HDRFusion: HDR SLAM using a low-cost auto-exposure RGB-D sensor

Shuda Li[1], Ankur Handa[2], Yang Zhang[1], Andrew Calway[1]

[1]University of Bristol, [2]University of Cambridge

Abstract. We describe a new method for comparing frame appearance in a frame-to-model 3-D mapping and tracking system using an low dynamic range (LDR) RGB-D camera which is robust to brightness changes caused by auto exposure. It is based on a normalised radiance measure which is invariant to exposure changes and not only robustifies the tracking under changing lighting conditions, but also enables the following exposure compensation perform accurately to allow online building of high dynamic range (HDR) maps. The latter facilitates the frame-to-model tracking to minimise drift as well as better capturing light variation within the scene. Results from experiments with synthetic and real data demonstrate that the method provides both improved tracking and maps with far greater dynamic range of luminosity.

Keywords: high dynamic range, 3-D mapping and tracking, auto exposure, RGB-D cameras

1 Introduction

Most existing methods for dense visual/RGB-D 3-D mapping and tracking rely on the brightness constancy assumption, i.e. the brightness of 3-D points observed from different viewing positions is constant. These can be categorized into using either a global or a local constancy assumption. The former assumes that any two over lapping frames from a sequence fulfil the condition [1], whilst the latter requires only that consecutive frames do[2,3]. The global assumption enables frame-to-model tracking which is known to accumulate less drift [4], whilst the local assumption is easier to meet in practice but means that the tracking is done frame-to-frame, with a consequent increase in drift.

However both of the above assumptions are broken in reality when using cameras equipped with automatic exposure (AE). AE is designed to map the high dynamic range of scene luminance into a narrow range for display devices while remain suitable for the human eye. When the camera moves from a bright to dark area, the exposure time is increased automatically so that more light can be captured by the camera sensor and vice versa when the camera moves from dark to bright regions. This breaks the global assumption since images viewing the same scene area from different viewing positions are seldom captured at the same auto-exposure. The local assumption is more likely to be met as exposure

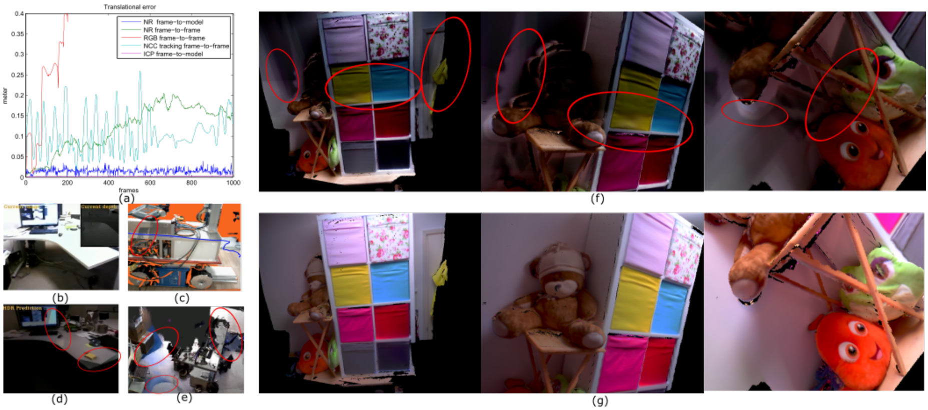

Fig. 1: (a) shows the proposed frame-to-model tracking using normalized radiance deliver best tracking accuracy using visual data. The tracking is performed using a challenging synthetic flickering RGB-D sequence. (b)-(e) are screen captures from video released with previous works. Specifically, (b) and (d) are from [5], where (b) is the raw input image. (d) is predicted scene textures. By contrast, the unrealistic artefacts, marked by red circles, indicate insufficent exposure compensation. (c) is predicted scene texture from [7]; (e) from [6]. Similar artefacts can be seen in these results. (f) in the top right shows the results from our implementation of [3] using a RGB-D video sequence, the artefacts are very strong due to large camera exposure adjustment when moving from bright area (top in the scene) to the dark area (bottom left in the scene). (g) in the bottom right are the predicted textures using the proposed HDRFusion. It can been that it is free of artefacts and its HDR textures are visualized using Mantiuk tone mapping operater [8].

usually changes smoothly, but this assumption also breaks when video flickering occurs. Video flickering artefacts, also known as brightness fluctuation, happen when a camera moves across the boundary between a bright and dark area or moves quickly back and forth between them: in these scenarios, the exposure changes dramatically in a short period of time and results in flickering. Turning AE off can ensure the brightness constancy, but it is often undesirable since it leaves bright areas over exposed and dark areas under exposed, leading to the loss of important visual detail.

