Optimal High-Dynamic-Range Image Acquisition for Humanoid Robots

D. Gonzalez-Aguirre, T. Asfour and R. Dillmann

Karlsruhe Institute of Technology, Adenauerring 2, Karlsruhe-Germany.

{*david.gonzalez, asfour, dillmann*}@kit.edu

Abstract— Humanoid robots should be able to visually recognize objects and estimate their 6D pose in real environmental conditions with their limited sensor capabilities. In order to achieve these visual skills, it is necessary to establish an optimal visual transducer connecting the scene layout with the internal representations of objects and places. This visual transducer should capture the noiseless visual manifold of the scene with high-dynamic-range in an efficient manner. Our endeavor is to develop such a visual transducer using the widespread LDR cameras in humanoid robots. In our previous work, the noiseless acquisition of continuous images [1] and the improved radiometric calibration [2] already enabled the humanoid robots to attain the desired visual manifold in terms of quality. However, since the radiance range of the scene can be very wide, the required amount of exposures to capture the visual manifold (robustly without radiance inconsistencies) turns impractically large in terms of scope, granularity and acquisition time. In this article, a method for estimating the minimal amount of exposures and their particular integration times is presented. This method integrates our previous work in order to synthesize HDR images with the minimal amount of exposures while ensuring the high quality of the resulting image. Conclusively, the minimal exposure set provides performance improvements without quality trade-off. Experimental evaluation is presented with the humanoid robots ARMAR-IIIa,b [3].

I. Introduction

Until now, HDR cameras are not widespread. Especially in applications with several tight constraints, for instance, high frame rate, light weight, reduced space, low power consumption and compliance with saccadic movements (see [4]). These are few of the restrictions in the eyes of most humanoid robots, for instance, ARMAR-IIIa,b [3] or the iCub robot [5] (see their detailed composition in [6], [7]). Due to these restrictions, an appropriate mechanism is introduced to capture HDR images by optimally employing the low-dynamic-range (LDR) cameras of humanoid robots, see figure 1.

Conceptually, the process of attaining HDR images is rather simple. During the HDR image synthesis, the short-exposed images sample the high radiance regions of the scene. Inversely, the long-exposed images sample the low radiance regions. The integration of this information in a sound manner is done by the inverse *radiometric response function* [8] of the whole optical system.

The proposed mechanism for attaining HDR images consists of the following elements. First, the nonlinear transformation mapping from the scene radiance to the discrete intensity values of the image is attained (the radiometric calibration). Second, the robust exposure control of the camera and the analysis of the available exposure distribution of the cameras are performed. The coordinated integration of these elements captures the intra-scene radiance range by fusing the *minimal collection of differently exposed images*.

Our strategy to efficiently synthesize high quality HDR images is to analyze the radiometric calibration and granularity of the integration times of the camera(s) in order to provide the means for optimal exposure selection. This reduces the total amount of images to be captured while ensuring the quality of the synthesized HDR image.

II. Related Work

During the last decades, various important contributions were made on high-dynamic-range scene capturing using digital cameras. In particular, the work of [9] was an important contribution to this field. It set the foundation for response function estimation solely based on images. The key idea of the approach was the modeling of the camera response function without using controlled light sources or other complex luminance devices. In [9] the authors presented exposure bracketing as controlling element during calibration. Despite the influential ideas of that work, the dependency on particular image content forbids its systematic application in general scenarios. Later on, the camera radiometric calibration was obtained without assumptions about image content. In [10], the method is based on digital cameras with precise integration timing. The key idea of that approach is to analyze the progression of the intensity values as a monotonic function depending on the integration time. This fundamental idea is found in almost all subsequent calibration methods (see [11], [8], [12]). Recently, novel methods have appeared (for instance [13]) which only use a single image to attain the radiometric calibration. Nevertheless, they assume certain image content (as in [9]) such as colored edges or fixed patterns. Thus, their application in humanoid robots remains limited. The approach of the method from P. Debevec et al. in [10] established a fundamental research reference in the field. Because this method has no image content dependency, it is possible to use any camera supporting the DCAM/IIDC specification [14] with precise integration timing. Despite being suitable for humanoid robots and due to the calibration artifacts, the Debevec method cannot be directly applied for structural feature extraction (see details in [2]). In order to simultaneously explain the negative issues and clearly separate the work in [10] from the improvements realized in our previous work (in section III, partially in [2]), a sequential presentation of the Debevec method is given followed by the proposed novel improvements presented for the first time in this article[1]. The proposed method presented in section IV

---

[1]The section III is not only a recapitulation from our previous work in [2] (necessary to clearly present the novel method in section IV) but it also provides the detailed and revisited algorithm for optimal radiometric camera calibration.

can be easily disambiguated from various existing works on optimal HDR reconstruction (see approaches [15], [16]). Because the aim of our method is, given a particular camera and lens, how to determine the minimal and consistent set of exposures (according to its radiometric calibration and exposure granularity) to robustly and efficiently cover the full reachable irradiance of the camera sensor. In contrast, the optimal reconstruction methods of related approaches focus on how to recover the absolute sensor irradiance using *suboptimal and arbitrarily* exposed images. This remark is fundamental to understand the structure, novelty and importance of the contributed method.

III. Improved Radiometric Calibration

The intensity values provided by a digital camera are ideally subject to a principle called *reciprocity*. The sensor exposure X is the collected energy per surface unit in Jm^{-2}. It is the product of the sensor *irradiance* E and the integration time Δt as $X = E\Delta t$. The reciprocity states that the total collected energy $X' = X$ remains constant if the sensor irradiance is scaled $E' = \alpha E$ as long as the integration time is inversely scaled $\Delta t' = \frac{1}{\alpha}\Delta t$. The reciprocity principle does not entirely hold in physical systems (see the Hurter-Driffield curve in [17]) particularly at the upper and lower bounds of the sensor irradiance. Once the sensor irradiance has been integrated, the electronic components of the camera convert the analog signal to its discrete and quantized version. This composed process is followed by a nonlinear mapping (usually gamma compression) for the high contrast image content to be contained within the dynamic limits of the m-bit representation. This nonlinear mapping is done in order to rescale the signals to better fit human perceptual metrics. The whole process is called the *sensor mapping function* and it is denoted as $\mathcal{F}$. The sensor mapping function can be expressed as $I_{\mathbf{x}}^i = \mathcal{F}(E_{\mathbf{x}}\Delta t_i)$, where the irradiance at the pixel location $\mathbf{x} \in \Omega := \{(0,0),...,(w-1,h-1)\} \subset \mathbb{N}^2$ is denoted as $E_{\mathbf{x}}$. The integration time Δt_i is controllable and the intensity values $I_{\mathbf{x}}^i \in \Theta := \{0,...,2^m-1\} \subset \mathbb{N}$ are provided by the camera interface. Since the mapping function $\mathcal{F}$ is monotonic, it is possible to find its inverse function such as $E_{\mathbf{x}}\Delta t_i = \mathcal{F}^{-1}(I_{\mathbf{x}}^i)$. In this manner, the ideal estimation of the sensor irradiance from the pixel $\mathbf{x}$ is $E_{\mathbf{x}} = \mathcal{F}^{-1}(I_{\mathbf{x}}^i)/\Delta t_i$. In [10], additional considerations were taken for modeling $g := \ln \mathcal{F}^{-1}$ (g is used to simplify notation). i) Since the codomain of $\mathcal{F}$ is the finite intensity set Θ, the estimation of the function g considers the function as a vector $g \in R^{2^m}$ in a least squares formulation. ii) The sensor noise is considered during the g estimation process. This is coherently formulated by kernel weighting in a least squares error sense. The least squares formulation proposed by Debevec et al. using ρ images with different integration times and selecting a subset of pixels $S \subset \Omega$ is expressed as

