SUPER-RESOLUTION RECONSTRUCTION OF HIGH DYNAMIC RANGE IMAGES WITH PERCEPTUAL WEIGHTING OF ERRORS

Tomas Bengtsson⋆ *Tomas McKelvey*⋆ *Irene Yu-Hua Gu*⋆

⋆ Department of Signals and Systems, Chalmers University of Technology

ABSTRACT

Super-Resolution and High Dynamic Range image reconstruction are two different signal processing techniques that share in common that they utilize information from multiple observations of the same scene to enhance visual image quality. In this paper, both techniques are merged in a common model, and the focus is to solve the reconstruction problem in a suitable image domain, which relates to the perception of the Human Visual System. Simulated results are presented, including a comparison with a conventional method, demonstrating the benefits of the proposed approach, in this case avoiding some severe reconstruction artifacts.

Index Terms— Super-Resolution, dynamic range, image reconstruction, human visual system, regularization

1. INTRODUCTION

Digital camera devices have limited achievable performance when it comes to spatial resolution as well as dynamic range of its sensor elements. Particularly, this is the case for cheap devices such as web cameras. Other imaging systems, such as medical image acquisition devices, suffer from similar imperfections.

The most natural way to increase spatial resolution is to reduce the size of the sensor pixel elements. However, the smaller the hardware sensors are, the longer they need to be exposed, and thus they become more susceptible to sensor noise and motion blur [1]. Video systems have requirements on high temporal resolution, which is also in contradiction to high spatial resolution, in terms of hardware. Super-Resolution (SR) methods provide a signal processing approach to enhance spatial image resolution, which can help to reduce hardware cost [2, 3]. In SR Reconstruction (SRR), multiple degraded Low Resolution (LR) observations of the same original scene are used to construct a single High Resolution (HR) representation of that scene.

Over- or underexposure in an image due to insufficient dynamic range of the camera sensor can be addressed by High Dynamic Range (HDR) image reconstruction, which is a two-step procedure. In step 1, which concerns acquiring information, a series of images with different exposure settings are merged into a single HDR image [4]. In step 2, the visualization of information, the HDR image needs to be *tonemapped* before it can be displayed on e.g. a monitor [5, 6].

Traditionally, the SRR model assumes a set of similarly exposed Low Dynamic Range (LDR) pixel valued images, that are shifted relative to each other on a subpixel scale in the capturing process, blurred by their respective blur functions and downsampled. An inverse problem is solved, including image registration and blur estimation, to reconstruct the desired HR image. Bayesian interpretations of the SRR problem that include a noise model and priors on the blur and HR image are common [7, 8], including variational Bayesian methods [9], as well as deterministic interpretations that also lead to minimizing an (regularized) objective function [10, 11, 12]. In recent years, the problem of Blind Super-Resolution (BSR) [12] is often addressed, which is to estimate both the unknown HR image as well as the blur and subpixel level shifts jointly, rather than in a sequential manner. BSR has clear similarities with Multi-Channel Blind Deconvolution [13], with the added extension of downsampling.

An HDR image is constructed from two or more aligned and differently exposed LDR images. The input images could potentially be produced in one photograph by custom hardware where the individual elements of the pixel sensor array are exposed for different time durations. Given a set of LDR images in the pixel value domain, their corresponding sensor exposures are given by a mapping with the inverse Camera Response Function (CRF), and illuminance domain images are subsequently retrieved by scaling with the respective exposure duration of each image. The images are merged by weighted average in a perceptual domain into a single HDR illuminance image [4]. Saturated image areas are naturally left out from the reconstruction by giving them zero weight.

Both the SRR and the HDR image reconstruction methods can be merged into a joint image reconstruction framework. Instead of performing SRR in the LDR pixel value domain, as in the traditional case, it may just as well be performed in the illuminance domain. The combined dynamic range of the images has no restrictions, as the exposure durations can be selected freely. For good image reconstruction results however, input images need to be accurately registered both geometrically and photometrically. This is a bit more challenging when faced with both tasks at the same time, as discussed in [14], where they decide to first do geometric registration followed by photometric registration. In [15, 16], the image acquisition process is considered to be performed in a controlled environment, such that geometric registration is not necessary. However, to perform the HDR SRR they estimate the CRF in order to achieve photometric image alignment. Finally, in [17], geometric alignment is performed using an optical flow approach that handles local motion within images, while they assume the CRF to be known.

