Super-resolution reconstruction of high dynamic range images in a perceptually uniform domain

Tomas Bengtsson
Tomas McKelvey
Irene Yu-Hua Gu

Super-resolution reconstruction of high dynamic range images in a perceptually uniform domain

Tomas Bengtsson
Tomas McKelvey
Irene Yu-Hua Gu
Chalmers University of Technology
Department of Signals and Systems
SE-412 96 Gothenburg, Sweden
E-mail: tomas.bengtsson@chalmers.se

Abstract. Super resolution is a signal processing method that utilizes information from multiple degraded images of the same scene in order to reconstruct an image with enhanced spatial resolution. The method is typically employed on similarly exposed pixel valued images, but it can be extended to differently exposed images with a high combined dynamic range. We propose a novel formulation of the joint super-resolution, high dynamic range image reconstruction problem, using an image domain in which the residual function of the inverse problem relates to the perception of the human visual system. Simulated results are presented, including a comparison with a conventional method, demonstrating that the proposed approach avoids some severe reconstruction artifacts. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.OE.52.10.102003]

1 Introduction

To obtain high visual quality in digital camera systems, a high spatial resolution is required. The most direct way to increase spatial resolution in a camera is to reduce the size of the sensor pixel elements or to increase the chip size. However, there are applications for which this may not be feasible. For example, to overcome sensor noise at low exposure levels, smaller pixels have to be exposed for a longer duration. Consequently, in terms of hardware, a high spatial resolution is in contradiction to having a high temporal resolution, which is a requirement for high-quality video systems. Additionally, a longer exposure duration means that frames become susceptible to motion blur effects. Super-resolution reconstruction (SRR), as an alternative, is a signal processing approach to enhance spatial resolution. Multiple low-resolution (LR) images, captured in a sequence, of the same original scene are used to construct one high-resolution (HR) image. A further introduction to super resolution is given in Sec. 1.3.

Camera sensors that use a single fixed-exposure duration for each captured image have dynamic range limitations. This is particularly the case for compact cameras that have a smaller pixel size. The limitation is experienced e.g., when capturing an image of an outdoor environment on a sunny day or an indoor scene with a window. Certain image areas will then be overexposed or underexposed, which has a large impact in terms of visual quality. The human visual system (HVS) is able to perceive detailed information in bright and dim areas simultaneously, and thus it is undesired of the camera system to impose this limitation. By taking two or more images of the same scene, using different exposure settings, a single high dynamic range (HDR) image can be created, which is free from saturation. Digital display devices, for their part, also have a low dynamic range (LDR) compared to real-world scenes. The dynamic range limitation for display devices, however, can practically be overcome by tonemapping HDR image information to a LDR image in

such a way that it, to the HVS, resembles the HDR image that it was created from.[1] Before proceeding to a review of relevant literature on HDR as well as SRR, the HVS and how light is interpreted is briefly introduced.

1.1 The Human Visual System

Understanding the main principles of the HVS is of crucial importance to researchers in image processing. When dealing with standard 8-bit pixel image formats, much is abstracted into how those images are coded. For example, in the standard red, green, and blue (RGB) color spaces, data is coded linearly with respect to the perception of brightness in the HVS. This implies that image operations or filters naturally work as expected. Also, storage, compression or filtering of image data in this domain makes sense, because of the property that pixel value errors have the same perceptual impact across the whole range of pixel values.

Light that is incident on the camera sensor is nonlinear with regard to the HVS in several respects. As will be presented, various camera components are designed to mimic the HVS in order to produce an output image, which is perceived to be of high fidelity. Consider a light spectrum observed by the human eye. The incident energy is registered by the photoreceptors of the eye, the three different types of cone cells (corresponding to, roughly, red, green and blue light), that are responsible for human color vision. The three cone types each have their respective spectral responses. Their combined spectral response can be described in terms of grayscale vision by a single function of wavelength, the luminous efficacy curve. For simplicity as well as to stay within the main scope of the paper, the discussion of the HVS is from here on reduced to grayscale vision. The luminous efficacy curve is a dimensionless function that tells what fraction of radiant power at any given wavelength contributes to the luminous properties of the HVS. It is nonzero

for wavelengths between, roughly, 380 to 700 nm, the so-called visible spectrum.[2]

The term for the incident power spectrum of light is irradiance. The mathematical inner product of irradiance with the luminous efficacy curve is denoted illuminance. As opposed to irradiance, a radiometric unit, illuminance is a term defined based on the HVS, and is thus a photometric unit. While illuminance is the light registered by the human photoreceptors, the cones, it is still nonlinear with respect to perceived brightness. Illuminance changes of an absolute measure have a much higher perceived impact at low illuminance levels as compared to at high illuminance levels. Perceived brightness is subjective. Based on tests of human vision on large sample sizes, standardized units such as lightness and luma are defined and used in color spaces for LDR image processing. Table 1 summarizes the terms presented in this section and how they relate to the HVS and digital cameras. An incident light spectrum, irradiance, on a camera sensor is filtered according to the same principle as that of the cones of the eye. Thus, light contributes differently, depending on its wavelength, to the amount of current that is excited from sensor elements. The term illuminance is used here also when referring to the sensor output due to the filtered light, even though the term is really defined for the HVS. An image domain that is an approximation of perceived brightness, i.e., which is approximately linear to perceived brightness, will henceforth be called a perceptually uniform (PU) domain. Examples of (grayscale) PU domains are lightness and luma; however, they are only designed for LDR images. Any PU domain will correspond to a concave nonlinear mapping from the illuminance domain.

1.2 High Dynamic Range Images

The HVS can adapt to illuminance levels of about 10 orders of magnitude, and at any given adaptation level about five orders of magnitudes are perceived. As mentioned, for a given exposure setting, camera sensors have a lower dynamic range than that of many natural scenes (and of the HVS), which will cause saturation of sensor elements, either from over- or underexposure. Thus, the sensors are said to have a LDR. In HDR image reconstruction, a single HDR image is constructed from two or more LDR images captured with different exposure settings. The LDR images must be aligned and should be captured in a quick sequence, such that local motion in the scene is kept to a minimum. The images are merged by weighted average in a PU image domain. Weights of saturated pixels are naturally set to zero. Given a set of

LDR images in the pixel value domain, their corresponding sensor exposures are retrieved by an inverse mapping with the camera response function (CRF). Illuminance domain images are subsequently obtained by scaling the images with the inverse of their respective exposure durations. Finally, the images are transformed to e.g., the log-illuminance domain, which is an approximation of lightness (because perceived brightness of the HVS is much closer to logarithmic than linear in illuminance), where the weighted average is taken.[3]

To reproduce an image of an HDR scene on a display device, e.g., a computer monitor, it has to be tonemapped to the lower dynamic range of the display device.[1,4] The tonemapping essentially compresses the dynamic range uniformly in a suitable domain, but more sophisticated tonemappers also do local processing. For example, image contrasts may be compressed more on edges between image segments than on details within image segments. Additionally, suitable color spaces are used in order to maintain high color fidelity. The output LDR image, typically stored in a standard 8-bit format such as *sRGB*, is perceived by the HVS to still contain HDR information. HDR applications are recently starting to reach the general public. For example, smartphone applications that take two images in a sequence and generate a merged HDR image have reached widespread use, and similar features are common in modern commercial cameras. The result is often satisfactory, although artifacts sometimes occur due to spatial misalignment.

