REGULARIZED OPTIMIZATION FOR JOINT SUPER-RESOLUTION AND HIGH DYNAMIC RANGE IMAGE RECONSTRUCTION IN A PERCEPTUALLY UNIFORM DOMAIN

Tomas Bengtsson[⋆] *Irene Yu-Hua Gu*[⋆] *Mats Viberg*[⋆] *Konstantin Lindström*[†]

[⋆] Department of Signals and Systems, Chalmers University of Technology, Sweden
[†] Volvo Cars AB, Sweden

ABSTRACT

This paper discusses resolution enhancement of a set of images with varying exposure durations, having a high combined dynamic range. So far, little has been said in relation to the Human Visual System when it comes to Super-Resolution and High Dynamic Range fusion, unlike the case for traditional Super-Resolution where errors are measured with respect to human perception in the pixel domain. We propose a Super-Resolution method in the L*a*b* domain to bridge that gap and present some image reconstruction results.

Index Terms— Super-Resolution, Dynamic Range, Image Reconstruction, Regularization, Human Visual System

1. INTRODUCTION

For good visual quality in digital images it is desirable that the image has a high spatial resolution as well as a high dynamic range of light intensities, the latter to avoid clipping at over- or underexposed image areas. Applications range from artistic photography to astronomy and tracking.

Image resolution may be enhanced by Super-Resolution (SR) reconstruction techniques as an alternative way of using more expensive cameras. SR image reconstruction utilizes multiple Low Resolution (LR) images that are degraded by blur and noise and slightly shifted relative to each other, see [1] for a good overview of SR techniques or the example software in [2]. Several SR methods sequentially estimate blur and motion before the SR Reconstruction (SRR) of the unknown High Resolution (HR) image. More sophisticated methods use Blind Super-Resolution (Blind Deconvolution after downsampling) where blur, motion and HR image are estimated simultaneously [3, 4].

High Dynamic Range (HDR) images are typically created from multiple Low Dynamic Range (LDR) images with varied exposure durations [5]. They contain more detail than a single image, that is inherently bound to be LDR due to sensor dynamic range limitations. To be visualized on a monitor, a HDR image needs to be processed by a so called tone-mapping operator (TMO) [6], that transforms an HDR image to an LDR image while trying to maintain the same perceived appearance for the Human Visual System (HVS). The Image Appearance Framework in [7] provides a TMO that makes an ambitious attempt to model the HVS. It is not until recently that HDR modeling has been in focus. Whereas color spaces for e.g. perceptual uniformity of brightness is based on large studies for LDR imagery, studies on extended luminance (HDR) levels for perceptual uniformity [8] and color constancy [9] are only from the recent years.

Few attempts have been made of combining SR and HDR imaging in the past [10, 11, 12, 13]. By performing the SRR in the illuminance domain (photometric exposure divided by exposure duration), before the camera maps the exposure to pixel values, it is straight forward to allow varying exposure durations. However, opposite to the pixel domain (sRGB, where standard SRR is performed), the illuminance domain is not perceptually uniform (PU) to brightness in the HVS. Thus, errors in illuminance will not be weighted according their to perceptual severity. Still, [10, 11, 12, 13] all run their SRR algorithms in the illuminance domain, where artifacts easily arise, as discussed further in section 2.2.

Recently, [10, 11] proposed methods for combining SR and HDR imaging. They extend traditional regularized SR to differently exposed images. The drawback is that they use a full set of LR, LDR images. A downsampling factor of 4 implies that 16 LR images (for each exposure duration) are used to recover a unique solution. This is unpractical in a real case, where we wish to limit the number of observations. In [12], the image to be reconstructed is segmented according to lightning conditions and only one exposure duration is used for SRR in each segment. Non-saturated information from images with other exposure durations are not used in the same image segment, thereby avoiding the problem of interlacing information from differently exposed images, but at the cost of throwing away valid, useful data.

Motivated by the above, we propose a novel method for joint SR and HDR image reconstruction from a set of LR, LDR images. It extends [10, 11] to not necessarily having a full set of LR, LDR images. Furthermore, we transform our images to a domain that is more related to the HVS and draw parallels to traditional (LDR images) Super-Resolution. In section 2, we formulate the method, in section 3 we present some results and finally we conclude the paper in section 4.

