Why can't we purchase a perfect single-shot 3D-sensor?

G. Häusler, C. Faber, F. Willomitzer, P. Dienstbier
Institute of Optics, Information and Photonics, University of Erlangen-Nuremberg, Germany
mailto: gerd.haeusler@physik.uni-erlangen.de

Although highly desired, there is – surprisingly - no optical 3D-sensor that would allow for the acquisition of high quality 3D-movies. Why can't we acquire dense 3D-point clouds within one single camera exposure? We will discuss fundamental physical and information-theoretical obstacles for such a sensor.

1 Introduction

Among the wide spectrum of available optical 3D-sensors there are sensors with high depth resolution, there are sensors that deliver a dense point cloud, and there are (a few) sensors that allow for single-shot acquisition. Why can't we buy a "3D-camera" that comprises all three features? With such a camera we could acquire complex objects just by guiding it free hand around the object or we could even make 3D-movies. We will discuss some fundamental obstacles for such a sensor.

A key for the discussion is the term "dense point cloud". "Dense" means: Each of the N_{pix} camera pixels will deliver a valid 3D-pixel. Thereby we assume for conceptual clarity that the image at the camera chip of the 3D-sensor satisfies the sampling theorem. So we can judge the efficiency η of a sensor by $\eta = N_{3D}/N_{pix}$ with N_{3D} = number of valid 3D-pixels. If $\eta < 1$, a "pseudo dense" surface reconstruction is only possible by posteriori interpolation which implies a reduced lateral resolution of the sensor.

Looking closer at the single-shot solutions claimed for example by Schmalz [1] or Artec [2], it turns out that these solutions indeed lack high lateral resolution and display an efficiency significantly smaller than unity. The resulting artifacts in the acquired point clouds are often non obvious, since many objects are smoothly shaped and display "low 3D-bandwidth". In any case, fine details cannot be acquired by these sensors. Is this fundamental? The bad news is: "yes".

2 The Space-Time-Bandwidth trap

Fringe projection [3] is the paradigm of a sensor with optimal lateral resolution (η=1). Unfortunately, fringe projection is not a single-shot principle; at least 3 exposures are required. The reason is that there are 3 unknowns (ambient light, object reflectivity, fringe phase) to be calculated for each camera pixel. There are real-time workarounds such as described by Takeda [4], which however demands a spatial bandwidth of the object smaller than 1/3 of the otherwise allowed bandwidth. Obviously, in order to buy speed we have to put space-bandwidth on the counter. This may explain why the "magic number 3" will come across more often in the further discussion.

We will now discuss another paradigm, the "perfect" single-shot sensor: light-sectioning triangulation. Instead of fringes, now only one single narrow line is projected onto the object. Along this line a perfect 3D-profile can be calculated from the camera image. Of course, this sensor displays a very small efficiency η. The obvious question is, to what extent can we improve η and what does it cost?

The basic line of thinking can be deduced from the "Flying Triangulation" principle [5]. There, many lines are projected, instead of only one single line. For the present, we will neglect the (severe but hopefully solvable) indexing problem [6] and estimate the maximum possible number of lines: In order to localize each line with sub-pixel accuracy, the lines must satisfy the sampling theorem.

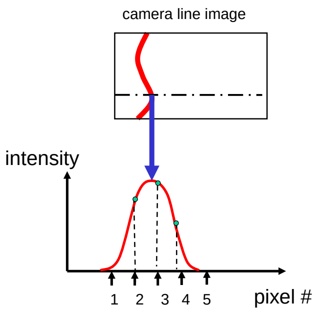

Fig. 1 *Nyquist sampling allows for accurate sub-pixel line localization / high longitudinal precision.*

To be on the safe side, the line width w must be wider than w=4p, with p=pixel pitch (in agreement with experimental experience). Figure 2 reveals that the distance l between two lines must be bigger than l=3p, in order to avoid crosstalk. The resolution of such a sensor in lateral direction (across the lines) will be three-times smaller than with fringe projection.

