Multi-view multi-exposure stereo

Alejandro Troccoli
Columbia University
New York, NY
atroccol@cs.columbia.edu

Sing Bing Kang
Microsoft Research
Redmond, WA
SingBing.Kang@microsoft.com

Steve Seitz
University of Washington
Seattle, WA
seitz@cs.washington.edu

Abstract

Multi-view stereo algorithms typically rely on same-exposure images as inputs due to the brightness constancy assumption. While state-of-the-art depth results are excellent, they do not produce high-dynamic range textures required for high-quality view reconstruction. In this paper, we propose a technique that adapts multi-view stereo for different exposure inputs to simultaneously recover reliable dense depth and high dynamic range textures. In our technique, we use an exposure-invariant similarity statistic to establish correspondences, through which we robustly extract the camera radiometric response function and the image exposures. This enables us to then convert all images to radiance space and selectively use the radiance data for dense depth and high dynamic range texture recovery. We show results for synthetic and real scenes.

1. Introduction

Most people capture images with their cameras set to auto-exposure—and without the use of tripods. Developing vision algorithms that can cope with and even exploit the properties of such differently exposed images is therefore an important practical problem. In this paper we address the problem of computing stereo depth maps and high dynamic range (HDR) textures from images taken at different viewpoints and with different exposures, and its application for high-quality view synthesis.

Multi-view stereo algorithms traditionally make use of the brightness constancy assumption in computing correspondences (with notable exceptions including [2, 10, 17, 7]). Similarly, most methods for computing HDR images from regular photographs assume a (nearly) stationary camera and a static scene [4, 13, 14]. An approach to produce HDR video from interleaved long and short exposure frames was described in [9], but it relies on registration techniques that would not work well in the presence of large motion.

In addition to extracting depth and HDR texture, we also estimate the radiometric camera response curve and the exposure settings of the camera[1]. We only assume the camera intrinsic and extrinsic parameters are known.

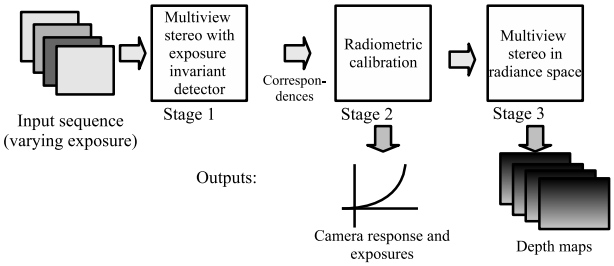

Figure 1. Overview of multi-view multi-exposure stereo.

2. Overview of our approach

The flow of our approach is depicted in Figure 1. The inputs are a set of images (different exposures) and their corresponding camera matrices. The final output is a depth map and HDR texture for each view, the camera response function, and the relative exposure values. First, we apply multi-view stereo using an exposure invariant feature detector similarity statistic to establish reliable correspondences. We use these correspondences to compute the camera response and relative exposures, after which all image intensities are converted to radiance values. We again run multi-view stereo, this time in radiance space, to extract dense depth maps (one per viewpoint).

Our main contribution is the merging of radiometric calibration techniques with a multi-view stereo algorithm to handle multiple exposures and output useful dense depth and HDR textures. In doing so, we have addressed some key issues that have not been looked at carefully in the past. First, we show that normalized cross correlation is invariant to exposure changes that can be characterized linearly or by gamma (or exponential) functions. In addition, we introduce a weighting scheme to robustly extract the radiometric

[1]The EXIF tags, if available, should contain the exposure values, but they are only estimates.

Proceedings of the Third International Symposium on
3D Data Processing, Visualization, and Transmission (3DPVT'06)
0-7695-2825-2/06 $20.00 © 2006 IEEE

curve. We also show that applying multi-view stereo on a common radiance space yields better depth estimates. Once we have computed a set of view-dependent depth maps and radiometrically calibrated the camera, we can then view interpolate in both (3D) space and exposure.

3. Previous work

Stereo is one of the most studied areas in computer vision (an excellent survey on 2-view stereo is given in [16]). In this section, we limit our discussion to approaches that extract HDR through motion estimation or stereo. The approaches closest to ours are those of Kang *et al*. [9] and Kim and Pollefeys [10].

