$ is the feasible space of the original signal given observed quantized signal $\mathbf{y}$, *i.e.*,

$$\mathbb{F}(\mathbf{y}) = \{\mathbf{x} \mid \; (\text{floor}(x_i/Q) + 0.5)\, Q = y_i, \; \forall i\} \tag{5}$$

The prior $f(\mathbf{x})$ reflects the statistics of the original signal. Like other inverse problems, the prior serves to regularize an otherwise ill-posed problem [32]. Next, we discuss the prior we employ in this work for bit-depth enhancement—the graph-signal smoothness prior.

[1]Another possible quantizer is $y_i = \text{round}(x_i^o/Q)\, Q$ where $Q = \frac{1}{2^b - 1}$.

C. Smoothness Prior for Image Signal

Smoothness is a widely used signal prior for inverse imaging problems such as denoising [32–34]. It assumes that the desired signal is slow-varying over a defined spatial domain where the signal exists [35]. Smoothness can be mathematically defined in a number of ways. For example, *total variation* (TV)—which assumes that the aggregate inter-pixel difference in l_1-norm in an image is small—is a well-known smoothness prior in image restoration [36]. One can also define smoothness as having mainly low-frequency energies in a chosen transform domain, such as discrete Fourier or wavelet transform [37, 38].

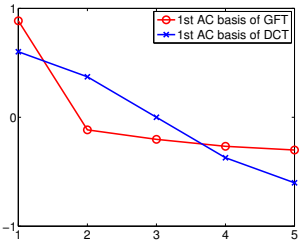
The appropriate choice of smoothness definition depends on the specific inverse problem. For many image restoration problems, smoothness defined as low-frequencies in fixed Fourier transforms such as Discrete Cosine Transform (DCT) may not be a good choice. Defined independent of the signal it sought to represent, DCT essentially assumes that every pair of adjacent pixels are equally similar. Thus, edges / discontinuities common in a natural image would translate to high frequencies in the Fourier domain, and enforcing smoothness will blur sharp edges.

Thus a good smoothness prior should account for the underlying *image structures* (discontinuities), so that over-smoothing of sharp edges in images does not occur. We argue that a *smoothness prior defined on a weighed graph* is suitable for our image bit-depth enhancement problem. This is because detected image structures (discontinuities) can be embedded as small edge weights in the graph; small edge weights prevent filtering across edges during graph-based filtering, which would otherwise lead to blurring [10]. In literature, graph-signal smoothness prior has been successfully used in other inverse problems such as depth image denoising [9], depth image interpolation [13, 14], natural image denoising [15, 16] and soft decoding of JPEG images [12, 18].

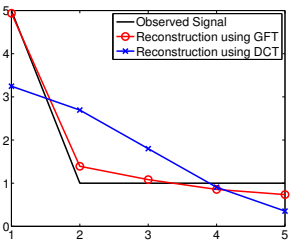
We first formally define a graph and graph-signals on top of the graph. We then define our notion of graph-signal smoothness prior for bit-depth enhancement.

1) Signal on Graphs: We consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{W})$ where $\mathcal{V}$ and $\mathcal{W}$ are the sets of vertices and edge weights, respectively. The graph is defined on a pixel block, where: 1) each vertex represents a pixel i with associated pixel value x_i; and 2) two vertices—representing adjacent pixels i and j in the block—are connected by an undirected edge, which in turn is labeled with a non-negative weight $w_{i,j} \in [0, 1]$. Edge weight reflects the correlation of—or similarity between—two pixels: 1 means highest correlation / similarity and 0 effectively means no correlation / similarity.

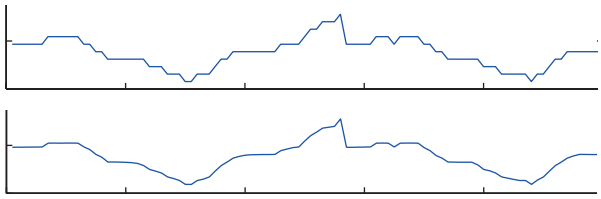
A graph-signal is defined as a vector $\mathbf{x} \in \mathbb{R}^N$ on a graph $\mathcal{G}$ with fixed $\mathcal{V}$ and $\mathcal{W}$. Hence the edge weights have to be determined prior to the processing of the graph-signal $\mathbf{x}$. Given only observed signal $\mathbf{y}$, typically an edge weight $w_{i,j}$ is assigned according to the similarity between the two corresponding observed pixels y_i and y_j, so that observable image structure can be embedded into the graph definition. Like the bilateral filter [39] where a Gaussian kernel is used to determine inter-pixel weights, one can use a Gaussian kernel



(a) The bases of the 1st AC component. In GFT, weight $w_{1,2} = 0.1$.



(b) Signal reconstruction using only DC and 1st AC component.



(c) Example of enforcing graph smoothness. Upper figure is an input row of pixel values, lower figure is the smoothed output.

Fig. 3: Advantage of graph smoothness over smoothness defined by DCT: enforcing graph smoothness leads to a better reconstructed signal. For (a) (b), the observed $\mathbf{y} = [5, 1, 1, 1, 1]$. Signal samples are connected to its immediate neighbor, with edge weights being 1 except for $w_{1,2} = w_{2,1} = 0.1$.

to compute $w_{i,j}$ [9, 15], *i.e.* $w_{i,j} = \exp\left(-(y_i - y_j)^2/\sigma_w^2\right)$, where σ_w is selected so that $w_{i,j}$ is close to 0 when y_i and y_j are on two sides of a discontinuity, and close to 1 when they are in a smooth region.

With the calculated edge weights, we define the *adjacency matrix* $\mathbf{W} \in \mathbb{R}^{N\times N}$ of the graph, where its (i, j)-entry is the weight $w_{i,j}$. The *degree matrix* $\mathbf{D} \in \mathbb{R}^{N\times N}$ is a diagonal matrix with entries $d_{i,i} = \sum_j w_{i,j}$. The *graph Laplacian matrix* is further defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$. Eigenvectors of $\mathbf{L}$ compose the rows of the transform matrix of the *Graph Fourier Transform* (GFT). GFT decomposes a graph-signal into its graph frequency components, where low frequency components correspond to the eigenvectors with smaller eigenvalues. Note the smallest eigenvalue is zero by definition of $\mathbf{L}$, which corresponds to the DC component.

2) Graph-Signal Smoothness Prior: A smooth graph-signal means that its low-frequency energy is dominant in the GFT domain. Thus we define a graph-signal smoothness prior $f(\mathbf{x})$ where a signal $\mathbf{x}$ is more probable if $\mathbf{x}$ contains less high-frequency energy in GFT domain. Since changing the DC component does not affect the smoothness of a signal, we write $f(\mathbf{x}) = f(x_D)f(\mathbf{x}_A)$ where x_D and $\mathbf{x}_A$ are AC and DC components of $\mathbf{x}$ respectively:

$$\mathbf{x} = x_D\mathbf{1} + \mathbf{x}_A, \quad x_D = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{6}$$

Prior of the DC signal is assumed uniform, *i.e.* $f(x_D) = C$ is a constant scalar. The AC prior PDF $f(\mathbf{x}_A)$ favors low-frequency components in GFT domain:

$$f(\mathbf{x}_A) = \frac{1}{K}\exp\left\{-\sigma\mathbf{x}_A^\intercal\mathbf{L}\mathbf{x}_A\right\} = \frac{1}{K}\exp\left\{-\sigma\sum_{i=2}^{N}\rho_i c_i^2\right\} \tag{7}$$

where $[\rho_2, \cdots, \rho_N]$ are $\mathbf{L}$'s non-zero eigenvalues in non-decreasing order. c_i is the GFT coefficient of the i-th graph frequency. K is the normalization factor for $f(\mathbf{x}_A)$[2].

By (7), an AC signal $\mathbf{x}_A$ with energy mostly in the low graph frequencies (non-zero coefficients c_i's only for small ρ_i's) will have higher probability $f(\mathbf{x}_A)$. Further, because $\mathbf{x}_A^\intercal\mathbf{L}\mathbf{x}_A = \frac{1}{2}\sum_{i,j} w_{i,j}(x_{A,i} - x_{A,j})^2$, a large inter-pixel difference $|x_{A,i} - x_{A,j}|$ will not incur high frequency energy if edge weight $w_{i,j}$ is pre-assigned a small value. Hence reconstructing a smooth graph-signal using prior in (7) will not blur sharp edges *if* the image structure is described by appropriately set edge weights. Fig. 3(b) shows an example where a smooth graph-signal containing only DC and 1st graph AC component better represents the original signal (with discontinuity between the first and second samples) compared to that using DCT basis. Fig. 3(c) shows an example of graph smoothing on a row of pixels, where we see that a smooth signal is restored without blurring the true edges.

D. Solving the MMSE Problem

Given Bayes' theorem $f(\mathbf{x}|\mathbf{y}) \propto f(\mathbf{x})f(\mathbf{y}|\mathbf{x})$, we multiply our graph-signal smoothness prior (7) and likelihood function (4) to obtain the posterior of $\mathbf{x}$ as follows:

$$f(\mathbf{x}|\mathbf{y}) \propto \begin{cases} f(\mathbf{x}), & \text{if } \mathbf{x} \in \mathbb{F}(\mathbf{y}) \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Given the posterior, we can in theory compute the MMSE solution using (3). However, directly calculating (3) is difficult: the prior (7) is a N-dimensional Gaussian, while the likelihood (4) is a binary-valued function defining a feasible signal space. Hence the resulting posterior (8) is a "cropped" N-dimensional Gaussian function. In that case, calculating (3), *mean* of the posterior, requires multi-dimensional integration. General high-dimensional integration is computationally expensive and is typically done using Monte Carlo methods [41, 42].

