



HAL
open science

Active High Dynamic Range Mapping for Dense Visual SLAM

Christian Barat, Andrew I. Comport

► **To cite this version:**

Christian Barat, Andrew I. Comport. Active High Dynamic Range Mapping for Dense Visual SLAM. IEEE/RSJ International Conference on Intelligent Robots and Systems, Sep 2017, Vancouver, Canada. hal-01635528

HAL Id: hal-01635528

<https://hal.science/hal-01635528v1>

Submitted on 15 Nov 2017

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.

Active High Dynamic Range Mapping for Dense Visual SLAM

Christian Barat¹ and Andrew I. Comport²

Abstract—Acquiring High Dynamic Range (HDR) photos from several images, with an active shutter providing different exposures (sensor integration periods), has been widely commercialised in photography for static camera positions. In the case of a mobile video sensor (as is the case in robotics), this problem is more difficult due to real-time motion of the sensor which transforms the perspective between the acquired images. HDR approaches for a set of images from different perspectives have therefore been significantly overlooked since this would require sophisticated dense mapping approaches to eliminate the motion component. Recent dense visual SLAM (Simultaneous Localization And Mapping) approaches provide this framework, however, few works have attempted to perform HDR visual SLAM. Current approaches are thus highly dependant on illumination conditions and camera shutter settings. In this paper a new approach is proposed that enables 3D HDR environment maps to be acquired *actively* from a dynamic set of images in real-time. The 6 DOF pose, the dense scene structure and the HDR texture map will be estimated simultaneously with the objective of maximising the dynamic range. This will allow to obtain a radiance map of the scene by fusing a real-time stream of low dynamic range images (LDR) into a graph of HDR key-frame images. In particular, a method is proposed to actively control the shutter based on information theory to optimise the information content of the 3D HDR environment map for RGB-D sensors. As will be shown in the results, a 3D HDR environment map allows robot localisation and mapping (visual SLAM) to take actively advantage of varying luminosity in different parts of the scene.

I. INTRODUCTION

In robotics, a fundamental problem is to localize the robot within its environment with precision and robustness. One of the most prominent solutions is to use a visual SLAM approach. Much computer vision research has been carried out to remain invariant to lighting conditions, but, inevitably these approaches rely on the intensity (brightness) of the acquired images. Few approaches have investigated actively controlling the camera exposure to obtain better quality high dynamic range maps. Generally, the chosen solution is to fix the shutter (exposure time) of the camera. Fixing the shutter can especially be a problem in an environment with a high dynamic radiance containing under illuminated zones where high sensitivity is required alongside high illumination where the sensor can saturate. In this work an *active* approach is proposed to control the variation of the shutter to obtain an optimal map of the dynamic radiance of the environment.

As a first approach, a simple method for extending the dynamic range of a camera is to combine multiple images of a scene taken with different exposures [1]. Often, the shutter is automatically controlled by the camera with proprietary methods developed by each manufacturer (and generally unknown). In the authors previous work, dense real-time localisation and mapping was performed simultaneously whilst reconstructing 3D High Dynamic Range environment maps [2]. In that initial work the shutter was not *actively* controlled and the black-box automatic camera shutter of the camera was used. Subsequently it was initially necessary to estimate the shutter value during the process since it was not provided by the driver. Following this first approach, [3] modified the approach slightly to perform tracking using the normalised radiance since it depends on luminance. This approach, however, also uses the default auto shutter.

To improve 3D HDR mapping, this paper proposes to optimise the dynamic range by actively controlling the shutter. To attain this objective it is necessary identify two models. The first model maps pixel intensities to irradiance values (the inverse model) [4] and the second maps pixel intensities between different ranges based on shutter values (the direct model). This second model provides a function to predict the value of the intensities when the shutter is varied.

In comparison to other works, this paper proposes to improve the dynamic of the radiance and not the dynamic of the pixel intensity. In [5], authors use a simple method based on the Mean Brightness Level ($B_{l_{mean}}$) and the Median Brightness Level ($B_{l_{med}}$). They use the difference between the ($B_{l_{mean}}$) and ($B_{l_{med}}$) to control the shutter, the objective is to obtain an image with a $B_{l_{mean}} \in [100, 130]$. In [6] they propose to fuse two images obtained with two different exposures. The system first runs an auto exposure algorithm to determine the initial guess X of the exposure setting. Then it takes two frames whose exposure are set as two times ($2X$) and half ($1/2X$) of the initial exposure. [7] proposes an AE (Auto-Exposure) control method for stereo cameras using Gaussian sampling, with more sampling in the center region such that the sensitivity to light from exterior areas decreases and the resource efficiency improves. The work presented in [8] proposes a computational method for compositing Low Dynamic Range (LDR) images into a High Dynamic Range by the use of Comparametric Camera Response Function. This approach uses a Look Up Table (LUT) to perform in real time, but, they use a fixed ratio of shutter values. [1] presents a method to select the exposure value to obtain an effective camera response. The camera response and the desired dynamic range is selected offline by the user.