1.3 Super-Resolution Reconstruction

The signal processing technique SRR makes use of multiple degraded LR images of the same original scene.[5,6] In order to be useful, each image must contain new information of the scene, which is the case if the images are shifted relative to each other on a subpixel scale (e.g., from natural camera movement) or if they are blurred by different blur functions. In the corresponding mathematical model, this means that the images each contribute with linearly independent observations of the desired HR image representation. To reconstruct an HR image, an inverse problem is solved, including image registration (alignment) with high precision as well as blur estimation, either as pre-processing or jointly with estimating the HR image. A property of SRR is that the resulting inverse problem is fundamentally ill-conditioned, which presents big challenges.[7,8] If a larger number of observations are used, the conditioning improves, but only slightly. Furthermore, because complex motion between frames is undesired, a relatively small number of images

Table 1 Description of light units in the context of the HVS and digital cameras.

Unit	Description	HVS	Camera
Irradiance (Radiometric)	Incident power spectrum $I(\lambda)$	Sensitive to light of wavelength $\lambda \in [380,700]$ nm	Filter $I(\lambda)$ according to approximated luminous efficacy
Illuminance (Photometric)	$\int_{-\infty}^{\infty} I(\lambda)V(\lambda)\mathrm{d}\lambda$, where $V(\lambda)$ is the luminous efficacy curve of the HVS	The power registered by the photoreceptors	Filtered light exposes the sensor to approximate illuminance of the HVS
Perceived brightness, e.g., Lightness, Luma (Photometric)	A nonlinear function of illuminance	A subjective measure	Standardized units such as lightness and luma are approximations of perceived brightness

is preferred so that they are close temporally. Thus, to improve the conditioning, a regularization term is often added to the problem formulation. The regularization function should enforce a solution that matches the statistics of the image, or rather of natural images in general, since the statistics of the unknown image are unavailable.

Bayesian interpretations, in which the unknown image is viewed as stochastic, of the SRR problem are common.[9,10] Similarly to deterministic methods, they lead to minimizing a (regularized) objective function. A prior is assumed on the HR image and often also on other system parameters that should be estimated. For example, fine-scale image registration can be cast as a stochastic problem, with a prior selected for the registration error that remains after a rough pre-processing registration. Blind super resolution (BSR) addresses the joint estimation of HR image and blur kernels (that include subpixel level shifts).[11] BSR has clear similarities with multichannel blind deconvolution,[12] with the added extension of downsampling. BSR is very challenging in terms of finding the global optimal point, due to its nonconvex structure. Often, simple parameterizations of the blur kernel are made, or it is assumed static to model only the blur effect of the sensor. Applications of SRR exist both for offline processing, such as astronomy and forensics, and for real-time processing, as for video streaming. Work on real-time implementations are increasingly common, including SRR implementations on custom hardware such as Field Programmable Gate Arrays. Custom sensors may also be used, that can e.g., increase pixel size for desired improvement of temporal resolution in situations with moving objects that are susceptible to motion blur.[13] The spatial resolution is then maintained by performing SRR on the frames with reduced resolution. Optical flow models are implemented to handle local object motion within the image frames.

1.4 Super-resolution Reconstruction of HDR Images

Traditionally, SRR is performed on similarly exposed LDR pixel valued images. However, during the past few years SRR has also been applied to HDR images.[14,15] The combined dynamic range of the reconstructed HR, HDR image has no restrictions, as the exposure durations of the input images can be selected freely. A good reconstruction result requires accurate image registration, both for spatial shifts and to photometrically align the images in an HDR image domain. Facing both geometric and photometric registration adds some extra challenges, as compared to only geometric registration.[16] Ambitious methods implement optical flow techniques to handle local spatial motion within the images.[17] If the input images to the HDR SRR are given as 8-bit pixel valued images, the CRF is needed in order to retrieve the illuminance information from the earlier stage of the camera processing chain. However, if the device was designed to perform HDR SRR, the illuminance information (of a higher bit depth) could be more readily available.

A common factor for the published work on SRR for HDR images is that the reconstruction takes place in the illuminance domain, which, as discussed in Sec. 1.1, is highly nonlinear to human perception.[18] This is opposed to the case of LDR SRR methods, where the reconstruction is performed in the PU pixel value domain. Because illuminance

errors at low illuminance levels have such a high perceived impact, reconstruction errors in dim image areas cause noticeable artifacts. Therefore, in this work, the HDR SRR is performed in a tonemapped, PU domain. The HDR illuminance information is thus mapped by a nonlinear tonemapping function, modifying the inverse SRR problem from a least squares (LS) problem to a nonlinear problem of equal dimension. Note that the choice of tonemapping function is nontrivial. They all aim to mimic the HVS but should merely be seen as approximations.[1,4] In the formulation of the proposed method, a simple tonemapping function is used, which nevertheless captures the main characteristics of the HVS well (but should not be the preferred choice for the application of visualizing HDR data). Methods for image registration or the estimation of other system parameters are outside the scope of this paper, the focus is on the formulation of the inverse problem used in the SRR. The remainder of this paper is organized as follows. Section 2 describes the camera model for the image data acquisition. In Sec. 3, a method for SRR on HDR images in a PU domain is proposed. Section 4 provides experimental results for the proposed method and a related illuminance domain approach. Finally, Sec. 5 concludes the paper.

2 Camera Model

When acquiring an image, the camera sensor is illuminated by a real-world scene for a given exposure duration. Let the image $\mathbf{X}$, of size $X_1 \times X_2$, denote an HR representation with dynamic range X_{max}/X_{min} of an original continuous scene, and let $\mathbf{x}$ be its corresponding vectorized representation of length $(X_1 X_2) \times 1 \triangleq n \times 1$. It is assumed that the camera provides a set of low-resolution and LDR images $\mathbf{y}_k$ according to the expression

$$\mathbf{y}_k = f[\Delta t_k (\mathbf{DC}_{\mathbf{H}_k} \mathbf{x} + \mathbf{n}_k)] + \mathbf{q}_k, \qquad k = 1, \dots, K. \qquad (1)$$

For each of the multiple observations, $\mathbf{C}_{\mathbf{H}_k}$, of size $n \times n$, performs 2-D convolution on the vectorized HR image $\mathbf{x}$. Its convolution kernel $\mathbf{H}_k$, of support $H_1 \times H_2$, represents blurring as well as planar small-scale spatial shifts (assuming that rough image registration can be performed as pre-processing) for $\mathbf{y}_k$ relative to a reference image (e.g., $\mathbf{y}_1$). The downsampling matrix $\mathbf{D}$, of size $(n/L^2) \times n$, decimates the spatial resolution a factor L in the x- and y-directions, and Δt_k is the exposure duration. Additive Gaussian noise $\mathbf{n}_k$ of variance σ_n^2 models the noise in the image sensor. In addition, by somewhat exaggerating the value of σ_n^2, other error sources can be contained in $\mathbf{n}_k$. The $\mathbf{y}_k$ images are quantized to 8-bit pixel values, with uniformly spaced quantization levels $Y \in \{0, \dots, 1\}$. This quantization effect is modeled by $\mathbf{q}_k$, while quantization in the A/D converter to a higher bit depth is considered to be part of $\mathbf{n}_k$.

The low-resolution exposure of the camera sensor is $\mathbf{e}_k = \Delta t_k (\mathbf{DC}_{\mathbf{H}_k} \mathbf{x} + \mathbf{n}_k)$. The $\mathbf{e}_k$ vector is, similar to $\mathbf{y}_k$, $\mathbf{n}_k$, and $\mathbf{q}_k$, of size $(n/L^2) \times 1$. At each pixel i, the exposure $[\mathbf{e}_k]_i$ is mapped by the pixelwise, nonlinear CRF, f, to a pixel value in the output image. A reasonable (simplified) model of a real CRF, adopted for the simulations in this paper, is

$$f(E) = \begin{cases} 0 & , E \leq E_{\min} \\ \left(\frac{E - E_{\min}}{E_{\max} - E_{\min}}\right)^{\gamma_{\mathrm{LDR}}} & , E_{\min} \leq E \leq E_{\max} \\ 1 & , E \geq E_{\max} \end{cases} \tag{2}$$

where $\gamma_{\mathrm{LDR}} < 1$ makes the curve concave in the unsaturated region. What causes the $\mathbf{y}_k$ images to be of LDR is the fixed operational range of exposure values of the CRF, $[E_{\min}, E_{\max}]$. Exposure values outside of this interval are clipped and can not be recovered (from that single image). The CRF is said to perform gamma compression on the exposure image, to an LDR, PU output domain.

