Published in proceedings of ICPR 2000, vol. 1, pages 778-781, 2000.

Range Image Acquisition of Objects with Non-uniform Albedo Using Structured Light Range Sensor*

Danijel Skočaj and Aleš Leonardis
University of Ljubljana, Faculty of Computer and Information Science
Tržaška 25, SI-1001 Ljubljana, Slovenia
{danijel.skocaj, ales.leonardis}@fri.uni-lj.si

Abstract

We present a novel approach to acquisition of range images of objects with non-uniform albedo using a structured light sensor. The main idea is to systematically vary the intensity of the light projector and to form high dynamic scale radiance maps. The range images are then formed from these radiance maps. We tested the method on the objects which have surfaces with very different reflectance properties. We demonstrate that the range images obtained from the high dynamic scale radiance maps are of much better quality than those obtained directly from the original images of a limited dynamic scale.

1. Introduction

Systems consisting of a structured light projector and a camera are commonly used for acquiring range images. A major limitation of such systems is their sensitivity to reflectance properties of the object's surfaces.

Most of the problems are due to the limited dynamic scale[1] of CCD cameras. This problem is evident on objects which have surfaces with very different reflectance properties. Since the dynamic scale is limited, we can not reliably capture both high and low reflective surfaces simultaneously. In this work we present a novel approach to overcome this problem. The main idea is to systematically vary the intensity of the light projector and to form high dynamic scale radiance maps. The range images are then formed from these radiance maps.

The paper is organized as follows. In section 2 we describe the problem, while in the section 3 we explain our approach to the solution of the problem. In section 4 we present some results of the new approach and compare them with the results obtained with the standard acquisition method. The last section concludes our paper and gives a short outline of a planned future work.

2. Problem description

For range image acquisition we use a coded light range sensor based on an LCD stripe projector. The projector projects stripe patterns over an object. A range image is obtained by analyzing the captured images of the object with projected stripe patterns. The accurate estimation of the 3-D coordinates depends on the accurate binarization and localization of the projected stripes [3].

Since we can control the illumination intensity of the stripe projector, we can acquire range images of an object under different lighting conditions. As we can see in Fig. 1, the quality of a range image depends on the illumination intensity of the stripe projector. The optimal results (i.e., minimal errors) are obtained, when the range image is taken under the suitable illumination intensity, otherwise the quality of the range image degrades.

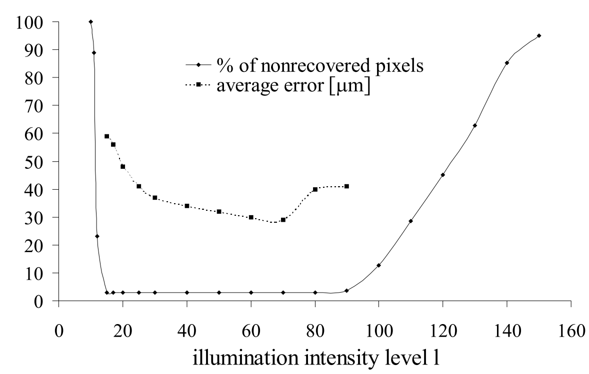


Figure 1. Percentage of the pixels where depth could not be recovered (solid line) and average distance between the obtained 3-D points on a planar surface and the fitted plane (dotted line).

*This work was supported by a grant from the Ministry of Science and Technology of Republic of Slovenia (Project J2-0414).

[1]The term *dynamic range* is commonly used in the literature, however, we rather use the term *dynamic scale* to avoid possible confusion with the term *range (i.e., depth) image*.

There are two reasons for such behavior. One is that the signal-to-noise ratio is decreasing with the decrease of the illumination intensity. Thus, when we acquire a range image at a low illumination intensity, signal-to-noise ratio is low, therefore there is a higher probability for erroneous binarisation of pixels on the borders of projected stripes and consequently wrong estimation of 3-D coordinates.

On the other hand, at high illumination intensities we face the problem of camera saturation. Some highly saturated pixels even spill over and affect values at neighboring pixels. This is known as "blooming" of CCD cells. At such pixels it is impossible to make a proper binarization and consequently estimate 3-D points.

