Research Article | Vol. 25, No. 19 | 18 Sep 2017 | OPTICS EXPRESS 23451
Optics EXPRESS

Single-shot 3D motion picture camera with a dense point cloud

FLORIAN WILLOMITZER* AND GERD HÄUSLER

Institute of Optics, Information and Photonics, University Erlangen-Nuremberg, Staudtstr. 7/B2, 91058 Erlangen, Germany
**florian.willomitzer@fau.de*

Abstract: We discuss physical and information theoretical limits of optical 3D metrology. Based on these principal considerations we introduce a novel single-shot 3D movie camera that almost reaches these limits. The camera is designed for the 3D acquisition of macroscopic live scenes. Like a hologram, each movie-frame encompasses the full 3D information about the object surface and the observation perspective can be varied while watching the 3D movie. The camera combines single-shot ability with a point cloud density close to the theoretical limit. No space-bandwidth is wasted by pattern codification. With *1-megapixel* sensors, the 3D camera delivers nearly *300,000* independent 3D points within each frame. The 3D data display a lateral resolution and a depth precision only limited by physics. The approach is based on multi-line triangulation. The requisite low-cost technology is simple. Only two properly positioned synchronized cameras solve the profound ambiguity problem omnipresent in 3D metrology.

OCIS codes: (110.6880) Three-dimensional image acquisition; (150.6910) Three-dimensional sensing; (100.6890) Three-dimensional image processing; (110.0110) Imaging systems; (120.6650) Surface measurements, figure; (120.3940) Metrology; (120.5800) Scanners; (150.2945) Illumination design.

References and links

1. N. L. Lapa and Y. A. Brailov, "System and method for three-dimensional measurement of the shape of material objects," U.S. Patent No. US 7,768,656 B2 (2010).
2. B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth mapping using projected patterns," U.S. Patent Application No. US 2010/0118123 A1 (2010).
3. H. Kawasaki, R. Furukawa, R. Sagawa, and Y. Yagi, "Dynamic scene shape reconstruction using a single structured light pattern," IEEE Conference on CVPR, 1–8 (2008).
4. R. Sagawa, R. Furukawa, and H. Kawasaki, "Dense 3D reconstruction from high frame-rate video using a static grid pattern," IEEE Trans. Pattern Anal. Mach. Intell. **36**(9), 1733–1747 (2014).
5. B. Harendt, M. Große, M. Schaffer, and R. Kowarschik, "3D shape measurement of static and moving objects with adaptive spatiotemporal correlation," Appl. Opt. **53**(31), 7507–7515 (2014).
6. S. Heist, P. Lutzke, I. Schmidt, P. Dietrich, P. Kühmstedt, A. Tünnermann, and G. Notni, "High-speed three-dimensional shape measurement using GOBO projection," Opt. Lasers Eng. **87**, 90–96 (2016).
7. N. Matsuda, O. Cossairt, and M. Gupta, "MC3D: Motion Contrast 3D Scanning," 2015 IEEE International Conference on Computational Photography (ICCP), Houston, TX, 2015, pp. 1–10.
8. W. Lohry and S. Zhang, "High-speed absolute three-dimensional shape measurement using three binary dithered patterns," Opt. Express **22**(22), 26752–26762 (2014).
9. W. Lohry, V. Chen, and S. Zhang, "Absolute three-dimensional shape measurement using coded fringe patterns without phase unwrapping or projector calibration," Opt. Express **22**(2), 1287–1301 (2014).
10. H. Nguyen, D. Nguyen, Z. Wang, H. Kieu, and M. Le, "Real-time, high-accuracy 3D imaging and shape measurement," Appl. Opt. **54**(1), A9–A17 (2015).
11. R. Ishiyama, S. Sakamoto, J. Tajima, T. Okatani, and K. Deguchi, "Absolute phase measurements using geometric constraints between multiple cameras and projectors," Appl. Opt. **46**(17), 3528–3538 (2007).
12. K. Zhong, Z. Li, Y. Shi, C. Wang, and Y. Lei, "Fast phase measurement profilometry for arbitrary shape objects without phase unwrapping," Opt. Lasers Eng. **51**(11), 1213–1222 (2013).
13. C. Bräuer-Burchardt, P. Kühmstedt, and G. Notni, "Phase unwrapping using geometric constraints for high-speed fringe projection based 3D measurements," Proc. SPIE **8789**, 878906 (2013).
14. K. Song, S. Hu, X. Wen, and Y. Yan, "Fast 3D shape measurement using Fourier transform profilometry without phase unwrapping," Opt. Lasers Eng. **84**, 74–81 (2016).
15. G. Häusler and S. Ettl, "Limitations of optical 3D sensors," in *Optical Measurement of Surface Topography*, R. Leach, ed. (Springer, 2011).

#301096 https://doi.org/10.1364/OE.25.023451
Journal © 2017 Received 5 Jul 2017; revised 24 Aug 2017; accepted 25 Aug 2017; published 15 Sep 2017

