HDR Imaging under Non-uniform Blurring

C.S. Vijay, Paramanand Chandramouli, and Rajagopalan Ambasamudram

Department of Electrical Engineering,
Indian Institute of Technology Madras, Chennai 600 036, India
{csv2000,paramanand}@gmail.com, raju@ee.iitm.ac.in

Abstract. Knowledge of scene irradiance is necessary in many computer vision algorithms. In this paper, we develop a technique to obtain the high dynamic range (HDR) irradiance of a scene from a set of differently exposed images captured using a hand-held camera. Any incidental motion induced by camera-shake can result in non-uniform motion blur. This is particularly true for frames captured with high exposure durations. We model the motion blur using a transformation spread function (TSF) that represents space-variant blurring as a weighted average of differently transformed versions of the latent image. We initially estimate the TSF of the blurred frames and then estimate the latent irradiance of the scene.

1 Background

The irradiance from a real world scene incident on a camera could range from very small to very large values. This variation is termed as the dynamic range. However, the CCD/CMOS sensors used in today's cameras are unable to accommodate the entire range. Over the past decade, this limitation has increasingly drawn the attention of vision researchers for applications in computer graphics, visual effect production etc. [1]. A widely followed approach for inferring scene irradiance is to optimally combine information from multiple images captured with different exposure times [2–6]. The focus in these methods is on modeling the camera pipeline and estimating the camera response function (CRF) to find the high dynamic irradiance of the scene. Another approach for HDR imaging is based on image fusion [7–9]. Here the focus is on obtaining a well-composed low dynamic range (LDR) image by finding a weighted average of the differently exposed intensity images.

For a real world scene, the exposure times can vary from values as small as $1/500s$ to as high as $5s$. Although the camera can be held still for the lower exposures, it would be an incredible feat for any photographer to hold the camera steady for long exposure times while taking multiple shots of a static scene. One could resort to tripods but they are bulky and difficult to carry around and setup. The issue of motion blurring has recently attracted a lot of attention. The restoration of uniformly blurred images was considered in [10, 11]. Fergus et al. [12] consider single image blind-deblurring for uniformly blurred images using natural image priors. However, as shown in [13, 14], camera rotation which induces space-variant blurring cannot be ignored in camera-shake.

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part II, LNCS 7584, pp. 451–460, 2012.

Fig. 1. Multiple exposed images where higher exposures are motion-blurred

The works in [14–16] consider single image deblurring of space-variantly motion-blurred images. Interestingly, none of the above works address the high dynamic (HD) nature of scenes. One of the effects of HD scenes is pixel saturation which causes ringing effects while deblurring. This problem was addressed in [17, 18] by masking outliers. Lu et al. [19] discuss the problem of obtaining HDR images from motion blurred observations and jointly estimate the CRF, blur kernels and the latent scene irradiance. However, they consider a space-invariant blur model which restricts the camera motion to in-plane translations. Though, this is a good step forward, it is limited in scope as the effect of rotation on blur is greater than that of translation [14] and hence cannot be ignored.

In this paper, we discuss a method to obtain the HDR image from a set of differently exposed images that are likely to be non-uniformly blurred. While capturing images with hand-held cameras, one can safely assume that no blurring will occur for low exposure times since the displacement of the camera in this short duration will be negligible. Hence, we consider a scenario with access to both blurred higher exposures and non-blurred lower exposures. The scene is assumed to be distant enough to preclude any parallax effects. The higher exposures (though blurred) contain information not present in the less-exposed images. Fig. 1 shows one such image set. One simplistic solution is to deblur the images and then reconstruct the HDR using existing methods. This is, of course, subject to the effectiveness of deblurring of these over-exposed images which i) may not be very accurate, and ii) may also contain artifacts. Here, we adopt the transformation spread function (TSF) [20] aka motion density function [15] to model space-variant blur. The TSF describes the blurring process as a weighted average of differently transformed versions of the latent scene irradiance. Studies have revealed that a general camera motion can be well-approximated by a 3D TSF (in-plane translation and rotation) when the depth variations are not significant [14, 15]. For the estimation of these TSFs, we assume that the blur is uniform locally, and use the local point spread functions (PSFs) of the blurred (high exposure) images, which are derived using patches of blurred and non-blurred (lower exposure) image pairs centered at the same location. The scene irradiance is then estimated using least square minimization along with total variation regularization. Thus, we simultaneously achieve the twin objectives of deblurring and HDR imaging.