AE also poses a problem when texturing a 3-D model of the scene. Overlapping images captured with inconsistent brightness will leave mosaic artefacts when projected back onto the model surface. This is a common problem for many state-of-the-art dense mapping systems as illustrated in Fig. 1. The problem has been widely addressed in conventional model texturing, panoramic imaging [9] and video tonal stabilization [10,11]. These works tackle the problem by compensating the global brightness of input images and blending colours along the boundaries between input images to create consistent texture. But these are

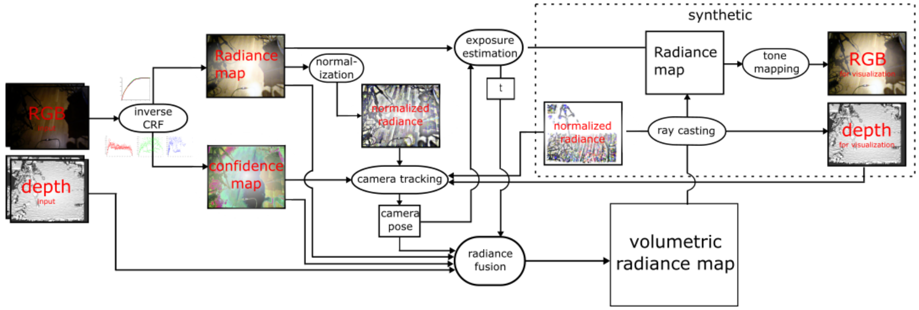

Fig. 2: Flow chart of HDRFusion. The boxes represent data structures, eclipses represent data transforming modules and arrows represent data flow. From left, input RGB frames are converted into radiance map. The camera pose is tracked in a frame-to-model style. Note that in confidence map is used in exposure estimation module but for simplitcity the data flow is not shown.

usually expensive offline approaches and are mainly aimed at delivering visually pleasing results rather than maintaining fidelity to the real world luminance.

In this paper, we introduce a novel technique for appearance based frame comparison which allows robust frame-to-model mapping and tracking using RGB frames with AE enabled. It is very robust to brightness fluctuation and is capable of capturing a consistent HDR texture on the 3-D surface of the model (Fig. 1(a)). The HDR range corresponding to real world luminance values is illustrated in Fig. 1(j) using Mantiuk tone mapping operation (TMO) [8].

The key assumption of the work is that the luminance of real world is globally constant and invariant to video brightness changes due to AE. The main challenge lies in how to build a real-time system, capable of tracking reliably with AE enabled so that HDR luminance can be captured by fusing low dynamic range (LDR) frames together. Instead of jointly tracking and compensating exposure like previous work [5] — which is not as robust and reliable as shown in our tests — we propose to track normalized radiance since it is a function which depends on only luminance. Radiance is the amount of luminance captured during the period of the exposure time. Another advantage of tracking normalized radiance lies in that the normalization operation can be efficiently implemented using down-sampled integral images.

Exposure compensation is therefore decoupled from tracking and greatly improves its accuracy as well. In the end, both the tracking and radiance fusion benefit from confidence maps derived from sensor noise level function which adaptively weighs radiance map at pixel level. Overall, the proposed HDRFusion achieves high quality radiance map and enables better visualization experience using TMOs. We will demonstrate the improvements in both qualitative and quantitative experiments.

1.1 Overview

We now give an overview of HDRFusion. The main algorithm is shown in the flow-chart in Fig. 2. The inputs are RGB-D frames from a Xtion sensor. Firstly, we estimate the inverse camera response function and noise level function for radiometric calibration (Section 4). The RGB frames are then converted into radiance map with estimated pixel-wise confidence. The camera poses are tracked by aligning incoming frame with the prediction coming from the 3D model built so far, *i.e.*, registering the live normalized radiance with predicted normalized radiance. The predicted normalized radiance is estimated by casting rays into the global volume. The confidence map is used to adaptively weigh error function for tracking, exposure compensation and radiance fusion. The ray casting module establishes predicted radiance, normalized radiance and depth. The predicted radiance and depth can be used for visualization through tone mapping on LDR devices or output as HDR data.

2 Related work

There is a huge wealth of literature on dealing with visual odometry or camera motion tracking. However, we will only focus on the direct approaches which can track and reconstruct a dense and textured 3-D model in real-time. Motion tracking using active sensor [4,12] is independent to lighting but leaves surfaces un-textured. Approaches combining appearance and depth [2,13,3,7] for camera tracking are the most relevant approaches. In all these approaches, it is assumed that brightness of consecutive frames is constant which is likely to fail when video flickering happens. In addition, [3] introduce a simple color blending method but as shown in our experiments, it is inadequate to deal with large exposure changes. Kerl *et al.* [7] introduce a key frame based approach by taking the rolling shutter effect into account. The approach relies on local brightness constancy when tracking live frames with a key frame, it is capable of producing sharp super-resolution frames involves no exposure compensation.

Maxime *et al.* [5] propose one of the first work in real-time 3-D HDR texture capturing. We follow the same approach of transforming raw RGB into radiance domain and tracking using radiance. It mainly focuses on re-lighting virtual specular objects. The differences between [5] and this paper are two-fold. First, in [5], a gamma function is adopted to approximate inverse camera response function (CRF). Gamma function may introduce error when radiance is high and the resulting radiance is not directly proportional to scene luminance (Fig. 3). Second, in [5], the exposure is estimated jointly with camera pose, but we find that the shape of error function when tracking using exposure compensated radiance bears shallow global minima even when exposure has been compensated for and, therefore, not as robust as normalized radiance based object function we proposed.(Fig. 4) Lastly, mosaic artefacts are clearly visible from the synthetic HDR mode which indicate inadequate exposure estimation (Fig. 1(d)).

Normalized cross correlation (NCC) has been widely applied in visual tracking [14] to deal with challenging lighting condition but its computational cost grows exponentially with the size of patch. Small patches are sensitive to image noise and bring many local minimum (Fig. 4). In addition, the 3-D HDR texture capturing is not addressed in the paper. HDR video capture using high-end stereo rig [15,16] is also relevant to the topic since it involves estimating disparity between binocular views so that LDR frames captured by both frame can be integrated into a single stream of HDR video [16], but the high quality HDR video is the main focus of the group of approach rather than a full 3-D model.