Therefore, we should aim to use as high projector illumination intensity as possible to achieve high signal-to-noise ratio, but at the same time to avoid pixel saturation. Having a scene containing multiple surfaces of varying reflectances, it is evident that high reflective surfaces require lower illumination intensities while surfaces with lower reflectance require higher illumination intensities to achieve optimal results. How can we satisfy both conditions simultaneously? Fig. 4 demonstrates that when we acquire a range image of an object with non-uniform albedo with low illumination intensity, we can not reliably acquire depth of surfaces with low reflectance (because of the low signal-to-noise ratio), while when we acquire a range image with high illumination intensity, we can not recover depth of surfaces with high reflectance (because of the saturation). As a solution to this problem we propose a new method which uses a high dynamic scale radiance maps of an object instead of its intensity images.

3. Our approach

We obtain a high dynamic scale radiance map by taking multiple images at different illumination intensity levels of the stripe projector (a similar effect can be achieved by varying exposure times [1] or iris settings). The question is how to combine multiple images into a single map.

Ideally, we can assume that a pixel value g is proportional to the radiance of the corresponding point in the scene (with the exception of saturated pixels). Since the stripe projector is the only light source in the system, the scene radiance is proportional to the illumination intensity of the projector. Therefore, a pixel value g should be proportional to the reflectance[2] r of the corresponding surface point and to the illumination intensity level l, thus $g \propto r \cdot l$. Since the radiance of a point is proportional to its reflectance and we are interested only in relative values of the radiance, we can neglect the scale factor, thus write $g = r \cdot l$ and take the value of r as an entry in the radiance map.

If g_{ij} is the grey level of the i-th pixel in the image taken under illumination intensity l_j, then $g_{ij} = r_i l_j$, where r_i is the relative reflectance of the corresponding scene point. Relative reflectance of the i-th pixel $r_i = g_{ij}/l_j$ should remain constant for all illumination intensities. In this way we could combine the information from several images into a single radiance map.

This simple method would yield correct results, if the strict linear relationship held. However, it turns out that the sensor system introduces nonlinearities. As a consequence, the calculated relative reflectance values of the i-th pixel $r_i = g_{ij}/l_j$ yield different values at different illumination intensities. This is plotted in Fig. 2. Thus, to combine the data from multiple images, we have to recover this nonlinear mapping.

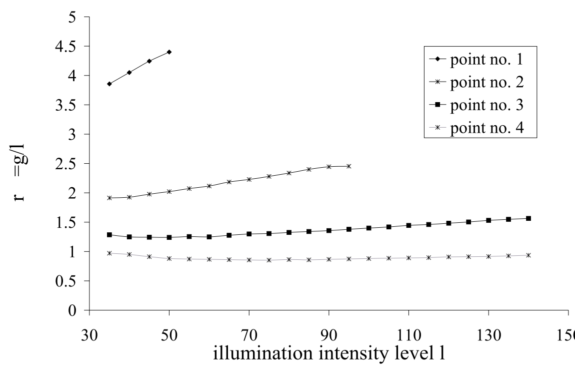


Figure 2. Relative reflectance values of four points computed as $r_i = g_{ij}/l_j$.

To achieve this, we have extended an approach of Debevec and Malik [1] which dealt with recovering high dynamic scale radiance maps from photographs.

We model the nonlinearities of the system with two nonlinear functions f_1 and f_2:

$$g_{ij} = f_1(r_i \cdot f_2(l_j)) \ . \qquad (1)$$

If we assume that f_1 is invertible, we can write:

$$\ln f_1^{-1}(g_{ij}) = \ln r_i + \ln f_2(l_j) \ . \qquad (2)$$

Let us define $h_1 = \ln f^{-1}$ and $h_2 = \ln f_2$. If we take m images under different illumination intensities and if we observe n pixels, then we obtain a set of equations:

$$h_1(g_{ij}) = \ln r_i + h_2(l_j); \ \ i = 1 \dots n, \ \ j = 1 \dots m, \qquad (3)$$

where pixel gray values g_{ij} and illumination intensity values l_j are known, while functions h_1 and h_2 and values of relative reflectances r_i are unknowns. We wish to recover

[2] We use the term *reflectance* for the product of two factors: geometric term expressing the dependence on the angles of light reflection and albedo [2].

these unknowns that best satisfy the set of equations arising from (3) in a least-squared error sense.

