Structured-light-based Depth Reconstruction Using Low-light Pico Projector

Thomas Rittler
Vienna University of Technology
Institute of Software Technology and Interactive Systems
Favoritenstrasse 9-11, 1040 Vienna, Austria
thomas.rittler@tuwien.ac.at

Florian Seitner
emotion3D GmbH
Gartengasse 21/3, 1050 Vienna, Austria
florian.seitner@emotion3d.tv

Margrit Gelautz
Vienna University of Technology
Institute of Software Technology and Interactive Systems
Favoritenstrasse 9-11, 1040 Vienna, Austria
margrit.gelautz@tuwien.ac.at

ABSTRACT

In this work we investigate an infrared structured light prototype which is intended for 3D reconstruction in resource-restricted mobile applications. We explore the constraints on working range and pattern resolution that are imposed by the low-light property of our single-shot set-up. While focusing on the most light-sensitive steps of the decoding workflow, we suggest adaptations of image rectification and pattern generation and segmentation algorithms that are tailored to the specific spatial and radiometric requirements of our system.

Categories and Subject Descriptors

I.4.5 [**Image Processing and Computer Vision**]: Reconstruction

General Terms

Application

Keywords

Structured light, low-light projector, image rectification, pattern segmentation

1. INTRODUCTION

Depth sensing finds its use in a wide range of applications such as 3D reconstruction/scanning, mobile multimedia applications and gaming, or automotive applications. Contrary to passive stereo approaches, active Structured Light (SL) methods substitute one of the cameras by an active device, for example a projector or laser. This active device emits a light pattern - named as *SL illumination* - that is reflected by the scene and captured by the camera. The projected illumination pattern eases the process of finding correspondences in the presence of conditions such as textureless surfaces, repetitive textures, or poor ambient illumination, which are known to significantly reduce the quality and accuracy of the retrieved depth information when passive stereo methods are applied [1, 5].

MoMM 2015, December 11-13, 2015, Brussels, Belgium
 ISBN 978-1-4503-3493-8/15/12 ...$15.00
DOI: http://dx.doi.org/10.1145/2837126.2837154.

In this paper, we focus on a near-infrared SL prototype which is intended to be used on resource-restricted mobile devices. The set-up consists of a single camera and pico projector. Specifically, we investigate the low-light properties of the projector regarding their effects on the image rectification, pattern codification and pattern segmentation algorithms.

The rest of this paper is organized as follows. Section 2 takes a closer look at the algorithmic aspects including the rectification process between the captured camera image and the projected pattern image, pattern codification strategies and pattern extraction. In Section 3 the working range of the proposed SL system with its low projector brightness is evaluated. The concluding Section 4 provides an outlook on further ongoing and future work.

2. ALGORITHMS

In this section, we discuss several algorithmic aspects of the proposed SL system under consideration of the constraints imposed by the low-light property and spatial arrangement of our set-up.

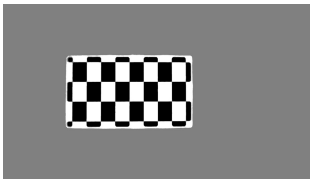
2.1 Image rectification

The first step of the SL decoding workflow comprises the rectification between the camera and projector image. Due to its simplicity and well-known accuracy, the method by Zhang [6] has become one of the most widespread approaches for camera calibration. Inspired by that earlier work, Moreno and Taubin [3] proposed a method for camera/projector calibration by modeling the projector as an inverse camera. A sequence of Gray code patterns is projected onto a static planar checkerboard placed within the working volume for finding correspondences between projector pixels and 3D world points. This approach requires a physical checkerboard to be placed in the scene at various positions, which has to remain static during projecting and capturing of the Gray code pattern sequence.

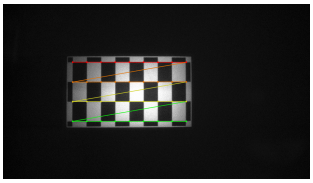
In the case of low projector brightness the accurate determination of chessboard corner locations is diminished or may

(a) Initial camera image.

