

Single Shot HDR Imaging via Compressed Sensing

Matthew A. McCready

Abstract—Single-shot high-dynamic-range (HDR) imaging can be performed using a spatially varying exposure (SVE). The SVE image is typically decoded into an HDR image using neural networks or standard interpolation with some information discarded. These techniques can be undesirable due to inflexibility in the case of neural net approaches, and loss of high frequency detail in interpolation approaches. Here we propose and demonstrate a method for SVE based single-shot HDR imaging using compressed sensing techniques as a decoder. This method discards no information and can be used in conjunction with any existing exposure fusion technique. We will see that this method can be used to accurately compute a busy HDR scene from a single-shot image with PSNRs of up to 28 dB.

Index Terms—Computational Photography, HDR, SVE, Single-Shot HDR, Compressed Sensing

1 INTRODUCTION

MANY scenes of interest exhibit a dynamic range beyond the capabilities of typical camera hardware to fully measure. With high-dynamic-range (HDR) imaging techniques, one can expand the measurement bit depth to fully encapsulate the details of the scene through modifications to hardware and/or software. The classic implementation of HDR imaging is multiple exposure fusion (MEF) where a set of low-dynamic-range (LDR) images of the same scene are acquired with different exposures and systematically combined based on metrics such as well-exposedness to produce an HDR image. However, MEF often requires multiple time consuming long-exposure images for poorly lit details. MEF also suffers from ghosting and misalignment artifacts should the camera or scene move at all between LDR images, which can occur even with a stable tripod.

A better solution would allow for capture of HDR images in a single camera shot. Such a method would reduce the total capture time to that of the maximum exposure required. Furthermore, there would be no misalignment artifacts in the image. This has been implemented in the past, often by introducing a spatially varying pixel exposure (SVE) (Figure 1). In this work we propose to decode a single grayscale SVE masked image using compressed sensing to produce multiple exposure images for input to traditional MEF.

2 RELATED WORK

2.1 Single-shot HDR imaging

Single shot HDR often relies upon the use of SVE to encode additional exposure information about the scene into a single image. To implement this one typically uses either a physical neutral density filter mask, spatially varying ISO, or direct spatial variation of exposure. These approaches use a variety of decoding techniques, some depending on the

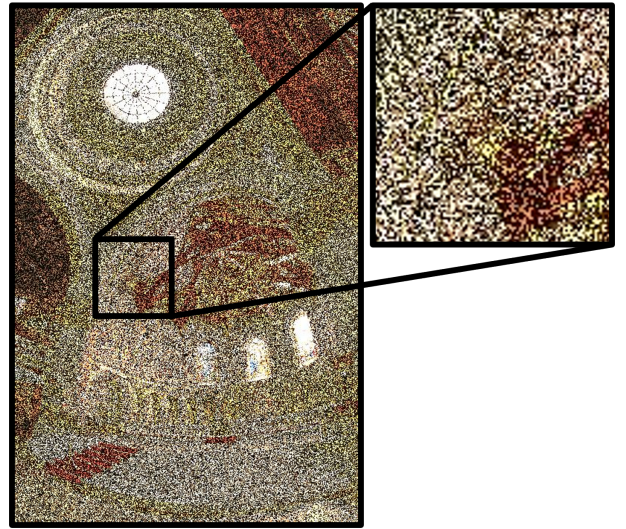


Fig. 1. Example SVE input image (simulated). Each pixel is randomly chosen to be one of three unique exposures with uniform sampling.

choice of mask. A previous EE 367 final project optimized their SVE encoding for the specific scene to be imaged, however this is not true single-shot HDR as it relies on some prior knowledge of the scene from metering or a previous frame in HDR video (1). Neural networks have been used for hallucination of HDR images from single saturated LDR images, however this is not always reliable to faithfully recreate the true scene (2). Neural networks have also been used to decode images taken with SVE or other optical filters (3,4). These methods can be end-to-end optimized and perform well but are specific to a chosen or optimized mask/filter and as with all neural networks, could be unpredictable when presented with unfamiliar data. Classic single-shot HDR imaging using SVE implements interpolation over unsaturated significant pixels (5). However, interpolation may struggle to retail high frequency details of the image, in particular if the resulting

• M. McCready is with the Department of Electrical Engineering, Stanford University, Stanford, CA, 94305.
E-mail: see <https://profiles.stanford.edu/matt-mccready>

sampling is below the Nyquist rate.