We list below the main contributions of this work:

1. This is the first attempt of its kind for recovering HDR from *non-uniformly* blurred images.
2. We propose TSF estimation in the irradiance domain in order to handle

in-plane translation as well as rotation. This is a generalization of Lu et al.'s work [19] which allows for only simple translations since it uses uniform PSF.
3. We use locally estimated PSFs to arrive at the TSF which has not been attempted before.

2 Blur Model and TSF

During the image formation process, the scene radiance passes through a camera pipeline [2], before getting converted into intensity values. The pixel intensity value in an image is a monotonic but nonlinear function of the irradiance and the exposure period. The light energy accumulated can be described as

$$X = \int_{t=0}^{\Delta t} E \; dt + n \tag{1}$$

where E is the latent irradiance, Δt is the exposure time and n is additive noise. This accumulated energy undergoes several nonlinear conversions before being represented as pixel intensity value. These nonlinearities collectively are termed as the CRF, resulting in the intensity $Z = f(X)$. Given the exposure time and the CRF, the irradiance can be obtained as $B = f^{-1}(Z)/\Delta t$ where f^{-1} is the inverse CRF. It is important to note that this inverse mapping is not linear. Since the CRF maps the response of a particular camera setting, it is independent of the scene and can be calculated off-line.

When there is camera motion, the latent irradiance image (considering a static scene) can be viewed as undergoing a series of transformations. The final observation is an accumulation of the result of all these transformations. i.e.,

$$Z = f\left(\int_{t=0}^{\Delta t} \mathbf{\Gamma}_t\left(E\right) \; dt\right) + n \tag{2}$$

where $\mathbf{\Gamma}_t$ refers to the homography transformation operation at instant t, and $\mathbf{\Gamma}_t\left(E\right)$ denotes the outcome of applying $\mathbf{\Gamma}_t$ on E. Each of these transformations contributes according to the time spent in that state (exposure time), and is represented by suitably derived weights. While averaging over time, the temporal information is lost. In the TSF model, the result (blurred image) is the average over all possible transformations.

The general camera motion can be well approximated by 3 degrees of motion (in-plane translation and rotation) [15] when depth variations in the scene are minimal. The transformation space can be sampled to obtain a finite set $\mathbf{H}$ of $N_{\mathbf{H}}$ different transformations, which contains the transformations to approximately form the blurred image. The TSF h_T is a function that maps the finite set of transformations $\mathbf{H}$ to a non-negative real number i.e., $h_T : \mathbf{H} \rightarrow \mathbf{R}$. For each transformation $\mathbf{\Gamma} \in \mathbf{H}$, the value of the TSF $h_T\left(\mathbf{\Gamma}\right)$ denotes the fraction of the total exposure duration for which the camera stayed in the position that caused the transformation $\mathbf{\Gamma}$. The final blurred intensity image in terms of the TSF is

$$Z = f\left(\sum_{k=1}^{N_{\mathbf{H}}} h_T\left(\mathbf{\Gamma}_k\right) \mathbf{\Gamma}_k(E)\right) \tag{3}$$

2.1 TSF Estimation

In this work, the input consists of a set of differently exposed images with non-uniformly blurred higher exposures. The differently exposed images are obtained from the same scene irradiance. Hence, the inverse map of each intensity image to the irradiance domain will be comparable. The image intensity and irradiance are related by $B^i = f^{-1}(Z^i)/\Delta t^i$, where i takes values $1 \dots N_E$ for N_E differently exposed images, Z^i is the i^{th} observation, and B^i is the corresponding irradiance map. Because of the non-linearity of the CRF, irradiance values obtained from pixels with very high or very low intensities may not be comparable. The influence of such pixels is restricted using a mask whose weights are obtained from a 256-point Tukey window.