Optics EXPRESS

16. V. Srinivasan, H. C. Liu, and M. Halioua, "Automated phase-measuring profilometry of 3-D diffuse objects," Appl. Opt. **23**(18), 3105–3108 (1984).
17. M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-D object shapes," Appl. Opt. **22**(24), 3977–3982 (1983).
18. G. Häusler and W. Heckel, "Light sectioning with large depth and high resolution," Appl. Opt. **27**(24), 5165–5169 (1988).
19. F. Willomitzer, S. Ettl, C. Faber, and G. Häusler, "Single-shot three-dimensional sensing with improved data density," Appl. Opt. **54**(3), 408–417 (2015).
20. H. O. Saldner and J. M. Huntley, "Temporal phase unwrapping: application to surface profiling of discontinuous objects," Appl. Opt. **36**(13), 2770–2775 (1997).
21. M. Servin, J. M. Padilla, A. Gonzalez, and G. Garnica, "Temporal phase-unwrapping of static surfaces with 2-sensitivity fringe-patterns," Opt. Express **23**(12), 15806–15815 (2015).
22. S. Ettl, O. Arold, Z. Yang, and G. Häusler, "Flying Triangulation: an optical 3D sensor for the motion-robust acquisition of complex objects," Appl. Opt. **51**(2), 281–289 (2012).
23. F. Willomitzer, S. Ettl, O. Arold, and G. Häusler, "Flying Triangulation - a motion-robust optical 3D sensor for the real-time shape acquisition of complex objects," AIP Conf. Proc. **1537**, 19–26 (2013).
24. X. Laboureux and G. Häusler, "Localization and registration of three-dimensional objects in space—where are the limits?" Appl. Opt. **40**(29), 5206–5216 (2001).
25. G. Häusler and D. Ritter, "Parallel three-dimensional sensing by color-coded triangulation," Appl. Opt. **32**(35), 7164–7169 (1993).
26. J. Geng, "Rainbow three-dimensional camera: New concept of high-speed three-dimensional vision systems," Opt. Eng. **35**(2), 376–383 (1996).
27. C. Schmalz and E. Angelopoulou, "Robust single-shot structured light," IEEE Workshop on Projector–Camera Systems, (2010).
28. M. Young, E. Beeson, J. Davis, S. Rusinkiewicz, and R. Ramamoorthi, "Viewpoint-coded structured light," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), (2007).
29. R. G. Dorsch, G. Häusler, and J. M. Herrmann, "Laser triangulation: Fundamental uncertainty in distance measurement," Appl. Opt. **33**(7), 1306–1314 (1994).
30. G. Häusler, "Ubiquitous coherence - boon and bale of the optical metrologist," Speckle Metrology, Trondheim. Proc. SPIE **4933**, 48–52 (2003).
31. G. Häusler, "Speckle and Coherence," in *Encyclopedia of Modern Optics*, B. Guenther, ed. (Elsevier, 2004).
32. J. Habermann, "Statistisch unabhängige Specklefelder zur Reduktion von Messfehlern in der Weißlichtinterferometrie," Diploma Thesis, University Erlangen-Nuremberg (2002).
33. F. Schiffers, F. Willomitzer, S. Ettl, Z. Yang, and G. Häusler, "Calibration of multi-line-light-sectioning," DGaO-Proceedings **2014**, 12 (2014).
34. YouTube-Channel of the authors' research group: *www.youtube.com/user/Osmin3D*
35. C. Wagner and G. Häusler, "Information theoretical optimization for optical range sensors," Appl. Opt. **42**(27), 5418–5426 (2003).
36. G. Häusler, C. Faber, F. Willomitzer, and P. Dienstbier, "Why can't we purchase a perfect single shot 3D-sensor?" DGaO-Proceedings **2012**, A8 (2012).

1. Introduction

Although highly desired, there is, surprisingly, no optical three-dimensional (3D) sensor that permits the single-shot acquisition of 3D motion pictures with a dense point cloud. There are approaches for real-time 3D data acquisition [1–14], but these are either multi-shot approaches, or they are not 'local', which means that the density of uncorrelated (independent) 3D points leaves room for improvement.

Obviously, the acquisition of a dense 3D point cloud within one single camera frame is difficult. We will discuss the physical and information limits involved. Eventually, we will introduce a novel 3D sensor principle and a 3D camera that indeed combines *single-shot ability* with a *dense 3D point cloud*. The camera works close to the discussed limits: The *lateral resolution* and *depth precision* are as good as physics allows [15]. In other words, the camera performance is limited only by the sampling theorem and by shot noise or speckle noise. Neither color- nor spatial encoding is exploited. The camera displays almost the best possible 3D point cloud density in relation to the available camera pixels. With *1-megapixel* cameras the sensor is able to deliver about *300,000* 3D points within each single shot. The single-shot ability allows for the very fast acquisition of 3D data by flash exposure. This enables the capture of 3D motion pictures in which each camera frame includes the full-field 3D information, with free choice of the viewpoint while watching the movie (see Visualization 1).

Optics EXPRESS

To the best of our knowledge, no such 3D camera is currently available. Along the wide spectrum of optical 3D sensors, there are sensors with high precision, there are sensors that deliver a dense point cloud, and there are (a few) sensors that allow for single-shot acquisition of only sparse data. What are the obstacles for a single-shot 3D camera that offers dense 3D data and physically limited precision at the same time?

The key term is the *"dense point cloud"*. We could naively demand that each of the N_{pix} camera pixels delivers a 3D point, completely independent of its neighbors. However, to avoid aliasing, we have to ensure that the image at the camera chip satisfies the sampling theorem, so a certain correlation between neighboring points is unavoidable. This is what has to be remembered when "*point cloud density*" is discussed: 100% is impossible because it contradicts linear systems theory. Indeed, all 3D sensors display artifacts at sharp edges, where the sampling theorem is violated. Satisfying the sampling theorem is also necessary in exploiting subpixel interpolation for high distance precision, as depicted in Fig. 1(a).

Keeping in mind that 100% is impossible, we nevertheless calculate the "*point cloud density*" ρ of a 3D sensor from $\rho = N_{3D}/N_{pix}$, were N_{3D} is the number of independent 3D pixels and N_{pix} is the number of pixels on the camera chip. For a *1-megapixel* camera (*1000 × 1000 pixels*), a density of, say, *30%* will yield *300,000* independent 3D points. A low-density point cloud implies a reduced lateral resolution of the sensor. For $\rho < 1$, and more so for $\rho << 1$, only a "pseudo dense" surface reconstruction is possible – which is commonly prettified via *a posteriori* interpolation and high-resolution texture mapping. Looking more close at existing single-shot solutions, such as those obtained by Artec [1] and Kinect One [2], we find that they lack high lateral resolution. The reason for the low density is that any type of triangulation requires the identification of corresponding points, whether this be for classical stereo or for the methods described above. The necessary encoding devours space-bandwidth that is lost for high lateral resolution. In [1], for example, the width of the projected stripes is encoded piecewise in the stripe direction. In [3,4], the period of projected lines is encoded and combined with different colors (the pros and cons of color encoding are discussed below). In [2], a pseudo-random pattern of dots is projected. Classical stereo exploits "natural" salient but spacious "features." Is this fundamental for single shot sensors? The bad news is: *yes*.

To better understand this phenomenon, let us start with the paradigm multi-shot principle that delivers a point cloud with virtually 100% density, the so-called "fringe projection triangulation", sometimes called "phase-measuring triangulation" (PMT) [16]. This principle is local. Each pixel delivers information widely independent from its neighboring pixels (within the limits of the sampling theorem). As mentioned, this feature has its price: At least *three* exposures are required for one 3D height map because the local distance has to be deciphered from the *three* unknowns—the ambient illumination, object reflectivity, and the fringe phase—individually for each camera pixel. The three (or more) exposures are commonly taken in a way to later decipher the data by phase shifting algorithms. From only one exposure, this is impossible. Single-shot workarounds, such as single sideband demodulation [17], were suggested. However, these demand a spatial bandwidth of the object smaller than *1/3* of the system bandwidth to avoid overlap of the base band with the modulated signal. This restriction is even more severe if the carrier frequency is not equal to the optimal *2/3* of the system bandwidth. The same argument, by the way, holds for holographic "3D imaging", where the carrier frequency must be sufficiently large, to avoid overlap of the reconstructions and the base band. The arguments above clarify: A single-shot 3D camera for arbitrary surface shapes with a pixel-dense point cloud is impossible if no additional modality is exploited. One camera pixel can principally not deliver sufficient information.