2.2 Compressed sensing

It has been shown that compressed sensing outperforms classical interpolation of randomly undersampled signals in regions of rapid change (6). In images this would correspond to regions with sharp edges in the scene. This is possible if a domain exists where the data can be represented sparsely, and the measurement and transformation matrices are incoherent with respect to one another.

3 PROPOSED METHODS

This work seeks to accomplish single shot HDR imaging by formulating N compressed sensing problems from an image taken with N unique SVE values. The resulting N LDR images are then treated as a typical MEF input image set to produce the final HDR image (Figure 2). The project will be implemented entirely through simulation, with sets of true MEF input images sampled via the chosen grayscale mask and patchworked together into the “single-shot” image.

3.1 Formulation of compressed sensing problem

Given our known SVE mask of N unique exposures, we will separate the input image into N undersampled images. For a given undersampled image b of K measured pixels, we will attempt to reconstruct the full image by solving the underdetermined inverse problem:

$$b = Cx \quad (1)$$

for the full image x of $M > K$ pixels, where C is the measurement or sampling matrix of size K by M . The process is carried out on each of the N under sampled images. We then combine the N fully sampled approximations through MEF techniques to produce an approximate HDR image.

Since the inverse problem is underspecified, we must include some regularizer based on prior knowledge of what the solution should look like. We will consider three different regularized solutions to this problem based on basis

pursuit denoising (BPDN), a denoising convolutional neural network (DnCNN), and anisotropic total variation (TV).

3.1.1 Basis pursuit denoising

BPDN works by solving for the full image in a basis where it is known to be sparse. In this implementation we will use the Fourier basis, where we now solve the inverse problem:

$$b = C\mathcal{F}^{-1}s \quad (2)$$

$$x = \mathcal{F}^{-1}s \quad (3)$$

for the image’s Fourier transform s , where $\mathcal{F}^{-1}$ is the inverse discrete Fourier transform (DFT) matrix, C is our measurement matrix, and x is the full LDR image. Given that the Fourier transform of natural images is known to be sparse, we can solve this problem for the sparsest s by minimizing:

$$\min_s \frac{1}{2} \|C\mathcal{F}^{-1}s - b\|_2^2 + \lambda \|s\|_1 \quad (4)$$

The resulting s can then be forward Fourier transformed to return an approximation of the fully sampled LDR image.

This formulation is solved by two methods. The first is to naively apply the Adam solver in PyTorch as a general approach. The second is to implement the Alternating Direction Method of Multipliers (ADMM) to minimize this convex objective (7). The update equations for the BPDN formulation are then:

$$s^{k+1} = (\mathcal{F}C^T C\mathcal{F}^{-1} + \rho)^{-1}(\mathcal{F}C^T b + \rho(z^k - u^k)) \quad (5)$$

$$z^{k+1} = S_{\lambda/\rho}(s^{k+1} + u^k) \quad (6)$$

$$u^{k+1} = u^k + s^{k+1} - z^{k+1} \quad (7)$$

Where we have used the fact that the inverse DFT matrix is unitary, so its conjugate transpose is the forward DFT matrix. Here, the operator $S_{\lambda/\rho}$ refers to the element-wise soft thresholding operator.

Implementation of this solution in Python uses the Fast Fourier Transform algorithms and LinearOperators rather than the DFT matrix directly. Note that in the code implementation and results of this work the BPDN formulation is also referred to as the L1 solution. Results of this work use a λ of 5×10^{-4} for the Adam solver and λ and ρ values of 1.0 for the ADMM solver.

3.1.2 Denoising convolutional neural network

If we use the prior knowledge that a natural image should not be noisy, and reformulate the problem in the primal (image) domain with a DnCNN regularization we derive the ADMM updates as:

$$x^{k+1} = (C^T C + \rho)^{-1}(C^T b + \rho(z^k - u^k)) \quad (8)$$

$$z^{k+1} = \mathcal{D}(x^{k+1} + u^k, \sigma^2 = \lambda/\rho) \quad (9)$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1} \quad (10)$$

where $\mathcal{D}$ represents application of our DnCNN. In our implementation, the DnCNN used is pre-trained using the methods of Zhang et al (8). The results of this work use a λ of 1.0 and a ρ of 1×10^{-4} .