While estimating the TSF, we first estimate the PSF at local patches where the blur may be assumed to be locally uniform. The estimation is carried out with respect to a non-blurred lower exposure by minimizing

$$\underset{h}{\operatorname{argmin}} \left\| \left(m_p \cdot B^s_{\mathbf{p}} - m_p \cdot \left(h * E^r_{\mathbf{p}} \right) \right) \right\|_2 \tag{4}$$

where h is a local PSF to be estimated, s is the observation index corresponding to one of the blurred higher exposures, r is the index of highest exposed non-blurred observation, $B^s_{\mathbf{p}}$ is the blurred irradiance patch of an exposure corresponding to s, $E^r_{\mathbf{p}}$ is the the non-blurred irradiance patch of an exposure corresponding to r, and m_p is the mask applied at location $\mathbf{p}$. We estimate $N_{\mathbf{p}}$ such PSFs $h_{\mathbf{p}_1}\ h_{\mathbf{p}_2} \dots h_{\mathbf{p}_{N_p}}$ at image locations $\mathbf{p}_1\ \mathbf{p}_2 \dots \mathbf{p}_{N_{\mathbf{p}}}$. If any misalignment is present among the observations, it is implicitly accounted for in the estimated TSF (globally) and PSF (locally).

Fig. 2 demonstrates PSF estimation for a non-uniformly blurred image captured with a hand-held camera. The space-variant nature of the blur is evident from the estimated PSFs. A PSF at a pixel (i, j) is related to TSF as

$$h\left(i, j; m, n\right) = \sum_{k=1}^{N_{\mathbf{H}}} h_T\left(\mathbf{\Gamma}\right) \delta_d\left(m - \left(i_{\mathbf{\Gamma}} - i\right), n - \left(j_{\mathbf{\Gamma}} - j\right)\right) \tag{5}$$

where $(i_{\mathbf{\Gamma}}, j_{\mathbf{\Gamma}})$ denotes the co-ordinates of the point when a transformation $\mathbf{\Gamma}^{-1}$ is applied on (i, j), and δ_d denotes the 2D Kronecker delta function. Each local PSF $\mathbf{p}_l$ can be considered as a weighted sum of the components of the TSF of the image. Such a linear relation can be expressed as $h_{\mathbf{p}_l} = M_l h_T$ for $l = 1 \dots N_p$, where M_l is a matrix whose entries are determined by the location $\mathbf{p}_l$ of the blur kernel and the interpolation coefficients. All the M_ls and $\mathbf{p}_l$s for $l = 1 \dots N_p$ can be stacked and related to the TSF as

$$\overline{h} = \mathbf{M} h_T \tag{6}$$

where $\overline{h}$ is a stack of the estimated blur kernels. A typical value for N_p is 5. To get an estimate of the sparse TSF that is consistent with the observed blur kernel, we minimize the following cost function

$$\underset{h_T}{\operatorname{argmin}} \left\| \overline{h} - \mathbf{M} h_T \right\|_2 + \lambda_s \left\| h_T \right\|_1 \tag{7}$$

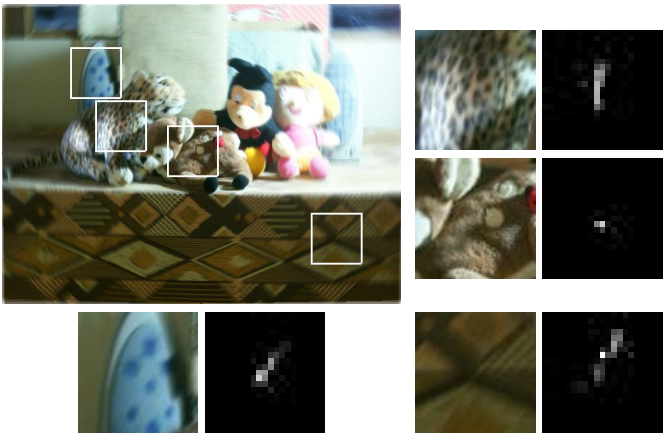


Fig. 2. The blurred image and some select patches are shown. The estimated PSFs are shown along-side each patch.

where a sparsity constraint is enforced by the l_1-norm term. To minimize the cost function, we use the toolbox available at [21]. The TSFs of all the blurred higher exposures are obtained using this method. For a detailed explanation of TSF estimation, we refer the reader to our supplementary material [22].

3 Scene Irradiance Estimation

Having obtained the TSFs of the blurred images, we proceed to invert the blurring process to obtain the actual scene irradiance which is present in the different exposures. We pose this as an optimization problem, where the solution can be obtained by minimizing the cost function

$$\sum_{i=1}^{N_E} \|K^i E - B^i\|^2 \tag{8}$$

Here E and B^i are the latent irradiance and the given i^{th} image irradiance of the scene, expressed in vector form and the sum is over all the observations. K^i is a large but sparse matrix, which represents the space-variant blurring operation and whose rows are a local blur filter acting on E.