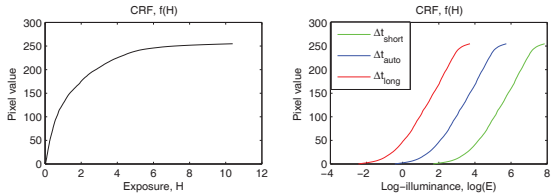

Fig. 1. Left: (a) CRF f (red color channel) as estimated from three differently exposed (real) LR, LDR images by the method in [5]. Right: (b) Total dynamic range of illuminances for 3 different exposure durations.

2. MATHEMATICAL FORMULATION

In this section we first describe the observation model for how the LR, LDR images are obtained. We then proceed with the model for joint SR and HDR image reconstruction.

2.1. Observation model

A real-world scene is observed through a set of LR and LDR images

$$\mathbf{y}_k = f(\mathbf{D}\mathbf{B}_k\mathbf{M}_k(\Delta t_k\mathbf{x}) + \mathbf{n}_k^a), \quad k = 1, .., p \qquad (1)$$

where $\mathbf{x}$ is a vector (for notational simplicity) of a desired HR image representing a continuous scene. The nonlinear Camera Response Function (CRF) f maps the sensor exposures $\mathbf{H} = \Delta t_k\mathbf{E}$, the product of exposure duration Δt_k and pixel-wise illuminance $\mathbf{E}$ to pixel values. An example CRF is shown in Fig. 1 (a). The images are shifted relative to a reference image $\mathbf{y}_{ref}$, as represented by the warping matrix $\mathbf{M}_k$, blurred by $\mathbf{B}_k$, downsampled by $\mathbf{D}$ a factor L in x- and y-dimension and corrupted by AWGN $\mathbf{n}_k^a$. Together, $\mathbf{M}_k$, $\mathbf{B}_k$ and $\mathbf{D}$ map a small region of HR pixels from $\mathbf{x}$ to a single LR pixel in $\mathbf{y}_k$. Contrary to traditional Super-Resolution, the observed images may be differently exposed, each with an exposure duration Δt_k, giving a higher combined dynamic range as shown in Fig. 1 (b). A common assumption for $\mathbf{M}_k$ is to only allow planar motion, e.g. global translational and rotational motion as in [14].

2.2. Model for Joint SR and HDR image reconstruction

The objective of the SRR is to reconstruct an estimate $\hat{\mathbf{x}}$ of the HR, HDR illuminance image $\mathbf{x}$ observed through the LR, LDR images $\mathbf{y}_k$. We propose to consider the properties of the HVS and take $\hat{\mathbf{x}}$ as the inverse $\tilde{f}^{-1}(\hat{\mathbf{z}})$ of the minimizer of the regularized weighted 2-norm of the pixel differences after pixel-wise transformation $\hat{\mathbf{z}} = \tilde{f}(\hat{\mathbf{x}})$ from the RGB illuminance domain to a domain that is closer to being linear to the perceptual sensation of the HVS. We minimize the cost function

$$C(\hat{\mathbf{z}}) = \sum_{k=1}^{p} |(\mathbf{D}\hat{\mathbf{B}}_k\hat{\mathbf{M}}_k\hat{\mathbf{z}} - \tilde{f}(g(\mathbf{y}_k)/\Delta t_k))|^2_{W(\mathbf{y}_k)} + \\ + \lambda\Psi(\hat{\mathbf{z}}), \qquad (2)$$

where $g(\mathbf{y}_k) \approx f^{-1}(\mathbf{y}_k)$ (approximate inverse, due to quantization, f being many-to-one) transforms the observations $\mathbf{y}_k(k = 1, .., p)$ in (1) to exposure values and thereafter to the illuminance domain by dividing with Δt_k, under the assumption that the camera-dependent response function f is known or estimated in advance. $\tilde{f}$ transforms the RGB illuminance images $g(\mathbf{y}_k)/\Delta t_k$ and $\hat{\mathbf{x}}$ in (2) to the L*a*b* color space [15], where the L*-channel of L*a*b* approximates lightness, a term for subjective perceived brightness (a grayscale measure) of the HVS. The a*- and b*-channels are designed to give good color consistency through color appearance uniformity and, importantly, to be perceptually orthogonal to L*. Notice that we have taken the operators on $\hat{\mathbf{x}}$, $\mathbf{D}\hat{\mathbf{B}}_k\hat{\mathbf{M}}_k$, outside of $\tilde{f}$. This is analogue to the traditional LDR SR approach, with $\tilde{f}$ being f in that case and Δt kept fixed.