3 Preliminaries

Start from direct tracking using visual data assuming brightness constancy, camera poses can be estimated by minimizing the intensity difference between a reference frame and a live frame. The object funciton F can be formulated as:

$$F(\mathbf{R},\mathbf{t}) = \int_{\Omega} \left\| I_r(\mathbf{u}) - I_l(\pi(\mathbf{R}\pi^{-1}(\mathbf{u}, D_r(\mathbf{u})) + \mathbf{t})) \right\|_2^2 d\mathbf{u} \tag{1}$$

where $I : \Omega \rightarrow \mathbb{R}_+$ and $D : \Omega \rightarrow \mathbb{R}_+$ denote the intensity and depth functions. The whole 2-D image domain is denoted as $\Omega \subset \mathbb{R}^2$ and $\mathbf{u} \in \mathbb{R}^2$ is the pixel coordinate. Subscript $_r$ and $_l$ denote the reference frame and live frame respectively. $\mathbf{R} \in \mathbb{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$ are the rigid body motion to transform a 3-D point defined in reference coordinate system to live coordinate system. $\pi : \mathbb{R}^3 \rightarrow \Omega$ and $\pi^{-1} : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}^3$ are projection function and its inverse. $\pi(.)$ projects a 3-D point to image plane and $\pi^{-1}(.)$ transforms 2-D point back into 3-D given the depth D.

Equation 1 works as long as brightness constancy holds. We can define the correspondent point in live frame as $\mathbf{u}' = \pi(\mathbf{R}\pi^{-1}(\mathbf{u}, D_r(\mathbf{u})) + \mathbf{t})$ and $e(\mathbf{u},\mathbf{u}') = I_r(\mathbf{u}) - I_l(\mathbf{u}')$. Equation 1 is rewritten as $F(\mathbf{R},\mathbf{t}) = \int_{\Omega} \left\| e(\mathbf{u},\mathbf{u}') \right\|^2 d\mathbf{u}$. NCC based tracking can be viewed as an extension from equation 1 by replacing $e(\mathbf{u},\mathbf{u}')$ with $\sqrt{1 - C^2(\mathbf{u},\mathbf{u}')}$. $C(.)$ is the NCC score and defined as following:

$$C(\mathbf{u},\mathbf{u}') = \frac{1}{|\Omega_N|^2} \int_{\Omega_N} \frac{(N_r(\mathbf{u},\mathbf{v}) - \mu)(N_l(\mathbf{u}',\mathbf{v}) - \mu')}{\sigma\sigma'} d\mathbf{v} \tag{2}$$

Where $N : \Omega \times \Omega_N \rightarrow \mathbb{R}_+$ defines a small image patch, a neighbourhood centred at $\mathbf{u}$. $\Omega_N \subset \mathbb{R}^2$ is the domain of the neighbourhood N and $\mathbf{v} \in \mathbb{R}^2$ is the coordinate w.r.t. N. μ and σ are mean and std. (standard deviation) of image intensity over N_r and μ' and σ' are mean and std. over N_l. The NCC-based tracking can be formulated as $F(\mathbf{R},\mathbf{t}) = \int_{\Omega} \left\| 1 - C^2(\mathbf{u},\mathbf{u}') \right\| d\mathbf{u}$.

4 Camera imaging process

The key observation we rely on in this paper is that the scene luminance is mostly constant and invariant to exposure settings. The idea is to replace $e(.)$ with a new error function dependent on luminance only. The luminance L is the

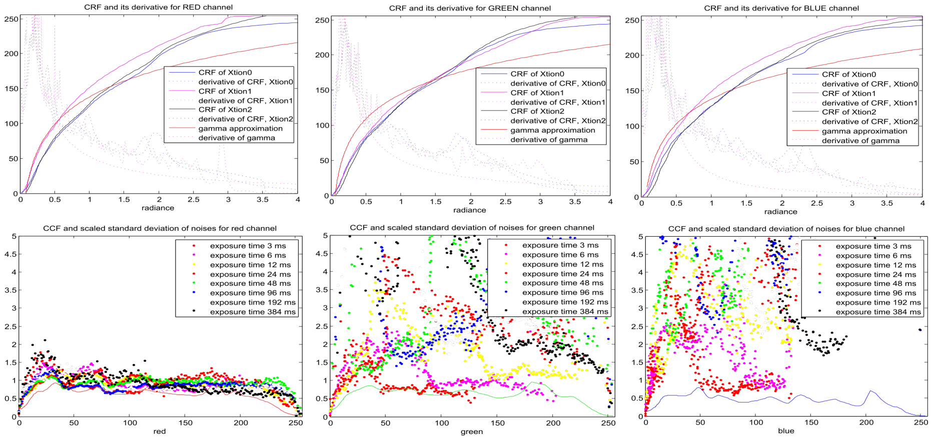

Fig. 3: The CRF and PCF of the RGB camera on 3 Xtion sensors. The figures in the top row are the CRF function and its derivative of RGB channels respectively. From the figure, we can see that the gamma approximation of CRF bears large error when the radiance is high. The figures in the bottom row are the PCF and scaled standard deviation of noise level captured as various exposure time.

radiance R received at the camera sensor per unit time $L = R/\Delta t$, where Δt is the exposure time. The relation between luminance and image intensity I can be described by the image formation model [17]:

$$I = f(R + n_s(R) + n_c) \tag{3}$$

where $f : \mathbb{R}_+ \to \mathbb{Z}_+$ is the camera response function (CRF) and $R = L\Delta t$. Essentially, it maps radiance R to LDR intensity level I, which is ranged from 0 to 255. n_s accounts for noise component dependent on the radiance, n_c accounts for the constant noise. The statistics of noise can be assumed as $E(n_s) = E(n_c) = 0$, $Var(n_s) = L\Delta t \sigma_s^2$ and $Var(n_c) = \sigma_c^2$.