In Kang *et al*.'s system [9], alternating long and short exposures were captured by their camera system. They precalibrate their camera so that both the exposure values and radiometric response curve are known *a priori*. To compute HDR for a given frame, its immediate neighbors are registered to it. Registration is facilitated by synthetically mapping lower exposures to higher exposures to more closely match intensity distributions. Artifacts are highly visible in areas of large motions (where registration fails). There are two key differences with our method: we do not precalibrate the camera, and we use all available views. The latter is possible for us because we assume static scenes, and Kang *et al*. necessarily limit the search because both camera and scene can be dynamic.

Kim and Pollefeys construct a 3D model from a video sequence under auto-gain [10]. They use correspondences across the input frames with varying gain to compute the camera response function and relative exposure ratios between frames, which are then used to remap color intensities in the final textured model. We build on their technique to estimate intensity correspondences and response curves, but our method is different in the following ways: we use all possible pairs of input images to obtain intensity correspondences, not just pairs of images that are consecutive in the input sequence; we use a different weighting in the formulation of the weighted least squares problem that accounts for the fact that the fit is computed in log-space instead of linear space; we compute view-dependent depth maps; and finally, in our approach, once we know the camera response and exposure ratios, we perform matching in radiance space to obtain better results.

The problem of radiometric calibration from a sequence of images has been addressed in various approaches [4, 14, 12, 6, 10, 11]. All of these methods use brightness correspondences between images acquired with different exposures to compute the camera response. These brightness correspondences define a brightness transfer function (BTF) [6]. For a pair of images (I_i, I_j), the BTF $T_{i,j}(B_i) = B_j$ states that a brightness value B_i in image I_i maps to B_j in I_j. Under reasonable assumptions, T is semi-monotonic, $T(0) = 0$ and depending on the exposure ratio $e_{i,j} = e_j/e_i$, $T(B) \leq B$ if $e_{i,j} \leq 1$ or $T(B) \geq B$, otherwise. For images acquired by a static camera, the BTF is typically computed from the brightness values of collocated pixels. It can also be computed from the histograms of the images, as shown by Grossberg and Nayar in [6]. This histogram based technique can also be applied to a moving camera provided the scene radiance stays almost constant. Such a constraint restricts the camera movement and is not desirable for our application. In addition, theoretical limitations on what is know about the camera response from an image sequence are enumerated in Grossberg and Nayar [6], who observed that it is not possible to recover the camera response and exposure ratios simultaneously and unambiguously from the BTF alone, without further assumptions. Therefore, most techniques require some knowledge about the exposures. In the context of our application, we will see later that this ambiguity will require us to either provide some information about the exposures or fix a point in the response curve. In addition, Grossberg and Nayar [15] developed an empirical model of camera response (EMOR). This model provides a low dimensional set of basis functions for describing the space of camera responses, which we use in our approach.

4. Multi-view multi exposure stereo

The inputs to our system are a collection of images $\mathcal{I} = \{I_k(x, y), k = 0...K\}$ acquired by a camera with unknown response function f and unknown exposures $\mathcal{E} = \{e_0, e_1, ..., e_K\}$, and their corresponding camera matrices $\{P_k, k = 0...K\}$. Our goal is to recover f, exposures $\mathcal{E}$, and depth maps $d_k(x, y)$, $k = 0...K$. To achieve our goal, we combine the multi-view stereo algorithm of Kang and Szeliski [8] (part of which is based on the Collins' plane sweep algorithm [3]) with the robust radiometric calibration technique of Kim and Pollefeys [10].

We first obtain a dense set of correspondences by running the multi-view stereo using normalized cross correlation (NCC) as the similarity statistic. We select, without loss of generality, I_0 as a reference image and find the depth map $d_0(x, y)$. From these correspondences we solve for f and $\mathcal{E}$, after which all input image color intensities are mapped to radiance space. Finally, we run the multi-view stereo algorithm again, this time on the radiance space set of images using sum of squared differences as the similarity statistic. We run the algorithm K times, on each time using a different image as the reference, to compute the rest of $d_k(x, y)$s.

