

# Eccentricity Edge-Graphs from HDR Images for Object Recognition by Humanoid Robots

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

Karlsruhe Institute of Technology, Adenauerring 2, Karlsruhe-Germany.  
{gonzalez,asfour,dillmann}@ira.uka.de

**Abstract**—A novel structural and statistical approach for model-based visual object recognition using geometric radiance saliencies is presented. The approach acquires accurate high dynamic range images to properly capture complex heterogeneously lighted scenes. Based on these images, the receptive radiance saliency is computed through a Gabor kernel set. This oriented saliency is used to extract and refine the radiance edge-graphs. Subsequently, the combination of two distributions, (i) the topological connectivity and (ii) the spatial arrangement of the subpixel nodes provides a propitious insight into the underlying geometrical composition of the radiance edges. The proposed characterization of the combined distribution profitably unveils and simultaneously segments the geometric edge primitives. Finally, uncovering the complementary geometric patterns and reinforcing the structural regularity is accomplished by the proposed extended perceptual organization. Experimental evaluation with the humanoid robot ARMAR-III is presented.

## I. INTRODUCTION

In everyday application, humanoid robots should be able to visually explore and recognize objects in their environment for self-localization [1], environmental status mapping [2], grasping and manipulation [3],[4]. In these active fields much research is being done and as a consequence, considerable results have been recently achieved and several humanoid robots expose elaborated visual recognition capabilities [5],[6]. However, these approaches assume restricted light conditions. This illumination assumption is explicitly attained either by controlling the light sources or restricting the 6D-pose of the robot during recognition [6],[7].

Due to these limitations, the flexibility and applicability of humanoid robots are strongly reduced in the presence of everyday high-contrast and/or dynamically varying illumination, for instance, the backlighting of a window, the interior of an oven, the complex structures inside a dishwasher or the metallic/reflective surface of a microwave, see Fig.1.

Avoiding the ubiquitous high-contrast image-content in everyday humanoid robot applications is not plausible. Neither is it possible to mend the dynamic range issue through adaptive exposure in order to allow recognition approaches to implicitly hold the illumination assumption. This issue arises because the adaptively selected exposure attempts to preserve the majority of the image-content at the cost of losing the extremal regions. This condition generates severe rate-distortion quantization effects which substantially diminish the image quality producing local over- or under-exposure.

Additionally, even if the objects are completely captured within an *unclipped* image region (where the discrete values

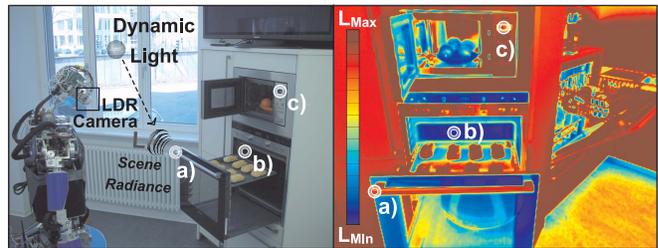


Fig. 1. The ubiquitous wide intra-scene radiance range. In *made-for-humans* environments, humanoid robots are confronted with high-contrast scenes, dynamically varying illuminations and complex materials. On the left side, the humanoid robot ArmAr-III [9] encounters a wide intra-scene illumination in a household environment. On the right side, the HDR image captured with the LDR (low dynamic range) camera of the robot using the method presented in Sec.II. The pseudo-color representation allows better visualization of the radiance levels in the a) bright, b) dark and c) complex material regions which require precise mechanisms to acquire HDR images.

are neither *white-* nor *black-clipped*), the resulting intensity values are still rarely a true measurement of the relative radiance in the scene. Under such *non-injective* circumstances, a single discrete intensity value comprises a wide range of scene radiance. This *many-to-one* mapping makes it unfeasible to robustly and accurately extract structured visual cues, namely geometric saliencies such as edges or corners, even when using computationally expensive approaches such as automatic space-scale selection [8]. Without these essential saliencies, the recognition with robust pose estimation becomes notably complicated or intractable.

These limitations can be overcome through this dependable *acquisition* and *processing* mechanism of HDR images. In Sec.II the acquisition mechanism robustly manages the intra-scene radiance range providing consistent measurements. Based on these HDR images, in Sec.III, the simultaneous extraction and segmentation of geometric consistent primitives is organized into an effective graph representation for object recognition. The experimental evaluation in Sec.IV supports the reliability of the approach. Finally, the Sec.V discusses the technical and theoretical contributions. See the schematic representation of the approach in Fig.2.

## II. HDR IMAGE ACQUISITION

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. These are the restrictions in the eyeball of the humanoid

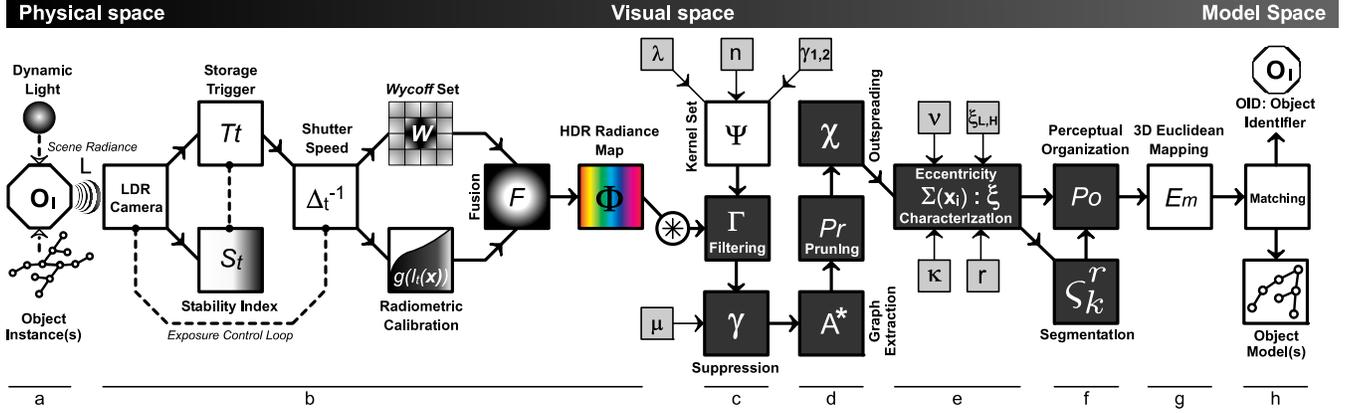


Fig. 2. Eccentricity edge-graphs from HDR images for object recognition. a) Complex high-contrast heterogeneously lighted scene as in Fig.1. b) HDR acquisition component. c) Oriented saliency is the receptive radiance energy for edge extraction and subpixel refinement. d) Edge-graph extraction, pruning and outspreading for the cohesiveness enhancement of the combined statistical distribution. e) Eccentricity characterization. f) Eccentricity segmentation and geometrical-statistical perceptual organization. g) 3D-Euclidean metric mapping. h) Object model(s) and matching elements.

robot ARMAR-III, see the detailed composition [10] and applications [11]. Due to these restrictions, an appropriate mechanism for capturing HDR images by suitably employing the humanoid robot LDR cameras is been introduced. The mechanism consists of two elements: First, the nonlinear transformation from the scene radiance to the discrete values of the image is attained, namely the radiometric calibration. Second, the proper exposure control of the camera by means of the shutter speed and the exposure-stability analysis was performed. Using these two elements, it is possible to capture the intra-scene wide radiance range by fusing a collection of  $m$  differently exposed images, the so called *Wycoff* set for *bracketing* in [12]. 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.