In the approach proposed here, the optimal shutters values

¹Christian Barat is with Laboratoire I3S ,UNS-CNRS, 2000, route des Lucioles, 06903 Sophia-Antipolis, FRANCE barat@i3s.unice.fr

²Andrew I. Comport is with Laboratoire I3S ,UNS-CNRS, 2000, route des Lucioles, 06903 Sophia-Antipolis, FRANCE Andrew.Comport@cnrs.fr

This work is financed by the H2020 EU project COMANOID <http://www.comanoid.eu/>, RIA No 645097.

are determined based on an information theory framework. The aim is to maximise the entropy of combined radiance images. Alternatively to [1] and [8], the objective is to determine optimal shutter values automatically by maximising information content in real-time.

The main contributions can be summarized as follows:

- Active control of the camera shutter in real-time to perform visual SLAM.
- Optimal 3D HDR environment mapping by maximising the information content in high dynamic range image based on entropy.
- Performing visual SLAM with a varying shutter and HDR maps allows to handle varying luminosity in different parts of the scene as a robot explores different parts of the scene (i.e. shaded areas to brightly lit environments, etc.).

In Section II our classic dense RGB-D visual tracking method for LDR images is first presented. Section III introduces the adaptation of the dense visual tracking to HDR images. Finally, Section V presents the results for a static camera and a moving cameras in a real application.

II. DENSE VISUAL TRACKING

For completeness, our classic real-time dense visual odometry algorithm [9] will be presented for Low Dynamic Range Images (LDRI). Whilst many more recent dense approaches have followed and been celebrated extensively, the dense approach remains essentially the same as in this seminal work.

Consider a calibrated RGB-D sensor (stereo, projective light, etc - see section V) with a color 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 $\Omega = [1, n] \times [1, m] \subset \mathbb{R}^2, \mathbf{P} = (p_1, p_2, \dots, p_{nm})^T \in \mathbb{R}^{mn \times 2} \subset \Omega$ 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 luminance range. Consider the set of measurements in vector form such that $\mathbf{I}(\mathbf{P}, t) \in \mathbb{L}^{nm \times 1}$ and $\mathbf{D}(\mathbf{P}, t) \in \mathbb{R}^{nm \times 1}$. In the following t and \mathbf{P} can be omitted for clarity. $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_{nm})^T \in \mathbb{R}^{nm \times 3}$ is defined as the matrix of 3D vertices corresponding 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.

With abuse of notation, \mathbf{V} will also be considered as 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 $\mathcal{I} = \{\mathbf{I}, \mathbf{V}\}$ is therefore defined to be an augmented image containing both intensities and vertices for each pixel.

A. Image-based 3D model

The 3D representation is based here on a graph of N augmented images $\mathcal{J} = \{\mathcal{I}_1, \dots, \mathcal{I}_N\}$ where each link of the graph is the 6 DOF twist $\mathbf{x} = (\boldsymbol{\nu}, \boldsymbol{\omega}) \in \mathbb{R}^6$ that connects two images in the graph. The twist is related to a 3D pose

$T = (\mathbf{R}, \mathbf{t}) \in \mathbb{SE}(3)$ via the exponential map as $\mathbf{T} = e^{[\mathbf{x}]^\wedge}$ with the operator $[\cdot]^\wedge$ as:

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

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.

This 3D model is built incrementally in a SLAM approach [10], and is used to predict a dense virtual image by rasterising and blending the k closest key-frames at a desired camera pose within the 3D model. The predicted augmented reference frame denoted $\mathcal{I}^* = \{\mathbf{I}^*, \mathbf{V}^*\}$ can then be used to perform a dense registration with a current live frame \mathcal{I} . The superscript $*$ will be used to designate the predicted reference view variables.

B. Low dynamic range registration

In the classic case for image-based registration it is assumed that the exposure of the camera remains fixed. The objective here is to register a *current image* \mathbf{I} with an augmented *reference image* \mathcal{I}^* 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 $\tilde{\mathbf{T}}$ (i.e. initialised from the previous frame). Assuming brightness consistency (Lambertian material) 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:

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

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

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

Note that for clarity, in this paper, only the photometric error term in Eq. (4) is presented. It is also possible to simultaneously minimise a geometric error based on a depth image to improve this estimate [11].

The matrix $\mathbf{\Pi} = [\mathbf{I}_{3 \times 3}, \mathbf{0}] \in \mathbb{R}^{3 \times 4}$ is a projection matrix from dimension 4 onto dimension 3. An overline ($\bar{\mathbf{v}}$) 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 within the range of values $[0, 1]$ instead of the discrete $[0, 255]$ for an 8 bit image.