2.1 Alternative Camera Models

It is interesting to note that the camera model used for LDR SRR on pixel domain images differs from the camera model for HDR SRR methods, without exception. The forward model assumed for LDR SRR is

$$\mathbf{y}_k = \mathbf{DC}_{\mathbf{H}_k} f(\Delta t \mathbf{x}) + \mathbf{n}_k, \qquad k = 1, \ldots, K, \tag{3}$$

for which the order of f and $\mathbf{DC}_{\mathbf{H}_k}$ is interchanged as compared to Eq. (1).[5] From the set of observations, $\{\mathbf{y}_k\}$, the desired pixel valued HR image $\mathbf{z} = f(\Delta t \mathbf{x})$ is to be reconstructed. The model in Eq. (3) is thus used in the formulation of the inverse problem, where it relates the unknown HR image to the actual observed LR data. The inconsistency in the choice of camera model gives rise to the question of what model, Eq. (1) or Eq. (3), is closer to the real physics of the camera. One of the two, LDR SRR or HDR SRR, has a larger model mismatch that comes into play when real-world data is considered. The choice of camera model is not investigated further in this paper.

3 Image Reconstruction in a Perceptually Uniform Domain

To reconstruct, or rather to find a good estimate of $\mathbf{x}$, an inverse problem should be formulated and solved. To do so, knowledge of the system parameters in $\mathbf{H}_k$, including blur functions and subpixel-precision relative motion between image frames, is necessary. Assuming these, as well as the CRF f, to be known or to have been estimated, a compact notation is introduced in the beginning of this section to aid the inverse problem formulations. Then, in the remainder of the section, three different HDR SRR formulations are described, one which has an objective function that is expressed in the illuminance domain, and the other two in which a PU domain is employed. For mathematical convenience, the impact of the quantization noise $\mathbf{q}_k$ will be neglected in the inverse problem formulation (quantization is nevertheless used when generating $\mathbf{y}_k$), as it typically has little impact on reconstruction results compared to other error sources, e.g., $\mathbf{n}_k$ (and in the more general case, to errors in system parameter estimates).

To align the observed images $\mathbf{y}_k$ photometrically, the corresponding illuminance domain images [of the same size $(n/L^2) \times 1$]

$$\mathbf{i}_k = g(\mathbf{y}_k)/\Delta t_k \tag{4}$$

are recovered by mapping each pixel value in $\mathbf{y}_k$ with the inverse (not in the strict sense, due to the clipping effect) CRF,

$$g(Y) = Y^{(1/\gamma_{\mathrm{LDR}})}(E_{\max} - E_{\min}) + E_{\min}, \tag{5}$$

followed by a scaling with $1/\Delta t_k$. To handle saturation in f, each observation $\mathbf{y}_k$ has a corresponding diagonal weight matrix $\mathbf{W}_k$, of size $(n/L^2) \times (n/L^2)$, which is zero for diagonal elements corresponding to saturated pixels in $\mathbf{y}_k$ and one for the remaining diagonal elements. Saturated pixels are those where $[\mathbf{e}_k]_i \notin [E_{\min}, E_{\max}]$. However, since only the $\mathbf{y}_k$ are at hand, saturated pixels are taken to be those where $[\mathbf{y}_k]_i$ is either 0 or 1, the lowest and the highest quantization levels respectively. Based on the camera model in Eq. (1),

$$\mathbf{W}_k \mathbf{i}_k = \mathbf{W}_k[g(\mathbf{y}_k)/\Delta t_k] =$$
$$= \mathbf{W}_k(\mathbf{DC}_{\mathbf{H}_k}\mathbf{x} + \mathbf{n}_k), \qquad k = 1, \ldots, K, \tag{6}$$

where the second equality holds if the quantization effect is neglected, which means that $g(f())$ is the identity function for the unsaturated region of f. By stacking the individual LR, LDR illuminance domain observations in $\mathbf{i} = [\mathbf{i}_1^T, \ldots, \mathbf{i}_K^T]^T$, the noise vector $\mathbf{n} = [\mathbf{n}_1^T, \ldots, \mathbf{n}_K^T]^T$, both of size $(nK/L^2) \times 1$, defining a block diagonal weight matrix $\mathbf{W} = \mathrm{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_K)$, of size $(nK/L^2) \times (nK/L^2)$, and $\mathscr{H} \triangleq [(\mathbf{DC}_{\mathbf{H}_1})^T, \ldots, (\mathbf{DC}_{\mathbf{H}_K})^T]^T$, of size $(nK/L^2) \times n \triangleq m \times n$, it follows that

$$\mathbf{Wi} = \mathbf{W}(\mathscr{H}\mathbf{x} + \mathbf{n}). \tag{7}$$

More generally, $\mathbf{W}$ could be designed to weigh observed data differently, as a function of e.g., its pixel value and exposure duration, according to the noise properties of Eq. (1).[14] In the image reconstruction, the objective is to minimize the difference, in some sense, between $\mathbf{W}\mathscr{H}\mathbf{x}$ and the faithfully (unsaturated) observed information $\mathbf{Wi}$. Thus, a distance measure of the residual function $\mathbf{r}(\mathbf{x})$ should be minimized in order to obtain

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \rho[\mathbf{r}(\mathbf{x})] = \sum_i h[r_i(\mathbf{x})], \qquad h_i[r_i(\mathbf{x})] \geq 0, \ \forall i, \tag{8}$$

where ρ is a nonnegative, norm-like function. The objective functions to be minimized in this paper all have the structure above. The rank of $\mathbf{W}\mathscr{H}$ depends on the number of observations K, and specifically the unsaturated regions of the images $\mathbf{y}_k$, as well as the selection of L. For example, if there is no saturation (which however implies that $\mathbf{x}$ cannot be an HDR image) and $L = 2$, $K = L^2 = 4$ images or more implies that $\mathbf{W}\mathscr{H}$ is full rank. However, due to the structure of $\mathbf{W}\mathscr{H}$, or more precisely due to the blur functions $\mathbf{H}_k$ contained within, the inverse problem is inherently ill-conditioned. Thus, even for a full-rank problem, the solution is very sensitive to noise. By adding a regularization term of a certain weight, the condition number of the overall problem improves, at the cost of removing fidelity from the data term. The remainder of this section presents three specific formulations of the general inverse problem in Eq. (8). These are then evaluated in Sec. 4.

3.1 The Least Squares Solution

Under the assumption of additive Gaussian noise $\mathbf{n}$, the LS estimate of $\mathbf{x}$ is optimal in the maximum likelihood sense. However, the corresponding inverse problem is equivalent

to minimizing the sum of squared reconstruction errors in the illuminance domain, a domain that is nonlinearly related to perceived error by the HVS, yet it is still used for SRR of HDR images.[14,15,17] The LS estimate is included mainly as a benchmark for the proposed method. In order to improve the condition number of the inverse problem and provide a stable solution, a regularization term is appended. Thus, the LS estimate is

$$\hat{x}_{\mathrm{LS}} = \arg\min_x \left\| \begin{bmatrix} \mathbf{W}\mathscr{H} \\ \sqrt{\lambda}\Gamma \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{Wi} \\ \mathbf{0} \end{bmatrix} \right\|_2^2 \triangleq \arg\min_{\mathbf{x}} \left\| \mathbf{r}_{\mathrm{LS}}(\mathbf{x}) \right\|_2^2, \tag{9}$$

for some linear regularization matrix Γ of size $n \times n$ with regularization weight parameter λ. Because images are typically piecewise smooth, Γ is selected as a matrix that performs 2D convolution on the vectorized image $\mathbf{x}$ with a 3×3 Laplacian kernel, that enforces a smooth solution by penalizing the 2nd order x- and y-derivative. The residual vector here is denoted $\mathbf{r}_{\mathrm{LS}}(\mathbf{x})$. In the above case, the objective function is

$$\mathcal{O}_1 = \|\mathbf{r}_{\mathrm{LS}}(\mathbf{x})\|_2^2 = \sum_{i=1}^{m+n} r_{\mathrm{LS},i}^2(\mathbf{x}), \tag{10}$$

where $r_{\mathrm{LS},i}(\mathbf{x})$ denotes the i'th component function in $\mathbf{r}_{\mathrm{LS}}(\mathbf{x})$. Because the system of equations is very sparse, an analytic solution is computationally feasible for a fairly large size of the problem.