1.1. Contribution

All the previous work on joint HDR, SRR minimize an objective function in the illuminance domain (which several authors call the irradiance domain, neither term is strictly correct, see section 2.1 in [4]), in which reconstruction errors do not relate linearly to perceived error in the Human Visual System (HVS). Therefore, in our recent paper, we took a heuristic approach to weigh reconstruction errors by perceived severity [18]. The objective function was alternated in such a way that the minimizer was no longer the original image, an approach that however proved to break down under even a moderate image degradation process. In this work, the SRR is performed in a tonemapped, perceptually uniform, domain. The full HDR illuminance information is thus mapped by a nonlinear function, modifying the objective function from a Least Squares (LS) problem to a nonlinear LS problem of equal dimension. Note that the choice of tonemapping function is difficult. Tonemapping functions all aim to mimic the HVS but should merely be seen as approximations [5, 6].

The remainder of this paper is as follows. In Section 2, the camera model used is described. Section 3 outlines the image reconstruction, Section 4 presents some experimental results and finally, in Section 5 concluding remarks are given.

2. CAMERA MODEL

When acquiring an image, the camera sensor is illuminated by a real-world scene for an exposure duration Δt. Let $\mathbf{X}$ of size $X_1 \times X_2$ denote the desired HR, HDR representation of an original scene, and let $\mathbf{x}$ be its $(X_1X_2)\times 1$ vector representation. Then, the input to the HDR SRR, where the objective is to reconstruct, or rather to estimate as good as possible the image $\mathbf{x}$, is a set

$$\mathbf{y}_k = f(\Delta t_k(\mathbf{D}\mathbf{C}_{\mathbf{H}_k}\mathbf{x} + \mathbf{n}_k)). \quad k = 1, \dots, K \tag{1}$$

of degraded observations of $\mathbf{x}$. For each of the multiple observations, $\mathbf{C}_{\mathbf{H}_k}$ performs 2d convolution on the vectorized HR image $\mathbf{x}$. Its convolution kernel $\mathbf{H}_k$ 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 matrix $\mathbf{D}$ downsamples the image a factor L in x- and y-direction, $\mathbf{n}_k$ is a noise term and Δt_k the exposure duration. The LR, LDR observations are typically stored, and available as input to the SRR algorithm, after its pixel exposures are transformed by the Camera Response Function (CRF) $f : [0, \infty) \to [0, 1]$ (for image display, the monitor driver maps back to a suitable range). The operational dynamic range of the sensor is however limited to be what is termed Low Dynamic Range, i.e. the sensor exposure of all pixel elements j are clipped by the CRF, in this work to $(\Delta t_k \, \mathbf{i}_k)_j \in [0.01, 10]$. The CRF then is a nonlinear mapping from (normalized) photometric sensor exposure to a perceptually coded domain followed by quantization of the signal to a certain bit-depth.

3. IMAGE RECONSTRUCTION IN A PERCEPTUAL DOMAIN

To attempt to reconstruct the underlying High Resolution, High Dynamic Range image, knowledge of the image degradation processes for the images $\mathbf{y}_k$ is needed. In a complete SR algorithm, relative geometric shifts as well as blur kernels of the observations need to be estimated with high precision, either as pre-processing or simultaneously with estimation of the desired image. For differently exposed observations (different Δt_k), the images also need to be photometrically aligned, e.g. by mapping with the approximate inverse CRF, $g : [0, 1] \to [0.01, 10]$, here assumed to be known. Due to saturation in f, g is not strictly speaking an inverse function. In the remainder of this paper, it is also supposed that the blur kernels $\mathbf{H}_k$ are known. For each observation $\mathbf{y}_k$, introduce a diagonal weight matrix $\mathbf{W}_k$, which is zero for diagonal elements corresponding to saturated pixels in $\mathbf{y}_k$, and one otherwise. Then, $\mathbf{i}_k = g(\mathbf{y}_k)/\Delta t_k$, and

$$\begin{aligned}\mathbf{W}_k\mathbf{i}_k &= \mathbf{W}_k(g(\mathbf{y}_k)/\Delta t_k) = \\ &= \mathbf{W}_k(\mathbf{D}\mathbf{C}_{\mathbf{H}_k}\mathbf{x} + \mathbf{n}_k). \quad k = 1, \dots, K\end{aligned} \tag{2}$$