The CRF of cameras is highly nonlinear at very high and very low pixel intensities. In order to restrict the influence of these nonlinearities, we mask these pixels from contributing to the estimation process. This is done to emphasize the usage of the linear part of the CRF. The mask is derived from a Tukey window, which has a smooth transition from 0 to 1. The mask applied is common to all 3 channels. This is done so as to maintain color balance. The parameter of the window used to obtain the mask depends on the CRF and image histogram (intensity distribution). This mask also acts as a pixel weighting function for different exposures. The summation can be removed from equation (8) by stacking

all the irradiance images (B^is) into a single matrix $\widetilde{B}$ and the K^i matrices into a single matrix $\widetilde{K}$ and can be expressed in a single matrix equation

$$\|\overline{m}\widetilde{K}E - \overline{m}\widetilde{B}\|^2 \tag{9}$$

where E is the latent irradiance to be estimated and $\overline{m}$ is a large diagonal matrix which is a collection of all the masks applied to different exposures. The mask values are arranged along the diagonal. While optimizing the above cost, total variation regularization is used to stabilize the solution, thus giving the energy

$$O = \underset{E}{\operatorname{argmin}} \|\overline{m}.\widetilde{K}E - \overline{m}.\widetilde{B}\|^2 + \lambda_{TV} \int |\nabla E| \tag{10}$$

We solve the above equation iteratively using conjugate gradient method. The derivative of the cost function is

$$\frac{\partial O}{\partial E} = \overline{m}.\widetilde{K}^T \left(\overline{m}.\left(\widetilde{K}E - \widetilde{B}\right)\right) + \lambda_{TV} \left(\nabla \frac{\nabla E}{|\nabla E|}\right) \tag{11}$$

where E is the irradiance image in t^{th} iteration, λ_{TV} is the total variation regularization parameter, and K^T represents the adjoint of the blurring operation and corresponds to inverse warping [16]. The term $\widetilde{K}E$ is obtained by blurring E using TSFs that are present in $\widetilde{K}$.

4 Experimental Results

For the synthetic case, we captured a set of observations at different exposures from a still camera. The higher exposures from this non-blurred dataset were subject to non-uniform motion blurring. The real data was captured with a hand-held DSLR Sony α and a compact Fujifilm HS10. While the blurring was found to be negligible for lower exposures, it was significant at higher exposures. All the datasets were captured from a sufficiently long distance. The CRF of these cameras were derived off-line using the code of [2] from images captured using a tripod (for stability) and for identical camera settings as used in the experiments. To recover the HDR image from the given set (which includes blurred observations), we first convert the intensities to irradiances. Next, we select patches in a blurred image and corresponding patches in a non-blurred frame (with maximum exposure) to evaluate the local PSFs. Using these estimated PSFs, the TSF corresponding to each of the blurred images is determined by solving Eqn.(7). The scene irradiance is finally estimated by minimizing Eqn.(10). For viewing purpose, the estimated irradiance image was tone-mapped using the technique of [23]. The quality of the final LDR image is subject to tone-mapping.

We initially applied our scheme on the synthetic image set shown in Fig. 3. The higher exposures (Fig. 4(a) and (b)) of $\frac{1}{2}s$ and $2s$ are blurred using known TSFs. These TSFs were chosen along the lines of blurs caused by real camera-shake in hand-held cameras so as to simulate a real-world scenario. The TSF dimensions t_x and t_y took integer values $\in$ [-12,12], and θ_z ranged $\in$ [-1.5,1.5]

Fig. 3. Synthetic image set

(a) (b)

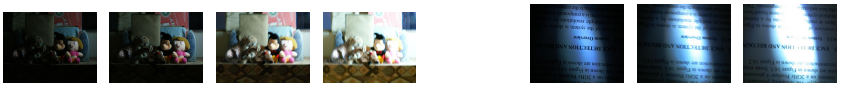
Fig. 4. Synthetically blurred higher exposures

(a) (b) (c) (d)

Fig. 5. (a) Original tone-mapped HDR image derived from non-blurred images. (b) HDR obtained from individually deblurred images. (c) Result obtained using [19]. (d) Estimated HDR tone-mapped image using our method.

degrees in steps of 0.25 degrees. The blurred images have visible regions that are not apparent in lower exposures. We hope to recover these while estimating the HDR irradiance but without blur.