So far about a multi-shot sensor with the best possible data density. At the other end of the spectrum of sensors is the "perfect single-shot camera": light-sectioning triangulation [18]. Instead of fringes, only one narrow line is projected onto the object. Along this line, a perfect 3D profile can be calculated from one camera image.

Optics EXPRESS

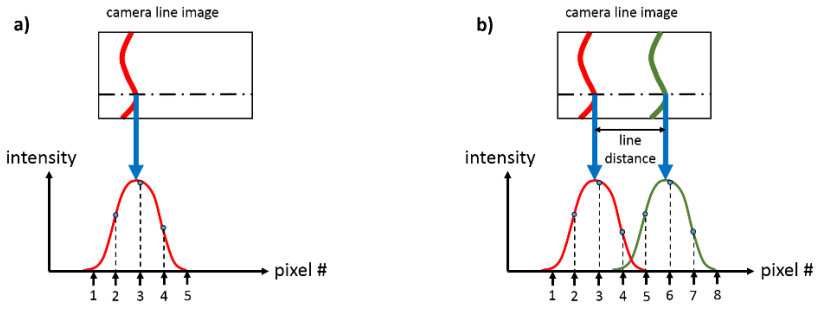

Fig. 1. (a) Nyquist sampling allows for precise sub-pixel line localization and high distance precision. (b) The minimum distance between projected lines is three times the pixel pitch.

Note that even light sectioning with only one line is not perfectly local, as the calculation of the line position with subpixel precision exploits at least three neighboring pixels perpendicular to the line direction, as illustrated in Fig. 1(a). It should be noted as well that light sectioning allows for high lateral resolution along the line direction. Of course, light sectioning with only one line displays a very low data density, e.g., $\rho \approx 0.1\%$, for the 1-megapixel camera example. According to the profound ambiguity problem [19] *multi-line light sectioning* principles are commonly not able to project more than about 10 lines at the same time ($\rho \approx 1\%$), if no line codification is exploited. It should be mentioned that the ambiguity problem occurs in multi-shot fringe projection as well. Here, the ambiguity is commonly solved by temporal phase unwrapping [20,21] which, however, requires even more exposures.

There is a workaround to this problem, we called it "Flying Triangulation" [22,23]. Flying Triangulation involves the use of a single-shot sensor with about *10* projected lines. The data from each exposure are sparse, but the gaps between the measured lines can be filled within seconds by on-line registration of subsequent exposures, while the sensor is guided (even by hand) around the object. Eventually, Flying Triangulation delivers dense high-quality data, even of moving objects. However, as the dense point cloud is accumulated by subsequent exposures, it demands for rigid objects: speaking or walking people cannot be acquired. The obvious question is: how many lines can be maximally projected in order to obtain a high point cloud density already within each single shot?

If the significant ambiguity problem is neglected for a moment, one can estimate the maximum possible number of lines. The considerations are illustrated by Fig. 1(b). To localize each line with the best subpixel accuracy, the line images must be as narrow as possible [24] but wide enough to satisfy the sampling theorem. And there must be sufficient space between the lines. With Fig. 1 we find that for subpixel interpolation, the linewidth must be wider (but not much wider, to avoid overlap) than $4p$ (where p is the pixel pitch). With a half-width of a little more than $2p$, precise subpixel interpolation is ensured. These numbers are consistent with theoretical and experimental experience [24]. We note as well that the line image at the three evaluated pixels (Fig. 1(a)) must not be disturbed by abrupt variation of the object height or texture. This tells us again that it is not possible to acquire completely independent data within a small area.

Figure 1(b) shows that the distance between two lines must be at least $3p$ for low crosstalk. Fine details between the lines are not resolved, and, with proper band limitation, should not occur. Consequently, a camera with N_x pixels in the x-direction permits a maximum of $L \approx N_x/3$ lines, and the sensor can acquire $N_{pix}/3$ valid 3D points within one single shot, yielding a density of $\rho \approx 33\%$ or *330,000* 3D pixels with a *1-megapixel* camera.

Why are we not surprised to find this limit for multi-line triangulation? Considerations in object space (see fringe triangulation) tell us that there are three unknowns to be deciphered;

Optics EXPRESS

considerations in Fourier space (single sideband encoding) tell us that we must limit the object bandwidth to less or equal than *33%* of the available system bandwidth.

Obviously, we have to sacrifice at least 2/3 of the available space-bandwidth. This may explain why the magic number *"3"* is frequently encountered in this paper.

However, the absolute limit of ρ ≈33% can only be reached for relatively flat, untilted surfaces. For highly tilted areas and a large triangulation angle, the line distance in the camera image may shrink according to perspective contraction. This requires a larger line distance of *6p* to *7p* to be projected. The resulting density of $\rho \approx 16\%$, leads to *160,000* 3D points for the *1-megapixel* camera. We demonstrate herein that this density is realistic for objects with significant depth variation and it can be technically achieved without extreme requirements for the calibration and mechanical stability.

As to the crucial question of how to correctly identify (i.e., how "to index"), say, *330* lines - or more modestly, *160* lines, - we conclude that this formidable problem cannot be solved by spatial encoding of the lines if we want to exploit the full channel capacity for the acquisition of precise, high-resolution 3D data.

The two core questions to be answered by this paper are therefore:

- To what extent can we improve ρ for light sectioning, and what does it cost?
- How to make a sensor that approaches the maximum possible data density, without loss of precision and lateral resolution.