$$O = \sum_{i}^{\rho}\sum_{\mathbf{x} \in S}\left\{w(I_{\mathbf{x}}^i)\left[g(I_{\mathbf{x}}^i) - \ln E_{\mathbf{x}} - \ln \Delta t_i\right]\right\}^2 \quad (1)$$
$$+ \lambda \sum_{u=1}^{2^m-2}\left(w(u)\left[g(u-1) - 2g(u) + g(u+1)\right]\right)^2,$$
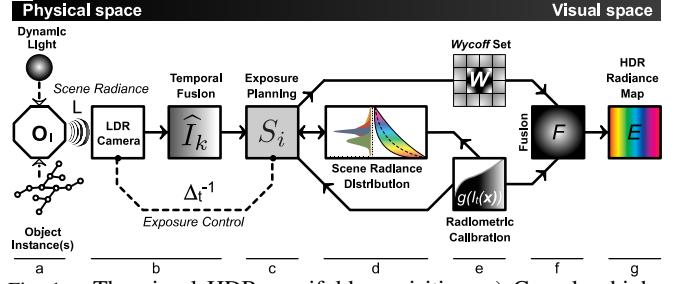

Fig. 1. The visual HDR manifold acquisition. a) Complex high-contrast heterogeneously lighted scene. b) The low-dynamic-range cameras capture differently exposed images (so-called Wycoff set). Those image sets are used for temporal fusion to obtain *"noiseless"* images $\widehat{I}_k^i$ by our method in [1]. c) The exposure plan (in section IV) determines the amount and integration time Δt_i of each of the images in these sets to optimally sample the scene using a particular camera. d) The radiance distribution of the scene soundly captured by the minimal amount of exposures $\mathcal{E}_{\min}$. e) The radiometric calibration of the composed optical system. f) The Wycoff set and the radiometric calibration are used to synthesize [F] the HDR image. g) The resulting HDR image E.

where the λ-smoothing term ensure the evenness of the estimated function g. The contributions are weighted according to their intensity with less weighting at the extrema by a symmetric kernel $w : \Omega \mapsto \mathbb{R}$ with maximal central value $w(\frac{1}{2}(2^m - 1)) = 1$. The computation of the overdetermined system requires $\rho(|S|-1) \geq 2^m$ pixel exposures. The Debevec radiometric calibration shows critical issues: i) The resulting g curves present noisy artifacts which cannot be removed by varying the calibration parameters (λ,S,ρ). ii) There is neither implicit nor explicit selection criteria for the smoothing factor λ nor for selection of pixels in the sampling set S. iii) The resulting noisy curves g produce detrimental effects near the upper and lower intensity extrema. These calibration artifacts appear because the reciprocity between exposure and intensities is not held. These are the jointed numerical effects (least squares fitting) and the intrinsic physical sensor behavior. The results are noisy calibrations propagating into salient artifacts in the HDR image, see figure 2.

A. Continuous Reciprocity-Consistent Calibration Model

The detrimental radiometric artifacts (see zooms in figure 2) resulting from [10] are unveiled by mapping the λ-*optimal smooth*[2] function g_λ from its *homomorphic*[3] domain [19] to the lineal domain through the exponential transformation $g_\lambda^*(u) := \mathbf{exp}_u[g_\lambda(u)]$, where the auxiliar variable $0 \leq u \leq (2^m - 1)$ covers the intensity set Θ. The function $g_\lambda^*(u)$ clearly illustrates the noise at the intensity extrema. Furthermore, the discrete domain of g_λ is a limitation when using real intensity values resulting from our noise removal and range enhancement fusion method [1]. Because of these limitations, the calibration curve g_λ^* is improved by estimating the continuous function g_λ^L which optimally (considering reliability of the regions according to their confidence by weighted regression) fits the tendency of g_λ^*.

[2]The optimal smoothing is attained in lines 7-15 of algorithm 1.
[3]Homomorphic filtering is a nonlinear mapping to a target domain where linear filtering is applied followed by back mapping, see example in [18].

Algorithm 1 *Extended-Radiometric-Calibration* $(S,\rho,\epsilon,\delta_\lambda)$

Require: $\rho(|S|-1) > (2^m-1)$ **and** $0 < \delta_\lambda \leq \frac{1}{10}$
1: $P \leftarrow [0]_{2^m x 2^m}, \ X \leftarrow [0]_{2^m x 2}$
2: $\lambda_s \leftarrow 0, \ \lambda_{\min} \leftarrow 0, \ E_{\min} \leftarrow \infty$
3: **for** $u = 0$ **to** (2^m-1) **do**
4:     $X[u] \leftarrow [u,1], \ P[u,u] = \exp\left[\frac{-3(u-2^{m-1})^2}{2^{(m+1)}-1}\right]$
5: **end for**
6: $K \leftarrow [X^T P X]^{-1}[X^T P]$
7: **while** $\lambda_s \leq 1$ **do**
8:     $g^*_{\lambda_s} \leftarrow \exp_u[Debevec(\lambda_s,S,\rho)], \ M \leftarrow K g^*_{\lambda_s}, \ \lambda \leftarrow \lambda + \delta_\lambda$
9:     **for** $u = 0$ **to** (2^m-1) **do**
10:        $\delta \leftarrow (uM[1,1] + M[1,2]) - g^*_{\lambda_s}(u), \ E \leftarrow E + \delta^2$
11:    **end for**
12:    **if** $(E < E_{\min})$ **then**
13:       $E_{\min} \leftarrow E, \ \lambda_{\min} \leftarrow \lambda_s$
14:    **end if**
15: **end while**
16: $\lambda_a \leftarrow \lambda_{\min} - \frac{\delta_\lambda}{2}, \ \lambda_b \leftarrow \lambda_{\min} + \frac{\delta_\lambda}{2}$
17: $g^*_{\lambda_a} \leftarrow \exp_u[Debevec(\lambda_a,S,\rho)], \ M_a \leftarrow K g^*_{\lambda_a}, \ E_a \leftarrow 0$
18: $g^*_{\lambda_b} \leftarrow \exp_u[Debevec(\lambda_b,S,\rho)], \ M_b \leftarrow K g^*_{\lambda_b}, \ E_b \leftarrow 0$
19: **for** $u = 0$ **to** (2^m-1) **do**
20:     $\delta_a \leftarrow (uM_a[1,1] + M_a[1,2]) - g^*_{\lambda_a}(u), \ E_a \leftarrow E_a + \delta_a^2$
21:     $\delta_b \leftarrow (uM_b[1,1] + M_b[1,2]) - g^*_{\lambda_b}(u), \ E_b \leftarrow E_b + \delta_b^2$
22: **end for**
23: **repeat**
24:     **if** $(E_a < E_b)$ **then**
25:        $\lambda_b \leftarrow \frac{1}{2}(\lambda_a + \lambda_b), \ g^*_{\lambda_b} \leftarrow \exp_u[Debevec(\lambda_b,S,\rho)]$
26:        $M_b \leftarrow K g^*_{\lambda_b}$
27:        **for** $E_b \leftarrow 0, \ u = 0$ **to** (2^m-1) **do**
28:           $\delta_b \leftarrow (uM_b[1,1] + M_b[1,2]) - g^*_{\lambda_b}(u), \ E_b \leftarrow E_b + \delta_b^2$
29:        **end for**
30:     **else**
31:        $\lambda_a \leftarrow \frac{1}{2}(\lambda_a + \lambda_b), \ g^*_{\lambda_a} \leftarrow \exp_u[Debevec(\lambda_a,S,\rho)]$
32:        $M_a \leftarrow K g^*_{\lambda_a}$
33:        **for** $E_a \leftarrow 0, \ u = 0$ **to** (2^m-1) **do**
34:           $\delta_a \leftarrow (uM_a[1,1] + M_a[1,2]) - g^*_{\lambda_a}(u), \ E_a \leftarrow E_a + \delta_a^2$
35:        **end for**
36:     **end if**
37: **until** $((\lambda_b - \lambda_a) > \epsilon)$
38: **return** $g^L_\lambda \leftarrow K \exp_u[Debevec(\frac{E_b\lambda_a + E_a\lambda_b}{\lambda_a + \lambda_b},S,\rho)]$

