



HAL
open science

3D High Dynamic Range Dense Visual SLAM and Its Application to Real-time Object Re-lighting

Maxime Meilland, Christian Barat, Andrew I. Comport

► **To cite this version:**

Maxime Meilland, Christian Barat, Andrew I. Comport. 3D High Dynamic Range Dense Visual SLAM and Its Application to Real-time Object Re-lighting. International Symposium on Mixed and Augmented Reality, Oct 2013, Adelaide, Australia. 10.1109/ISMAR.2013.6671774 . hal-01357357

HAL Id: hal-01357357

<https://hal.science/hal-01357357>

Submitted on 27 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

3D High Dynamic Range Dense Visual SLAM and Its Application to Real-time Object Re-lighting

Maxime Meilland^{*}, Christian Barat[†], Andrew Comport[‡]
CNRS-I3S,
University of Nice Sophia-Antipolis



Figure 1: Realistic HDR reflections on several virtual objects with shadow casting, rendered in real-time from live video stream. The reflection maps have been computed from a dense live reconstruction of the scene's light-field.

ABSTRACT

Acquiring High Dynamic Range (HDR) light-fields from several images with different exposures (sensor integration periods) has been widely considered for static camera positions. In this paper a new approach is proposed that enables 3D HDR environment maps to be acquired *directly* from a dynamic set of images in real-time. In particular a method will be proposed to use an RGB-D camera as a dynamic light-field sensor, based on a dense real-time 3D tracking and mapping approach, that avoids the need for a light-probe or the observation of reflective surfaces. The 6dof pose and dense scene structure will be estimated simultaneously with the observed dynamic range so as to compute the radiance map of the scene and fuse a stream of low dynamic range images (LDR) into an HDR image. This will then be used to create an arbitrary number of *virtual* omni-directional light-probes that will be placed at the positions where virtual augmented objects will be rendered. In addition, a solution is provided for the problem of automatic shutter variations in visual SLAM. Augmented reality results are provided which demonstrate real-time 3D HDR mapping, virtual light-probe synthesis and light source detection for rendering reflective objects with shadows seamlessly with the real video stream in real-time.

Index Terms: Real-time rendering, photo-realistic rendering, vision-based registration and tracking, MR/AR for entertainment.

^{*}e-mail: meilland@i3s.unice.fr

[†]e-mail: barat@i3s.unice.fr

[‡]e-mail: andrew.comport@cnrs.fr

1 INTRODUCTION

Rendering virtual objects seamlessly in real-time video streams requires both visually coherent and efficient rendering techniques. Achieving this goal depends on several major criteria including accurate estimation of the pose between the camera and the scene, correctly handling occlusions between objects and generating realistic lighting conditions. Whilst large advances have been made in both computer graphics for illumination and in computer vision for structure and motion, fewer papers have looked at the difficult topic performing 6D pose, dense structure and complex illumination estimation in real-time. Efficient lighting models should not only make pose estimation and mapping approaches more robust and accurate but also enable realistic illumination of virtual objects for Augmented Reality (AR). In particular, real-time AR with photo-realistic rendering of virtual objects is far from current state-of-the-art computer graphics approaches.

Lighting is one of the most important subjects in the field of 3D graphics and modelling it correctly adds a lot to the visual appeal of the rendered scene. Modern computer graphics approaches exist for real-time realistic rendering and illuminating scenes based on accurate models for lighting, textures and bidirectional reflectance distribution functions (BRDFs). The basic light encompasses Ambient, Diffuse and Specular elements and rendering reflective surfaces is often handled by an efficient *reflection-map* whereby a texture is used to store the image of the environment surrounding the object [14]. Note that in this paper the term "reflection-map" is used to define the "2D environment-map" as widely known in the computer graphics literature, whereas, "scene-map" and "3D environment-map" will be used to refer to a 3D image-based model of the world. In computer graphics, the reflection-map is usually assumed to be distant or at infinity so that parallax effects can be neglected. One of the best ways of obtaining high-quality data is through measurements of scene attributes from real photographs, known as image-

based rendering or light-field rendering [19]. Inevitably, there is always a trade-off between real-time rendering performance and highly realistic rendering such that efficient image-based rendering approaches have rarely been considered in real-time.

In computer graphics, the process of estimating the various rendering attributes from photos is known as *inverse rendering*. In [31, 32] a mathematical framework was proposed which describes reflected light as a convolution of the lighting with the BRDF and expresses it mathematically as a product of spherical harmonic coefficients. This approach is only well conditioned if a reflective or specular surface is observed, however, it provides an efficient representation for real-time diffuse shading. Several surveys on inverse rendering are available including [29] and more recently [17] where the various approaches to illuminating mixed reality scenes are classified. Inverse rendering approaches have not only considered recovering lighting parameters but also 3D structure estimation. In [5] a useful "voxel" representation was proposed for representing 3D worlds based on extending classic 2D image grids to 3D and much effort has been invested in efficiently representing surfaces by truncated signed distance functions.

The problem of representing the *dynamic range* of a scene has existed since the beginning of photography, however, early work on compositing multiple LDR images to form an HDR image using electronic CCD sensor was pioneered by [21]. In this approach multiple images, each with a different exposure times or shutter speeds are used to capture a single radiance map covering a large dynamic range. A common method of compositing is to estimate the radiance by independently transforming each of the input images and combining the results using a weighted sum or per-pixel non-linear optimisation [8]. Much of this research assumes that the images have been acquired at the same time instant and real-time computation is not considered. In [1], the paracomparametric camera response function (CRF) was used to model the response of a virtual HDR camera to multiple inputs using a lookup table. Whilst efficiency was largely improved compared to direct computation methods, it is still not real-time (20sec per image). More recently HDR videos have been considered with processing achieved on specialised hardware. Finally, in [20] a real-time hardware based approach was proposed that triggers the acquisition of images with different dynamic range sequentially in the video [20, 22].

In [6], scene radiance and illumination are measured to render synthetic objects in real scenes with realistic lighting. The approach uses an HDR image-based map of the scene for relighting rather than synthetic light sources. In many approaches such as [6, 18, 28], the global illumination is measured using a light probe. These approaches are non-passive and involve estimating the lighting using probes such as omni-directional cameras that use mirrored balls or wide angle fish-eye cameras. A stereo omni-directional light probe was used in [34, 4] to acquire reflection-maps with depth. Only static cameras are considered and it is assumed that the exposure time of the cameras can be varied manually in order to acquire a large dynamic range. This class of techniques is very cumbersome and too invasive to be used in a real-time AR setting. Furthermore, the light probe only measures the light-field from one point in space and therefore the lighting environment is only approximated locally around that point. If several virtual objects are required or a single virtual object is displaced from the probe location then it is necessary to have placed the light probe at all these positions beforehand, otherwise the reflectance map is completely wrong.

More recently several approaches [42, 38] have been proposed using multiple views with multiple illumination conditions allowing virtual relighting of a scanned object. A multi-view 3D reconstruction algorithm is first applied using traditional multi-view stereo algorithm. After this, the reconstructed model is relit through an image-based relighting scheme for each camera view, followed with view-independent texture mapping procedure. Whilst these

approaches show the fundamental properties of image-based relighting, they are prohibitive and require complex apparatus along with manual procedures to acquire an object. Whilst compositing several real scans together (ie. a scanned scenes with scanned objects) is of high interest, the present paper does not aim to cover this problem and the focus will be made on augmenting reality with virtual objects and realistic relighting using dense 3D HDR environment-maps.