The weight matrix $\mathbf{W}_k$ is necessary for the second equality to hold, because some areas in $\mathbf{x}$ may be saturated in $\mathbf{i}_k$. Each $\mathbf{i}_k$ has its own non-saturated illuminance interval depending on its respective exposure duration, thus a higher combined dynamic range can be achieved. By stacking the individual LR, LDR illuminance domain observations in $\mathbf{i} = [\mathbf{i}_1^T, ..., \mathbf{i}_K^T]^T$, the noise vector $\mathbf{n} = [\mathbf{n}_1^T, ..., \mathbf{n}_K^T]^T$ and defining a block diagonal weight matrix $\mathbf{W} = diag(\mathbf{W}_1, ..., \mathbf{W}_K)$ and $\boldsymbol{\mathcal{H}} \triangleq [(\mathbf{D}\mathbf{C}_{\mathbf{H}_1})^T, ..., (\mathbf{D}\mathbf{C}_{\mathbf{H}_K})^T]^T$, it follows that

$$\mathbf{W}\mathbf{i} = \mathbf{W}(\boldsymbol{\mathcal{H}}\mathbf{x} + \mathbf{n}), \tag{3}$$

where $\boldsymbol{\mathcal{H}}$ has size $(X_1X_2K/L^2) \times (X_1X_2)$. More generally, $\mathbf{W}$ could give different weights depending on e.g. pixel value and exposure duration according to noise properties [15].

3.1. Solving the inverse problem

Solving the inverse problem involves finding a solution $\mathbf{x}$ to (3). The rank of $\mathbf{W}\boldsymbol{\mathcal{H}}$ depends on the number of LR, LDR observations K, and how many pixels in these that are saturated. For the case $L = 2$, and no saturation (which however implies $\mathbf{x}$ cannot be an HDR image), a minimum of

$K = L^2 = 4$ images are required for a full rank problem. Even for high K, there is often some image area that is saturated in most HDR images, thus creating a nullspace for $\mathbf{W}\boldsymbol{\mathcal{H}}$. To make the problem full rank, some type of regularization is required. Furthermore, the full rank case is in itself ill-conditioned [19]. Adding regularization then makes the problem less ill-conditioned. The Tikhonov regularized LS problem

$$\hat{\mathbf{x}}_{LS} = \arg\min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{W}\boldsymbol{\mathcal{H}} \\ \sqrt{\lambda}\mathbf{\Gamma_S} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{W}\mathbf{i} \\ \mathbf{0} \end{bmatrix} \right\|_2^2 \tag{4}$$

has a unique solution. The matrix $\mathbf{\Gamma_S}$ of size $(X_1X_2) \times (X_1X_2)$ performs 2d convolution on the vectorized image $\mathbf{x}$ with a 3×3 Laplacian kernel, $\mathbf{S}$, that enforces a smooth solution (images are typically piecewise smooth) by penalizing the 2nd order x- and y-derivative. A constant parameter λ tunes the amount of smoothing.

3.2. Perceptual weighting of reconstruction errors

We propose to measure reconstruction errors according to perceptual impact in a tonemapped domain. In image capture with a camera, the LDR sensor raw data is mapped by the CRF, f (also essentially a tonemapping function, but designed for the given LDR operational exposure range of the camera), to a perceptually uniform domain (typically the *sRGB* space). Similarly, $\tilde{f}$ as introduced here maps the HDR illuminance information to a perceptually uniform (still HDR) domain. Thus, the LS problem (4) becomes

$$\begin{aligned} \hat{\mathbf{x}}_{PU} &= \arg\min_{\mathbf{x}} \left\| \begin{matrix} \mathbf{W}(\tilde{f}(\boldsymbol{\mathcal{H}}\mathbf{x}) - \tilde{f}(\mathbf{i})) \\ \sqrt{\lambda}\mathbf{\Gamma_S}\tilde{f}(\mathbf{x}) \end{matrix} \right\|_2^2 \\ &\triangleq \arg\min_{\mathbf{x}} \|\mathbf{r}(\mathbf{x})\|_2^2, \end{aligned} \tag{5}$$

where $\mathbf{r}(\mathbf{x})$ denotes the residual to be minimized. To minimize $\tilde{f}(\mathbf{W}(\boldsymbol{\mathcal{H}}\mathbf{x} - \mathbf{i}))$, i.e. instead the tonemapped difference, would be incorrect, since the absolute illuminance level determines how sensitive the HVS is. The proposed objective function (5) is a nonlinear LS problem that needs to be solved iteratively. The simple, frequently used, gradient descent method has a very slow convergence rate in some directions for the discussed problem. Therefore, in the update step