### A. Camera Radiometric Calibration

In contrast to [12] due to the image-content independence, the radiometric calibration  $g_\lambda : \mathbb{N} \mapsto \mathbb{R}$  was initially computed as described by *Debevec* [14], see Fig.3-a. However, the *Debevec* approach has two critical issues: i) There is no explicit criterion for the selection of the  $\lambda$ -smoothing regression factor. ii) The essential assumption of reciprocity between exposure and intensity values is not held on the extremal of the radiometric calibration curve. The results are noisy calibrations propagating into salient artifacts in the HDR image, see Fig.3. In order to anticipate these issues, an experimental evaluation of the smoothing factor for optimal selection was performed (see Fig.4) and an improvement extension for calibration is presented:

1) *Continuous Reciprocity-Consistent Calibration Model*: The exponential domain transformation of the discrete calibration curve

$$g_\lambda^*(I_t(\mathbf{x})) := \exp(g_\lambda(I_t(\mathbf{x}))) \quad (1)$$

leads to a weighted regression within this convenient domain in order to obtain a continuous model, see Fig.5,

$$g_\lambda^L : \mathbb{R} \mapsto \mathbb{R} ; g_\lambda^L(I_t(\mathbf{x})) \approx g_\lambda^*(I_t(\mathbf{x})),$$

in contrast to the results from [12] and [14], this calibration function properly holds the reciprocity assumption even at the extremal intensity, see Fig.6. In addition, the calibration model  $g_\lambda^L$  also allows the proper estimation of the sensor response function at real value intensities arising during the blending of  $n$  images with the same  $j := E_t \propto \Delta_t$  exposure in order to decrease the sensor noise, namely

$$\bar{I}_n^j(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n I_t^j(\mathbf{x}) \quad (2)$$

$$g_\lambda^M(\bar{I}_n^j(\mathbf{x})) = \log(g_\lambda^L(\bar{I}_n^j(\mathbf{x}))), \quad (3)$$

where  $\bar{I}_n^j(\mathbf{x})$  is the  $j$ -iso-exposed mean of  $n$  images and  $g_\lambda^M$  is the standard domain of the radiometric calibration model.

### B. Camera Exposure Control

When capturing the *Wycoff* set of a scene, the exposure value was precisely controlled by the shutter speed of the camera  $\Delta_t^{-1}$  according to [13] in a closed loop, see Fig.2-b. The storage trigger  $T_t$  was fired by means of the exposure-stability analysis, see Fig.7. This component takes into account the overall latency<sup>1</sup> of the capture system. This is done by the normalized *exposure-stability index*  $S_j(t) : \mathbb{N}^2 \mapsto \mathbb{R}^+$

$$S_j(t) = \frac{1}{\omega} \sum_{\mathbf{x}} I_t^j(\mathbf{x}), \quad (4)$$

which integrates the image intensity while controlling the exposure value  $j$ . The smoothed-gradient analysis of this index helps to determine reliable capture intervals  $D_n$ . Without this component, the storage trigger could be dispatched within an unstable interval, e.g., in Fig.7-b the instantaneous flickering between  $t_1$  and  $t_4$  with the peak at the index value  $S_2$ .

### C. HDR Image Synthesis

Finally, the fusion of the captured *Wycoff* set using the model calibration  $g_\lambda^M$  is expressed as

$$\Phi(\mathbf{x}) = \exp \left[ \frac{\sum_{j=1}^m \mathcal{N}(\bar{I}_n^j(\mathbf{x})) (g_\lambda^M(\bar{I}_n^j(\mathbf{x})) - \log(\Delta_{t_j}))}{\sum_{j=1}^m \mathcal{N}(\bar{I}_n^j(\mathbf{x}))} \right], \quad (5)$$

<sup>1</sup>Slightly irregular delays resulting from the data bus and the non-real-time humanoid robot modular controller architecture.

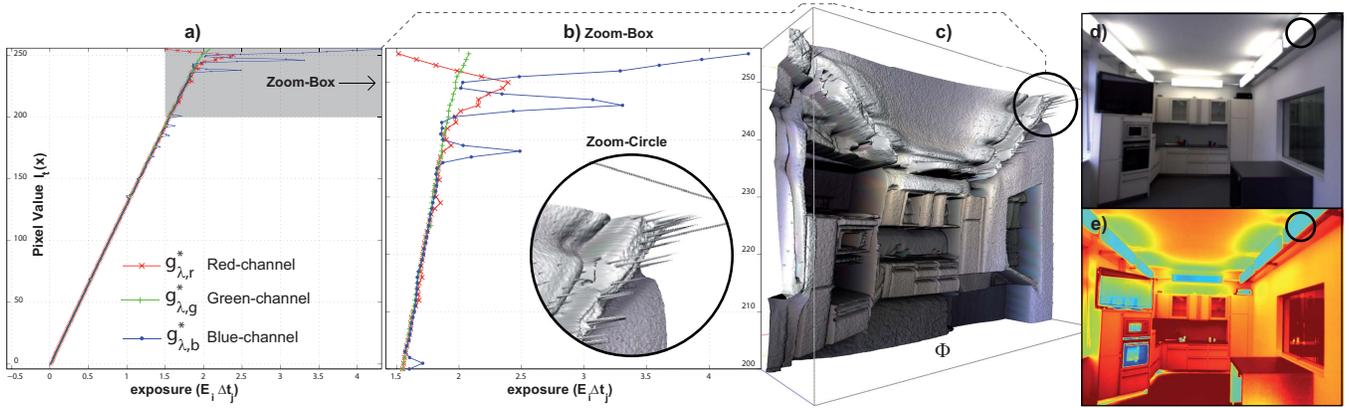


Fig. 3. The exponential domain of the Debevec radiometric calibration allows the detection of the detrimental artifacts. a) The direct output of the Debevec calibration curves  $g_{\lambda}^*(I_i(\mathbf{x}))$  per color channel. The top right image shows the source scene with a very wide intra-scene radiance range. This high range is ideal for calibration purposes. Notice the light gray masked region, within this region the straightforward application of the Debevec calibration produces detrimental artifacts. b) The zoom-box and the zoom-circle showing the resulting artifacts. c) Prominent artifact in the HDR image. Notice that this distortion is not easily detected by the human perception in the bottom right radiance map. However, these kind of artifacts severely impact the structural feature extraction for object recognition. d) The source Wycoff image element  $E_2$  with shutting speed  $\Delta_t = 4.375\text{ms}$  [13]. e) The HDR radiance image. Notice in particular the region where the peaks were produced by the high radiance of the kitchen lamps.