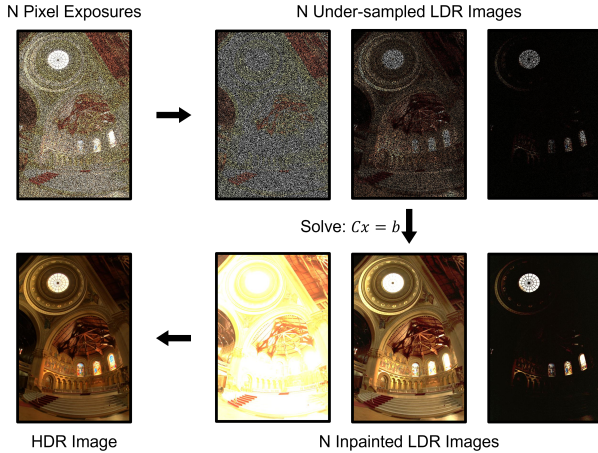


Fig. 2. Single-shot HDR imaging pipeline via compressed sensing. The SVE of N unique exposures is separated into N undersampled LDR images. The N full LDR images are inpainted as a compressed sensing problem and fused to create an HDR image.

3.1.3 Anisotropic total variation

Another useful image prior is that natural images have sparse gradients. To enforce this in our solution an anisotropic TV regularizer is included in the problem formulation.

$$\min_x \frac{1}{2} \|Cx - b\|_2^2 + \lambda \|Dx\|_1 \quad (11)$$

$$D = \begin{bmatrix} \nabla_x \\ \nabla_y \end{bmatrix} \quad (12)$$

If we apply ADMM to solve this problem our update equations become:

$$x^{k+1} = (C^T C + \rho)^{-1} (C^T b + \rho(z^k - u^k)) \quad (13)$$

$$z^{k+1} = S_{\lambda/\rho}(Dx^{k+1} + u^k) \quad (14)$$

$$u^{k+1} = u^k + Dx^{k+1} - z^{k+1} \quad (15)$$

noting that u and $z \in \mathbb{R}^{2M}$ where M is the number of pixels in the full image x . The results of this work use a λ of 2×10^{-3} and a ρ of 1.0.

3.1.4 Initial solution guesses

While introduction of a single-shot HDR method will decrease the time required for image capture, the computation time will greatly increase due to inclusion of iterative methods to solve the compressed sensing problems. It is therefore advantageous to select an initial guess at the solution x_0 that can be used to “skip” a large number of the early iterations.

Four possible initial conditions are considered in this work. The first two of these are naively initializing the solution as all zeros or initializing it as a constant half-intensity value. That is, if the measurements are scaled to the range $[0, 1]$, the initial guess would be all 0.5s. The second two conditions attempt to use some initial knowledge from our measurements to guess towards the solution. We can initialize the solution by randomly sampling from a gaussian distribution of the standard deviation and mean values given by our measurements. In a more advanced technique, we can apply a zero-order hold to the measurements to provide a filled in initial guess.

It is important to note that these initial guesses are not set directly as x prior to ADMM, as the first update of x overwrites the initial choice making it arbitrary. Instead, z is set to the initial guess prior to the first updates and u is initialized as all zeros. In the case of the Adam solver, the initial guess can be set directly as x . Note also that when applying BPDN the initial guesses are the two-dimensional Fourier transforms of those mentioned here. The convergences of the solvers with these initial guesses are compared in later sections.

3.2 Creating and tonemapping the HDR image

Once we have solved for the N inpainted LDR images we can apply standard MEF techniques to generate and tonemap the HDR image. In this work we will consider two techniques for doing this; Debevec et al’s method for HDR computation (9), and Mertens et al’s “exposure fusion” technique for direct generation of a high-quality LDR image (10).