3.2 The Proposed Objective Function

For the proposed objective function, the residual function $\mathbf{r}(\mathbf{x})$ is formulated in such a way that errors perceived by the HVS are penalized uniformly in the reconstruction. Thus, for the estimate of $\mathbf{x}$,

$$\hat{x}_{\mathrm{PU}} = \arg\min_x \left\| \begin{matrix} W[\tilde{f}(\mathscr{H}x) - \tilde{f}(i)] \\ \sqrt{\lambda}\Gamma\tilde{f}(\mathbf{x}) \end{matrix} \right\|_2^2 \triangleq \arg\min_{\mathbf{x}} \|\mathbf{r}_{\mathrm{PU}}(\mathbf{x})\|_2^2, \tag{11}$$

reconstruction errors are measured in a PU domain. Similar to the CRF, f, discussed in Sec. 2, f here is taken to be a power law function but with a different exponent, γ_{HDR}. Because the method that is used to generate an HDR ground truth image (from a set of LDR images) only provides relative illuminance information, and not the true absolute values,[3] the illuminance information, $\mathbf{i}$, is normalized before applying f. The same notation is kept. The normalized illuminance information is then tonemapped to a PU domain by $f(\cdot) = (\cdot)^{\gamma_{\mathrm{HDR}}}$. In this work, no special care is taken to customize the value of γ_{HDR} depending on the dynamic range of $\mathbf{i}$. It should be noted that, to instead minimize $f[\mathbf{W}(\mathscr{H}\mathbf{x} - \mathbf{i})]$, i.e., the tonemapped difference, would be incorrect, since the absolute illuminance level determines how sensitive the perception of the HVS is. The objective function of Eq. (11) is

$$\mathcal{O}_2 = \|\mathbf{r}_{\mathrm{PU}}(\mathbf{x})\|_2^2 = \sum_{i=1}^{m+n} r_{\mathrm{PU},i}^2(\mathbf{x}), \tag{12}$$

where $r_{\mathrm{PU},i}(\mathbf{x})$ denotes the i'th component function in $\mathbf{r}_{\mathrm{PU}}(\mathbf{x})$.

To solve the nonlinear minimization problem in Eq. (11), an iterative nonlinear programming method must be employed. An initial estimate, $\mathbf{x}^{(0)}$, may be taken as an interpolated version of an $\mathbf{i}_k$ image. Furthermore, let $\mathbf{x}^{(k)}$ be an estimate of $\mathbf{x}$ at iteration k, then the estimate is updated according to

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{d}^{(k)}, \tag{13}$$

where $\mathbf{d}^{(k)}$ is the search direction and $\alpha^{(k)}$ is the step length at iteration k. The search direction is of great importance with regard to the convergence rate. The 1st order gradient descent method is typically very slow to converge for the SRR problem. Because of that, higher order information is considered in order to determine $\mathbf{d}^{(k)}$. A 2nd order Taylor approximation of $\mathcal{O} = \rho[\mathbf{r}(\mathbf{x})]$ in the neighborhood of $\mathbf{x}^{(k)}$, as used in Newton methods, is

$$\mathcal{O}(\mathbf{x}^{(k)} + \mathbf{d}^{(k)}) \approx \mathcal{O}(\mathbf{x}^{(k)}) + \mathbf{G}^T(\mathbf{x}^{(k)})\mathbf{d}^{(k)} + \mathbf{d}^{(k)T}\mathbf{H}(\mathbf{x}^{(k)})\mathbf{d}^{(k)}, \tag{14}$$

where $\mathbf{G}$ is the gradient vector and $\mathbf{H}$ is the Hessian matrix of $\mathcal{O}$. For the specific case of a (possibly) nonlinear LS objective function, as for $\mathcal{O}_2$ (and $\mathcal{O}_1$), the gradient and Hessian can be written in terms of $\mathbf{r}$ as

$$\mathbf{G}(\mathbf{x}^{(k)}) = [\nabla_x \mathcal{O}(\mathbf{x})]|_{\mathbf{x}^{(k)}} = 2\sum_{i=1}^{m+n} r_i(\mathbf{x})\nabla_x r_i(\mathbf{x})$$
$$= 2\mathbf{J}_\mathbf{r}^T(\mathbf{x}^{(k)})\mathbf{r}(\mathbf{x}^{(k)}), \tag{15}$$

and

$$\mathbf{H}(\mathbf{x}^{(k)}) = [\nabla_x^2 \mathcal{O}(\mathbf{x})]|_{\mathbf{x}^{(k)}}$$
$$= 2\sum_{i=1}^{m+n} [\nabla_x r_i(\mathbf{x})\nabla_x^T r_i(\mathbf{x}) + r_i(\mathbf{x})\nabla_x^2 r_i(\mathbf{x})]|_{\mathbf{x}^{(k)}}$$
$$= 2\mathbf{J}_\mathbf{r}^T(\mathbf{x}^{(k)})\mathbf{J}_\mathbf{r}(\mathbf{x}^{(k)}) + \mathbf{S}(\mathbf{x}^{(k)}). \tag{16}$$

Note that $\mathcal{O}(\mathbf{x})$, $\mathbf{r}(\mathbf{x})$, $\mathbf{G}(\mathbf{x})$, and $\mathbf{H}(\mathbf{x})$ all depend on $\mathbf{x}$, and thus need to be re-evaluated for each update in the iterative optimization method. However, for compactness, this dependency will not always be emphasized by writing out the argument. The commonly used Gauss-Newton method fits nicely for the optimization problem in Eq. (11). To see this, consider the nonlinear residual function

$$\mathbf{r}_{\mathrm{PU}}(\mathbf{x}) = \begin{bmatrix} W_{11}\left[\left(\sum_{j=1}^n \mathscr{H}_{1j}x_j\right)^\gamma - i_1^\gamma\right] \\ \vdots \\ W_{mm}\left[\left(\sum_{j=1}^n \mathscr{H}_{mj}x_j\right)^\gamma - i_m^\gamma\right] \\ \sqrt{\lambda}\sum_{j=1}^n \Gamma_{1j}(x_j^\gamma) \\ \vdots \\ \sqrt{\lambda}\sum_{j=1}^n \Gamma_{nj}(x_j^\gamma) \end{bmatrix}, \tag{17}$$

which consists of a data term $\mathbf{r}_{\mathrm{PU},1:m}(\mathbf{x})$ and a regularization term $\mathbf{r}_{\mathrm{PU},m+1:m+n}(\mathbf{x})$. The data term elements have small residual values for $\mathbf{x}$ close to the true image. The regularization term is designed such that the true image should be penalized as little as possible, but in practice its contribution