Fig. 5(a) shows the tone-mapped HDR image obtained from the original set of non-blurred images using [2]. This result is used for comparison and to evaluate the performance of our method. Fig. 5(b) shows the HDR image obtained from a set of deblurred images. The deblurring was performed using the code provided by Whyte et al. [14]. It is quite apparent that the image is still blurred to some extent (as can be seen around the numbers) and it contains artifacts along the edges. This is understandable as [14] is not designed for the HDR problem. Hence a naive and computationally intensive solution of deblurring images and finding the HDR is not very effective. Fig. 5(c) shows the result obtained using [19] for our non-uniformly blurred data. Although the image has been deblurred in certain regions such as the illuminated parts and the book title, residual blur can be seen in the outer region where the effect of rotation is more. For instance, the digits in the calender and the flower pattern near the left-hand-side boundary of the image appear blurred. This is not surprising since [19] is not equipped to handle non-uniform blur. Fig. 5(d) shows the recovered HDR image using our method. The details in the scene become more apparent in our output. The front-illuminated part is visible and so are the darker regions in the scene such as the book title and the top image on the booklet. The numbers which become better visible with increased exposure have been uniformly lit. The result is in-fact comparable to Fig. 5(a) which was obtained without any blur.



(a) (b)

Fig. 6. Two different real data sets

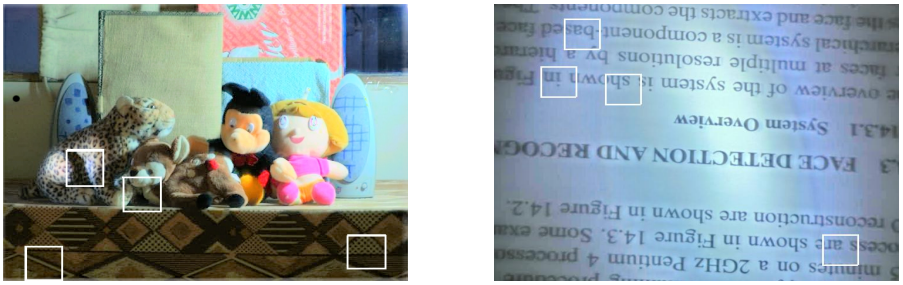


Fig. 7. Recovered HDR images for the datasets in Fig. 6. using our method

(a) (b) (e) (f)

(c) (d) (g) (h)

Fig. 8. Close-up of blurred data (available only in the higher exposures) and the corresponding patch in the output image

Fig. 6 shows two real image sets with the highest exposure being $2s$. The first set shown in Fig. 6(a) has images of size 640×480 captured at exposure times $\frac{1}{15}, \frac{1}{8}, \frac{1}{2}$ and 2s. The two highest exposures are motion-blurred. These frames contain information about the lower portion of the scene that is not visible in the less-exposed frames. The second set in Fig. 6(b) shows images captured with exposure durations $\frac{1}{8}, \frac{1}{3}$ and 1s. The observations corresponding to $\frac{1}{3}$ and 1s are blurred. The difference in the information available among these observations is clearly evident.

In Fig. 7, the HDR outputs of both the sets are shown for our model. The results have been rid of saturation effects and the darker regions are well-lit. The blurred information in the higher exposures has been restored while retaining the information from lower exposure, such as the illuminated regions in the scene. In Fig. 8, we show close-up patches from the most exposed images and the estimated result (side-by-side). Figs. 8(a-d) correspond to the first set while Figs. 8(e-h) correspond to the second. Fig. 8(a) reveals reduced smudging in the image. Also, the texture on the toy has been recovered. Patch sets 8(b) and (c) bring out the pattern present in the lower part of the scene clearly. These

(a) (b) (c) (d)

Fig. 9. (a) Our Result. (b) Result of [19]. (c) Close-ups of (a). (d) Close-ups of (b).

patterns are not quite visible in the lower exposures. From Fig. 8(d), we can also observe that the result is shifted with respect to the input blur patches. This shift is due to the compensation for the misalignment between input patches. Fig. 8(e-h) further reinforce the fact that our method satisfactorily achieves the twin objectives of HDR imaging and deblurring.