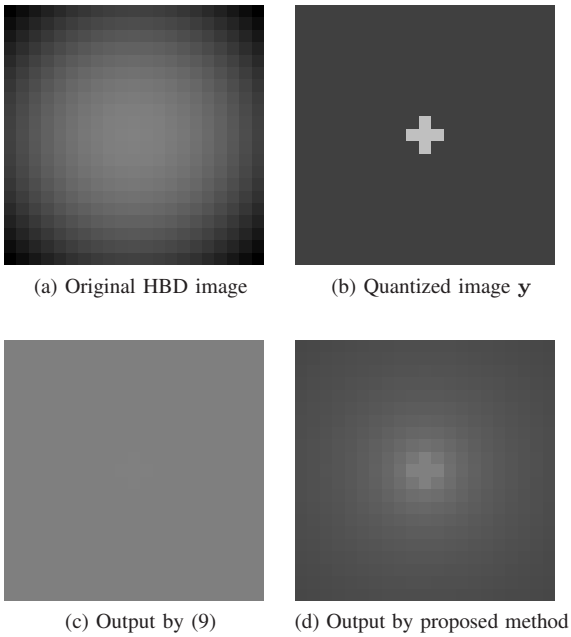
One alternative to directly solving the MMSE problem is to solve the MAP problem as an approximation; *e.g.*, total variation denoising [36] can be computed using a MAP estimator assuming the signal gradient follows a Laplace distribution. MAP estimation is easier to compute since maximization can be efficiently solved using convex optimization [43].

By definition, MAP solution is the *mode* of the posterior PDF:

$$\hat{\mathbf{x}}^{\text{MAP}} = \arg\max_{\mathbf{x}} f(\mathbf{x}|\mathbf{y}) \tag{9}$$

MAP solution $\hat{\mathbf{x}}^{\text{MAP}}$ is clearly different from MMSE solution $\hat{\mathbf{x}}^{\text{MMSE}}$: MMSE is the location of center mass of the posterior distribution, whereas the MAP is the location of the peak of the posterior distribution. There are cases where the MAP solution (mode) is far from the MMSE solution (mean).

[2]In some applications, the same smoothness prior can alternatively be interpreted statistically using Gaussian Markov random field [40].



(a) Original HBD image (b) Quantized image $\mathbf{y}$

(c) Output by (9) (d) Output by proposed method

Fig. 4: Example showing direct MAP estimation leads to a DC output (c) with large MSE, while proposed two-step approach outputs a smooth signal (d) with smaller MSE.

In particular, we show that in our bit-depth enhancement problem, using MAP directly as a proxy can lead to arbitrarily poor MSE performance.

By (8), MAP estimator (9) finds the signal $\mathbf{x}$ with the largest prior probability $f(\mathbf{x})$ in $\mathbb{F}(\mathbf{y})$: $\hat{\mathbf{x}}^{\text{MAP}} = \arg\max_{\mathbf{x}\in\mathbb{F}(\mathbf{y})} f(\mathbf{x})$. Given a smoothness prior $f(\mathbf{x})$ is used, the MAP estimator identifies the "smoothest" signal in $\mathbb{F}(\mathbf{y})$. In the example shown in Fig. 4, MAP identifies a solution in (c) that is essentially DC, regardless of quantization step size Q, with MSE $= Q^2/3$. In contrast, MSE of simply using $\mathbf{y}$ as the estimation is $Q^2/12$. For increasingly large Q, the MAP solution is increasingly worse than $\mathbf{y}$. Hence directly using MAP as a proxy to MMSE is not a good choice.

A natural question is then: what are the conditions for the MAP solution to have good MSE performance? Derived in Appendix A, the estimation error of MMSE solution and MAP solution are respectively:

$$\begin{aligned} &\text{MSE}(\hat{\mathbf{x}}^{\text{MMSE}}) = \mathbf{E}(\|\mathbf{x} - \mathbf{E}(\mathbf{x}|\mathbf{y})\|_2^2|\mathbf{y}), \\ &\text{MSE}(\hat{\mathbf{x}}^{\text{MAP}}) = \text{MSE}(\hat{\mathbf{x}}^{\text{MMSE}}) + G(\mathbf{x},\mathbf{y})^2 \end{aligned} \tag{10}$$

where $G(\mathbf{x},\mathbf{y}) = \|\hat{\mathbf{x}}^{\text{MAP}} - \hat{\mathbf{x}}^{\text{MMSE}}\|_2$ is the gap between MAP estimation and MMSE estimation. Small gap means good MSE performance. There are two sufficient conditions for a small gap $G(\mathbf{x},\mathbf{y})$: 1) the gap equals to zero if posterior distribution $f(\mathbf{x}|\mathbf{y})$ is symmetric around the peak (*e.g.* Gaussian distribution); 2) the gap is small when the posterior distribution is concentrated around the peak.

Above analysis explained the bad MSE performance of direct MAP estimation MSE($\hat{\mathbf{x}}^{\text{MAP}}$): 1) the gap is larger due to the asymmetry of the posterior distribution; 2) the variance of the posterior distribution $f(\mathbf{x}|\mathbf{y})$ is large. This is because the posterior $f(\mathbf{x}|\mathbf{y})$ in our problem is a "cropped" version of Gaussian function (7). Although the prior probability $f(\mathbf{x})$ is symmetric in space $\mathbb{R}^N$, it is not symmetric in the feasible subspace $\mathbb{F}(\mathbf{y}) \in \mathbb{R}^N$ defined by likelihood (4), thus the posterior distribution is asymmetric and the gap is large.

In summary, *MMSE solution is optimal in MSE but is hard to compute, while MAP solution is easy to compute but can have bad MSE performance*. We next present our two-step approach that achieves a good tradeoff between the two: we first estimate the AC signal using MAP which is shown to have good MMSE performance, followed by a MMSE estimation of the DC signal based on the estimated AC signal.

IV. Proposed ACDC Method

A. Approximation to the MMSE Solution

As discussed previously, directly computing the MMSE solution (1) is computationally expensive, while MAP does not yield good approximations in general. So instead of estimating the signal $\mathbf{x}$ directly, we propose to estimate the AC and DC signals separately: AC component via MAP, then DC component via MMSE given computed AC component. For notation convenience we call our proposed algorithm *ACDC*.

As in (6), we decompose the N-dimensional signal $\mathbf{x}$ into AC and DC components: $\mathbf{x} = x_D\mathbf{1} + \mathbf{x}_A$, where the DC component $x_D = \frac{1}{N}\sum_{i=1}^N x_i$ is a scalar, and the AC component $\mathbf{x}_A$ is a vector satisfying $\sum_{i=1}^N x_{A,i} = 0$.

By the above AC-DC decomposition, the MMSE problem (1) can be rewritten as:

$$\hat{\mathbf{x}}^{\text{MMSE}} = \arg\min_{\hat{x}_D,\hat{\mathbf{x}}_A} \int \|\hat{x}_D\mathbf{1} + \hat{\mathbf{x}}_A - x_D\mathbf{1} - \mathbf{x}_A\|_2^2 f(\mathbf{x}|\mathbf{y})d\mathbf{x} \tag{11}$$

Because $\|\hat{x}_D\mathbf{1} + \hat{\mathbf{x}}_A - x_D\mathbf{1} - \mathbf{x}_A\|_2^2 = N(\hat{x}_D - x_D)^2 + \|\hat{\mathbf{x}}_A - \mathbf{x}_A\|_2^2 + 2(\hat{x}_D - \hat{x}_D)\mathbf{1}^\intercal(\hat{\mathbf{x}}_A - \mathbf{x}_A)$ and $\mathbf{1}^\intercal(\hat{\mathbf{x}}_A - \mathbf{x}_A) \equiv 0$, equation (11) becomes

$$\arg\min_{\hat{x}_D,\hat{\mathbf{x}}_A} \int (N(\hat{x}_D - x_D)^2 + \|\hat{\mathbf{x}}_A - \mathbf{x}_A\|_2^2) f(\mathbf{x}|\mathbf{y})d\mathbf{x} \tag{12}$$

Problem (12) is still hard to solve with two variables $\hat{x}_D$ and $\hat{\mathbf{x}}_A$, so we first estimate the AC signal $\hat{\mathbf{x}}_A$, followed by DC estimation. The reason for this decomposition is because our graph-signal smoothness prior only characterizes the AC signal, so estimating AC signal first in general is more accurate.

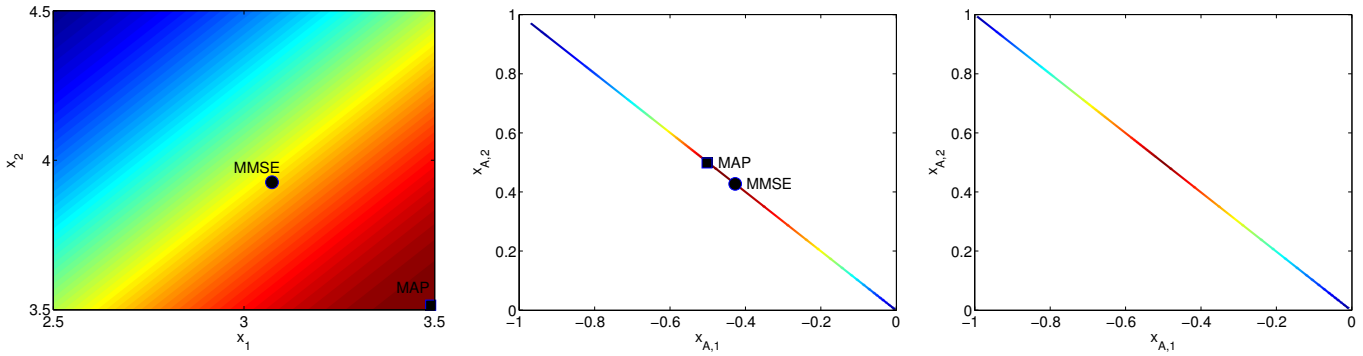
Mathematically, we first solve (12) with $\hat{\mathbf{x}}_A$ as the only variable:

$$\begin{aligned} &\arg\min_{\hat{\mathbf{x}}_A} \int (N(\hat{x}_D - x_D)^2 + \|\hat{\mathbf{x}}_A - \mathbf{x}_A\|_2^2) f(\mathbf{x}|\mathbf{y})d\mathbf{x} \\ &= \arg\min_{\hat{\mathbf{x}}_A} \int \|\hat{\mathbf{x}}_A - \mathbf{x}_A\|_2^2 \; f(\mathbf{x}_A|\mathbf{y}) \; d\mathbf{x}_A \end{aligned} \tag{13}$$

This is still a MMSE problem that cannot be solved directly. In the following Section IV-B, we will solve it approximately using MAP; the solution is denoted as $\hat{\mathbf{x}}_A^{\text{ACDC}}$.