In implementing Debevec’s MEF technique, we calculate a weighting for each pixel in our LDR image set, based

on the well-exposedness of said pixel. These weights are normalized across the image set, and the images are fused according to:

$$w_k = \exp \left(-4 \frac{(I_{lin_k} - 0.5)^2}{0.5^2} \right) \quad (16)$$

$$\text{HDR} = \exp \left(\frac{\sum_k w_k (\log(I_{lin_k}) - \log(t_k))}{\sum_k w_k} \right) \quad (17)$$

Here, I_{lin} refers to the linearized LDR image scaled to the range $[0, 1]$ obtained by inverting the camera response function and t_k is the exposure for image k . To view this HDR image a simple scaling and gamma were applied for tonemapping.

$$\text{HDR}_{\text{LDR}} = (s \times \text{HDR})^\gamma \quad (18)$$

Merten’s exposure fusion method is implemented in the cv2 library for Python. Rather than directly calculating the HDR image based on well-exposedness as Debevec’s MEF does, exposure fusion calculates its weights based on three quality metrics. Contrast calculated via a Laplacian mask, saturation calculated as the standard deviation of a pixel across colour channels, and finally well-exposedness as calculated in Debevec’s method. The product of these metrics is used as a weight for image fusion and normalized at each pixel across the image set. A weighted sum of Gaussian pyramid decomposed weights and Laplacian pyramid decomposed LDR images is then collapsed to give a high-quality LDR image. In this work, exposure fusion results were further tone mapped with a gamma correction of $(1/1.2)$ as it was found that this produced more pleasing images. Any PSNR calculations were made prior to this gamma correction.

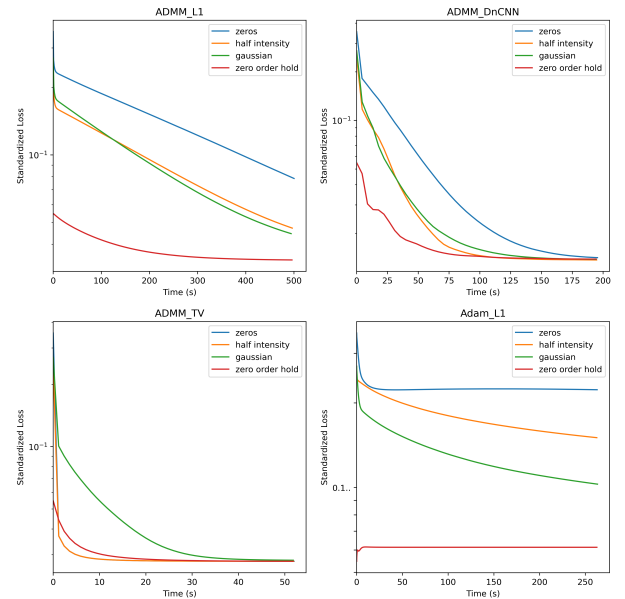


Fig. 3. Summed absolute difference per pixel between inpainted and true LDR images vs time for each solver and initial guess combination. Top-left ADMM_L1 shows ADMM implementation of BPDN. Top-right ADMM_DnCNN and bottom-left ADMM_TV show ADMM implementations of DnCNN and anisotropic TV regularizers respectively. Bottom-left Adam_L1 shows Adam solver applied to BPDN.

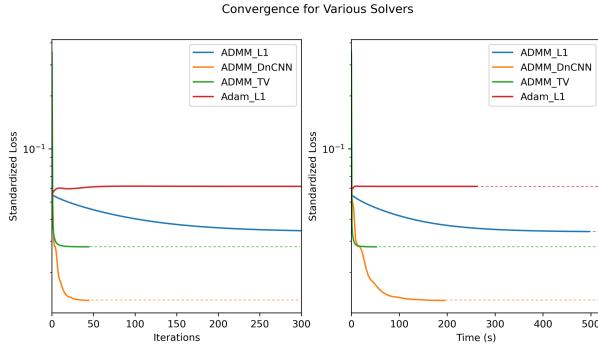


Fig. 4. Standardized loss vs optimization iteration (left) and time (right) for solvers using their fastest initial guesses. L1 solvers refer to the BPDN problem.

4 EXPERIMENTAL RESULTS AND DISCUSSION

4.1 LDR inpainting via compressed sensing

We first compare the convergence of our four compressed sensing solutions using the various initial conditions. Rather than computing convergence based on the residual $\|Cx - b\|_2$ a standardized loss is computed from the known full LDR image as the summed absolute difference per pixel. From the convergence plots of (Figure 3) for inpainting an $N = 3$ compression, we find that the BPDN (L1) and DnCNN solvers achieve their fastest convergence with the zero-order hold initial condition. However, the anisotropic TV solver converges fastest with constant initial conditions. The TV solution performs worst with the Gaussian noise initial condition, possibly due to amplification of this noise in the gradient calculations.

