in: Proc. 55th Int. Scientific Colloquium, Ilmenau, Germany, 2010, pp. 33-37, Verlag ISLE 2010

ESTIMATING LIGHT REGIONS IN INDOOR ENVIRONMENTS FOR A MOBILE ROBOT CAMERAMAN

Sandra Helsper and Horst-Michael Gross*

Neuroinformatics and Cognitive Robotics Lab
Ilmenau University of Technology, Germany

ABSTRACT

This paper presents an approach for determining an illumination model used by a mobile robot in an indoor living environment. Detecting light sources and estimating the illumination situation are important basic principles in order to find an optimal pose for observing a person, because lighting is one of the fundamental aspects which affects image quality. The robot has to cope with high dynamic range of scenes and with changing lighting conditions. In a first step, irradiances of the environment are estimated. Sections with high irradiance values are defined as light regions. Unlike other approaches which describe illumination, in this paper not only the lighting situation at one particular position shall be estimated but also for several poses in a room in relation to the observed person.

***Index Terms*—** autonomous robot cameraman, illumination estimation, light source estimation

1. INTRODUCTION

In these days mobile robot systems start to become a part of the daily life. One important field is the development of autonomous systems for Ambient Assisted Living application areas. In this case the mobile robot has to assist elderly people and supporting them in their indoor living environments. In addition to services like reminding of taking medicine or keys and physical and mental stimulation exercises, there are tasks which include recording pictures and videos of the person to care for. The information of these images can be used to recognize unusual behavior and critical situations, but video conferences can also be useful to consult doctors for first medical diagnoses or to keep in touch with family and friends.

In each of these tasks a good image quality is a prerequisite to ensure that the required information is observable. If the robot has to regard the image quality, it has to imitate actions and reflections of a real cameraman who achieves a good quality by combining the influences of camera pose, image composition and the illumination of the scene.

The focus of this paper lies in the estimation of the illumination situation of an environment in order to find optimal observation poses for the robot cameraman. It is important to localize the positions of the light sources in one room, especially the ones which can influence the image quality. This includes for example windows and lamps. With the knowledge of the positions of the light sources it can be determined from which direction the person is illuminated. In addition it is essential to analyze the light sources regarding glare. In this context two different kinds of glare are defined which have to be prevented by the cameraman. A false positioning on the one hand can cause glaring of the observed person and leads to reduced visual performance and indisposition, and on the other hand also the robot can be glared, which means that the camera can not compensate the direct incidence of light and produces overexposed images. Such situations can occur because the robot interacts with the person and thus the person turns the face towards the robot as in video conferences. Therefore not only the positions of light sources have to be estimated but also their intensities, so that the robot cameraman can position itself at the optimal position.

In this paper the basis for an illumination model is provided. Therefore it is necessary to estimate the illumination situation of the environment and localize light sources which can cause glare. In a following step it is planned to describe a complete illumination model by means of these illumination estimations. It should be possible to react to changing lighting situations which can occur due to changing time of day or sudden modifications like switching a light on or off or drawing a curtain. So the model has to be flexible enough to detect important changes and adapt its description of the illumination situation. Furthermore it is required to construct the model iteratively so that it can also be continually enhanced and updated.

This paper is structured as follows. Section 2 refers to existent approaches which deal with mobile camera systems and the estimation of light sources. Section 3 presents a first possibility to detect light regions. This is improved by using estimated irradiances introduced

*The work of Sandra Helsper is supported by a scholarship of the Carl-Zeiss Foundation.

in section 4. Following the last section explains how an illumination model can be created.

2. RELATED WORK

The notation autonomous camera(man) system is often used in relation to digital technology for broadcasting where intelligent systems control the camera work so that persons or objects can be tracked. The main focus is the camera work while following moving objects, which is more exact for camera robots but less accepted by audience, so a control system learns to imitate human cameramen techniques ([1] and [2]). In the majority of cases the cameras are fixed at positions where the lighting conditions are good. So the control system do not have to find an optimal position for itself. Camera work in general is not the focus of the robot cameraman.