III. HIGH DYNAMIC RANGE REGISTRATION

This section now introduces the HDR registration approach by replacing the 3D key-frames with high dynamic ranges images. Now the reference image intensities are no longer clamped: $\mathbf{I}_{HDR}^* \in [0, \text{inf}]$. The light-exposure is

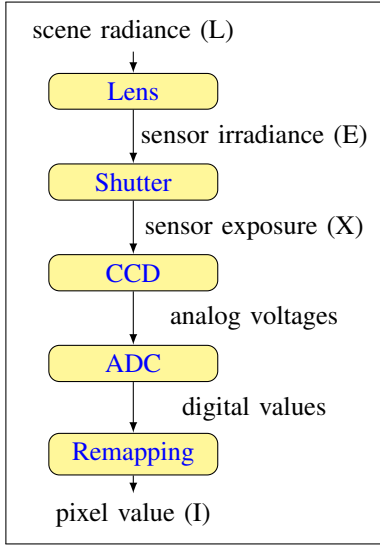


Fig. 1. Image Acquisition Pipeline. The scene radiance \mathbf{L} traverses the optics to illuminate the sensor with irradiance \mathbf{E} . The electronic shutter accumulates the irradiance during the time Δt before being sampled into discrete intensity values \mathbf{I} .

defined as the product of the irradiance \mathbf{E} at the sensor and exposure time Δt , see Eq. 6. The pixel intensity \mathbf{I} is obtained after transformation by the non-linear function f (representing the camera response function - CRF). The radiance \mathbf{L} is proportional to \mathbf{E} for any particular pixel, but the proportional factor can be different at different places on the sensor. In the following the irradiance \mathbf{E} will be used, considering that it is always possible to obtain \mathbf{L} after a CRF calibration step [12]. To obtain HDR Images it is necessary to estimate the inverse of the camera response to obtain the irradiance.

In [13], multiple photographs of the scene are taken with different amounts of exposure. The algorithm then uses these differently exposed photographs to calibrate the CRF of the imaging process, up to a scale factor, by using the assumption of reciprocity.

$$\mathbf{I}_{ij} = f(\mathbf{E}_i \Delta t_j) \quad (6)$$

If f is monotonic, then it is invertible, (6) becomes:

$$f^{-1}(\mathbf{I}_{ij}) = \mathbf{E}_i \Delta t_j \quad (7)$$

Using the logarithm gives:

$$\ln f^{-1}(\mathbf{I}_{ij}) = \ln \mathbf{E}_i + \ln \Delta t_j \quad (8)$$

Finally with $g = \ln f^{-1}$:

$$g(\mathbf{I}_{ij}) = \ln \mathbf{E}_i + \ln \Delta t_j \quad (9)$$

where i ranges over pixels and j ranges over exposure durations.

In the basic case with static images of the same scene (without sensor motion), but with a shutter varying from 1ms to 30ms, the model presented in Fig. 2 is obtained by using the optimisation method presented in [13]. It can be seen that the responses g of the channels are similar. In the following

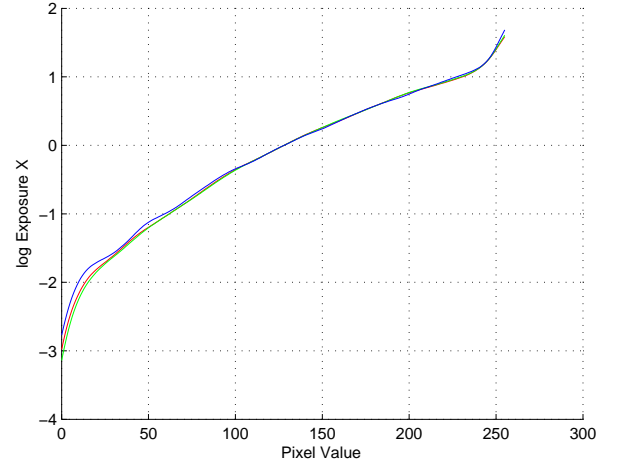


Fig. 2. Inverse Model for the three colors R-G-B. It can be noted that this function is non-linear and therefore the HDR fusion is equally non-linear.

an inverse model will be used for the intensity of the pixel. This is defined by the mean response of the three channels:

$$g_{intensity} = (g_{red} + g_{green} + g_{blue}) / 3$$

IV. DYNAMIC IRRADIANCE OPTIMIZATION

During SLAM the 3D model (key-frame graph) is updated with new log irradiance key-frame images: $\ln(E)$. When the 3D model is updated, the log irradiance images are blended. Each key-frame image \mathbf{I}_{HDR}^* and its cumulative weights \mathbf{C}_{HDR}^* (defined by Eq. 10) are subsequently updated, for each point p , incrementally between time $t - 1$ and time t as:

$$\begin{aligned} \mathbf{C}_{HDR}^*(t) &\leftarrow \mathbf{C}_{HDR}^*(t-1) + h(\mathbf{I}^w(t)) \\ \mathbf{I}_{HDR}^*(t) &\leftarrow \frac{h(\mathbf{I}^w(t))\mathbf{I}_{HDR}^w(t) + \mathbf{C}_{HDR}^*(t-1)\mathbf{I}_{HDR}^*(t-1)}{\mathbf{C}_{HDR}^*(t)} \end{aligned}$$

where \mathbf{I}^w and \mathbf{I}_{HDR}^w are respectively the current LDR and HDR images warped onto the reference frame. Compared to our previous work [2], this paper introduces a new weighting function inspired from [14] which is based on the slope of the camera response function. The slope indicates how quickly the pixel intensity value varies from the given input. Consequently the most reliable weight is given to the most sensitive part of the sensor:

$$h(\mathbf{I}_{ij}) = \frac{d\mathbf{I}_{ij}}{dg(\mathbf{I}_{ij})} \quad (10)$$

where g is defined in Eq.9.