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - \alpha^{(n)}\mathbf{d}^{(n)}, \tag{6}$$

the Gauss-Newton method is used to find the search direction

$$\mathbf{d}^{(n)} = \arg\min_{\mathbf{d}} \|\mathbf{J_r}(\mathbf{x}^{(n)})\mathbf{d} - \mathbf{r}(\mathbf{x}^{(n)})\|_2^2, \tag{7}$$

where $\mathbf{J_r}(\mathbf{x})$ is the Jacobian of $\mathbf{r}(\mathbf{x})$ in (5). In (6), $\alpha^{(n)}$ is obtained from a line search procedure in the direction of $\mathbf{d}^{(n)}$. Due to the large size of the Jacobian matrix $\mathbf{J_r}(\mathbf{x})$, the same as the size of LS problem in (4), $\mathbf{d}$ in the inner Gauss-Newton step (7) generally also needs to be computed iteratively, which is done using the same method of solving (4). Because $\mathbf{J_r}(\mathbf{x})$ is sparse, the *LSQR* method [20], in which $\mathbf{J_r}(\mathbf{x}^{(n)})\mathbf{d}$ is calculated without explicitly forming the matrix $\mathbf{J_r}(\mathbf{x}^{(n)})$, is the choice for the experiments in this article.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In this Section, two experimental results on SRR of HDR scenes using the proposed method are presented. Examples of scenes that are generally HDR are outdoor scenes with bright sky as well as shadow areas and indoor scenes with a daylit window. Here, the two original images used are Memorial Church and Mount Tam West[1]. Given these respective ground truth images, $\mathbf{x}$, semi-synthetical observations $\mathbf{y}_k$ are first generated according to (1). For the datasets in each experiment, $\mathbf{H}_k$ contains approximately Gaussian blur kernels of size 5×5 and small sub-pixel level relative shifts, so that new linearly independent equations are added to the reconstruction with each $\mathbf{y}_k$. Images are downsampled a factor $L = 2$, and their exposures mapped, after normalization to numerical values in $[0, 1]$, by the pixelwise function $f() = ()^{\gamma_{LDR}}, \gamma_{LDR} = 1/2.2$, a simplified yet realistic CRF. Through mapping with g and scaling with the respective Δt_k, $\{\mathbf{i}_k\}$ are subsequently given as in (2).

SRR is performed using both the conventionally employed illuminance domain approach in (4), as well as the proposed approach in (5). Here, the simple pixelwise tonemapping function $\tilde{f}() = ()^{\gamma_{HDR}}, \gamma_{HDR} = 1/6$ is applied after normalizing the full dynamic range of illuminance values of $\mathbf{i} \in [min(i), max(i)]$ to $[0, 1]$. In this domain, the magnitude in numerical error corresponds equally to perceived error across all of the dynamic range. For experiment 1 on the Memorial Church image (of size 384×256), $K = 5$ LR, LDR observations with $\Delta t = \{2^{-6}, 2^{-1}, 2^{-1}, 2^4, 2^4\}$ are used and the smoothing parameter is $\lambda = 10^{-3}$ for both objective functions. For experiment 2 on the Mount Tam West image (of size 182×302), there are $K = 4$ observations, $\Delta t = \{2^0, 2^0, 2^4, 2^4\}$ and $\lambda = 10^{-4}$. An initial estimate $\mathbf{x}^{(0)}$, to the iterative minimization (6), is produced from interpolating one of the $\mathbf{i}_k$ to the higher resolution of $\mathbf{x}$.

In Fig. 1 (a), the two original HDR images are displayed. These are degraded according to (1), including downsampling. Fig. 1 (b) shows the (zoomed-in) LR, LDR images used for initialization. Fig. 1 (c) and (d) show the reconstructed results $\hat{\mathbf{x}}_{LS}$ and $\hat{\mathbf{x}}_{PU}$, from the illuminance domain and the tonemapped domain respectively. In $\hat{\mathbf{x}}_{LS}$, clearly visible artifacts are present in image segments with sharp step edges from dim to bright areas. These arise from insufficient observations of non-saturated data in the area, due to e.g. over- or underexposure in some of the $\mathbf{i}_k$, necessitating some type of regularization. Enforcing smoothness of the image in the tonemapped domain gives a favorable

[1] Images are courtesy of Greg Ward, available at his website http://www.anyhere.com/gward/hdrenc/pages/originals.html

