HIGH DYNAMIC RANGE IMAGING BY A RANK-1 CONSTRAINT*

Tae-Hyun Oh, Joon-Young Lee, In So Kweon

Robotics and Computer Vision Lab, KAIST, Korea

ABSTRACT

We present a high dynamic range (HDR) imaging algorithm that utilizes a modern rank minimization framework. Linear dependency exists among low dynamic range (LDR) images. However, global or local misalignment by camera motion and moving objects breaks down the low-rank structure of LDR images. The proposed algorithm simultaneously estimates global geometric transforms to align LDR images and detects moving objects and under-/over-exposed regions using a rank minimization approach. In the HDR composition step, structural consistency weighting is proposed to generate an artifact-free HDR image from an user-selected reference image. We demonstrate the robustness and effectiveness of the proposed method with real datasets.

Index Terms— HDR, Rank Minimization, Alignment.

1. INTRODUCTION

Radiance of a scene has far wider dynamic range than the dynamic range of consumer cameras. Due to the limited dynamic range of cameras, the photographer should adjust the range to focus on the region of interest of a scene by controlling the exposure time. Modern digital cameras provide photographers with convenient functions such as auto-exposure to a focused region. However, the user cannot avoid under- or over-saturation if the captured scene has a large difference between low and high radiance. The issue of HDR composition therefore must be addressed to overcome the inherent difference between the dynamic range of human visual perception and that of electronic imaging devices in practice.

To recover the wider range of scene radiance, conventional approaches inversely follow an image acquisition pipeline and estimate sensor irradiance from multi-exposure images. These HDR composition methods with multi-exposure LDR images recover radiance maps of static scenes well. However, there are difficulties in directly applying the conventional methods because ghost artifacts and alignment errors occur due to both global transformations among LDR images by camera shake and local misalignment by moving objects and under-/over-saturation.

In this paper, we present an Intensity Observation Model (IOM) that describes an intensity acquisition pipeline from sensor irradiance to image intensity. From the IOM, a linear dependent relationship (rank-1 representation) among differently exposed images naturally arises. We exploit the linearity in the IOM to handle geometric transformations and misalignment in a unified rank minimization framework with a rank-1 constraint.

2. RELATED WORK

Mann and Picard [1] and Debevec and Malik [2] are considered pioneer works in the field of HDR imaging. In [1, 2], they estimate the camera response function (CRF) and compose radiance maps from multi-exposure images with the assumption; that a static scene is captured with a fixed camera. Ward [3] presents a translation based alignment algorithm to account for camera motion among multi-exposure images. The method searches translational motions along the X/Y-axis. In practice, real camera motions by hand shake include rotational motions that cannot be modeled with translations. Also, captured scenes may contain moving objects, which cause ghost artifacts in the HDR image.

There have recently been some efforts to make artifact-free HDR images. Gallo *et al*. [4] detect artifact regions using a linear property of log radiance values by a block-wise comparison. While their approach can handle ghost artifacts, blocking artifacts may remain near block boundaries. Heo *et al*. [5] propose a ghost-free HDR imaging framework by using a joint bilateral filter approach. They align LDR images by homographies and detect ghost regions using graph cuts [6]. The method shows appealing results, but there are many parameters in each step and due to the heuristic combination of various algorithms the results strongly depend on the performance of each algorithm. For instance, ghost detection results tend to be sensitive to threshold parameters and the method fails to recover ghost-free HDR images if any unsuitable parameter is selected.

Contrary to the above methods, our work is based on the inherent property from an image acquisition model, and also provides a joint optimization approach for registration and outlier detection simultaneously.

3. PROPOSED ALGORITHM

In this section, we formulate the HDR composition as a rank minimization problem that simultaneously estimates a set of geometric transformations to align LDR images and detects both moving objects and under/over-saturated regions.