to the residual is certainly not zero. Because the residual elements are small near $\mathbf{x}$, the Hessian $\mathbf{H}$ in Eq. (16) is well approximated by $\mathbf{H}_{\text{GN}} = 2\mathbf{J}_{\mathbf{r}}^T \mathbf{J}_{\mathbf{r}}$, dropping the term $\mathbf{S}$,

$$
\mathbf{J_r} =
\begin{bmatrix}
\frac{\partial r_1}{\partial x_1} & \cdots & \frac{\partial r_1}{\partial x_n} \\
\vdots & & \vdots \\
\frac{\partial r_p}{\partial x_1} & \cdots & \frac{\partial r_m}{\partial x_n} \\
\frac{\partial r_{m+1}}{\partial x_1} & \cdots & \frac{\partial r_{m+n}}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial r_{m+1}}{\partial x_1} & \cdots & \frac{\partial r_{m+n}}{\partial x_n}
\end{bmatrix}
=
\begin{bmatrix}
W_{11}\gamma\mathscr{H}_{11}\left(\sum_{j=1}^n \mathscr{H}_{1j}x_j\right)^{\gamma-1} & \cdots & W_{11}\gamma\mathscr{H}_{1n}\left(\sum_{j=1}^n \mathscr{H}_{1j}x_j\right)^{\gamma-1} \\
\vdots & & \vdots \\
W_{\text{mm}}\gamma\mathscr{H}_{m1}\left(\sum_{j=1}^n \mathscr{H}_{1j}x_j\right)^{\gamma-1} & \cdots & W_{\text{mm}}\gamma\mathscr{H}_{\text{mn}}\left(\sum_{j=1}^n \mathscr{H}_{1j}x_j\right)^{\gamma-1} \\
\sqrt{\lambda}\Gamma_{11}\gamma x_1^{\gamma-1} & \cdots & \sqrt{\lambda}\Gamma_{1n}\gamma x_n^{\gamma-1} \\
\vdots & \ddots & \vdots \\
\sqrt{\lambda}\Gamma_{n1}\gamma x_1^{\gamma-1} & \cdots & \sqrt{\lambda}\Gamma_{nn}\gamma x_n^{\gamma-1}
\end{bmatrix}
\tag{18}
$$

is computed. From here on, $\mathbf{J_r}$ refers to the Jacobian matrix of the residual function $\mathbf{r} = \mathbf{r}_{\text{PU}}$. By minimizing the Taylor expansion in Eq. (14) with regard to $\mathbf{d}^{(k)}$, and replacing $\mathbf{H}^{(k)}$ with $\mathbf{H}_{\text{GN}}^{(k)}$,

$$
\mathbf{d}_{\text{GN}}^{(k)} = -[\mathbf{J_r}^T(\mathbf{x}^{(k)})\mathbf{J_r}(\mathbf{x}^{(k)})]^{-1}[\mathbf{J_r}^T(\mathbf{x}^{(k)})\mathbf{r}(\mathbf{x}^{(k)})],
\tag{19}
$$

is obtained as the search direction to be used in the iterative minimization in Eq. (13). By taking the inner product of the gradient and the search direction $\mathbf{d}_{\text{GN}}$, it is seen that the GN search direction is indeed a descent direction. This is a nice property of the GN method, and it is a consequence of the positive definiteness of $\mathbf{H}_{\text{GN}}$. Standard GN does not include any step size $\alpha^{(k)}$ in its updates. However, while $\mathbf{H}_{\text{GN}} = 2\mathbf{J_r}^T \mathbf{J_r}$ is positive definite (by construction), $\mathbf{J_r}$ is only valid in a region around $\mathbf{x}^{(k)}$ for which it was evaluated, thus the update step may be taken too large. By including a line search method to determine the step size at each iteration, the robustness of the method is increased. The extended method is famously known as damped Gauss-Newton, and it is used here to minimize $\mathcal{O}_2$.

3.3 Reconstruction Using Robust Norm and Robust Regularization

So far, the methods of reconstructing $\mathbf{x}$ have involved minimizing an objective function that is the L2-norm of a residual vector. The LS solution is suitable when the residual, at the optimum, has a Gaussian distribution. This was the case for the (unsaturated) observations $\mathbf{i}$ in the LS formulation of $\mathbf{r}_{\text{LS}}$ in Eq. (9). The residual function $\mathbf{r}_{\text{PU}}(\mathbf{x})$ in Eq. (11), however, does not follow a Gaussian distribution. Thus, using another norm function could give a more suitable estimate. A second reason to consider other functions than the L2-norm is that any errors in the system parameters of $\mathscr{H}$ (which would be the case for a practical setup where system parameters themselves need to be estimated), as used in the inverse reconstruction model, give a model mismatch with the observation model.

As mentioned, a regularization term is usually added to avoid noise amplification and thus stabilizes the solution. Depending on the amount of noise, the weight parameter λ for the regularization term should be adjusted accordingly. The larger λ is, the more important it becomes that the solution preferred by the regularization term matches the true

which contains the second order derivatives of r_i. In order to obtain $\mathbf{d}_{\text{GN}}$, at each iteration

statistical properties of $\mathbf{x}$. For this purpose, large values of $\mathbf{r}_{m+1:m+n}(\mathbf{x})$ should be interpreted as edges, which are a natural part in images, and should not be penalized. The Lorentzian "norm" (not really a norm since it violates the triangle inequality),

$$
\rho_{\text{Lor}}(\mathbf{r}) = \sum_i h_{\text{Lor},i}(r_1) = \sum_i \log\left[1 + \tfrac{1}{2}\left(\tfrac{r_i}{T_i}\right)^2\right],
\tag{20}
$$

where $T_i > 0$ are constant parameters, has been introduced and showed promising results for SRR in which several different noise models were tested.[19] The T_i are interpreted as threshold values. For r_i smaller than T_i, $\rho_{\text{Lor}}(\cdot)$ acts as the L2-norm and penalizes residual components quadratically, while for r_i greater than T_i, $\rho_{\text{Lor}}(\cdot)$ acts as the L1-norm, making the reconstruction less sensitive to large residuals that correspond to outliers in the data term or edges in the regularization term. In this regard, the simple linear Tikhonov regularization with the Lorentzian norm achieves the same edge-preserving property, which is the motivation for the use of the bilateral total variation regularization.[6,19]

Using the same residual function $\mathbf{r} = \mathbf{r}_{\text{PU}}$ as for $\mathcal{O}_2$, combined with the Lorentzian norm, a third estimate of $\mathbf{x}$ is

$$
\hat{x}_{\text{Lor}} = \arg\min_x \rho_{\text{Lor}}[r_{\text{PU}}(x)].
\tag{21}
$$

The corresponding objective function is

$$
\mathcal{O}_3 = \rho_{\text{Lor}}[\mathbf{r}_{\text{PU}}(\mathbf{x})] = \left\{ \sum_{i=1}^m \log\left[1 + \tfrac{1}{2}\left(\frac{r_{\text{PU},i}(\mathbf{x})}{T_1}\right)^2\right] \right.
$$
$$
\left. + \sum_{i=m+1}^{m+n} \log\left[1 + \tfrac{1}{2}\left(\frac{r_{\text{PU},i}(\mathbf{x})}{T_2}\right)^2\right] \right\},
\tag{22}
$$

where T_1 and T_2 are the threshold values of the Lorentzian norm for the data term and the regularization term respectively. The nonlinear objective is minimized iteratively. The GN method used for Eq. (11) does not fit this problem. Instead, the update equation for the iterative minimization of Eq. (22) is

$$
\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{P}^{(k)}\mathbf{d}_{\text{Grad}}^{(k)},
\tag{23}
$$

where $\mathbf{P}^{(k)}$ is a linear transformation performed on the gradient vector

$$\mathbf{d}_{\text{Grad}}^{(k)} = \nabla_x \mathcal{O}_3(\mathbf{x}) = \sum_{i=1}^{m+n} h_i'(r_i) \nabla_x r_i(\mathbf{x}) = \mathbf{J}_\mathbf{r}^T(\mathbf{x})\psi(\mathbf{x}) \quad (24)$$

$$\psi = [h_{\text{Lor},1}'(r_1), \ldots, h_{\text{Lor},m+n}'(r_{m+n})]^T$$
$$= \left[\frac{2r_1}{2T_1^2 + r_1^2}, \ldots, \frac{2r_{m+n}}{2T_2^2 + r_{m+n}^2} \right]^T. \quad (25)$$