There are a few nearly perfect solutions exploiting color or time of flight as an additional modality. The so-called "rainbow sensors" [25,26] use a projected color spectrum to encode the distance via triangulation. A color camera decodes the shape from the hue of each pixel. Color-encoded triangulation may have a density of $\rho = 100\%$ if a three-chip color camera is employed for the acquisition (no spatial interpolation of the Bayer-pattern). We notice, by the way, that there are three color channels in a three-chip color camera, which permit faster measurement by virtue of greater space-bandwidth (= more pixels). Although the concept of rainbow sensors (and other color encoding approaches) has long been known, it is not yet well established, possibly because it prevents color texture acquisition in many cases [10,27]. Another, nearly "single-shot" solution exploits the time of flight (TOF) from the camera to the object and back. For each pixel, the distance is deciphered from a fast temporal sequence of exposures by temporal phase shifting. TOF is not a genuine single shot principle, but the image sequence can be taken quite fast. So the method is virtually suited for non static objects such as walking people. A popular implementation is the latest Microsoft Kinect sensor, which, however displays poor precision in the millimeter range and a limited number of pixels, not allowing photo realistic 3D images or precise 3D metrology.

From the *"three chip"* camera for color encoding it is a small step to ask whether we can replace the three (red, green, and blue) sensors using a couple of synchronized black-and-white cameras. With a multitude of cameras, the identification of each projected "line" or "pixel" may become much easier.

The idea of using many cameras was proposed in [28] approximately 10 years ago, and a principal solution was demonstrated. The authors project a pattern with binary stripes onto the object, and C cameras are required to distinguish 2^C depths. By selecting the triangulation angles properly (in exponential sequence), each stripe can be uniquely identified.

This was (to the best of our knowledge) the first "proof of principle" for a single-shot 3D camera with a potentially dense point cloud. It demonstrates that unique triangulation can be achieved from several images *obtained at the same time* - as opposed to a timed sequence of images. As explained above, this method may improve the density of the 3D point cloud without however reaching the *100%* of phase-measuring triangulation.

Approaches which exploit several synchronized cameras have been suggested for the purpose of reducing the number of sequential exposures without attempting to obtain a single-shot solution. For example, multiple cameras are exploited for PMT, to speed up measurements by avoiding multi-frequency phase shifting [11–13]. The method described in

Optics EXPRESS

[14], based on Fourier transform profilometry, exploits two cameras to facilitate the phase matching.

How does the idea proposed in [28] match our considerations above? For a setup comparable to ours (*160* lines, *>500* distinguishable distances), the method described in [28] requires a multitude of cameras. We demonstrate in this paper that *two* cameras are sufficient to measure up to *300,000* 3D pixels. Moreover, due to proper subpixel interpolation, the precision of our method is limited only by coherent noise or electronic noise and not by the number of cameras.

2. A single-shot 3D movie camera with unidirectional lines

As discussed in the previous section, a single-shot principle does not supply sufficient information to provide data with *100%* density (disregarding "rainbow triangulation"). If pattern encoding comes into play, the density will be even less. Our single-shot camera is based on *multi-line triangulation*, without any encoding to identify the lines - the decoding is performed just by combining the images of two properly positioned cameras. Compared to approaches that rely on line encoding, our approach preserves fine details and edges because no additional space bandwidth is consumed.

Two approaches with different projected patterns are described. The first approach exploits a projected pattern of straight, narrow lines (*160* lines for a 1-megapixel sensor). The object is observed from different triangulation angles by *two* cameras. The projected slide displays binary lines. The projected lines at the object surface are low-pass filtered by the projector lens, which helps to satisfy the sampling theorem, as discussed above.

The next question is how to manage the necessary "indexing" of that many lines. The corresponding ambiguity problem was discussed in a previous paper [19]. As the novel solution is an extension of these results, they are briefly summarized:

For common multi-line triangulation, the achievable line density $L/\Delta x$ (where L is the number of projected lines and Δx is the width of field; see Fig. 2) is related to the triangulation angle θ and the unique measurement depth Δz as described by the following expression:

$$\frac{L}{\Delta x} \le \frac{1}{\Delta z \cdot \tan\theta}. \tag{1}$$

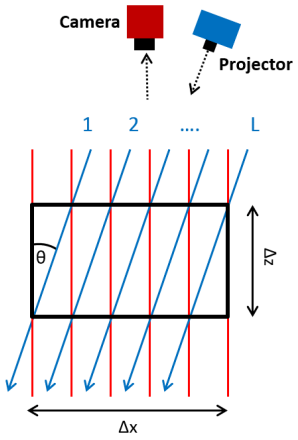

Fig. 2. The achievable number of lines L depends on the triangulation angle θ and the unique measurement depth Δz.

A violation of Eq. (1) results in false 3D data, observed as outliers (see Fig. 4(e)). Reference [19] explains how these outliers can be detected and corrected using data from a

Optics EXPRESS

second camera positioned at a second angle of triangulation. The basic idea is to (virtually) project the data from the first camera back onto the camera chip of the second camera. The correctly evaluated data can be detected, as they necessarily coincide at the camera chip (but, commonly, not the outliers). However, with increasing line density, more outliers from one camera accidentally coincide with data from the second camera, and the achievable (unique) line density is only moderate. As in Flying Triangulation, registration of several frames is required for sufficient density.

Here an effective improvement of the "back-projection" idea is introduced which allows for an about 10 times higher line density without generating any outliers, by the proper choice of a small and a large triangulation angle. In contrast to [19], where outliers were detected with a probabilistic method, this new approach allows for a deterministic identification of the line indices. We demonstrate that thoughtfully designed optics for illumination and observation, in combination with moderately sophisticated software, can solve the problem.

The basic idea is as follows: a narrow line pattern with L lines is projected onto the object. The object is observed by two cameras (C_1 and C_2) at two triangulation angles θ_1 and θ_2 (see Fig. 3). The first camera C_1 and the projector P create a triangulation sensor with a very small triangulation angle θ_1. This first sensor delivers noisy but unique data within the demanded measurement volume Δz_1, according to Eq. (1). The data are noisy, as the precision is $\delta z \sim 1/(SNR \sin\theta_1)$ [29], with SNR being the signal-to-noise ratio (the dominant source of noise for line triangulation is, commonly, speckle noise [29–32]). The second camera C_2 and the projector create a second triangulation sensor with a larger triangulation angle θ_2. This second sensor delivers more precise but ambiguous data. As both sensors look at the same projected lines, the first sensor can "tell" the second sensor the correct index of each line, via a back-projection mechanism similar to that described in [19].

