Multi-View Stereo Reconstruction with High Dynamic Range Texture

Feng Lu, Xiangyang Ji, Qionghai Dai, and Guihua Er

TNList and Department of Automation, Tsinghua University, Beijing, China

Abstract. In traditional 3D model reconstruction, the texture information is captured in a certain dynamic range, which is usually insufficient for rendering under new environmental light. This paper proposes a novel approach for multi-view stereo (MVS) reconstruction of models with high dynamic range (HDR) texture. In the proposed approach, multi-view images are firstly taken with different exposure times simultaneously. Corresponding pixels in adjacent viewpoints are then extracted using a multi-projection method, to robustly recover the response function of the camera. With the response function, pixel values in the differently exposed images can be converted to the desired relative radiance values. Subsequently, geometry reconstruction and HDR texture recovering can be achieved using these values. Experimental results demonstrate that our method can recover the HDR texture for the 3D model efficiently while keep high geometry precision. With our reconstructed HDR texture model, high-quality scene re-lighting is exemplarily exhibited.

1 Introduction

3D model is created as a mathematical representation of any 3D surface of objects in the scene via specialized methodologies. It can be displayed with adequate appearance information of virtual objects through 3D rendering. Therefore, 3D model is widely used in applications such as industrial design, 3D video game, digital film-making, and advanced education. With the rapid development of computer vision and computer graphics, a lot of novel technologies for reconstructing 3D models of real-world objects are coming into sight, in which multi-view stereo (MVS) has drawn more and more attentions from researchers in recent years [17].

MVS reconstructs watertight 3D model from multiple images captured at different viewpoints. Existing MVS solutions only capture multi-view images in a fixed low dynamic range (LDR). These images are unable to present complete information of scene radiance due to limited dynamic range determined by the exposure settings. Thus, the 3D model recovered from these images only has LDR texture, which is not applicable in specified applications, e.g. whose require rendering with environmental light variation.

High dynamic range (HDR) imaging can be fulfilled via cameras with specified design [16]. However, it is too expensive and cannot be widely used in practice. As a result, most research works focused on fusing temporally captured

multi-exposure images to generate HDR images [5, 7]. In these approaches, pixel alignment for moving object is a hard work. As a result, it is usually inefficient for response function recovery and HDR image fusing. Besides, these methods cannot be directly applied to generate HDR texture in MVS scenario. Multi-camera setup throws fresh light on HDR texture recovery based on inter-view correspondence. However, as for now, no research work in terms of HDR texture based MVS has been reported.

This paper proposes a novel MVS algorithm allowing HDR texture recovery from multi-view LDR images. The approach mainly goes through three phases: multi-view multi-exposure image capture, film response recovery and image calibration, and HDR texture 3D model reconstruction, as shown in Fig. 1. As a result, without any additional hardware cost compared to traditional multi-cameras based MVS, HDR texture which contains more information of the original scene radiance can be generated.

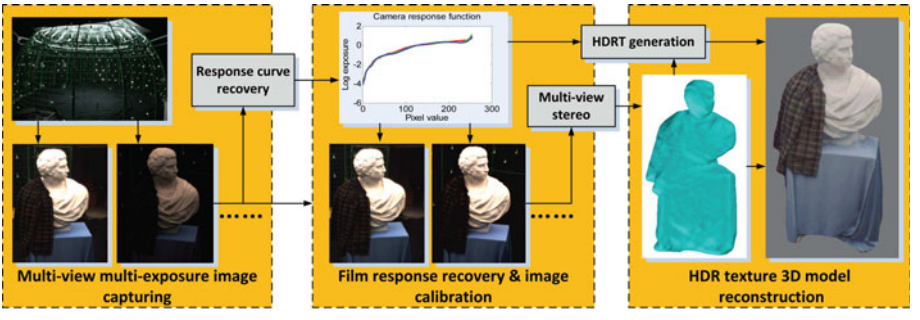

Fig. 1. Our proposed approach: the three phases to reconstruct 3D model with HDR texture

The remainder of the paper is organized as follows: Section 2 provides the background of HDR image recovery. In Section 3, HDR texture 3D model reconstruction approach will be described in detail. Section 4 presents experimental results and Section 5 concludes the paper.

2 Background

This section overviews the imaging principle of conventional LDR imaging systems. Then several related works on radiance map recovery and HDR images/ videos generation are discussed.