If $\mathbf{P}^{(k)}$ is the Hessian matrix

$$\mathbf{H}_{\mathcal{O}_3}(\mathbf{x}) = \nabla_x^2 \mathcal{O}_3(\mathbf{x}) = \sum_{i=1}^{m+n} \{ h_{\text{Lor},i}''(r_i) \nabla_x r_i(\mathbf{x}) \nabla_x^T r_i(\mathbf{x})$$
$$+ h_{\text{Lor},i}'(r_i) \nabla_x^2 r_i(\mathbf{x}) \}, \quad (26)$$

the iterative Eq. (23) update is the Newton method. However, $\mathbf{P}^{(k)}$ may be any linear transformation, as long as it gives a higher convergence rate as compared to that of the gradient

direction it is beneficial. It is preferred to be positive definite, as this guarantees to give a descent direction in the neighborhood of $\mathbf{x}^{(k)}$. The choice $\mathbf{P}^{(k)} = (\mathbf{J}_\mathbf{r}^T \mathbf{J}_\mathbf{r})^{-1}$ in Eq. (23) is empirically proven to give a high convergence rate for Eq. (21), and is thus used for the experiments.

4 Experimental Results and Discussion

In this section, the three SRR formulations presented in Sec. 3 are evaluated in terms of experimental results. To compare the quality of the estimates, $\hat{\mathbf{x}}_{\text{LS}}$, $\hat{x}_{\text{PU}}$, and $\hat{x}_{\text{Lor}}$, their respective tonemapped results are presented. This provides a subjective comparison. As objective measures, PSNR and MSSIM values are given for the tonemapped results relative to the tonemapped original image $\mathbf{x}$.[20] The tonemapping function in MATLAB is used to display all HDR images, as it gives more visually appealing images than when tonemapping with f.

Two original images are used in the presented experiments, Memorial Church and Mount Tam (Images are courtesy of Greg Ward, available at Ref. 21), displayed in Figs. 1(a) and 2(a), respectively. From each of these images, a set of
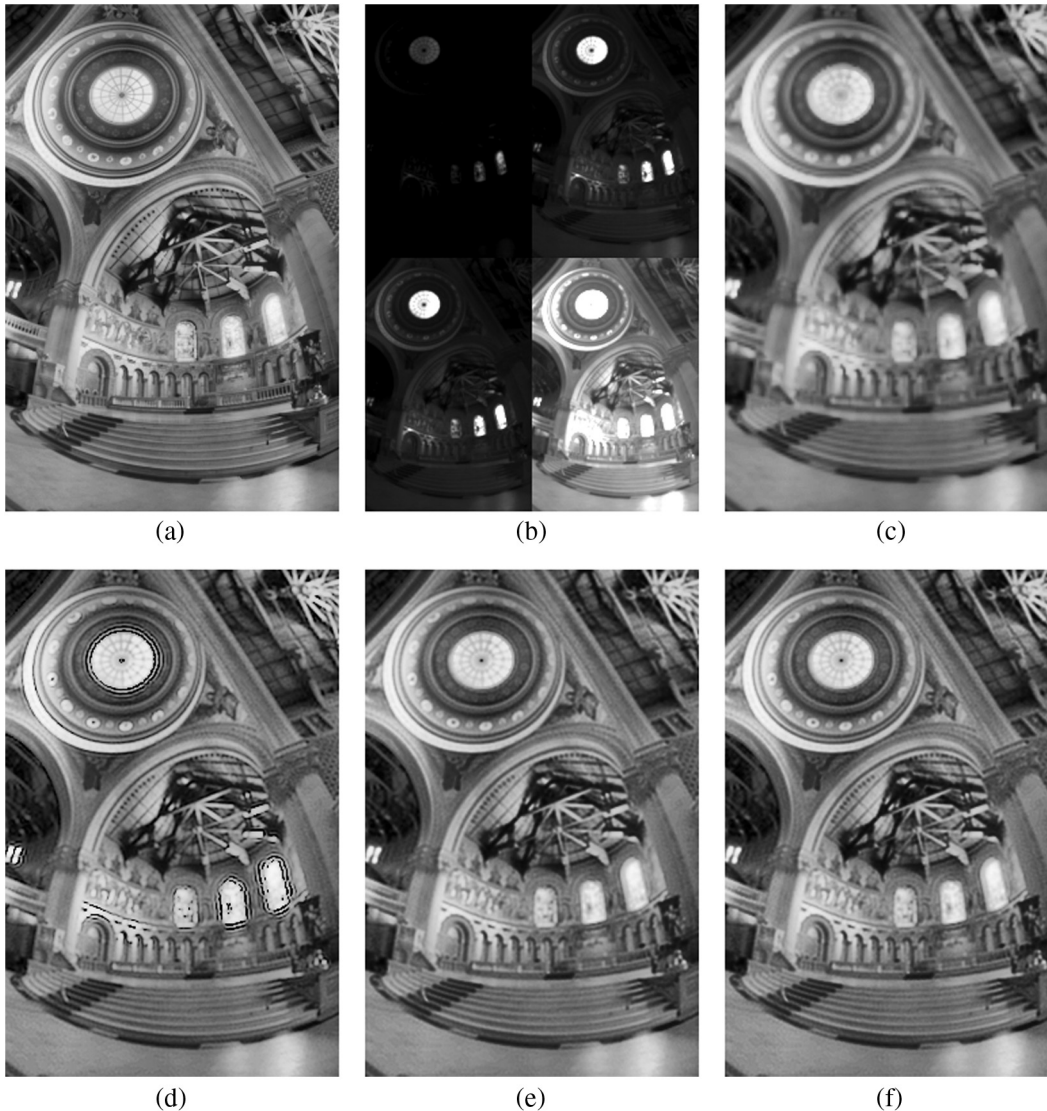

