Illumination-Invariant Image-Based Novelty Detection in a Cognitive Mobile Robot's Environment

Werner Maier, Fengqing Bao, Elmar Mair, Eckehard Steinbach and Darius Burschka

Abstract— **Image-based scene representations enable a mobile robot to make a realistic prediction of its environment. Hence, it is able to rapidly detect changes in its surroundings by comparing a virtual image generated from previously acquired reference images and its current observation. This facilitates attentional control to novel events. 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 the robot computes an illumination-invariant image-based environment model. With this representation and statistical models about the illumination behavior, the robot is able to robustly detect texture changes in its environment under different lighting. Experimental results show high-quality images which are free of illumination effects as well as more robust novelty detection compared to state-of-the-art methods.**

I. Introduction

Traditionally, the environment of a mobile robot is represented by geometry data which is used for collision detection during navigation or during a grasping process. A sparse set of texture images stored with the model allows for the visualization of the surroundings and hence provides information about the appearance of the scene. However, this modeling approach can get computationally expensive for translucent or filigree objects – due to the high level of detail and the refraction of light. Image-based rendering techniques, in turn, provide photorealistic virtual images while maintaining a low complexity during the rendering process.

The enviroment of a mobile robot is usually not static. Hence, the robot has to be notified about changes such as the removal or sudden appearance of objects. In [1], we presented a technique for the computation of per-pixel surprise maps using the robot's current observation and a set of reference views from its internal image-based environment representation. Surprise is a very important means in cognitive systems to direct the attention to unexpected events [2] and to segment objects from the known familiar part of the environment. However, if surprise detection is based on the raw intensity values provided by the robot's camera, robust attentional control to objects cannot be ensured if the illumination changes. Many robots, like service robots, act in a closed environment and return to the same place over and over again. Hence, the robot can collect several observations over a longer period of time and under different lighting conditions and infer the intrinsic colors (reflectance) of the objects in the scene.

In this work, we present a method for the computation of illumination-invariant image-based models from multiple spatial image sequences acquired of the same scene under varying illumination. The reflectance of the environment is recovered at a dense series of viewpoints from virtual images which are rendered from the captured sequences. Using the reflectance images at these viewpoints together with local geometry and camera pose information, a robot is able to predict the reflectance of the scene in a continuous viewpoint space. Statistical knowledge gathered from illumination images enables the robot to reliably distinguish novel objects from illumination changes.

This paper is structured as follows. In Section II we outline related work. Next, in Section III, we briefly describe the visual localization technique used in this work and the registration of several acquired image sequences with respect to a common coordinate frame. Section IV presents our framework for the generation and visualization of image-based environment models. Section V presents the main contributions of this work, namely the computation of an illumination-invariant image-based environment representation for cognitive mobile robots as well as a method for novelty detection under varying lighting conditions. After showing experimental results in Section VI, we conclude this paper in Section VII.

II. Related Work

In [6], the huge body of work in the field of image-based rendering is presented. Different approaches are classified in terms of the amount of geometry information they incorporate.

Several methods for the computation of intrinsic images have been presented. In [4], a reflectance and an illumination image is recovered by entropy minimization. While this technique uses only one intensity image, several constraints are imposed on the camera sensor and on the lighting which are not always fulfilled in a real-world environment. The method in [5] also recovers the reflectance and the illumination of the scene from only one image. However, classifiers have to be trained which distinguish between gradients in the intensity image due to illumination effects and gradients due to reflectance changes. A computationally cheap approach was presented in [3] where intrinsic images are recovered

W. Maier, F. Bao and E. Steinbach are with the Institute for Media Technology, Technische Universität München, 80333 München, Germany {werner.maier, eckehard.steinbach}@tum.de, fengqing.bao@mytum.de

E. Mair and D. Burschka are with the Institute for Robotics and Embedded Systems, Technische Universität München, 85748 Garching, Germany {maire,burschka}@in.tum.de

from multiple intensity images which show the scene under different illumination. While this approach provides high-quality reflectance images, the viewpoint of the camera has to be static.