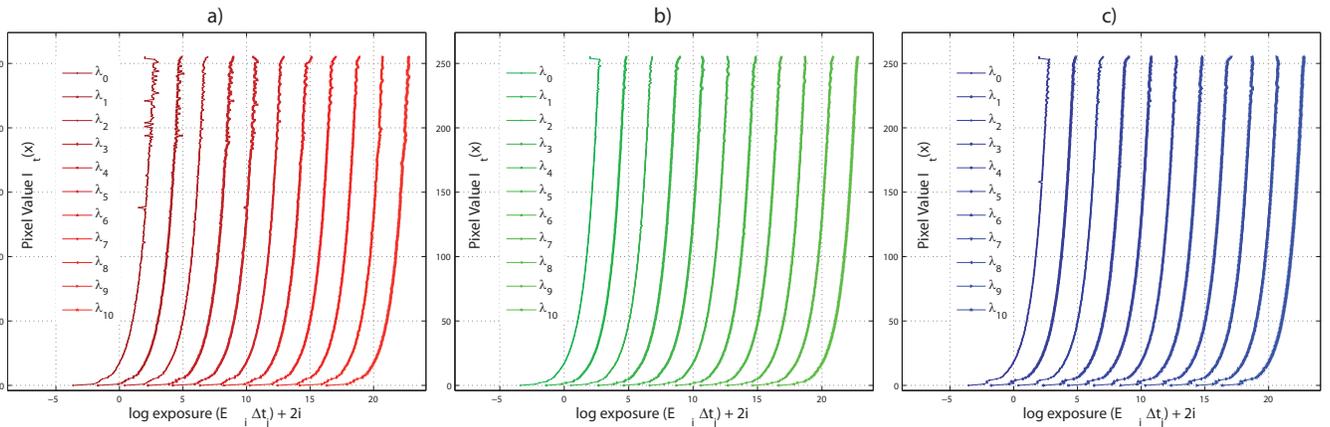


Fig. 4. The Debevec  $\lambda$ -optimal radiometric camera calibration by experimental evaluation of the  $\lambda$ -smoothing factor. a) The  $\lambda$ -smoothing curves of the red-channel shows that this is the most noisy channel of the camera. b) The  $\lambda$ -smoothing curves of the green-channel. This channel has the higher *signal-to-noise* ratio of all three channels. It is not surprising since the color camera has a Bayer-pattern sensor. c) The blue-channel  $\lambda$ -smoothing curves. Notice that for each  $\lambda_i$  plot, the horizontal axes were  $2i$  units right shifted to conveniently display the comparative effects of the  $\lambda$ -smoothing factor.

where  $\bar{I}_n^j(\mathbf{x})$  is the  $j$ -exposed blending intensity from Eq.2 and  $\mathcal{N}$  is the Gaussian kernel used for regression in Fig.5. The acquired spatial discrete non-quantized radiance map, namely the HDR image  $\Phi: \mathbb{N}^2 \mapsto \mathbb{R}^3$  is a consistent *up-to-scale* manifold of the scene radiance  $L$ , see Fig.8.

### III. ECCENTRICITY EDGE-GRAPHS

Because of its nature, the continuous intensity values of the HDR image cannot be reliably processed with edge detectors like  $n$ -order directional derivatives such as the *Sobel*, *Prewitt*, *Canny* filters or even more elaborate approaches as in [8]. This occurs because the parameters that these methods require are image-content dependent and can hardly be estimated automatically. In other words, the magnitude of the  $n$ -order directional derivative of a HDR image has a wide range multimodal distribution where the local maxima are indiscriminately spread. Moreover, when using global methods, the non-prominent local maxima cannot be substantially

distinguished due to the inherent effects (kernel/bandwidth selection resulting in over- or under-smoothing) produced during the density estimation used to represent the magnitude distribution. In addition, when using local methods, the parameter-varying batch analysis such as scale-space methods do not perform satisfactory and are neither computationally efficient nor geometrically consistent. This lack of consistency arises when geometric structured elements on the images have widely dispersed directional-derivative magnitudes along them, resulting in split segments since neither hysteresis nor adaptive-thresholding can adequately cope with such dynamic range. As a consequence, slightly salient structures are neglected, despite of their sound geometrical significance.

In order to manage these issues, a novel parameterless and non-iterative method for geometric edge extraction and representation is introduced. The method extracts the geometric primitives based on the characterization of the

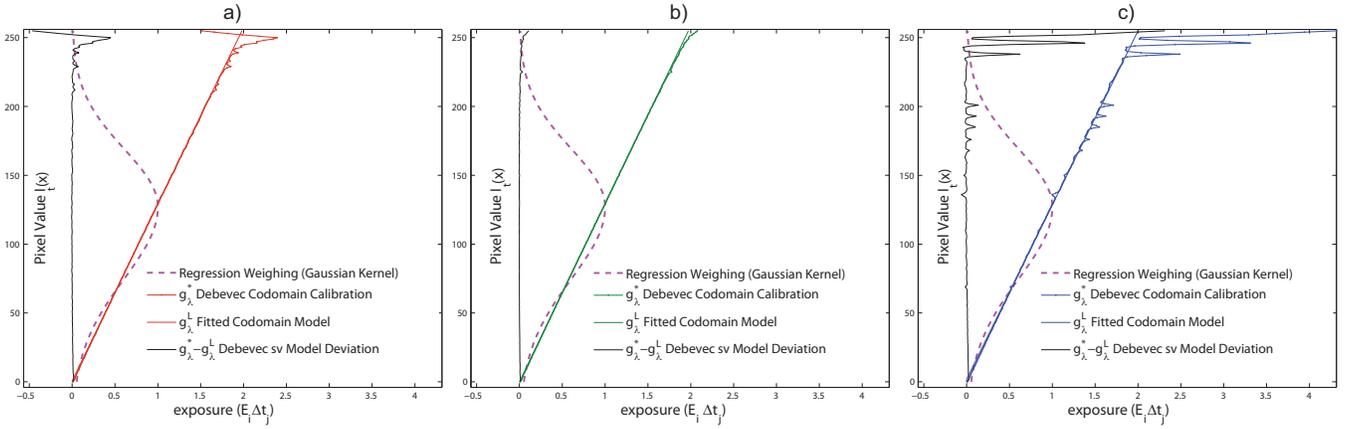


Fig. 5. The model calibration  $g_{\lambda}^L$  is accomplished by a weighted linear regression in the exponential domain. The magenta dashed weighting Gaussian kernel  $\mathcal{N}(\mu = 128, \sigma^2 = 1779.34)$  ponderates the Debevec  $\lambda$ -optimal  $g_{\lambda}^*$  Debevec calibration in order to estimate the linear model calibration  $g_{\lambda}^L$ . a) The resulting red-channel model-calibration is  $g_{\lambda,r}^L(I_{i,r}(\mathbf{x})) = \log(0.007961 * I_{i,r}(\mathbf{x}) - 0.001419)$ . b) The resulting green-channel model-calibration is  $g_{\lambda,g}^L(I_{i,g}(\mathbf{x})) = \log(0.007602 * I_{i,g}(\mathbf{x}) + 0.003388)$ . c) The resulting blue-channel model-calibration is  $g_{\lambda,b}^L(I_{i,b}(\mathbf{x})) = \log(0.007798 * I_{i,b}(\mathbf{x}) + 0.001609)$ . The kernel selection and its bandwidth allow the proper integration of the information within the smooth region of the curve  $g_{\lambda}^*$  while gradually disregard the broken-reciprocity regions. In this manner, the calibration deviations and detrimental artifacts are correctly removed. These results were obtained using 16 images with 273 pixels per image. The location of the pixels were selected according to the maximal intensity variance heuristic with a 32 pixels radius dominance.