Alternatively to inverse-rendering, if the light sources have been detected in a scanned 3D map of the scene, it is then possible to consider illuminating a virtual object placed in the scene. In this forward lighting case, virtual light sources are placed in the scene and used to cast light and shadows onto the virtual objects. An image-based lighting (IBL) of the environment using a finite number of point lights is a solution proposed in the literature that provides realistic illumination. [7] proposes to use the Median Cut Algorithm to subdivide an HDR image into regions of equal light energy (or equal sum of pixel intensities), whereby each region represents one estimated light source. Another approach is to use classic K-means clustering [2] to segment the light sources from the HDR image. In [10] both clustering and Median-cut methods are combined. First a pre-processing step is used based on fuzzy C-means to segment the image into three subtypes: Bright, Medium and Dim. The Median Cut Algorithm is then applied independently on each of the three images analyse the distribution of the light intensities. With a real-time constraint, however, increasing the number of light sources decreases the performance of these methods.

Whilst computer graphics has focused on inverse-rendering, the geometrical aspects of visual perception have been much more widely studied in the computer vision literature where the fundamental problem of structure and motion estimation from a set of images is well defined. Image-based photometric rendering techniques have previously been studied in the context of multi-view structure from motion such as in [12], however, this approach is not efficient enough to be run in real-time. With the advent of more powerful hardware and improved methods, real-time approaches, known as 3D tracking and mapping have been demonstrated which allow new applications in Augmented-reality. Many techniques for real-time localisation and mapping estimate the geometry from a sparse set of features [15, 11, 37, 36]. Approaches in dense localisation and mapping [3, 27, 33, 41, 25], however, have shown that dense 3D tracking can be performed in real-time using the full-image. Few approaches, however, minimise both photometric and geometric errors directly in the sensor space [39, 25]. In the present context, knowing the scene geometry and the camera pose is quite useful for relighting a virtual scene and casting shadows correctly using image-based 3D environment-map.

Few approaches have been proposed which consider reflectance, light source positioning, occlusions and shadow mapping simultaneously in real-time with a moving camera. In [18] and [30] it is assumed that the camera observes a planar scene for which shadows and occlusions are estimated for an augmented reality application. More recently, [16] proposed to capture the light-field from a single hand-held camera observing a specular planar surface. The estimated environment map is then used to augment a planar surface with an object in real-time. Dense mapping approaches [25], on the other hand, are especially suited to handling arbitrary geometry and provide much information for the inverse rendering pipeline, however, they have been little exploited. In[13], a dense mapping approach was proposed and the environment map was decomposed using spherical harmonics, as proposed in [31]. This approach focuses on obtaining only very rough reflection-maps because reflective surfaces are used to replace the light-probe in order to acquire a reflected illumination map at 5Hz frame-rate. In that case the reflection-map is approximate and depends on observing reflective surfaces in the environment which is not always possible.

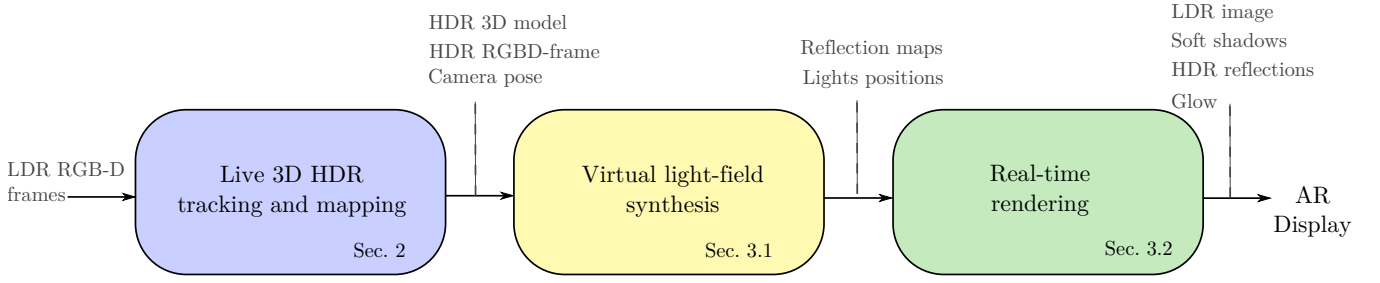


Figure 2: Augmented reality pipeline: A live stream of LDR images is input and a real-time AR video stream is output. Section 1 describes a new approach to dense tracking and mapping that fuses LDR views into a 3D HDR Model. Section 2.1 presents how individual reflection-maps are computed for any number of virtual objects. Section 2.2 details the rendering pipeline including reflective objects, shadows and glow.

Contributions

In this paper an approach is proposed which allows to directly estimate a dense 3D HDR environment-model (up to scale) in real-time and exploit that map for augmented reality relighting. Contrary to classic 2D HDR approaches, which assume that the camera is static, the approach proposed here will create a dense HDR scene-map in 3D from a live set of moving images. Furthermore, no light probe is required, nor is it necessary to compute reflection-maps indirectly via specular surfaces of the scene. Instead the approach consists in extending recent real-time dense mapping and 6D tracking approaches based on an image-based key-frame model as proposed in [3, 24] to *directly* observe and acquire the 3D environment-map and subsequently the entire 4D light-field covering the amount of light faring in every direction through every point in space. The resulting dense 3D HDR map is therefore of much higher quality than a single light probe and can be used to render virtual spherical reflectance maps at any given virtual object positions. Moreover, only a single 3D environment-map is required for the set of all virtual object locations within that scene which differs to a light-probe which has to be individually placed for each virtually rendered object. Whilst the present paper is formulated around using stereo and projective light RGB-D sensors, we also note highly relevant work in [26] which proposes a graph of *spherical* RGB-D key-frames which are ideally suited to the purpose of measuring light-fields. Another contribution of the approach proposed here is that automatic shutter variations are modelled and exploited to capture dense HDR maps without the need to manually vary the exposure settings.

Additionally, to performing reflection mapping of virtual objects in real-time, the dense 3D map is also used to detect light sources for shadow rendering. Given a detected light source, virtual light sources can be placed in the same pose so that shadows can be cast from the virtual objects onto the dense map of the real world. In particular the approach proposed will estimate the position, direction and intensity of a reduced number of light sources from an HDR image. Two solutions proposed in the literature are implemented, optimised for real-time performance and compared. The first being the Median-cut algorithm for image-based lighting proposed in [7]. The second being to used a classic K-means algorithm [2] to segment the brightest regions in the HDR image.

The reader is invited to refer to Figure 2 for a global overview of the full Augmented reality pipeline. In the next section an the dense visual tracking and 3D HDR mapping approach will be introduced and an illustration is given in Figure 3.

2 DENSE VISUAL TRACKING AND 3D HDR MAPPING

2.1 Sensor model

Consider a calibrated RGB-D sensor with a colour brightness function $\mathbf{I} : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{L}; (\mathbf{p}, t) \mapsto \mathbf{I}(\mathbf{p}, t)$ and a depth function $\mathbf{D} : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}^+; (\mathbf{p}, t) \mapsto \mathbf{D}(\mathbf{p}, t)$, where $\Omega = [1, n] \times [1, m] \subset \mathbb{R}^2$,

$\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{nm})^\top \in \mathbb{R}^{nm \times 2} \subset \Omega$ are pixel locations within the image acquired at time t , and $n \times m$ is the dimension of the sensor's images and $\mathbb{L} = [0, 1]$ is the normalised low dynamic range luminance interval.