*THIS WORK WAS SUPPORTED BY THE NATIONAL RESEARCH FOUNDATION OF KOREA(NRF) GRANT FUNDED BY THE KOREA GOVERNMENT(MEST) (NO.2013-006603)

3.1. Low-Rank Structure of Multi-Exposure Images

For a static scene, multi-exposure images taken from a fixed camera are linearly proportional to the exposure time Δt under a linear CRF, because the sensor irradiance R of the scene is constant [7, 8]. While a linear relationship is imposed among the images, we cannot directly observe the ideal relationship in practice due to artifacts such as camera motions, moving objects, and intensity saturation.

We model a forward-intensity acquisition pipeline, called an IOM, in consideration of the artifacts. The artifacts from moving objects or saturation occupy a relatively small region in the LDR images. They are modeled as sparse error E_S with sparse non-zero entries and large magnitudes. The camera motion is represented as g in the homography transformation group $\mathbb{G}$, which is a p-parameter group. Therefore, the intensity image I is observed though the IOM as

$$I = f\left(k(R + E_S)\,\Delta t\right) \circ g^{-1}, \tag{1}$$

where f denotes CRF, $\circ$ is the element-wise mapping operator, and k denotes a constant scaling factor. We assume a linear CRF because it can be estimated by various calibration methods even if images are unaligned (*e.g.*Grossberg *et al.* [9]). Under this assumption, each observed image is represented as

$$\begin{aligned} I_i \circ g_i &= f\left(k\big(R + E_S^i\big) \cdot \Delta t_i\right) \\ &= kR \cdot \Delta t_i + kE_S^i \cdot \Delta t_i \quad \text{(if } f \text{ is linear)} \\ &= A_i + E_i \qquad\qquad (A_i = kR \cdot \Delta t_i, \\ &\qquad\qquad\qquad\quad E_i = kE_S^i \cdot \Delta t_i), \end{aligned} \tag{2}$$

where i denotes an image index of multiple input images. By stacking the vectorized images I_i, we construct the observed intensity matrix $\mathbf{O} = [\mathrm{vec}(I_1)|\cdots|\mathrm{vec}(I_n)] \in \mathbb{R}^{m\times n}$, where m and n are the number of pixels and images, respectively. We then use matrix representation for $\mathbf{A} = [\mathrm{vec}(A_1)|\cdots|\mathrm{vec}(A_n)]$, $\mathbf{E} = [\mathrm{vec}(E_1)|\cdots|\mathrm{vec}(E_n)]$, and $\mathbf{g} = \{g_1, \ldots, g_n\} \in \mathbb{R}^{p\times n}$. The IOM from Eq. (2) is represented as a matrix form $\mathbf{O} \circ \mathbf{g} = \mathbf{A} + \mathbf{E}$. Each column of the matrix $\mathbf{A}$ is spanned by sensor irradiance R. This means that the aligned observation $\mathbf{O} \circ \mathbf{g}$ is equal to the background irradiance $\mathbf{A}$, which is the rank-1 matrix, if there is no artifact in a scene ($\mathbf{E} = \mathbf{0}$). In practice, the rank of $\mathbf{O} \circ \mathbf{g}$ is higher than 1 due to the aforementioned artifacts. Therefore, HDR composition that is robust to outliers becomes a problem to decompose the observation matrix $\mathbf{O} \circ \mathbf{g}$ into the rank-1 matrix $\mathbf{A}$ and sparse error matrix $\mathbf{E}$, and simultaneously to estimate transformations $\mathbf{g}$ that make the matrix $\mathbf{O}$ possibly close to rank-1.

3.2. Rank Minimization Approach

Our formulation is inspired by Peng *et al.* [10], which is a batch image alignment task [11] that utilizes low-rank and sparsity of the matrices $\mathbf{A}$ and $\mathbf{E}$. We observed that the solution from [10] includes some outliers as inliers and vice versa, and the alignment accuracy is degenerated when the number of inputs in $\mathbf{O}$ is very limited. Such limited observations are common in HDR problems due to a practical reason. Common HDR methods capture images with only 2-5 exposures, and the use of five exposures could be enough to cover almost all the informative dynamic range of a scene.