A survey on various image change detection algorithms is given in [9]. Recently, an approach for detecting object and motion changes under varying illumination was presented in [10]. It is based on the assumption that the gradient structure within a pixel block does not change with illumination. Hence, the normalized correlation between corresponding pixel blocks in two images is used as a measure for change. Another popular technique is the transformation of an image to an illumination-invariant color space. In [11], the Spherical Coordinate Transform ([12]) (SCT) is a preprocessing step for color learning in a mobile robot.

One of the main contributions of this paper is an approach which provides a reflectance representation of the environment not only at one static viewpoint but in a given 3D viewpoint space. Another core contribution is a method for the robust detection of novel objects under varying lighting conditions, which is based on statistical models of the illumination behavior. These models are trained from illumination images which separate the lighting effects from the intrinsic appearance of the scene.

III. Visual Localization and Registration of Multiple Image Sequences

The visual localization algorithm used in our work extracts salient features from stereo images according to [15]. The distance to the corresponding scene points is initially estimated by the triangulation of sub-pixel accurate stereo matches. This 3D structure is retained unchanged for the whole run. During navigation, the features are tracked with the Kanade-Lucas-Tomasi (KLT) tracker [13] in the images of the left camera. Based on a robust variant of the visual GPS algorithm (RVGPS), the camera motion is computed w.r.t the local camera coordinate frame at the starting position. An M-estimator weights drifting features and rejects outliers. More details on the pose estimation algorithm can be found in [14].

The acquisition of multiple image sequences can happen over a longer period of time. However, the visual localization algorithm looses the tracked features after one image sequence. Hence, the localization is initialized for each new image sequence and the poses of the images in a particular sequence are determined with respect to its first image. However, if we want to render a virtual image at the same viewpoint in space from different sequences, the relationship between the image sequences is required.

Let us assume throughout this work that a robot acquires M image sequences. The images in a given sequence are indexed by j. We use the coordinate frame of the first sequence I_1 as the common coordinate frame of all sequences. In order to register a particular sequence $I_m, \quad m = 2, ..., M$ with respect to this coordinate frame, a sparse set of N_S support views is taken from it. For illustration, three support views are represented by the camera viewing frusta in Figure 1.

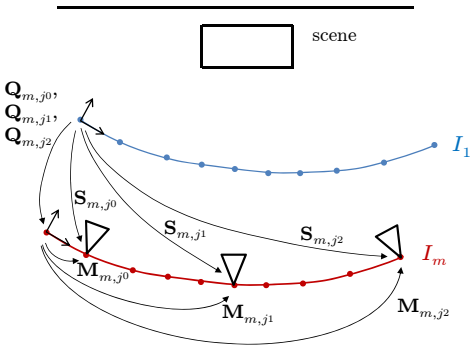


Fig. 1. All acquired image sequences are registered in a common coordinate frame. Using several support views in an image sequence I_m, the transformation between the sequence and the common coordinate frame is determined.

The support views are inserted into the image sequence I_1 between similar images which results in an augmented image sequence I'_1. The transformation matrices of the support views $\mathbf{S}_{m,j_n}$ $(n = 0, \ldots, N_\mathrm{S})$ w.r.t the common coordinate frame are determined by applying the visual localization technique in [14] to I'_1.