4.1. Cross correlation is (approximately) exposure invariant

It turns out that normalized cross correlation (NCC) is invariant to exposure changes when a camera has a gamma

Proceedings of the Third International Symposium on
3D Data Processing, Visualization, and Transmission (3DPVT'06)
0-7695-2825-2/06 $20.00 © 2006 IEEE

response function. (This includes the special linear case where gamma equals one.) *Interestingly, while NCC has been used in the past, its exposure invariance property has not been observed nor proven (that we are aware of).*

Consider a window around a pixel (x, y) over two images I_0 and I_k. The normalized cross-correlation is defined as

$$E_{NCC}(I_0, I_k) = \frac{\sum(I_0(u,v) - \bar{I}_0)(I_k(u,v) - \bar{I}_k)}{\sqrt{\sum(I_0(u,v) - \bar{I}_0)^2}\sqrt{\sum(I_k(u,v) - \bar{I}_k)^2}}, \quad (1)$$

where (u, v) are coordinates relative to the window and $\bar{I}$ is the mean value over the window. Normalized cross correlation yields a value of 1 when the pixel intensities over the window around (x, y) are in exact correspondence. If I_0 and I_k correspond to the same scene but captured with different exposures on a camera that has a gamma response, then pixel intensities are related to sensor irradiance I by

$$I_0(x, y) = (I(x, y)e_0)^{1/\gamma} \quad (2)$$

and

$$I_k(x, y) = (I(x, y)e_k)^{1/\gamma}, \quad (3)$$

from which the following relationship between recorded intensities is derived:

$$I_k(x, y) = I_0(x, y)e_{k0}^{1/\gamma}, \quad (4)$$

where e_{k0} is the exposure ratio e_k/e_0. Using this relationship, the mean over I_k is

$$\bar{I}_k = \sum_{u,v} e_{k0}^{1/\gamma} I_0(u, v) = e_{k0}^{1/\gamma} \bar{I}_0. \quad (5)$$

Substituting (5) into (1), we get

$$E_{NCC}(I_0, I_k) = \frac{\sum(I_0(u,v) - \bar{I}_0)(I_0(u,v)e_{k0}^{1/\gamma} - \bar{I}_0 e_{k0}^{1/\gamma})}{\sqrt{\sum(I_0(u,v) - \bar{I}_0)^2}\sqrt{\sum(I_0(u,v)e_{k0}^{1/\gamma} - \bar{I}_0 e_{k0}^{1/\gamma})^2}}. \quad (6)$$

Notice that the factor $e_{k0}^{1/\gamma}$ vanishes; this concludes our proof that normalized cross correlation is invariant to exposure changes if the radiometric response is exactly a gamma curve (ignoring noise and quantization effects). While the real camera response is typically not exactly a gamma curve, it is close enough that normalized cross correlation is practically invariant.

4.2. Initial correspondence estimation

For a given reference image I_k in $\mathcal{I}$, the multi-view stereo algorithm described in [8] computes the depth map $d_k(x, y)$ that maximizes photo consistency between all pixels $I(x, y)$ and their corresponding projections on the other images. For better results near depth boundaries, where stereo algorithms typically have problems due to some pixels in one view being occluded on the other views, shiftable windows and temporal view selection are used (which we implemented).

However, rather than using direct sum-of-squared-differences (SSD), we used normalized cross correlation (NCC). As shown in the previous section, NCC is invariant to exposure changes for cameras with gamma response curves. However, there is a price to pay for the invariance: ambiguity within the equivalence class of gamma response curves. To handle this problem, once the radiometric curve has been robustly estimated, we convert all color intensities to radiance space and re-apply the multi-view stereo algorithm, this time using SSD on radiance values (Section 4.4).

For color images, we compute NCC for each channel independently and then average. In computing NCC, we ignore under- and over-exposed pixels (tagged as invalid). We make an overly-conservative choice of ignoring an entire color band within the window if some pixels are under- or over-exposed in that band. If all the color bands within the window have some pixels that invalid, then we ignore that center pixel (tagged as having unreliable depth). We use a simple winner-take-all approach to extract an initial depth map for an arbitrarily chosen reference image.

4.3. Radiometric calibration

We use the depth map computed in the first stage to warp each input image to the same viewpoint, creating a new set of *stabilized* images $\hat{\mathcal{I}}$. Pixels are sampled using bicubic interpolation. This set of images is used for radiometric calibration. The view stabilization is done for convenience; all corresponding pixels now share the same (x, y) coordinate frame.