The noise level function [18,19] measures how reliable sensor response is at given intensity level. For convenience, we convert it to a probability function by scaling the noise level function using a scalar m, where m is the maximum standard deviation over 3 colour channels.

$$p(I) = \frac{1}{m} \left.\frac{\partial f(r)}{\partial r}\right|_{r=R} \sqrt{R\sigma_s^2 + \sigma_c^2} \tag{4}$$

where $p : \mathbb{Z}_+ \to (0, 1)$. $R = f^{-1}(I)$ is the radiance. In the right column of Fig. 5, the probability maps are shown. Each channel represents the probability of the channel at the pixel location: dark areas show the low probability pixels which usually occur around exposed or under exposed parts of the image. We can also define variants of this probability function based on equation 4. $p_0(I) = 1$, $p_i(I) = \sqrt{p(I)}$, $p_2(I) = p(I)$, and $p_3(I) = p(I)^2$. For clarity, the family of

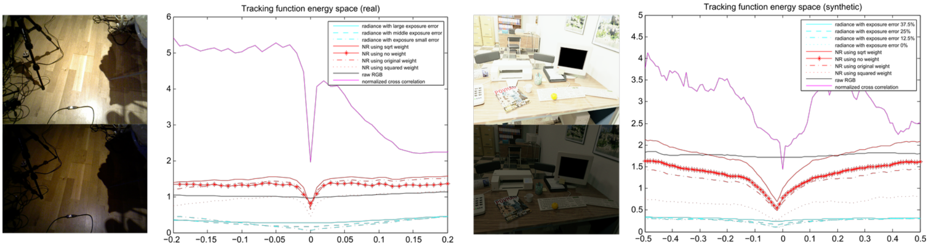

Fig. 4: The comparison of tracking errors. The family of error function using normalized radiance (red) gives the most ideal global minimum over the ground truth. The NCC based error function [14] also presents a strong convex but bears a lot of local minimum. Tracking using exposure compensated radiance [5] looks better than tracking raw RGB but its global minimum are shallow even when the exposure is compensated with high accuracy. The left plotting uses real flickering pair and the right uses simulated flicking pairs based on [19].

probability functions are named as pixel confidence functions (PCF) from now on since these are different from noise level functions. Their effects will be discussed in section 5.

The CRF and PCF depends on specific type of camera sensor. They can be pre-calibrated before performing the HDRFusion [20,21,22]. Specifically, our CRF is estimated by putting the RGB-D sensor at fixed position. A sequence of images at different exposure time are captured [20] and the noise level function and PCF are estimated using [19]. The CRF, its derivative and PCF are shown in Fig. 3. With this estimated CRF, its inverse $f^{-1}(.)$ and PCF can be calculated straightforwardly: inverse CRF, allows us to convert intensity to radiance efficiently and PCFs allow us to weigh the error terms appropriately in tracking, exposure compensation and fusion stage.

5 Normalized radiance

Now we derive a novel error function dependent on luminance alone. The normalization of the radiance in a patch of neighbourhood N, centred at pixel location $\mathbf{u}$, can be formulated in the following:

$$\bar{R}_N(\mathbf{u}) = \frac{R_N(\mathbf{u}) - E(R_N)}{\sqrt{Var(R_N)}} = \frac{L_N(\mathbf{u})\Delta t - E(L_N \Delta t)}{\sqrt{Var(L_N \Delta t)}} = \frac{L(\mathbf{u}) - E(L_N)}{\sqrt{Var(L_N)}} \qquad (5)$$

where $R_N : \Omega_N \to \mathbb{R}_+$ and $L_N : \Omega_N \to \mathbb{R}_+$ denote radiance map and luminance map in N, respectively. From above equation, it can be seen that $\bar{R}_N(\mathbf{u})$ is independent from exposure Δt. This value is also invariant to viewpoint due to the fact that the luminance distribution in the local region corresponding to N is roughly constant to viewing position, as long as the surface is Lambertian. Fig. 5 shows the mean, standard deviation, normalized radiance and confidence map

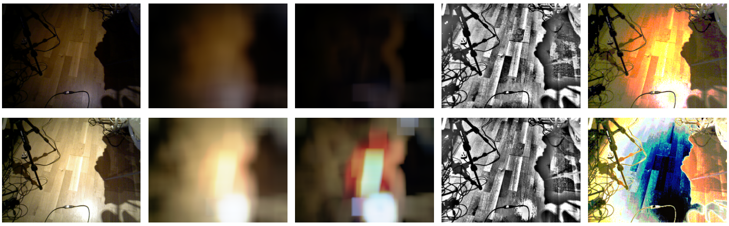

Fig. 5: Radiance normalization. From left to right, figures correspond to raw RGB, mean, standard deviation, normalized radiance, and confidence map. The 1^{st} and 2^{nd} row correspond to 2 consecutive frames when flickering happens. Although the image brightness changes significantly, the normalized radiance map is pretty similar thank to equation 5. The mean, standard deviation maps and normalized radiance are tone mapped from HDR.