Using the transformation matrices of the support views and their poses $\mathbf{M}_{m,j_n}$ $(n = 0, \ldots, N_\mathrm{S})$ w.r.t the coordinate frame of I_m, the transformation between the two coordinate frames of the sequences is calculated, respectively, for each support view by

$$\mathbf{Q}_{m,j_n} = \mathbf{S}_{m,j_n} \cdot \mathbf{M}^{-1}_{m,j_n} \quad n = 0, \ldots, N_\mathrm{S} - 1 \tag{1}$$

Although describing the same relationship, these matrices can be different from each other due to the error propagation in the visual localization over the image sequence I_m. Using $\mathbf{Q}_{m,j_n}$, the transformation $\mathbf{Q}_{m,j}$ between the two coordinate frames is interpolated over the whole image sequence I_m by a Lagrange polynom. In order to fulfill the orthonormality constraint on the rotation matrices, the column vectors of the rotation matrix in $\mathbf{Q}_{m,j}$ are normalized to unit length. The poses $\mathbf{S}_{m,j}$ of the images in I_m w.r.t the common coordinate frame are then calculated by

$$\mathbf{S}_{m,j} = \mathbf{Q}_{m,j} \cdot \mathbf{M}_{m,j} \tag{2}$$

where $\mathbf{M}_{m,j}$ is the pose of a given image in I_m w.r.t. its first image.

IV. Image-based Environment Modeling and Rendering

For image-based modeling, only the left images of the captured stereo pairs are stored. In the following, these images will be called reference images. Similar to [7] and [8] we use local geometry information instead of a global geometric model of the environment for the interpolation of virtual views. The recovery of the view-dependent geometry is done off-line while the computation of virtual images is performed on-line by the graphics processing unit.

In order to render novel virtual views from captured image data, correspondences between the pixels in the images have

to be established. Hence, for each stored image, a per-pixel depth map is calculated. To this end, we define multiple planes which lie parallel to the image plane of the camera and quantize its viewing frustum within a given distance range. Similar to the method in [16], a cost volume is computed by a plane-sweep. In order to handle occlusions, we use two other images from the sequence for the computation of the depth map. These images lie at a certain baseline from the reference view. For a given layer of the cost volume the matching cost is determined by the minimum absolute difference between the luminance values of the reference image and the luminance of the other warped images. In order to get smooth disparity maps in non-textured areas, the loopy belief propagation algorithm in [17] is applied. In a geometry validation step the values of a depth map are corrected using the redundancy in neighboring depth maps. The meshes reconstructed from the validated depth maps are simplified in order to reduce memory consumption.

For the interpolation of novel virtual images of the environment all reference cameras are ranked in terms of their distances to the current position of the virtual camera and the angular deviation of their viewing directions. The seven closest reference cameras are selected for view synthesis and their data (meshes, textures) is loaded into graphics memory. The view interpolation is done in several passes. First, the reference images are warped on the image plane of the virtual camera using the view-dependent meshes and are stored as textures in the off-screen memory of the graphics hardware. These textures are the input to the pixel shader program which finally calculates the virtual image.

V. Illumination-Invariant Image-Based Models for Novelty Detection

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 at identical viewpoints around the scene. This provides virtual image sequences which are denoted by $I_{\mathrm{v},m}$ in Figure 2. The index $m = 1, \ldots M$ here indicates the image sequence which provides the eight reference images for view interpolation. The reflectance at a given viewpoint is then recovered using the virtual images rendered at this position.

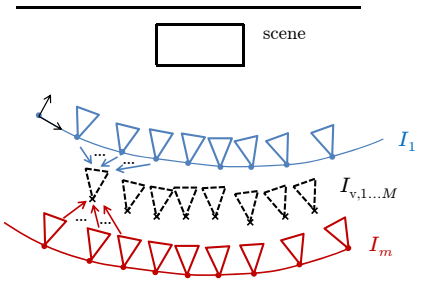


Fig. 2. The crosses with the dashed viewing frusta illustrate the interpolation of virtual images at a series of viewpoints. At a given viewpoint, a virtual image is rendered from each acquired image sequence, respectively.

For the recovery of a reflectance image from several intensity images taken under different illumination, an efficient method was proposed in [3] which we also use here. This technique is briefly presented in the following.