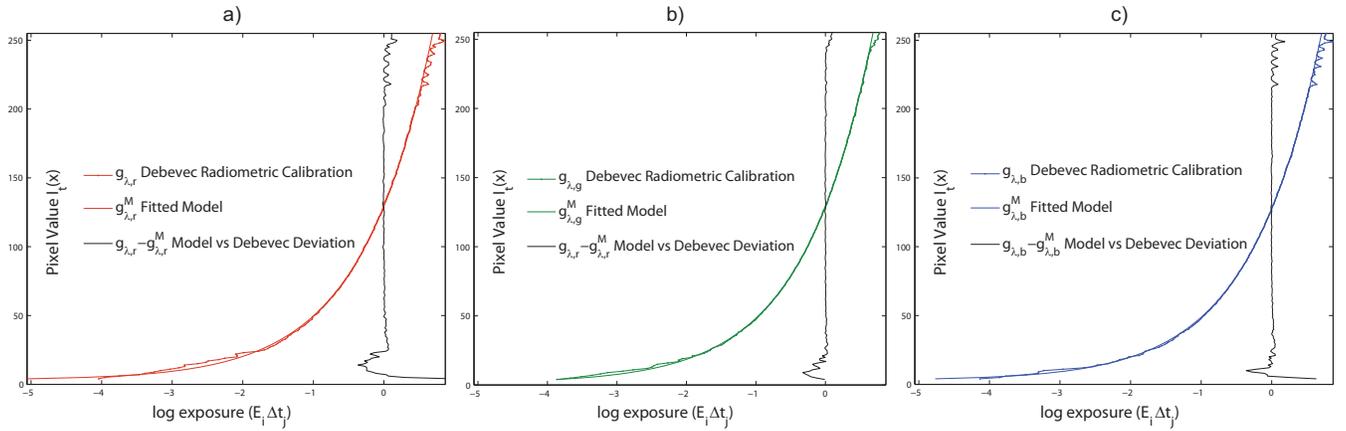


Fig. 6. The  $\lambda$ -factor selection is based on the minimal deviation of the regression model, namely  $\operatorname{argmin}\{\lambda : \sum_{i=0}^{255} (g_{\lambda}^M(i) - g_{\lambda}(i))^2\}$ . The plots of the Debevec  $\lambda$ -optimal, the continuous model  $g_{\lambda}^M$  and their mutual deviations. a) Red-channel  $\lambda_r = 0.94726$ . b) Green-channel  $\lambda_g = 0.95421$ . c) Blue-channel  $\lambda_b = 0.96014$ . These results were obtained using the left peripheral camera of the humanoid robot Armar-III.

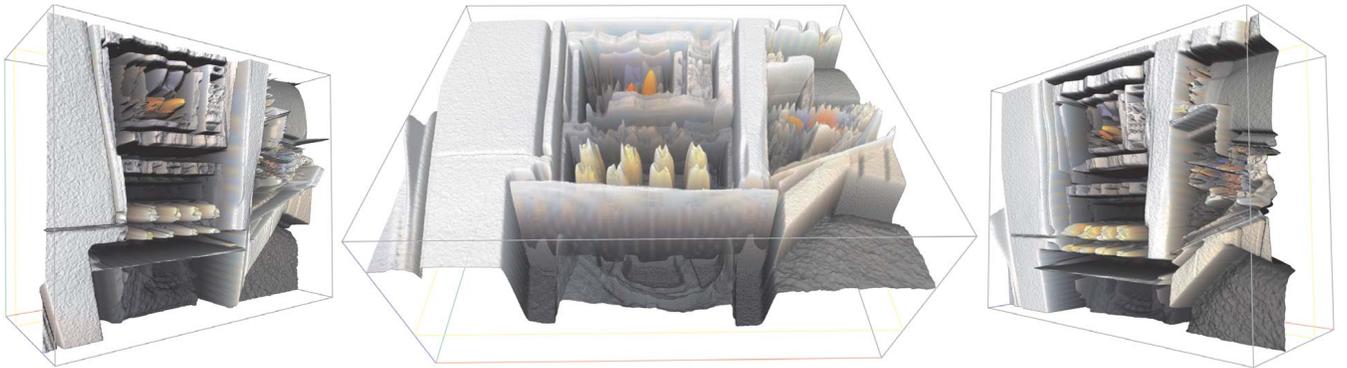


Fig. 8. The  $\Phi$  HDR image. This representation is  $\log(\Phi)$  compressed in order to allow a suitable visibility of the sensibility and consistency of the acquisition component. All three images correspond to a single HDR image captured by the humanoid robots Armar-III with its left peripheral camera in the scene setup of Fig.1. The variation of the perspective displays the discussed regions in Sec.I, compare these superior results with those in the Fig.3.

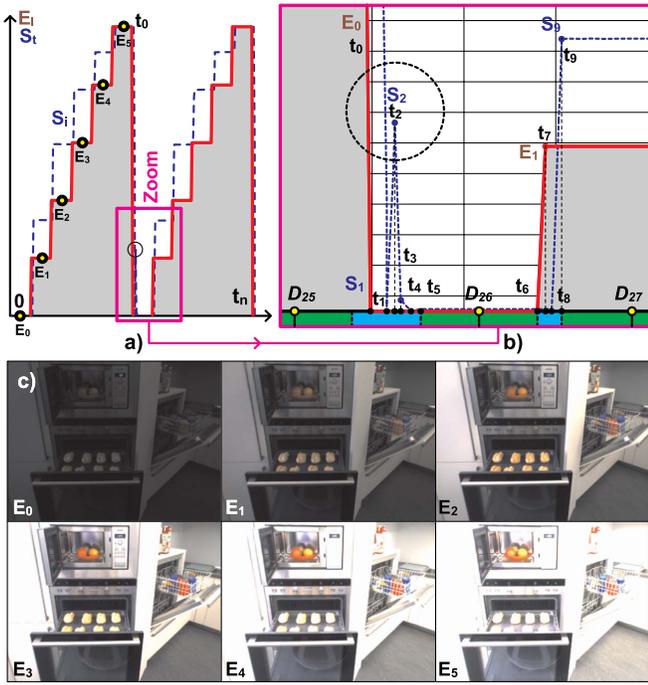


Fig. 7. Camera exposure control. a) The normalized exposure-stability index  $S_t$  in blue dashed line is closely correlated to the commanded exposure  $E_t$  shown in a continuous red line. Notice that the exposures  $E_0 - E_5$  are the time stamps where the Wycoff set was captured. b) The zoom window shows the importance of the exposure-stability analysis in order to consistently capture HDR images. The  $D_i$  reliable exposure intervals are  $70 \pm 12$ ms away from each other when capturing at 15 fps. c) The Wycoff set of the scene setup in Fig.1, see the HDR synthesis of these images by the Eq.5 into the Fig.8.