2.1 Scene Radiance and Camera Response Function

The "scene brightness" in a real-world scene refers to the radiance L of the object surface patches in the scene, whereas what a typical imaging system really

measures is the sensor irradiance E, which indicates the "image brightness", and is given in [4]:

$$E = L\frac{\pi}{4}(\frac{d}{f})^2 \cos^4 \alpha \tag{1}$$

where d is the diameter of the aperture, f is the focal length, and α is the angle that the optical axis makes with the ray from the object patch to the center of the lens. With the reasonable assumption that the terms of $\cos^4 \alpha$ and $(\frac{d}{f})^2$ don't vary from pixels, the image irradiance E is actually proportional to the scene radiance L.

During the process of filming, within a particular time referred as exposure time Δt, the total energy collected by the sensor for a pixel should be:

$$X = E\Delta t = L\frac{\pi}{4}(\frac{d}{f})^2 \cos^4 \alpha \Delta t = kL\Delta t \tag{2}$$

Thus X is the exposure amount. Imaging systems record the exposure amount and convert it into image value Z, as $Z = f(X)$. Here, the function f is called the response function, and is designed to be non-linear. It is reasonable to assume that function f is monotonically increasing, so f^{-1} can be defined. Hereby, knowing the exposure time Δt, X can be calculated from image value Z by $X = f^{-1}(Z)$. And since $X/\Delta t$ is proportional to the scene radiance L, $f^{-1}(Z)/\Delta t$ can be regarded as relative scene radiance.

2.2 HDR Recovering

The increasing demand for HDR images and videos has inspired many solutions in the past years. For typical digital system, the available number of values for a pixel is limited, e.g. 256 levels in 8bit system, and thus the dynamic range of the captured scene radiance is narrowed. The most straightforward idea is to develop novel sensors with built-in HDR functions instead of conventional cameras [14, 15]. However, these special sensors are always expensive and not widely available. As a result, most research works focus on how to convert pixel values to relative scene radiance values, by using the camera response function obtained from multiple images captured with different exposures for the same scene .

Mann and Picard model the response function by $Z = \alpha + \beta X^{\gamma}$ [12]. Their solution is simple and feasible, but the recovery precision is unsatisfactory because of the highly restricted function form. Debevec and Malik propose a method only requires a smoothness assumption [2]. With the known exposure times for each image, their method achieves most accurate results. Mitsunaga and Nayar consider the situation where the exposure times are unobtainable, and model the response functions with high-order polynomials which can approximate the response functions accurately enough [13].

These methods require highly accurate pixel alignments across different images, which usually cannot be guaranteed in dynamic scene capture because

of object motion. Jacobs et al. proposes a LDR image alignment method to deal with both object and camera movements [5]. Nevertheless they have to assume that the movements are small and only introduce Euclidean transformation. Kang et al. recover HDR video by well implemented motion estimation [7]. However, artifacts such as ghost phenomenon still exist in fast moving scenes. Recently, Lin et al. present a novel method to generate HDR image for stereo vision [9]. Motion estimation is avoided in this method. However, similar problem of pixel alignment between different viewpoint images exists, which makes it difficult to handle the multi-view images with large disparities.

3 The Approach

This section presents our MVS approach for recovering HDR texture 3D model from multi-view LDR images. The approach is mainly composed of three stages (see Fig. 1): multi-view multi-exposure image capture, film response recovery and image calibration, and HDR texture 3D model reconstruction.

3.1 Multi-view Multi-exposure Image Capture

In our approach, multi-camera system is used to capture the multi-view images for MVS reconstruction. And the exposure times are set to be different for adjacent cameras, in order to jointly recover the HDR texture for the scene objects.

Capture system. To support the multi-view stereo approach, we have developed a multi-camera multi-light dome for multi-view capture. Inside the dome, 20 cameras are available which located evenly on a ring as shown in Fig. 2(a). Multi-view images/videos with spatial resolution of 1024×768 can be captured simultaneously. What's more, the exposure time of each camera can be individually configured, which meets the requirement of HDR texture recovery.