The fastest solutions for each of the four solvers are shown together in (Figure 4). We can see that the DnCNN solver achieves the highest degree of convergence, with TV converging by far the fastest. Based on the loss vs iterations graph, it is possible that the DnCNN solution could converge in time closer to the TV solution if a GPU was used for its calculations. We can see that the Adam solver is unable to converge further than the zero-order hold initial condition, instead working to enforce sparsity in the Fourier domain to create a more natural but poor image. This is not unexpected as the Adam solver is very generalized and therefore exhibits poor convergence. We can also see

in (Figure 3) that when initialized with zeros it encounters some local minimum with very poor convergence.

Looking directly at the inpainted LDR results (Figure 5), we can see that denoising gives the highest PSNR at 29.63 dB and was able to cleanly capture many of the finer details of the scene including etchings on the columns and the various artworks on the walls and ceiling. TV produced a strong result for much lower computational resources, however the image is slightly blurred with finer details lost. The BPDN solution using ADMM recreates the scene well but with significant noise. It is possible to improve this result with denoising techniques, but this has been set aside in the interest of a narrower scope. The Adam implementation of BPDN does not converge well with very high noise and blurring.

As the ADMM DnCNN solver was able to recover the $N = 3$ result in high detail, it was further tested on higher compression measurements. A $N = 5$ compression result is presented in (Figure 6) covering a segment of the full image for easier comparison of detail. Here we can see that ADMM DnCNN is able to accurately recover the general features of the scene, though we begin to lose high frequency details such as those in the angel paintings and the designs around the window. This reconstruction achieved a PSNR of 27.04 dB while the ADMM BPDN solver achieved 21.78 dB, the TV solver 23.22 dB, and the Adam BPDN solver 21.17 dB.

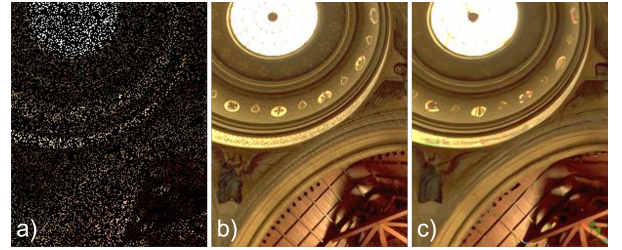


Fig. 6. LDR inpainting results for the DnCNN solver with a compression of $N = 5$. a) The undersampled LDR input ($N = 5$). b) The ground truth LDR image. c) The inpainted DnCNN result exhibiting a PSNR of 27.04 dB.

4.2 HDR image results

The HDR image fusions performed similarly to the LDR inpaintings they were constructed from. The HDR images

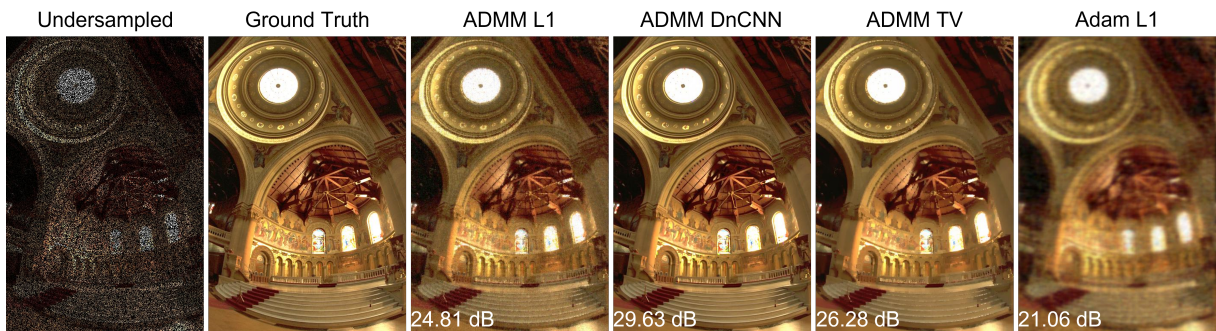


Fig. 5. LDR inpainting results for $N = 3$ compression. To the left we see the undersampled input and the ground truth image. We then have from left to right the ADMM BPDN, DnCNN, anisotropic TV, and Adam BPDN solutions. These performed with PSNRs of 24.81, 29.63, 26.28, and 21.06 dB respectively as compared to the ground truth image.