of two consecutive frames captured at different exposure time. It can be seen that the normalized radiance maps extracted from frames captured at different exposure are strikingly similar while the mean and standard deviation maps are smooth and blurry which indicates good resistance to viewpoint changes. Therefore, the new error function can be defined as:

$$e'(\mathbf{u}, \mathbf{u}') = (\bar{R}_r(\mathbf{u}) - \bar{R}_l(\mathbf{u}'))p(I_l(\mathbf{u}')) \tag{6}$$

where the probability $p(I_l(\mathbf{u}'))$ serves as a dynamic weight to balance the noise introduced during image formation such that less reliable pixel will be assigned with a smaller weight. $p(.)$ can be chosen from the family of PCFs we defined before. $p(.) \in \{p_0(.), p_1(.), p_2(.), p_3(.)\}$.

The error functions using NCC, raw intensity, radiance with exposure compensated and the proposed normalized radiance are compared by plotting against the ground truth along x-axis in Fig. 4. Pairs of flickering consecutive frames are chosen, where one is real and the other is synthetic. It can be seen that our proposed error function using normalized radiance and weighted by square root PCF $p_1(.)$ gives the most ideal error space for optimization.

The camera poses can then be solved out by optimizing the error functions using the forward compositional approach described in [3].

5.1 Exposure compensation

When the camera pose is estimated, the exposure will then be compensated using the follow equation:

$$t = \frac{1}{|\Omega|} \int_{\Omega} p_l(\mathbf{u}) \frac{R_r(\mathbf{u})}{R_l(\mathbf{u}')} d\mathbf{u} \tag{7}$$

where $p_l(.)$ is the PCF of live frame. After t is estimated, the radiance map of live frame will be scaled by t.

6 Radiance Fusion

The exposure compensated radiance map tR_l will then be fused into a global volume using an fast parallel approach similar to [3]. The volumetric data structure stores not only the truncated signed distance function (TSDF) and its weights, but also the 3 channels of radiance and normalized radiance and radiance weights. The normalized radiance is also fused into the global volume so that synthetic normalized radiance map can be efficiently extracted using ray casting. Note that the radiance weight is different from TSDF weights. The fusion of radiance with depth for each voxel is shown in the following equations:

$$F = \frac{w_F * F + w'_F * F'}{w_F + w'_F} \tag{8}$$

$$R = \frac{w_R * R + w'_R * R'}{w_R + w'_R} \tag{9}$$

$$\bar{R} = \frac{w_R * \bar{R} + w'_R * \bar{R}'}{w_R + w'_R} \tag{10}$$

$$w_F = w_F + w'_F \tag{11}$$

$$w_R = w_R + w'_R \tag{12}$$

where F and R are TSDF values and radiance in global volume; F' and R' are those from live frame. Similarly, w_F and w_R are the global weights. w'_F and w'_R are weights from live frame. $w_F = |\mathbf{n}^T\mathbf{v}|$ is the absolute cosine values between surface normal $\mathbf{n}$ and viewing direction $\mathbf{v}$ at the live pixel location where $\mathbf{n}, \mathbf{v}$ are unit vectors. It down weight the TSDF values captured at high angle between the normal and viewing direction. Its effect is illustrated in Fig. 6 $w_R = \frac{p_r+p_g+p_b}{3}$, where p_r, p_g and p_b are the PCF values of 3 colour channels respectively. In experiments, we find that storing individual PCF of 3 colour channel is the global volume is unnecessary and may introduce color distortion as well.

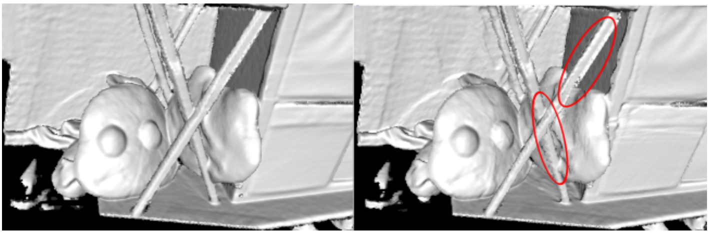

Fig. 6: Weight TSDFs according the angle between viewing direction and surface normal improves the geometry quality around thin and corner structures.

To ensure the quality of radiance, only the pixels whose maximum PCFs are above a threshold τ_0 and the angle between surface normal and viewing direction is above threshold τ_1 are allowed to be fused into the volume.

$$\{R|max(p_0, p_1, p_2) > \tau_0) \bigcap \mathbf{n}^T\mathbf{v} < \tau_1\} \tag{13}$$

7 Experiments

In all experiments, we have used 3 Xtion RGB-D sensors whose exposure can be specified. Calibrated CRFs of them are plotted in Fig. 3. Except CRF and noise level function, no other parameters need to be calibrated. Camera intrinsics are set as default values as in OpenNI library. A C++ implementation and testing data for both the main HDRFusion and its calibration are available in [1]. The codes are tested on two commodity system, PC0 equiped with NVIDIA GTX 680 and PC1 NVIDIA GTX Titan Black GPU. Both PC are hosted by an i7 quad-core CPU. The volume resolution are set as 256^3 and 480^3 for PC0 and PC1 respectively with volume size ranges from 2^3 to 3^3m according to the size of the scene. Frame resolution are set as QVGA for PC0 and VGA for PC1. Both of them operates at about 10Hz. We present a qualitative comparison with [3] and demonstrate the quality or recover HDR radiance map in an accompanying video: `https://youtu.be/ehwiFkmFQ7Q`.