Fig. 2. Multi-view multi-exposure image capture: (a) dome system, cameras in the near-side are marked with circles; (b) camera setups: "*A*" and "*B*" represent cameras with two different exposure times; (c) captured multi-view multi-exposure images

Camera setup. All of the 20 cameras are divided into 2 groups and utilized for multi-view capture. Cameras in the first group have exposure time of $\Delta t1$, while those in the second group have exposure time of $\Delta t2$. The values of $\Delta t1$ and $\Delta t2$ are chosen to capture as large as possible the contiguous dynamic range of the scene radiance.

Selecting cameras for each group is another key point. As shown in Fig. 2(b), every 4 neighboring cameras starting from position $2k-1, k = 1, 2, \ldots$ constitute a subgroup, which contains three cameras with the same exposure time and one camera with another exposure time. For instance, the first 3 subgroups in Fig. 2(b) are "$AABA$", "$BABB$" and "$BBAB$". Notice that the camera with unique exposure time in the subgroup never appears at the subgroup border. And the total camera numbers of the two groups remain the same. Examples of captured multi-view multi-exposure images are shown. In Fig. 2(c), images from 4 adjacent viewpoints from image set "plaster figure" captured under the introduced camera setup are illustrated.

3.2 Film Response Recovery and Image Calibration

Under our camera setup, multi-view multi-exposure images are captured. In different images, the corresponding pixels which record the same scene point are very different in values and image positions. A multi-projection method is designed to extract corresponding pixels from different multi-view multi-exposure images accurately. And then film response recovery and image exposure calibration can be done.

Here we assume a controlled environment in which the cameras can be easily geometrically calibrated. And in this case the camera response functions may be recovered by some other pre-calculation methods. However, considering the uncontrolled environment such as outdoor scene capture, our technique will be essential together with other techniques such as self-calibration to achieve the goal.

Pixel correspondences extraction. Response function recovery requires highly accurate pixel alignment. For single viewpoint HDR imaging, positions of corresponding pixels in different images are the same. However, for multi-view imaging, the filming positions of the same scene point vary through different images.

For our captured multi-view multi-exposed images, take one subgroup defined in Section 3.1 for example. As shown in Fig. 3, images A, B, and D have the same exposure time which is different from that of image C. For any specified pixel p_b in image B, an optical ray starting from p_b and passing through the camera center can be calculated using the camera parameters. It is obvious that the real-world scene point x recorded by p_b is located on this ray. The only unknown variable to determine the real-world coordinate of x is the distance d from the camera center to x along the optical ray. To validate effective search for the value of d, visual hull [3] of the target object is firstly computed to restrict the search range. Since generation of visual hull only requires silhouette information of the object, the multi-exposed images can be directly used before exposure calibration.

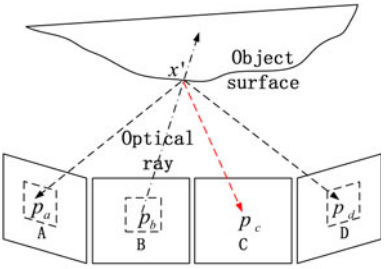

Fig. 3. Correspondence extraction in multi-exposed images

Within the restricted range along the optical ray, possible 3D points are evenly sampled first. Each candidate x' is projected back to pixels p_a and p_d in images A and D. Since p_a, p_b and p_d are filmed with the same exposure time, pixel consistency can be easily evaluated. Here, zero mean normalized cross correlation ($ZNCC$) is used as an evaluation metric. For two square windows of $N \times N$ centering on pixels p and q in two images, $ZNCC$ is computed by:

$$ZNCC(p,q) = \frac{\sum_{i=1}^{N^2} (z_p(i) - \overline{z_p}) \cdot (z_q(i) - \overline{z_q})}{\sqrt{\sum_{i=1}^{N^2} (z_p(i) - \overline{z_p})^2 \cdot \sum_{i=1}^{N^2} (z_q(i) - \overline{z_q})^2}} \tag{3}$$

where $z_p(i)$ and $z_q(i)$ are the values of the i-th pixels inside the $N \times N$ regions centering on p and q, while $\overline{z_p} = (\sum_{j=1}^{N^2} z_p(j)) \Big/ N^2$ and $\overline{z_q} = (\sum_{j=1}^{N^2} z_q(j)) \Big/ N^2$ denote the average pixel values in these regions. In our case, with different x', two $ZNCC$ curves are calculated: one is for areas centering on p_b and p_a and the other is for areas centering on p_b and p_d. According to [19], a local maximum method is used to find the most corresponding solution from the $ZNCC$ curves. Finally, with the best x', corresponding pixel values of p_b and p_c are restored as a corresponding value pair (z_b,z_c).