The goal of radiometric calibration is to find the mapping between the amount of light falling on the camera sensor, or sensor irradiance, and the actual brightness measurement recorded. This mapping is a non-linear function f of the scene irradiance and camera exposure given by

$$B = f(Ie). \quad (7)$$

This is the image formation equation, with B being the measured brightness, I the irradiance at the corresponding sensor pixel, and e the camera exposure. The exposure depends on the lens aperture, shutter speed, and camera gain.

To compute the BTF between a pair of images, all the required information is given in form of pairs of corresponding intensity values (B_1, B_2). This intensity correspondences can be collected in a two-variable joint histogram J, in which a given entry $J(B_1, B_2)$ stores the number of pixels that in the first image have an intensity value of B_1

Proceedings of the Third International Symposium on
3D Data Processing, Visualization, and Transmission (3DPVT'06)
0-7695-2825-2/06 $20.00 © 2006 IEEE

and map to intensity B_2 in the second image. These correspondences, however, may be incorrect due to ambiguity or mismatches using NCC. To estimate the BTF, we implemented the dynamic programming technique of Kim and Pollefeys that works on the joint histogram to find a BTF T that satisfies the properties enumerated earlier: semimonotonicity, $T(0) = 0$, and either $T(i) \leq i$ or $T(i) \geq i$. To decide which of these last two conditions should hold, we partition the joint histogram $J(B_i, B_j)$ into two triangles along the line $B_i = B_j$. We sum over all entries in each of these triangles, and compare the resulting sums; if the sum from the upper traingle is larger than the sum from the lower triangle, then $T(i) \geq i$ must hold, otherwise, $T(i) \leq i$ will be true.

In our approach, we recover *two* BTFs for each pair of images in $\hat{\mathcal{I}}$. For images $(\hat{I}_1, \hat{I}_2)$ we compute the BTFs $T(\hat{I}_1, \hat{I}_2)$ and $T(\hat{I}_2, \hat{I}_1)$. Once the BTF functions have been computed for all pairs of images, we use the EMOR model of Grossberg and Nayar [5] to find the inverse camera response, which describes an empirical basis for camera response functions. Using this model, the camera response can be written as

$$f(x) = f_0(x) + \sum_{i=1}^{N} c_n h_n(x), \tag{8}$$

where f_0 is the mean of a set of measured camera response functions and h_i form a basis obtained through PCA. A small number of basis functions, between three and five, are enough to represent the whole space of camera responses with reasonable accuracy.

To solve for the inverse camera response once the BTF is known, we invert the image formation equation (7) and take the logarithm of each side, giving

$$\log f^{-1}(B) = \log I + \log e. \tag{9}$$

For simplicity, we will use g to denote $\log f^{-1}$. For a given pair of images $(\hat{I}_i, \hat{I}_j)$ and their corresponding BTF $T_{i,j}$, the following brightness transfer constraint applies:

$$g(B_i) - g(T_{i,j}(B_i)) - \log e_i + \log e_j = 0, \tag{10}$$

with e_i and e_j being the unknown exposure values. To solve for g using the EMOR, we computed the mean and basis vectors for the space of log inverse camera response functions, to obtain the following model:

$$g(x) = g_0(x) + \sum_{n=1}^{N} c_n h'_n(x). \tag{11}$$

Replacing this model in the brightness transfer constraint results in

$$g_0(B_i) - g_0(T_{i,j}(B_i)) + \sum_{n=1}^{N} c_n (h'_n(B_i) - h'_n(T_{i,j}(B_i)) - \log e_i + \log e_j = 0. \tag{12}$$

This is the constraint that we will minimize to find the coefficients c_n of our model and the set of exposures $\mathcal{E}$. However, there is an inherent ambiguity in the radiometric calibration problem, as described in [5]. It is not possible to compute the exposures and the camera response simultaneously without any further assumptions. In our implementation, we either specify at least one exposure ratio by setting the values of two exposures e_i and e_j, or fix the point $g(128)$ in the inverse response curve to pass through $g_0(128)$. Both of these techniques will break the ambiguity, and the choice of one over the other depends on the amount of data provided.