Fig. 2. Model calibration g^L_λ accomplished by kernel weighted regression in the exponential domain. Gaussian kernel weighting $\mathcal{N}$ ponderates the λ-optimal g^*_λ calibration to estimate the model g^L_λ. The model calibration is $g^L_{\lambda,R}(I_r) = \ln(0.007961 * I_r) - 0.001419$ @ $\lambda_R = 0.94726$. Plots of the continuous reciprocity-consistent calibration model g^M_λ and the comparison to the results from [10]. The deviation plot shows the removed detrimental effects.

This is done by kernel weighted regression obtaining a continuous model $g^L_\lambda : \mathbb{R} \mapsto \mathbb{R}$; $g^L_\lambda(u) \cong g^*_\lambda(u)$, see algorithm 1 and figure 2. The applied Gaussian kernel $\mathcal{N}(\mu = \frac{1}{2}(2^m-1), \sigma = \frac{1}{3}(2^m-1))$ (dotted line in figure 2) properly integrates the smooth region of the curve g^*_λ while gradual disregarding the broken reciprocity regions. Because of this gradually tendency integration, the calibration deviations and detrimental artifacts were soundly removed. This extended radiometric calibration is superior to [8], [9], [10], [12] because the model g^L_λ is reciprocity consistent even at intensity extrema. Moreover, the calibration model g^L_λ estimates the sensor response function at continuous intensities of k iso-exposed images by our method in [1] as:

$$g^M_\lambda(\widehat{I}^i_k) := \ln(\widehat{I}^i_k \cdot g^L_\lambda[1,1] + g^L_\lambda[1,2]), \quad (2)$$

where $\widehat{I}^i_k$ is the i-exposed temporal fusion of k images and g^M_λ is the standard homomorphic domain of the radiometric calibration function. The improved radiometric calibration is unique per color channel and (due to the logarithm in eq. 2) is restrictively defined within the interval:
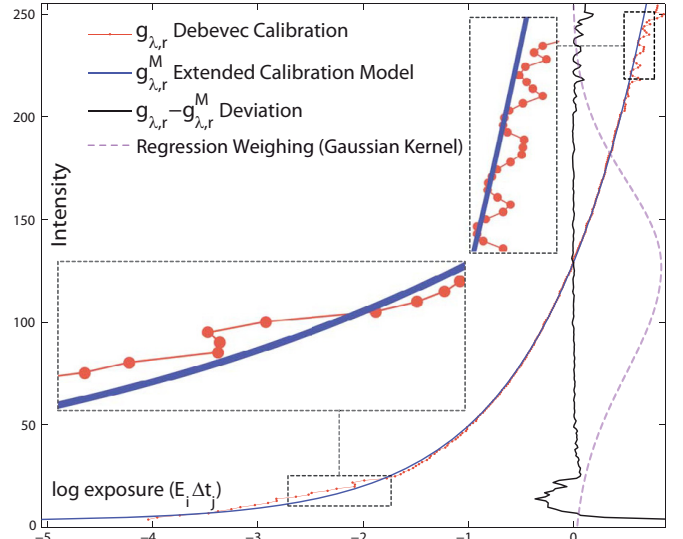
$$g^M_\lambda : \left(\mathbf{max}(-g^L_\lambda[1,2]/g^L_\lambda[1,1], 1), (2^m-1)\right] \in \mathbb{R}^+ \mapsto \mathbb{R}. \quad (3)$$

In summary, the radiometric calibration model g^M_λ is an improvement and extension of the method in [10]. The model calibration does not only address the parameter selection in an optimal manner. It also provides a continuous and computationally efficient radiometric function which enables the integration of the whole reachable radiance without producing detrimental artifacts. Since the reciprocity is not physically held at sensor level, the kernel weighting $\mathcal{N}$ used during the regression must be also applied during the HDR synthesis as follows: i) For an initial integration time Δt_1 a set of k_1 images is captured in order to estimate the temporal fusion image $\widehat{I}^1_k$. ii) Next, the following integration time Δt_2 is set and the next set of images with cardinality k_2 is taken in order to obtain the fusion image $\widehat{I}^2_k$. iii) The collection of n temporal fused images $\{\widehat{I}^1_k,...,\widehat{I}^n_k\}$ and the radiometric calibration model g^M_λ are synthesized into a HDR image as

$$\ln(E_{\mathbf{x}}) = \frac{\sum_{i=1}^n \mathcal{N}\left(\widehat{I}^i_k(\mathbf{x})\right) \ln\left(\widehat{I}^i_k(\mathbf{x}) \cdot \frac{g^L_\lambda[1,1]}{\Delta t_i} + \frac{g^L_\lambda[1,2]}{\Delta t_i}\right)}{\sum_{i=1}^n \mathcal{N}\left(\widehat{I}^i_k(\mathbf{x})\right)}. \quad (4)$$

The acquired spatial discrete and non-quantized radiance map, namely the HDR image $E : \mathbb{N}^2 \mapsto \mathbb{R}$ is a consistent (up-to-scale) manifold of the scene radiance L, see figure 5.