Note that under our camera setup, three images with the same exposure time can be used to determine the scene point position, which increases the robustness of the algorithm. Also to note that only pixel values of p_b and p_c are finally kept. This is because images B and C are most correlated and closed to the optical ray, and thus they are most efficient for film response recovery.

Fig. 4(a) illustrates the corresponding pixels extracted by our method for multi-view multi-exposure images. Only small portion (1%) of the total pixels correspondences are drawn to make them easy to be identified. Fig. 4(b) gives another result by using SIFT descriptor [8] which is used to match pixels for gray-scale stereoscopic images in [9]. It is obvious that our method achieves much higher precision and the number of effective points is more than sufficient. On the other hand, although the incorrect matchings in SIFT-based method can be partially removed by adding other constraints, it will result in inadequate

Fig. 4. Correspondence matching results: (a) our method and (b) SIFT-based method

number of the corresponding points for film response function recovery. This result proves the effectiveness of our geometry constraint based method.

Film response recovery. To recover the camera response function, we use the method proposed by Debevec and Malik [2] since the exposure times in our system are available. The extracted corresponding pixels offer large number of corresponding value pairs, denoted as $Z = \{(z_{\Delta t1}, z_{\Delta t2})\}$. It is usually not necessary to use all the values for recovering response function. On the other hand, due to noise and mismatch, not all the value pairs are valid. In order to remove the inaccurate value pairs, for each value pair, outliers are detected by examining if $z_{\Delta t1}$ or $z_{\Delta t2}$ has the deviation over certain thresholds. Considering the filming characteristics and saturate phenomenon, for exposure times $\Delta t1 > \Delta t2$, reasonable thresholds may be:

$$\begin{aligned} threshold(z_{\Delta t1}) &= \begin{cases} c\ , & for\ 0 < z_{\Delta t1} < 200 \\ c + (z_{\Delta t1} - 200), & for\ z_{\Delta t1} > 200 \end{cases} \\ threshold(z_{\Delta t2}) &= \begin{cases} c\ , & for\ z_{\Delta t2} > 50 \\ c + (50 - z_{\Delta t2}), & for\ 0 < z_{\Delta t2} < 50 \end{cases} \end{aligned} \tag{4}$$

where c is a small constant to indicate the basic tolerance of the inaccuracy. After the inaccuracy removal, N value pairs are randomly selected from Z to form a new set $Z^* = \{(z^*_{\Delta t1}(i), z^*_{\Delta t2}(i))\}, i = 1, \ldots, N$, with the only constraint that they should be well-distributed. Using the data in Z^*, camera response function $g(z)$ can be computed by solving the problem:

$$\begin{aligned} &\min\{\sum_{i=1}^{N}\sum_{j=1}^{2}[w(z^*_{\Delta t_j}(i)) - \ln(E_i) - \ln(\Delta t_j)]^2 \\ &+\lambda \sum_{z=1}^{254}[w(z)(g(z-1) - 2g(z) + g(z+1))]^2\} \end{aligned} \tag{5}$$

where $w(z)$ is the weight function introduced in [2]. Here we have assumed that all the cameras in our capture system have the same response function $g(z)$. This assumption is reasonable in our case because all the cameras in our dome system are with the same type and color calibration is done before capturing.

Image calibration. With the recovered response function $g(z)$, exposure calibration is easy to process. For one image with the exposure time Δt, the relative radiance E for any pixel value z can be estimated by $\ln(E) = g(z) - \ln(\Delta t)$. Similarly, all the differently exposed images can be converted to relative radiance maps. In practice, to ensure compatibility with the following steps, we just convert images from exposure time $\Delta t1$ to $\Delta t2$ by modifying each pixel value by: $z_{\Delta t2} = g^{-1}(\ln(\Delta t2) + g(z_{\Delta t1}) - \ln(\Delta t1))$. It is also feasible to convert all the images to a new exposure time $\Delta t3$. The exposure calibration results will be included in Section 4.1, which demonstrate that the exposure calibrated multi-view images can be satisfactorily obtained. And thus the effectiveness of 3D model reconstruction via MVS can be guaranteed.