Then we fix the AC signal as $\hat{\mathbf{x}}_A^{\text{ACDC}}$ and solve the MMSE DC given the feasibility constraint:

$$\begin{aligned} &\arg\min_{\hat{x}_D} \int (\hat{x}_D - x_D)^2 \; f(x_D + \hat{\mathbf{x}}_A^{\text{ACDC}}|\mathbf{y}) \; dx_D \\ &\text{s.t.} \;\; \hat{x}_D\mathbf{1} + \hat{\mathbf{x}}_A^{\text{ACDC}} \in \mathbb{F}(\mathbf{y}) \end{aligned} \tag{14}$$



(a) Posterior PDF $f(\mathbf{x}|\mathbf{y})$ in direct MAP (9). Likelihood function (4) defines a square feasible space where the MAP solution is at the corner.

(b) Posterior PDF of AC signal $f(\mathbf{x}_A|\mathbf{y})$ in MAP AC estimation (23). Note MAP solution is close to MMSE solution.

(c) Likelihood function of AC signal (20) in MAP AC estimation (23). Feasible space of AC signal (21) is a line segment.

Fig. 5: An example of estimating signal $\mathbf{x} = [x_1, x_2]$ ($N = 2$) given its quantized signal $\mathbf{y} = [3, 4]$ and $Q = 1$. Prior PDF is defined to be $f(\mathbf{x}) \propto \exp(-\sigma(x_1 - x_2)^2)$. The colormap stands for the probability density (red for higher value). We see that due to the non-uniform AC likelihood (c), the posterior PDF of AC signal (b) is more concentrated than that of the whole signal (a), thus MAP for AC signal has good MSE performance.

The solution of (14) is denoted as $\hat{x}_D^{\text{ACDC}}$, and will be discussed in Section IV-C. Next we elaborate how we calculate the MAP AC signal $\hat{\mathbf{x}}_A^{\text{ACDC}}$ and MMSE DC signal $\hat{x}_D^{\text{ACDC}}$ in turn.

B. AC Signal Estimation

We propose to estimate the AC signal $\hat{\mathbf{x}}_A$ using MAP estimation. We first derive the exact likelihood function for the AC signal, then we explain why MAP is a good proxy for the MMSE AC problem (13).

Seeking MAP estimate of AC signal means solving:

$$\max_{\mathbf{x}_A} f(\mathbf{x}_A|\mathbf{y}) = \max_{\mathbf{x}_A} f(\mathbf{y}|\mathbf{x}_A)\, f(\mathbf{x}_A) \tag{15}$$

where the prior for the AC signal is defined in (7). By total probability theorem, the likelihood can be written as:

$$f(\mathbf{y}|\mathbf{x}_A) = \int f(\mathbf{y}|\mathbf{x}_A, x_D) f(x_D)\, dx_D \tag{16}$$

where $f(\mathbf{y}|\mathbf{x}_A, x_D)$ is 1 if $x_D + \mathbf{x}_A \in \mathbb{F}(\mathbf{y})$ and 0 otherwise. Recall that $f(x_D) = C$. Hence the likelihood (16) becomes:

$$f(\mathbf{y}|\mathbf{x}_A) = \int_{x_D | x_D + \mathbf{x}_A \in \mathbb{F}(\mathbf{y})} C\, dx_D \tag{17}$$

1) Deriving Likelihood of AC Signal: Next we derive the upper and lower integral bounds for the above equation. By definition of the feasible space $\mathbb{F}(\mathbf{y})$, we have N per-pixel quantization constraints as follows:

$$\begin{aligned} x_{A,i} + D^{\text{up}} &\le y_i + \frac{Q}{2}, \forall i \\ x_{A,i} + D^{\text{dn}} &\ge y_i - \frac{Q}{2}, \forall i \end{aligned} \tag{18}$$

Solving the above simultaneous inequalities leads to the upper and lower bounds of the integral:

$$\begin{aligned} D^{\text{up}} &= \min_i (y_i + \frac{Q}{2} - x_{A,i}) \\ &= \min_i (y_{A,i} - x_{A,i}) + y_D + \frac{Q}{2} \\ D^{\text{dn}} &= \max_i (y_i - \frac{Q}{2} - x_{A,i}) \\ &= \max_i (y_{A,i} - x_{A,i}) + y_D - \frac{Q}{2} \end{aligned} \tag{19}$$

where $i \in \{1, \ldots, N\}$. Defining function $\text{range}(\mathbf{x}) \triangleq \max_i(x_i) - \min_i(x_i)$, the likelihood of $\mathbf{x}_A$ (17) becomes:

$$f(\mathbf{y}|\mathbf{x}_A) = \begin{cases} C(Q - \text{range}(\mathbf{y}_A - \mathbf{x}_A)), & \text{if } \mathbf{x}_A \in \mathbb{F}_A(\mathbf{y}) \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

Feasible space $\mathbb{F}_A(\mathbf{y})$ of the AC signal is determined by $D^{\text{up}} \ge D^{\text{dn}}$, *i.e.* $\text{range}(\mathbf{y}_A - \mathbf{x}_A) \le Q$. Together with the AC definition $\mathbf{1}^\intercal \mathbf{x}_A = \sum_i x_{A,i} = 0$, we have:

$$\mathbb{F}_A(\mathbf{y}) = \{\mathbf{x}_A \mid \text{range}(\mathbf{y}_A - \mathbf{x}_A) \le Q,\ \mathbf{1}^\intercal \mathbf{x}_A = 0\} \tag{21}$$

2) MAP Estimation of AC Signal: By Bayes' theorem, the posterior of the AC signal is proportional to the product of (7) and (20):

$$f(\mathbf{x}_A|\mathbf{y}) \propto \begin{cases} C(Q - \text{range}(\mathbf{y}_A - \mathbf{x}_A)) f(\mathbf{x}_A), & \text{if } \mathbf{x}_A \in \mathbb{F}_A(\mathbf{y}) \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

Hence, our proposed MAP AC estimation becomes:

$$\hat{\mathbf{x}}_A^{\text{ACDC}} = \arg \max_{\mathbf{x}_A \in \mathbb{F}_A(\mathbf{y})} (Q - \text{range}(\mathbf{y}_A - \mathbf{x}_A))\, f(\mathbf{x}_A) \tag{23}$$

3) Approximation Error: We now argue that the MAP solution of AC signal by (23) does give a good approximation of MMSE solution (13), *i.e.* the gap $G_{\mathbf{x}|\mathbf{y}}$ is small. In contrast to the uniform likelihood (4) of the whole signal $\mathbf{x}$, the non-uniform likelihood for AC signal (20) contributes to a posterior PDF (22) that is more symmetric *and* more concentrated.

Above facts can be easier to understand using an example in Fig. 5. The posterior (8) of the whole signal $\mathbf{x}$—which is proportional to the smoothness prior (7) multiplied by a binary likelihood (4)—is shown in Fig. 5(a). We see that the gap between MMSE and MAP solutions is large in this case.

The AC likelihood $f(\mathbf{y}|\mathbf{x}_A)$, shown in Fig. 5(c), is the result of integrating $f(\mathbf{y}|\mathbf{x})$ over feasible DC values. Mathematically, this integral is implied in the multiplication with value $Q - \text{range}(\mathbf{y}_A - \mathbf{x}_A)$ in (23). Due to this integration, the AC likelihood $f(\mathbf{y}|\mathbf{x}_A)$ favors the AC signals closer to $\mathbf{y}_A$ (note the non-uniform likelihood function in 5(c)). Note this behavior is very different from direct MAP estimation of the whole signal (Fig. 5(a)) where the likelihood is uniform within the feasible space.

The posterior of AC signal (22) is proportional to the product of the smoothness prior (23) and the AC likelihood function (20). Due to the non-uniform (20), we have a much smaller gap between MAP and MMSE solution as shown in Fig. 5(b). Following our previous discussion, a small gap means that our MAP AC solution $\hat{\mathbf{x}}_A^{\text{ACDC}}$ has similar MSE performance as the MMSE solution.

C. DC Signal Estimation

Given that the AC signal is estimated as $\hat{\mathbf{x}}_A^{\text{ACDC}}$ via MAP, together with $f(x_D + \hat{\mathbf{x}}_A^{\text{ACDC}}|\mathbf{y}) \propto f(\mathbf{y}|x_D + \hat{\mathbf{x}}_A^{\text{ACDC}}) f(x_D)$, problem (14) becomes:

$$\begin{aligned} &\arg\min_{\hat{x}_D} \; C \int (\hat{x}_D - x_D)^2 \; f(\mathbf{y}|x_D + \hat{\mathbf{x}}_A^{\text{ACDC}}) \; dx_D \\ &\text{s.t.} \;\; \hat{x}_D \mathbf{1} + \hat{\mathbf{x}}_A^{\text{ACDC}} \in \mathbb{F}(\mathbf{y}) \end{aligned} \tag{24}$$

where $f(\mathbf{y}|x_D, \hat{\mathbf{x}}_A^{\text{ACDC}}) = 1$ so long as x_D leads to quantized $\mathbf{y}$ according to (5), *i.e.*

$$y_i - \frac{Q}{2} \le \hat{x}_{A,i}^{\text{ACDC}} + x_D \le y_i + \frac{Q}{2}, \quad \forall i \tag{25}$$

The integral bounds of (24) is determined by (19), so our estimated DC signal is:

$$\begin{aligned} \hat{x}_D^{\text{ACDC}} &= \arg\min_{\hat{x}_D} \int_{\max_i(y_{A,i} - \hat{x}_{A,i}^{\text{ACDC}}) + y_D - \frac{Q}{2}}^{\min_i(y_{A,i} - \hat{x}_{A,i}^{\text{ACDC}}) + y_D + \frac{Q}{2}} \|\hat{x}_D - x_D\|_2^2 \; dx_D \\ &= y_D + \frac{1}{2}(\min_i(y_{A,i} - \hat{x}_{A,i}^{\text{ACDC}}) + \max_i(y_{A,i} - \hat{x}_{A,i}^{\text{ACDC}})) \\ &= y_D - \frac{1}{2}(\min_i(\hat{x}_{A,i}^{\text{ACDC}} - y_{A,i}) + \max_i(\hat{x}_{A,i}^{\text{ACDC}} - y_{A,i})) \end{aligned} \tag{26}$$

where $i \in \{1, \ldots, N\}$.