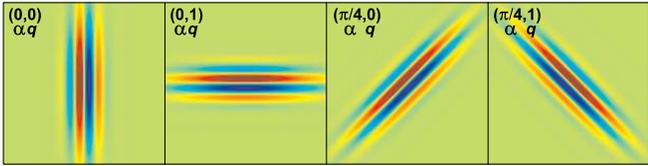


Fig. 9. The Gabor kernel set  $\Psi(\mathbf{x}, \alpha + \frac{q\pi}{2}, \gamma_1, \gamma_2, \lambda)$ ,  $q \in \{0, 1\}$  is composed of orientation-complementary band-pairs.

subpixel-nodes by means of their combined topological-saliency distribution.

#### A. Oriented Saliency Extraction

The receptive radiance saliency  $\Gamma: \mathbb{N}^2 \mapsto \mathbb{R}^2$  is computed by the composed Gabor kernel, see Fig.2-c,

$$\Psi(\mathbf{x}, \alpha, \gamma_1, \gamma_2, \lambda) = \exp\left(-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x}\right) \sin\left(\pi \frac{U_{(\alpha)} \cdot \mathbf{x}}{\lambda}\right), \quad (6)$$

in contrast to [15], this formulation provides the means to conveniently control the impulse response function by

- **Kernel Orientation:** In Eq.6 and Eq.7, the semi-axes

$$U_{(\alpha)} = [\cos \alpha \ \sin \alpha]^T \text{ and } V_{(\alpha)} = [-\sin \alpha \ \cos \alpha]^T,$$

allow the saliency extraction in a  $\alpha$ -target direction.

- **Kernel Support:** It is adjusted by the ratio  $\gamma_1 : \gamma_2$  in the Eq.6 by the covariance matrix  $\Sigma$ ,

$$\Sigma = \text{diag}[\gamma_1 \ \gamma_2] \begin{bmatrix} U_{(\alpha)} & V_{(\alpha)} \end{bmatrix}, \quad (7)$$

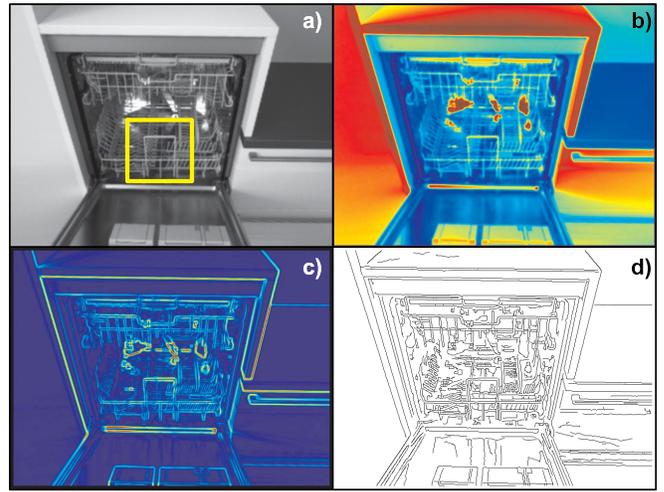


Fig. 10. Saliency extraction process. a) Wide intra-scene radiance range with complex structures and metallic surfaces. Notice the selected yellow window used in Fig.11. b) The HDR image  $\Phi$ . c) Receptive radiance saliency  $\|\Gamma\|$  from Eq.8. d) Active map  $\Upsilon$  from Eq.9. The coherency factor  $\mu = \cos(\pi/8)$  improves the quality of the map and is tightly related to the image discretization, thus higher values require to increase the image resolution by interpolation.

this adjusts the smoothing extraction trade-off.

The computation of the receptive radiance saliency  $\Gamma$  requires  $k \geq 1$  band-pairs  $(\alpha, q)$  of orientation-complementary kernels (see Fig.9), namely

$$\Gamma(\mathbf{x}) = \sum_{p=1}^k \sum_{q=0}^1 U_{(p\pi/2k)} \left\{ \Phi(\mathbf{x}) \otimes \Psi\left(\mathbf{x}, p\pi/2k + \frac{q\pi}{2}, \gamma_1, \gamma_2, \lambda\right) \right\}, \quad (8)$$

where  $\otimes$  denotes the convolution operator, see Fig.10.

#### B. Dual Non-maximal Suppression

Based on the saliency direction  $\hat{\Gamma}(\mathbf{x}) = \Gamma(\mathbf{x}) \cdot \|\Gamma(\mathbf{x})\|^{-1}$ , the pixel location  $\mathbf{x} \in \mathbb{N}^2$  has two neighbors  $\tilde{\mathbf{x}}_{\pm} \in \mathbb{N}^2$  along  $\hat{\Gamma}(\mathbf{x})$ ,

$$\tilde{\mathbf{x}}_{\pm} = \mathbf{x} \pm \omega \left[ \hat{\Gamma}(\mathbf{x}) \cdot \eta \left( \hat{\Gamma}(\mathbf{x}) \right)^{-1} \right],$$

where the function  $\eta: \mathbb{R}^2 \mapsto \mathbb{R}^+$  attains the largest absolute component of the unitary vector  $\hat{\Gamma}(\mathbf{x})$  and the function  $\omega: \mathbb{R}^2 \mapsto \mathbb{N}^2$  rounds its vector argument by component.