The main objective (and contribution of this paper) is to optimize the range of the log irradiance during real-time acquisition. This equivalent to maximising the log irradiance entropy (11) as:

$$H(p) = \sum p(\mathbf{E}) \ln(p(\mathbf{E})) \quad (11)$$

The idea is to select the next best shutter that will maximise the entropy of the blended images (see Fig. 3).

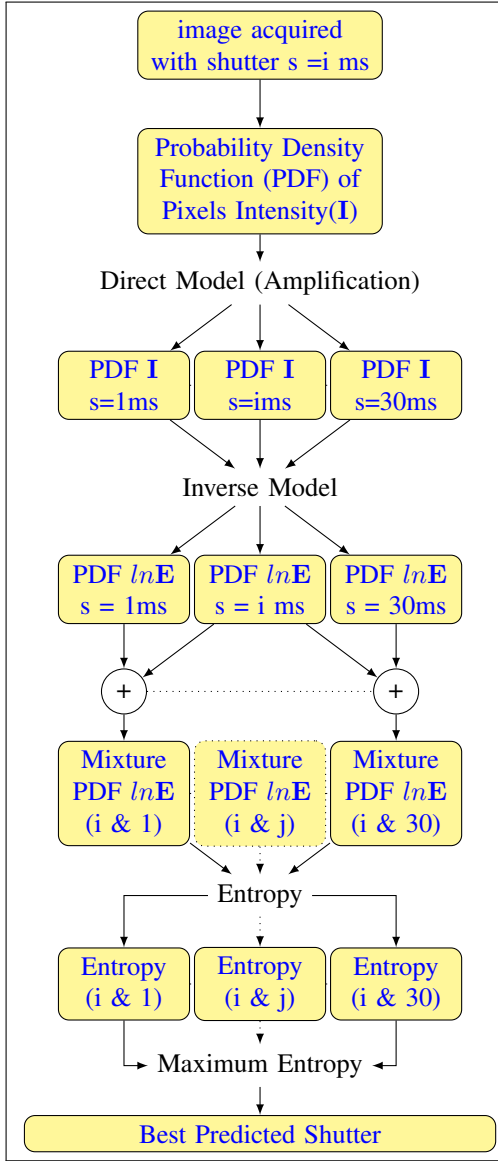


Fig. 3. The method to select the next shutter.

If it is possible to predict the response of the sensor for different shutter values based on an image, it is also possible to predict the best shutter which maximises the entropy of the irradiance. It remains to identify a direct model to estimate images for different shutter values based on the current image. The following section will describe this model.

A. Direct Model

In this section an approach is proposed to estimate the image transformation function between two successive shutter values. This is necessary for active shutter control in order to predict which new shutter value will provide more information. The transformation function corresponds to an amplification of the pixel intensity I_{ij} due to a change in shutter value. This amplification can be estimated from a series of image pairs acquired with different exposure times ranging from 1ms to 30ms with a step of 1ms, and for each

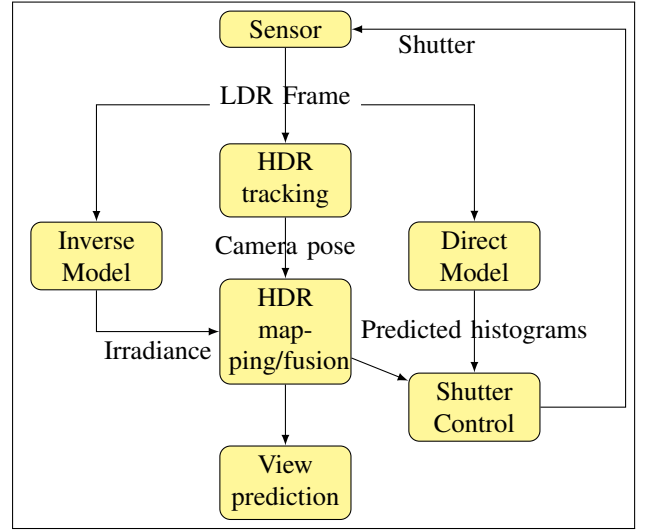


Fig. 4. High level HDR tracking an mapping with active shutter control.

level of grey-scale intensity. The linear relationship between the intensity of an image obtained with a shutter s_1 versus an image obtained with a shutter s_2 is subsequently a matrix A of size 256×30 (intensity level \times shutter values).

$$\mathbf{I}_{ij}^{s_2} = A(\mathbf{I}_{ij}^{s_1}, s_1) \mathbf{I}_{ij}^{s_1} \quad (12)$$

with $s_1 = n$, $s_2 = n + 1$ and $n \in [1, 30]$.