The regularization function $\Psi(\hat{\mathbf{z}})$ in (2) is needed to limit the solution space of the typically underdetermined SRR. For images, it should typically enforce smoothing with preservation of strong edges[1] [16]. In [13], an anisotropic regularization function that does not smooth in edge gradient directions is attempted. Estimates $\hat{\mathbf{M}}_k$ and $\hat{\mathbf{B}}_k$ are obtained from pre-processing. The relative shifts $\mathbf{M}_k$ between the LR images may be estimated e.g. using [14] (assuming global planar motion) on mutually non-saturated image parts. The $\hat{\mathbf{B}}_k$ is taken as a 2-d Gaussian with variance σ^2.

Gradient Descent is used to solve (2), iterating

$$\hat{\mathbf{z}}^{(n)} = \hat{\mathbf{z}}^{(n-1)} - \beta\,\nabla C(\hat{\mathbf{z}}^{(n-1)}), \qquad (3)$$

where

$$\nabla C(\hat{\mathbf{z}}) = \sum_{k=1}^{p} \hat{\mathbf{M}}_k^T\hat{\mathbf{B}}_k^T\mathbf{D}^T\mathbf{W}(\mathbf{y}_k)(\mathbf{D}\hat{\mathbf{B}}_k\hat{\mathbf{M}}_k\hat{\mathbf{z}} - \\ - \tilde{f}(g(\mathbf{y}_k)/\Delta t_k)) + \lambda\nabla\Psi(\hat{\mathbf{z}}), \qquad (4)$$

until convergence. The step size β can be set adaptively as in [10]. In general, if the inverse SRR problem is solved using $p = L^2$ non-saturated observations, such that a unique solution exists, $\mathbf{x}$ is recovered exactly for $\mathbf{n}_k^a = 0$. If however the SRR problem is underdetermined, which is the usual case, artifacts easily arise near image edges. These artifacts are a result of the smoothness enforced on the solution, and the inability of edge-preserving regularization functions to fully overcome smoothing of strong edges. For HDR SRR in the illuminance domain, with $C(\hat{\mathbf{x}}) = \sum_{k=1}^{p} |(\mathbf{D}\hat{\mathbf{B}}_k\hat{\mathbf{M}}_k\hat{\mathbf{x}} - g(\mathbf{y}_k)/\Delta t_k|^2_W + \lambda\Psi(\hat{\mathbf{x}})$ as the cost function, prominent edge artifacts would frequently occur at low illuminance levels, due to the large perceptual impact in the HVS (and thus in any suitably designed TMO) of small errors in the region of low illuminance values. On the contrary, the artifacts are less visible for SRR in the L*a*b* domain, thanks to the transformation of the channels to perceptual uniformity.

[1] Edge-preserving filtering in Super-Resolution reconstruction is not as simple to implement as when just filtering a given image. This is evident from experiments and e.g. the results in [16] where "edge-preserving" regularization methods perform worse than pure smoothing regularization.

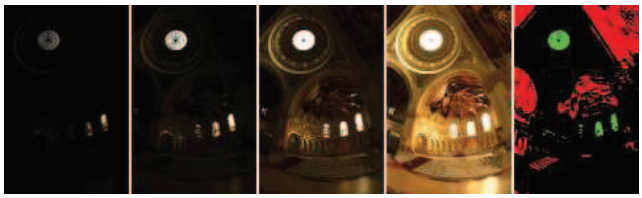

Fig. 2. Left: 4 different exposures to cover the full dynamic range of the scene in Fig. 4 (d). Selection of optimal exposure duration is outside the scope of this paper. Right (5th image): Underexposed (red) and overexposed areas (green) for the image in (c). These areas have weight zero in the SRR.

Fig. 3. Top-left: Zoom-in on the tone-mapped SRR result in Fig. 4 (c), from reconstruction in the proposed L*a*b* domain. Top-right: Reconstruction of the same original image but in the illuminance domain. Note the edge artifacts. Bottom: SSIM quality maps generated by the algorithm in [17] between the full versions of the above images and the original scene in 4 (a).