It is convenient to consider the set of measurements in vector form such that $\mathbf{I}(\mathbf{P}, t) \in \mathbb{L}^{nm \times 1}$ and $\mathbf{D}(\mathbf{P}, t) \in \mathbb{R}^{nm \times 1}$. Note that t and \mathbf{P} may be omitted in these functions for clarity.

$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{nm})^\top \in \mathbb{R}^{nm \times 3}$ is defined as the matrix of 3D vertices related to the surface according to the following point-depth back-projection:

$$\mathbf{v}_i = \mathbf{K}^{-1} \bar{\mathbf{p}}_i \mathbf{D}(\mathbf{p}_i), \quad (1)$$

where $\mathbf{K} \in \mathbb{R}^{3 \times 3}$ is the intrinsic matrix of the camera. $\bar{\mathbf{p}}_i$ are the homogeneous pixels coordinates. For clarity, we will also consider \mathbf{V} to be a 3D vertex function such that $\mathbf{V} : \Omega \times \mathbb{R}^+ \rightarrow \mathbb{R}^3; (\mathbf{p}, t) \mapsto \mathbf{V}(\mathbf{p}, t)$. The set $\mathcal{S} = \{\mathbf{I}, \mathbf{V}\}$ is therefore defined to be an augmented image containing both intensities and vertices for each pixel.

2.2 Image-based 3D model

The 3D representation considered here is based on a graph of N augmented images $\mathcal{G} = \{\mathcal{S}_1^*, \dots, \mathcal{S}_N^*\}$, where each node of the graph is the 6 dof twist $\mathbf{x} = (\mathbf{v}, \boldsymbol{\omega}) \in \mathbb{R}^6$ that connects two images.

The twist is related to a 3D pose $\mathbf{T} = (\mathbf{R}, \mathbf{t}) \in \mathbb{SE}(3)$ via the exponential map as $\mathbf{T} = e^{[\mathbf{x}]_\wedge}$ with the operator $[\cdot]_\wedge$ as:

$$[\mathbf{x}]_\wedge = \begin{bmatrix} [\boldsymbol{\omega}]_\times & \mathbf{v} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (2)$$

where $[\cdot]_\times$ is the skew symmetric matrix operator, $\mathbf{R} \in \mathbb{SO}(3)$ is a rotation matrix and $\mathbf{t} \in \mathbb{R}(3)$ a translation vector.

As proposed in [25], this 3D model is build incrementally in a SLAM approach, and is used to predict a dense virtual image by rasterising and blending nearby key-frames at a desired camera pose within the model. The predicted augmented reference frame denoted $\mathcal{S}^* = \{\mathbf{I}^*, \mathbf{V}^*\}$ can then be used to perform a dense registration with a current live frame \mathcal{S} . A superscript $*$ will be used throughout to designate the predicted reference view variables.

2.3 Low dynamic range registration

This is the classic case used for image-based realignment for which it is assumed that the exposure of the camera remains fixed [3]. The objective here is to register a *current* image \mathbf{I} with an augmented *reference* image \mathcal{S}^* predicted from the 3D model, where \mathbf{I} is undergoing a full 3D transformation $\tilde{\mathbf{T}}$. The aim is to estimate the incremental twist transformation \mathbf{x} that satisfies

$$\tilde{\mathbf{T}} = \hat{\mathbf{T}} \mathbf{T}(\mathbf{x}), \quad (3)$$

where $\hat{\mathbf{T}}$ is an initial pose estimate of $\tilde{\mathbf{T}}$ (ie. initialised from the previous frame).

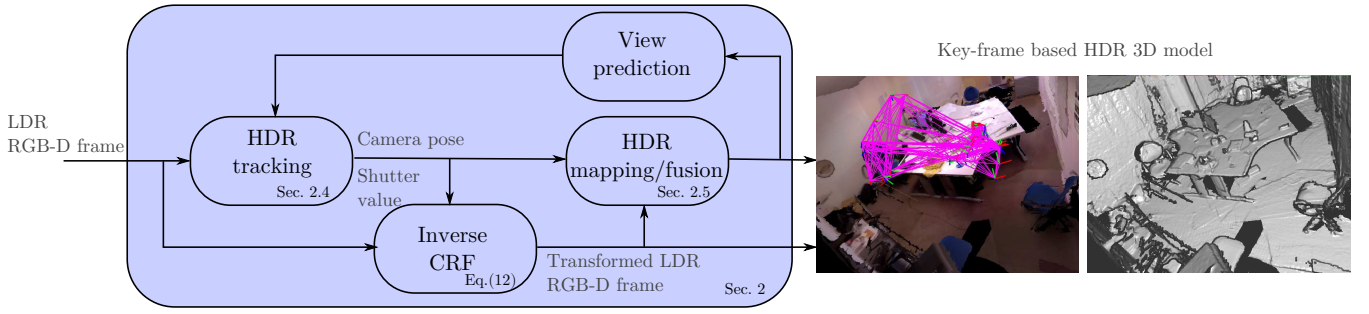


Figure 3: High level HDR tracking and mapping pipeline. A live RGB-D frame is registered to a 3D HDR map by estimating the pose and shutter values. The incoming LDR images are then fused with the HDR map to improve the dynamic range. A HDR environment map is available as output for rendering purposes. On the right, a graph of key-frames is displayed in the image, with the edges of the graph in violet.

Assuming brightness consistency and that both \mathbf{I} and \mathbf{I}^* have the same exposure, the 6 dof unknown \mathbf{x} can be estimated by minimising the following non-linear intensity error:

$$\mathbf{e}(\mathbf{x})_{ldr} = \left[\mathbf{I} \left(w(\hat{\mathbf{T}}\mathbf{T}(\mathbf{x}), \mathbf{V}^*) \right) - \mathbf{I}^*(\mathbf{P}^*) \right], \quad (4)$$

where the warping function $w(\hat{\mathbf{T}}\mathbf{T}(\mathbf{x}), \mathbf{V}^*)$ warps the vertices \mathbf{V}^* , associated with the back-projected pixels \mathbf{P}^* from equation (1), with the transformation $\hat{\mathbf{T}}\mathbf{T}(\mathbf{x})$ onto the normalised image plane:

$$\bar{\mathbf{p}}^w = \mathbf{K}\mathbf{\Pi}\tilde{\mathbf{v}}^*. \quad (5)$$

The matrix $\mathbf{\Pi} = [\mathbf{I}_{3 \times 3}, \mathbf{0}] \in \mathbb{R}^{3 \times 4}$ projects 4 vectors onto 3 space. An overline will be used to indicate homogeneous coordinates normalised wrt. the last component. One difference with respect to classic approaches is that each LDR image is normalised in a range $[0, 1]$ instead of $[0, 255]$ for an 8 bit image.

2.4 High dynamic range registration

Now consider that the 3D key-frame model contains high dynamic range images, hence the reference image intensities are not clamped: $\mathbf{I}_{hdr}^* \in [0, \infty]$. The scene irradiance E , which is independent of the camera transformations, is related to its corresponding pixel intensities \mathbf{I} by the following camera response function:

$$\mathbf{I} = cr(\mathbf{E}\Delta t) \quad (6)$$

where the inner term corresponds to the light-exposure $\mathbf{E}\Delta t$ in units Jm^2 , received by the sensor during the exposure time, Δt . As in [8] we use the term irradiance, however, the quantity is actually weighted by the spectral response at the sensor site.