Equation (12) is solved in a weighted least-squares fashion. In [10], a Gaussian weighting scheme with mean 128 and an empirically chosen standard deviation was used. Through experimentation with ground truth data, we have found that using this weighting mechanism does not always yield optimal results. There are several reasons that explain this observation. Firstly, as pointed by Grossberg and Nayar in [5], PCA is not optimal for the log-space model with regard to least squares in the (linear) space of camera responses. In addition, the log space model is very sensitive to variations in the lower range of pixel intensities (typically having low SNR), which can adversely affect the fitting process. The Gaussian weighting scheme only partially addresses these issues. We have found that the following weighting function, as defined by Litvinov and Schechner in [11], yields better results:

$$w(B_i, B_j) = \frac{B_i B_j}{(B_i^2 + B_j^2)^{1/2}}, \tag{13}$$

where B_i and B_j are the corresponding brightness values as defined by the BTF. We will refer to this weighting scheme as the *product weighting scheme*. These weights penalize lower brightness pixels with a low SNR and emphasizes those that have a high SNR.

4.4. Final depth map computation

After we have computed the log inverse camera response function g and the exposure values, we convert all images into radiance space and run the multi-view stereo algorithm again, this time using a modified version of sum of squared differences (SSD) to compute the matching cost. When computing the SSD for a pair of windows, we cannot just simply use the radiance values within a window—we must

also take into account the pixel intensity in the original image to make sure that we do not include radiance values that came from saturated or under-exposed pixels.

In this stage we run the multi-view stereo algorithm several times, every time setting the reference image to be a different one. The result will be a depth map for each view.

5. Experimental results

To validate our approach, we applied it on images of synthetic and real scenes. Note that all our depth maps were recovered using simple winner-take-all (taking the depth associated with the lowest error at each pixel). We expect cleaner-looking results using global optimization techniques with regularization (e.g., using graph cut[1]).

5.1. Synthetic data

We first tested our algorithm on a synthetic scene that we built with a ray-tracing package and that is shown in Figure 2. We generated images from different views and exposures by applying one of the response functions in the database used for the EMOR computation; the same response was applied to each channel.

First, we ran the radiometric calibration algorithm on a subset of those images taken from the same viewpoint but with varying exposure. In our run, we fixed two of the exposures and let the radiometric calibration find the remaining ones. Then, we repeated the experiment, this time using images from different views and the correspondences found after running multi-view stereo using NCC. The computed curves, plotted against the ground truth data, are shown in Figure 3. The left and middle graphs show the curves obtained over the fixed-viewpoint set of images (i.e., same viewpoints at different exposures as done for conventional HDR computation); the curves in the left graph were obtained using Gaussian weights and the curves in the middle with the product weighting scheme of equation (13). It can be seen that the latter scheme results in a better approximation. The rightmost graph shows the curves that were obtained when running the radiometric calibration over a set of images from different viewpoints, after warping the images to the same view using the correspondences computed in stage 1 of our system. The computed curve is a very good approximation of the ground truth data.

Figure 4 shows the depth maps obtained for this sequence. The leftmost depth map is provided as a reference and was obtained using a sequence of same exposure images. The middle is the result obtained over a sequence of images with varying exposure as obtained after stage 1, using NCC, and the rightmost is the result of running multi-view stereo once the camera response had been recovered and all images were converted to radiance space. Some pixels in the depth map are completely black, which correspond to areas in which the original reference images was either under- or over-exposed. It should be noted that the quality of the depth map varies from stage 1 (middle image) to stage 3 (right image). This shows that the radiance space multiview stereo iteration improves the results.

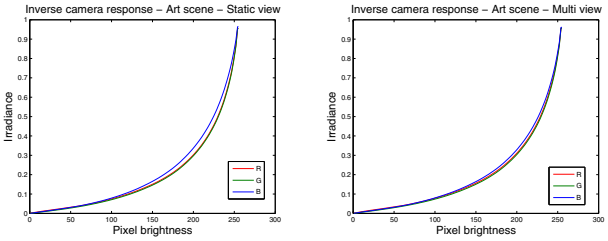

Figure 8. Radiometric calibration curves obtained from the Art sequence. Left: over a set images acquired from the same viewpoint; Right: over a set of viewpoint varying images after running stereo with NCC.