To estimate an image with a $s_2 - s_1 = m > 1$ it is necessary to apply the amplification m times. Once, this direct model has been estimated, it is possible to predict an image of the same scene with a different shutter value. To speed up the process for real-time performance, the predicted histogram is directly computed. To obtain a smooth estimate, a Kernel Density Estimator using a Gaussian kernel is employed.

$$\widehat{f}_h(\mathbf{I}) = \frac{1}{Nb} \sum_{i=1}^N K\left(\frac{\mathbf{I} - \mathbf{I}_i}{b}\right) \quad (13)$$

where N is the number of kernels (256 levels of intensity), b is a smoothing parameter (called bandwidth), K is a gaussian kernel:

$$K(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad (14)$$

B. Histogram Prediction

The direct model is applied with shutter values varying from 1ms to 30ms to the current histogram to obtain predicted histograms. The predicted entropy of the combined image can then be computed from the weighted sum of the current image and the predicted image. The estimation is carried out directly on the histogram. The new histogram for a shutter value $s \in [1..30]$ is expressed using the histogram from the previous shutter $s + 1$ (15) or $s - 1$ (16) based on

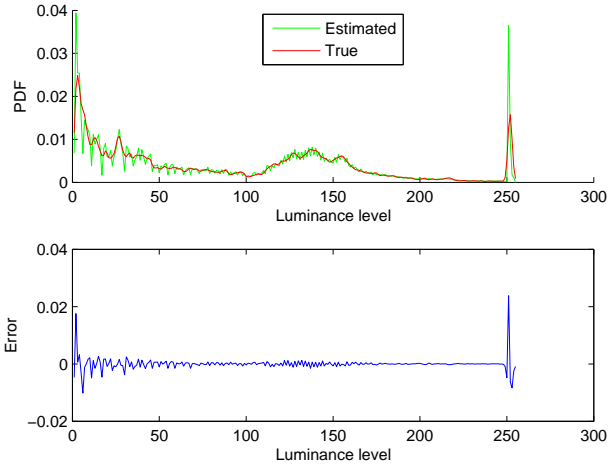


Fig. 5. Comparison of the estimated PDF with a shutter of 18ms and the baseline shutter of 15ms along with the true PDF for a shutter= 18ms.

the direct model $A(\mathbf{I}, s)$ with \mathbf{I} the pixel intensity.

$$\hat{f}_s(\mathbf{I}) = \frac{1}{256} \sum_{\mathbf{I}=0}^{255} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(A(\mathbf{I},s)\hat{f}_{s+1}(\mathbf{I})-\mathbf{I})^2} \quad (15)$$

$$\hat{f}_s(\mathbf{I}) = \frac{1}{256} \sum_{\mathbf{I}=0}^{255} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\hat{f}_{s-1}(\mathbf{I})}{A(\mathbf{I},s)}-\mathbf{I}\right)^2} \quad (16)$$

With \hat{f} expressed in (13) for a shutter value $s \in [2..29]$. Once we have the luminance histogram, using Eq.9 we obtain the log irradiance histogram. It is better to work in the log irradiance domain because is less sparse than in the irradiance domain and this improves the entropy estimation.

C. Results on Histogram Prediction

To evaluate the prediction of information obtained with various shutter values, predicted histograms of pixel intensity (\mathbf{I}) are compared here to the real histograms obtained with the true shutter. In Fig.5, the estimated histogram for a shutter of 18ms, predicted from the histogram for a shutter of 15ms is presented. The estimated and real histograms are very similar. To obtain a quantitative value, the predicted histogram is compared to the reference histogram by using the divergence between the two distributions. Generally the Kullback-Leibler Divergence is used, but as it is not symmetric, a symmetric version is used [15]:

$$J(p_1, p_2) = \sum_{z \in Z} p_1(z) \log \frac{p_1(z)}{\frac{1}{2}p_1(z) + \frac{1}{2}p_2(z)} \quad (17)$$

with z a discrete random variable and p_1 and p_2 two probability distributions of z . The divergence result for a baseline shutter= 15ms and an estimated shutter ranging from $\in [1..30]$ ms is presented Fig.6. It can be seen that the histograms have been correctly predicted for the various trial shutter values.

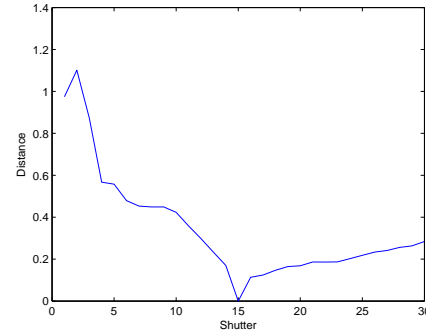


Fig. 6. Divergence between Predicted and ground truth for a reference shutter value=15ms.