In this work it is assumed that the exposure setting Δt of the camera is unknown. If the electronic shutter or exposure time is varied (automatically or manually) over time between different captured images then it is still possible to estimate the relative exposure time $\alpha = \frac{\Delta t_1}{\Delta t_2}$ between two subsequent frames. If an exact exposure time is available then it is possible to recover the irradiance.

For real-time performance it is assumed here that the camera response function (CRF) is simple non-linear function:

$$\tilde{\mathbf{I}} = (\mathbf{E}\Delta t)^\gamma \quad (7)$$

where γ is the gamma correction factor which can be chosen manually or is usually provided by the constructor. For the remainder of the paper it will be assumed that the gamma correction has been removed from the sensor images such that $\mathbf{I} = \tilde{\mathbf{I}}^\frac{1}{\gamma}$.

The classic LR intensity error can then be re-formulated as:

$$\mathbf{e}(\mathbf{x})_{hdr} = \left[\mathbf{I} \left(w(\hat{\mathbf{T}}\mathbf{T}(\mathbf{x}), \mathbf{V}^*) \right) - (\hat{\alpha} + \alpha)\mathbf{I}_{hdr}^*(\mathbf{P}^*) \right], \quad (8)$$

where $\hat{\alpha} + \alpha$ is the unknown shutter estimate that linearly maps an HDR intensity to an LDR intensity. $\hat{\alpha}$ is the initial value (initialised from the last estimate) and α is the increment (initially $\alpha = 1$). Note that in the formulation of this objective function, the error is bounded by the LDR image space (even if the values remain floating point and thus HDR). This choice is somewhat arbitrary (the error could have been minimised in HDR space) but it allows in practice to maintain a known bound on the error $\mathbf{I} \subset [0, 255]$. In practice a depth error can also simultaneously be minimised with the intensity (see [25]) but for clarity this will not be detailed here. In particular, choosing a bounded error function allows to avoid tuning the weighting between intensity and depth components.

The unknowns \mathbf{x} and α are estimated using a standard re-weighted Gauss-Newton approach:

$$\begin{bmatrix} \mathbf{x} \\ \alpha \end{bmatrix} = -(\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{W} \mathbf{e}_{hdr}, \quad (9)$$

where \mathbf{W} is a diagonal weighting matrix of dimensions $nm \times nm$ obtained by robust M-estimator based on Huber's influence function which rejects un-modelled data such as moving object occlusions and local illumination changes. The $nm \times 7$ Jacobian matrix \mathbf{J} is computed using an efficient second order approach as in [35]

$$\mathbf{J} = \frac{1}{2} [\mathbf{J}_{esm} \quad \mathbf{I}^w + \mathbf{I}_{hdr}^*] \quad (10)$$

where $\mathbf{J}_{esm} \in nm \times 6$ is the Jacobian of \mathbf{x} computed as in [3]. $\mathbf{I}^w + \mathbf{I}_{hdr}^* \in nm \times 1$ is the Jacobian matrix of α and \mathbf{I}^w is the warped image of equation (8).

The homogeneous pose estimate $\hat{\mathbf{T}}$ and the linear gain $\hat{\alpha}$ are finally updated incrementally using the following updates:

$$\begin{cases} \hat{\mathbf{T}} \leftarrow \hat{\mathbf{T}}\mathbf{T}(\mathbf{x}) \\ \hat{\alpha} \leftarrow \hat{\alpha} + \alpha \end{cases} \quad (11)$$

and the minimisation is iterated until \mathbf{x} and α are sufficiently small.

2.5 3D High dynamic range mapping

Once the current camera pose and shutter increments have been estimated with respect to the HDR space, the goal is to update the 3D model with new HDR measurements. The current image can be converted to the common HDR space using the estimated gain α :

$$\mathbf{I}_{hdr} = \frac{\mathbf{I}}{\alpha}. \quad (12)$$

Similarly to the super-resolution fusion approach in [25], the current frame in HDR space $\mathcal{I}_{hdr} = \{\mathbf{I}_{hdr}, \mathbf{V}\}$ is fused into a 3D model represented as a graph of key-frames (see Figure 4). In this process, the current image is warped onto the M closest key-frames in

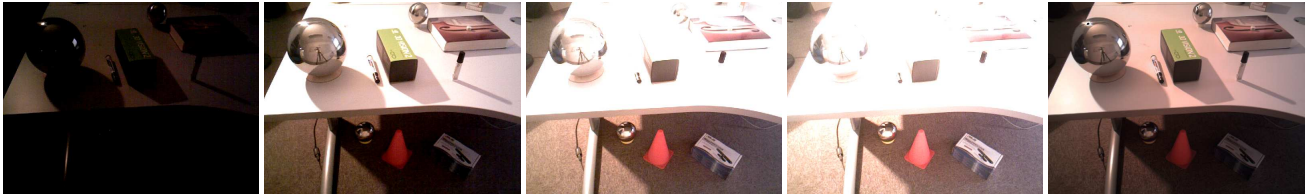


Figure 4: HDR fusion: From left to right same scene observed with increasing shutter values. The last image on the right is obtained using the HDR blending of equation (13) and is displayed using the tone-mapping of [23].

the graph using the estimated pose and the transformed intensities from (12). Integration of the intensities is then performed using an incremental weighted average of each pixel \mathbf{p} .

Each key-frame image \mathbf{I}_{hdr}^* and its cumulative weights \mathbf{C}_{hdr}^* are updated incrementally between time $t - 1$ and time t as:

$$\begin{aligned} \mathbf{C}_{hdr}^*(\mathbf{p}, t) &\leftarrow \mathbf{C}_{hdr}^*(\mathbf{p}, t-1) + f(\mathbf{I}^w(\mathbf{p}, t)) \\ \mathbf{I}_{hdr}^*(\mathbf{p}, t) &\leftarrow \frac{f(\mathbf{I}^w(\mathbf{p}, t))\mathbf{I}_{hdr}^w(\mathbf{p}, t) + \mathbf{C}_{hdr}^*(\mathbf{p}, t-1)\mathbf{I}_{hdr}^*(\mathbf{p}, t-1)}{\mathbf{C}_{hdr}^*(\mathbf{p}, t)} \end{aligned} \quad (13)$$

where \mathbf{I}^w and \mathbf{I}_{hdr}^w are respectively the current LDR and HDR images warped onto the reference key-frame. $f(\mathbf{I}(\mathbf{p}))$ is the weighting function of [8] which is used to compute a weight from the original LDR measurements $\mathbf{I}(\mathbf{p})$, defined as

$$f(\mathbf{I}(\mathbf{p})) = \begin{cases} \mathbf{I}(\mathbf{p}) & \text{if } \mathbf{I}(\mathbf{p}) \leq 0.5 \\ 1 - \mathbf{I}(\mathbf{p}) & \text{if } \mathbf{I}(\mathbf{p}) > 0.5 \end{cases} \quad (14)$$

This blending function is a simple hat function that gives a lower weight when an intensity value is close to the sensor saturation bounds of $[0, 1]$ and favours intensities in images that are closer to the center of the dynamic range of that image. This allows to continually integrate new information into the key-frame images of the graph by weighting higher dynamic ranges better.