3. EXPERIMENTAL RESULTS

In this section, we present some HR, HDR reconstruction results using semi-synthetically generated observations. Given an original illuminance image $\mathbf{x}$, a set of observations $\mathbf{y}_k$ are generated by (1), using a simple but realistic power law gamma encoding function $f(H) = H^{\gamma}, \gamma = 1/2.2$ after normalization of the non-saturated exposure to $H \in [0, 1]$. The dynamic range was set to $g(255)/g(0) = 10/0.01 = 10^3$, similar to what [5] gives for the CRF estimate in Fig. 1 (a). $\mathbf{B}_k$ is set to have a Gaussian kernel with $\sigma = 0.75$ and the downsampling factor is $L = 4$. The noise $\mathbf{n}_k^a$ is assumed to be negligible and is set to zero. It should be set with regard to the scene illuminance, for which [5] provides no absolute, only relative, values. Two original HR, HDR images, Belgian House and Memorial Church [18] in Fig. 4 (a) and (d) respectively, are used for generating sets of LR, LDR observations. For Belgian House, $\Delta t_k \in \{1/250, 1/30, 1/4\}$ and for Memorial Church $\Delta t_k \in \{1/32, 1/4, 2, 16\}$. Fig. 2 shows examples of observed images $\mathbf{y}_k$ for the 4 different exposure durations for Memorial Church. Fig. 3 demonstrates the edge artifacts discussed in section 2.2. Artifacts are clearly visible for SRR in the illuminance domain, but not for SRR in the proposed L*a*b* domain using (2).

Fig. 4 (b), (c) and (e) show tone-mapped SRR results, using the propsed method, from the observation sets. In (b), 16 LR images per exposure duration are used. In (c) and (e), 7 LR images are used for each exposure duration. Fig 4 (f) provides a comparison with a result from [11] that uses a full set of LR, LDR images and a downsampling factor $L = 2$ (just to give some idea of the performance, as the full resolution of reconstructed images do not fit in this paper). $\mathbf{M}_k$ is assumed to have been estimated perfectly and $\mathbf{B}_k$ are assumed to be known in advance, thus $\hat{\mathbf{M}}_k = \mathbf{M}_k, \hat{\mathbf{B}}_k = \mathbf{B}_k$. A Tikhonov regularization $||\mathbf{\Gamma z}||^2$ is used for $\Psi(\mathbf{z})$ with a 3×3 Laplacian kernel ($\mathbf{\Gamma}$ a 2-d convolution matrix), as suggested by the results in [16]. All three L*a*b*-channels are normalized to [0,1] to allow simpler tuning of β and the regularization coefficient λ, here set empirically to be 0.3 and 0.1 respectively.

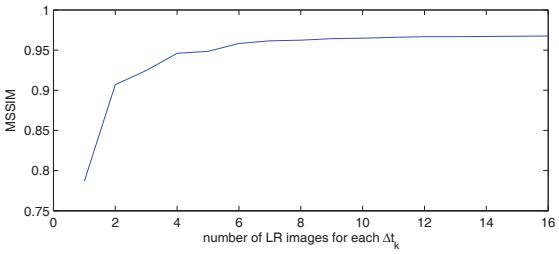

Fig. 5. MSSIM for tone-mapped reconstructions of the original image in Fig. 4 (d) as a function of number of LR images used per exposure duration. Each MSSIM value is an average of 20 simulations with randomly chosen LR warpings. Bicubic interpolation of one LR, HDR image has MSSIM 0.65.

The weighting matrix $\mathbf{W}(\mathbf{y}_k)$ should give higher weight to observations with low perceptual errors. Since each of the LDR images $\mathbf{y}_k$ are quantized in the PU sRGB domain specific for LDR images, the perceived quantization errors become a function of Δt_k after mapping to a domain of higher combined dynamic range. How to set $\mathbf{W}$ optimally is left out for now and we simply set weight 1 for non-saturated pixels and 0 for saturated pixels. This would be the case if SR HDR was implemented in a camera; quantization to only 8-bit channel depth would not be implemented for the LDR images.

Performance evaluation: Objective quality measures are less established for HDR images than for LDR images. We apply MSSIM scores [17] between the tone-mapped original HR, HDR image $\mathbf{x}$ in Fig. 4 (d) and reconstructed images $\hat{\mathbf{x}}$ as a function of the number of LR images used for each Δt_k, as presented in Fig. 3. It gives some indication that the proposed SRR method gives reasonable performance even for a relatively low number of observations.