The basic principles for estimating the light sources and illumination of a scene can be found in computer graphics. An application area which depends on illumination estimation is e.g. Augmented Reality (AR) where artificial objects are embedded into shootings of real scenes ([3] and [4]). Here, the light information is used to simulate illumination effects for the artificial objects.

In [3] an overview of different common illumination models is found. Various approaches define constraints for their illumination models, like the assumption that in a scene there only exists one light source or one distant light with parallel rays, but this is no suitable assumption for real scenes. Additionally, most approaches are only interested in the illumination situation of one particular point in a scene, namely the point at which an artificial object is inserted. Therefore special configurations are used to measure the lighting situation such as photographing a mirrored ball. In contrast, the requirement for the mobile robot cameraman is to use only the available camera.

Kölzer et al. [5] use unordered images of a standard camera which are arranged in environment maps realized as cube maps, where six maps are arranged parallel to the coordinate axes. The cube map is rendered into a single texture which is used to extract lights by means of a median cut algorithm introduced in [6]. The composition of single images can be omitted if an omnidirectional camera is available for the robot cameraman. In this paper we also apply the median cut algorithm and the method which converts pixel values to irradiance. Both approaches are adapted and extended for the requirements of the robot cameraman and are described in detail in the next section.

3. LIGHT REGIONS

The detection of light sources is realized in [5] by using the median cut algorithm introduced by Debevec [6].

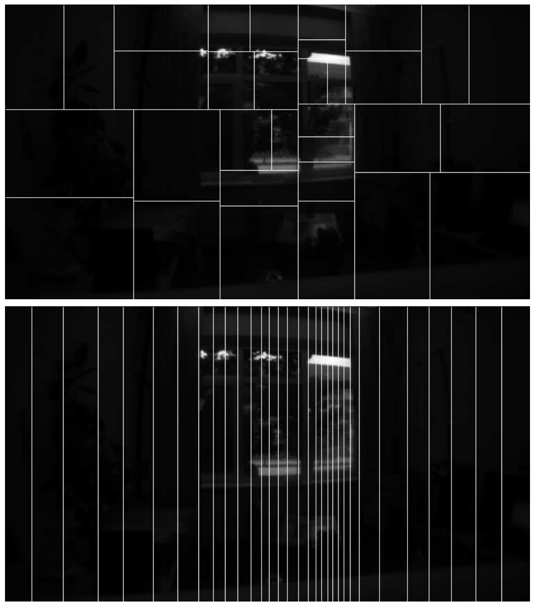

Fig. 1. *Median cut algorithm for underexposed images. Small subsections center on the bright window area.*

Here a light source constellation is approximated by dividing the image in sections with equal light energy. In n iterations the algorithm creates 2^n subsections by dividing respectively along the longest dimension. In order to optimize the computation of the light energy of a section the image can be transformed into a summed area table described in [7]. Finally a light source is placed at the center or centroid of each subsection and the light source color is set to the sum of pixel values of the region.

As a first approach the median cut algorithm is applied to single images. In this case the gray values are considered as light energy because they are a direct non-linear mapping of irradiances described by the camera response function. The computation of sections with equal light energy using the original partition along the longest dimension is shown in figure 1 and 2 in the upper images. Two different exposure times were chosen, in figure 1 an underexposed situation and in figure 2 a sufficiently good exposure time. In both images smaller subsections concentrate on the window area, the brightest region in the image. However, the window itself is not as simple to localize because there are no consistent boundaries between particular sections. In this case it is appropriate to split only along the horizontal dimension so that only vertical columns occur (see fig. 1 and 2).

For a first basic determination of light sources underexposed images can be used. The assumption is that light sources exceed a certain gray value, other regions can be ignored because they have no fundamental effect on the illumination of the scene. Figure 3 shows the mean gray values of each estimated column. As a comparison the image shows the subdivision into 16

Fig. 2. *Median cut algorithm for sufficiently exposed images. The subsections are distributed more evenly over the image.*

columns while figure 1 shows 32 columns. The two diagrams illustrate the mean gray values of both distributions. Using more columns results in more nuances of the gray values, but it is sufficient to use a smaller number because in the most cases the following results do not differ if the number of columns changes. This conclusion is of course relative to the used image sizes, so our camera resolution provides an image width of 752 pixels.