V. RESULTS

For practical purposes, the Asus Xtion [16] RGB-D sensor is used for the experiments this paper. This sensor possesses a 640×480 RGB Camera and a 640×480 depth camera. The driver interface allows the exposure time to be controlled, however, it only allows a discrete set of shutter values starting from $s = 1$ ms, to 3000ms. To maintain the frame rate at 30Hz the possible shutter values are limited here to $s \in [1, 30]$ ms.

To better analyse the proposed active shutter control method, results are firstly presented using a static camera to better illustrate the shutter prediction before finally presenting the results in a real SLAM application with a moving camera.

A. Static camera with three images

Firstly results are presented for an optimisation with three images. A base image is acquired with a shutter $s = 15$ ms and the proposed algorithm is used to estimate the next best shutter. Fig.8 shows the entropy estimated for the mixture of images (in the log irradiance domain), the best shutter (providing the most information) is obtained for the value $s = 1$ ms. Fig.7 shows the evolution of the entropy of the mixture (base image plus new image) in the log irradiance domain. It is noted that the minimum is correctly positioned at the shutter value of the reference image ($s = 15$ ms). The same algorithm is next applied to a third image providing the results presented in Fig.9. The improvement is clear on the light part which is over-exposed in the first image and is corrected with the addition of the second and third image. Fig.8 shows the improvement of the entropy estimated on the mixture histogram in the log irradiance domain. Others results are presented in Figures 10, ??.

If a static video camera with an internal auto-shutter is compared to the proposed algorithm, it is clear that the information content is improved because the shutter in auto mode seeks to simply compute the best constant shutter value for a given time instant. The proposed algorithm acts more like a HDR photographic camera which takes several images with different exposures and combines them. As will be seen in the next section, this is not possible if the camera or scene is in motion.

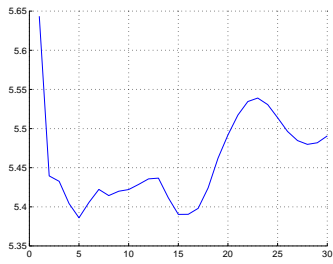


Fig. 7. Variation of the entropy for a mixture (in log irradiance domain) with the base image (with shutter= 15ms) and an image with a shutter varying from 1ms to 30ms.

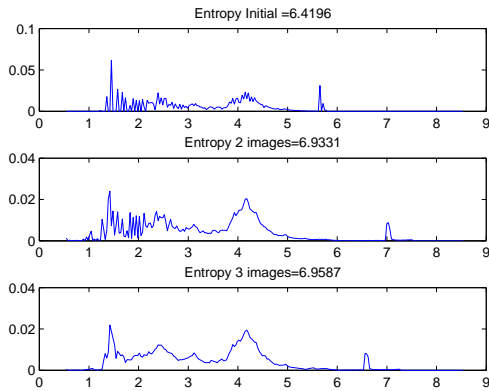


Fig. 8. Improvement of the entropy after the fusion of additional images into the HDR image. First the initial entropy is shown, then the mixture of 2 and then 3 images.

B. Moving camera: HDR SLAM

After validating the method with a static camera, results are now presented on images with a moving camera. Fig.11 shows the images chosen by the algorithm, the key-frame(reference image) is acquired with a shutter =15ms. For the next images we apply our method to select the best shutter.

After the acquisition of 6 images we obtain the following HDR image in Fig.12. In this final image, we can see the improvement, for example the whiteboard and the windows are not saturated compare to the reference image. The next example Fig.14 shows the improvement in the dark zone of the reference image Fig.13 (books are now visibles).

C. Comparison LDR/HRD Tracking

To validate the method, tracking results on synthetic images (30 images) are presented. For a LDR tracking with a fixed shutter (15ms) and HDR tracking with varying shutter the errors ($e = \sum (\tilde{\mathbf{T}}\mathbf{T}_{true}^{-1} - Id)^2$, with Id the identity matrix) of tracking are very similar, see Fig.15. We obtain a little improvement with the HDR approach. On Fig.16 is presented the convergence, for 2 images, in 3D for a fixed shutter=4ms for the LDR tracking and a varying shutter for HDR tracking. The reference image is very dark with this shutter value. Then, the LDR tracking diverges, when the HDR tracking converges to the true position. The



Fig. 9. Static camera : visual improvement of the HDR mixture image obtained from three LDR source images. Since the printer (or screen) dynamic range is not capable of displaying the extended dynamic range the HDR image is remapped to LDR using global tone mapping.