To derive a more satisfying solution, we utilize the prior rank information as a constraint. We encourage the constraint as an inequality constraint to robustly deal with residual factors, which cannot be modeled with sparsity with large magnitudes (*e.g.* Gaussian noise has dense entries with small magnitudes). With the rank constraint, our rank minimization problem is formulated as follows:

$$\begin{aligned} \mathbf{A}^*, \mathbf{E}^* &= \arg\min_{\mathbf{A},\mathbf{E}} \ \mathrm{rank}(\mathbf{A}) \ + \ \lambda\|\mathbf{E}\|_0, \\ &\text{subject to } \mathbf{O} \circ \mathbf{g} = \mathbf{A} + \mathbf{E}, \mathrm{rank}(\mathbf{A}) \geqq 1, \end{aligned} \tag{3}$$

where $\|\cdot\|_0$ denotes l^0-norm (the number of non-zero entries in matrix), and λ is the weight for sparse error. Unfortunately, solving Eq. (3) is known to be intractable and optimization with an inequality constraint is not easy. We approximate Eq. (3) by the convex relaxation [12, 13, 10] and encourage the constraint by replacing the rank$(\cdot)$ and inequality rank constraint with the sum of singular value ratios of $\mathbf{A}$, similar to the definition in [8]. Minimizing the sum satisfies the constraint and is equal to minimizing the residual rank of $\mathbf{A}$, because the first ratio is always 1. Our convex relaxed objective function is given by

$$\begin{aligned} \mathbf{A}^*, \mathbf{E}^*, \mathbf{g}^* &= \arg\min_{\mathbf{A},\mathbf{E},\mathbf{g}} \ \sum\nolimits_{i=2}^{\min(m,n)} \frac{\sigma_i(\mathbf{A})}{\sigma_1(\mathbf{A})} \ + \ \lambda\|\mathbf{E}\|_1, \\ &\text{subject to } \mathbf{O} \circ \mathbf{g} + \textstyle\sum_{j=1}^{n} J_j \boldsymbol{\Delta}\mathbf{g}\epsilon_j\epsilon_j^T = \mathbf{A} + \mathbf{E}, \end{aligned} \tag{4}$$

where $\sigma_i(\mathbf{A})$ denotes the i-th singular value of $\mathbf{A}$, $J_i = \frac{\partial}{\partial \varsigma}\mathrm{vec}(I_i \circ \varsigma)|_{\varsigma=g_i} \in \mathbb{R}^{p\times n}$ is the Jacobian of the i-th image with respect to the transformation g_i (we invite the reader to [10] for details about the Jacobian representation.), and $\{\epsilon_i\}$ denotes the standard basis for $\mathbb{R}^n$. To avoid trivial solutions, we use the ratio of singular values instead of directly using singular values [8].

3.3. Optimization

The proposed objective function in Eq. (4) is a constrained optimization problem. Lin *et al.* [14] propose an augmented Lagrange multipliers (ALM) method to minimize high dimensional nuclear norm, and Peng *et al.* [10] adapt the ALM method to solve the similar problem of Eq. (4). These approaches are known as scalable and fast convergence methods. We follow the optimization procedure in [10]. Let us define $h(\mathbf{A}, \mathbf{E}, \boldsymbol{\Delta}\mathbf{g}) = \mathbf{O} \circ \mathbf{g} + \sum_{j=1}^{n} J_j \boldsymbol{\Delta}\mathbf{g}\epsilon_j\epsilon_j^T - \mathbf{A} - \mathbf{E}$. The proposed Lagrangian function of Eq. (4) is then given by