In contrast to edge detectors (see Fig.11), the dual non-maximal suppression  $\Upsilon: \mathbb{N}^2 \mapsto \{0, 1\}$  selects pixels close to the edge by considering not only the saliency norm  $\|\Gamma(\mathbf{x})\|$  but also the saliency direction coherency  $\mu$ , see Fig.10-d,

$$\Upsilon(\mathbf{x}) = \begin{cases} 1, & \langle \|\Gamma(\mathbf{x})\| > \|\Gamma(\tilde{\mathbf{x}}_+)\| \rangle \wedge \langle \|\hat{\Gamma}(\mathbf{x}) \cdot \hat{\Gamma}(\tilde{\mathbf{x}}_+)\rangle > \mu \rangle \wedge \\ & \langle \|\Gamma(\mathbf{x})\| > \|\Gamma(\tilde{\mathbf{x}}_-)\| \rangle \wedge \langle \|\hat{\Gamma}(\mathbf{x}) \cdot \hat{\Gamma}(\tilde{\mathbf{x}}_-)\rangle > \mu \rangle \\ 0, & \text{Otherwise.} \end{cases} \quad (9)$$

#### C. Edge Subpixel Optimization

In order to properly extract the edge-graphs, it is necessary to optimize the edge-pixels to their subpixel position. Every active pixel  $\mathbf{x}$  on the binary map  $\Upsilon(\mathbf{x})$  is located at most  $\sqrt{2}$  units away from the local maximum edge on  $\|\Gamma(\mathbf{x})\|$ . In order to reach this maximum, the discrete location  $\mathbf{x}$  is refined to its subpixel counterpart  $\hat{\mathbf{x}} \in \mathbb{R}^2$  by a gradient ascent along the radiance saliency direction using the bicubic convolution algorithm  $\Lambda$  [17], see Fig.12-a,

$$\hat{\mathbf{x}} = \Lambda[\mathbf{x}, \hat{\Gamma}(\mathbf{x})] \Rightarrow \langle \|\Gamma(\hat{\mathbf{x}})\| \geq \|\Gamma(\mathbf{x})\| \rangle \wedge \langle \|\hat{\mathbf{x}} - \mathbf{x}\| < \sqrt{2} \rangle. \quad (10)$$

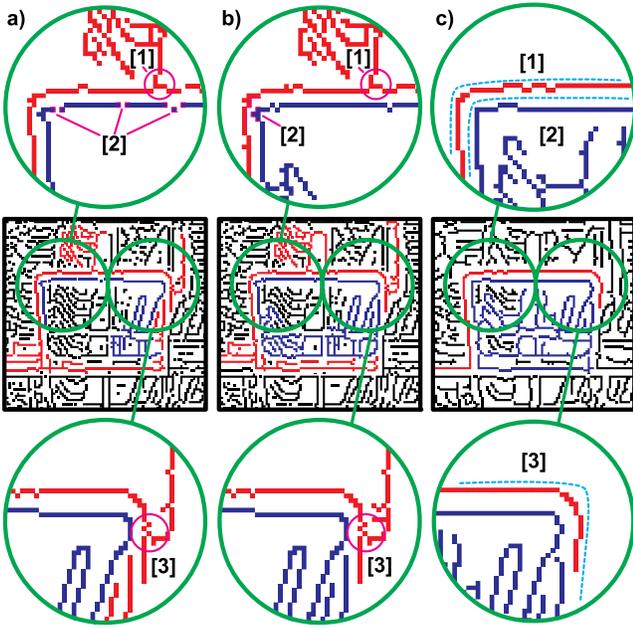


Fig. 11. Comparative edge-extraction results from the yellow window in Fig.10-a. The Canny filter applied with maximal sensitivity (the lower and the higher threshold set to 0, i.e., only non-maximal suppression). a) Column results from the implementation in IVT [16]. b) Column results from the implementation in OpenCV. c) Edge resulting from the method presented in Sec.III supporting the greater quality of the approach. Notice that the red edges obtained in a-1, a-3, b-1 and b-3 present connectivity to residual clutter which corrupts the underlying geometric structure. Besides, the broken segments in a-2 forbid the primitive detection. In b-2, the fast but deficient Canny filter poorly reflects the inherent edge structure.

After this refinement, it is possible to computationally profit from the discretization grid of the pixels for the expansion while establishing the node links according to the subpixel distances. In this manner, the structure of the radiance edge is correctly embodied by the graph incidence.

#### D. Edge-Graph Extraction

Now, the graph extraction proceeds analogously to the  $A^*$  algorithm in a region growing fashion. The edge-graphs  $G_k$  are extracted by a heuristic recursive expansion. The extraction traversal priority stack is sorted by the subpixel distances  $\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\|$ , see Fig.12-b.

#### E. Edge-Graph Characterization

The edge-graphs  $G_k$  are acyclic linked collections of active subpixel nodes, namely edge-trees. They extend along the radiance edges revealing their distinctive structural composition. In order to properly characterize this composition, a two-stage preprocessing was performed;

- *Pruning*: It improves the representativeness of the linkage by eliminating the superfluous leaf nodes of the trees. This is done by a tree-depth analysis.
- *Outspreading*: It enhances the cohesiveness of the graph linkage with the Euclidean distances of the nodes. The outspreading is performed by a uniform arc-length arrangement  $\chi: \mathbb{R}^2 \mapsto \mathbb{R}^2$  of the  $\hat{\mathbf{x}}_i$  nodes contained within the tree branch paths  $\beta_{u,v}$ , see Fig.12-c,

$$\tilde{\mathbf{x}}_i = \chi(\hat{\mathbf{x}}_i, \beta_{u,v}). \quad (11)$$

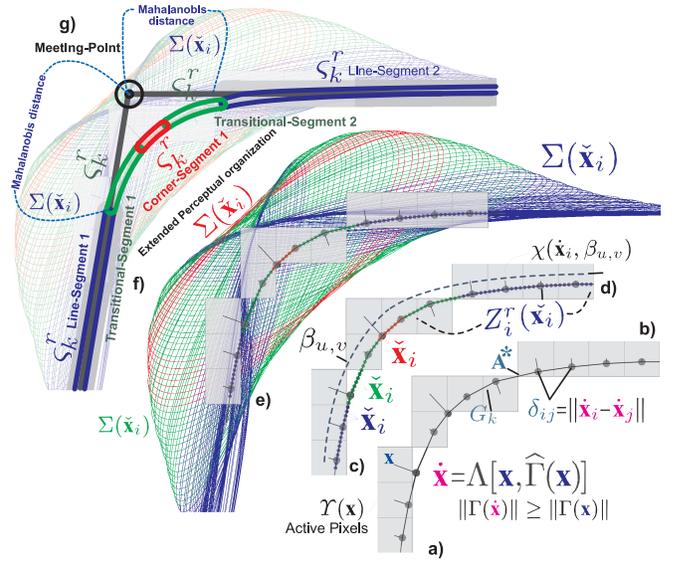


Fig. 12. The eccentricity edge-graph processing stages. a) Subpixel optimization from Eq.10. b) Graph extraction. c) Outspreading from Eq.11. d)  $r$ -Zone node set. e) Eccentricity ellipses from the Eigenvalues and Eigenvectors of the covariance matrix in Eq.13. f) Geometric primitive segmentation. g) Evidence-based grouping transform.

During this arrangement, the branch-paths are treated as parametric curves where the supporting points are the subpixel node locations  $\hat{\mathbf{x}}_i$ . This way, the implicit parametric curve of the edge remains unaffected, while the graph and Euclidean distance are cohesively correlated.