D. Final Output

Our estimated HBD image is $\hat{\mathbf{x}}^{\text{ACDC}} = \hat{\mathbf{x}}_A^{\text{ACDC}} + \hat{x}_D^{\text{ACDC}} \mathbf{1}$. Note that (25) guarantees that $\hat{\mathbf{x}}^{\text{ACDC}}$ satisfies the quantization constraints (19). Therefore our ACDC bit-depth enhancement method gives a solution with small MSE satisfying all constraints. Calculation of DC signal (26) is trivial. We discuss how MAP AC estimation (23) can be efficiently solved next.

V. Solving the MAP AC Estimation

We now discuss how MAP estimation of the AC signal can be computed efficiently via convex optimization. We insert the graph-signal smoothness prior (7) into (23), and given $\text{range}(\mathbf{x}) = \text{range}(-\mathbf{x})$, we can write MAP AC estimation as:

$$\begin{aligned} &\hat{\mathbf{x}}_A^{\text{ACDC}} = \arg \max_{\mathbf{x}_A \in \mathbb{F}_A(\mathbf{y})} \\ &[\frac{1}{K} \exp\{-\sigma \mathbf{x}_A^\intercal \mathbf{L} \mathbf{x}_A\} \cdot (Q - \text{range}(\mathbf{x}_A - \mathbf{y}_A))] \end{aligned} \tag{27}$$

According to Appendix B, above optimization problem is convex and can be well approximated by a quadratic programming (QP) problem with variables $\mathbf{x}_A^\intercal$, s, and t. By defining $\mathbf{z} = [\mathbf{x}_A^\intercal, s, t]^\intercal \in \mathbb{R}^{N+2}$, the QP is in standard form:

$$\begin{aligned} \min_{\mathbf{z}} \quad & \mathbf{z}^\intercal \mathbf{M} \mathbf{z} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{z} \le \mathbf{b}, \\ & \mathbf{c}^\intercal \mathbf{z} = 0 \end{aligned} \tag{28}$$

where

$$\mathbf{M} = \begin{bmatrix} & & & 0 & 0 \\ & \frac{Q^2}{2\lambda}\mathbf{L} & & \vdots & \vdots \\ & & & 0 & 0 \\ 0 & \ldots & 0 & 1 & -1 \\ 0 & \ldots & 0 & -1 & 1 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} & & & 1 & 0 \\ & -\mathbf{I}_N & & \vdots & \vdots \\ & & & 1 & 0 \\ & & & 0 & -1 \\ & & & 0 & -1 \\ & \mathbf{I}_N & & \vdots & \vdots \\ & & & 0 & -1 \\ 0 & \ldots & 0 & -1 & 1 \end{bmatrix},$$

$$\mathbf{b} = [-\mathbf{y}_A^\intercal, \mathbf{y}_A^\intercal, Q]^\intercal, \quad \mathbf{c} = [1, \cdots, 1, 0, 0]^\intercal$$

QP can be solved very efficiently using, for example, interior-point algorithms [44]. Although matrix $\mathbf{M}$ is typically large in size ($N \times N$), it is very sparse (the number of nonzero entries is less than $5N + 4$ for a 4-connected graph). The sparsity structure ensures efficient computation of QP.

A. Boundary Conditions

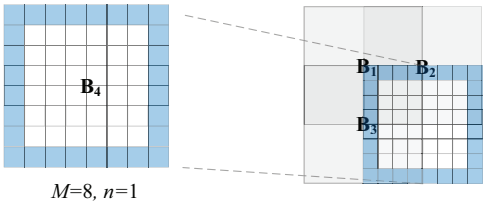


Fig. 6: Illustration of boundary pixels and boundary condition. Suppose the gray blocks have been processed, $\mathbf{B}_4$ is the current 8×8 block with 28 boundary pixels (marked in blue).

Because we apply bit-depth enhancement block by block, there may be block artefacts along block boundaries. We propose to enforce the following boundary condition for each block. First, we define *boundary pixels* to be pixels on the circumference of a block of width n. We then constrain the boundary pixels to be exactly the same as the most recent estimated values. If a boundary pixel i does not belong to any previously optimized block, then its most recent estimated

value is simply y_i. Fig. 6 shows an example of boundary pixels with $M = 8, n = 1$.

To achieve this, we need to add a set of equality constraints to the QP formulation: $\mathbf{S}\mathbf{x}_A + \hat{x}_D^{\mathrm{ACDC}}(\mathbf{x}_A) = \mathbf{S}\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is the latest estimated signal (for a pixel i, if x_i has not been estimated by previous blocks, use y_i as the latest estimate); $\mathbf{S}$ is a matrix whose each row, $\mathbf{s}_k \in \mathbb{R}^N$, is a selection vector with form $[0, 0, \cdots 0, 1, 0 \cdots, 0]$ indicating the k-th boundary pixel. By (26) we know $\hat{x}_D^{\mathrm{ACDC}}(\mathbf{x}_A) = -\frac{1}{2}(s+t) + y_D$, so above constraint can be rewritten as:

$$[\mathbf{s}_k, -0.5, -0.5]\mathbf{z} = \mathbf{s}_k\hat{\mathbf{x}} - y_D, \quad \forall \text{ row } k \tag{29}$$

Therefore with boundary conditions the formulation (28) becomes:

$$\begin{aligned} \min_{\mathbf{z}} \quad & \mathbf{z}^\intercal \mathbf{M} \mathbf{z} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{z} \leq \mathbf{b}, \\ & \mathbf{C}^\intercal \mathbf{z} = \mathbf{d} \end{aligned} \tag{30}$$

where

$$\mathbf{C} = \begin{bmatrix} 1 & \cdots & 1 & 0 & 0 \\ & & & -0.5 & -0.5 \\ & \mathbf{S} & & \vdots & \vdots \\ & & & -0.5 & -0.5 \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} 0 \\ \mathbf{S}\hat{\mathbf{x}} - y_D\mathbf{1} \end{bmatrix}$$

Compared to (28), solving the above QP leads to the estimated AC signal with boundary consistency. The boundary condition guarantees continuous transition from block to block, thus alleviates block artefacts.

VI. Experiments

A. Experiment Setup

To validate our proposed bit-depth enhancement algorithm experimentally, we use a collection of test images [45] as shown in Fig. 7. All test images have bit-depth greater than 8. The HBD test images serve as the ground-truth $\mathbf{x}^o$, and the input $\mathbf{y}$ to our bit-depth enhancement problem is the quantized version of $\mathbf{x}^o$ with low bit-depth $b = 4$ or 6 or 8.

We compare the objective performance of our proposed bit-depth enhancement algorithm with the following competing schemes: 1) ANC, the anchor method that simply uses $\mathbf{y}$ as the output; 2) EPF, which estimates a high bit-depth pixel value from neighboring pixels through an edge preserving filter [22]; 3) DEC, a filtering-based method for removing false contours [7]; 4) INT, a spatial interpolation method in spatial domain [5]; 5) MRC, a bit-depth enhancement method using minimum risk based classification [23]; and 6) ACDC, our proposed two-step bit-depth enhancement method.

Given that we enhance block-by-block, we set block size $M = 64$ with overlapped width of 32 pixels and a boundary pixel width of $n = 2$. To compute a graph Laplacian $\mathbf{L}$ to define the graph-signal smoothness prior, we first construct a 4-connected graph for pixels in the target block. We then compute the weight of an edge connecting two horizontally / vertically adjacent pixels as: $w_{i,j} = 1$ if $|y_i - y_j| \leq Q$, otherwise $w_{i,j} = 0$. For color images, the graph weights are calculated using the maximal absolute differences in three channels, and different color channels share the same graph Laplacian. Because our MMSE objective is defined statistically, for fixed λ there is no guarantee that our computed solution is close to a deterministic ground-truth image $\mathbf{x}^o$. For optimal quality, we empirically adjust λ by estimating the degree of smoothness in the signal.

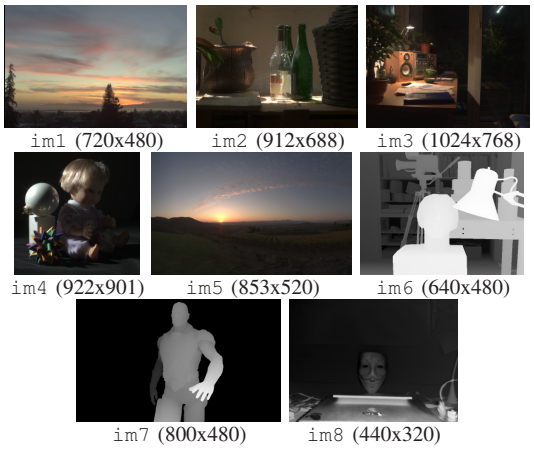


Fig. 7: HBD test images that serve as the ground-truth in objective bit-depth enhancement experiments.