IV. OPTIMAL EXPOSURE CONTROL

In addition to the continuous reciprocity-consistent calibration model g^M_λ, the consistency of the resulting HDR images depends on: i) The proper selection of integration times. ii) The amount of images to be fused per exposure. iii) The precise exposure timing.

Therefore, a detailed analysis of the exposure control is presented. Afterwards, a novel method is introduced for optimal selection of the amount and particular integration times for optimal radiance sampling. Finally, the robust and accurate image acquisition of experimental scenarios is presented.

A. Embodiment Aspects

When capturing the Wycoff set of a scene, the integration time Δt_i is controlled according to the camera specification [20], see figure 1. However, there might be issues when using different camera models or even different firmware versions of the same camera. The first problem occurs when the integration time is not consistent for the complete image. This happens systematically when using rolling shutter sensors or (in case of global shutter) it may occur while dynamically changing the exposure settings. This issue is anticipated by the inclusion trigger. This trigger is fired by means of exposure stability analysis. This analysis takes into account the latency of the capture system. The exposure stability index $S_i(t) := \frac{1}{wh} \sum_{\mathbf{x} \in \Omega} I_t^i(\mathbf{x})$ integrates the image intensities while controlling the exposure time Δt_i. Its differential analysis $\delta(S_i(t))/\delta t$ determines reliable capture intervals. Without this analysis, the storage trigger could be eventually fired within unstable intervals. Furthermore, cameras supporting the specification in [14] provide an indexed set of exposure times $\mathcal{E}$. However, depending on the particular camera, frame rate and firmware the indexing integration times are not always regularly distributed. Without this consideration, the HDR synthesis (eq. 4) is corrupted.

B. Scene Aspects

Each scene has a singular radiance distribution depending on the lighting, materials and viewpoint. In order to optimally acquire the visual manifold of each scene, the humanoid robot can use our method [2] to fuse every available integration time of the camera to create a highly accurate HDR image. However, in practice this is neither desirable nor feasible for online applications. The limitation is the long period of time (in the order of minutes) necessary to complete the HDR image acquisition. In this process, the amount and distribution of the exposures plays a critical role. Slightly differing integration times capture *"almost"* the same radiance segment. This redundancy can be removed without negative effects[4] during the HDR synthesis. Based on the radiometric calibration model and the exposure granularity of a particular camera, it is possible to determine the minimal set of exposures $\mathcal{E}_{\min} \subsetneq \mathcal{E}$ necessary to sample the complete reachable radiance range of the sensor. When using a Bayer pattern camera, the minimal set of exposures depends on the radiometric calibration models ($g_{\lambda,R}^M$, $g_{\lambda,G}^M$ and $g_{\lambda,B}^M$) of the color channels, the regression kernel $\mathcal{N}$ used during calibration and the intrinsic minimal sampling density κ_0.

[4] The noise reduction by integrating redundant exposures [10] is not necessary in our pipeline (see figure 1) due to the *"noiseless"* images attained by our optimal temporal fusion [1] during the exposure bracketing.

C. Sampling Density

The intensity extrema of the images do not hold the reciprocity principle. The comparative results of the proposed calibration model (in figure 2) experimentally support these observations. Despite the attained continuous and reciprocity consistent calibration model for HDR image synthesis (eq. 2), the physical nature of the sensor still produces unreliable measurements at the intensity extrema. The kernel weighting $\mathcal{N}$ compensates these effects by gradually disregarding these intensity regions. Consequently, by considering the calibration interval (in eq. 3), the weighting associated with the pixel location $\mathbf{x}$ resulting from the exposure with integration time Δt_i is expressed as

$$\kappa(\mathbf{x}, \Delta t_i) = \begin{cases} \mathcal{N}\left(\widehat{I^i}(\mathbf{x})\right), & \text{if } \widehat{I^i}(\mathbf{x}) > \mathbf{max}\left(\frac{-g_\lambda^L[1,2]}{g_\lambda^L[1,1]}, 1\right) \\ 0, & \text{else.} \end{cases} \quad (5)$$

This piecewise weighting is the density indicator of the representativeness of the value $\widehat{I^i}(\mathbf{x})$ at the exposure Δt_i. Ideally, all pixels in the image have to be (at least once) sampled with the maximal density $\forall \mathbf{x} \in \Omega, \exists \Delta \hat{t}_i \in \mathcal{E} : \kappa(\mathbf{x}, \Delta \hat{t}_i) = 1$. However, since the radiance range can be rather wide and the pixel intensities are continuously and arbitrary spread along the sensor irradiance, the required amount of exposures is impractical large in terms of scope and granularity. The required exposure set may have a cardinality larger than one available from the particular camera $\mathcal{E}$. Nevertheless, the accumulated sampling contributions attained by the exposure times ($\Delta t_j \approx \Delta \hat{t}_i$) are close enough to the ideal exposure time $\Delta \hat{t}_i$. These exposures provide the necessary information to approximate the ideal radiance sampling. This accumulated collection of sampling contributions (in terms of kernel weighting) is called *radiance sampling density* δ_Φ and is expressed as

$$\delta_\Phi(\mathbf{x}) := \sum_{i}^{n} \kappa(\mathbf{x}, \Delta t_i). \quad (6)$$

Based on this concept, the minimal set of exposures $\mathcal{E}_{\min}$ ensuring the sampling density $\delta_\Phi(\mathbf{x}) > \kappa_0, \forall \mathbf{x} \in \Omega$ is the key to reduce the total visual manifold capturing time without decreasing the HDR image quality. Therefore, the minimal exposure set $\mathcal{E}_{\min}$ is also subject to $\forall \mathbf{x} \in \Omega, \exists \Delta \check{t}_i \in \mathcal{E}_{\min} \Rightarrow \kappa(\mathbf{x}, \Delta \check{t}_i) \geq \kappa_0$, this guarantees the minimal exposure overlapping removing the exposure redundancy. Moreover, if not all pixels captured from a scene can be sampled with the minimal sampling density κ_0, it can be affirmed that the lighting conditions are beyond the physical sensor capabilities. This is an important fact for reasoning and planning for robots.

D. Combined Radiance-Exposure Model

The radiometric image formation process $g(I^i)$ and radiometric calibration model $g_\lambda^M(I^i)$ are consistently combined

$$\overbrace{\ln E + \ln \Delta t_i}^{g(I^i)\,:\,\text{Image formation}} \cong \overbrace{\ln\left(I^i \cdot g_\lambda^L[1,1]\right) + g_\lambda^L[1,2]}^{g_\lambda^M(I^i)\,:\,\text{Radiometric calibration model}},$$

$$E \cong I^i \cdot \frac{g_\lambda^L[1,1]}{\Delta t_i} + \frac{g_\lambda^L[1,2]}{\Delta t_i}. \quad (7)$$

This model associates the intensity values I^i according to the integration time Δt_i by means of the radiometric calibration model g_λ^M. Eq. 7 may lead the reader to assume its direct applications for image synthesis. However, its straightforward application produces severe artifacts due to the broken reciprocity effects of the physical sensor. These effects have to be properly managed by kernel weighted fusion of various exposures (in the homomorphic domain as presented in eq. 4). Fortunately since this model depicts the ideal relationship between radiance and exposure, it can be exploited to delineate the boundary curves of the radiance functions of the specific camera integration times. This representation merges the radiometric calibration with the integration times in the exposure set $\mathcal{E}$ allowing to determine the minimal subset $\mathcal{E}_{\min}$ as follows.