3.3 HDR Texture 3D Model Reconstruction

In this step, the exposure calibrated multi-view multi-exposure images are used to finally reconstruct the 3D model of the real-world object by MVS algorithm. And then, HDR texture of the model is generated by comprehensively utilizing the reconstructed geometry information, the recovered camera response function and the color information from the original LDR images.

Geometry reconstruction. We adopt the point cloud based multi-view Stereo algorithm (PCMVS) [10] to reconstruct the 3D model. The PCMVS algorithm belongs to the multi-stage local processing MVS and produces outstanding performance among current MVS algorithms under spares viewpoint setups. PCMVS consists of three main steps: point cloud extraction, merging and meshing. The decoupled designing of these steps allows modification and utilization of any of the modules independently, which is another advantage for PCMVS being adopted in our approach.

Exposure calibrated images usually contain "textureless" and noisy areas which may result in failure in MVS algorithm. It is similar with the problem caused by shading effects or lack-of-texture which traditional MVS methods are facing. In PCMVS, robust methods are designed to solve this kind of problems. Thus, we simply extend the functional range of these methods to cover the new problem. Results will be shown in Section 4.2 to demonstrate the effectiveness of the adopted PCMVS algorithm in exposure calibrated images based geometry reconstruction.

HDR texture recovery. Once the 3D model is reconstructed, it can be used to generate HDR texture. More precisely, we recover the color information for each vertex v on the 3D mesh.

Color value for a certain vertex v on the mesh can be calculated by fusing its corresponding pixel colors in multiple images. Note that it cannot be foreknown which images contain the corresponding pixels of the vertex v, therefore the visibility of v in each image has to be computed first. Then, knowing that v is visible in images $\{p_j\}$, and the corresponding pixel values are $\{z_j\}$, the relative

radiance value of v is computed as the weighted average of the relative radiance values of each p_j:

$$\ln(E_v) = \frac{\sum_j w(z_j)(g(z_j) - \ln(\Delta t(p_j)))}{\sum_j w(z_j)} \tag{6}$$

where $w(z_j)$ is the same weight function used in film response recovery.

In our implementation, two stages of pre-processing are implemented to improve the recovery quality. First, images $\{p_j\}$ in which v is visible are sorted according to their view angles made by the vertex normal of v and the optical ray from v to their camera centers. Image with smaller view angle is moved closer to the top of the queue. This is because small view angle is more likely to guarantee color consistency. In practice, 2-4 images with least view angles are used to calculate the relative radiance value.

The second stage is introduced to deal with the ghost phenomenon. Ghost phenomenon is one of the most common problems in HDR image/video generation and is frequently discussed [5, 6]. It is basically caused by misalignment of pixels in multiple images with object motion. In our approach all the images are captured simultaneously, and thus motion is not a problem. Besides, multiple candidates of corresponding pixels are available to allow robust color recovery by selecting valid pixel values and discarding the outliers from $\{z_j\}$. Then the misaligned pixels are excluded and ghost phenomenon is efficiently removed.

Our method computes the relative radiance value for vertex v properly. The relative radiance map is used as the HDR texture of the 3D model directly or through certain processes. To further generate LDR textures under desired illuminance, just look up pixel values in the camera response curve with hypothetic exposure times or refer to existing tone-mapping methods [1, 11].

4 Results

In this section, the performance of our proposed approach for HDR texture model reconstruction is tested and the results are presented. Experiments use multi-view multi-exposure images captured for static scenes. Because to carry out comparative experiments, it is hard to involve dynamic scenes in respect that they are basically unrepeatable. However, our approach can handle dynamic scene directly since each temporal model is constructed independently.

4.1 Recovery of Camera Response Curve

Response function is firstly recovered based on our captured images. 20 images are divided into 2 groups by their exposure times. Then algorithm introduced in Section 3.2 is applied to recover the camera response function. The multi-view multi-exposure image set "plaster figure" (Fig. 2(c)), which is captured in two

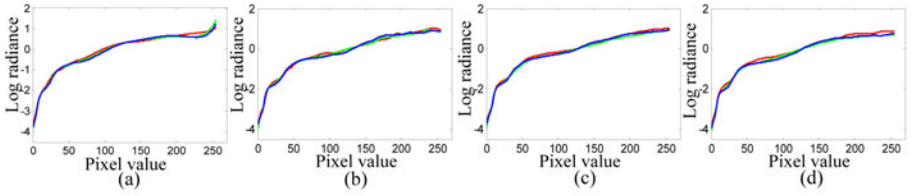

Fig. 5. Camera response curves: (a) our result; (b) average curve of 20 cameras; (c) curve of view 1; (d) curve of view 2

exposure times of 30ms and 10ms, is used as an example. The recovered average response curve for each color channel (R, G, and B) of 20 cameras are illustrated in Fig. 5(a).