B. Objective Comparisons

We use two objective metrics for rendered image evaluation: peak signal to noise ratio (PSNR) and an alternative local metric—*segmental SNR* (SSNR). Analogous to the segmental SNR in audio processing literature [46, 47], the SSNR for image is defined as the average SNR of non-overlapping patches (we set patch size to be 100×100 in our experiments). Different from the traditional SNR definition, the signal energy here is the energy of the AC signal for each patch. The intuition is that human eyes can adapt to the DC level of each local block [48], thus the SNR using AC signal would better characterize the perceived image quality by the human visual system. The numerical results are summarized in Table I, which shows our proposed ACDC performs consistently well in 4-/6-/8-bit enhancement experiments because it not only utilizes quantization constraints into the likelihood term, but also embeds known image structures into graph weights, enabling the reconstruction of edge-aware smooth signals.

Fig. 8, Fig. 9 and Fig. 10 show the visual results of competing methods for $b = 4, 6, 8$ respectively. Given the b-bit LBD input image, we re-quantize the output HBD image in h-bit ($h > b$) to simulate the effect of displaying the output images on a $h-$bit display device. When $h > 8$ one cannot visually discern HBD images in a PDF document, so for output images in $b = 6$ or 8 we instead show the local regions whose intensity levels can be represented as 8-bit numbers. We see that output images of proposed ACDC have no contouring artifacts or blocky artifacts, in which true edges are well-preserved from being smoothed out.

C. Subjective Comparisons

We conducted subjective testing to verify the effectiveness of our proposed ACDC method compared to other ones. We

Fig. 8: Output HBD images of competing methods in 4-bit experiment ($b = 4$) for test image `im1`.

TABLE II: Subjective evaluation of different bit-depth enhancement methods for 4-bit and 6-bit images. Each table cell shows the vote statistics and the corresponding p-value of two-sided χ^2 test.

	`duck` (4-bit)	`bill` (4-bit)	`lion` (4-bit)	`chair` (4-bit)	`ball` (4-bit)	`apple` (4-bit)	`banana` (4-bit)	`sack` (4-bit)	`bill` (6-bit)	`lion` (6-bit)	`chair` (6-bit)	`banana` (6-bit)
ACDC:ANC	18:0	14:1	18:0	13:1	17:0	18:0	13:1	17:0	15:0	14:0	13:1	10:4
p-value	2.2e-5	7.9e-4	2.2e-5	1.3e-3	3.7e-5	2.2e-5	1.3e-3	3.7e-5	1.1e-4	1.8e-4	1.3e-3	0.11
ACDC:MRC	15:0	18:0	15:2	17:0	14:0	11:6	15:0	17:1	18:0	15:2	14:3	12:3
p-value	1.1e-4	2.2e-5	1.6e-3	3.7e-5	1.8e-4	0.23	1.1e-4	1.6e-4	2.2e-5	1.6e-3	7.6e-3	0.02
ACDC:DEC	17:0	16:1	11:3	13:2	13:2	11:3	18:0	12:2	12:2	9:6	18:0	9:9
p-value	3.7e-5	2.7e-4	3.3e-2	4.5e-3	4.5e-3	3.3e-2	2.2e-5	7.5e-3	7.5e-3	0.44	2.2e-5	1.0
ACDC:GT	1:13	8:6	7:8	2:16	5:13	3:10	1:16	5:10	8:9	3:10	5:10	9:8
p-value	1.3e-3	0.59	0.80	9,7e-4	5.9e-2	5.2e-2	2.7e-4	0.20	0.81	0.05	0.20	0.81

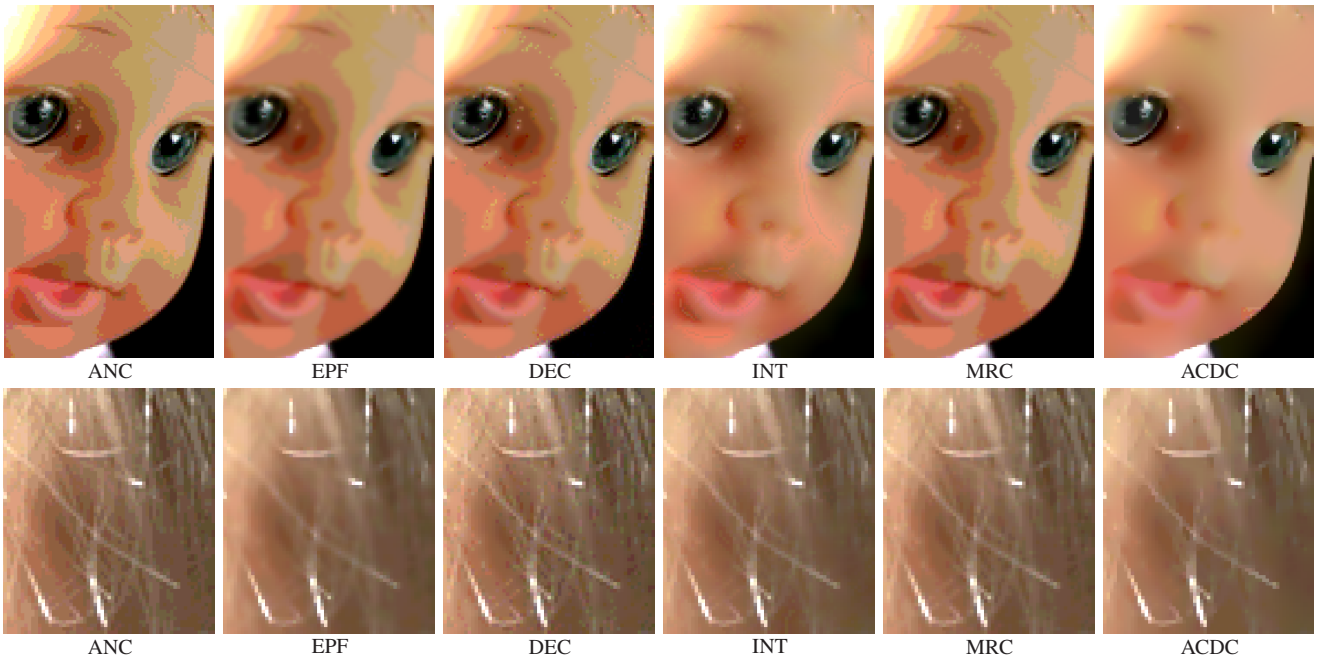


Fig. 9: Simulation of displaying the output images in 6-bit experiment ($b = 6$) on HBD monitor (cropped regions of `im4`).

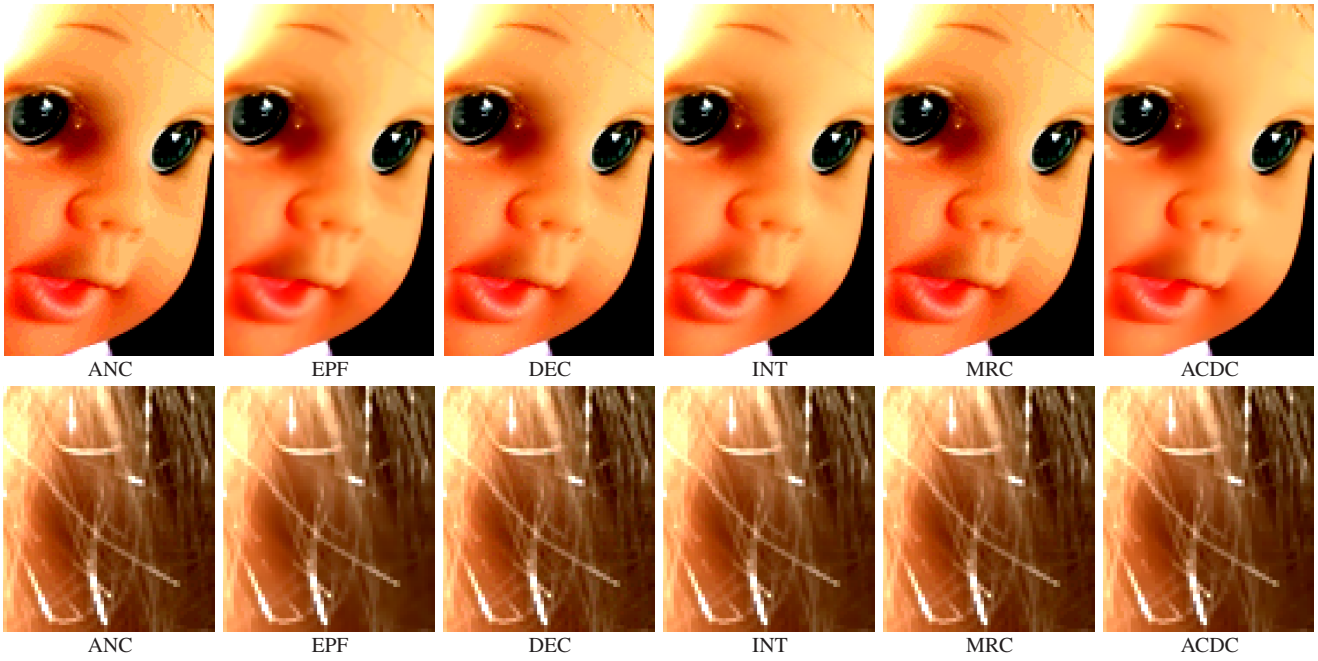


Fig. 10: Simulation of displaying the output images in 8-bit experiment ($b = 8$) on a HBD monitor (cropped regions of `im4`).

used eight HBD test images selected from the HBD image database in [49] for testing, which had the same resolution and aspect ratio for ease of side-by-side comparison on a 16:9 display monitor in landscape mode. They are shown in Fig. 11. For each image, ACDC competed against four methods—ANC, EPF, DEC and GT (the ground truth)—in four separate side-by-side comparison tests (also called *two alternative forced choice* (2AFC) [50]); *i.e.*, for each test pair a human test subject was asked to select the better of the two versions of the same image rendered side-by-side. The four side-by-side comparisons were randomly slotted into four test sequences, where for each sequence, the ordering of the eight test image pairs was also randomized. Each test subject randomly chose two test sequences to observe back-to-back and enter scores. Each image pair was observed for 10 seconds, and the score was entered in the next 5 seconds. This testing procedure followed closely guidelines provided by ITU-R BT.500 [51].