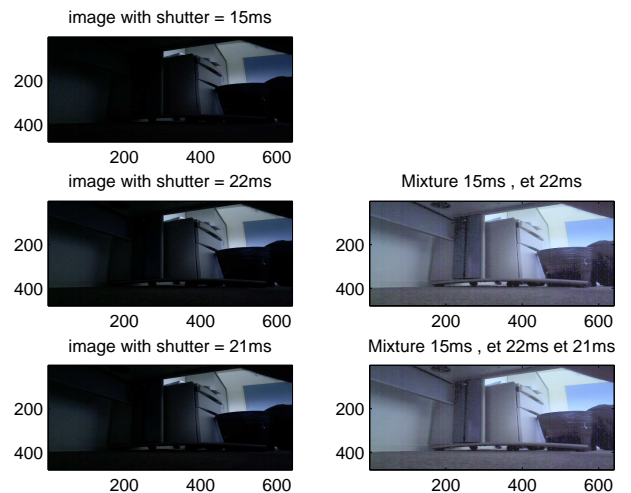


Fig. 10. A second example of static camera HDR mapping.

same experimentation is presented now with a reference image acquired with a shutter=25ms on Fig.17 Both methods converge to the true position. But, LDR tracking needs more iterations to converge, this is true also for the whole tracking, LDR tracking needs more iterations to converge.

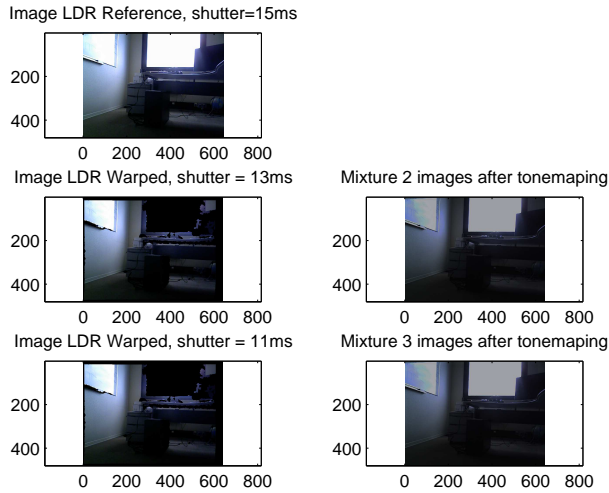


Fig. 11. Moving camera : visual improvement of the HDR mixture image obtained from three LDR source images. Since the printer (or screen) dynamic range is not capable of displaying the extended dynamic range the HDR image is remapped to LDR using global tone mapping.

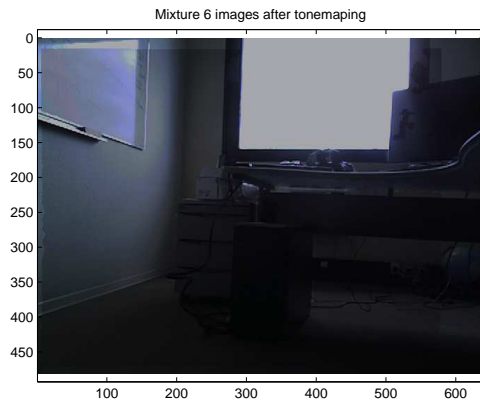


Fig. 12. HDR Mixture of 6 images mapped to the LDR domain.

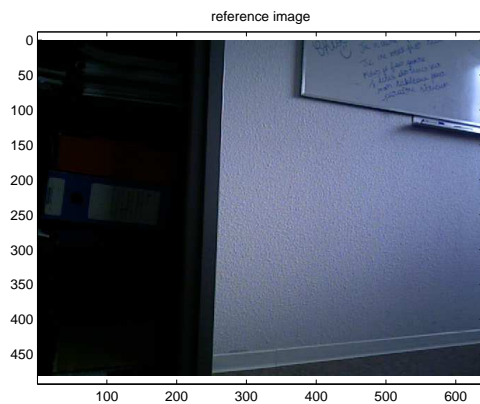


Fig. 13. HDR Mixture of 6 images mapped to the LDR domain.

VI. CONCLUSION

This paper has presented an information theory method to *actively* control the shutter of a RGB camera in real-time.

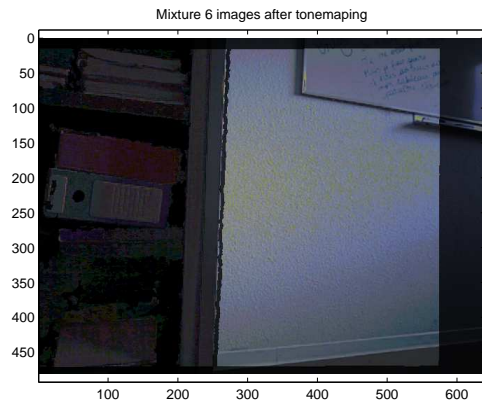


Fig. 14. HDR Mixture of 6 images mapped to the LDR domain.

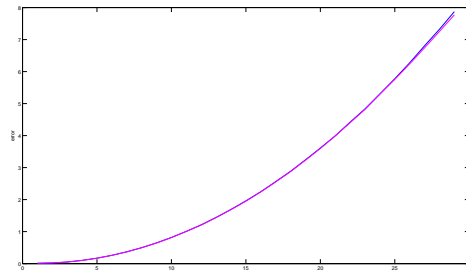


Fig. 15. Tracking error on synthetic images for LDR in blue and HDR in magenta.