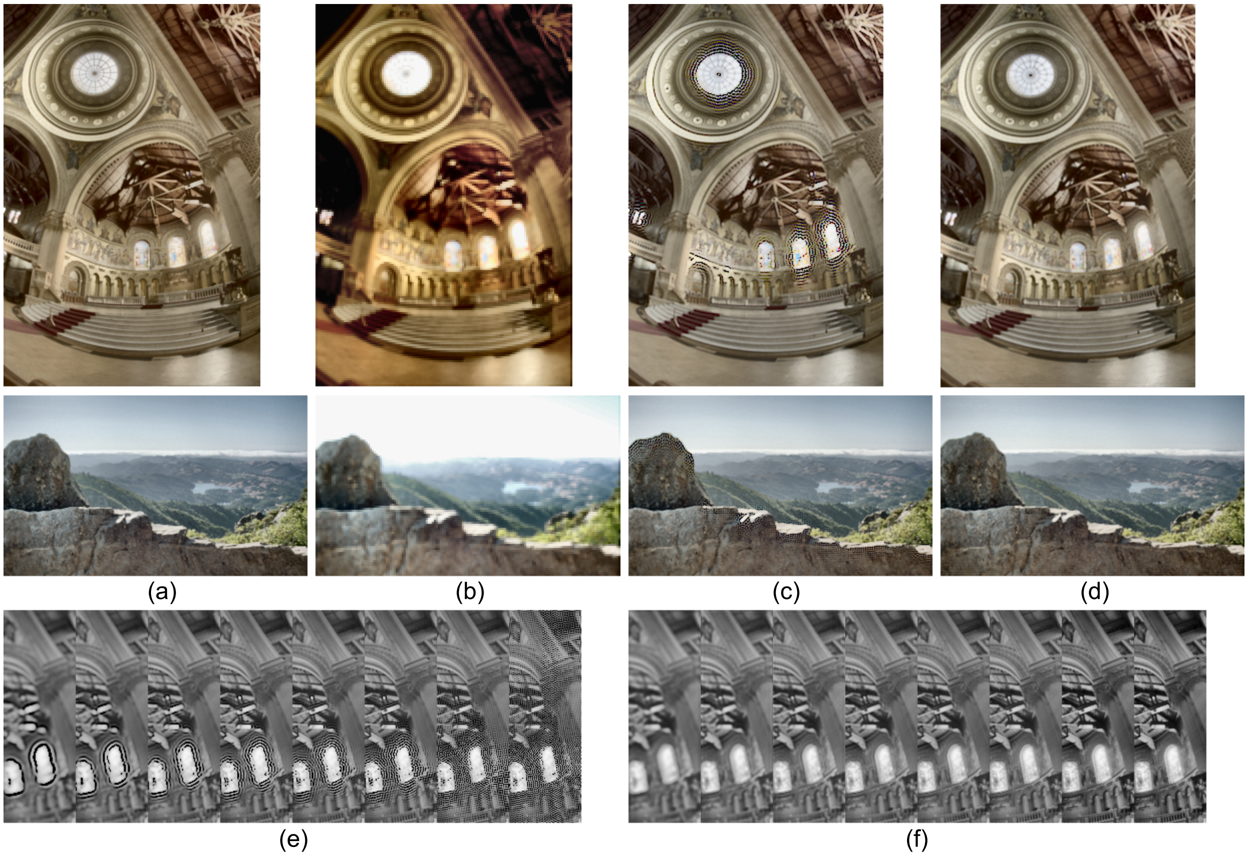

Fig. 1. HDR images Memorial Church and Mount West Tam. (a) Originals ,(b) interpolated LR, LDR $\mathbf{y}_k$ image used as initial guess, (c) Reconstructions in illuminance domain, (d) Reconstructions in tonemapped domain, (e) Illuminance domain approach for varying λ, (f) Tonemapped domain approach for varying λ. From left to right, $\lambda = \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-7}, 10^{-9}\}$

result. To use the saturated data in $\mathbf{i}$ would reduce the need for regularization, but would also distort the solution in those areas because of severely incorrect data. In practice, it is desired to keep the number of LR images used fairly low, due to temporal changes in the scene, as well as hardware constraints. Fig. 1 (e) and (f) show a part of the Memorial Church image reconstructed using different amount of smoothing regularization, from left to right, $\lambda = \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-7}, 10^{-9}\}$. The illuminance domain artifacts differ depending on the λ value, from producing edge artifacts for high λ to being unable to sufficiently smooth dim areas for low values. The proposed domain is fairly insensitive to the choice of λ.