(b) Camera image after contrast enhancement and denoising.



(c) Camera image with binary chessboard pattern formed after combining the regular and inverse projected chessboard patterns.



(d) Detected chessboard corners overlaid on contrast enhanced input image.

Figure 1: Chessboard-based image rectification.

not be possible at all due to blurry delineations of chessboard tiles and an increased presence of image noise. Hence, since the vertical alignment between projector and camera image denotes a crucial step in order to extract depth information efficiently, we propose a computationally fast method for uncalibrated image rectification. Instead of placing the chessboard physically into the scene, a chessboard image is projected onto a planar surface.

In our processing chain, we first enhance the contrast of the camera image and reduce the amount of image noise by local filtering operators (see Figure 1b). To increase the robustness of the corner detection process, the inverse chessboard is projected additionally and both captured camera images are combined to form a binary image of the chessboard pattern as illustrated in Figure 1c. This binary image serves as the basis to fit the initial chessboard pattern (i.e. projector image) into the camera image via cross-correlation by adapting the scale and orientation of the pattern image. Finally, the artificially created camera image is used to compute the chessboard corner correspondences. After the inner chessboard corners have been detected, a mapping between coordinates of projector and camera pixels is established based on the chessboard corner correspondences. This mapping is then extrapolated in the area of the outer chessboard tiles.

As the correspondences between chessboard corners of the projector image and the camera image are established based on the inner chessboard corner points, the size of the exterior chessboard tiles is reduced so that the set of inner corner points covers a larger part of the projector image, as can be seen in Figure 1.

2.2 Pattern codification strategy

Since our aim is to construct a SL system using a near infrared light source (i.e. invisible to the human eye), color encoding of the pattern image is not an option. Additionally, to minimize the computational complexity and to ease the decoding process in case of low projector brightness, the encoding of codewords based on different shapes appears not to be a good strategy. Thus, we have chosen a binary pseudorandom array using circles as shape primitives (hereinafter also referred to as dot pattern) as the proposed SL pattern.

A widespread algorithm in the computer vision community for constructing such a pseudorandom array is given by Morano et al. [2]. The authors propose an iterative process based on a brute-force approach that allows to define the length of the alphabet, the size of the window, the dimensions of the array and the Hamming distance between different windows to allow error correction. Hence, this approach is easily adjustable to different resolutions of the pattern image. Since the initial Morano algorithm was designed for generating color patterns, an additional connectivity constraint was introduced to ensure that those pixel positions that are adjacent (in an 8-neighbourhood) to an 'on' pixel are set to an 'off' status in our binary patterns. This isolation of 'on' pixels constitutes a important requirement to robustly segment a single dot during the decoding stage.

We use a dot pattern of size 1280×720 pixels which is based on an 80×45 binary pseudorandom array, 9×9 window property and minimum Hamming distance of 3. Currently, finer resolutions are not applicable as smaller dot sizes will further reduce the overall image brightness and as dot segmentation is prone to fail due to reduced sharpness of the pattern image.

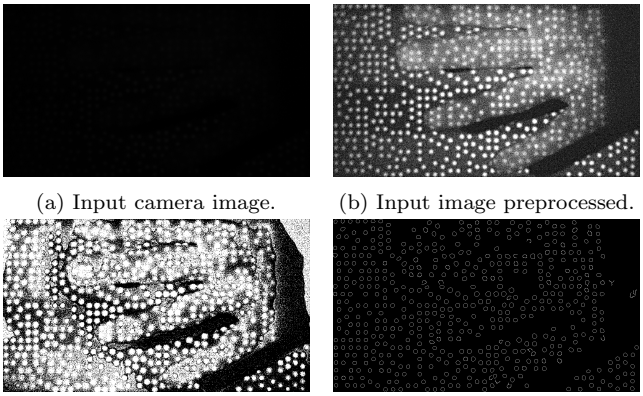
2.3 Pattern segmentation

A crucial step in the workflow of a SL system is the segmentation of the pattern image. After the scene has been illuminated by the projector, the aim is to extract the altered pattern from the camera image in order to compute correspondences between the projected and the captured image pattern. In the following, we discuss the segmentation of the individual dot regions, as shown in Figure 2, in order to extract the distorted pseudorandom pattern. Two methods based on absolute image intensities and image gradients are presented.