In Fig. 9, we give a comparison of the results obtained by our method (Fig. 9(a)) with that of [19] (Fig. 9(b)) on the dataset of Fig. 6(a). Both the results are well lit but Fig. 9(b) has significant residual blur. While Fig. 9(c) shows the close-up patches from our result, those obtained using [19] are shown in Fig. 9(d). Since the upper portion of the scene is well illuminated, the non-blurred lower exposures contribute to the final result. Hence, in the first patch, we see that the outputs from both the algorithms are focused. However, since [19] uses a convolution model, blur persists in their result as highlighted in second and third patches of Fig. 9(d). In contrast, the corresponding patches from our output in Fig. 9(c), show that the latent irradiance image has been completely recovered. This example serves to clearly bring out the advantages of our approach over the state-of-the-art [19] in HDR imaging with blur.

5 Conclusions

In this paper, we proposed an approach to obtain high dynamic scene irradiance from a set of frames captured with a hand-held camera, where the higher exposures are non-uniformly blurred due to camera-shake. A two-step procedure was proposed in which we first obtain the TSF of each blurred image followed by regularization to solve for scene irradiance. The method was validated on both synthetic and real datasets. It will be interesting to extend our framework to include the possibility of all the frames being blurred (as in the case of night scenes) and for dynamic scenarios.

References

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

2. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: ACM SIGGRAPH (1997)
3. Mann, S., Picard, R.W.: On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. Citeseer (1995)
4. Mitsunaga, T., Nayar, S.K.: Radiometric self calibration. In: Proc. CVPR (1999)
5. Fattal, R., Lischinski, D., Werman, M.: Gradient domain high dynamic range compression. ACM Trans. on Graph. 21 (2002)
6. Reinhard, E., Ward, G., Pattanaik, S., Debevec, P.: High dynamic range imaging. Elsevier (2006)
7. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion. In: Pacific Conf. on Computer Graph. and App. (2007)
8. Raskar, R., Ilie, A., Yu, J.: Image fusion for context enhancement and video surrealism. In: ACM SIGGRAPH 2005 Courses (2005)
9. Ward, G.: Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. Journal of Graphics Tools 8, 17–30 (2003)
10. Rav-Acha, A., Peleg, S.: Two motion-blurred images are better than one. Pattern Recognition Letters 26 (2005)
11. Yuan, L., Sun, J., Quan, L., Shum, H.: Image deblurring with blurry/noisy image pairs. ACM Trans. Graph. (26)
12. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. ACM Trans. on Graphics 25 (2006)
13. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Understanding and evaluating blind deconvolution algorithms. In: Proc. CVPR (2009)
14. Whyte, O., Sivic, J., Zisserman, A., Ponce, J.: Non-uniform deblurring for shaken images. In: Proc. CVPR (2010)
15. Gupta, A., Joshi, N., Lawrence Zitnick, C., Cohen, M., Curless, B.: Single Image Deblurring Using Motion Density Functions. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 171–184. Springer, Heidelberg (2010)
16. Tai, Y., Tan, P., Brown, M.S.: Richardson-lucy deblurring for scenes under projective motion path. IEEE Trans. PAMI 33 (2011)
17. Whyte, O., Sivic, J., Zisserman, A.: Deblurring shaken and partially saturated images. In: IEEE Workshop CPCV, with ICCV (2011)
18. Cho, S., Wang, J., Lee, S.: Handling outliers in non-blind image deconvolution. In: Proc. ICCV (2011)
19. Lu, P.Y., Huang, T.H., Wu, M.S., Cheng, Y.T., Chuang, Y.Y.: High dynamic range image reconstruction from hand-held cameras. In: Proc. CVPR (2009)
20. Chandramouli, P., Rajagopalan, A.N.: Inferring image transformation and structure from motion-blurred images. In: Proc. of the BMVC (2010)
21. Liu, J., Ji, S., Ye, J.: Slep: sparse learning with efficient projections (2009), `http://www.public.asu.edu/~jye02/Software/SLEP`
22. Supplementary material: Transformation spread function estimation (2012) Supplied as additional material `SupMaterial.pdf`
23. Reinhard, E., Stark, M., Shirley, P., Ferwerda, J.: Photographic tone reproduction for digital images. ACM Transactions on Graphics 21 (2002)