E. Full Reachable Radiance by Minimal Exposure Set

The extraction of the minimal exposure set $\mathcal{E}_{\min}$ requires a unified representation of i) the combined radiance-exposure model (eq. 7), ii) the valid interval of the calibration model (eq. 3) and iii) the sampling density (eq. 6). The estimation of the set $\mathcal{E}_{\min}$ should also consider that Bayer pattern cameras have three different color radiometric calibrations.

1) Radiance-Exposure Bounding Curves: The radiometric calibration is defined only within the interval expressed in eq. 3. Since the scene radiance $L \geq 0$ cannot be negative (see figure 1), the minimal valid intensity value (for all three color channels) is called the *"lower reachable"* intensity as

$$\eta := \mathbf{max}\left(\left\lceil \frac{-g_{\lambda,R}^L[1,2]}{g_{\lambda,R}^L[1,1]} \right\rceil , \left\lceil \frac{-g_{\lambda,G}^L[1,2]}{g_{\lambda,G}^L[1,1]} \right\rceil , \left\lceil \frac{-g_{\lambda,B}^L[1,2]}{g_{\lambda,B}^L[1,1]} \right\rceil , 1 \right),$$
$$(8)$$

where $\eta \geq 1$ for eq. 2 to be defined. This implies an irradiance *lower bounding curve* E^l function of the integration time

$$\overbrace{E^l(\Delta t_i)}^{\text{Lower bounding}} = \overbrace{(\eta \cdot g_\lambda^L[1,1] + g_\lambda^L[1,2])}^{\text{Fixed lower exposure}} \overbrace{\Delta t_i^{-1}}^{\text{Shutter speed}} . \quad (9)$$

There is also an *upper bounding curve* E^u defined by the maximal intensity minus the *saturation margin* $\nu > 1$

$$\overbrace{E^u(\Delta t_i)}^{\text{Upper bounding}} = \overbrace{((2^m - \nu) \cdot g_\lambda^L[1,1] + g_\lambda^L[1,2])}^{\text{Fixed upper exposure}} \overbrace{\Delta t_i^{-1}}^{\text{Shutter speed}} . \quad (10)$$

Following this pattern, the ideally sampled intensity $I_{\text{mid}} = \frac{1}{2}(2^m - 1)$ defines the curve E^{mid}. At this middle intensity (eq. 6) the maximal sampling density is found. The curve is

$$\overbrace{E^{\text{mid}}(\Delta t_i)}^{\text{Optimal sampling}} = \overbrace{((2^m - 1)/2 \cdot g_\lambda^L[1,1] + g_\lambda^L[1,2])}^{\text{Fixed central exposure}} \overbrace{\Delta t_i^{-1}}^{\text{Shutter speed}} . \quad (11)$$

Eq. 11 is called *optimal sampling curve* (see figure 3).

2) Calibrated Exposures: The previous bounding curves describe the continuous range which can be sampled depending on the integration time. Each integration time available on the camera has an associated irradiance range.

In figure 3 (at marker 1), the shortest available integration time $\Delta t_1 = 3.0994415 \ \mu s$ captures the highest reachable irradiance. The maximal intensity value with $(2^{(m:=8)} - 1) = 255$ obtained at the time interval Δt_1 is (using for example the calibration $g_\lambda^L = [0.00811, -0.0289]^T$ from figure 2) produced by the sensor irradiance $E((2^m - 1), \Delta t_1) = (2^m - 1) \cdot g_\lambda^L[1,1] + g_\lambda^L[1,2])(\Delta t_1)^{-1} = 657,908.9$, meanwhile its lower reachable intensity η corresponds to the irradiance $E(\eta, \Delta t_1) = 2,616.6$. Notice the nonlinear behavior at the middle intensity value $E(\frac{1}{2}(2^m - 1), \Delta t_1) = 324,292.3$. Unfortunately, these extreme values are usually corrupted due to the broken reciprocity in physical sensors (see section II). In order to address this, the kernel weighting (in terms of the sampling density of eq. 6) enables the integration of the information robustly. This means, beyond a cut-off sampling density κ the irradiance cannot be reliably captured. For example, the region within the cut-off sampling density $\kappa = 0.001$ is bounded with the intensities $[16, 239]$ (see figure 3 at (1)). By applying kernel weighting, the sensor irradiance range can be robustly sensed to $[32015.2, 616569.4]$ at Δt_1 from a larger (only ideal) range of $[2616.6, 657908.9]$. The kernel-integrated range is only 89.2% of the ideal capacity of the sensor. In figure 3 (at marker 2), the same situation is found at the longest integration time $\Delta t_{651} = 24.491072$ ms. In figure 3 (at marker 3), the lower irradiance captured by one exposure with the configuration $(\Delta t_{217} = 6.359 \ @ \ \kappa = 0.001)$ is denoted as $E(I = 16, \kappa = 0.001) = 15.6$, whereas the upper reachable irradiance as $E(I = 239, \kappa = 0.001) = 300.5$. This range emphasized in figure 3 (at marker 4) solely depends on the radiometric calibration and the sampling density cut-off.

Since the radiometric calibration is constant (for a fixed camera-lens combination), the variation of the cut-off (sampling density) narrows or expands the reachable irradiance range at each particular exposure. The lower the cut-off the wider the range. Accordingly, with a lower cut-off less exposures are needed to cover the whole irrandiance range of the camera sensor. However, a low cut-off would also produce holes in the radiance distribution of the scene. Hence, the cut-off estimation (in algorithm 2) ensures the whole irradiance range by covering at least one exposure within the minimal sampling density κ_0. The irradiance range captured with all available integration times (using a low cut-off sampling density $\kappa = 0.001$) is $[4.1, 616569.4]$. The intensity value requires at least $\lceil \ln_2(616, 569.4 - 4.1) \rceil = 20$ bits for its integer representation. Thus, the radiance representation is 32 bit floating point per color channel.