To verify the feasibility of the recovered response curve, contrast experimental results are provided as references. Response functions for each camera are computed separately by taking multiple exposed images with a single camera and thus precise results can be obtained. Fig. 5(b) shows the average of the 20 referential camera response curves. Although there are slight differences, curves in Fig. 5(a) and Fig. 5(b) indicate that our method achieves fairly satisfactory result in camera response recovery without requiring temporal multi-exposure capturing.

Furthermore, Fig. 5(c) and (d) show the referential camera response curves of viewpoint 1 and viewpoint 2. In fact, note that each individual camera response curve looks similar to others, thus they can be substituted by the average curve. Therefore a conclusion can be reached that using one recovered curve for all cameras in our approach is reasonable and feasible.

With the recovered response curve, image exposure calibration can be done according to Section 3.2. Fig. 6 illustrates the calibration results of "plaster figure". Images (a) is captured in exposure time of 30ms while (b) and (d) are in 10ms. Images (c) and (e) are the results of exposure calibration, which now have the same exposure intensity with (a).

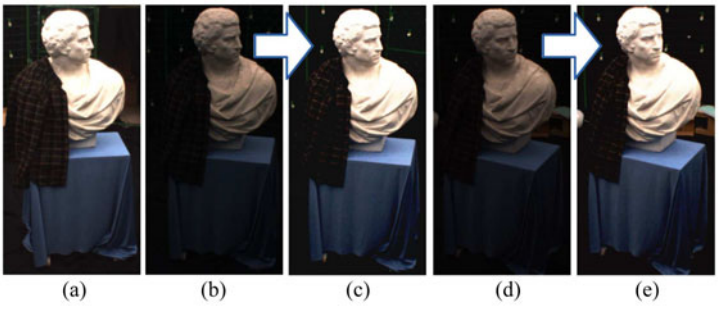

Fig. 6. Exposure calibration: (a), (b) and (d): original captured images in two different exposure times; (c) and (e): exposure calibration results of (b) and (d)

Clearly, the exposure calibration can be done accurately enough for 3D reconstruction. However, it should still be pointed out that because of the low dynamic range, images through exposure calibration cannot recover all the color information, and texture in some regions especially the saturated areas appears to be artificial and noisy. Generating HDR texture for 3D models aims to solve this problem and is just the motivation of our work.

4.2 Reconstruction of Geometry

Fig. 7 compares the geometry reconstruction results between visual hull [3], PCMVS [10] and our method. Our method uses the exposure calibrated images as inputs, while both visual hull and PCMVS use the images captured with carefully determined single exposure time to ensure the reconstruction quality. Obviously, visual hull performs the worst among these methods. And for PCMVS and our results, it is hard to tell the difference between them even with careful observation. It proves that our method achieves the same geometry reconstruction accuracy as PCMVS while using multi-exposed images as inputs.

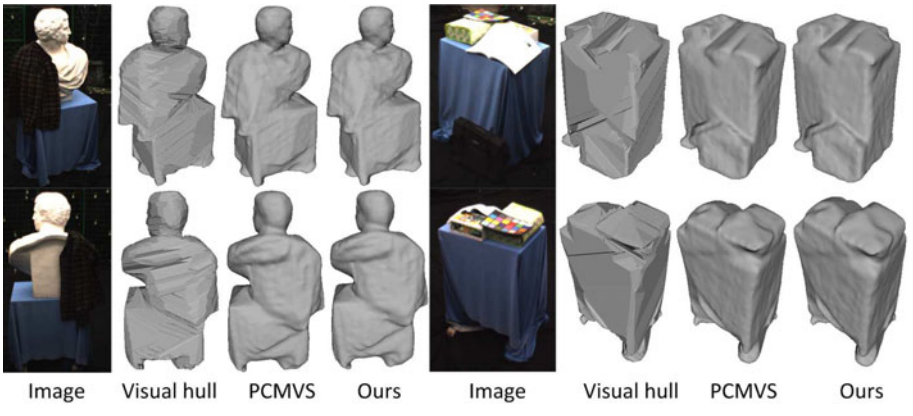

Fig. 7. Geometry reconstruction results. From left to right: original captured images; visual hull results; PCMVS results; our results.