An intensity image captured by a camera is the product of a reflectance image and an illumination image. We also use this model here for the rendered images. In the logarithmic domain, a virtual image rendered from eight real images in sequence I_m is thus given by

$$i_m = r_m + l_m \tag{3}$$

where r_m is the logarithmic reflectance and l_m the logarithmic illumination image. The convolution of the virtual image with the filters $f^{\mathrm{h}} = [1, -1]$ and $f^{\mathrm{v}} = [1; -1]$ provides the horizontal and vertical gradients of the image, respectively.

$$i_m^{\mathrm{h}} = f^{\mathrm{h}} * i_m \tag{4}$$
$$i_m^{\mathrm{v}} = f^{\mathrm{v}} * i_m \tag{5}$$

Although not explicitly expressed in our notation, the computation of the gradients is done in RGB domain, separately for each color channel. Since illumination gradients are sparse and since it is assumed that the reflectance at a viewpoint remains constant over all image sequences, the reflectance image in gradient domain can be recovered by

$$\hat{r}^{\mathrm{h}} = \underset{m=1,\ldots,M}{\mathrm{median}} \left(i_m^{\mathrm{h}}\right) \tag{6}$$
$$\hat{r}^{\mathrm{v}} = \underset{m=1,\ldots,M}{\mathrm{median}} \left(i_m^{\mathrm{v}}\right) \tag{7}$$

As the convolution in (4) and (5) is a linear operation, we get the following overconstrained system of equations for the recovered reflectance gradients.

$$\hat{r}^{\mathrm{h}} = f^{\mathrm{h}} * \hat{r} \tag{8}$$
$$\hat{r}^{\mathrm{v}} = f^{\mathrm{v}} * \hat{r} \tag{9}$$

The eqs. (8) and (9) are solved for the reflectance $\hat{r}$ using pseudoinverse filtering.

$$\hat{r} = \left(\hat{r}^{\mathrm{h}} * \tilde{f}^{\mathrm{h}} + \hat{r}^{\mathrm{v}} * \tilde{f}^{\mathrm{v}}\right) * g \tag{10}$$

The filters $\tilde{f}^{\mathrm{h}}$ and $\tilde{f}^{\mathrm{v}}$ are reversed versions of the respective gradient filters in (8) and (9). The filter g only depends on the gradient filters and is chosen such that it fulfills the equation

$$\left(f^{\mathrm{h}} * \tilde{f}^{\mathrm{h}} + f^{\mathrm{v}} * \tilde{f}^{\mathrm{v}}\right) * g = \delta \tag{11}$$

where δ is the Dirac impulse.

Using the reflectance of the scene at a given viewpoint and the rendered virtual images $i_m, m = 1 \ldots M$, a sequence of images l_m can be recovered by (3) which only contain information about the lighting conditions in i_m. An example for an illumination image is shown in Figure 3(c), which is amplified for visualization. It appears greyish and with little color saturation while the corresponding reflectance image in Figure 3(b) gives information about the intrinsic colors of the objects which are ideally independent of the illumination. In the linear domain the illumination image indicates for each color channel the scaling factor for the intrinsic colors in

Figure 3(b) which is necessary to produce the intensities of the pixels in the camera image in Figure 3(a). In shadow regions this factor is very small while in the rest of the image it is close to one or larger if the light intensity is very strong.

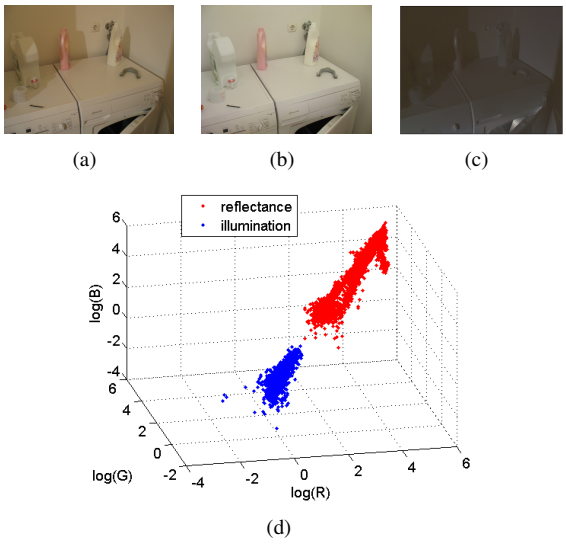


Fig. 3. (a) Camera image. (b) Reflectance image. (c) Illumination image. (d) Reflectance and illumination image plotted in logarithmic RGB space.