Figure 4 illustrates the HDR fusion process of a single reference image. The first 4 images are captured with increasing shutter values. It can be seen that, depending on the shutter time, each individual image is not able to capture the full intensity range of the scene because several regions are either under- or over-exposed. The last image is the HDR reconstruction obtained with equation (13). For display purposes it has been converted to LDR space with non-linear tone mapping [23]. It can be seen that the HDR reconstruction allows to represent the full scene range without saturations.

In order to map large environments, the amount of occluded pixels between the current frame and the predicted 3D model frame are computed and monitored. This allows to detect unvisited areas and extend the 3D environment-map with new HDR key-frames (see [24] for more detail).

3 AUGMENTED REALITY

The main underlying objective of this work is to perform photo-realistic augmented reality in such a way that the virtual objects are indistinguishable from the real scene. Most state-of-the-art works in augmented reality focus on positioning and orientating the virtual object(s) w.r.t the viewer producing promising results, however, photo-realistic rendering is rarely considered in live AR. Indeed visual cues are present for real objects at many levels, ranging from environment reflections to shadows and this needs to be integrated into the rendering of virtual objects for high realism.

In this paper the augmented reality pipeline is divided in two steps. The first step, illustrated in Figure 5 consists in reconstructing reflection maps at arbitrary locations and detecting the light positions. Assuming a static environment, it only necessary to perform this once when the 3D model is reconstructed. The second

step, illustrated in Figure 6, is performed for each live image and consists in augmenting the live images with virtual objects containing HDR environment reflections and shadows.

3.1 Virtual light-field synthesis

As shown on Figure 5, the input to the virtual light-field pipeline is the graph of HDR key-frames and the virtual objects positions. The output is a set of HDR reflection maps (one for each object) and a set of light positions with their pre-computed depth-maps. The following sections will describe each step of the light-field synthesis.

3.1.1 Generating reflection maps

When a sufficiently complete 3D model of the environment is reconstructed, classic 2D HDR reflection-maps can be generated within the scene at arbitrary 3D locations. For a desired rendering location (typically specified by the user-application in the current image), each reflection-map is obtained by rasterising and blending the entire graph of HDR key-frames onto the six faces of a cube-map positioned at the selected 3D location. The blending function of equation (13) is used to obtain un-saturated measurements in HDR space. To take into account reflections between virtual objects, the reflection maps are generated recursively for each object. Additionally the scene depth is stored in a second cube-map in order to detect lights positions as described in Section 3.1.2 and also to allow shadow casting on the reflection maps.

3.1.2 Detecting light sources

The aim here is to detect light sources for casting shadows onto the dense scene map. Two approaches have been implemented, the K-means method [2] and the Median Cut approach. To simplify computation, the light sources are only detected in a single generated reflection map (corresponding to a single virtual object).

For the K-means method, each sample is considered as being composed of three values: the position of the pixel in the image (u, v) and the intensity of the pixel $I(u, v)$. First the images are filtered for high intensity pixels (above a given threshold). Then n clusters corresponding to n light sources are estimated using K-means. Finally the light source position is given as the centre of each cluster and its direction is chosen to point toward the centre of the HDR map (which is the centre of the virtual object).

The K-means method aims to minimise the sum of squared distances between all points and the cluster centre:

1. Choose K initial cluster centres.
2. Assign each image observation to the cluster whose mean is closest to it.
3. Estimate the new centres of the new clusters.
4. Iterate steps two and three until the assignments no longer change.

The Median Cut algorithm is computed as follows:

1. The image is divided by a line \mathbf{l} (along the longest dimension) into 2 sub-images \mathbf{I}_1 and \mathbf{I}_2 .
2. The splitting line is moved until the optimal line position is found : $\arg\min\{\mathbf{l}\} (\sum \mathbf{I}_1 - \sum \mathbf{I}_2)$.

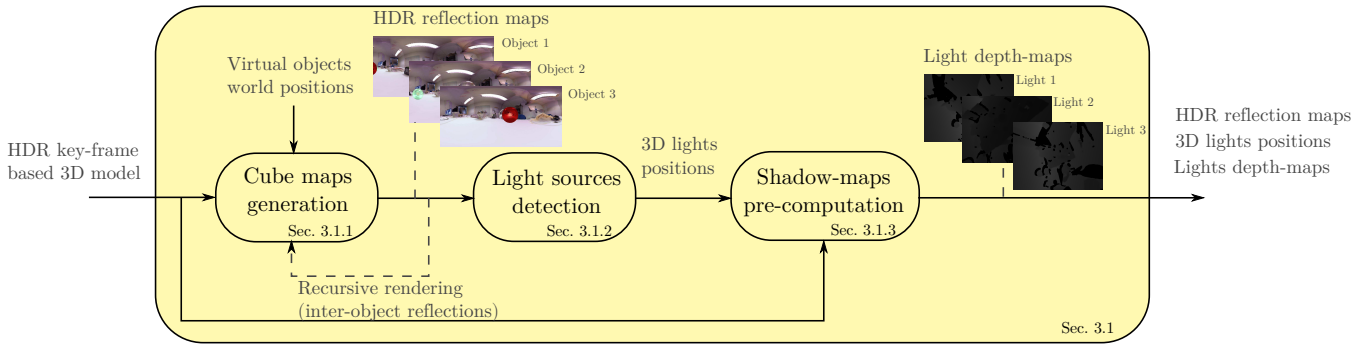


Figure 5: Virtual light-field composition pipeline. The 3D HDR model is used to create virtual light probes, detect light sources and to perform pre-computations for real-time shadow mapping. This provides necessary scene information for photo-realistic AR.

3. Recurse to step one with each of the 2 new sub-images until the initially fixed number of regions is reached.
4. Estimate the position of the light (which is the barycentre) in the sub-images.

The Median Cut is accelerated using Integral Images [40].

3.1.3 Shadow-maps pre-computation

Real-time shadow casting is performed using variance shadow maps [9] which allow to generate smooth shadows with a minimal computational cost. To efficiently compute light casting onto the dense 3D surface, the lights' depth-maps are pre-computed and pre-filtered once (since a quasi-static background 3D environment-map is considered). The 3D model (several key-frame's depth maps) is rendered and blended at each light position and the corresponding depth values and variances are stored in a floating point texture. One of the major drawbacks of shadow maps is that their computational cost grows linearly with the number of lights, which restricts the maximum number of lights for real-time constraints (≈ 10 lights for the hardware described in Section 4).

3.2 Rendering

As shown on Figure 6, the rendering pipeline input is the current RGB-D frame converted to HDR space and its 3D pose. The output is an augmented image, containing virtual objects, HDR reflections and shadows projected onto the real 3D scene. The pipeline can be divided in several steps which are described in the following section, whilst Figure 7 shows the intermediate results of the pipeline.

3.2.1 Reflection mapping

The generated reflection-maps (2D omni-directional environment-maps) can be used to perform classic HDR reflection mapping [6] of virtual objects positioned in the scene (at the same position as the light-field maps). The objects are rendered in the scene using classic rasterisation and a depth test is performed with the current depth-map in order to handle occlusions. The standard reflection model is computed in the fragment shader (screen space), for each given incident vector \mathbf{v} and surface normal \mathbf{n} . This model returns the reflection direction calculated as $\mathbf{v} - 2\mathbf{n}^T \mathbf{v}\mathbf{n}$. The reflected direction allows to perform efficient texture look-up in the reflection-map associated to the virtual object.