7.1 Tracking under flickering

We first use synthetic dataset ICL to evaluate our approach [19]. The high quality CG HDR frames and ground truth camera poses are available. First, photo realistic LDR RGB frames are simulated using real CRF and noise level function of a randomly chosen Xtion sensor. We generates two sequences of video to simulate video flickering and smooth AE behaviour. The flickering sequence is simulated by randomly choosing exposure time from the set $3, 6, 12, ..., 96$ (ms). The second sequence is generated using the equation $\Delta t = C/L$, where $C = 4.8 \times 10^5$ and L is the average HDR intensity of the 10 by 10 patch in the center of the original HDR frames. The exposure simulated in the second way are changing smoothly. The Kinect like depth noise is also added using the approach from [23]. Typical flickering pairs are illustrated in Fig. 4. The tracking approach using normalized intensity, NCC object function based on [14] and approach similar to the tracking of [5] are used as baseline approaches. For fairness, the ICP-based frame-to-model tracking are disabled for all above methods. The tracking accuracy in terms of rotational and translational error are plotted in Fig. 7.

We also performed a qualitative comparison using real data between the proposed tracking and tracking using the approach from [3]. Two sequences of RGB-D video with flickering are captured. In these sequences, the sensor is overlooking a floor and a white board respectively. As the camera moving

[1] `https://lishuda.wordpress.com`

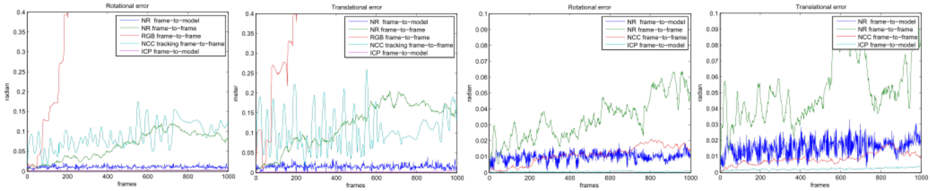

Fig. 7: Tracking synthetic sequence. The left two figures are the rotational and translational error using synthetic flickering sequence. The right two figures are using synthetic smooth AE sequence. In the flickering sequence, we can see that raw RGB based tracking quickly get lost, while the NCC and the proposed frame-to-frame tracking (NR) and frame-to-model tracking using normalized radiance remains working well. The tracking NR in frame-to-model mode gives the best performance in the flickering sequence. Due to the rich geometric variance, the ICP-based frame-to-model tracking give the best results. In smooth sequences, the ncc and ICP performs better but the proposed tracking remain working reasonably accurate. The frame-to-model tracking is within 3cm meter in the 1000 frames testing sequence.

from dark to bright areas, video flickering happens. [3] fails to tracking when flickering happens, while the proposed method remain tracking effectively. The reconstructed floor and white board using proposed approach are shown in Fig. 8. The tracking comparison between our approach and [3] is also available in the accompanying video.

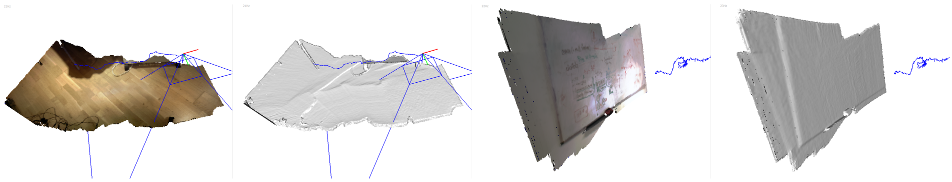

Fig. 8: Tracking under flickering using real data. The blue curves are the camera trajectories. The frustums in the left figure show the camera pose.

7.2 HDR Radiance map

The HDR radiance are shown both in the screen shots attached in the paper and in the accompanying video. In Fig. 1, 9, 10 and 11, we perform the proposed HDRFusion in three scenes, namely ’Bear’, ’Desk’ and ’Sofa’. The bear sequence is illuminated by indirect sun light. The desk sequence is illuminated by Fluorescent. The sofa sequence is illuminated by both fluorescent lighting and

Dedolight-400D metal halide lamp. In Fig. 9, HDR scene textures are compared with the ground truth. The ground truth is captured using a Canon 5D MarkII SLR camera. Three exposure LDR images with a 2-fstop interval of the scene were captured and then merged to form an HDR image. Both are rendered using tone maping operator(TMO) [8].