2.3.1 Intensity based pattern segmentation

The first approach is based on the idea that the pattern projected onto the objects to be measured is significantly brighter than the basic brightness of the scene. The goal is now to find an appropriate gray level threshold to separate the illumination pattern from its surrounding background. A common approach for image binarization is based on Otsu's method [4]. This algorithm assumes that the image contains two classes of pixels, i.e. foreground and background pixels, and calculates the optimum threshold separating the two classes by maximizing the inter-class variance of the corresponding gray level distribution. In the proposed approach, the threshold resulting from Otsu's method is used as a starting point. In particular, the mean size of the connected components, also referred to as Binary Large OBjects (BLOBs), of the binary image according to the calculated threshold is compared to the size of the dots in the initial pattern image. Depending on the deviation of the dot size in the camera image compared to the original dot size, the threshold is incremented or decremented iteratively until the average difference between the dot sizes lies beneath a given confidence level.

However, the use of a global threshold can lead to poor binarization results due to vignetting artifacts of the cam-



(a) Input camera image. (b) Input image preprocessed.

(c) Intensity thresholding. (d) Gradient-based edge detection.

Figure 2: Dot segmentation.

era sensor, uneven illumination of the projector or varying reflectivity properties of the measured objects. Thus, the image is divided into smaller blocks and local thresholds are computed for each block center with missing thresholds obtained by bilinear interpolation to avoid blocking-artifacts in the resulting binary image.

2.3.2 *Gradient based pattern segmentation*

Instead of using absolute intensity levels, the second approach is based on intensity changes in the input image. As sharp brightness changes typically occur at object boundaries, an edge detection algorithm is applied to segment the individual dot regions. In particular, due to its low error rate and its accurate edge point localization, the well-known Canny edge detector is selected for our processing framework. In our experiments the gradient-based segmentation has proven to be more robust in low-light environments compared to intensity thresholding due to significant presence of image noise (cf. Figure 2 c and d).

3. EVALUATION

In this section, we evaluate the working range of our SL set-up in conjunction with the proposed processing algorithms. The available projector brightness should guarantee that the projected SL pattern can be segmented and extracted from the captured camera image to enable the decoding process. For that purpose, a completely white frame (i.e. all intensities are set to 255 using 8 bit depth) is projected on a plain white wall at three different distances (50 cm, 100 cm and 150 cm) under various indoor lighting conditions. As can be seen in Figure 3, the captured intensity values at a distance of 50 cm deviate by 94.90% from the originally projected intensities. For distances of 100 cm and 150 cm naturally the brightness of the captured camera image further decreases leading to a narrower distribution of pixel intensities. In particular, values only range from 1 to 8 and 0 to 5 reaching their peaks at 4 and 3, respectively. This corresponds to a brightness reduction of almost 99%, and the narrow histogram distributions hinder the setting of suitable threshold values for segmentation. It is worth noting that a plain white wall nearly represents an ideal test case and materials like plastic or fabric additionally diminish the captured pixel brightness, as experiments have indicated. Since our evaluation has been conducted indoors using artificial light sources (i.e. very small amount of infrared radiation), changing ambient light conditions have only small impact on the intensity distributions.

To quantify the quality of the pattern segmentation, a reference pattern - i.e. a pattern where every second grid position is set 'on' - based on an 80x45 random array is projected and the number of retrieved dots is measured. Additionally, to investigate the influence of decreasing projector brightness, the captured image is divided into a 4×4 grid of equally sized rectangles and the detection rates of the 12 rectangles at the marginal area and the remaining 4 rectangles in the center area are analyzed separately. This procedure is motivated by the fact that the brightness of our projector cannot be altered manually and that the currently available brightness of the projector is already very low.