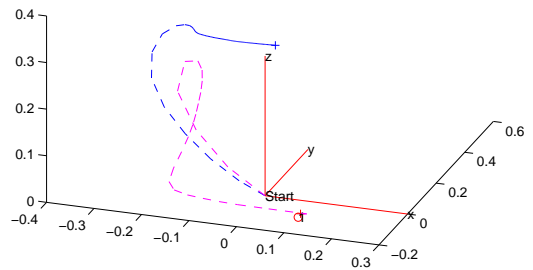


Fig. 16. Tracking convergence on synthetic images (reference image acquired with shutter=4ms) for LDR in blue and HDR in magenta. The red circle represents the objective.

The proposed approach effectively acquires optimal high dynamic range 3D texture maps from a moving sensor using a dense visual localisation and mapping framework. The proposed method is based on the entropy of the resulting map and aims to maximise the information content in the dynamic range of the texture map. The 3D photometric model is represented via a HDR keyframe graph which facilitates real-time operation and loop closure. This novel approach has been shown to improve localisation and mapping in

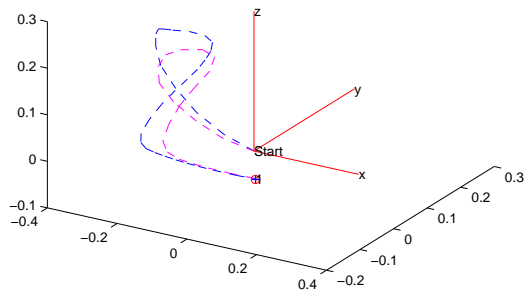


Fig. 17. Tracking convergence on synthetic images (reference image acquired with shutter=25ms) for LDR in blue and HDR in magenta. HDR tracking is faster.

complex lighting situations. The method has been validated by comparing the estimated histogram to real histogram and the improvement in entropy via the new fusion approach.

In the future, we aim to adapt the method to include other constraints like the coupling between motion blur and the exposition time.

REFERENCES

- [1] M. Grossberg and S. Nayar, "High dynamic range from multiple images: Which exposures to combine?" in *ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, Oct 2003.
- [2] M. Meilland, C. Barat, and A. Comport, "3d high dynamic range dense visual slam and its application to real-time object re-lighting," in *Proceedings of International Symposium on Mixed and Augmented Reality, ISMAR*, Adelaide, Australia, 2013.
- [3] S. Li, A. Handa, Y. Zhang, and A. Calway, "Hdrfusion: HDR SLAM using a low-cost auto-exposure RGB-D sensor," *CoRR*, vol. abs/1604.00895, 2016. [Online]. Available: <http://arxiv.org/abs/1604.00895>
- [4] 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*, ser. SIGGRAPH '97. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 369–378. [Online]. Available: <http://dx.doi.org/10.1145/258734.258884>
- [5] Q. K. Vuong, S. hwan Yun, and S. Kim, "A new auto exposure and auto white-balance algorithm to detect high dynamic range conditions using cmos technology," in *Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, USA, 2008.
- [6] W.-C. Kao, C.-C. Hsu, C.-C. Kao, and S.-H. Chen, "Adaptive exposure control and real-time image fusion for surveillance systems." in *ISCAS*. IEEE, 2006.
- [7] H.-W. Kim, S. Kwon, J. Je-Kyo, and J. Ha, "Auto-exposure control method for a stereo camera robust to brightness variation," *International Journal of Control and Automation*, vol. 7, No 1, pp. 321–330, 2014.
- [8] M. A. Ali and S. Mann, "Comparametric image compositing: Computationally efficient high dynamic range imaging," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 913–916.
- [9] A. I. Comport, E. Malis, and P. Rives, "Accurate quadri-focal tracking for robust 3d visual odometry," in *IEEE International Conference on Robotics and Automation*, Rome, Italy, 2007.
- [10] M. Meilland and A. I. Comport, "On unifying image and model-based real-time dense mapping at large scales," in *Int. Conf. on Intelligent Robots and Systems*. IEEE, 2013.
- [11] F. Ireta and A. Comport, "Point-to-hyperplane rgb-d pose estimation: Fusing photometric and geometric measurements," in *IEEE International Conference on Intelligent Robots and Systems*, Deajeon, South Korea, 2016.
- [12] N. Asada, A. Amano, and M. Baba, "Photometric calibration of zoom lens systems," in *Proceedings International Conference Pattern Recognition*, 1996.
- [13] M. Grossberg and S. Nayar, "Determining the Camera Response from Images: What is Knowable?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1455–1467, Nov 2003.
- [14] S. Mann, "Comparametric equations with practical applications in quantigraphic image processing," *Image Processing, IEEE Transactions on*, vol. 9, no. 8, pp. 1389–1406, 2000.
- [15] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Transactions on Information theory*, vol. 37, pp. 145–151, 1991.
- [16] "Xtion," <http://www.asus.com/Multimedia/Xtion.PRO.LIVE/overview/>.