3.2.2 Real-time shadow casting

Real-time shadow casting is achieved using the pre-computed light depth-maps, and the current frame depth-map. Each vertex of the current frame is projected onto each light depth-map, in order to perform a depth comparison which indicates if the pixel is shadowed or illuminated. To avoid aliasing and also to obtain more realistic shadows, we use variance shadow maps. Instead of storing

a single depth value in the shadow map, an overlapping Gaussian distribution (mean and variance) is maintained for each pixel. This allows to weight the shadow value with the variance of the shadow map, which creates smooth shadows edges (see Figure 7(c)). The interested reader can refer to [9] for additional details.

3.2.3 Forward CRF

Since the inputs of the rendering pipeline are in HDR space (image and HDR reflections), it is necessary to convert the output augmented image into LDR space for on-screen display. Whilst any kind of tone-mapping could be performed, it is proposed here to simply apply the estimated α from equation (8) followed by the desired gamma correction, which corresponds to transforming back the image to its original colour space.

3.2.4 Glow mapping and display

For additional realism a glow-map is also used to blur the saturated reflections in the final image which simulates lens blooming. This is achieved by first down-sampling the augmented image, then extracting the pixels with an intensity value above a saturation threshold (e.g ≥ 1.0) and finally blurring the segmented image with a Gaussian filter. The final glow-map is then blended with the current image using an additive blending function and the result is displayed on screen. Figure 7(e) shows the glow-map obtained from the image of Figure 7(d). It can be seen that only saturated pixels (i.e the lights) are extracted and blurred.

4 RESULTS

The entire 3D HDR tracking and mapping pipeline along with augmented reality rendering is implemented on the GPU using the OpenCL library with OpenGL interoperability and GLSL shaders. The entire algorithm runs robustly and accurately at frame-rate (30Hz) on a standard desktop PC running Ubuntu 12.10 with an Intel Core i7 2700k and a 2GB nVidia GeForce GTX670.

The approach has been successfully tested on various environments with an Asus Xtion RGB-D sensor in automatic shutter mode. According to the constructor, the RGB image delivered by the sensor is by default non-linearly corrected with a gamma correction of 2.2. In order to avoid non-linear intensity changes induced by shutter variations, the gamma correction is removed from the images before processing and re-applied before display as in (7).

Figure 8 shows the resulting HDR reflection-maps acquired after performing dense mapping of an entire room in real-time using an image-based key-frame 3D model. The automatic shutter of the camera varies naturally as the camera scans the dark and lighter parts of the scene allowing to capture the full dynamic range of the office. The LDR images with different shutter settings have been incrementally warped and fused into the same map in a dense

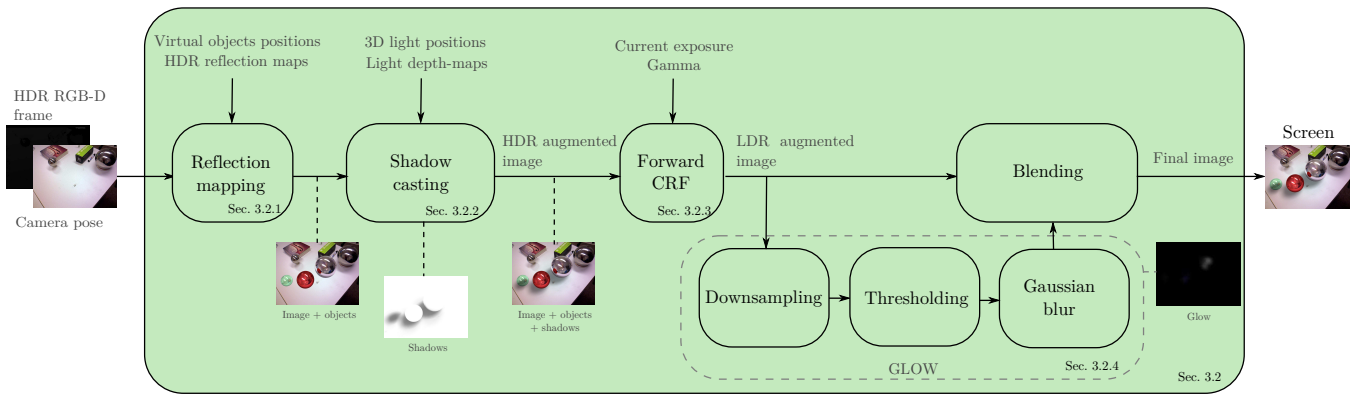


Figure 6: Augmented reality rendering pipeline. The AR pipeline provides all necessary information to perform real-time reflection mapping from each cube-map, shadow casting from the detected lights, conversion to LDR for display and application of a realistic glow filter.

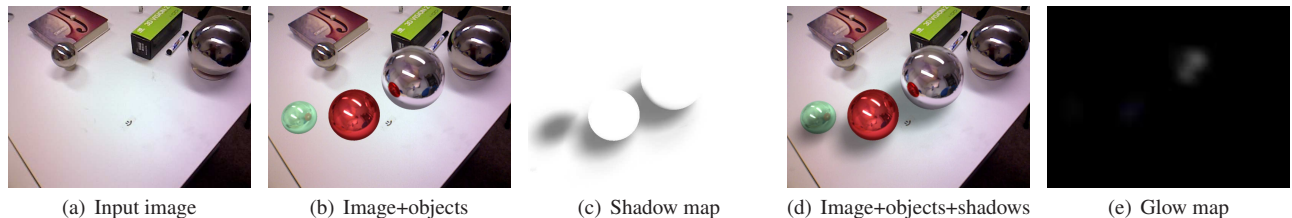


Figure 7: Enlarged images of intermediate results from Figure 6. The final output is shown in Figure 1.

manner. In Figure 8(a) an HDR reflection map is shown after non-linear tone mapping allowing the full range to be rendered into the image. In Figure 8(b) a second image of the same scene is rendered to the right that shows the degraded performance when an LDR image observes only part of the dynamics of the scene.

4.1 Light source estimation

The dense HDR light-field image is used as input to a light estimation algorithm that looks to obtain the 3D position, orientation and intensity of several unknown light sources. Two methods, the Median-cut [7] and K-means [2] have been implemented and their results are shown in Figure 9. It can be seen visually that the principal lights in the HDR image are detected with the K-means approach, whilst the Median-cut method clearly has larger regions for the darker regions and smaller regions for the brightest ones (each region is divided according to equivalent intensity).

When analysing these two approaches, the Median-cut clearly provides correct lighting, however, it requires a much larger number of lighting sources (mostly artificial) to achieve this. The solution to segment the brightest regions in the image using a K-means algorithm is much more efficient, however, it requires a manually chosen threshold to do a first pass and choose the bright pixels in the image. If the number of lights is greater than the K-means becomes computationally expensive.

An important difficulty for each of these approaches is being able to detect the number of light sources automatically. The Median-cut method does not try to detect light sources but simply represents equal regions of illumination. In that case reflections and other light properties are nicely modelled at the expense of a large number of light emitting regions. The problem with the K-means is that it needs to determine the number of lights. This requires beginning with a high number of clusters and once the centres are computed they are merged to reduce the number of regions. This can handle large lights such as fluorescent tubes.