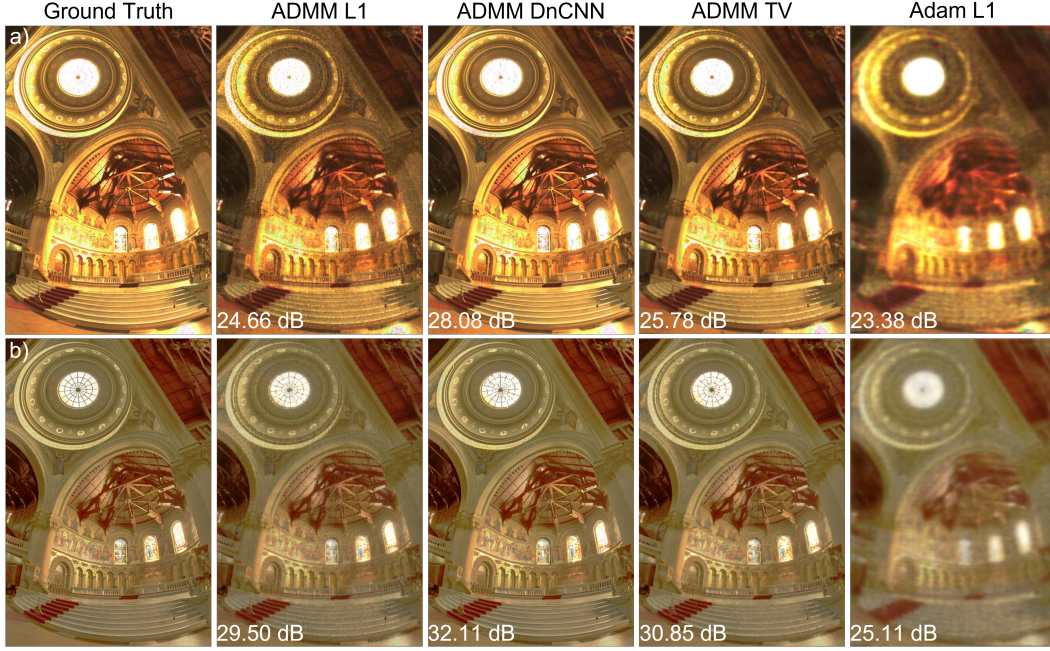


Fig. 7. HDR reconstruction results for $N = 3$ exposure single-shot SVE input. a) HDR fusions using Debevec's MEF technique tonemapped with equation (18) with s and γ values of 0.7 and 0.5 respectively. b) High-quality LDR image fusions using Merten's exposure fusion technique gamma corrected with a value of $(1/1.2)$

computed following the methods of Debevec's MEF were tone mapped according to equation (18) with scaling and gamma values of 0.7 and 0.5 respectively and can be found in (Figure 7a). The high-quality LDR images generated through Merten's exposure fusion were gamma corrected with a value of $(1/1.2)$ and can be found in (Figure 7b). The Merten's results provide a slightly better tonemapping than Debevec's likely due to the use of only three exposures. It is possible that a small set of exposures could have greater relative deviation from linearity than a large set even after inverting the camera response function. This would affect Debevec's method, but not Merten's as it does not require linearized images.

In both Debevec and Merten's methods we can see that the DnCNN approach gives the highest PSNR result with values of 28.08 dB and 32.11 dB respectively. We can see from looking at any of the solutions that a true HDR scene has been recovered given the captured window frames and upper left-side rafters. We can therefore say that single-shot HDR imaging is possible with this technique. These results were constructed using $N = 3$ exposures, but as was shown in the previous section the DnCNN approach can accurately recreate LDR scenes from higher compressions. A comparison of HDR scenes recreated using $N = 3$ and 4 exposures through the combination of the DnCNN solver with Merten's exposure fusion is shown in (Figure 8). The four-exposure result provides more depth to the scene as we can see by comparing the shading of the angel paintings, column details, and statue lighting. However, this comes at a cost to high frequency details as we should expect given the increased compression in the LDR inpainting problem. We can see this loss manifested in the ceiling window frame, railing and its reflection on the altar, and aliasing of the floor vents.