Based on the gray values those columns are chosen which exceed a threshold, in this case about 40% of the maximum gray value. The columns are additionally grouped to light regions if the difference between two neighboring columns is lower than about 10% of the maximum gray value. Figure 4 illustrates an example for light regions based on the underexposed image of figure 3 but the limitations of the regions are displayed at a normal exposed image for better visualization. In the image the solid edges separate light regions from these which are categorized as too dark, the dashed edges symbolize the neighborship of two light regions.

In indoor environments the decisive light sources are normally found along walls or as ceiling illumination. In the context of glare, ceiling illumination can be ignored because it lies clearly above eye level. In addition its function is to illuminate the room constantly so it influences the entire light level. These light sources become more interesting if the environment is too dark and no good observation pose can be found. In this case the robot can start an interaction and ask for switching the light on.

In this work the localization of the mobile robot and

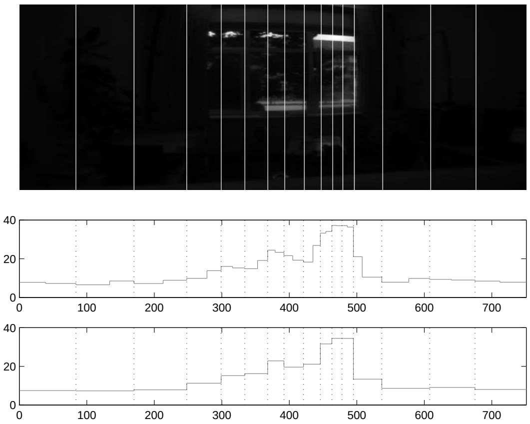

Fig. 3. *Median cut distribution in 16 columns and the associated average gray values of each section (lower image). As comparison the average gray values for a 32 column distribution is shown in the middle.*

Fig. 4. *Light regions based on a threshold applied to the mean gray values of an underexposed image shown on a normal exposed image.*

the availability of an environment map are considered as given requirements. The knowledge of the location of walls and the distance to them helps to restrict input images to important regions. In this way the ceiling can be hidden in images and examined separately. The environment map is also helpful for saving light information which is discussed in a later section.

In this section it is demonstrated that underexposed images can give a first indication for regions with light sources. Images with higher exposure times cannot be used because of the restriction of image capture devices. They can only handle a limited dynamic range and so with a higher exposure time more pixel values around 255 occur because the irradiances of bright regions are restricted. Until now there is no significant information about the intensity of the light regions because the gray values of the underexposed images are not sufficient. Therefore the irradiances have to be estimated which are, unlike the gray values, independent of the exposure time. In the next section the approach of Robertson [8] is presented with which the irradiances can be calculated.

4. EXTRACTING LIGHTING INFORMATION

Standard digital camera systems are strongly limited in their ability to reproduce scenes with high dynamic range. In comparison the human eye can cope with a considerably wider dynamic range. In indoor living environments such high contrasts can appear in particular close to windows in the daytime. In order to estimate the possible glare for the person, the robot cameraman has to get an impression of the intensity of illumination.

During the generation of images the irradiances of the environment are accumulated at an image sensor as energy which is converted to pixel values. The amount of collected light energy depends on the exposure time which limits the accumulation. For a high exposure time the collected light energy of many pixels exceed the maximum limit which produces a gray value around 255, reciprocally, a low exposure time many pixels are mapped to low gray values. For this reason pixel values in the range of 0 and 255 are unsuitable for reconstructing the irradiance values. The camera response function describes the non-linear mapping between irradiances and gray values, so if an image is grabbed the irradiances can be reconstructed as long as the response function is known. If it is unknown it can be estimated using an exposure series. Images with different exposures are necessary to compensate the described uncertainty of the low and high pixel values.

