IMAGE-BASED OBJECT DETECTION UNDER VARYING ILLUMINATION IN ENVIRONMENTS WITH SPECULAR SURFACES

Werner Maier, Michael Eschey, and Eckehard Steinbach

Technische Universität München
Institute for Media Technology
80333 Munich, Germany

ABSTRACT

Image-based environment representations capture the appearance of the surroundings of a mobile robot and are useful for the detection of novelty. However, image-based novelty detection can be impaired by illumination effects. In this paper we present an approach for the image-based detection of novel objects in a scene under varying lighting conditions and in the presence of objects with specular surfaces. The computation of an illumination-invariant image-based environment representation allows for the extraction of the shading of the environment from camera images. Using statistical models inferred from the luminance and the saturation component of the shading images, specularities and shadows are detected and suppressed in the process of novelty detection. Experimental results show that the proposed method outperforms two recently presented reference approaches for illumination-invariant change detection in images.

Index Terms— novelty detection, image-based rendering, illumination modeling

1. INTRODUCTION

Image-based scene representations [1] allow for the rapid synthesis of photorealistic virtual views of real-world environments with complex illumination and transparent or specular objects. In [2] an approach is presented for the acquisition of an unstructured image-based model with a hand-held camera. Recently, an unstructured image-based scene representation has also been proposed as a component of the internal environment model of mobile cognitive robots [3]. While a geometric representation of the objects is indispensable for a robot in manipulation tasks like grasping, an image-based model is useful for the detection of novelty in the robot’s environment. The probabilistic appearance representation in [4] is inspired by image-based representations but models the robot’s expectation of the appearance and its uncertainty in terms of probability distributions. This facilitates the computation of surprise maps and the direction of the robot’s attention to surprising events in new camera images. The representation in [4] is inferred from low-level cues like the luminance and chrominance of a pixel in a captured image. These cues are subject to the current illumination in the environment. If the lighting conditions change, this leads to elevated surprise levels all across the image. Hence, the robot needs a mechanism which enables it to distinguish relevant novelty from irrelevant novelty. New objects in the environment which are relevant for the robot’s current tasks are of greater importance than changes in illumination. The detection of environmental changes in images taken under varying illumination conditions is addressed in [5]. It is assumed that illumination changes do not diminish the correlation between the gradients of two corresponding image blocks. However, the approach does not propose a method which copes with shadow borders or specularities. In [6] a method was presented for the removal of specularities in facial images with the goal of improved skin and face detection. The concept of intrinsic images was introduced to represent the illumination and the reflectance of objects in a scene in separate images. A method for the computation of intrinsic images from multiple images which show a scene under different illumination was presented in [7]. This work has inspired our approach in [8] where we present an approach for the illumination-invariant detection of novel objects in a mobile robot’s environment. By first computing an image-based environment representation which is free of time-varying illumination effects, the shading of the environment can be extracted from the camera images. A multivariate Gaussian model is inferred which describes the intensity variation of the shading. Novel objects are identified as outliers in this model. While this approach facilitates a robust detection of new objects in an environment with Lambertian surfaces, its performance can drop if objects with specular surfaces are present since they also form outliers in this model.

Hence, in this paper, we present a method for the detection of novel objects under varying lighting conditions and in presence of specular surfaces. One of the main contributions is a sophisticated statistical model for the luminance and the saturation of the shading images which is used to identify bright specularities and dark shadows in training images. These regions are described by their binary shapes and stored in the environment representation. The second main contribution is the suppression of specularities and heavy shadows in the novelty maps computed from new camera images by a shape matching technique based on Zernike descriptors.

The remainder of this paper is structured as follows. Section 2 briefly revisits our method for the computation of illumination-invariant image-based environment representations. In Section 3, we present the inference of statistical models for the shading of the environment and the shape-based description of dominant illumination effects. Section 4 presents our novel method for object detection. Before we conclude this work, we present experimental results in Section 5.

2. ILLUMINATION-INVARIIANT ENVIRONMENT REPRESENTATION

As described in [8] and [9], an illumination-invariant image-based environment representation can be computed from image sequences

\[
I_m, \ m = 1, ..., M
\]

which are captured at a series of densely spaced
viewpoints from a static scene under different lighting conditions. At each viewpoint, both a depth map is computed and the 6D pose (3D position + 3D orientation) of the camera in the environment is estimated.

From each captured image sequence $I_m$, a series of virtual views $I_{v,m}$ is interpolated at a defined set of densely spaced viewpoints, as illustrated in Fig. 1. The median over the gradients of all virtual images at an identical viewpoint removes the gradients which stem from varying illumination effects like shadow borders or specularities.

In an integration step, an illumination-invariant image is recovered from the gradient image which is then free of varying illumination effects.