Comparing the ratio between the maximum brightness of the center and the marginal area, experiments have shown that at 50 cm the intensity at border regions reaches between 89.47% and 95.46% compared to the center area for varying lighting conditions. At 100 cm and 150 cm, the ratio falls off to 40% and 30.77%, respectively. Figure 4 shows the precision/recall results of the dot detection based on the segmentation approaches discussed in the previous section. As can be seen, the detection rate drastically decreases with increasing distance. In particular, from 50 cm to 150 cm the recall rate using local image thresholding drops from 90% to 3%, falling already below 45% at 100 cm. Compared to global or local image thresholding, edge detection shows more robust segmentation results, i.e. up to 77% recall at 100 cm while preserving almost 100% precision. Additionally, it can be noticed that there is only a minor difference between precision/recall rates at central and marginal areas when edge detection is used. For example, while for image thresholding the recall score differs by 74% at 100 cm under low ambient light conditions, edge detection results vary only by 5%. In summary, it can be stated that the currently available projector brightness solely allows for acceptable dot detection rates (95% recall at 100% precision on average) up to a distance of 50 cm in this ideal experimental assembly. We also observed that varying ambient lighting conditions generated by indoor light sources had only minor impact on the detection rate at this distance.

4. CONCLUSION

In this paper, we have investigated the properties of a depth sensing system based on near-infrared SL illumination. Based on the projection of a so-called 'single-shot' pseudorandom binary pattern, techniques for projector / camera image rectification with a suitable calibration pattern and dot segmentation under low-light projector illumination were explored. Experiments have shown that currently a limited working range of 50 cm and a pattern image based on an 80×45 pseudorandom array are feasible.

For future work, we will evaluate different strategies to accomplish the disparity computation between the original and captured pattern image using codeword lookup tables, foreground/background segmentation or active stereo matching techniques. Moreover, enhancement of the projector brightness would allow to use modulations of gray values instead of strictly binary patterns to enrich the underlying codeword alphabet and enable the application of smaller window sizes and denser pattern images.

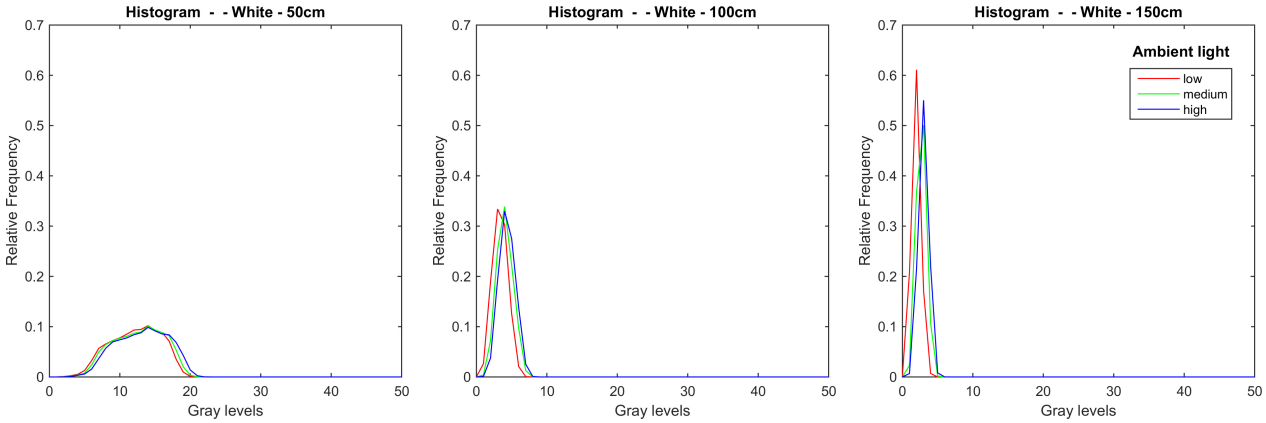


Figure 3: Histograms of a white frame captured under different lighting conditions and distances.