Robertson et. al. introduce in [8] an approach for estimating the camera response function while simultaneously estimating the irradiances. For N images with different exposures $t_i, i = 1, \ldots, N$ the set $\{y_{ij}\}$ defines the known observations, where y_{ij} is the j^{th} pixel value of the i^{th} exposed image. In this context x_j describe the related irradiances.

In case of static cameras the quantity of light for a pixel accumulated of the sensor is described by $t_i x_j$ and an additional noise term.

The quantity $z = t_i x_j + N_{ij}$ is mapped by the non-linear camera response function to the related pixel value y_{ij}.

$$f(z) = \begin{cases} 0 & \text{if } z \in [0, I_0] \\ m & \text{if } z \in (I_{m-1}, I_m], m = 1, \ldots, 254 \\ 255 & \text{if } z \in (I_{254}, \infty) \end{cases} \tag{1}$$

The values I_m define the camera response function which are not evenly spaced because of its non-linearity. These values are estimated by means of a form of Gauss-Seidel relaxation. In [8] the objective function to be minimized is defined as follows:

$$O(I, x) = \sum_{i,j} w_{ij} (I_{y_{ij}} - t_i x_j)^2 \tag{2}$$

The term w_{ij} weights values in the middle of the output range higher than those close to 0 or 255 because the uncertainty of pixel values in the boundary area. If there are only the different exposed images are given the I_m values as well as the irradiances x_j are unknown. Therefore both values have to be estimated in the objective function (2). The Gauss-Seidel relaxation minimizes an objective function with respect to a single value and can than use these new values for minimizing with respect to the second value. For minimization with respect to one value the respective partial derivative of (2) is used. The whole process and the corresponding formulas can be found in [8].

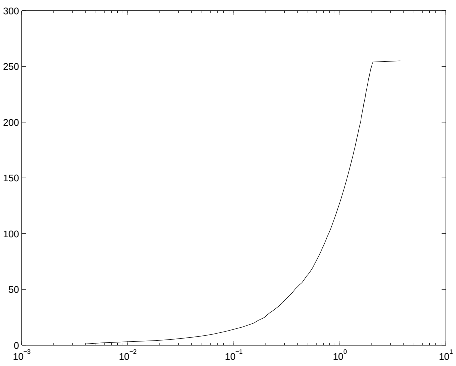

Fig. 5. *Estimated response function using eight images with different exposure times.*

Figure 5 shows the used exposure series and the resulting camera response function. We use gray scale images for estimating only the response function for gray values because we are not interested in different color values.

After the I_m values are known the irradiance values x_j can be estimated.

$$\hat{x}_j = \frac{\sum_i w_{ij} t_i I_{y_{ij}}}{\sum_i w_{ij} t_i^2} \tag{3}$$

The estimation of the x_j becomes more precise the more input images with different exposure times are used because low exposed images contain information about bright regions and high exposed images accordingly contain information about dark regions.

We choose three exposure times 1/250 sec., 1/60 sec., and 1/15 sec. for a static position of the robot cameraman. The pixel values and exposure times are used in (3) to compute the irradiances x_j. Instead of the pixel values, now the irradiances are used as input for the median cut algorithm. In figure 6 the estimated columns of equal light energy are shown at the image grabbed with exposure time 1/60 sec. Like in figure 3 the smaller columns concentrate on the window area. The mean values refer to the irradiances and because of the additional information of the different exposed images these values are more representative than the single gray values. The exposure series are grabbed with a cloudy sky, but on sunny days the x_j values are much higher. So these values are an indication for the

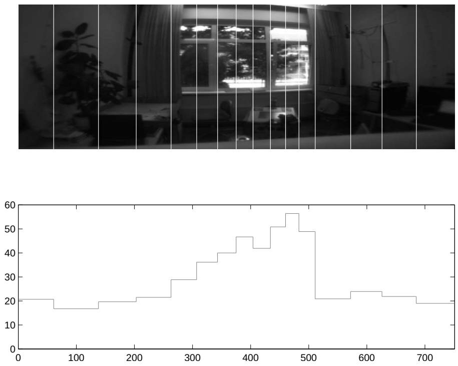

Fig. 6. *Distribution into columns with equal light energy using estimated irradiances. The upper diagram shows the mean irradiance of each column.*

intensities in the light regions. These results are the basis for an illumination model whose basic principles are explained in the next section.