Finally, the ADMM DnCNN solver was demonstrated on additional $N = 3$ exposure sets obtained from Merianos et al (11). Merten's exposure fusion was used once more, and the single-shot results yielded PSNRs of 30.90 dB and 30.53 dB (Figure 9). We can see that the differences between the three-shot and single-shot HDR images are almost imperceptible in these natural scenes. Only some very slight high frequency details on leaves or the rockface are lost or softened.

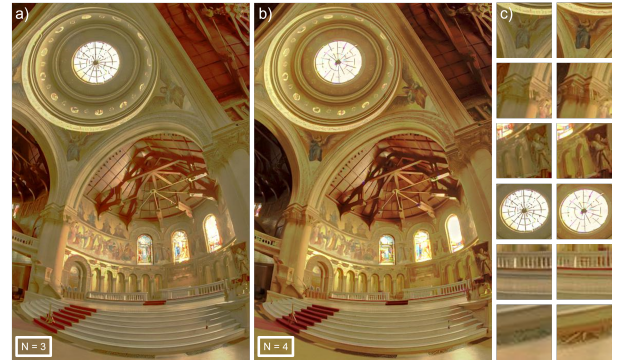


Fig. 8. Comparison of $N = 3$ exposure and $N = 4$ exposure single-shot SVE images using Merten's exposure fusion. a) $N = 3$ exposure result. b) $N = 4$ exposure result. c) Comparison of details in 3 exposure (left column) and 4 exposure (right column) results. We see increased depth of the 4 exposure result in the angel painting, column details, and statue while the high frequency details deteriorate in the window frame, railing, and floor vent.

5 CONCLUSIONS AND FUTURE WORK

Single-shot HDR significantly reduces acquisition time for HDR scenes as compared with standard MEF techniques,

and removes any potential for misalignment artifacts. We have shown that a single SVE input image of N exposures can be separated into N compressed sensing problems to reconstruct the full LDR scenes of each exposure. We have shown that these problems can be solved accurately by use of a DnCNN regularizer in particular, and that they can be solved rapidly with a TV regularizer. We have shown that the choice of initial solutions can greatly impact the time needed for convergence. Finally, we have shown that a true and accurate HDR scene can be recovered from a single SVE image via compressed sensing.

In the future, this work could be extended and improved upon in a number of ways. The current method requires the solving of N compressed sensing problems each of which is compressed by factor N . The standard interpolation method for single-shot HDR with SVE discards over- or underexposed pixels, then linearizes and normalizes the pixels by their exposure times before interpolating to recover the HDR image directly. The same approach could be taken with our technique by replacing the interpolation step with compressed sensing. This would require solving only a single compressed sensing problem and with a much lower compression factor. This was attempted during the project but could not be made to operate correctly in time for submission. However, this technique can run into issues in regions where all pixels are over- or underexposed, leaving an entire region unsampled. Another improvement that could be made to this work is careful design of the SVE mask. The mask could be end-to-end optimized with the decoding technique as it has been for neural network decoders. A mask could also be designed to allow individual pixels to depend on more than one point in the scene. With the current method, a change in a pixel which is unmeasured for compressed sensing problem k will not be reflected in the solution to that problem.

ACKNOWLEDGMENTS

The author would like to thank Manu Gopakumar for mentoring on this project, as well as Dr. Gordon Wetzstien

and Mark Nishimura for an excellent learning experience in EE 367.