Note that the geometry reconstruction module in our approach is totally decoupled from other modules, and hereby our approach is flexible to adopting any better MVS algorithms (e.g. [18]) to improve the reconstruction result.

4.3 Recovery of HDR Texture

The method for generating HDR texture is introduced in Section 3.3 and the recovered relative radiance map is used as the HDR texture after being taken the logarithm of is values. Fig. 8 illustrates the models of "plaster figure" with HDR texture (remapped to 0-255) generated by our method and LDR texture generated by the single exposure method. It is obvious that the HDR texture contains more color information in both bright and dark regions.

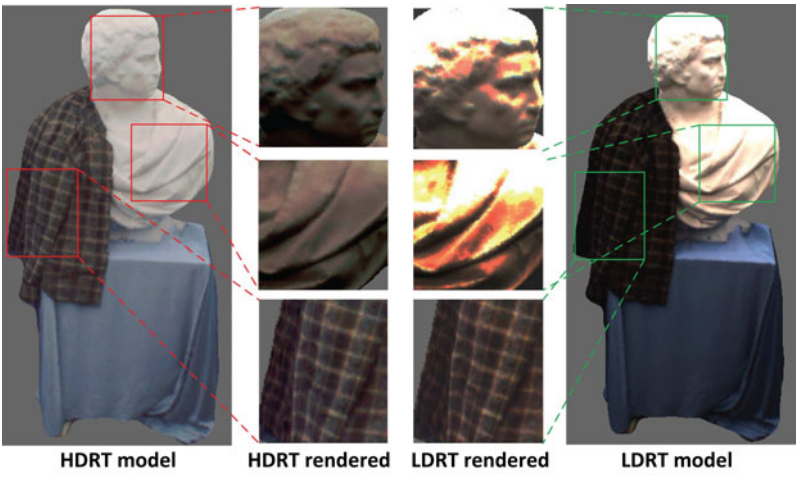

Fig. 8. Results of HDR texture recovery and rendering for "plaster figure". For the rendering results, the top and middle re-map the bright region, the bottom re-map the dark region.

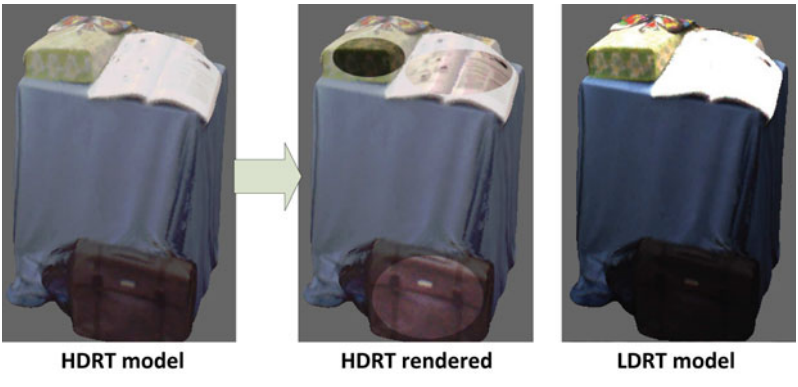

Fig. 9. Another result of HDR texture recovery and rendering for "article objects". Ellipse regions in the middle image show the local rendering results from HDRT model.

Luminance rendering results are further compared in Fig. 8. For HDR texture, different value ranges of the logarithmic relative radiance map are remapped to 0-255 to form the new textures; and for LDR texture, same ranges of the original 8bit color values are scaled to 0-255 as well. As shown in Fig. 8, the first two rendering settings aim to extract the details in the bright areas. Clearly, HDR texture rendered results derive plenty of details while the LDR texture fails. This is because the color information which LDR texture contains is highly limited and depends on the exposure setting. Thus bright areas with most pixel values saturated cannot offer any extra information in re-lighting. The last rendering results in Fig. 8 represent the textures under relatively strong illuminance. This

time the LDR texture rendered result is pretty good since the original LDR images are captured in a relatively long exposure time (30ms). On the other hand, HDR rendered result is still satisfactory.