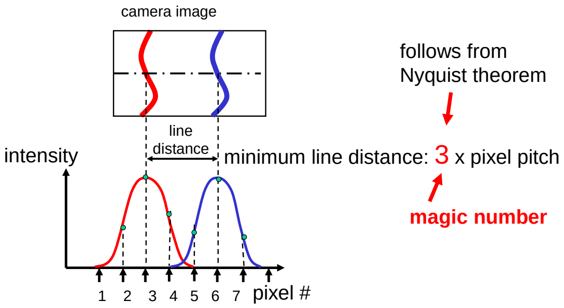

Fig. 2 *Minimum distance between projected lines*

So, a camera with N_x pixels in x-direction will allow for $N_{3D} \sim N_x / 3$ lines, and the sensor can acquire $N_{pix} / 3$ valid 3D-points, within one single shot. Why are we not surprised to find again the magic number 3? It should be noted that for non-flat objects the line distance may appear smaller, by perspective. For inclinations up to about 45°, a reasonable distance is $l \sim 6p$. So multi-line triangu- lation for realistic objects would have a (pessimistic) efficiency not better than $\eta \sim 16\%$, resulting in ~166,000 valid 3D-points (at a square 1-Mpix camera).

What is the consequence of the reduced line density? Obviously, we cannot acquire object details which are “smaller than 3p” (or even 6p) in x-direction.

3 On the way to the optimal single shot sensor

We learned that there is a fundamental limitation for the lateral object bandwidth of single-shot sensors. But we did not yet discuss how to realize a sensor that works at this limit. The crucial question is how to correctly identify (to “index”), e.g., 166 or 333 lines? There are attempts to solve this well-known ambiguity problem, e.g., by encoding the lines [2]. However, line encoding consumes further bandwidth which reduces the maximum number of lines even more.

Another solution does not show this limitation: The “rainbow sensor” [7] projects a color-spectrum onto the object. With a color camera, the correct distance can be acquired for each pixel. So, color-encoded triangulation has an efficiency $\eta=1$. This is possible because we use the three (!) color channels of a three-chip color camera. So we buy faster measurement by more space-bandwidth product (=more pixels). Although the concept has been long known, it is not yet well established - possibly because it is difficult, until now, to project bright saturated spectra. The situation may change by new fiber lasers.

From the three-chip camera it is a small step to ask if we can avoid three modalities and the expensive rainbow projection by using several monochrome cameras and the projection of only one (monochrome) pattern, as in Figure 3. With several cameras, the identification of each projected “line” or “pixel” will become much simpler, if not unique. The implementation of a multi-camera single-shot system will be presented in [8].

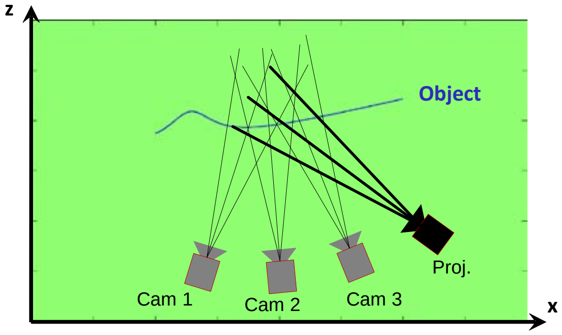

Fig. 3 *From multi-exposure- to multi-camera sensor*

A further system based on Flying Triangulation [5] is presented in [6].

Why are not just two cameras sufficient as for stereo-photography? Passive stereo needs spatial structure to identify corresponding points which again requires additional bandwidth that cannot be used for high-resolution 3D-metrology. For the objective sketched in the introduction, only active systems (with projected patterns) are properly adapted.

It is the aim of our research to understand the physical and information-theoretical limits of such a single-shot 3D-camera, heading for the real-time 3D-video camera.

References

[1] C. Schmalz. Robust single-shot structured light 3D scanning. Dissertation University of Erlangen (2011)

[2] *http://www.artec3d.com/*

[3] M. Halioua, H. Liu, V. Srinivasan. Automated phase-measuring profilometry of 3-D diffuse objects. Appl. Opt. 24 (1984) 3105-3108

[4] M. Takeda and K. Mutoh. Fourier transform profilometry for the automatic measurement of 3-D object shapes. Appl. Opt. 22 (1983), 3977-3982

[5] S. Ettl, O. Arold, Z. Yang and G. Häusler, Flying Triangulation – an optical 3D sensor for the motion-robust acquisition of complex objects, Applied Optics 51 (2012), 281-289

[6] F. Willomitzer, S. Ettl, C. Faber, G. Häusler. Flying Triangulation – towards the 3D-movie camera. arXiv:1305.4168, 2013

[7] G. Häusler und D. Ritter. Parallel 3D-sensing by color coded triangulation. Appl. Opt. 32, No. 35 (1993) 7164-7169

[8] G. Häusler, F. Willomitzer, P. Dienstbier, C. Faber. Tomographic Triangulation. DGaO annual meeting 2013, to appear in Proceedings of the DGaO, 2013