Thus, if we take a look at the values of an illumination image in the logarithmic domain, we see that they cluster around zero (see Figure 3(d)). In order to describe this behavior in a statistical way, we use a multivariate Gaussian model. For each viewpoint in our image-based representation, we learn the parameters of this model by Maximum Likelihood (ML) estimation, using the pixel values of several illumination images at the given viewpoint as a set of training samples. Consequently, we obtain the mean vector μ_{C} and the covariance matrix $\mathbf{\Sigma}_{\mathrm{C}}$ as follows

$$\mu_{\mathrm{C}} = \frac{1}{N_{\mathrm{p}}} \sum_{n=1}^{N_{\mathrm{p}}} \mathbf{x}_n \tag{12}$$

$$\mathbf{\Sigma}_{\mathrm{C}} = \frac{1}{N_{\mathrm{p}} - 1} \sum_{n=1}^{N_{\mathrm{p}}} (\mathbf{x}_n - \mu_{\mathrm{C}}) (\mathbf{x}_n - \mu_{\mathrm{C}})^{\mathrm{T}} \tag{13}$$

Here, $\mathbf{x}_n$ denotes a column vector which contains the RGB values of a pixel in an illumination image in logarithmic domain. The total number of training pixels taken from the set of illumination images is N_{p}. The mean vector and the covariance matrix are stored together with the reflectance image, the depth map and the pose information at each viewpoint so that the illumination-invariant image-based representation is enriched by statistical knowledge about the view-dependent illumination characteristics in the environment.

When a mobile robot returns to a known part of the environment it first has to register the currently acquired camera views with respect to its internal representation. A virtual reflectance image is predicted from the illumination-invariant environment model at the estimated current pose of the robot's camera. Using this virtual image, the robot is able to compute an illumination image for the new observation by (3). For each pixel $\mathbf{p}_{u,v}$ the robot evaluates the squared Mahalanobis distance

$$\Delta^2_{u,v} = (\mathbf{p}_{u,v} - \mu_{\mathrm{C,ref}})^{\mathrm{T}} \mathbf{\Sigma}^{-1}_{\mathrm{C,ref}} (\mathbf{p}_{u,v} - \mu_{\mathrm{C,ref}}) \tag{14}$$

where u and v denote the column and the row index of the illumination image, respectively. The Gaussian mean $\mu_{\mathrm{C,ref}}$ and the covariance matrix $\mathbf{\Sigma}_{\mathrm{C,ref}}$ are chosen from one of the reference views used for the interpolation of the virtual reflectance image. If the Mahalanobis distance of a given pixel is very large, the RGB triple lies far from the region where the values of the training set of illumination images cluster. Thus, it is very likely that this pixel indicates a reflectance change due to a newly added or removed object in the environment.

VI. Experimental Results

In order to test our methods, we acquired 9 image sequences I_m, $m = 1, ...9$ with a Pioneer 3-DX robot. The robot was equipped with two cameras with a baseline of 9 cm. Each sequence consisted of 100 images. The robot was controlled to follow a quarter-circle with the stereo camera looking towards its center. Due to the inaccurate internal odometry of the robot, however, the trajectories were not perfect quarter-circles. Furthermore, the robot was manually steered to the starting point of the trajectories. Consequently, the trajectories were similar but never completely identical.