Similar results to the ones presented here are obtained when replacing the simple Tikhonov regularization with non-linear, edge-preserving, regularization functions as proposed in e.g. [11]. There are improvements to be made with regard to choosing robust regularization- and norm functions, however these are complementary methods, rather than competing, to using a suitable image domain as proposed here. A robust objective function becomes crucial in cases with noisy data and imperfect image registration, where the inherently ill-conditioned nature of SRR becomes apparent. A more stable solution is achieved by increasing λ, but that always comes with the cost of a certain degree of oversmoothing. It is not clear-cut which regularization- and norm functions should be used, results are much dependent on the assumed degradation model and its noise properties. To our knowledge, no complete survey has been made for real data, but [21] offers a comprehensive study on various simulated noise scenarios.

5. CONCLUSION AND FUTURE WORK

A method to perform SRR on HDR images in which reconstruction errors correspond to perceived impact has been proposed., and a clear example of when the proposed method is beneficial has been presented. A framework has been used in which no restrictions are implied on the dynamic range of the desired image. For SRR methods in general, computational complexity is a big issue. Much consideration is required for sophisticated practical implementations, and is part of ongoing research.

6. REFERENCES

[1] M. Angelopoulou, C-S. Bouganis, P. Cheung, and G. Constantinides, "Robust real-time super-resolution on fpga and an application to video enhancement," *ACM Trans. Reconfigurable Technol. Syst.*, vol. 2, no. 4, pp. 22:1–22:29, 2009.

[2] S.C. Park, M.K. Park, and M.G. Kang, "Super-resolution image reconstruction: a technical overview," *Signal Processing Magazine, IEEE*, vol. 20, no. 3, pp. 21 – 36, May 2003.

[3] S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47–57, 2004.

[4] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1997, SIGGRAPH '97, pp. 369–378, ACM Press/Addison-Wesley Publishing Co.

[5] M. Cadik, M. Wimmer, L. Neumann, and A. Artusi, "Evaluation of hdr tone mapping methods using essential perceptual attributes," *Computers Graphics*, vol. 32, no. 3, pp. 330 – 349, 2008.

[6] J. Kuang, G. Johnson, and M. Fairchild, "icam06: A refined image appearance model for hdr image rendering," *J. Vis. Comun. Image Represent.*, vol. 18, pp. 406–414, Oct. 2007.

[7] R. Hardie, K. Barnard, and E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, pp. 1621–1633, 1997.

[8] L. Pickup, D. Capel, S. Roberts, and A. Zisserman, "Bayesian methods for image super-resolution," *The Computer Journal*, 2007.

[9] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Variational bayesian super resolution," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 984 –999, April 2011.

[10] Y. Chen, Y. Luo, and D. Hu, "A general approach to blind image super-resolution using a PDE framework," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2005.

[11] S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327 –1344, Oct. 2004.

[12] F. Sroubek, G. Cristobal, and J. Flusser, "A unified approach to superresolution and multichannel blind deconvolution," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2322 –2332, Sep. 2007.

[13] G. Harikumar and Y. Bresler, "Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 202 –219, Feb. 1999.

[14] M. Gevrekci and B. Gunturk, "Superresolution under photometric diversity of images," *EURASIP J. Appl. Signal Process.*, vol. 2007, pp. 205–205, Jan. 2007.

[15] J. Choi, M.K. Park, and M.G. Kang, "High dynamic range image reconstruction with spatial resolution enhancement," *Comput. J.*, vol. 52, pp. 114–125, Jan. 2009.

[16] F. Schubert, K. Schertler, and K. Mikolajczyk, "A hands-on approach to high-dynamic-range and superresolution fusion," in *WACV*, 2009, pp. 1–8.

[17] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand hdr imaging of moving scenes with simultaneous resolution enhancement," *Computer Graphics Forum*, vol. 30, no. 2, pp. 405–414, 2011.

[18] T. Bengtsson, I.Y.-H. Gu, M. Viberg, and K. Lindström, "Regularized optimization for joint super-resolution and high dynamic range image reconstruction in a percetually uniform domain," *ICASSP*, March 2012.

[19] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646 –1658, Dec. 1997.

[20] C. Paige and M. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Trans. Math. Softw.*, vol. 8, no. 1, pp. 43–71, March 1982.

[21] 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 Processing*, vol. 2007, no. 2, pp. 21–21, 2007.