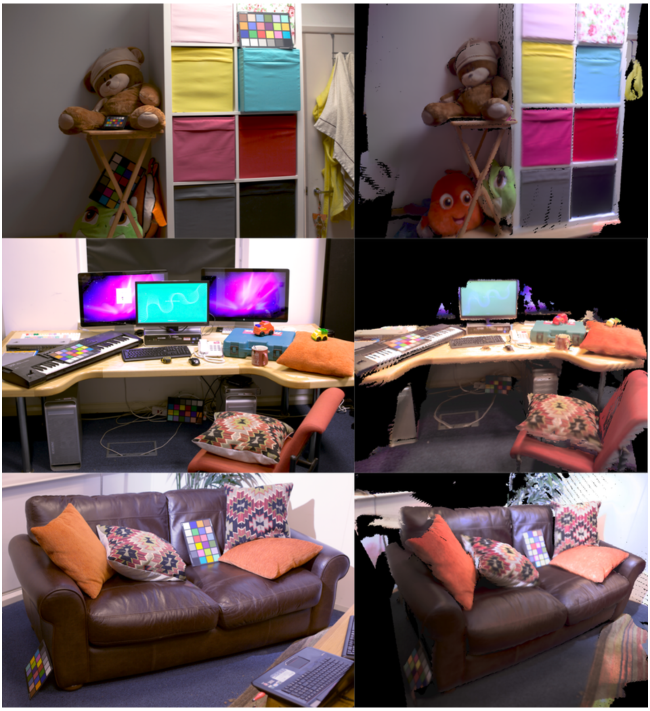

Fig. 9: The left are the ground truth HDR radiance and HDR radiance generated using HDRFusion are rendered using [8] where the colour saturation is set as 1. We can see that estimated HDR texture closely matches the HDR radiance captured using the high-end SLR camera.

8 Conclusion

In this paper, we propose a novel HDRFusion system capable of capturing high quality HDR scene texture using a low cost RGB-D sensor. Tracking normalized radiance allows decouple the tracking from exposure compensation which improves the accuracy of both. Tracking normalized radiance is also shown to be robust to video flickering due to camera AE adjustment. The tracking is runing in frame-to-model mode which accumulates less drift. In future work, calibrating the CRF function online will be investigated as in some sensors the exposure time can not be changed by user. Another limitation of the system lies in its large memory footprint. Storing both the normalized radiance and radiance seems unnecessary. Reducing the size of memory cost by combining the both will also be investigated.

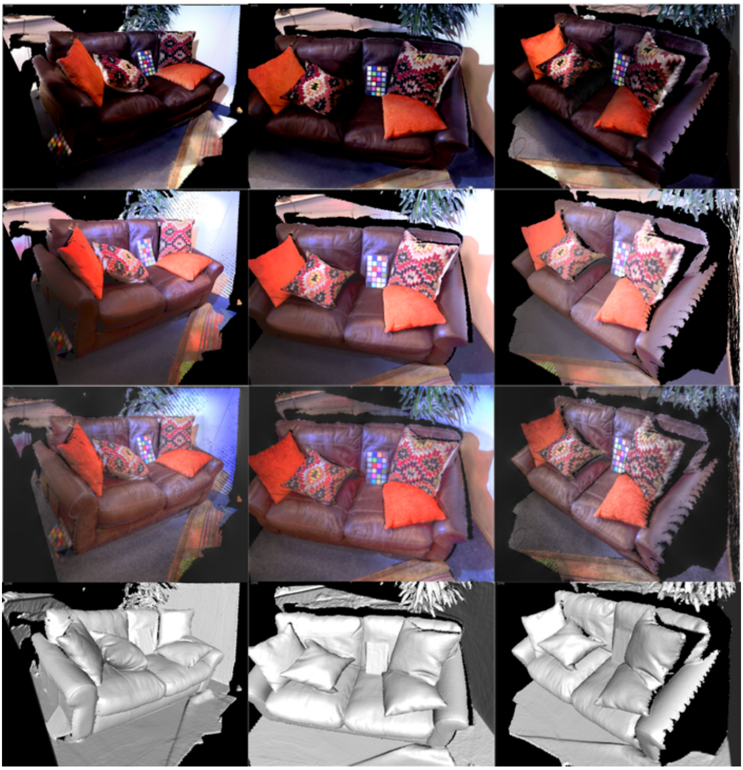

Fig. 10: Sofa. The LDR frames generated using [3] are shown in the first row and HDR frames produced by HDRFusion are shown in the second row and third row. The second row is generated using [8] where the colour saturation is set as 1. Comparing with raw RGB fusion [3], the dynamic range of the radiance texture is much higher. The details in dark area are well preserved. The third row is generated using [24], where the colour saturation is set as 1.25. [24] visualizes the rich details captured by HDRFusion. The bottom row shows the recovered surface geometry.

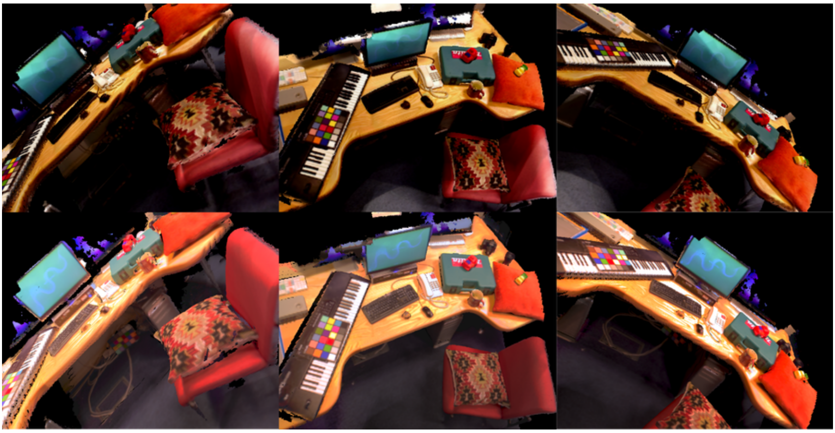

Fig. 11: Desk. The LDR frames generated using [3] are shown in the first row and HDR frames produced by HDRFusion are shown in the second row. The HDR radiance is rendered using [8], where the colour saturation is set as 1.5. The luminance under the desk is very low but are well preserved in the HDR radiance map.