4.2 Real-time rendering

The previous results have shown how to obtain a dense HDR light-field map in real-time using a dense key-frame localisation and mapping approach. This was then used to estimate light positions. Now that all the essential photometric properties such as scene geometry, illumination and reflectance properties have been estimated it is possible to render virtual objects seamlessly into the real video sequence for a realistic augmented reality experience.

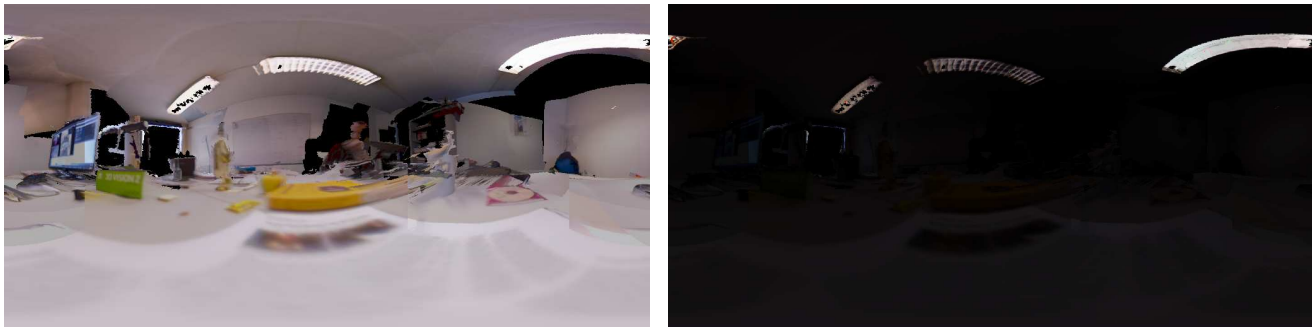
4.2.1 Comparison with ground truth

A qualitative comparison has been performed using real and virtual light-probes that were both placed manually at the same location in the scene. Figure 10(a) shows the image obtained using the real-light probe and Figure 10(b) shows that of the synthetic light probe. It can be seen that the HDR reflections are quite similar, the 3 lights are correctly reflected, as well as the rest of the far environment. The close environment is, however, less accurate, since the proposed approach uses a cube-map (or sphere) centred on the object in order to perform fast texture look-ups in the reflection-map. This does not consider the parallax between the object surface and the centre of projection of the cube-map, which yield to false reflections for close surfaces. Another difference between the two images is the colour of the reflected environment. Indeed the real light-probe that was used for the experiment has its own colour response, whilst the virtual light-probe is a perfect mirror.

4.2.2 Live results

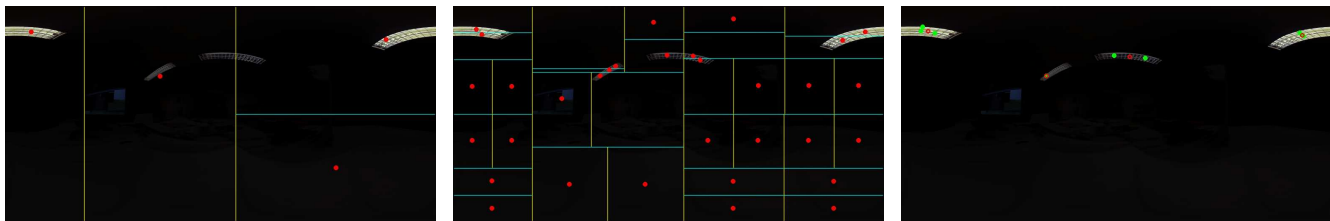
Figure 1 from the teaser along with Figure 11 depict different augmented scenes that were rendered and interacted with in real-time. The associated video with this article demonstrates the real-time interactive nature along with the realism of the rendering. The viewer can also note the stability of the localisation and mapping approach with respect to fast movements.

In Figures 11(a) and 11(b) a rendered teapot can be seen along with two reflective balls. Several real reflective objects such as CDs



(a) HDR reflection-map after non-linear tone mapping with full visible range. (b) The same image with an 8 bit range corresponding to a single input image.

Figure 8: An HDR reflection-map



(a) Median Cut for 4 lights.

(b) Median Cut for 32 lights.

(c) K-means merged from 10 to 4 lights.

Figure 9: In (a) and (b) the regions detected by the Median-cut algorithm are shown. In (c) the green points correspond to the points which have been detected before merging and the red points are the final result with merging.

and aluminium packaging have also been placed in the scene for comparison. It can be seen that the reflectance map provides very high realism and it is possible to identify the various light sources in the room within the reflections on both real and virtual objects. The various surrounding objects can also be identified.

In Figures 11(c) and 11(d) a second augmented scene is presented. It contains four virtual objects: a teapot, a green sphere, a red sphere and a full reflective sphere. Three real light probes and a CD are also in the scene. It can be seen that the virtual objects are well integrated in the scene, thanks to the reflection mapping and shadow mapping. Virtual inter-object reflections are also visible on the teapot and the larger sphere, where the green and red spheres are reflected along with the real CD.

4.3 Limitations

The proposed 3D HDR mapping pipeline assumes that the camera has a simple non-linear CRF. However [8] showed that radiance may map in complex non-linear ways to the pixel values given by a camera. For professional cameras, this CRF can be calibrated by taking several images with different exposure settings and minimising a non-linear error function which yields the CRF. For low-cost sensors, manually setting the exposure time is, however, rarely possible which means that the environment illumination has to be uniformly controlled in order to calibrate automatic shutter variations making the entire procedure much more complex. Another limitation comes from the combination of low frequency fluorescent lights (ie. 50 Hz) with a rolling shutter sensor: this creates a periodic light change across the image when the shutter time is smaller than a light period (flickering).

As stated in Section 4.2.1, another limitation of the proposed approach (and all real-time approaches) is that the reflection maps capture the light-field at the centre of the object. This does not consider the parallax between the object surface and the centre of projection of the cube-map, which yield to false reflections for close surfaces. In the given experiments, those artefacts are visible where

the reflected spheres touch the surface of the table, however, as can be seen in Figure 10, this does not affect the photo-realism.

Finally, only static objects were considered so that rendering remained real-time and efficient. This is due to the fact that with static objects it is simple to pre-compute the reflection and shadow maps. Future works will aim at generating dynamic virtual light-fields for moving objects as well as global illumination.

5 CONCLUSION

This paper has demonstrated a method for acquiring a 3D HDR environment-map and using it to synthesise virtual light-fields at arbitrary positions in a scene at high-quality in real-time. The proposed approach does not require any light-field probe nor does it need to capture an approximate light-field through the observation of surface reflectance. Additionally, it is assumed that it is not possible to manually control the shutter as is the case in low-cost cameras. As such the dynamic range of each LDR image is estimated whilst simultaneously localising a moving camera in 6 dof and mapping an environment densely to performing 3D HDR compositing. Estimating the sensor's exposure settings also solves the problem of dense tracking and mapping in the presence of unknown shutter variations. In the results it is shown that the approach successfully recovers a dense 3D HDR environment-map that can be interactively modified by choosing a desired dynamic range for the rendered image. Subsequently, this approach allows to transform an RGB-D camera sensor into a light-field sensor and perform 3D image-based rendering in real-time. In the results a virtual spherical light-field sensor is used to render reflective virtual objects for an augmented reality scenario. This provides a highly realistic rendering of the virtual objects in the scene in real-time. This realism is further augmented via the automatic detection of light sources in HDR maps which are then used to cast shadows and deforming them correctly onto a dense 3D map of the scene.