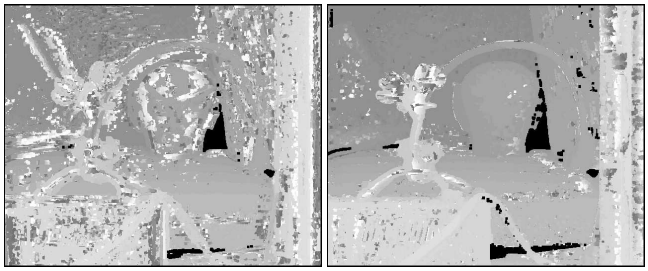

Figure 9. Depth maps computed for the Reindeer sequence. Left: computed over a set of images with varying exposure using normalized cross correlation. Right: computed after converting all images to radiance space and running multi-view stereo with sum of squared differences.

5.2. Real data

We also tested our system on sets of real images. The first one is the Art sequence, shown in Figure 5. For this sequence we used a set of 5 different views acquired by a translating camera. To create a reference depth map that could help us evaluate the results, we first ran our system on a subset of images with a fixed exposure. We then conducted a different experiment, this time allowing the exposures to vary from one view to the next. Figure 6 shows the resulting depth maps. Once more, it can be seen that there is an improvement achieved when running the multi-view stereo algorithm in radiance space using SSD. However, the final depth map obtained from the varying exposure sequence (right) shows some differences with that obtained using the sequence with fixed exposure (left). These differences occur in areas where the intensity of blue channel is very low. The computed response curves are shown in Figure 8. The left plot shows the inverse response curves obtained from three images taken from the same view; the right one shows the curves resulting from stage 2 of our

Proceedings of the Third International Symposium on
3D Data Processing, Visualization, and Transmission (3DPVT'06)
0-7695-2825-2/06 $20.00 © 2006 IEEE

Figure 2. Four images of the synthetic scene with varying view and exposure.

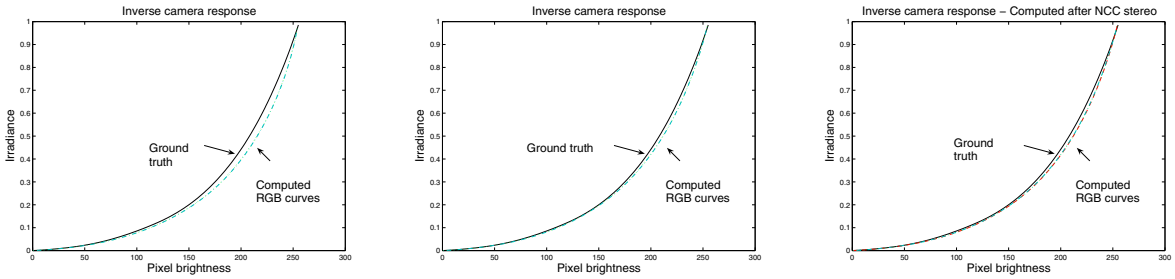

Figure 3. Computed camera inverse response functions plotted against ground truth. Left: result from an input set of images from same view, equations weighted with Gaussian curve (RGB RMS: 0.51, 0.52, 0.54). Middle: result from an input set of images from same view, using the product weighting scheme of equation 13 (RGB RMS: 0.27, 0.28, 0.29). Right: result from an input set of images from multiple views warped to the same viewpoint, using the product weighting scheme (RGB RMS: 0.30, 0.29, 0.28). Note that the left and middle curves were obtained using the simple condition of multiple exposures at the same viewpoint (correspondences are known exactly); these curves are used for comparison with our technique under a more difficult condition of multiple exposures at different viewpoints.

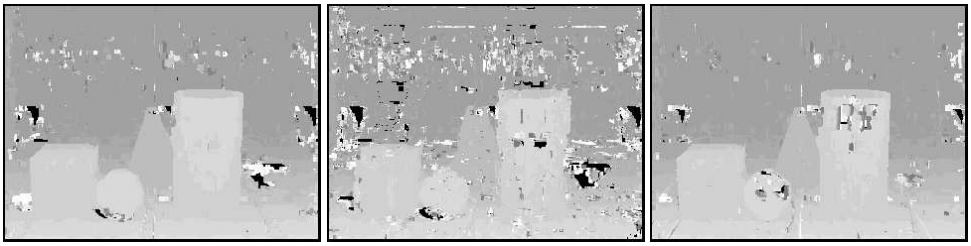

Figure 4. Depth maps obtained for the synthetic scene using simple winner-take-all. Left: computed over a set of images with same exposure using SSD. Middle: computed over a set of images with varying exposure using NCC. Right: computed after converting all images to radiance space and running multi-view stereo with SSD.