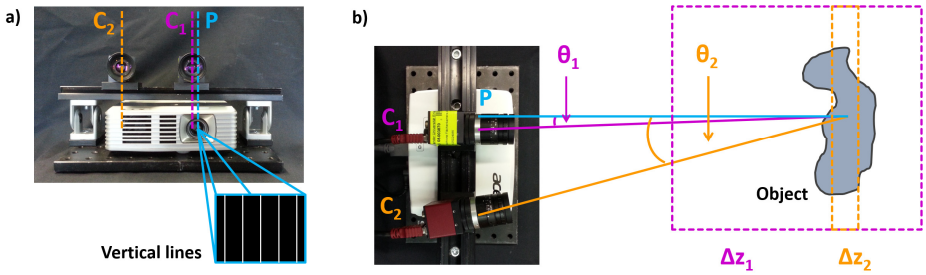

Fig. 3. Setup comprising a projector P that projects a static line pattern and two cameras C_1 and C_2. For a vertical line pattern, the horizontal distance between the nodal points of the projector and the cameras define the related triangulation angles θ_1 and θ_2. (a) Front view of the setup. (b) View from top, illustrating related angles and measurement volumes.

The evaluation procedure is illustrated in Fig. 4. The observed line images deviate from a straight line, depending on the triangulation angle. The lines seen by camera C_1 are nearly straight and can be easily indexed, as shown in Fig. 4(b). In the sketch, the index is illustrated by a color code. The directly calculated 3D model (Fig. 4(d)) displays correct indexing but high noise. 3D points, directly calculated from the image of C_2, display low noise but ambiguity errors (Fig. 4(e)). To solve this problem, both sets of information are merged: the points of Fig. 4(d), including their index information, are back-projected onto the chip of C_2. With precise calibration [33], the back-projections overlap with the line signal (Fig. 4(f)). Eventually, the back-projected line indices of C_1 are assigned to the corresponding lines on the chip of C_2 (Fig. 4(g)), leading to unique data with high precision (Fig. 4(h)).

Optics EXPRESS

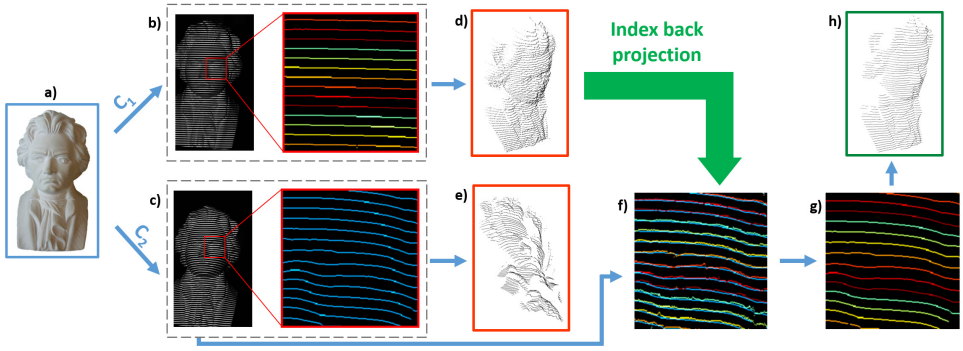

Fig. 4. Unique indexing by combining two camera images: (a) Object. (b) and (c) Images of the object with two cameras, seen from different triangulation angles (known indices in (b) are color coded). (d) and (e) 3D data, directly calculated from (b) and (c). (f) Noisy 3D data from (d), correctly indexed (color coded), back-projected onto the chip of C_2, together with the line image of C_2. (g) Correctly indexed lines of C_2, assigned from the indices delivered by C_1 (color coded). (h) Final 3D model evaluated from (g) with correct indices and low noise.

As the reader might guess from Fig. 4(f), θ_1 and θ_2 cannot be chosen independently. Noise has to be taken into account. The correct index can be assigned uniquely if the back-projected noisy line images of C_1 do not crosstalk with the neighboring line images on C_2. More precisely, the back-projected (noisy) lines should not overlap with lines other than the corresponding lines seen by the second camera. This is the case if Eq. (2) is satisfied:

$$d_2' > 2\delta x' \cdot \frac{\sin\theta_2}{\sin\theta_1}. \qquad (2)$$

Here, $\delta x'$ represents the image-sided uncertainty of the line localization and d'_2 is the line distance in the camera image of C_2.

To demonstrate the robustness of the principle against locally varying object texture and reflectivity, we measured a "natural" object: human faces. Figure 5 displays raw data from a single-shot acquisition of a human face, with *160* projected lines, acquired with a camera resolution of *1024 × 682 pixels*. This corresponds to a 3D point density of $\rho \approx 16\%$. Figure 5(b) displays the acquired 3D data of the object (Fig. 5(a)). Note that all perspectives are extracted from the same single video frame. Black-and-white texture information is included in the 3D data. Examples comprising color texture are shown in Fig. 11 and in [34]. In Fig. 5(c), a close-up view illustrates the low noise, which will be discussed in detail in the next section. We emphasize that the displayed 3D models are *not* post-processed, no interpolation or smoothing was applied. Each displayed 3D point was measured independently from its neighbors. As discussed, the absence of spatial encoding strategies allows for edge preservation, as demonstrated in Fig. 6.

Fig. 5. Raw data (without post-processing) from a single-shot measurement with a line density of *~160 lines/field.* (a) Camera image of human face with projected lines. (b) 3D model from different perspectives, evaluated from one single video frame. (c) Close view of the 3D data, illustrating the low noise (*~200µm*).

Optics EXPRESS

With the novel single-shot method, object surfaces can be measured with considerable density and a precision that is only limited by coherent or electronic noise [24,29]. The time required for a single measurement is as short as the exposure time for one single camera frame. A static binary pattern (which can be a simple chrome-on-glass slide) is projected. No electronically controlled pattern generator is required. Hence the 3D data can be acquired in milliseconds or microseconds, only limited by the available illumination. Each camera frame delivers a 3D model, so 3D motion pictures can be acquired. Examples are shown in section 4, in Visualization 2, and in [34].

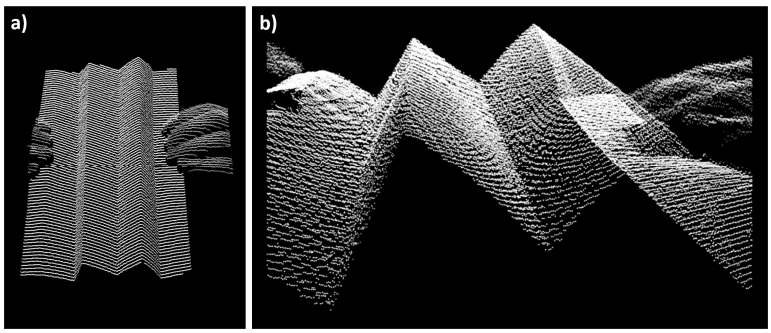

Fig. 6. Single-shot measurement (raw data) of a folded piece of paper. Edges are preserved, as the full space bandwidth can be exploited, there is no spatial line encoding. (a) 3D model of the folded paper. (b) Close-up view. The object was continuously moved during the measurement. A video can be found in Visualization 3 or [34].