We invited 31 participants for subjective testing, with age 23 to 48 and with normal or corrected to normal vision. Each image pair was shown on a 8-bit Eizo EV2750-BKR 27" monitor with contrast and bright set at 75%. The distance from

TABLE I: Objective results. For each cell, the number on the top and bottom are the PSNR and SSNR in dB, respectively.

(a) 4-bit experiment ($b = 4$)

	ANC	EPF	DEC	INT	MRC	ACDC
im1	34.87 13.53	36.68 15.26	35.53 14.16	36.53 15.18	36.49 15.07	**37.84** **16.57**
im2	35.04 9.59	36.39 10.87	35.60 10.16	35.78 10.06	36.35 10.81	**37.37** **11.70**
im3	34.39 6.47	35.29 7.30	34.61 6.69	33.88 5.82	35.34 7.35	**37.66** **9.94**
im4	35.04 5.35	35.96 5.64	35.32 5.72	33.95 3.99	36.06 6.30	**36.93** **6.85**
im5	34.87 6.20	36.31 7.69	35.29 6.55	37.77 9.24	36.17 7.60	**38.29** **9.69**
im6	34.88 16.37	35.20 16.72	34.88 16.34	33.52 14.59	35.15 16.67	**36.25** **17.68**
im7	35.41 9.10	36.18 9.67	35.83 8.97	34.85 9.26	36.00 9.62	**37.31** **10.96**
im8	34.33 7.46	35.37 8.87	34.78 8.05	35.14 9.73	35.34 8.93	**37.90** **10.79**
Ave.	34.85 9.26	35.92 10.25	35.23 9.58	35.18 9.73	35.86 10.29	**37.44** **11.77**

(b) 6-bit experiment ($b = 6$)

	ANC	EPF	DEC	INT	MRC	ACDC
im1	46.80 25.43	48.18 26.93	45.92 24.57	48.31 26.95	48.83 27.38	**49.03** **27.67**
im2	46.87 21.08	48.05 22.50	46.39 20.84	47.95 22.21	46.68 22.94	**49.09** **23.43**
im3	46.89 18.81	47.47 19.62	46.32 18.47	47.66 19.76	48.10 20.03	**48.90** **21.00**
im4	47.26 16.60	47.30 16.75	46.69 16.01	47.26 16.77	48.33 17.53	**49.09** **18.26**
im5	46.87 18.09	46.89 18.16	45.97 17.16	48.82 20.38	48.37 19.49	**49.21** **20.50**
im6	46.88 27.78	47.63 28.62	47.01 28.02	46.65 27.59	47.61 28.55	**48.79** **29.64**
im7	46.96 20.59	49.43 23.48	48.50 21.99	50.94 25.10	48.84 22.71	**51.12** **25.01**
im8	46.86 19.61	48.33 21.06	46.79 19.40	49.99 22.52	48.34 21.24	**50.64** **23.22**
Ave.	46.92 21.00	47.91 22.14	46.70 20.81	48.45 22.66	48.14 22.48	**49.48** **23.59**

(c) 8-bit experiment ($b = 8$)

	ANC	EPF	DEC	INT	MRC	ACDC
im1	58.82 37.44	55.55 34.54	56.01 34.74	57.73 36.46	58.85 37.46	**58.99** **37.63**
im2	58.91 33.12	57.44 32.00	56.49 30.83	58.42 32.84	58.97 33.17	**59.57** **33.89**
im3	58.89 30.78	58.24 30.73	57.11 29.34	58.89 31.38	58.89 30.78	**60.19** **32.45**
im4	58.92 28.07	58.89 28.28	57.62 26.83	59.18 28.86	58.89 28.28	**60.83** **30.38**
im5	58.93 30.14	56.53 27.99	56.56 27.80	58.40 29.72	58.92 30.10	**59.44** **30.72**
im6	58.91 39.56	60.15 40.86	59.09 39.85	60.86 41.75	58.87 39.52	**61.08** **41.96**
im7	59.03 32.46	**63.43** **36.58**	61.24 34.12	62.39 35.34	58.98 32.42	62.55 35.62
im8	58.94 31.70	58.25 30.86	57.27 29.78	59.71 32.11	58.55 31.05	**60.25** **32.79**
Ave.	58.92 32.91	58.56 32.73	57.67 31.66	59.45 33.56	58.87 32.85	**60.36** **34.43**

Fig. 11: HBD test images used for subjective experiments. The spatial resolution is 1200×1200.

Fig. 12: Cropped test images used for subjective experiments.

the subject to the monitor is approximately twice the monitor's height (335.7mm). The illumination of the room was in the 300-320 Lux range. Each participant was familiarized with the testing procedure before the start of the experiment.

Table II summarizes the subjective test results, where we show the participates' votes for different methods in 4-bit and 6-bit experiments. We see that for all 4-bit test images, the participants selected ACDC over competing schemes by a clear margin. This validates the superior visual quality of ACDC over competing schemes. When compared to GT, we see that ACDC was comparable in two images, but was inferior in the remaining six images. This shows that there are even cases where ACDC is indistinguishable from the original high bit-depth images (GT).

We also conducted a similar subjective experiment where four 6-bit images were enhanced for cropped image patches; the images were cropped to isolate spatial regions with gradations that are known to be problematic for bit-depth enhancement algorithms, resulting in false contours. The cropped images are shown in Fig. 12. We observe that ACDC outperformed ANC and MRC for all images. For DEC, ACDC is superior for two images and indistinguishable for the other two images. Overall, ACDC is still better than DEC. Compared to the ground-truth GT, ACDC is comparable in two images. That indicates the reconstructed images by our proposed method are of high visual quality.

We used the two-sided chi-square χ^2 test [52] to examine the statistical significance of the results. The null hypothesis is that there is no preference for proposed ACDC and a competing methods. Under this hypothesis, the expected number of votes should be equal. The p-value [52] is also indicated in the table. As a rule of thumb in experimental sciences, the null hypothesis is rejected when $p < 0.05$. As seen in Table II, the majority of the p-values are much smaller than 0.05. We can

thus conclude that subjects showed a statistically significant preference for our proposed method ACDC.

VII. Conclusion

We proposed a computation-efficient algorithm for the image bit-depth enhancement problem, where an image of low bit-depth (LBD) is reconstructed to one of high bit-depth (HBD). Specifically, after formulating an optimization problem with a minimum mean squared error (MMSE) objective and arguing that the problem is difficult to solve directly, we propose to first use MAP to estimate the AC signal—efficiently solved via quadratic programming—and then compute the DC signal with an MMSE criterion. Experiments show that our proposed two-step method outperforms competing methods.

Appendix A
MSE of MMSE and MAP Solution

Proof. By (12), the estimation error of an estimator $\hat{\mathbf{x}}$ is: $\text{MSE}(\hat{\mathbf{x}}) = \int \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \; f(\mathbf{x} \,|\, \mathbf{y}) \; d\mathbf{x} = \mathbf{E}(\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 | \mathbf{y})$.

For a random variable $\mathbf{r}$, the mean-variance decomposition $\mathbf{E}(\|\mathbf{r}\|_2^2) = \mathbf{E}(\|\mathbf{r} - \mathbf{E}(\mathbf{r}) + \mathbf{E}(\mathbf{r})\|_2^2) = \mathbf{E}(\|\mathbf{r} - \mathbf{E}(\mathbf{r})\|_2^2) + \|\mathbf{E}(\mathbf{r})\|_2^2$ always holds. Replacing $\mathbf{r}$ by $\mathbf{x} - \hat{\mathbf{x}}$ yields $\mathbf{E}(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2) = \mathbf{E}(\|\mathbf{x} - \mathbf{E}(\mathbf{x})|_2^2) + \|\mathbf{E}(\mathbf{x}) - \hat{\mathbf{x}}\|_2^2$. Given the quantized signal $\mathbf{y}$, above equation becomes: $\text{MSE}(\hat{\mathbf{x}}) = \mathbf{E}(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 | \mathbf{y}) = \mathbf{E}(\|\mathbf{x} - \mathbf{E}(\mathbf{x}|\mathbf{y})\|_2^2 | \mathbf{y}) + \|\mathbf{E}(\mathbf{x}|\mathbf{y}) - \hat{\mathbf{x}}\|_2^2$.