Since the camera has a limited dynamic scale, a pixel gray level g_{ij} can only be an integer number from a finite interval $1 \dots g_{max}$. Similarly, there are only a limited number (l_{max}) of illumination intensity levels. Therefore, we only need to recover a finite number of values of the functions $h_1(g)$ and $h_2(l)$.

Since the signal-to-noise ratio is higher when the values of acquired pixels are higher, we additionally introduce a weighting function w, which during minimization process puts more weight on equations with higher values of g_{ij} (with exception of saturated pixels, which are excluded).

We can formulate the problem as one of finding the g_{max} values of $h_1(g)$, l_{max} values of $h_2(l)$ and n values of $\ln r_i$, that minimize the following function:

$$C_1 = \sum_{i=1}^{n} \sum_{j=1}^{m} [w(g_{ij})(h_1(g_{ij}) - \ln r_i - h_2(l_j))]^2 \quad . \qquad (4)$$

As an additional constraint we simultaneously minimize the functions

$$C_2 = \sum_{g=2}^{g_{max}-1} h_1''(g)^2 \ \text{ and } \ C_3 = \sum_{l=2}^{l_{max}-1} h_2''(l)^2 \quad , \qquad (5)$$

which ensure the smoothness of the functions h_1 and h_2. For calculation of h'' we use $h''(k) = h(k-1) - 2h(k) + h(k+1)$.

Finally, to solve the problem, we also have to fix at least two values of r_i. Therefore, we have to know the real relative reflection values of at least two pixels. If we have n_R pairs of numbers $\langle p_k, R_k \rangle$, where p_k are indices of pixels for which the relative reflections are known, and R_k are values of these reflections, we obtain additional set of equations $r_{p_k} = R_k;\ \ k = 1 \dots n_R$. Therefore, we can formulate the last constraint of our minimization problem as:

$$C_4 = \sum_{k=1}^{n_R} (\ln r_{p_k} - \ln R_k)^2 \quad . \qquad (6)$$

The combination of Eqs. (4), (5), and (6) leads to the final minimization problem:

$$C = \lambda_1 C_1 + \lambda_2 C_2 + \lambda_3 C_3 + \lambda_4 C_4 \quad , \qquad (7)$$

where λ_1, λ_2, λ_3 and λ_4 are parameters which put different weights on the terms of the equation.

The minimization of the function C is a least squares problem. The overdetermined system of linear equations is solved using the singular value decomposition method as proposed in [1].

By obtaining the functions h_1 and h_2, we can calculate the relative reflectance of every point in every scene from the image taken under an arbitrary known illumination intensity. From (3) we obtain

$$r_i = e^{h_1(g_{ij}) - h_2(l_j)} \quad . \qquad (8)$$

Using (8), all images taken under different illumination intensities yield almost the constant values of r_i for a scene point (Fig. 3). The average normalized standard deviation of calculated values of r_i is $4.22 \cdot 10^{-3}$ while without nonlinear mapping, in Fig. 2, it was $7.48 \cdot 10^{-2}$. Thus, the nonlinearity of the sensor system has been successfully removed.

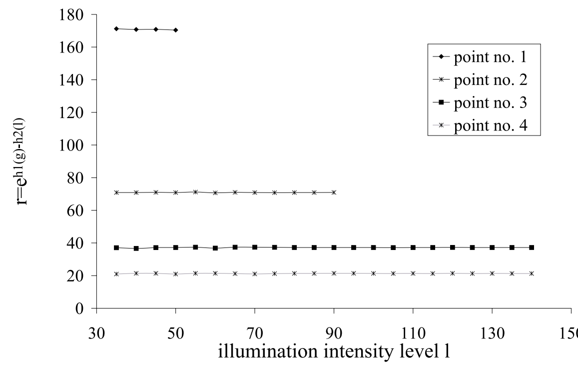


Figure 3. Relative reflectance values of four points computed by Eq. (8).