$$\begin{aligned} L(\mathbf{A}, \mathbf{E}, \boldsymbol{\Delta}\mathbf{g}, \mathbf{Z}, \mu) &= \sum\nolimits_{i=2}^{\min(m,n)} \sigma_i(\mathbf{A})/\sigma_1(\mathbf{A}) + \lambda\|\mathbf{E}\|_1 \\ &\quad + < \mathbf{Z}, h(\mathbf{A}, \mathbf{E}, \boldsymbol{\Delta}\mathbf{g}) > + \frac{\mu}{2}\left\|h(\mathbf{A}, \mathbf{E}, \boldsymbol{\Delta}\mathbf{g})\right\|_F^2, \end{aligned} \tag{5}$$

where μ is a positive scalar, $\mathbf{Z} \in \mathbb{R}^{m\times n}$ is an estimate of the Lagrange multiplier matrix, $<,>$ denotes the matrix inner product, and $\|\cdot\|_F$ denotes the Frobenius norm. To solve this easily, Eq. (5) is divided into three sub-problems for $\mathbf{A}, \mathbf{E}$, and $\mathbf{\Delta g}$ and iteratively minimized. The sub-problems for $\mathbf{E}$ and $\mathbf{\Delta g}$ are identical with [10], and therefore those problems can be solved by [15, 10]. The reader can refer to papers [15, 10] for details. The sub-problem for $\mathbf{A}$ is updated by fixing the other variables, and it is derived by

$$\begin{aligned}\mathbf{A}_{t+1} &= \arg\min_{\mathbf{A}} L(\mathbf{A}, \mathbf{E}_t, \mathbf{\Delta g}_t, \mathbf{Z}_t, \mu_t) \\ &= \arg\min_{\mathbf{A}} \sum\nolimits_{i=2}^{\min(m,n)} \sigma_i(\mathbf{A})/\sigma_1(\mathbf{A}) \\ &\quad + <\mathbf{Z}_t, \mathbf{O}' - \mathbf{A} - \mathbf{E}_t> + \frac{\mu_t}{2}\|\mathbf{O}' - \mathbf{A} - \mathbf{E}_t\|_F^2 \\ &= \arg\min_{\mathbf{A}} \mu_t^{-1} \sum\nolimits_{i=2}^{\min(m,n)} \sigma_i(\mathbf{A})/\sigma_1(\mathbf{A}) \\ &\quad + \frac{1}{2}\left\|\mathbf{A} - (\mathbf{O}' - \mathbf{E}_t + \mu^{-1}\mathbf{Z}_t)\right\|_F^2, \end{aligned} \tag{6}$$

where $\mathbf{O}' = \mathbf{O} \circ \mathbf{g} + \sum_{j=1}^{n} J_j \mathbf{\Delta g} \epsilon_j \epsilon_j^T$, and t indicates the iteration index.

We normalize $\mathbf{A}$ by l^2-norm in the initial step. Then, $\sigma_1(\mathbf{A})$ becomes 1 and Eq. (6) can be solved by the Partial Singular Value Thresholding (PSVT) operator [16] (rank-1 case):

$$\begin{aligned}\mathbb{P}_\tau[\mathbf{Y}] =& \mathbf{U}(\mathbf{D}_{Y1} + \mathcal{S}_\tau[\mathbf{D}_{Y2}])\mathbf{V}^{\mathrm{T}} \\ &= \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 + \tau \sum\nolimits_{i=2}^{\min(m,n)} \sigma_i(\mathbf{X}), \\ &\text{where } \mathbf{D}_{Y1} = \mathrm{diag}(\sigma_1, 0, \cdots, 0), \\ &\qquad \mathbf{D}_{Y2} = \mathrm{diag}(0, \sigma_2, \cdots, \sigma_l),\end{aligned} \tag{7}$$

where $\tau > 0$ and $\mathcal{S}_\tau[\mathbf{X}] = \{\max(0, x - \tau)\}$ denotes the entry-wise soft-thresholding operator [15]. $\mathbf{U}, \mathbf{V}$ and $\mathbf{D}(= \mathbf{D}_{Y_1} + \mathbf{D}_{Y_2})$ correspond to the singular value decomposition of $\mathbf{Y}$. In Oh *et al*. [16], the PSVT operator provides the closed-form solution of Eq. (7).