References

1. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM : Dense Tracking and Mapping in Real-Time. In: Intl. Conf. on Computer Vision (ICCV). (2011)
2. Kerl, C., Sturm, J., Cremers, D.: Robust odometry estimation for RGB-D cameras. In: IEEE Intl. Conf. on Robotics and Automation (ICRA). (2013) 3748–3754
3. Whelan, T., Kaess, M., Johannsson, H., Fallon, M., Leonard, J.J., McDonald, J.: Real-time large-scale dense RGB-D SLAM with volumetric fusion. Intl. Journal on Robotics Research (IJRR) **34**(4-5) (2015) 598–626
4. Newcombe, R.A., Molyneaux, D., Kim, D., Davison, A.J., Shotton, J., Hodges, S., Fitzgibbon, A.: KinectFusion : Real-Time Dense Surface Mapping and Tracking. In: IEEE/ACM Intl. Symposium on Mixed and Augmented Reality (ISMAR). (2011)
5. Meilland, M., Barat, C., Comport, A.: 3D High Dynamic Range dense visual SLAM and its application to real-time object re-lighting. In: IEEE/ACM Intl. Symposium on Mixed and Augmented Reality (ISMAR). (2013) 143–152
6. Whelan, T., Leutenegger, S., Salas-moreno, R.F., Glocker, B., Davison, A.J.: ElasticFusion : Dense SLAM Without A Pose Graph. Robotics: Science and Systems (RSS) (2015)
7. Kerl, C., Cremers, D., Universit, T.: Dense Continuous-Time Tracking and Mapping with Rolling Shutter RGB-D Cameras. In: Intl. Conf. on Computer Vision (ICCV). (2015)
8. Mantiuk, R., Daly, S., Kerofsky, L.: Display adaptive tone mapping. ACM Trans. on Graphics (ToG) **27**(3) (2008) 1
9. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. Intl. Journal of Computer Vision (IJCV) **74**(1) (2007) 59–73
10. Aydin, T., Stefanoski, N., Croci, S.: Temporally coherent local tone mapping of HDR video. ACM Trans. on Graphics (ToG) **33**(6) (2014)
11. Farbman, Z., Lischinski, D.: Tonal stabilization of video. ACM Trans. on Graphics (ToG) **30**(4) (2011) 1
12. Serafin, J., Grisetti, G.: NICP : Dense Normal Based Point Cloud Registration. In: Intl. Conf. on Intelligent Robot Systems (IROS). (2015) 8
13. Whelan, T., Johannsson, H., Kaess, M., Leonard, J.J., McDonald, J.: Robust real-time visual odometry for dense RGB-D mapping. In: IEEE Intl. Conf. on Robotics and Automation (ICRA). (2013) 5724–5731
14. Scandaroli, G.G., Meilland, M., Richa, R.: Improving NCC-based direct visual tracking. In: European Conf. on Computer Vision (ECCV). (2012) 442–455
15. Heo, Y.S., Lee, K.M., Lee, S.U.: Robust Stereo matching using adaptive normalized cross-correlation. IEEE Trans. Pattern Anal. Machine Intell. (PAMI) **33**(4) (2011) 807–822
16. Batz, M., Richter, T., Garbas, J.U., Papst, A., Seiler, J., Kaup, A.: High dynamic range video reconstruction from a stereo camera setup. Signal Processing: Image Communication **29**(2) (2014) 191–202
17. Hasinoff, S.W., Durand, F., Freeman, W.T.: Noise-optimal capture for high dynamic range photography. In: IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR). (2010) 553–560
18. Liu, C., Szeliski, R., Kang, S.B., Zitnick, C.L., Freeman, W.T.: Automatic estimation and removal of noise from a single image. IEEE Trans. Pattern Anal. Machine Intell. (PAMI) **30**(2) (2008) 299–314

19. Handa, A., Newcombe, R.A., Angeli, A., Davison, A.J.: Real-Time Camera Tracking : When is High Frame-Rate Best? In: European Conf. on Computer Vision (ECCV). (2012)
20. Debevec, P.E., Malik, J.: Recovering High Dynamic Range Radiance Maps from Photographs. In: ACM SIGGRAPH (SIGGRAPH). Number August (1997) 1–10
21. Kim, S.J., Pollefeys, M.: Radiometric Self-Alignment of Image Sequences. In: IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR). (2004) 645–651
22. Kim, S.J., Frahm, J.M., Pollefeys, M.: Joint feature tracking and radiometric calibration from auto-exposure video. In: IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR). Volume 2. (2007)
23. Handa, A., Whelan, T., Mcdonald, J., Davison, A.J.: A Benchmark for RGB-D Visual Odometry , 3D Reconstruction and SLAM. In: IEEE Intl. Conf. on Robotics and Automation (ICRA). (2014)
24. Mantiuk, R., Myszkowski, K., Seidel, H.P.: A perceptual framework for contrast processing of high dynamic range images. ACM Transactions on Applied Perception (2006) 87 – 94