Fig. 1 Top row, from left to right: (a) Original image Memorial Church. (b) The observed images $\mathbf{y}_k$ (c) Bicubic interpolation of an LR, HDR version of Memorial Church. Bottom row: Reconstruction results using (d) $\mathcal{O}_1$, (e) $\mathcal{O}_2$, and (f) $\mathcal{O}_3$.
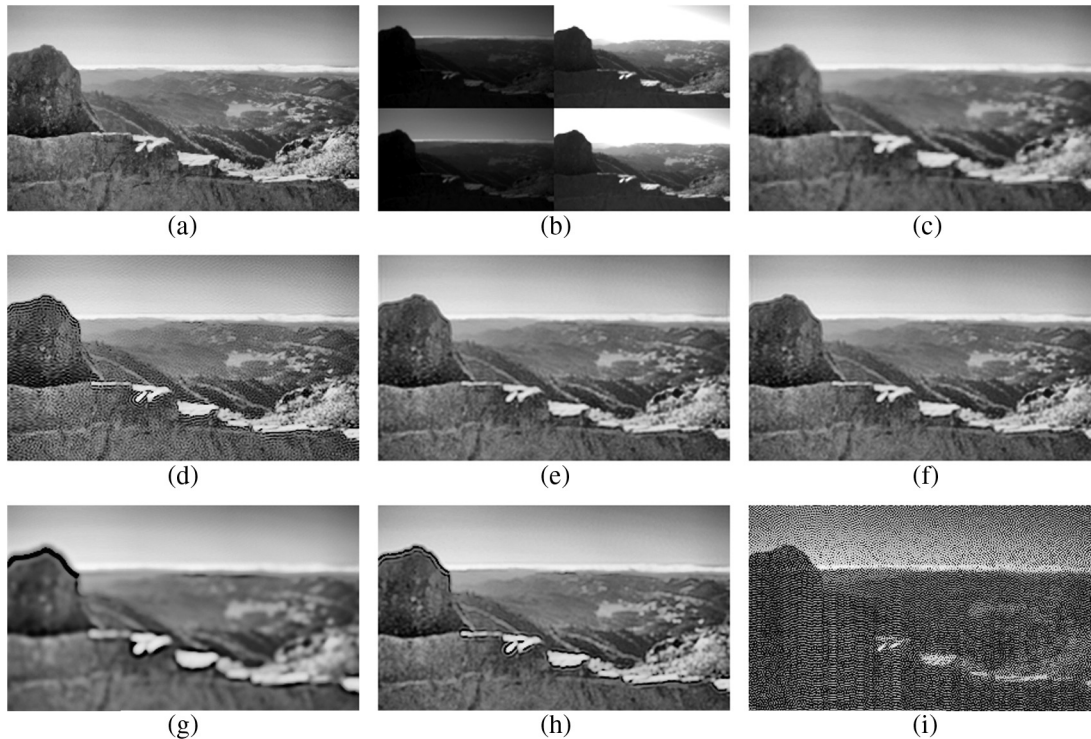
Fig. 2 Top row, from left to right: (a) Original image Mount Tam. (b) The observed images $\mathbf{y}_k$. (c) Bicubic interpolation of an LR, HDR version of Mount Tam. Middle row: Reconstruction results using (d) $\mathcal{O}_1$, (e) $\mathcal{O}_2$, and (f) $\mathcal{O}_3$. Bottom row: More reconstruction results using $\mathcal{O}_1$, with λ set to (g) 10, (h) 0.1, and (i) 10^{-6}.

semisynthetical LR, LDR observations are generated according to the model in Eq. (1). Parameters used for the CRF are $\gamma_{\text{LDR}} = 1/2.2$ and $[E_{\text{min}}, E_{\text{max}}] = [0.1, 100]$. The pixel values of $\mathbf{y}_k$ are quantized to 8 bits. The corresponding illuminance domain images $\mathbf{i}_k$, contained in $\mathbf{i}$, are obtained as in Eq. (4). For both experiments, $K = 4$ images are used to reconstruct HR images with a resolution enhancement factor of $L = 2$. This setup would give a full rank inverse problem if all images were free from saturation, which is however not the case here due to the original images being HDR. The differently exposed $\mathbf{y}_k$ images are shown in Fig. 1(b) as well as in Fig. 2(b), with saturation visible in all of them. The blur kernels used to generate $\mathbf{y}_k$, contained in the respective $\mathbf{H}_k$, all have a Gaussian shape with standard deviation of 1 and a support of 4×4 pixels. The kernels of $\mathbf{H}_2$, $\mathbf{H}_3$, and $\mathbf{H}_4$ are shifted spatially [0.5,0], [0,0.5], and [0.5,0.5] pixels, respectively, on the HR pixel grid relative to that of $\mathbf{H}_1$, such that the corresponding $\mathbf{y}_k$ is shifted accordingly, providing linearly independent observations of $\mathbf{x}$. Finally, $\sigma_n^2 = 0.01$ is the noise variance of $\mathbf{n}_k$. In experiment 1 on the Memorial Church image (of size 384×256), the exposure durations are $\Delta t = \{2^{-11}, 2^{-6}, 2^{-6}, 2^{-1}\}$. For experiment 2 on the Mount Tam image (of size 182×302), $\Delta t = \{2^{-5}, 2^{-5}, 2^{-1}, 2^{-1}\}$.

The pixelwise function $f(\cdot) = (\cdot)^{\gamma_{\text{HDR}}}$ is used for $\mathcal{O}_2$ and $\mathcal{O}_3$ in order to tonemap the normalized illuminance domain data, thus achieving a residual function $\mathbf{r}_{\text{PU}}$ that is PU with respect to the HVS. There is no established value for the exponent γ_{HDR} of f. Instead, it is selected here as $\gamma_{\text{HDR}} = 1/6$ based on empirical tests. Thus, it is lower than that of the LDR case, $\gamma_{\text{LDR}} = 1/2.2$. It is noted that the authors of the image appearance model iCAM, similarly, suggest that an exponent of $1/6$ is required in order to clearly perceive details in dim areas (that is, to make that data roughly

perceptually uniformly coded), when using a power law function such as f for visualization of an HDR image.[22] In their updated model, iCAM06, the HDR data is raised by an exponent of of $1/3$ at one step and later by an additional exponent of 0.7. Due to use of several color spaces and additional manipulation, however, no clear value for the total gamma is given.[1]

As a lowest acceptable performance limit for SRR methods, consider the interpolated images in Figs. 1(c) and 2(c). These have been obtained by first creating an LR, HDR image using perfectly registered LR, LDR images, one for each unique exposure duration, and the exact g function for photometric alignment, followed by bicubic interpolation in order to increase the resolution a factor $L = 2$. Figure 1(d)–1(f) shows the reconstructed images, $\hat{x}_{\text{LS}}$, $\hat{x}_{\text{PU}}$, and $\hat{x}_{\text{Lor}}$, obtained from minimizing $\mathcal{O}_1$, $\mathcal{O}_2$, and $\mathcal{O}_3$, for the Memorial Church image. The regularization weight parameter used is $\lambda = 0.02$, 0.03, and 0.02 respectively. Similarly, the results for the Mount Tam image are shown in Fig. 2(d)–2(f), all acquired with $\lambda = 0.05$. The threshold values for the Lorentzian norm in $\mathcal{O}_3$ are set to $T_1 = T_2 = 0.1$ for both experiments, selected to give the highest possible MSSIM value, the same criterion used as for selecting λ (parameter values were selected empirically). PSNR and MSSIM values for the tonemapped reconstruction results are given in Table 2. Since PSNR is a crude measure for image quality, emphasis should be on the MSSIM values. Edge artifacts are clearly visible in the dim regions of the LS illuminance domain estimate $\hat{x}_{\text{LS}}$. These artifacts are not present for the SRR results in the PU domain $f(\cdot)$. Thus, visual inspection is needed to complement the PSNR or MSSIM values that do not fully convey the results of the reconstruction. However, the proposed objective function $\mathcal{O}_2$ and its

Table 2 Experimental results. PSNR and MSSIM values are given for tonemapped SRR estimates $t(\hat{x})$ relative to the tonemapped original image $t(\mathbf{x})$, where $t(\cdot)$ denotes the MATLAB tonemapping function. The highest PSNR and MSSIM values for each of the original images are highlighted in bold.

Image	Method	Parameters	PSNR$(t(\hat{x}), t(\mathbf{x}))$	MSSIM$(t(\hat{x}), t(\mathbf{x}))$
Memorial Church	$\mathcal{O}_1$	$\lambda = 0.02$	35.659	0.8813
	$\mathcal{O}_2$	$\lambda = 0.03$	36.435	0.9259
	$\mathcal{O}_3$	$\lambda = 0.02$, $T_1 = 0.1$, $T_2 = 0.1$	**36.707**	**0.9275**
Mount Tam	$\mathcal{O}_1$	$\lambda = 0.05$	35.901	0.8771
	$\mathcal{O}_2$	$\lambda = 0.05$	**37.307**	0.9054
	$\mathcal{O}_3$	$\lambda = 0.05$, $T_1 = 0.1$, $T_2 = 0.1$	37.290	**0.9055**

alternated version $\mathcal{O}_3$ do result in higher objective quality measures as well. There is a slight improvement by using the Lorentzian norm of $\mathbf{r}_{\text{PU}}$ as in $\mathcal{O}_3$ rather than the L2-norm in $\mathcal{O}_2$, but it is minor. Simulations using imperfectly estimated $\mathbf{H}_k$ in the SRR do not seem to further enhance the performance difference of using the robust Lorentzian norm as compared to using the L2-norm; however, more rigorous investigations for this case would be interesting.