Equation (10) is obtained by substituting $\hat{\mathbf{x}}$ with the MMSE solution $E(\mathbf{x}|\mathbf{y})$ and MAP solution $\hat{\mathbf{x}}^{\text{MAP}}$. □

Appendix B
Reformulating (27) to a QP

Proof. Taking the negative log of (27) leads to the following convex optimization problem:

$$\begin{aligned} \min_{\mathbf{x}_A} \quad & \mathbf{x}_A^\intercal \mathbf{L} \mathbf{x}_A - \lambda \log(Q - \text{range}(\mathbf{x}_A - \mathbf{y}_A)) \\ \text{s.t.} \quad & \text{range}(\mathbf{x}_A - \mathbf{y}_A) < Q, \\ & \mathbf{1}^\intercal \mathbf{x}_A = 0 \end{aligned} \tag{31}$$

where $\lambda = 1/\sigma$. Convexity of the above problem is proved in Appendix C. By introducing two extra scalar variables s and t, the above problem can be rewritten as:

$$\begin{aligned} \min_{\mathbf{x}_A, s, t} \quad & \mathbf{x}_A^\intercal \mathbf{L} \mathbf{x}_A - \lambda \log(Q - (t - s)) \\ \text{s.t.} \quad & s\mathbf{1} \le \mathbf{x}_A - \mathbf{y}_A \le t\mathbf{1}, \\ & t - s \le Q, \\ & \mathbf{1}^\intercal \mathbf{x}_A = 0 \end{aligned} \tag{32}$$

We note that $\log(Q - (t-s))$ promotes a small $t - s$ value. Towards efficient computation, we replace $\log(Q + s - t)$ with its second-order approximation $-\frac{2}{Q^2}(s-t)^2$ (See Appendix D for details). Then problem (32) becomes a Quadratic Programming (QP) problem:

$$\begin{aligned} \min_{\mathbf{x}_A, s, t} \quad & \mathbf{x}_A^\intercal \mathbf{L} \mathbf{x}_A + \frac{2\lambda}{Q^2}(t-s)^2 \\ \text{s.t.} \quad & s\mathbf{1} \le \mathbf{x}_A - \mathbf{y}_A \le t\mathbf{1}, \\ & t - s \le Q, \\ & \mathbf{1}^\intercal \mathbf{x}_A = 0 \end{aligned} \tag{33}$$

□

Appendix C
Proof of Convexity of (31)

Proof. Let $f(\mathbf{x}_A) = Q - \text{range}(\mathbf{x}_A - \mathbf{y}_A)$, $g(x) = -\log(x)$. Because quadratic function $\mathbf{x}_A^\intercal \mathbf{L} \mathbf{x}_A$ is convex, we only need to prove $g(f(\mathbf{x}_A))$ is convex. According to the composition rule "if f is concave and g is convex and non-increasing, then $h(x) = g(f(x))$ is convex" [43], the problem becomes to prove $f(\mathbf{x}_A)$ is concave, *i.e.* $\text{range}(\mathbf{x}_A - \mathbf{y}_A)$ is convex. Since $\mathbf{y}_A$ is constant, we next prove the convexity of $\text{range}(\mathbf{x}_A)$.

Denote $\theta \in [0,1]$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$, by definition $\text{range}(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) = \max_i(\theta x_{1,i} + (1-\theta)x_{2,i}) - \min_i(\theta x_{1,i} + (1-\theta)x_{2,i})$. For any two vectors $\mathbf{v}_1, \mathbf{v}_2$, apparently $\max_i(v_{1,i} + v_{2,i}) \le \max_i(v_{1,i}) + \max_i(v_{2,i})$ and $\min_i(v_{1,i} + v_{2,i}) \ge \min_i(v_{1,i}) + \min_i(v_{2,i})$. Letting $\mathbf{v}_1 = \theta \mathbf{x}_1$ and $\mathbf{v}_2 = (1-\theta)\mathbf{x}_2$, we have $\text{range}(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \le \theta \max_i(x_{1,i}) + (1-\theta)\max_i(x_{2,i}) - \theta \min_i(x_{1,i}) - (1-\theta)\min_i(x_{2,i}) = \theta \text{range}(\mathbf{x}_1) + (1-\theta)\text{range}(\mathbf{x}_2)$. By definition of convexity, function $\text{range}(\mathbf{x}_A)$ is convex, so is our objective $\mathbf{x}_A^\intercal \mathbf{L} \mathbf{x}_A - \lambda \log(Q - \text{range}(\mathbf{x}_A - \mathbf{y}_A))$. □

Appendix D
Second-order Approximation of log Function

Proof. The Taylor expansion of function $\log(x)$ at position a is: $\log(x) = \log(a) + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (\frac{x-a}{a})^k$. Because $t - s \in [0, Q]$, variable $x = Q - (t-s) \in [0, Q]$. Letting $a = \frac{Q}{2}$ be the center of the domain, we have $\frac{x-a}{a} = \frac{x}{a} - 1 \in [-1, 1]$. That means we can dismiss the high order ($k \ge 3$) polynomials and get the second-order approximation of $\log(x)$ as $\log(x) \approx \log(a) - \frac{x-a}{a} + \frac{1}{2}(\frac{x-a}{a})^2 = \text{const} + \frac{2x}{a} - \frac{x^2}{2a^2} = \text{const} + \frac{4x}{Q} - \frac{2x^2}{Q^2}$. Substituting $x = Q + s - t$ into above equation, we get $\log(Q + s - t) \approx \text{const} - \frac{2}{Q^2}(s-t)^2$. □

References

[1] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High dynamic range imaging: acquisition, display, and image-based lighting*, Morgan Kaufmann, 2010.

[2] P. Milanfar, *Super-Resolution Imaging (Digital Imaging and Computer Vision)*, CRC Press, September 2010.

[3] Min-Ho M.-H. Park, J. W. Lee, Rae-Hong R.-H. Park, and J.-S. Kim, "False contour reduction using neural networks and adaptive bi-directional smoothing," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 870–878, 2010.

[4] X. Jin, S. Goto, and K. N. Ngan, "Composite model-based dc dithering for suppressing contour artifacts in decompressed video," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2110–2121, 2011.

[5] P. Wan, O. C. Au, K. Tang, Y. Guo, and L. Fang, "From 2D extrapolation to 1D interpolation: Content adaptive image bit-depth expansion," in *Proc. IEEE International Conference on Multimedia & Expo (ICME)*, Melbourne, Australia, 2012.

[6] P. Wan, O. C. Au, K. Tang, and Y. Guo, "Image de-quantization via spatially varying sparsity prior," in *IEEE International Conference on Image Processing*. IEEE, 2012, pp. 953–956.

[7] S. Daly and X. Feng, "Decontouring: Prevention and removal of false contour artifacts," in *Proc. SPIE Human Vision and Electronic Imaging IX*, 2004, vol. 5292, pp. 130–149.

[8] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," in *IEEE Signal Processing Magazine*, May 2013, vol. 30, no.3, pp. 83–98.

[9] W. Hu, X. Li, G. Cheung, and O. Au, "Depth map denoising using graph-based transform and group sparsity," in *IEEE International Workshop on Multimedia Signal Processing*, Pula, Italy, October 2013.

[10] W. Hu, G. Cheung, A. Ortega, and O. Au, "Multi-resolution graph Fourier transform for compression of piecewise smooth images," in *IEEE Transactions on Image Processing*, January 2015, vol. 24, no.1, pp. 419–433.

[11] W. Hu, G. Cheung, and A. Ortega, "Intra-prediction and generalized graph Fourier transform for image coding," in *IEEE Signal Processing Letters*, November 2015, vol. 22, no.11, pp. 1913–1917.

[12] W. Hu, G. Cheung, and M. Kazui, "Graph-based dequantization of block-compressed piecewise smooth images," in *IEEE Signal Processing Letters*, February 2016, vol. 23, no.2, pp. 242–246.

[13] Y. Mao, G. Cheung, A. Ortega, and Y. Ji, "Expansion hole filling in depth-image-based rendering using graph-based interpolation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013.

[14] Y. Mao, G. Cheung, and Y. Ji, "Image interpolation during DIBR view synthesis using graph Fourier transform," in *3DTV-Conference*, Budapest, Hungary, July 2014.

[15] J. Pang, G. Cheung, W. Hu, and O. C. Au, "Redefining self-similarity in natural images for denoising using graph signal gradient," in *APSIPA ASC*, Siem Reap, Cambodia, December 2014.

[16] J. Pang, G. Cheung, A. Ortega, and O. C. Au, "Optimal graph Laplacian regularization for natural image denoising," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Brisbane, Australia, April 2015.

[17] X. Liu, G. Cheung, and X. Wu, "Joint denoising and contrast enhancement of images using graph laplacian operator," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Brisbane, Australia, April 2015.

[18] X. Liu, G. Cheung, X. Wu, and D. Zhao, "Inter-block soft decoding of JPEG images with sparsity and graph-signal smoothness priors," in *IEEE International Conference on Image Processing*, Quebec City, Canada, September 2015.

[19] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Inverse tone mapping," in *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*. ACM, 2006, pp. 349–356.

[20] A. Rempel et. al, "Ldr2hdr: on-the-fly reverse tone mapping of legacy video and photographs," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp. 39, 2007.

[21] Robert Ulichney, , Robert Ulichney, and Shiufun Cheung, "Pixel bit-depth increase by bit replication," in *Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III, Proc. SPIE*, 1998.

[22] Seungsin Lee, Du-Sik Park, and Chang-Yeong Kim, "Visual bit resolution enhancement algorithm for digital tvs," in *The 12th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, 2006.

[23] Gaurav Mittal, Vinit Jakhetiya, Sunil Prasad Jaiswal, Oscar C. Au, Anil Kumar Tiwari, and Dai Wei, "Bit-depth expansion using minimum risk based classification," in *IEEE Visual Communications and Image Processing(VCIP)*, San Diego, US, 2012, pp. 1–5.

[24] L. Schuchman, "Dither signals and their effect on quantization noise," *Communication Technology, IEEE Transactions on*, vol. 12, no. 4, pp. 162–165, 1964.

[25] Scott J. Daly and Xiaofan Feng, "Bit-depth extension using spatiotemporal microdither based on models of the equivalent input noise of the visual system," in *Proc. SPIE*, 2003, vol. 5008, pp. 455–466.

[26] P. Wan, O. C. Au, J. Pang, K. Tang, and R. Ma, "High bit-precision image acquisition and reconstruction by planned sensor distortion," in *Proc. IEEE International Conference on Image Processing*, Paris, France, 2014.

[27] C. H. Cheng, O. C. Au, C. H. Liu, and K. Y. Yip, "Bit-depth expansion by contour region reconstruction," in *Proc. of IEEE Int. Sym. on Circuits and Systems*, 2009, pp. 944–947.

[28] P. Wan, G. Cheung, D. Forencio, C. Zhang, and O. C. Au, "Image bit-depth enhancement via maximum-a-posteriori estimation of graph ac component," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Paris, France, 2014.

[29] C. Zhang, D. Florencio, and C. Loop, "Point cloud attribute compression with graph transfrom," in *IEEE International Conference on Image Processing*. IEEE, October 2014.