REFERENCES

- [1] Wang H, Jiang J, Lyu Z. One Shot High Dynamic Range Photography Based on Hardware-level Tunable Mask. *EE367 Comput. Imaging* 2018;22.
- [2] Eilertsen G, Kronander J, Denes G, Mantiuk RK, Unger J. HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.* 2017;36:1–15 doi: 10.1145/3130800.3130816.
- [3] Metzler CA, Ikoma H, Peng Y, Wetzstein G. Deep Optics for Single-Shot High-Dynamic-Range Imaging. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2020:1372–1382 doi: 10.1109/CVPR42600.2020.00145.
- [4] Martel JNP, Muller LK, Carey SJ, Dudek P, Wetzstein G. Neural Sensors: Learning Pixel Exposures for HDR Imaging and Video Compressive Sensing with Programmable Sensors. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020;42:1642–1653 doi: 10.1109/TPAMI.2020.2986944.
- [5] Nayar SK, Mitsunaga T. High dynamic range imaging: spatially varying pixel exposures. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*. Vol. 1. IEEE Comput. Soc; pp. 472–479. doi: 10.1109/CVPR.2000.855857.
- [6] Usman K, Ramdhani M. Comparison of Classical Interpolation Methods and Compressive Sensing for Missing Data Reconstruction. In: *2019 IEEE International Conference on Signals and Systems (ICSigSys)*. IEEE; 2019. pp. 29–33. doi: 10.1109/ICSIGSYS.2019.8811057.
- [7] Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* 2010;3:1–122 doi: 10.1561/22000000016.
- [8] Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans. Image Process.* 2017;26:3142–3155 doi: 10.1109/TIP.2017.2662206.

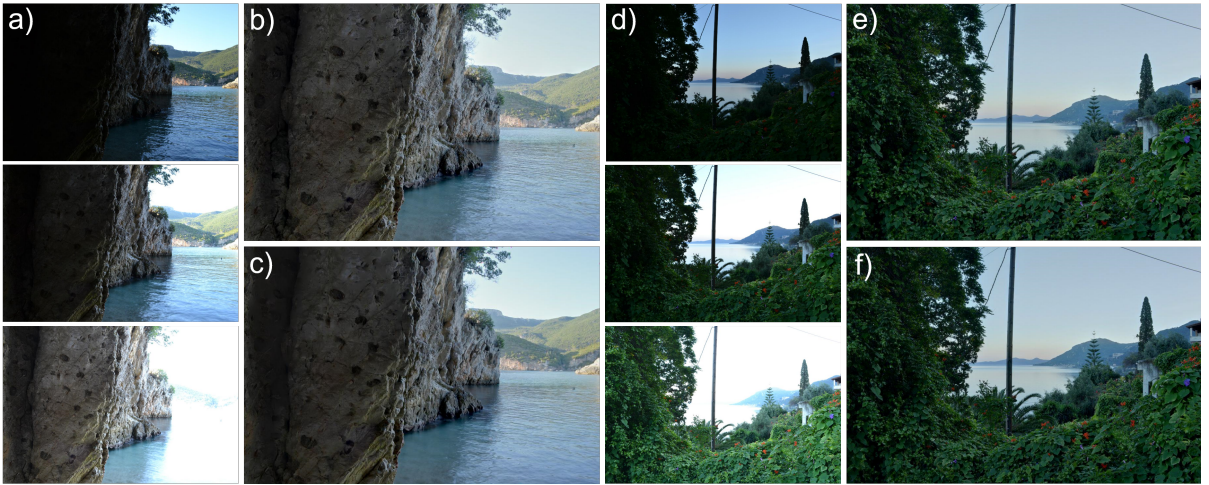


Fig. 9. Additional HDR results using $N = 3$ exposure data sets from Merianos et al (11). a) and d) MEF LDR data sets. b) and e) Ground truth exposure fusion results using the full 3 LDR images. c) and f) Single-shot SVE exposure fusion results using our technique with PSNRs of 30.90 and 30.53 respectively.

- [9] Debevec PE, Malik J. Recovering high dynamic range radiance maps from photographs. In: Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97. New York, New York, USA: ACM Press; 1997. pp. 369–378. doi: 10.1145/258734.258884.
- [10] Mertens T, Kautz J, Van Reeth F. Exposure Fusion. In: 15th Pacific Conference on Computer Graphics and Applications (PG'07). IEEE; 2007. pp. 382–390. doi: 10.1109/PG.2007.17.
- [11] Merianos I, Mitianoudis N. Multiple-Exposure Image Fusion for HDR Image Synthesis Using Learned Analysis Transformations. *J. Imaging* 2019;5:32 doi: 10.3390/jimaging5030032

ADDITIONAL COMMENTS

The source-code for this project has been made available through a github repository and can be found at: <https://github.com/mattmcstanford/CompressedSensingSingleShotHDR>