For the Mount Tam image, additional reconstruction results for the objective function $\mathcal{O}_1$ are shown in Fig. 2(g)–2(i), with $\lambda = 10$, 0.1 and 10^{-6} used respectively. It is seen that the increased value of λ from 0.05 to 0.1 gives a less noisy result compared to the LS reconstruction in Fig. 2(d); however, the edge artifacts are magnified. A too-high value for λ leads to oversmoothing, while a too-low value for λ results in severe noise amplification due to the poor condition number of the inverse problem. For the reconstruction result in Fig. 2(i) with $\lambda = 10^{-6}$, $K = 8$ images are used, four for each of the two different exposure durations, which means that the problem is full rank. Thus, it is clear that due to the ill-conditioned structure, regularization is necessary, which in turn gives numerical reconstruction errors particularly at image edges. These errors have a large perceptual impact for low-illuminance levels, as results show when performing SRR in the illuminance domain as for $\mathcal{O}_1$. For $\mathcal{O}_2$ and $\mathcal{O}_3$, as an example, setting λ for SRR of the Mount Tam image to 0.01 or 0.1 instead of the current choice of 0.05 does not cause much noticeable difference in the results. The SRR method presented in this paper could straightforwardly be applied to color images, by treating each RGB color channel individually. This gives satisfactory results, even though it is not optimal in terms of color fidelity. The RGB channels should then be coded jointly, such as for the LDR L*a*b color space.

5 Conclusions

This paper treats SRR of differently exposed LDR images. A formulation of the HDR SRR problem in which perceived difference is minimized in an appropriate, PU image domain has been proposed. As shown in experimental results, even small numerical reconstruction errors will have a large perceptual impact in dim image areas if the illuminance domain is used in the objective function, due to the nonlinearity of the illuminance domain to perceived brightness of the HVS.

This issue is overcome by using the proposed approach. Because the proposed residual function is non-Gaussian, an alternative to the L2-norm was suggested, which, however, only gave a minor improvement in the presented experiments.

References

1. J. Kuang, G. Johnson, and M. Fairchild, "icam06: a refined image appearance model for HDR image rendering," *J. Vis. Comun. Image Represent.* **18**(5), 406–414 (2007).
2. E. Allen and S. Triantaphillidou, *The Manual of Photography*, 10th edn., Focal Press, Oxford (2011).
3. P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. 24th Ann. Conf. Computer Graphics and Interactive Techniques, SIGGRAPH '97*, pp. 369–378, ACM Press/Addison-Wesley Publishing Co., New York (1997).
4. M. Cadik et al., "Evaluation of HDR tone mapping methods using essential perceptual attributes," *Comput. Graph.* **32**(3), 330–349 (2008).
5. S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *Signal Process. Mag.* **20**(3), 21–36 (2003).
6. S. Farsiu et al., "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.* **13**(10), 1327–1344 (2004).
7. M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Process.* **6**(12), 1646–1658 (1997).
8. S. Farsiu et al., "Advances and challenges in super-resolution," *Int. J. Imaging Syst. Technol.* **14**(2), 47–57 (2004).
9. L. Pickup, "Bayesian methods for image super-resolution," *Comput. J.* **52**(1), 101–113 (2009).
10. S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian super resolution," *IEEE Trans. Image Process.* **20**(4), 984–999 (2011).
11. F. Sroubek, G. Cristobal, and J. Flusser, "A unified approach to super-resolution and multichannel blind deconvolution," *IEEE Trans. Image Process.* **16**(9), 2322–2332 (2007).
12. G. Harikumar and Y. Bresler, "Perfect blind restoration of images blurred by multiple filters: theory and efficient algorithms," *IEEE Trans. Image Process.* **8**(2), 202–219(1999).
13. M. Angelopoulou et al., "Robust real-time super-resolution on FPGA and an application to video enhancement," *ACM Trans. Reconfig. Technol. Syst.* **2**(4), 22:1–22:29 (2009).
14. J. Choi, M. K. Park, and M. G. Kang, "High dynamic range image reconstruction with spatial resolution enhancement," *Comput. J.* **52**(1), 114–125 (2009).
15. F. Schubert, K. Schertler, and K. Mikolajczyk, "A hands-on approach to high-dynamic-range and superresolution fusion," in *IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 1–8 (2009).
16. M. Gevrekci and B. Gunturk, "Superresolution under photometric diversity of images," *EURASIP J. Appl. Signal Process.* **2007**, 036076 (2007).
17. H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement," *Comput. Graph. Forum* **30**(2), 405–414 (2011).
18. T. Bengtsson et al., "Regularized optimization for joint super-resolution and high dynamic range image reconstruction in a perceptually uniform domain," in *Int. Conf. on Acoustics, Speech, and Signal Proc., ICASSP 2012*, pp. 1097–1100 (2012).

19. V. Patanavijit and S. Jitapunkul, "A Lorentzian stochastic estimation for a robust iterative multiframe super-resolution reconstruction with Lorentzian-Tikhonov regularization," *EURASIP J. Adv. Signal Process.* **2007**(2), 1–21 (2007).
20. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
21. G. Ward, "High dynamic range images," http://www.anyhere.com/gward/hdrenc/pages/originals.html.
22. M. Fairchild and G. Johnson, "iCAM framework for image appearance, differences, and quality," *J. Electron. Imaging* **13**(1), 126–138 (2004).

Tomas Bengtsson is a PhD student in the signal processing research group at the Chalmers University of Technology, in Gothenburg, Sweden. He obtained a BS degree in electrical engineering and an MS degree in communication engineering at Chalmers University of Technology between 2004 and 2009. His research lies in signal processing and sensing, focusing on image processing, in particular high dynamic range imaging and super-resolution techniques. He is involved in a project on visual quality measures, with the aim to explore algorithms for the enhanced display of images and videos with applications toward the automotive industry, and in particular car safety.

Tomas McKelvey obtained an MS degree in electrical engineering at Lund University between 1987 and 1991 and received his PhD in automatic control at Linköping University in 1995. Between 1995 and 1999, he held research and teaching positions at Linköping University and became docent in 1999. Between 1999 and 2000 he was a visiting researcher at University of Newcastle, Australia. Since 2000, he has been with Chalmers University of Technology, becoming a full professor in 2006 and the head of the signal processing group at Chalmers in 2011. Professor McKelvey's research interests are model based in statistical signal processing, system identification, and image processing with applications to biomedical engineering and combustion engines.
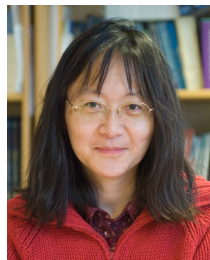
Irene Yu-Hua Gu received her PhD degree in electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1992. From 1992 to 1996, she was a research fellow at Philips Research Institute IPO, Eindhoven, The Netherlands, and did post-doctorate work at Staffordshire University, Staffordshire, United Kingdom. She was also a lecturer at the University of Birmingham, Birmingham, United Kingdom. Since 1996, she has been with the Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden, where she has been a full professor since 2004. Her research interests include statistical image and video processing, object tracking and video surveillance, pattern classification, and signal processing with applications to electric power systems. She has been an associate editor with the EURASIP *Journal on Advances in Signal Processing* since 2005, and on the editorial board for the *Journal on Ambient Intelligence and Smart Environments (JAISE)* since 2011.