3) Minimal Calibrated Exposure Set: Now with the proper unified representation for the extended radiometric calibration and integration time in figure 4, it is possible to introduce the Algorithm 2 which formally describes the applied method to select the minimal exposure set $\mathcal{E}_{\min}$. The strategy is to select the exposures producing the minimal sensor irradiance overlapping (see zoom in figure 5). This process starts from the shortest integration time towards the longest integration time. See the resulting set $\mathcal{E}_{\min}$ and minimal safe sampling density κ_0 in figure 4.
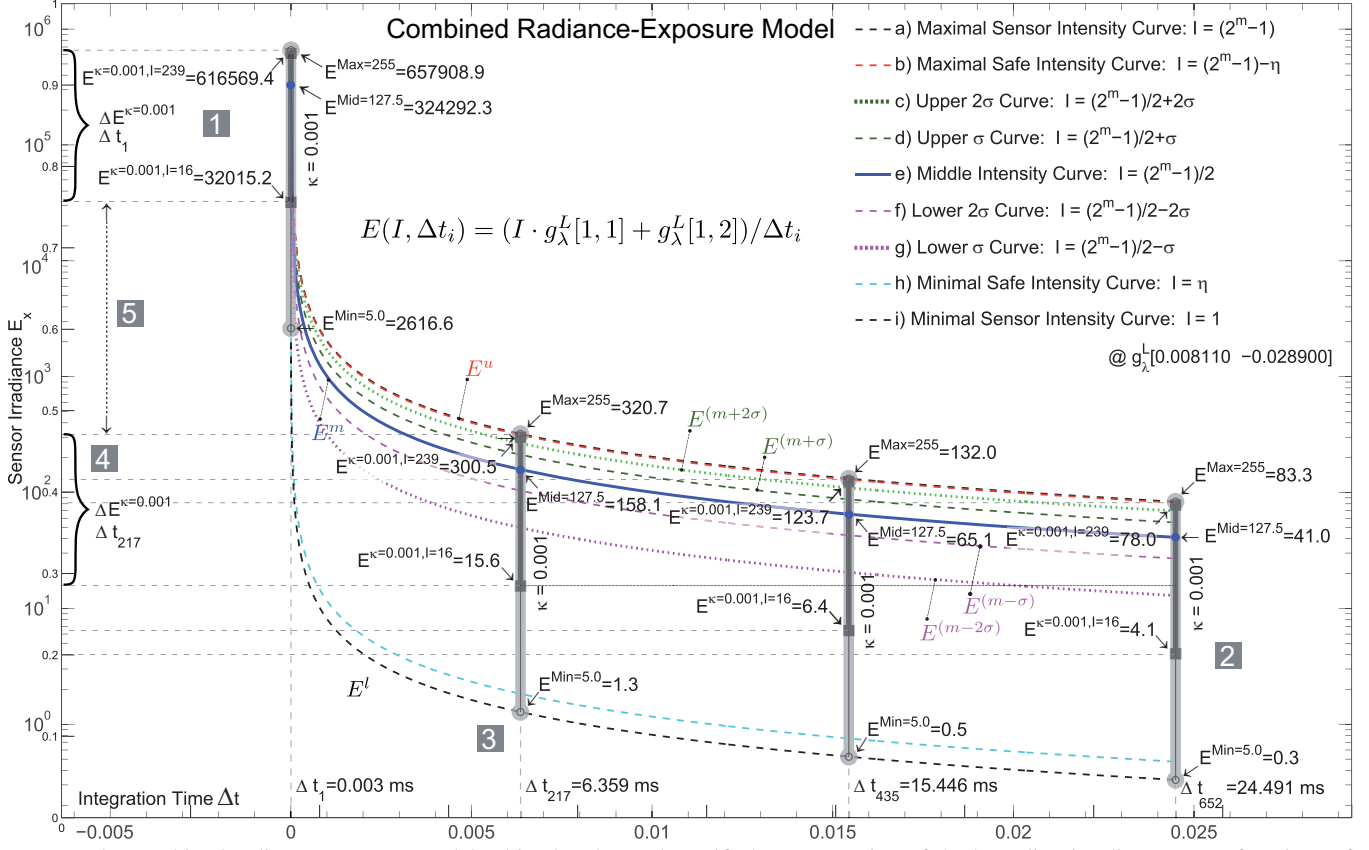
Fig. 3. The combined radiance-exposure model. This plot shows the unified representation of the bounding irradiance curve functions of the exposure times. This representation merges the radiometric image formation process with the extended radiometric calibration while considering the kernel sampling density. Four integration examples are marked and discussed along section IV. Since the sampling density is defined by a Gaussian kernel, thus it is possible to obtain the envelope curves at the lower and upper density quantiles (each at the first and second standard deviations $E^{(\text{mid}+q\sigma)} \mid q \in \{-2,-1,1,2\}$.

Algorithm 2 *Extract-Minimal-Exposure-Set* $(\kappa,\sigma,g_\lambda^L,\eta,\nu,\mathcal{E})$

Require: $\kappa > 0$ **and** $|\mathcal{E}| > 1$
1: $\mathcal{E}_{\min} \leftarrow \Delta t_1$
2: $I_l \leftarrow \mathbf{max}(\frac{1}{2}(2^m - 1) - \ln(\kappa\sigma^2)^{\frac{1}{2}}, \eta)$
3: $I_u \leftarrow \mathbf{min}(\frac{1}{2}(2^m - 1) + \ln(\kappa\sigma^2)^{\frac{1}{2}}, \nu)$
4: **for** $i = 2$ **to** $|\mathcal{E}|$ **do**
5: $\quad E_i^l \leftarrow (I_l g_\lambda^L[1,1] + g_\lambda^L[1,2])/\Delta t_i$
6: $\quad E_i^u \leftarrow (I_u g_\lambda^L[1,1] + g_\lambda^L[1,2])/\Delta t_i$
7: $\quad O_{\min} \leftarrow \infty, \ k = 0$
8: $\quad$ **for** $j = (i+1)$ **to** $|\mathcal{E}|$ **do**
9: $\quad\quad E_j^l \leftarrow (I_l g_\lambda^L[1,1] + g_\lambda^L[1,2])/\Delta t_j$
10: $\quad\quad E_j^u \leftarrow (I_u g_\lambda^L[1,1] + g_\lambda^L[1,2])/\Delta t_j$
11: $\quad\quad$ **if** $(E_j^u > E_i^l) \wedge ((E_j^u - E_i^l) < O_{\min})$ **then**
12: $\quad\quad\quad O_{\min} \leftarrow (E_j^u - E_i^l), \ k \leftarrow j$
13: $\quad\quad$ **end if**
14: $\quad$ **end for**
15: $\quad$ **if** $(k > 0)$ **then**
16: $\quad\quad \mathcal{E}_{\min} \leftarrow \mathcal{E}_{\min} \cup \Delta t_k$
17: $\quad$ **end if**
18: **end for**
19: **return** $\mathcal{E}_{\min}$

V. EXPERIMENTAL EVALUATION

The coordinated integration of both methods (our previous work [1] and the novel method presented in this article in section IV) enables humanoid robots to acquire the visual manifold in a robust, efficient and high quality manner (see figure 5 and 6). This high signal-to-noise ratio HDR image results from integrating multiple images while precisely controlling the exposure. The minimal amount of exposures is soundly attained by the minimal exposure set which is extracted based on the improved radiometric calibration and the exposure granularity of the cameras. This determines the minimal amount of images to be captured while simultaneously avoiding redundant sampling and ensuring high quality.