The resulting cohesively combined statistical distribution is the key to characterize the nodes  $\hat{\mathbf{x}}_i$  by the proposed eccentricity  $\xi_i$  of their  $r$ -zone<sup>2</sup>, see Fig.12-e,

$$\xi(\hat{\mathbf{x}}_i) = \kappa \left[ \frac{2}{\pi} \arctan(v \frac{\sigma_1}{\sigma_2}) - \frac{1}{2} \right], \quad (12)$$

where  $\sigma_1$  and  $\sigma_2$  are the Eigenvalues of the  $\hat{\mathbf{x}}_i$ -centered covariance matrix,

$$\Sigma(\hat{\mathbf{x}}_i) = \frac{1}{|Z_i^r|} \sum_{j=1}^{|Z_i^r|} (\hat{\mathbf{x}}_j - \hat{\mathbf{x}}_i)(\hat{\mathbf{x}}_j - \hat{\mathbf{x}}_i)^T. \quad (13)$$

In Eq.12, the saturation factor  $v$  shifts the clipping Eigenratio, whereas the normalization, offset and amplitude factors conveniently shape the eccentricity response function  $\xi$  in order to suitably distinguish the  $k := \{C, T, E\}$  eccentricity types, see Tab.I.

#### F. Edge-Graph Segmentation

Geometric primitives emerge by grouping nodes according to their  $k$ -type. These cliques  $\mathcal{C}_k^r := \{\hat{\mathbf{x}}_i\}$  constitute nodes in a geometric primitive graph  $G_i^r$ , see Fig.12-f.

#### G. Edge-Graph Extended Perceptual Organization

In contrast to [18] and [19], *proximity* and *continuity* among geometric primitives are not detected exclusively based on the length, position, orientation and distance between their end-points. The improved detection incorporates structural likelihood by the *Mahalanobis* distance induced

<sup>2</sup>The node subset  $Z_i^r(\hat{\mathbf{x}}_i) := \{\hat{\mathbf{x}}_j\}$  reachable within a maximal  $r$  graph-distance, see Fig.12-d. Its cardinality is denoted as  $|Z_i^r|$ .

$k$ - Type	$\xi$ - Range	$\xi$ - Level	$\sigma_1 : \sigma_2$	$\tilde{x}_i$ - Ridge Segment	$Z'_i$ - Subtree
<b>E</b> - Elongated	$[\xi_H = 40, \infty)$	High	$\sigma_1 \gg \sigma_2$	Line Segments or Low Curvature	Paths
<b>T</b> - Transitional	$[\xi_L, \xi_H = 40)$	Medium	$\sigma_1 > \sigma_2$	Middle Curvature	Intermediate Zone
<b>C</b> - Compact	$[1, \xi_L = 10)$	Low	$\sigma_1 \approx \sigma_2$	Corners or Junctions	Leaves or Splitters

TABLE I  
THE ECCENTRICITY  $\xi$  CHARACTERIZATION.

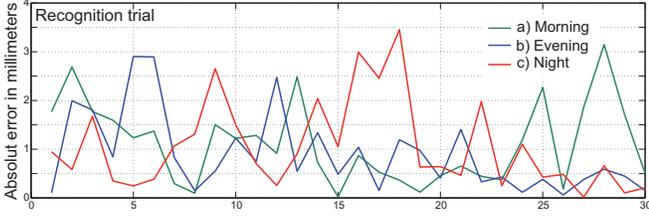


Fig. 13. Precision analysis by absolute error plots. The ground truth is the 116mm length of the physically measured segment. a) The plot shows the precision attained with artificial lighting plus the natural morning illumination incoming from the windows. b) Plot with artificial lighting plus the natural evening illumination. c) Only artificial lighting, see Tab.II.

Set	Mean Length	Std. Deviation	Max-Abs Error	Distance
Morning	116.91	1.07	3.15	1032.62
Evening	116.31	1.18	2.9	1043.38
Night	116.57	1.28	3.45	1064.23

TABLE II

THE PRECISION PERFORMANCE IN DIFFERENT ILLUMINATION SETUPS.

between the edge-node combined statistical distributions  $\Sigma(\tilde{x}_i)$  (from Eq.13) of the periphery of the clique and the intersection point, see Fig.12-g. This morphologic evidence-based grouping transforms the geometric primitives from  $\zeta_k^r$  nodes to consistently matchable geometric percepts.

#### H. Spatial Invariant 3D Features for Matching

The extracted geometric primitives from the left and right HDR images from the calibrated stereo-rig of the humanoid robot are used to calculate their 3D pose [20],[16]. The resulting Euclidean length and angles of the percepts are reasonably pose-invariant. These geometric primitives organized and grouped into a 3D spatial graph representation allows inexact graph matching methods [21] to identify model instances in the field of view. In this paper, the focus is placed on the reliable low-to-middle level processing of the object recognition. Thus, in Sec.IV a simple but effective matching method was used, see Fig.14.

#### IV. EXPERIMENTAL EVALUATION

The presented method for radiometric calibration requires about 5-7 minutes per trial and usually there are 15-25 convergence iterations per color channel for each camera. This process is done only once per camera.

In Fig.15, the HDR images were captured with 8mm focal length lenses with 10 equally exposed images for blending and 6 differently exposed images for fusion. Each capture takes 5-7s using the cameras [13] at 15 fps with 640x480 pixels. The fusion phase takes 250-350ms. The saliency extraction using HDR images for edge-graph requires 1.68s with 2 kernel band-pairs. The graph extraction and eccentricity characterization take 1.46s with a standard deviation of

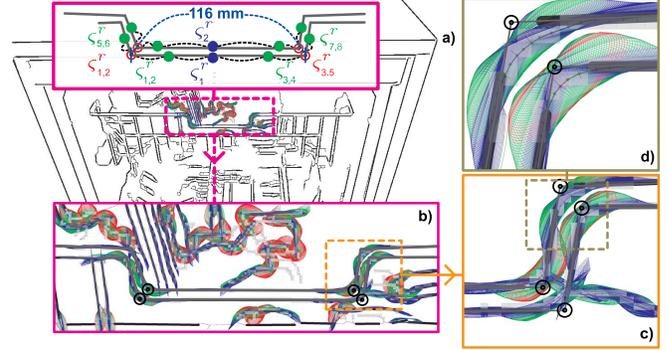


Fig. 14. The recognition of a dishwasher basket is a perception challenge for the complexity of the overlapping structures and the wide intra-scene radiance. a) Scalable vector graphics output of the recognition component. The selected region shows the grasping target region to be recognized. The upper schematic shows the geometric graph to be matched. b) First zoom. c) Second zoom. d) At the third zoom, it is possible to see the eccentricity types in red, green and blue covariance ellipses from Eq.13. The subpixel graph configuration along the curve of the metal wire of the basket shows the simultaneous extraction and geometric primitive segmentation. See Fig.12 for the sequential detailed formation of the edge-graphs.

3.52s depending on the image-content. Finally, the geometric regressions and perceptual organization estimate the line segments and meeting points in 25-50ms, see Fig.12.

In the presented experiments (see Fig.14), the matching method used to recognize the grasping-target region of the dishwasher exploits the stereo Euclidean metric to find a line-segment node with a known length (116mm), i.e., the grasping segment between two incident line-segment nodes.