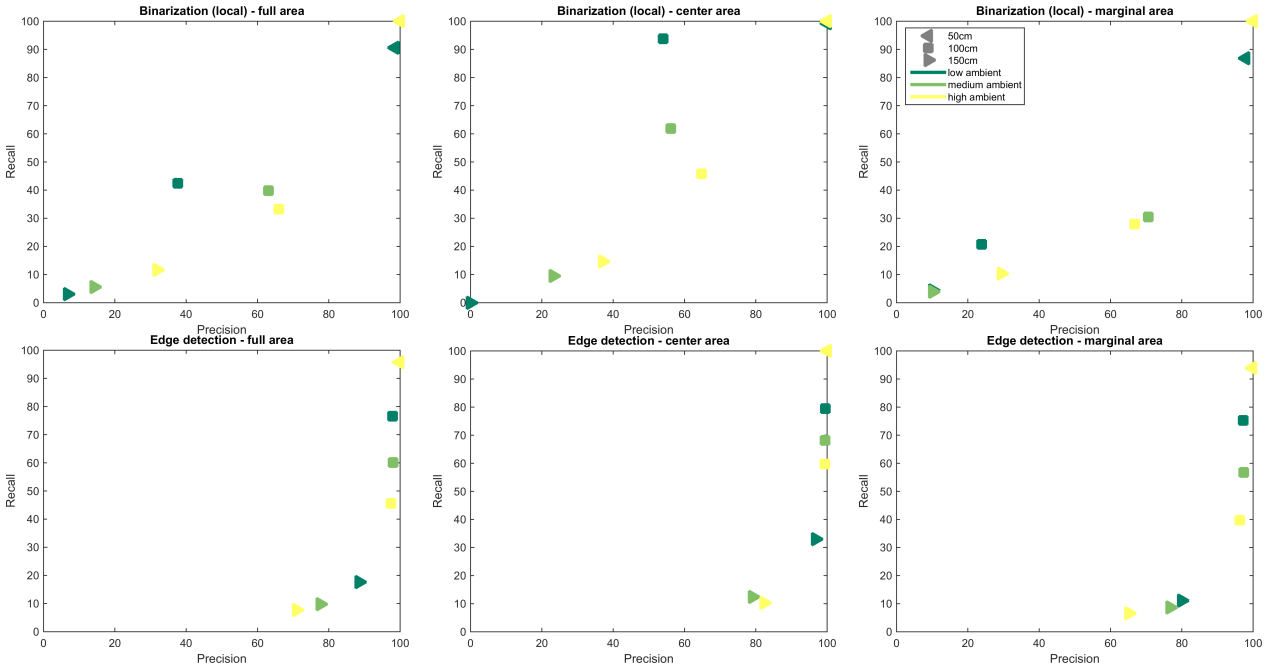


Figure 4: Precision/Recall plots based on dot retrieval using reference pattern. In the first, second and third column, results are presented regarding the whole image, the center and the marginal area, respectively. Results are generated by using local intensity thresholding and edge detection.

5. ACKNOWLEDGMENT

This work has been supported by the Austrian Research Promotion Agency (FFG) under project Robust Depth 3D (project no. 848163).

6. REFERENCES

[1] H. Kawasaki, R. Furukawa, R. Sagawa, and Y. Yagi. Dynamic scene shape reconstruction using a single structured light pattern. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2008.

[2] R. A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, and J. Nissanov. Structured light using pseudorandom codes. *IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI)*, 20(3):322–327, Mar. 1998.

[3] D. Moreno and G. Taubin. Simple, accurate, and robust projector-camera calibration. In *Proceedings of the Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pages 464–471, Oct. 2012.

[4] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, Jan 1979.

[5] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*, 43(8):2666 – 2680, 2010.

[6] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(11):1330–1334, Nov. 2000.