For each iteration, $\mathbf{A}_{t+1}$ can be updated with $\mathbb{P}$ as

$$\mathbf{A}_{t+1} = \mathbb{P}[\mathbf{O}' - \mathbf{E}_t + \mathbf{Z}_t/\mu_t, \mu_t^{-1}]. \tag{8}$$

3.4. HDR composition

Since different exposure images capture different dynamic range of a scene, taking multiple exposure images and combining them may create a more informative image that captures all details of the scene. LDR images are combined by $H(x) = \sum_{i=1}^{n} W^i(x) \cdot R^i(x)/\Delta t_i$, where n represents the number of input images and $H(x)$, $R^i(x)$, and $W^i(x)$ denote the estimated radiance, the sensor irradiance, and the weight of the pixel located at x in the i_{th} exposure, respectively.

Instead of direct use of a low-rank matrix, we use radiometrically calibrated intensity $R^i(x)$, and utilize the estimated sparse error as a weight, due to the following reason. A photographer might want to compose a HDR image with a moving object that appears in one of multi-exposure images. Hence, the composition quality relies on the weight term W, which can help to remove undesired artifacts and to leave desired objects by selecting a reference image.

In this paper, we use two different weighting terms given as

$$W^i(x) = \frac{W_E^i(x)}{\sum_{i=1}^{N} W_E^i(x)} \cdot \frac{W_S^i(I^i(x))}{\sum_{i=1}^{N} W_S^i(I^i(x))}, \tag{9}$$

where W_E and W_S denote structure consistency and saturation weighting terms, respectively.

Structural Consistency Assessment In images taken at different times, each image may have different foreground content. Therefore, it is useful to let the user select the reference image frame. We can generate HDR images with any reference image because we already know the exact information for inconsistent regions from the sparse error matrix $\mathbf{E}$. We define the structural consistency weight term as

$$\begin{aligned}W_E^i(X) &= \begin{cases} 1 & \text{, if } i = i_{\text{ref}}, \\ \exp\left(-\|\mathbf{E}_i(x)\|_2^2/\sigma\right) & \text{, otherwise,} \end{cases} \\ &\text{where } \mathbf{E}_i(x) = \mathrm{vec}\Big(E_i{}^r(x), E_i{}^g(x), E_i{}^b(x)\Big),\end{aligned} \tag{10}$$