3. Precision

In this paper we essentially discuss the key feature of our novel sensor: the single shot ability. The discussion so far was about data density and speed, not about precision. The reader, however, might suspect that the high data density is dearly bought by sacrificing precision. The following section demonstrates that the sensor is able to reach a precision close to the limit of what physics allows and that it is competitive to the paradigm "phase-measuring triangulation". We refer to earlier research about the physical limits of 3D sensing, see e.g [15,29–32,35,36].

The ultimate limit of the distance precision δz_{coh} for triangulation at rough surfaces [29] is caused by coherent noise, according to Eq. (3):

$$\delta z_{coh} = \frac{C_s}{2\pi} \cdot \frac{\lambda}{\sin u_{obs} \cdot \sin \theta}, \tag{3}$$

where C_s is the speckle contrast or inverse *SNR* (C_s = *1* for laser triangulation), u_{obs} is the observation aperture and θ is the triangulation angle. The origin of Eq. (3) is the principal uncertainty to localize the (speckled) image of a projected laser spot better than

$$\delta x'_{coh} = \frac{C_s}{2\pi} \cdot \frac{\lambda}{\sin u'_{obs}}, \tag{4}$$

with $\delta z_{coh} = \delta x'_{coh}/(\beta' \sin\theta)$. The image sided observation aperture is u'_{obs} and β' is the magnification. The limit $\delta x'_{coh}$ is independent of the specific sensor geometry, so we can easily compare sensors with different stand-off and field of view.

The influence of coherent noise is illustrated in Fig. 7. Two line profiles are shown, acquired with the same image sided observation aperture: One line, projected with coherent laser illumination (Fig. 7(a)) and one line with reduced coherence (Fig. 7(b)). For both lines, the raw image and the lateral variation of the line maximum (with sub pixel precision) is

Optics EXPRESS

shown, the latter leading to the depth precision. The depth precision is calculated for the parameters of our prototype setup (see below).

The line image in Fig. 7(b) displays much lower speckle contrast, which leads to better distance precision. We note that *"incoherent illumination"* means more than just applying a broad band light source. At metal surfaces (which are pure surface scatterers) an effective reduction of the speckle contrast can only be achieved by reducing the *spatial coherence* [31]. Measurements of volume scatterers such as skin or plastics can be further improved by exploiting low temporal coherence [31].

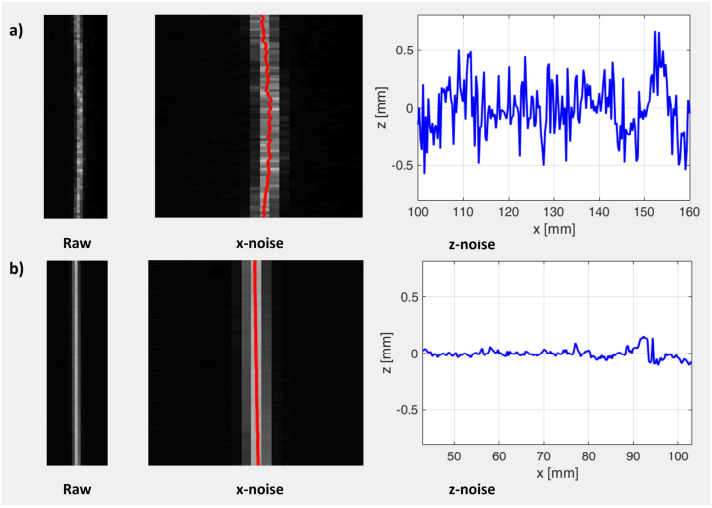

Fig. 7. Line images and noise. Left column: (a) Line image with laser illumination. (b) Line image with reduced spatial and temporal coherence. Center column: magnified line images with sub-pixel maximum location. Right column: calculated distance with uncertainty.

The prototype of our 3D camera displays an observation aperture of sin $u_{obs} = 0.00225$, a triangulation angle $\theta_2 = 9°$ and a horizontal field width of $\Delta x = 300\ mm$ at the center of the measurement range (at $z = 500\ mm$). With fully coherent illumination ($C_s = 1$), the achievable distance precision would be not better than $\delta z_{coh} = 249\ \mu m$ (see Eq. (3)).

The distance precision after optimization of the setup is found from the following experiment: A planar screen (laminated with thin white paper) is oriented perpendicular to the optical axis of the projector, within the measurement volume. The screen is illuminated with a pattern of *142* projected vertical lines. Measurements are taken at *11* different depths within the measurement volume (from $z = 450\ mm$ to $z = 550\ mm$) and the precision δz of each line is calculated from the standard deviation $\delta x'$ along each line. Around the distance $z = 500\ mm$, the best precision along an entire line is found to be $\delta z_{min} = 23\mu m$ (the very best precision in the entire measurement volume is $\delta z_{min} = 17\mu m$).

So far we achieved this high precision only around the center of the field, because the sub pixel precise evaluation of the line maximum is sensitive to line broadening by aberrations [24]. With low aberration optics the physical precision limit will be achievable in the entire field. The "worst" precision in the entire volume is better than $\delta z_{max} \leq 180\mu m$ for the planar screen and $\delta z_{max} \leq 200\mu m$ for human skin (the latter due to line broadening by volume scattering).