Figure 5. Four of the input images in the art sequence (courtesy of Daniel Scharstein).

multi-view system on a set of five images. Similar results were obtained from the Reindeer sequence of Figure 7 The stage 1 and stage 2 depth maps for one of the views are shown in Figure 9.

6. HDR texture computation

Once we have computed the depth maps we can warp all input images to the reference view point of the multi-view stereo stage and create and HDR texture. We compute the final output value as a weighted combination of the input

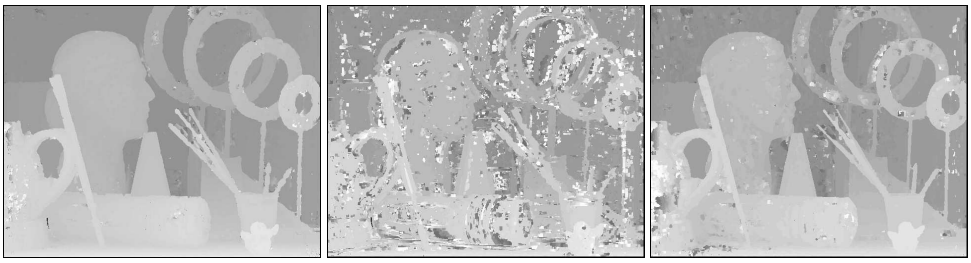

Figure 6. Depth maps computed for the Art sequence. Left: computed over a set of fixed-exposure images, Middle: computed over a set of images with varying exposure using NCC. Right: computed after converting all images to radiance space and running multi-view stereo with SSD.

Figure 7. Four of the input images in the Reindeer sequence (courtesy of Daniel Scharstein).

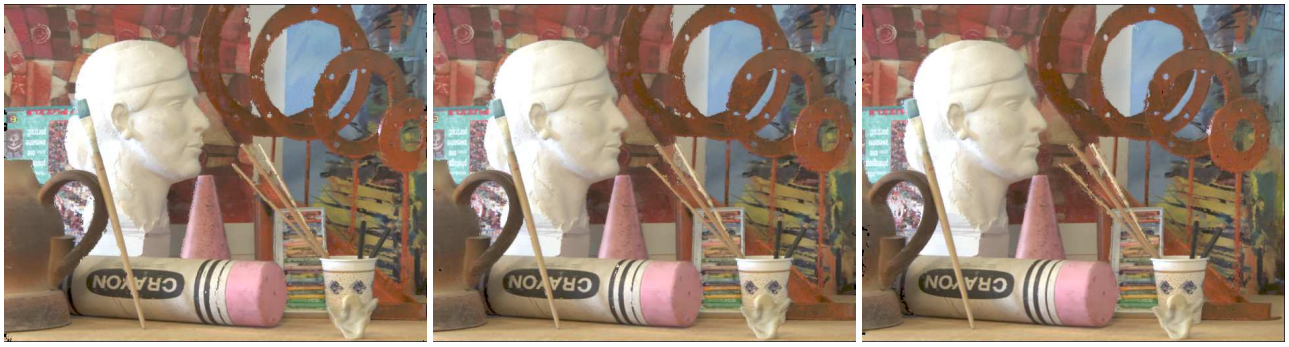

Figure 10. View interpolation results. Three interpolated views obtained from the Art sequence.

Figure 11. View interpolation results. Three interpolated views obtained from the Reindeer sequence.

pixels, using the following weighting function evaluated at the original pixel intensities:

$$w(B) = f(B)/f'(B), \tag{14}$$

where f is the camera response. This function was introduced in [14], based on the fact that image measurements can be trusted most when the signal-to-noise ratio (SNR) and sensitivity to radiance changes are maximum.