Now we can combine the images taken under different illumination intensities into a single radiance map. To increase the robustness we should use all gray values g_{ij} captured under various illumination intensities l_j weighted with the weighting function $w(g)$:

$$\ln r_i = \frac{\sum_{j=1}^{m} w(g_{ij})(h_1(g_{ij}) - h_2(l_j))}{\sum_{j=1}^{m} w(g_{ij})} \quad . \qquad (9)$$

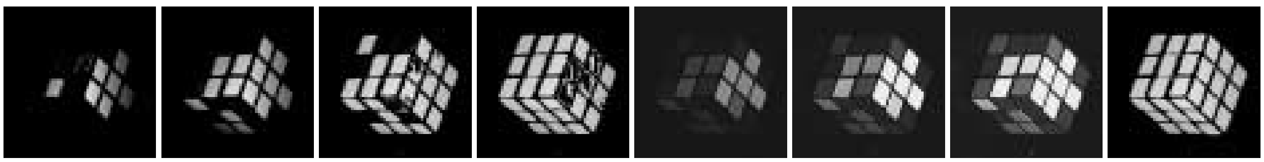
We store the values of r_i as floating point numbers, which means that they are not limited to a finite number of levels any more. These values are always proportional to the radiances of the scene points. Thus, by calculating r_i for all pixels, we obtain a high dynamic scale radiance map.

The remaining issue is how many images are necessary to estimate r_i. Already two images are often enough to obtain a good radiance map. We take one image with low and one with high illumination intensity level of the stripe projector. If we take more images we achieve a higher robustness and noise insensitivity.

Having obtained the radiance maps, we can calculate the range image. The basic algorithm remains the same as the original one [3], only the input data changes. Instead of taking only one image of the object under every stripe pattern, we take several images illuminated by different intensities and form radiance maps, which are then used as an input

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

Figure 4. (a) The intensity image of a toy, rendered models obtained from range images taken under (b) low and (c) high illumination intensity and (d) from the range image obtained using high dynamic scale radiance maps.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5. Four range images of a Rubic's cube obtained under various single illumination intensities (from (a) low to (d) high), three intensity images (from (e) to (g)), and (h) range image obtained using high dynamic scale radiance maps.

to the algorithm for range image formation. In such a way we overcome the problem caused by the limited scale of intensity images and we can reliably capture range images of objects with non-uniform albedo.

4. Experimental results

We tested the algorithm which uses the high dynamic scale radiance maps on objects with non-uniform albedo and compared the results with those obtained with the standard range image acquisition algorithm.

Fig. 4(a) shows an object consisting of parts in different colors. It is evident that we can not acquire the range image of the entire object under a single illumination intensity (Figs. 4(b),(c)). However, when we used the high dynamic scale radiance maps, we reliably acquired the depth of all surfaces (Fig. 4(d)).

As the second example we have chosen a well-known Rubic's cube. The cube is covered with squares of six different colors. Figs. 5(a)–(d) show that we can not recover the depth of all squares simultaneously, however, the depth can be recovered using high dynamic scale radiance maps. Each radiance map was formed from three images taken under low, medium, and high illumination intensity. Figs. 5(e)–(g) show the images used for the calculation of one high dynamic scale radiance map. These radiance maps were then used to create a range image, shown in Fig. 5(h). The depth of all colored squares was successfully estimated.

5. Conclusion

In this paper we have presented a solution to one of the main problems which we face during acquisition of range images of objects with non-uniform albedo. Because of the limited sensing range of the CCD sensor it is impossible to acquire the depth of both high and low reflective surfaces of the object with good quality simultaneously. We presented a method for forming a high dynamic scale radiance maps which solves this problem.

The high dynamic scale radiance maps contain more precise information about the surface properties than ordinary images taken under a single illumination intensity. They can also be used to produce adequate texture maps that are subsequently mapped onto the geometric model of the object. These ideas remain to be investigated in the future work.

References

[1] P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *SIGGRAPH'97 Conference Proceedings*, pages 369–378, August 1997.

[2] B. Horn. *Robot Vision*. The MIT Press, 1986.

[3] M. Trobina. *Error Model of a Coded-Light Range Sensor, Technical report*. Communication Technology Laboratory, ETH Zentrum, Zürich, 1995.