VI. CONCLUSIONS

The noiseless images captured by the temporal fusion method are systematically merged into high-dynamic-range images by optimal exposure bracketing. This dynamic range expansion is precisely archived through the improved radiometric calibration and meticulous exposure control. The proposed radiometric calibration model properly holds the reciprocity principle even at the intensity extrema with an efficient computational representation for continuous images. Based on these elements, the optimal combination of temporal fusion and exposure bracketing is properly controlled. This is concretely reflected in the high quality images for feature extraction with minimal acquisition time. These observations are supported by experimental evaluation. In figure 6, the results of the method introduced in this article show the minimal exposure set with 11 out of 651 available exposures (**only 1.69% of the exposures are actually needed**). The results also expose the particular
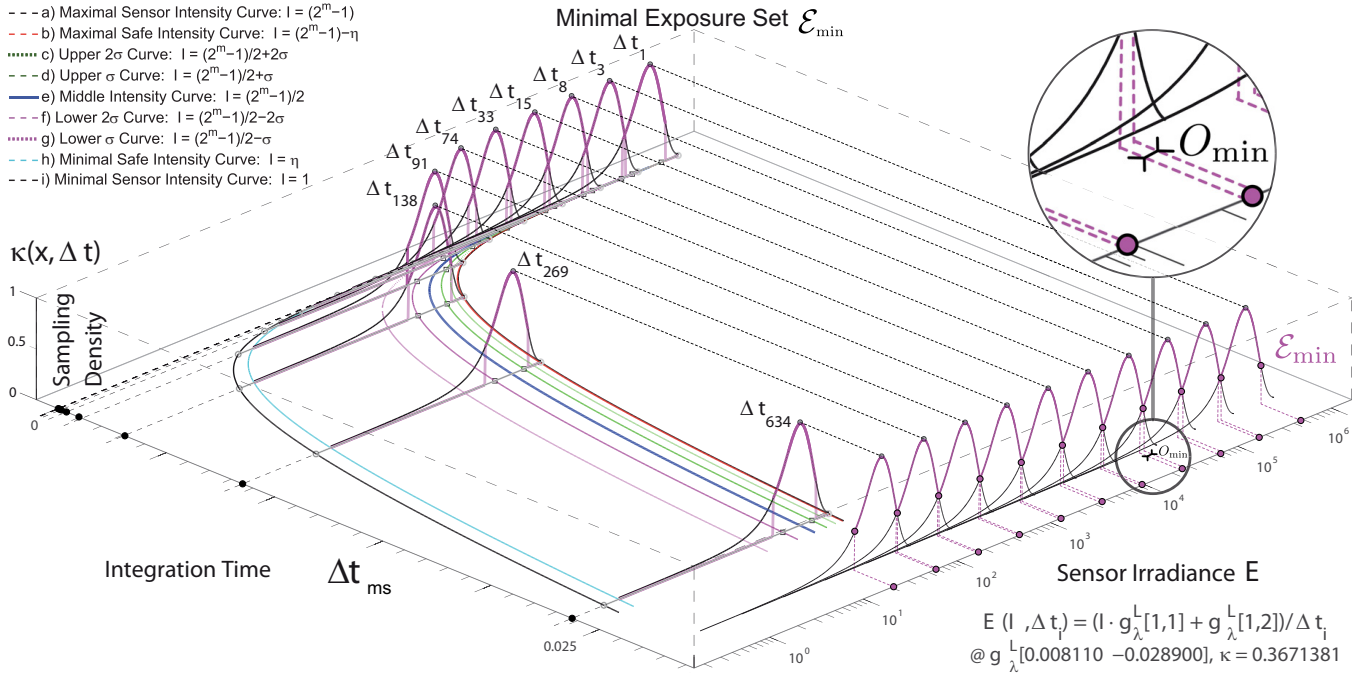
Fig. 4. The minimal calibrated exposure set containing 11 exposures (out of 652 available in $\mathcal{E}$). $\mathcal{E}_{min} := \{\Delta t_1 = 0.0031, \Delta t_3 = 0.0083, \Delta t_8 = 0.0212, \Delta t_{15} = 0.0550, \Delta t_{33} = 0.1488, \Delta t_{74} = 0.3989, \Delta t_{91} = 1.1075, \Delta t_{138} = 3.0665, \Delta t_{269} = 8.5269, \Delta t_{634} = 23.7408, \Delta t^*_{651} = 24.4911\}$ in ms (Δt_{651} is not depicted to avoid overlapping). This results from algorithm 2 with $\kappa_0 = 0.3671381$.

Fig. 5. Visual manifold acquisition. The HDR mesh is the $\ln(E)$ compressed representation of the radiance scene acquired with VGA resolution using the 4mm lens, see also the integrated HDR acquisition pipeline in figure 1). The total acquisition time of this HDR image is 6.672 sec fusing 11 exposures with 16 samples during the temporal fusion per iso-exposure. When fusing only one out of ten (1/10) exposures of the total 651 (available in camera) it takes 34.236 sec to acquired the HDR image. When capturing with less than (1/10) exposures the result presents holes in the high range regions of the images, for instance at laps or in the ceiling, see marked zone in figure 6. This means a **513.11% time performance improvement** compared to the 1/10 regular distributed (naive exposure) plan.