The first image sequence I_1 was captured under outdoor daylight which fell through a window behind the robot. Next, we acquired two sequences I_2 and I_3 under indoor illumination. Here we used the lamps mounted on the ceiling of the laboratory which provided white light. For the remaining 6 image sequences I_4 to I_9, the scene was illuminated by daylight and by the light of two lamps which were mounted on a tripod and which were placed at various positions in the laboratory. The lamps provided yellow-white light so that not only the positions but also the spectra of the light sources varied between the different runs.

Figures 4(a), 4(b), 4(d) and 4(e) show four virtual images rendered at different viewpoints from the image sequences I_4 and I_5. These virtual images are photorealistic and hardly show up 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 4(c) and 4(f). Obviously, the shadows in the virtual images cast by the objects on the table are largely gone in the reflectance images.

In order to evaluate our method for illumination-invariant novelty detection we acquired another image sequence I_{ob}. We positioned the two lamps in a way that the illumination was different from all the runs before. Furthermore, we removed one roll from the plate. One image, which is depicted in Figure 5(a), was chosen as the robot's observation for our experiments. The corresponding virtual reflectance

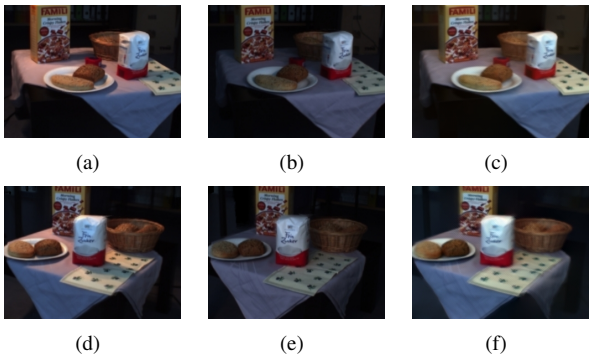


Fig. 4. (a),(d): Virtual images rendered at different viewpoints from images acquired under dominant artificial illumination. (b),(e): Virtual images rendered at the same viewpoints from images taken under dominant daylight illumination. (c),(f): Reflectance images recovered from all virtual images at these viewpoints. The illumination effects which are visible in (a),(b),(d) and (e) are largely removed.

image rendered at the observation's pose is shown in Figure 5(b).

Fig. 5. Table scene: (a) Observation which was taken after one roll had been removed from the plate. (b) Reflectance of the scene computed from the environment model. (c) Reference image for NGC and SCT.

The squared Mahalanobis distance computed by our method for each pixel is shown in Figure 6(a). It clearly indicates a region of high novelty around the missing roll (values between 60 and 120) while the shadow regions have a very low Mahalanobis distance (up to 20). Elevated values along the edges of the objects and in the background are due to a slight inaccuracy in the estimation of the observation's pose. We compare change detection methods based on Normalised Gradient Correlation (NGC) ([10]) and Spherical Coordinate Transform (SCT) ([11], [12]) to ours whose results are shown in Figures 6(b) and 6(c). In SCT, a point in cartesian RGB space is represented in spherical coordinates. One coordinate is the magnitude of the vector while the others represent the azimuth and elevation angles. The distance d in our experiments is the sum of squared differences between the azimuth and elevation values at a given pixel in two images. For both NGC and SCT we chose a virtual image rendered from the image sequence I_4 as a reference image (see Figure 5(c)). High correlation coefficients ρ obtained by NGC indicate blocks with no changes while a change is likely if the coefficient is low (see Section II). For coherent visualization we chose the measure $1-\rho$. As proposed in [10], we computed and combined the gradient correlation coefficients from three resolution layers. Compared to our method, NGC is much more sensitive to pose inaccuracies and clearly indicates the object edges as regions of high novelty since there the gradient structure between blocks from the observation and from the reference image is different. Hence, this method is less suitable for mobile robot applications since slight pose inaccuracies are inevitable. The SCT method reliably detects the missing roll but also shows up comparably high chromaticity differences d in the rest of the change map.