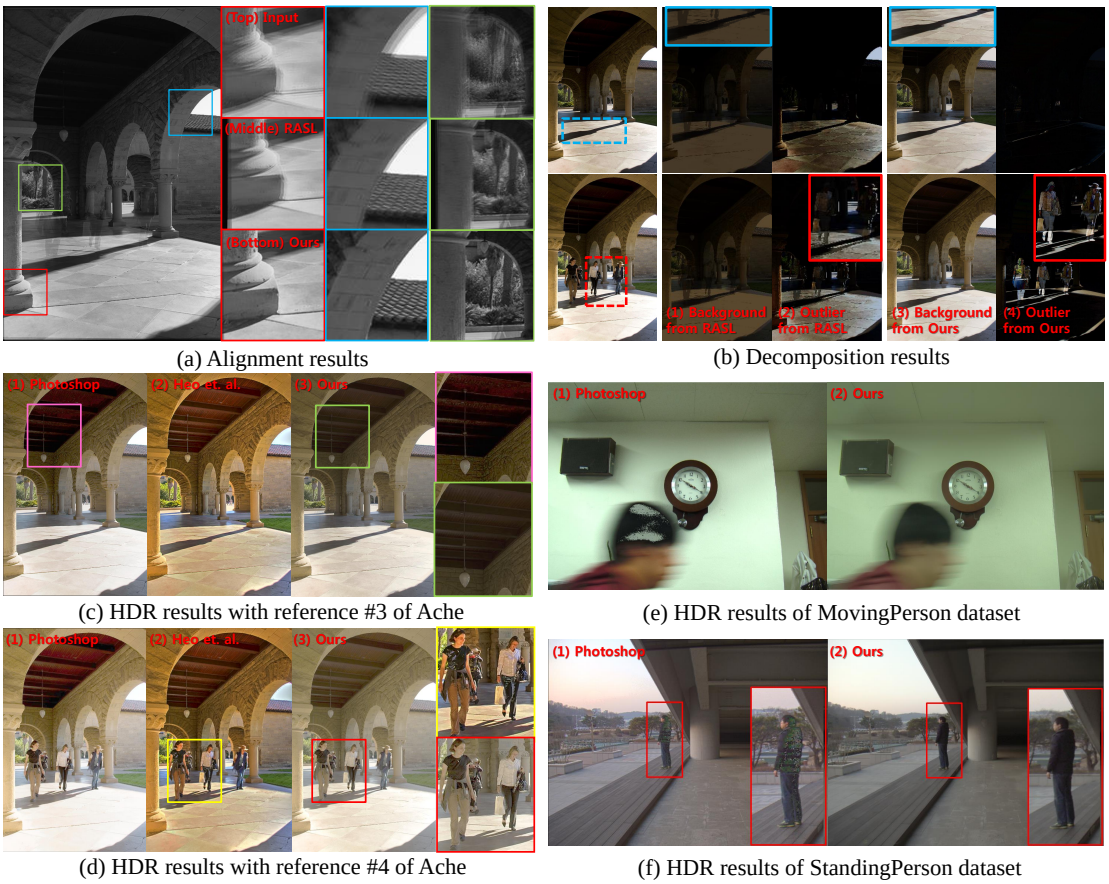
where σ is a variance and fixed to 3/255 in all experiments, i_{ref} represents the image index selected as a reference, and $E_i^c(x)$ represents the magnitude of the sparse error matrix of the c-color channel (in RGB space) at pixel location x.

Saturation Assessment The sparse error $\mathbf{E}$ is obtained based on a majority of observations. Namely, if some regions are over-exposed for more than half the input images, a low-rank system may consider the over-exposure regions as inliers. Thus, it is necessary to design another weight for penalizing the undesired saturation. We observe that a simple penalty such as $W_S(I) \approx 0$ for $I = \{0, 255\}$, $W_S(I) = 1$ otherwise is sufficient, rather than a precise weighting function. We use the weighting function in [17], which has the above property.

4. EXPERIMENTS

To show the performance of our method, we perform experiments with real datasets. We set $\lambda = 1/\sqrt{\max(m,n)}$ where m and n are row and column size of the matrix $\mathbf{O}$ during all experiments; therefore there is no manual parameter. We apply a coarse-to-fine approach with a scale pyramid to avoid local minima. The result from a coarser step is used as an initial estimate of the next finer step. We compare our results to results of Photoshop CS5 and Heo *et al*. [5]. For display, a simple gamma function is applied to the results of Photoshop and ours. For the results of Heo *et al*. [5], their tone mapping is applied.

First, we perform global alignment with the Ache dataset [4]. The dataset consists of only five images and includes moving



(a) Alignment results

(b) Decomposition results

(c) HDR results with reference #3 of Ache

(e) HDR results of MovingPerson dataset

(d) HDR results with reference #4 of Ache

(f) HDR results of StandingPerson dataset

Fig. 1. Experiment results. (a) Alignment results (Top: Unaligned Input, Middle: RASL [10], Bottom: Ours). (b) Results of decomposition to background and outliers from RASL and ours (c-d) HDR results from differently selected references of Ache dataset. (e-f) Additional HDR results of Photoshop CS5 (left) and Ours (right).

objects. We additionally add geometric transformations with a maximum of 5 degree rotation and 20 pixel translation to each image for simulating camera motion. We compare our alignment result to the result of RASL [10] and a resulting average image is shown in Fig. 1-(a). Our method estimates all the transformations accurately while RASL fails to estimate transformations due to the limited number of samples for directly applying their rank minimization approach.