![Fig. 1. The computation of an illumination-invariant image-based environment representation by interpolation of virtual images from different image sequences.](image)

### 3. ILLUMINATION MODELING

As in prior work on intrinsic images [7], we model a camera image by the product of an illumination-invariant image and a shading image. Hence, using the illumination-invariant image computed at a viewpoint of the sequence $I_v$, the shading can be extracted from the virtual camera images interpolated from the image sequences $I_m, m = 1, ..., M$ at that viewpoint. As already noted in [8], shading images mainly resemble gray-scale images and lack color information. Looking at an example of a shading image in Fig. 2(b), which shows the shading component of the camera image in Fig. 2(a), one sees, however, that there are also regions of higher color saturation, e.g., in the regions of the specularities on the green object.

We took 9 camera images of the scene in Fig. 2(a) under varying lighting conditions. In HSV color space, the histograms of the luminance and saturation components of their shading images show that the majority of the values lie around 1 (see Fig. 2(c) and Fig. 2(d)). The histograms drop quickly towards lower and higher luminance and saturation values and exhibit a skew towards high luminance or saturation values. A probability distribution which approximates well the histograms of the luminance and the saturation of shading images is the gamma distribution. Using the $N$ intensity values $x_v$ and the $N$ saturation values $x_s$ of the shading images computed from all $M$ interpolated virtual images at a given viewpoint in $I_v$, the parameters of the gamma distributions can be inferred by Maximum Likelihood estimation. The index $k = \{Y, S\}$ represents either the luminance or saturation component and $\Gamma(\cdot)$ is the gamma function. The parameter $a_k$ is computed by solving the nonlinear equation

\[
\log (a_k) - \psi (a_k) = \log \left( \frac{1}{N} \sum_{j=1}^{N} x_k, j \right) - \left( \frac{1}{N} \sum_{j=1}^{N} \log x_k, j \right)
\]

using the Newton-Raphson method [10]. $\psi (a_k) = \frac{d}{da_k} \Gamma(a_k)$ is the digamma function. For the parameter $b_k$ there exists the closed-form solution

\[
b_k = \frac{a_k \cdot N}{\sum_{j=1}^{N} x_k, j}. \tag{3}
\]

Since specularities usually go along with high luminance values and shadows with very low luminance values in the shading images we use the probability distribution $p(x_v)$ in order to detect these illumination effects as outliers. We compute the information of each luminance pixel $(u, v)$ in the $M$ shading images as

\[
i_Y(u, v) = -\log P(x_Y(u, v)) \tag{4}
\]

where $P(x_Y(u, v))$ is the probability of the luminance value $x_Y(u, v)$. Next, we look for regions which encompass information values which lie above the threshold

\[
\tilde{T}_Y = \frac{i_Y}{1 - \kappa} \tag{5}
\]

where $i_Y$ is the average information over all shading images. Throughout this work we use $\kappa = 0.2$. The information maps are binarized by setting all values above this threshold to 1 and all values below to 0. Images containing the binary shapes are stored at each viewpoint in $I_v$, together with the parameters $a_Y, b_Y, a_S$ and $b_S$. This data forms part of the environment representation.
4. DETECTION OF NOVEL OBJECTS

If the color and/or the luminance of new objects in the scene contrasts to the color and luminance of the environment, the corresponding regions in the shading image contain, as in case of specularities and shadows, luminance and saturation values which are outliers in the illumination models in (1) and provide high information values of shading regions in the shading image contain, as in case of specularities. If the color and/or the luminance of new objects in the scene contrast to the new objects, then illumination effects have to be identified and their information values have to be attenuated.

In this work, we propose a matching between binary shapes in order to identify illumination effects in a new camera image. This is motivated by the fact that both the shapes of specularities and shadows are subject to the geometry of the objects in the environment and thus have similar shapes under similar lighting conditions. Furthermore, specularities can only be present on objects with specular surfaces, not in the rest of the environment. For shape matching we use descriptors based on Zernike moments as proposed in [11]. The Zernike moment of order \( n \) with repetition \( m \) for an image region \( f(u, v) \) inside the unit circle is computed as

\[
A_{nm} = \frac{n + 1}{\pi} \sum_{u} \sum_{v} f(u, v)V_{nm}(u, v), \quad u^2 + v^2 = 1
\]

where \( V_{nm}(u, v) \) are complex Zernike polynomials. The magnitudes of the Zernike moments up to order \( n = 120 \) are summarized in a vector and build the descriptor \( d \).