cut-off of the sampling density ($\kappa_0 = 0.3671381$) which is determined by the larger gap (O_{min} in lines 7-12 of algorithm 2) between consecutive exposures using the kernel weighting of eq. 6. A lower cut-off $\kappa < \kappa_0$ would generate either sampling holes in the irradiance domain because the minimal sampling density would not be thoroughly ensured. Thus, the minimal exposure set is fully determined by our method based exclusively on the radiometric calibration and integration time granularity for the camera. Based on this method, it is possible to acquire the visual manifold for complex visual tasks using standard cameras. This is the missing component for a robust, precise and efficient visual transducer for robot vision.
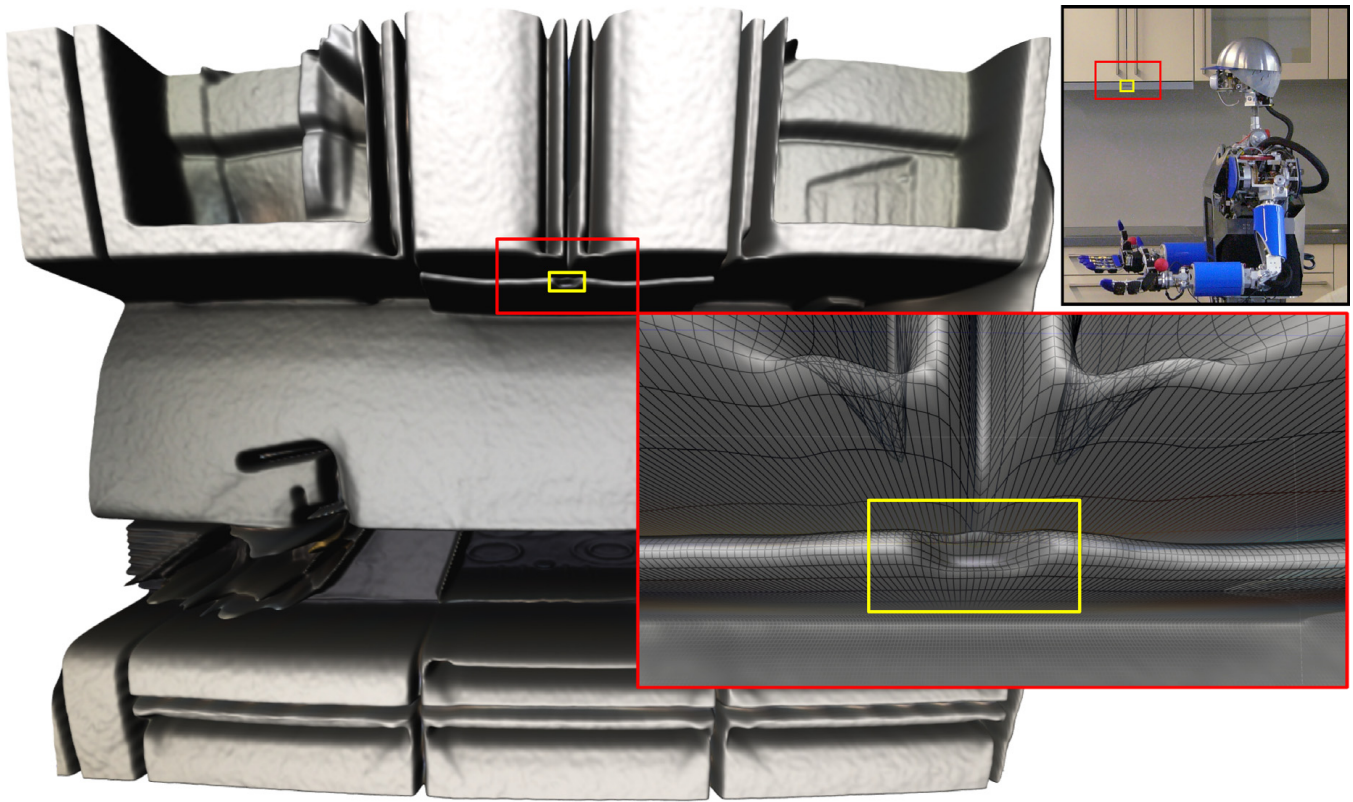
VII. ACKNOWLEDGMENT

Fig. 6. Everyday scene where the visual perception of the humanoid robot should recognize environmental elements for attaining its own 6D pose. This viewpoint contains high-dynamic-range produced by large radiance difference between the doors and stove. The left HDR mesh $\ln(E)$ of the region marked with a rectangle in figure 5. The (12mm lens) visual manifold, the large focal length significantly improves the spatial resolution. The red rectangle shows the synergistic integration of both temporal and exposure fusion methods.

REFERENCES

[1] D. Gonzalez-Aguirre, T. Asfour, and R. Dillmann, "Robust Image Acquisition for Vision-Model Coupling by Humanoid Robots," in *IAPR Conference on Machine Vision Applications*, 2011.

[2] D. Gonzalez-Aguirre, T. Asfour, and R. Dillmann, "Eccentricity Edge-Graphs from HDR Images for Object Recognition by Humanoid Robots," in *Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference on*, dec. 2010, pp. 144 –151.

[3] T. Asfour, K. Regenstein, P. Azad, J. Schroder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control," in *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, 2006, pp. 169 –175.

[4] K. Welke, T. Asfour, and R. Dillmann, "Bayesian Visual Feature Integration with Saccadic Eye Movements," in *Humanoid Robots, 2009. Humanoids 2009. 9th IEEE-RAS International Conference on*, 7-10 2009, pp. 256 –262.

[5] D. Vernon, G. Metta, and G. Sandini, "The iCub Cognitive Architecture: Interactive Development in a Humanoid Robot," in *Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on*, july 2007, pp. 122 –127.

[6] T. Asfour, K. Welke, P. Azad, A. Ude, and R. Dillmann, "The Karlsruhe Humanoid Head," in *Humanoid Robots, 2008. Humanoids 2008. 8th IEEE-RAS International Conference on*, 1-3 2008, pp. 447 –453.

[7] G. Metta, L. Natale, F. Nori, and G. Sandini, "The iCub Project: An open Source Platform for Research in Embodied Cognition," in *Advanced Robotics and its Social Impacts (ARSO), 2011 IEEE Workshop on*, oct. 2011, pp. 24 –26.

[8] M. D. Grossberg and S. K. Nayar, "What can be known about the radiometric response function from Images," in *In Proc. of the ECCV*, 2002, pp. 189–205.

[9] Mann, Picard, S. Mann, and R. W. Picard, "On Being 'undigital' With Digital Cameras: Extending Dynamic Range By Combining Differently Exposed Pictures," in *Proceedings of IS&T*, 1995, pp. 442–448.

[10] P. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," in *SIGGRAPH '97: 24th annual conference on Computer graphics and interactive techniques*, 1997, pp. 369–378.

[11] T. Mitsunaga and S. Nayar, "Radiometric Self Calibration," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 1, 1999, pp. 2 vol. (xxiii+637+663).

[12] G. Krawczyk, M. Goesele, and H.-P. Seidel, "Photometric Calibration of High Dynamic Range Cameras," Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany, Research Report MPI-I-2005-4-005, April 2005.

[13] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum, "Radiometric Calibration from a Single Image," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, june-2 july 2004, pp. II–938 – II–945 Vol.2.

[14] IIDC2, *IIDC2 Digital Camera Control Specification Ver.1.0.0*, Japan Industrial Imaging Association Standard JIIA CP-001-2011 / 1394 Trade Association Specification TS2011001, 2-10-15 Nakameguro, Yamate Ave. K Bldg., Meguro Tokyo 153-0061 Japan, 2012.

[15] M. Granados, B. Ajdin, M. Wand, and C. Theobalt, "Optimal HDR Reconstruction with Linear Digital Cameras," *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 215–222, 2010. [Online]. Available: http://www.computer.org/portal/web/csdl/doi/10.1109/CVPR.2010.5540208

[16] M. A. Robertson, S. Borman, and R. L. Stevenson, "Estimation-Theoretic Approach to Dynamic Range Enhancement Using Multiple Exposures," *Journal of Electronic Imaging*, vol. 12, p. 2003, 1999.

[17] A. J. P. Theuwissen, *Solid-state Imaging with Chargecoupled Devices*, Dordrecht, Ed. Kluwer Academic Publishers, 1995.

[18] B. Mackiewich, "Intracranial boundary detection and radio frequency correction in magnetic resonance images," 1995.

[19] V. Madisetti and D. Williams, *The Digital Signal Processing Handbook*, ser. Electrical engineering handbook series. CRC Press, 1998. [Online]. Available: http://books.google.de/books?id=14RXPgAACAAJ

[20] PTGrey, *Dragonfly Technical Reference Manual*, 5-8 2008.