4. CONCLUSION

We have presented a novel SR and HDR image reconstruction method in the Perceptually Uniform L*a*b* domain, where reconstruction errors are weighted by their perceptual severity. The results indicate that we can use a smaller set of LR, LDR observations and achieve comparable results to other methods using a full set of observations. Interesting future work includes to incorporate Blind Super-Resolution for HDR images to estimate unknown blur and subpixel motion as a combined convolution kernel, making the model more suitable for real data, and to use other methods to weigh errors with respect to the HVS.

Fig. 4. Upper row, left to right: (a) Original image Belgian House [18] with a dynamic range of 1.26e4. (b) SRR from 16 LR images for each of the 3 exposure durations. (c) SRR from 7 LR images for the 3 exposure durations. Bottom row, left to right: (d) Original image Memorial Church [18] with a dynamic range of 1.80e6. (e) SRR from 7 LR images for the 4 exposure durations. (f) Reconstruction result from [11] using a full set of LR, LDR images for a downsampling factor $L = 2$. (g, h) Zoomed-in window of images in (e) and (f) respectively, showing comparable results despite using a reduced set of images in (e). The big difference in brightness is a result of different tone-mapping operators used. The *tonemap* function in MATLAB was used for all of our results.

5. REFERENCES

[1] S. C. Park, M. K. Park, M. G. Kang, "Super-resolution image reconstruction: a technical overview", in *IEEE Signal Proc. Mag.*, 20 (3), pp. 21-36, Mar. 2003.

[2] P. Vandewalle, http://lcav.epfl.ch/software/superresolution/index.html

[3] F. Šroubek, J. Flusser, "A unified approach to superresolution and multichannel blind deconvolution", in *IEEE Trans. Image Proc.*, 16 (9), 2007.

[4] S. Harmeling, S. Sra, M. Hirsch and B. Scholkopf, "Multiframe blind deconvolution, super-resolution, and saturation correction via incremenental EM", in *Proc. Int. Conf. Image Processing*, pp. 3313 - 3316, 2010.

[5] P. Debevec, J. Malik, "Recovering high dynamic range radiance maps from photographs", in *SIGGRAPH 1997*, Aug. 1997.

[6] J. Kuang et al., "Evaluating HDR Rendering Algorithms", in *ACM Trans. Appl. Percept.*, 4 (2), July 2007.

[7] J. Kuang, G. Johnson, M. Fairchild, "iCAM06: A refined image appearance model for HDR image rendering", in *J. Vis. Commun. Image R.*, 18 (5), pp. 406-414, June 2007.

[8] T. O. Aydm, R. Mantiuk, H-P. Seidel, "Extending Quality Metrics to Full Luminance Range Images", in *Proc. SPIE*, Jan. 2008.

[9] M. Kim, T. Weyrich, J. Kautz, "Modeling Human Color Perception under Extended Luminance Levels", in *ACM Trans. Graph.*, 28 (3), 2009.

[10] B. K. Gunturk, M. Gevrecki, "High-Resolution Image Reconstruction From Multiple Differently Exposures Images", in *Signal Processing Letters, IEEE*, 13 (4), April 2006.

[11] J. Choi, M. K. Park and M. G. Kang, "High dynamic range image reconstruction with spatial resolution enhancement", in *The Computer Journal*, 52 (1), pp. 114-125, Jan. 2009.

[12] F. Schubert, K. Schertler and K. Mikolajczyk, "A hands-on approach to high-dynamic-range and superresolution fusion", in *Workshop on Applications of Computer Vision*, 2009.

[13] H. Zimmer, A. Bruhn, J. Weickert, "Freehand HDR Imaging of Moving Scenes with Simultaneous Resolution Enhancement", http://www.mia.uni-saarland.de/Publications/zimmer-pp286.pdf, 2010.

[14] P. Vandewalle, S. Süsstrunk and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution", in *EURASIP J. Applied Signal Proc.*, pp 233-233, Jan. 2006.

[15] E. Allen, S. Triantaphillidou, "The Manual of Photography", ISBN10 024052037, Ch. 5, p. 91, 2011.

[16] V. Patanavijit, S. Jitapunkul, "A Lorentzian stochastic estimation for a robust iterative multiframe super-resolution reconstruction with Lorentzian-tikhonov regularization", in *EURASIP J. Adv. SP*, 2007.

[17] Z. Wang et al., "Image Quality Assessment: From Error Visibility to Structural Similarity, in *IEEE Trans. Image Proc.*, 13 (4), April 2004.

[18] G. Ward, "High Dynamic Range Image Examples", http://www.anyhere.com/gward/hdrenc/pages/originals.html, 2003.