Illumination-invariant Image-based Environment Representations for Cognitive Mobile Robots Using Intrinsic Images

Werner Maier[1], Fengqing Bao[1], Eckehard Steinbach[1], Elmar Mair[2] and Darius Burschka[2]

[1] Institute of Media Technology, Technische Universität München
Email: {werner.maier,eckehard.steinbach}@tum.de, fengqing.bao@mytum.de

[2] Robotics and Embedded Systems Group, Technische Universität München
Email: {maire,burschka}@in.tum.de

1 Introduction

Image-based scene representations enable a cognitive robot to make a realistic prediction of its environment. Hence, it is able to rapidly detect changes by comparing a virtual image generated from previously acquired reference images and its current observation. This facilitates surprise detection and attentional control to novel events ([1]). However, illumination effects can impair attentional control if the robot does not take them into account. To address this issue, we present in this paper an approach for the acquisition of illumination-invariant scene representations. Using multiple spatial image sequences which are captured under varying illumination conditions we render virtual images from each sequence at a series of dense viewpoints and recover intrinsic images of the scene. Experimental results show high-quality images which are free of illumination effects.

2 Acquisition of Image-based Representations

We use a Pioneer 3-DX robot, which is equipped with a stereo camera, as a mobile platform for model acquisition. When an image sequence is captured, the poses of the camera views w.r.t the coordinate frame of the first camera are estimated using a robust version of the visual GPS algorithm (RVGPS) ([4]). For the registration of the multiple image sequences w.r.t each other we insert the camera views from the different sequences into a reference sequence I_{ref} between two similar images. The RVGPS algorithm then provides the poses of the camera views w.r.t the coordinate frame of the first view in I_{ref}. For each reference view in a given image sequence we estimate a per-pixel depth map. The algorithm for depth estimation consists of two steps. Firstly, a cost volume is computed by a plane-sweep. Secondly, the cost volume is passed to a belief propagation algorithm which provides smooth and dense depth maps. In a refinement step we validate the value of each pixel in a depth map with corresponding depth estimates from depth maps of neighboring views. From each refined depth map a mesh is generated.

3 Illumination-invariant Image-based Representations

In order to obtain an image-based environment representation which is free of illumination effects, our idea is to render virtual images from each acquired image sequence $I_m, m = 1, ...M$ at identical viewpoints around the scene (see Figure 1). This provides virtual image sequences which are denoted by $I_{\mathrm{v},m}$. We apply a GPU-based view synthesis technique which, for each viewpoint, selects eight reference views from an image sequence and warps them into the virtual view. In a blending pass the virtual image is computed by averaging the warped colors at each pixel. Following the idea of intrinsic images, each virtual image can be decomposed in a reflectance image and an illumination image in log domain. The reflectance is assumed to be static in all virtual images at a given viewpoint. Similar to [2], we compute a horizontal and a vertical gradient image from each virtual image at a given viewpoint. By computing the median of the horizontal and vertical gradients at each pixel, the gradients which stem from the borders of shadows vanish and

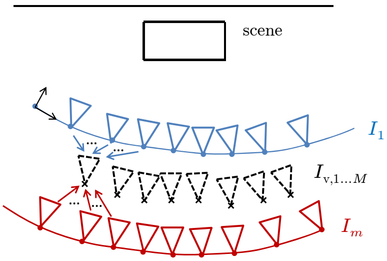


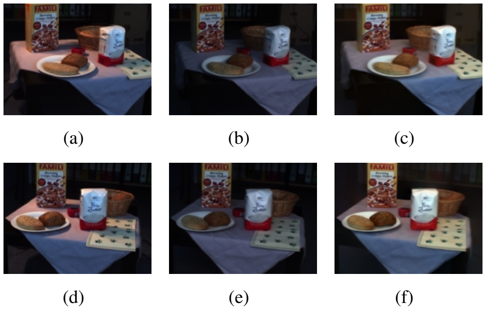
Figure 1: At a given viewpoint (crosses), a virtual image is rendered from each acquired image sequence.

only the static gradients which come from the texture and the borders of the scene objects remain. The reflectance can then be recovered from the medians of the horizontal and vertical gradients.

4 Experimental Results

In order to test our method, we acquired 9 image sequences which were captured under daylight and artificial light. By moving two lamps mounted on a tripod around the scene, the direction of the artificial light was varied between the sequences. Figures 2(a), 2(b), 2(d) and 2(e) show virtual images rendered at different viewpoints from two image sequences. These virtual images are photorealistic and hardly show artifacts due to erroneous poses of the reference images or depth maps. The reflectance images recovered from all 9 virtual images at the two viewpoints are shown in Figures 2(c) to 2(f). Obviously, the shadows in the virtual images cast by the objects on the table are largely gone in the reflectance images. The only shadow that remains is the one beneath the plate. This part of the table is very hard to illuminate since the plate is very close to the table. The reflectance images still show a high realism and do not contain any artifacts due to shadow removal.

5 Conclusion and Future Work



(a) (b) (c)

(d) (e) (f)

Figure 2: (a),(b),(d) and (e): Virtual images rendered at two viewpoints from two different image sequences. (c) and (f): Reflectance images recovered from all virtual images at these viewpoints.

In this paper, we proposed a novel method to render image-based scene representations with largely removed illumination effects. Using multiple image sequences acquired by a moving camera under varying lighting conditions, a sequence of reflectance images is computed. Experimental results show high-quality and shadow-free images. The next step in future work will be to use these illumination-invariant representations for robust attentional control. With a statistical model trained from the illumination images the robot is supposed to be able to distinguish between illumination effects and reflectance changes in a new observation.

6 Acknowledgement

This work is supported within the research cluster *Cognition for Technical Systems – CoTeSys.*

References

[1] W. Maier, E. Mair, D. Burschka and E. Steinbach. Visual Homing and Surprise Detection Using Image-Based Environment Representations. In *Int. Conf. on Robotics and Automation,* 2009.

[2] Y. Weiss. Deriving Intrinsic Images from Image Sequences. In *Int. Conf. on Computer Vision,* 68–75, 2001.

[3] C. Tomasi and T. Kanade. Detection and Tracking of Point Features. In *CMU Technical Report CMU-CS-91-132,* 1991.

[4] E. Mair, K. H. Strobl, M. Suppa and D. Burschka. Efficient Camera-Based Pose Estimation for Real-Time Applications. In *Int. Conf. on Intelligent Robots and Systems,* 2009.