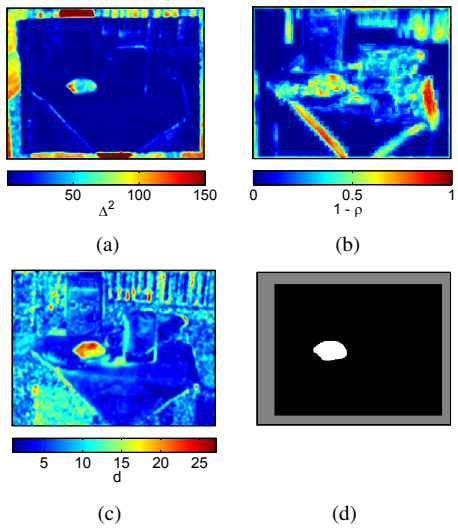


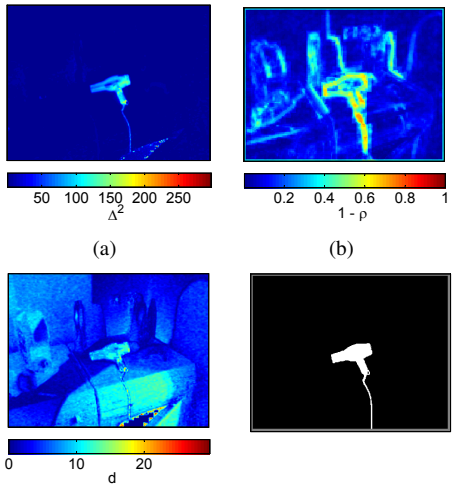
Fig. 6. (a) Squared Mahalanobis distance computed by our method, (b) Normalized Gradient Correlation $(1-\rho)$, (c) Spherical Coordinate Transform, (d) Ground truth.

We evaluated our method for novelty detection for another scene which is shown in Figure 7. In contrast to the table scene all images of the washing machine scene were taken at a static viewpoint. We again compare the performance of our method with respect to the two reference approaches NGC and SCT. Figure 8(a) shows the novelty map obtained from our method. Again, it provides high Mahalanobis distance values in the region of the observation image that shows the hair dryer that is added to the scene. Remarkably, our algorithm is able to distinguish between a black object and a shadow which is difficult for the other two methods. The change from the white color of the washing machine to the deep black of the hair dryer leads to an attenuation of the intensity that is not typical for a shadow. That is why it is classified as a reflectance change in our scheme. The result of the NGC shows that the edges of the shadows produce high values in the map in Figure 8(b). Due to the faint chromaticity change the values in the change map in Figure 8(c) in the region of the hair dryer are not higher than at many other pixels.

For a quantitive comparison of the three methods, we computed the corresponding receiver operating characteristic (ROC) curve. ROC curves are a common means to analyze the reliability of change detection algorithms. For the table scene we evaluated the false positive rate (FPR) vs. the true positive rate (TPR) using the ground truth in Figure 6(d), which we determined manually. The gray regions in the ground truth map are excluded from evaluation since

(a) (b) (c)

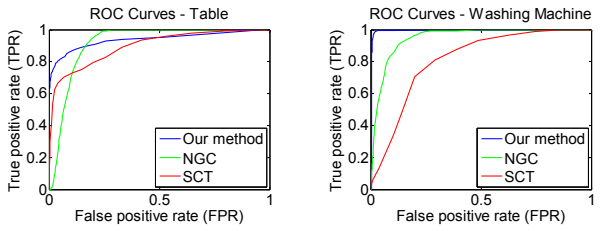
Fig. 7. Washing machine scene: (a) Observation which shows the hair dryer as a newly added object. (b) Reflectance of the scene. (c) Reference image for NGC and SCT.