5. FUTURE WORK

In combination with the environment map the regions can be projected on the boundaries of the room. This is advantageous for recognizing light regions if images where taken from different positions in the room, because the recognition is the basic principle for combining several measurements. More tests are necessary in order to evaluate if it is advantageous to save single light regions or just one region where adjacent light regions are combined.

Under certain conditions the size of the light region can be an indication of the type of the light source. In indoor environments wide regions mostly represent windows. So for such a region it is evident that the intensity changes by reason of time of day or weather, and that the intensity is depending on the viewing direction because of the position of the sun. Intensities which are direction-dependent must be described through different observation poses.

So the complete illumination model has to store the positions and widths of the light regions projected on the walls (with a range of tolerance for combining light regions for different viewing directions), the intensities based on the estimated x_j values and additionally a time stamp for the measurement. According to [9] the assumption can be made that for regions with only one estimation this value is equal for each direction. Further measurements can be stored according to their directions and the intensity values can be interpolated.

In this context more research work is necessary to define sensitive thresholds which indicate whether further intensity estimates differ from previous values just because of the dependency of the light source intensity on the viewing direction or because of an entire change of the intensity through outside influences. This is required to decide if the measurement is integrated in the existent description of the light source or if the whole description has to be rebuilt.

6. CONCLUSION

In this paper an approach for estimating light regions in indoor environments is presented and also a possibility to construct a model which includes decisive information. By means of the adapted median cut algorithm light regions can be detected which are the important regions for the entire lighting situation. The intensity of the light region is described by using estimated irradiance values to make it possible to evalute the regions with regard to glare. Further research work is necessary to improve the combination and modification of the intensity values of light regions so that a complete illumination model can be used to find optimal observation poses.

7. REFERENCES

[1] D. Kato, A. Ishikawa, T. Tsuda, S. Shimoda, and H. Fukushima, "Automatic control of a robot camera for broadcasting and subjective evaluation and analysis of reproduced images," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2000, vol. 3959, pp. 468–479.

[2] M. Okuda, S. Inoue, and M. Fujii, "Machine learning of shooting technique for controlling a robot camera," in *Robotics and Biomimetics (ROBIO), 2009 IEEE International Conference on*, 2009, pp. 1111 –1116.

[3] Jürgen Stauder, *Schätzung der Szenenbeleuchtung aus Bewegtbildfolgen*, Fortschritt-Berichte VDI, Reihe 10 Informatik / Kommunikationstechnik Nr. 584. Düsseldorf: VDI Verlag, 1999.

[4] Paul Debevec, "Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, 1998, pp. 189–198.

[5] Konrad Kölzer, Frank Nagl, Bastian Birnbach, and Paul Grimm, "Rendering virtual objects with high dynamic range lighting extracted automatically from unordered photo collections," in *ISVC '09: Proceedings of the 5th International Symposium on Advances in Visual Computing*. 2009, Lecture Notes in Computer Science, pp. 992–1001, Springer-Verlag.

[6] Paul Debevec, "A median cut algorithm for light probe sampling," in *SIGGRAPH '05: ACM SIGGRAPH 2005 Posters*, 2005, p. 66.

[7] Franklin C. Crow, "Summed-area tables for texture mapping," in *SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, 1984, pp. 207–212.

[8] Mark A. Robertson, Sean Borman, and Robert L. Stevenson, "Dynamic range improvement through multiple exposures," in *In Proc. of the Int. Conf. on Image Processing (ICIP 99)*, 1999, vol. 3, pp. 159–163.

[9] Christof Schroeter, Matthias Hoechemer, Steffen Mueller, and Horst-Michael Gross, "Autonomous robot cameraman – observation pose optimization for a mobile service robot in indoor living space," in *ICRA'09: Proceedings of the 2009 IEEE international conference on Robotics and Automation*, 2009, pp. 2199–2204.