For the further discussion we consider it justified to use the best precision, as we are interested in the physical limit (rather than in the technical imperfections of our projector- and camera lens). The measured precision $\delta z_{min} = 23\ \mu m$ at $z = 500\ mm$ is about *10* times better than the coherent limit of *249* μm. This has been achieved by reduction of speckle noise, according to the optimization steps described below:

Optics EXPRESS

i) Reduction of spatial coherence: by exploiting an illumination aperture (sin $u_{ill} = 0.0075$), about three times larger than the observation aperture (sin $u_{obs} = 0.00225$), the speckle contrast is reduced [31] by a factor

$$c_{spat} = \frac{\sin u_{obs}}{\sin u_{ill}} = 0.3. \tag{5}$$

ii) At a depolarizing surface (such as our screen), illuminated with unpolarized light (LED projector), the speckle contrast is reduced [32] by a further factor of

$$c_{pol} = 0.5. \tag{6}$$

iii) If the pixels are larger than the speckle size, the speckle noise is reduced via averaging (at the expense of lateral resolution). We chose a pixel size ($d'_{pix} = 4.65\mu m$) approximately the size of the subjective speckles, not to lose resolution and get

$$c_{pix} = \frac{\lambda}{d'_{pix} \cdot \sin u'_{obs}} = 0.95. \tag{7}$$

Combining steps i) – iii), we achieve a reduction of speckle noise by a factor of $c_{spat}\, c_{pol}\, c_{pix} = 0.14$, which explains a seven times improvement of the precision, compared to the coherent limit. In fact, the achieved precision is even better. This further improvement is due to volume scattering in connection with temporal incoherence. The quantitative contribution of temporal incoherence could be calculated in principle [31], but in fact, electronic noise is taking over the role of the dominant source of noise, if the speckle contrast is very low. Moreover, line broadening due to volume scattering partially counterbalances the positive effects of speckle reduction [24].

We conclude: In the center of the measurement volume, our prototype setup is able to display a precision of $\delta z_{min} = 23\mu m$, about ten times better than the coherent limit. This means that within the measuring range of $\Delta z = 100\ mm$, more than *4000* depth steps could be distinguished which competes with the most high quality 3D cameras available at the market. Note that sensors with better precision commonly exploit many exposures for noise averaging.

4. With crossed lines toward higher point density

In the introduction, the maximum number of lines was estimated to be about *160 lines* for a *1-megapixel* camera and realistic 3D scenes. From Fig. 8(a) it is obvious that it might become very difficult to implement more lines.

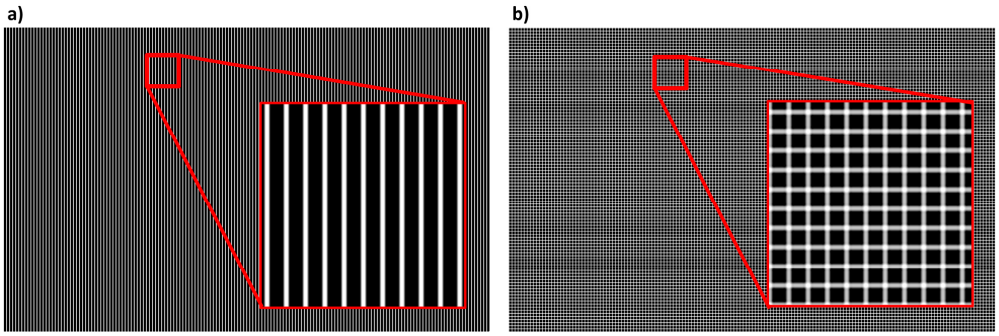

Fig. 8. Original patterns to be projected onto the object visualize the density of 3D points (with magnification windows). (a) Unidirectional line pattern (see section 2). (b) Crossed lines for higher point density. The reader can zoom in to resolve the patterns.

Optics EXPRESS

However, there is another option for more 3D points: our first approach can be upgraded by the projection of *crossed lines*. Figure 8(b) displays the *original* pattern to be projected, with *160* vertical lines and *100* horizontal lines, based on the aspect ratio (*16:10*) of the projector.

After image acquisition, the two line directions are identified, isolated, and separately evaluated. Principally, a second pair of cameras could be added to evaluate the second perpendicular line direction. However, there is a simpler and more cost effective solution, requiring *only two cameras* instead of four. As shown in Fig. 3(a), only the distance from the camera and the projector perpendicular to a line direction defines the triangulation angle. For a crossed line pattern, we can generate two different triangulation angles for each camera. The resulting setup (see Fig. 9) has *four* triangulation angles - one large and one small angle for each line direction. With only two cameras, we create *four* triangulation sensors.

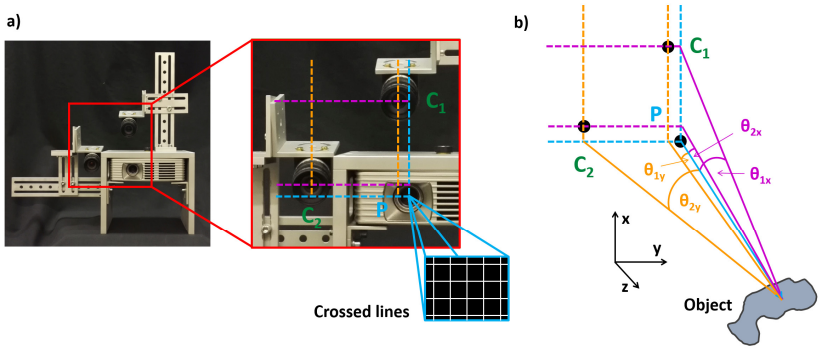

Fig. 9. Setup with crossed line projection: Two cameras C_1 and C_2 and the projector P produce *four* independent triangulation angles: θ_{1x}, θ_{1y}, θ_{2x}, and θ_{2y}. Each line direction is evaluated separately. (a) Front view of the setup. (b) Perspective sketch with triangulation angles.

Principally, this can be performed with even more cameras and more line directions. Such a setup with C cameras and D line directions could produce $C \times D$ triangulation sub-systems. A setup as shown in Fig. 9 ($C = D = 2$) with *160* vertical and *160* horizontal projected lines is able to acquire nearly *300,000* 3D points from a single frame of a *1-megapixel* camera (crossing points are counted only once). Again, it turns out that a proper optical setup makes things easy.

What is the cost of the increased number of 3D pixels? The identification of the line direction requires some (not too serious) restriction of the surface shape: to distinguish between different directions, a small line segment has to be visible, which requires some neighborhood and a certain "smoothness" of the surface. This means that not all measured 3D points are completely independent of the neighborhood anymore. We add that it is advantageous to increase the intensity of the projected pattern at the crossing points. So, the line position can still be evaluated in both directions. However, this reduces the signal-to-noise ratio at the other line segments, which reduces the precision.

Figure 10 displays frames of a 3D movie, acquired with the setup of Fig. 9. The different perspectives are each extracted from one video frame. Figure 10(b) and (c) illustrate the low noise with a close up view. Again, all figures display unprocessed raw data.

At least one color camera must be used to for the acquisition of color texture. It is possible to acquire color texture with an auxiliary color camera, using additional flash exposure. This does not constitute "single-shot" acquisition, which is why we acquire color texture directly from the line images. The simplest solution is to replace the two black-and-white cameras by two color cameras. In this case we condone spatial interpolation by the RGB Bayer pattern cameras. Figure 11 displays three different video frames, taken from a color 3D movie. The full movie can be seen in Visualization 5 and [34].