The complete single threaded execution time of 10-12s can be reduced by at least two thirds. Notice that there is an upper performance boundary by the delay of the camera response to exposure commands.

In the evaluation, there were no false positives. However, the accomplished recognition rate of 93.04% can be improved by using more compelling matching methods like in the geometric hashing approaches [22],[18].

When computing the length of the grasping segment, a maximal absolute error of 3.45mm was obtained and no detrimental performance was measured by lighting variations, see Fig.13 and Tab.II.

#### V. CONCLUSIONS

The main contribution of the approach is to manage the illumination issues by means of the HDR image capturing and the novel processing method. The technical contribution lies in the HDR exposure control module, whereas the theoretical contributions are: i) The continuous reciprocity-consistent radiometric calibration model. ii) The automatic extraction and segmentation of geometric primitives from the radiance edge by the combination of the topological and

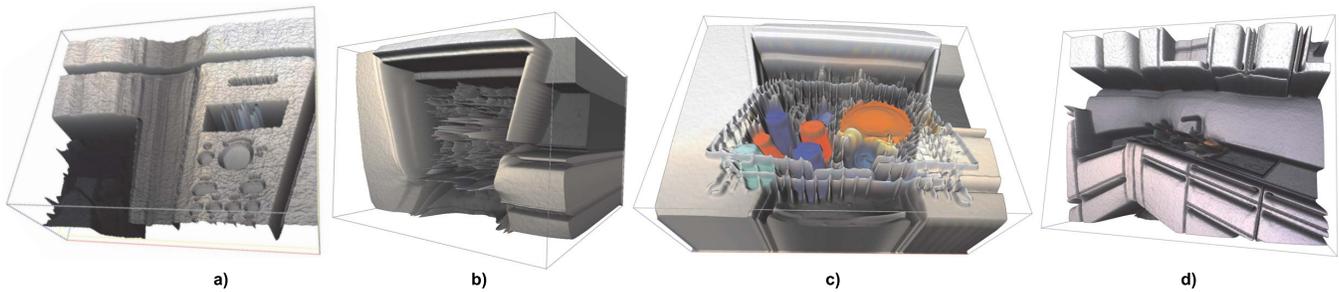


Fig. 15. HDR image acquisition results. a) The microwave metallic surface and LCD display. b) The interior of the empty dishwasher. c) The dishwasher loaded with glittering plastic elements. d) The sink and stove. Notice the saliencies inside the cupboard.

statistical distributions of the subpixel edge nodes. iii) The evidence based perceptual organization includes the novel structural eccentricity edge information in order to reinforce the structural regularity.

The presented methods have no parameters which depend on the image-content, whereas the content-independent parameters discussed here have only well-bounded impact on the performance of the execution. In the experiments, the exhaustive geometric primitives search took at most 15s using a Intel-CPU@1.8 GHz with no code optimization. Technical improvements such as parallelization using a GPU and faster interfaces to HDR-enabled cameras will improve the performance to real-time execution.

Surprisingly, the use of HDR images for model-based object recognition using stereoscopic vision has not been done before in robotic applications with no controlled light conditions and accurate pose estimation.

This approach attains the robust acquisition of visual cues for recognition allowing new algorithms to dependably solve the visual recognition assertions required for autonomous intelligent humanoid robots. Additional material including HDR images and SVG output of the algorithms is available at <http://i61www.ira.uka.de/users/gonzalez>.

## VI. ACKNOWLEDGMENTS

This work was partially conducted within the German Humanoid Research project SFB588 funded by the German Research Foundation (DFG: Deutsche Forschungsgemeinschaft) and the EU GRASP project (IST-FP7-IP-215821) funded by the European Commission.

## REFERENCES

- [1] D. Gonzalez-Aguirre, T. Asfour, E. Bayro-Corrochano, and R. Dillmann, "Model-based visual self-localization using geometry and graphs," in *ICPR*, 2008, pp. 1–5.
- [2] D. Gonzalez-Aguirre, S. Wieland, T. Asfour, and R. Dillmann, "On environmental model-based visual perception for humanoids," in *CIARP*, 2009, pp. 901–909.
- [3] M. Kojima, K. Okada, and M. Inaba, "Manipulation and recognition of objects incorporating joints by a humanoid robot for daily assistive tasks," in *2008. IROS 2008. IEEE/RSJ International Conference on*, 2008, pp. 1564–1569.
- [4] T. Asfour, P. Azad, N. Vahrenkamp, K. Regenstein, A. Bierbaum, K. Welke, J. Schröder, and R. Dillmann, "Toward humanoid manipulation in human-centred environments," *Robot. Auton. Syst.*, vol. 56, no. 1, pp. 54–65, 2008.
- [5] S. Tokutsu, K. Okada, and M. Inaba, "Environment situation reasoning integrating human recognition and life sound recognition using dbn," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009, pp. 744–750.
- [6] K. Okada, M. Kojima, S. Tokutsu, Y. Mori, T. Maki, and M. Inaba, "Task guided attention control and visual verification in tea serving by the daily assistive humanoid HRP2JSK," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, 2008, pp. 1551–1557.
- [7] S. Wieland, D. Gonzalez-Aguirre, T. Asfour, and R. Dillmann, "Combining force and visual feedback for physical interaction tasks in humanoid robots," in *Humanoid Robots, 2009 9th IEEE-RAS International Conference on*, 2009, pp. 439–446.
- [8] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, pp. 465–470, 1996.
- [9] T. Asfour, K. Regenstein, P. Azad, J. Schröder, N. Vahrenkamp, and R. Dillmann, "Armar-iii: An integrated humanoid platform for sensory-motor control," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, 2006, pp. 169–175.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] PTGrey, *Dragonfly Technical Reference Manual*, 5-8 2008.
- [14] 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.
- [15] C. Grigorescu, N. Petkov, and M. Westenberg, "Contour detection based on nonclassical receptive field inhibition," *Image Processing*, vol. 12, no. 7, pp. 729–739, July 2003.
- [16] "The Integrating Vision Toolkit," <http://ivt.sourceforge.net/>.
- [17] R. Keys, "Cubic convolution interpolation for digital image processing," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, no. 6, pp. 1153–1160, dec 1981.
- [18] D. G. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artificial Intelligence*, vol. 31, pp. 355–395, 1987.
- [19] S. Sarkar and K. Boyer, "Perceptual organization in computer vision: a review and a proposal for a classificatory structure," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 23, no. 2, pp. 382–399, 1993.
- [20] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.
- [21] R. Marcondes and P. Larrañaga, "Inexact graph matching for model-based recognition: Evaluation and comparison of optimization algorithms," *Pattern Recognition*, vol. 38, no. 11, pp. 2099–2113, 2005.
- [22] H. Wolfson and I. Rigoutsos, "Geometric hashing: An overview," *CALSE*, vol. 4, no. 4, pp. 10–21, October 1997.