Fig. 9 shows contrast experimental results with reconstructed models "article objects". The models with HDR and LDR texture are shown in the left and right, and the HDR texture rendered results can be seen in the three ellipses in the middle image. Notice the regions around the box and the book, most pixel values there are saturated in the LDR texture, which indicates that no details can be recovered. If shorter exposure time is used to capture the detail, the bag at the bottom would be totally invisible. On the contrary, the HDR texture keeps details both in the bright and dark regions within its high dynamic value range.

5 Conclusion

In this paper, we present a efficient approach for reconstructing 3D model with high dynamic range texture (HDRT) via multi-view stereo (MVS). Although both multi-view stereo and high dynamic range imaging attract lots of attentions from researchers, no joint research has been reported by now. Our approach includes three stages: multi-view multi-exposure image capturing; high-precision camera response recovery and image calibration; and finally 3D model reconstruction with HDR texture. Experimental results have demonstrated the effectiveness of our approach, and also the superiority of using these HDR texture models in illuminance rendering. Future works will include enhancing the geometry reconstruction precision by adopting new MVS algorithms or even utilizing our multi-exposure images to improve the existing MVS algorithms.

Acknowledgement. This work has been supported by the National Basic Research Project of China (973 Program No.2010CB731800), the Key Program of NSFC (No.60933006), and the Program of NSFC (No.60972013).

References

1. Drago, F., Martens, W.L., Myszkowski, K., Chiba, N., Rogowitz, B., Pappas, T.: Design of a tone mapping operator for high dynamic range image based upon psychophysical evaluation and preference mapping. In: HVEI 2003, pp. 321–331 (2003)
2. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: ACM SIGGRAPH 1997, pp. 369–378 (1997)
3. Franco, J.S., Lapierre, M., Boyer, E.: Visual shapes of silhouette sets. In: 3DPVT 2006, pp. 397–404 (2006)
4. Horn, B.K.: Robot Vision. McGraw-Hill Higher Education, New York (1986)
5. Jacobs, K., Loscos, C., Ward, G.: Automatic high-dynamic range image generation for dynamic scenes. IEEE Computer Graphics and Applications 28, 84–93 (2008)
6. Khan, E.A., Akyuz, A.O., Reinhard, E.: Ghost removal in high dynamic range images. In: ICIP 2006, pp. 2005–2008 (2006)

7. Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R.: High dynamic range video. ACM Transactions on Graphics 22, 319–325 (2003)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110 (2004)
9. Lin, H.Y., Chang, W.Z.: High dynamic range imaging for stereoscopic scene representation. In: ICIP 2009, pp. 4305–4308 (2009)
10. Liu, Y., Dai, Q., Xu, W.: A point cloud based multi-view stereo algorithm for free-viewpoint video. IEEE Transactions on Visualization and Computer Graphics 16, 407–418 (2009)
11. Mantiuk, R., Daly, S., Kerofsky, L.: Display adaptive tone mapping. ACM Transactions on Graphics 27, 1–10 (2008)
12. Mann, S., Picard, R.W.: Being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures. In: IST's 48th Annual Confrence, pp. 422–428 (1995)
13. Mitsunaga, T., Nayar, S.K.: Radiometric self calibration. In: CVPR 1999, pp. 1374–1380 (1999)
14. Nayar, S., Branzoi, V.: Adaptive dynamic range imaging: optical control of pixel exposures over space and time. In: Proc. ICCV 2003, pp. 1168–1175 (2003)
15. Nayar, S., Branzoi, V., Boult, T.: Programmable imaging using a digital micromirror array. In: CVPR 2004, pp. 436–443 (2004)
16. Nayar, S., Mitsunaga, T.: High dynamic range imaging: spatially varying pixel exposures. In: Proc. CVPR 2000, pp. 472–479 (2000)
17. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: CVPR 2006, pp. 519–528 (2006)
18. Vlasic, D., Peers, P., Baran, I., Debevec, P., Popović, J., Rusinkiewicz, S., Matusik, W.: Dynamic shape capture using multi-view photometric stereo. ACM Transactions on Graphics 28, 1–11 (2009)
19. Vogiatzis, G., Hernández, C., Torr, P., Cipolla, R.: Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 2241–2246 (2007)