(a) Reflections on a true light-probe (ground truth).



(b) Reflections on a synthetic light-probe.

Figure 10: A qualitative comparison between a true light probe and a synthetic one. Whilst highly realistic (especially in the moving sequence from the video), the parallax effects due to near objects can be observed (i.e. the CD on the table). The lights and various objects in the scene can be clearly identified in both real and virtual objects at the same positions. A difference in material colour can also be noticed.



(a)



(b)



(c)



(d)

Figure 11: Augmented reality results obtained in real-time from the 3D HDR mapping approach. (a) and (b) show images from the live AR sequence. The reflections and shadows provide a highly noticeable change in realism when compared with AR images without these effects (see Figure 7(d)). (c) and (d) show more images with ground truth light probes. Additionally inter-object reflections have been computed and can be observed in the virtual objects (not the real ones).

ACKNOWLEDGEMENTS

The present work is funded by the French DGA Rapid project Fraudo on dense visual SLAM in real-time.

REFERENCES

- [1] M. Ali and S. Mann. Comparametric image compositing: Computationally efficient high dynamic range imaging. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012.
- [2] M. J. B. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1967.
- [3] A. I. Comport, E. Malis, and P. Rives. Accurate quadrifocal tracking for robust 3d visual odometry. In *IEEE Int. Conf. on Robotics and Automation*, 2007.
- [4] M. Corsini, M. Callieri, and P. Cignoni. Stereo light probe. *Comput. Graph. Forum*, 27(2), 2008.
- [5] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Annual Conference on Computer graphics and interactive techniques*, 1996.
- [6] P. Debevec. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, 1998.
- [7] P. Debevec. A median cut algorithm for light probe sampling. In *ACM SIGGRAPH Posters*, New York, NY, USA, 2005.
- [8] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '97, 1997.
- [9] W. Donnelly and A. Lauritzen. Variance shadow maps. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, 2006.
- [10] J. Doug Yoo and K. H. Lee. Light source estimation for realistic shadow using segmented hdr images. In *International Symposium on Ubiquitous VR*, 2007.
- [11] N. Engelhard, F. Endres, J. Hess, J. Sturm, and W. Burgard. Real-time 3d visual slam with a hand-held rgb-d camera. In *RGB-D Workshop on 3D Perception in Robotics*, 2011.
- [12] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.
- [13] L. Gruber, T. Richter-Trummer, and D. Schmalstieg. Real-time photometric registration from arbitrary geometry. In *IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2012.
- [14] W. Heidrich and H.-P. Seidel. View-independent environment maps. In *Proceedings of the ACM SIGGRAPH workshop on Graphics hardware*, 1998.
- [15] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In *International Symposium on Experimental Robotics*, 2010.
- [16] J. Jachnik, R. Newcombe, and A. Davison. Real-time surface light-field capture for augmentation of planar specular surfaces. In *IEEE International Symposium on Mixed and Augmented Reality*, 2012.
- [17] K. Jacobs and C. Licos. Classification of illumination methods for mixed reality. *Computer Graphics Forum*, 25(1), 2006.
- [18] M. Knecht, C. Traxler, O. Mattausch, W. Purgathofer, and M. Wimmer. Differential instant radiosity for mixed reality. In *IEEE/ACM International symposium on mixed and augmented reality*, 2010.
- [19] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996.
- [20] R. C. H. Lo, S. Mann, J. Huang, V. Rampsad, and T. Ai. High dynamic range (hdr) video image processing for digital glass. In *Proceedings of the 20th ACM international conference on Multimedia*, 2012.
- [21] S. Mann. Compositing multiple pictures of the same scene. In *the 46th Annual IS&T Conference*, 1993.
- [22] S. Mann, R. Lo, J. Huang, V. Rampsad, and R. Janzen. Hdrchitecture: real-time stereoscopic hdr imaging for extreme dynamic range. In *ACM SIGGRAPH 2012 Emerging Technologies*, 2012.
- [23] R. Mantiuk, K. Myszkowski, and H. P. Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept.*, 2006.
- [24] M. Meilland and A. Comport. Super-resolution 3D Tracking and Mapping. In *IEEE Int. Conf. on Robotics and Automation*, Karlsruhe, Germany., 2013.
- [25] M. Meilland and A. I. Comport. On unifying image and model-based real-time dense mapping at large scales. In *IEEE Int. Conf. on Intelligent Robots and Systems*, 2013.
- [26] M. Meilland, A. I. Comport, and P. Rives. A spherical robot-centered representation for urban navigation. In *IEEE Int. Conf. on Intelligent Robots and Systems*, 2010.
- [27] R. A. Newcombe, S. Lovegrove, and A. J. Davison. Dtam: Dense tracking and mapping in real-time. In *IEEE Int. Conf. on Computer Vision*, 2011.
- [28] D. Nowrouzezahrai, S. Geiger, K. Mitchell, R. Sumner, W. Jarosz, and M. Gross. Light factorization for mixed-frequency shadows in augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality*, 2011.
- [29] G. Patow and X. Pueyo. A survey of inverse rendering problems. *Computer Graphics Forum*, 22(4), 2003.
- [30] J. Pilet, A. Geiger, P. Laguer, V. Lepetit, and P. Fua. An all-in-one solution to geometric and photometric calibration. In *IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2006.
- [31] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001.
- [32] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001.
- [33] H. Roth and M. V. J. McDonald. Moving volume kinectfusion. In *British Machine Vision Conference*, 2012.
- [34] I. Sato, Y. Sato, and K. Ikeuchi. Acquiring a radiance distribution to superimpose virtual objects onto a real scene. *IEEE Transactions on Visualization and Computer Graphics*, 5(1), 1999.
- [35] G. Silveira and E. Malis. Real-time visual tracking under arbitrary illumination changes. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2007.
- [36] J. Stückler and S. Behnke. Model learning and real-time tracking using multi-resolution surfel maps. In *AAAI Conference on Artificial Intelligence*, 2012.
- [37] J. Sturm, K. Konolige, C. Stachniss, and W. Burgard. 3d pose estimation, tracking and model learning of articulated objects from dense depth video using projected texture stereo. In *RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras*, 2010.
- [38] B. Tunwattapong, G. Fyffe, P. Graham, J. Busch, X. Yu, A. Ghosh, and P. Debevec. Acquiring reflectance and shape from continuous spherical harmonic illumination. In *ACM SIGGRAPH 2013*, 2013.
- [39] T. Tykkala, C. Audras, and A. I. Comport. Direct iterative closest point for real-time visual odometry. In *ICCV Workshop on Computer Vision in Vehicle Technology*, 2011.
- [40] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition*, 2001.
- [41] T. Whelan, H. Johannsson, M. Kaess, J. Leonard, and J. McDonald. Robust real-time visual odometry for dense RGB-D mapping. In *IEEE Int. Conf. on Robotics and Automation*, 2013.
- [42] G. Yang and Y. Liu. 3d object relighting based on multi-view stereo and image based lighting techniques. In *IEEE International Conference on Multimedia and Expo*, 2009.