Background estimations by decomposing the low-rank matrix and sparse outliers from RASL and our approach are shown in Fig. 1-(b). We apply both algorithms to each color channel independently and transform all channel results into a single canonical coordinate to align the results of RGB channels. Ideally, the decomposed background (low-rank matrix $\mathbf{A}$'s) in Fig. 1-(b)-(1,3) should have similar intensities with inputs where moving objects or saturation artifacts are removed. In Fig. 1-(b)-(1), the brightness of background from RASL has a large difference with our estimated background in Fig. 1-(b)-(3). The degenerated outliers (sparse error matrix $\mathbf{E}$'s) of RASL in Fig. 1-(b)-(2) yield dense non-zero entries that should be originally sparse. In contrast, our method shows the correct background scene in Fig. 1-(b)-(3) and successfully detects outlier regions in Fig. 1-(b)-(4).

Figs. 1-(c,d) show HDR results with differently selected references. In the results of Photoshop CS5 in Fig. 1-(c,d)-(1), under-saturated radiance is observed. The result of Heo *et al*. [5] in Fig. 1-(d)-(2) has ghost artifacts in moving object regions. This originates from the performance of the unstable moving object detection. In contrast, our results clearly reconstruct the HDR images of scenes without any artifacts. Figs. 1-(e,f) show additional results. The results from Photoshop CS5 have artifacts due to the aforementioned reason while our results properly recover HDR images including moving objects.

5. DISCUSSION AND CONCLUSION

We show that the low-rank (especially rank-1) and sparsity models offer key information to analyze dynamic scenes containing camera motion, moving objects, and saturation. By virtue of advanced optimization methods, the artifacts are effectively decomposed by our unified method without any manual parameters. The performance and robustness of the proposed method are demonstrated with real datasets. There is room to improve the perceptual quality of HDR. As future work, we will investigate preserving color balancing and include outcomes in our unified framework.

6. REFERENCES

[1] S. Mann and R. W. Picard, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proceedings of IS&T*, 1995, pp. 442–448. 1

[2] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH*, 1997. 1

[3] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from handheld exposures," *Journal of Graphics Tools*, 2003. 1

[4] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *ICCP*, 2009. 1, 3

[5] Y. S. Heo, K. M. Lee, S. U. Lee, Y. Moon, and J. Cha, "Ghost-free high dynamic range imaging," in *ACCV*, 2010. 1, 3, 4

[6] Y. Y. Boykov, O. Veksler, and R. Zabin, "Fast approximate energy minimization via graph cuts," *TPAMI*, 2001. 1

[7] T. Mitsunaga and S.K. Nayar, "Radiometric self calibration," in *CVPR*, 1999. 2

[8] J.-Y. Lee, Y. Matsushita, B. Shi, I. S. Kweon, and K. Ikeuchi, "Radiometric calibration by rank minimization," *TPAMI*, 2013. 2

[9] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: what is knowable?," *TPAMI*, 2003. 2

[10] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *TPAMI*, 2012. 2, 3, 4

[11] L. G. Brown, "A survey of image registration techniques," in *ACM Computing Surveys*, 1992. 2

[12] J. Wright, A. Ganesh, S. Rao, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *NIPS*, 2009. 2

[13] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, 2011. 2

[14] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Tech. Rep. UILU-ENG-09-2215, UIUC, 2009. 2

[15] E.T. Hale, Wotao Yin, and Yin Zhang, "Fixed-point continuation for l_1-minimization: Methodology and convergence," *SIAM Journal on Optimization*, 2008. 3

[16] T.-H. Oh, "A novel low-rank constraint method with the sparsity model for moving object analysis," *M.S. Thesis, KAIST*, 2012. 3

[17] B. Goossens, H. Luong, J. Aelterman, A. Pizurica, and W. Philips, "Reconstruction of High Dynamic Range images with Poisson noise modeling and integrated denoising," in *ICIP*, 2011. 3