The steps of our algorithm for the detection of novel objects are:

1. Compute the shading image from a new camera image using an illumination-invariant image interpolated from the environment representation in Section 2.
2. Identify regions in the shading image with information values \( i_Y(u, v) > T_Y \). For this, the parameters of the gamma distributions are also interpolated from nearby reference viewpoints.
3. For each region \( R' \):
   a. Compute the intersection \( BB_1 \) of the bounding box of \( R' \) and the bounding box of a shape \( R_M \) stored in the representation. The binary images containing the shapes \( R_M \) are interpolated from the representation at the viewpoint where the new image is taken.
   b. If \( BB_1 \neq \emptyset \), compute the Zernike descriptors \( d_{i}^{1} \) and \( d_{M,1} \) of the partial regions \( R_{i}^{1} \) and \( R_{M,1} \) which lie inside \( BB_1 \).
   c. Compute the distance of the two shapes using the Euclidean distance \( \Delta = \|d_{i}^{1} - d_{M,1}\| \).
   d. Store with the pixel region covered by \( R_{i}^{1} \) the minimum distance \( \Delta \) obtained from the region \( R_{M,1} \) in the environment representation which matches the region \( R_{i}^{1} \) best.
4. Compute for each pixel \( (u, v) \) in the new camera image the novelty measure

\[
\nu(u, v) = \frac{\Delta(u, v)}{\Delta} \cdot \frac{1}{\sqrt{i^{2}_Y(u, v) + i^{2}_S(u, v)}}.
\]

\( \Delta \) denotes the maximum shape distance across all pixels.

5. EXPERIMENTAL RESULTS

For the experimental validation of our proposed approach we acquire 8 image sequences \( I_m, m = 1, \ldots, 8 \) of a scene using the mobile robot platform in Fig. 3(a). The robot is controlled to move along a trajectory, which is similar to the quarter circle between the points \( Q_1 \) and \( Q_2 \) depicted in Fig. 3(b). In each run the scene is illuminated by two lamps mounted on a tripod whose position is changed around the scene from run to run. The robot is equipped with a Bumblebee XB-3 which has three cameras. The center camera captures the images for the image sequences \( I_m, m = 1, \ldots, 8 \) and the left and the right camera are used for depth estimation. The 6D pose of the Bumblebee XB-3 is estimated for each captured image using an optical tracking system on the ceiling of the laboratory.

Before the acquisition of a new sequence of 70 images a white cup is added to the scene. Fig. 4(a) shows frame 20 of the image sequence and Fig. 4(b) shows the illumination-invariant image interpolated from the environment representation. Despite the strong reflections on the tin plate and the shadow which the robot casts on the scene near the lower right image corner, the values in the novelty map in Fig. 4(c) are on average higher for the region of the cup than for the rest of the image. For a quantitative analysis we label in all 70 frames the region covered by the new white cup. The receiver operating characteristic (ROC) curves are obtained by sweeping a threshold through the value range of the novelty maps of all 70 frames. They illustrate the true positive rates (TPR) vs. the false positive rates (FPR) and show that our approach proposed in this paper outperforms the approach in [8]. For high FPRs the performance of our approach is similar to the performance of the normalised gradient correlation (NGC) [5]. At FPRs smaller than 10%, however, our approach also provides higher TPRs than the NGC method.

We further validate our approach with a series of images which we acquire with a camera mounted on a tripod at a single viewpoint. 16 images are taken while the position of a light source is varied half around the scene. A white object is placed into the scene and 6 new images are taken with the position of the light source changing. One of these images is shown in Fig. 5(a). The novelty map in Fig. 5(c) clearly indicates the novel object in the scene while illumination effects are not exhibited. In the region of the strong peculiarity on the plate the novelty map shows very low values. This is because its shape matches very well a shape which has been extracted from the 16 shading images before. Considering the ROC curves in Fig. 5(d) which are obtained by evaluating all six new images, the NGC
algorithm shows the highest TPRs if high FPRs are permitted. This is because the unicolored object strongly changes the gradient structure of the table surface with the line and flower pattern. However, below a FPR of 7% our method again outperforms both the NGC algorithm and the approach in [8].

6. CONCLUSION

In this paper we present an approach for the detection of novel objects under varying lighting conditions in an environment which contains specular objects. Using statistical models of the luminance and the saturation of shading images, outlier regions which stem from shadows and specularities are detected and later suppressed in new camera images by shape matching. Compared to our method presented in [8] the novel approach in this paper provides a more reliable detection of new objects at low false positive rates. Although the normalized gradient correlation is a very fast method to detect changes in images under varying illumination, it exhibits lower true positive rates at small false positive rates in our experiments. Small false positive rates are important for attentional selection since the number and the size of regions in the image which distract the robot’s attention from the actual target are small. In future work we plan to investigate more sophisticated methods for shape matching and we will consider the detection of novel objects in new images in situations where their shapes overlap with shapes of specularities.

7. ACKNOWLEDGEMENT

We thank Daniel Althoff from the Institute of Automatic Control Engineering (TU München) for the navigation of the robot.

8. REFERENCES