Proceedings of the Third International Symposium on
3D Data Processing, Visualization, and Transmission (3DPVT'06)
0-7695-2825-2/06 $20.00 © 2006 IEEE

7. View interpolation

From the previous steps, we have obtained an HDR image and a depth map for each input camera. With this representation we can generate new views by view interpolation. Our current implementation allows us to trace a path between two existing views, generating a new frame along the specified path at a user defined interval. The user fixes the number of frames to synthesize along the path and the exposure variation (if any, the exposure can also be kept constant). To generate a new view, we warp the closest two real views to the virtual view and then create a weighted average image. The weight used for each real view is inversely proportional to the distance between itself and the virtual view. Results of our view interpolation are shown in Figures 10 and 11. These views have been synthesized from the original Art and Reindeer sequences. The shown views have been stabilized to the same exposure. (Reviewers, please see the video sequence submitted as additional material).

8. Discussion

One problem that could be better treated relates to over- and under-exposed pixels. We currently handle these pixels by completely ignoring the offending color channel within a window if at least a pixel has that color out of range. However, this approach is very conservative, since these pixels do provide some information that can be used as constraints during the radiance space stereo matching. For example, if a certain pixel is saturated in the reference image, then that imposes a restriction on the radiance; more specifically, it cannot be below the radiance associated with the saturated value at the reference exposure. Our work currently does not take such constraints into account.

As mentioned earlier, we use the simple winner-take-all approach to extracting depth maps. We chose this to illustrate worst case scenarios, since winner-take-all is sensitive to noise. Despite this, we were able to recover good depth maps; obviously, more spatially-coherent depth maps can be obtained using global optimization techniques in conjunction with regularization (such as graph cuts [1]).

9. Conclusion

We have described a technique for estimating dense depth maps and HDR textures from a set of different exposures. Our technique consists of three stages: correspondence using normalized cross correlation (NCC), radiometric calibration to compute camera response function and exposure values, and view-dependent dense depth maps using SDD on radiance values. We proved that NCC is practically exposure-invariant, and showed view interpolation results to validate our approach.

References

[1] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In *ICCV*, pages 377–384, Sept. 1999.

[2] R. L. Carceroni and K. N. Kutulakos. Multi-view scene capture by surfel sampling. *Int. J. Comput. Vision*, 49(2):172–214, 2002.

[3] R. T. Collins. A space-sweep approach to true multi-image matching. In *CVPR*, pages 358–363, June 1996.

[4] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. of ACM SIGGRAPH '97*, pages 369–378, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[5] M. D. Grossberg and S. K. Nayar. What can be known about the radiometric response from images? In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part IV*, pages 189–205, London, UK, 2002. Springer-Verlag.

[6] M. D. Grossberg and S. K. Nayar. Determining the camera response from images: What is knowable? *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(11):1455–1467, 2003.

[7] H. Jin, S. Soatto, and A. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *Int. J. Comput. Vision*, 63(3):175–189, July 2005.

[8] S. B. Kang and R. Szeliski. Extracting view-dependent depth maps from a collection of images. *Int. J. Comput. Vision*, 58(2):139–163, 2004.

[9] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Trans. Graph.*, 22(3):319–325, 2003.

[10] S. J. Kim and M. Pollefeys. Radiometric self-alignment of image sequences. In *CVPR*, 2004.

[11] A. Litvinov and Y. Y. Schechner. Addressing radiometric nonidealities: A unified framework. In *CVPR*, pages 52–59, 2005.

[12] S. Mann. Comparametric equations with practical applications on quantigraphic image processing. *IEEE Trans. Image Proc.*, 9(8):1389–1406, August 2000. ISSN 1057-7149.

[13] S. Mann and R. Picard. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Procs. Imaging Science and Technology 46th Annual Conference*, pages 422–428, May 1995.

[14] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *CVPR*, pages 1374–1380, 1999.

[15] S. K. Nayar. Modeling the space of camera response functions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(10):1272–1282, 2004. Member-Michael D. Grossberg.

[16] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. of Computer Vision*, 47(1):7–42, May 2002.

[17] G. Vogiatzis, T. P., and R. Cipolla. Multi-view stereo via volumetric graph-cuts. In *CVPR*, volume II, pages 391–398, June 2005.

Proceedings of the Third International Symposium on
3D Data Processing, Visualization, and Transmission (3DPVT'06)
0-7695-2825-2/06 $20.00 © 2006 IEEE