Optics EXPRESS

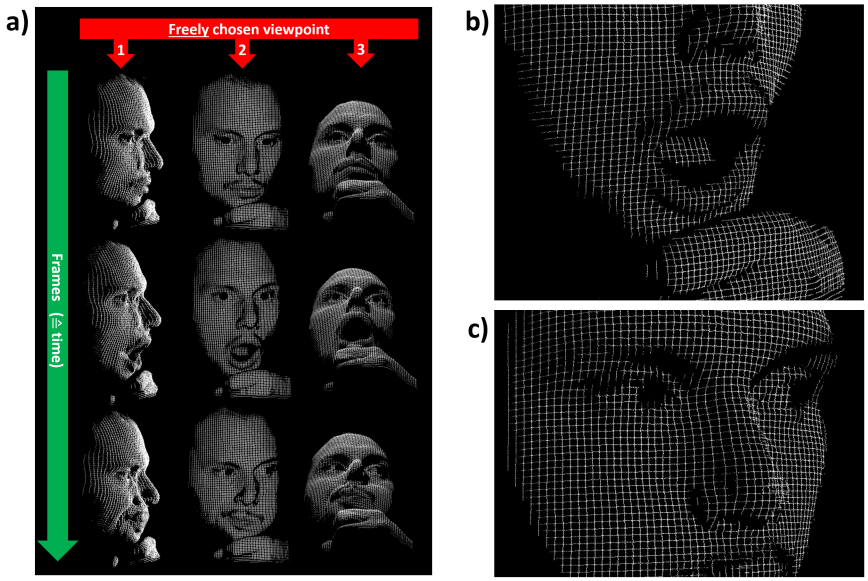

Fig. 10. (a) Single frames of a 3D movie of a "talking face", acquired with crossed lines single-shot triangulation. The different perspectives (viewpoints 1, 2, and 3) are all taken from the same corresponding video frame. (b) and (c) Close look at the 3D data, illustrating relatively low noise. The corresponding motion picture can be seen in Visualization 4 and in [34].

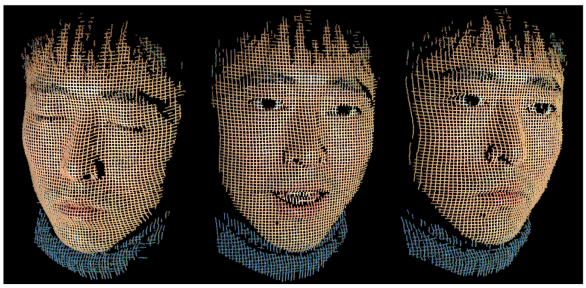

Fig. 11. Three frames from a color 3D movie. The monochrome cameras were replaced by color cameras, and the color texture was acquired along the projected lines from the same frame as the 3D data. The movie can be seen in Visualization 5 and in [34].

5. Summary

This paper presents a single-shot 3D camera concept and device for the acquisition of up to *300,000* 3D points within each single camera frame of two synchronized *1-megapixel* video cameras. This number is close to the possible maximum, as theoretical estimations reveal. The 3D camera exploits triangulation with a pattern of *160* unidirectional lines or with a pattern of crossed lines with the same pitch. The fundamental problem of unique line identification is solved by combining the images from two cameras: one with a very small triangulation angle and the other with a large triangulation angle.

The 3D camera is technically simple. Special care is given to the proper geometry of the optics and illumination and to obeying the sampling theorem. The precision is limited only by coherent or electronic noise. The precision is better than *1/500* of the distance measuring range ($\delta z \leq 200\ \mu m$ for the prototype setup).

The time for the acquisition of a 3D scene is limited only by the camera exposure time and the available illumination level (for static pattern projection). Visualization 6 illustrates the motion of a bouncing ball recorded with a camera frame rate of *30 Hz* and an exposure time of *~5 ms*. More videos are available on our YouTube channel [34].

Optics EXPRESS

The computational effort is moderate: *30 Hz* recording and display with interactive choice of the perspective seems possible in real time.

6. A retrospective aha-experience

By reviewing the images of both cameras (see Fig. 12), a striking resemblance with image plane holograms can be noticed. Indeed, the phase of the lines ("fringes") encodes the depth, as in a hologram. The 3D image from Fig. 12(a) could even be optically reconstructed by laser illumination. After eliminating the base band and the second diffraction order, the phase of the reconstruction represents the surface in 3D space. Of course, the phase has to be re-scaled: a 2π phase shift corresponds to a distance $\Delta z_1 = \Delta x/(L \tan\theta_1)$ (see Fig. 2 and Eq. (1)). The virtual "wavelength" is $\lambda_1 = 2\Delta z_1$.

This will work for Fig. 12(a), but not for the image in Fig. 12(b). Due to the large triangulation angle, the phase modulation in Fig. 12(b) is much larger than 2π, and a unique object reconstruction is impossible without "phase unwrapping".

After all, the first sensor, with the small triangulation angle θ_1, is the key component: it serves as a *"phase compressor"* that enables the acquisition of objects with large depth variation. The first sensor, in combination with the second sensor, exploits concepts of holography and two-wavelength interferometry, *here for rough, macroscopic objects* (the second wavelength is $\lambda_2 = 2\Delta z_2 = 2\,\Delta x/(L \tan\theta_2)$.

Fig. 12. Line images seen by C_1 (small θ_1) and C_2 (large θ_2). (a) The line image of C_1 displays a unique "phase distortion" ($< 2\pi$) and can be considered a perfect image-plane hologram of the object surface. (b) The line image of C_2 displays a large phase modulation ($> 2\pi$) that does not allow for simple unique decoding.

We conclude with a heavy heart, that the space-bandwidth constraints of single-shot principles have to be accepted. There is a little comfort, as state-of-the-art video cameras supply a plethora of pixels. Figures 5, 8, 10 and 11 indicate that even a *1-megapixel* camera can yield more than a hundred thousand 3D pixels which is sufficient for 3D metrology with significant lateral resolution. Cameras with many more pixels are available, and full-HD quality will be achievable.

Acknowledgments

This paper is essentially about principles and limits, as the sensor works with simple technology. However, the sensor would not work at the limits, without precise calibration. We want to acknowledge the invaluable contributions of Florian Schiffers, who was involved in many fruitful discussions, and we wish to acknowledge specifically his assistance in the calibration [33].