(a) (b) (c) (d)

Fig. 8. (a) Squared Mahalanobis distance computed by our method, (b) Normalized Gradient Correlation $(1 - \rho)$, (c) Spherical Coordinate Transform, (d) Ground truth.

there the virtual reflectance image does not provide any information. In Figure 9(a) the ROC curves are obtained by varying a threshold between the minimum and the maximum value of the maps in Figures 6(a) to 6(c). The ROC curve for our method shows higher TPRs at low FPRs than the other methods. Analyzing the ROC curves in Figure 9(b) that we get for the washing machine scene we see that the one for our method almost shows ideal behavior while the perfomance of the other methods is clearly worse.



(a) (b)

Fig. 9. ROC curves for the table scene (a) and the washing machine data set (b). They show the performance of the three approaches in terms of the true positive rate vs. the false positive rate.

VII. Conclusion

In this paper, we presented a method for the acquisition of illumination-invariant image-based representations for mobile robots. Virtual reflectance images predicted from this representation and statistical knowledge of the illumination behavior in the scene allow for a robust detection of object changes under varying lighting conditions. In experimental results our method shows better detection performance at low false positive rates compared to other state-of-the-art techniques. Our future work will focus on the extension of the current illumination models for specular surfaces and on novelty detection in the presence of varying spectra of the light sources.

VIII. Acknowledgement

This work is supported in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*, see also `www.cotesys.org`.

References

[1] W. Maier, E. Mair, D. Burschka and E. Steinbach, "Visual Homing and Surprise Detection Using Image-Based Environment Representations", *in IEEE Int. Conf. on Robotics and Automation*, 2009, pp. 807–812.

[2] L. Itti and P. Baldi, "Bayesian Surprise Attracts Human Attention", *in Vision Research*, vol. 49, no. 10, 2009, pp. 1295–1306.

[3] Y. Weiss, "Deriving Intrinsic Images from Image Sequences", *in Int. Conf. on Computer Vision*, 2001, pp. 68–75.

[4] G. D. Finlayson, M. S. Drew and C. Lu, "Intrinsic Images by Entropy Minimization", *in 8th Europ. Conf. on Computer Vision*, 2004, pp. 582–595.

[5] M. F. Tappen, W. T. Freeman and E. H. Adelson, "Recovering Intrinsic Images from a Single Image", *in IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 9, 2005, pp. 1459–1472.

[6] H.-Y. Shum, S.-C. Chan and S. B. Kang, *Image-Based Rendering*, Springer, 2007.

[7] C. L. Zitnick et al., "High-quality Video View Interpolation Using a Layered Representation", *in ACM Trans. on Graphics*, vol. 23, no. 3, 2004, pp. 600–608.

[8] J.-F. Evers-Senne and R. Koch, "Image-based Interactive Rendering with View Dependent Geometry", *in Computer Graphics Forum*, vol. 22, no. 3, 2003, pp. 573–582.

[9] R. J. Radke et al., "Image Change Detection Algorithms: A Systematic Survey", *in IEEE Trans. on Image Processing*, vol. 14, no. 3, 2005, pp. 294–307.

[10] R. O'Callaghan and T. Haga, "Robust Change-Detection by Normalised Gradient-Correlation", *in IEEE Conf. on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[11] M. Sridharan and P. Stone, "Structure-Based Color Learning on a Mobile Robot Under Changing Illumination", *in Autonomous Robots Journal*, vol. 23, no. 3, 2007, pp. 161–182.

[12] B. W. Minten et al., "Low-Order-Complexity Vision-Based Docking", *in IEEE Trans. on Robotics and Automation*, vol. 17, no. 6, 2001, pp. 922–930.

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

[14] 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, pp. 2696–2703.

[15] J. Shi and C. Tomasi, "Good Features to Track", *in Int. Conf. on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

[16] R. T. Collins, "A Space-Sweep Approach to True Multi-Image Matching", *in Int. Conf. on Computer Vision and Pattern Recognition*, 1996, pp. 358–363.

[17] P. Felsenszwalb and D. Huttenlocher, "Efficient Belief Propagation for Early Vision", *in Int. Conf. on Computer Vision and Pattern Recognition*, 2004, pp. I261–I268.