[30] Wikipedia, "Mid-riser Quantizer," http://en.wikipedia.org/wiki/Quantization_(signal_processing), accessed April 2015.

[31] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Fundamentals of Statistical Signal Processing. PTR Prentice-Hall, 1993.

[32] Tomaso Poggio, Vincent Torre, and Christof Koch, "Computational vision and regularization theory," *Nature*, vol. 317, no. 26, pp. 314–319, Sept. 1985.

[33] Priyam Chatterjee and Peyman Milanfar, "Is denoising dead?," *Image Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 895–911, 2010.

[34] Qi Shan, Jiaya Jia, and Aseem Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 73:1–73:10, Aug. 2008.

[35] Stan Z Li, *Markov random field modeling in image analysis*, Springer Science & Business Media, 2009.

[36] Leonid I Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[37] Anil K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.

[38] Y. Xu, J.B. Weaver, Jr. Healy, D.M., and J. Lu, "Wavelet transform domain filters: a spatially selective noise filtration technique," *Image Processing, IEEE Transactions on*, vol. 3, no. 6, pp. 747–758, Nov 1994.

[39] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, 1998.

[40] Cha Zhang and Dinei Florêncio, "Analyzing the optimality of predictive transform coding using graph-based models," *Signal Processing Letters, IEEE*, vol. 20, no. 1, pp. 106–109, 2013.

[41] Wikipedia, "MMSE," http://en.wikipedia.org/wiki/Minimum_mean_square_error, accessed April 2015.

[42] Marcelo GS Bruno, "Sequential monte carlo methods for nonlinear discrete-time filtering," *Synthesis Lectures on Signal Processing*, vol. 6, no. 1, pp. 1–99, 2013.

[43] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[44] Wikipedia, "Quadratic Programming," http://en.wikipedia.org/wiki/Quadratic_programming, accessed April 2015.

[45] P. Wan, "Details on the test images," http://ihome.ust.hk/~leoman/test_images.txt, accessed April 2015.

[46] Schuyler R Quackenbush, Thomas Pinkney Barnwell, and Mark A Clements, *Objective measures of speech quality*, Prentice Hall, 1988.

[47] John R Deller, John G Proakis, and John HL Hansen, *Discrete-time processing of speech signals*, IEEE New York, NY, USA:, 2000.

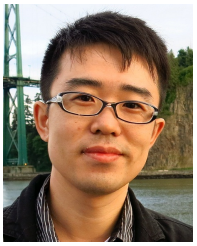
[48] Horace B. Barlow, "Dark and light adaptation: Psychophysics," in *Visual Psychophysics*, vol. 7 / 4 of *Handbook of Sensory Physiology*, pp. 1–28. Springer Berlin Heidelberg, 1972.

[49] Asuni N and Giachetti A, "Testimages: a large-scale archive for testing visual devices and basic image processing algorithms," 2014.

[50] M. Taylor and C. Creelman, "PEST: Efficient estimates on probability functions," in *J. Acoustical Society of America*, 1967, vol. 41, pp. 782–787.

[51] ITU-R, *Recommendation ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures*, January 2012.

[52] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, 2007.



Pengfei Wan is with Meitu, Inc. He received the B.E. degree in Electronic Engineering and Information Science from the University of Science and Technology of China (USTC), Hefei, China in 2010, and the Ph.D. degree in Electronic and Computer Engineering from the Hong Kong University of Science and Technology (HKUST) in 2015.

His research interests include image/video signal processing, computational photography, and machine learning techniques for computer vision.

Gene Cheung (M'00—SM'07) received the B.S. degree in electrical engineering from Cornell University in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1998 and 2000, respectively.

He was a senior researcher in Hewlett-Packard Laboratories Japan, Tokyo, from 2000 till 2009. He is now an associate professor in National Institute of Informatics in Tokyo, Japan. He has been an adjunct associate professor in the Hong Kong University of Science & Technology (HKUST) since 2015.

His research interests include 3D image processing, graph signal processing, and signal processing for sleep analysis. He has served as associate editor for IEEE Transactions on Multimedia (2007–2011), DSP Applications Column in IEEE Signal Processing Magazine (2010–2014) and SPIE Journal of Electronic Imaging (2014–2016). He currently serves as associate editor for IEEE Transactions on Image Processing (2015–present), IEEE Transactions on Circuits and Systems for Video Technology (2016–present) and APSIPA Journal on Signal & Information Processing (2011–present), and as area editor for EURASIP Signal Processing: Image Communication (2011–present). He is a distinguished lecturer in APSIPA (2016–2017). He served as a member of the Multimedia Signal Processing Technical Committee (MMSP-TC) in IEEE Signal Processing Society (2012–2014), and a member of the Image, Video, and Multidimensional Signal Processing Technical Committee (IVMSP-TC) (2015–2017). He has also served as technical program co-chair of International Packet Video Workshop (PV) 2010 and IEEE International Workshop on Multimedia Signal Processing (MMSP) 2015, and symposium co-chair for CSSMA Symposium in IEEE GLOBECOM 2012. He is a co-author of best student paper award in IEEE Workshop on Streaming and Media Communications 2011 (in conjunction with ICME 2011), best paper finalists in ICME 2011, ICIP 2011 and ICME 2015, best paper runner-up award in ICME 2012 and best student paper award in ICIP 2013.

Dinei Florencio (M'96—SM'05—F'16) received the BS and MS degrees from University of Braslia, and the PhD from Georgia Tech, all in Electrical Engineering. He is a researcher with Microsoft Research since 1999. From 1996 to 1999, he was a member of the research staff at the David Sarnoff Research Center.

Dr. Florencios current research focus includes signal processing and computer security. His research has enhanced the lives of millions of people, through high impact technology transfers to many Microsoft products, including Internet Explorer, Live Messenger, Exchange Server, RoundTable, and the MSN toolbar.

Dr. Florencio has authored over 80 referred papers, and holds 57 granted US patents. His papers received awards at ICME2010, SOUPS2010, MMSP09, MMSP12, and ICIP14. Dr. Florencio was general co-chair of CBSP08, MMSP'09, Hot3D10 and 13, and WIFS11, and technical co-chair of WIFS10, ICME11, and MMSP13.

Dr. Florencio is a Fellow of the IEEE, and was Chair of the IEEE SPS Technical Committee on Multimedia Signal Processing and a member of the IEEE SPS Technical Directions Board (2014-2015).

Cha Zhang (M'04—SM'09) is a Principal Researcher in the Multimedia, Interaction and eXperience Group at Microsoft Research. He received the B.S. and M.S. degrees from Tsinghua University, Beijing, China in 1998 and 2000, respectively, both in Electronic Engineering, and the Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University, in 2004. His current research focuses on applying various audio/image/video processing and machine learning techniques to multimedia applications. Dr. Zhang has published more than 80 technical papers and holds 20+ U.S. patents. He won the best paper award at ICME 2007, the top 10% award at MMSP 2009, and the best student paper award at ICME 2010. He currently serves as an Associate Editor for IEEE Trans. on Circuits and Systems for Video Technology, and IEEE Trans. on Multimedia.

Oscar C. Au received his B.A.Sc. from Univ. of Toronto in 1986, his M.A. and Ph.D. from Princeton Univ. in 1988 and 1991 respectively. After being a postdoc in Princeton for 1 year, he joined Hong Kong Univ. of Science and Technology (HKUST) as an Assistant Professor in 1992 and promoted to Associate Professor and then Full Professor. He left HKUST in 2014 and migrated to the USA. His main research contributions are on video/image coding and processing, watermarking/light weight encryption, speech/audio processing. Research topics include fast motion estimation for H.261/3/4/5, MPEG-1/2/4, and AVS, optimal and fast sub-optimal rate control, mode decision, transcoding, denoising, deinterlacing, post-processing, multi-view coding, view interpolation, depth estimation, 3DTV, scalable video coding, distributed video coding, subpixel rendering, JPEG/JPEG2000, HDR imaging, compressive sensing, halftone image data hiding, GPU-processing, software-hardware co-design, etc. He has published 70 technical journal papers, 370+ conference papers, 3 book chapters, and 70+ contributions to international standards. His fast motion estimation algorithms were accepted into the ISO/IEC 14496-7 MPEG-4 international video coding standard and the China AVS-M standard. His lightweight encryption and error resilience algorithms are accepted into the China AVS standard. He was Chair of Screen Content Coding AdHoc Group in JCTVC for HEVC. He has 25+ granted US patents and is applying for 70+ more on his signal processing techniques. He has performed forensic investigation and stood as an expert witness in Hong Kong courts many times.

Dr. Au is a Fellow of IEEE and HKIE. He is/was Associate/Senior Editors of several IEEE journals (TCSVT, TIP, TCAS1, SPL, JSTSP, SigView) and non-IEEE journals (JVCIR, JSPS, TSIP, JMM, JFI, and SWJ). He is guest editor of some special issues in JSTSP and TCSVT. He was BoG member and Vice President Technical Activity of APSIPA. He was Chair of 3 technical committees: IEEE CAS MSA TC, IEEE SPS MMSP TC, and APSIPA IVM TC. He was a member of 5 other TCs: IEEE CAS VSPC TC, DSP TC, IEEE SPS IVMSP TC, IFS TC, and IEEE ComSoc MMC TC. He served on 2 steering committees: IEEE TMM, and IEEE ICME. He also served on organizing committee of many conferences including ISCAS 1997, ICASSP 2003, ISO/IEC 71st MPEG in Jan 2005, ICIP 2010, etc. He was General Chair of several conferences: PCM 2007, ICME 2010, and PV 2010. He won 5 best paper awards: SiPS 2007, PCM 2007, MMSP 2012, ICIP 2013, and MMSP 2013. He was IEEE Distinguished Lecturer (DL) in 2009 and 2010, APSIPA DL in 2013 and 2014, and has been keynote speaker multiple times.