

Optical Engineering

SPIDigitalLibrary.org/oe

Fast data acquisition for three-dimensional shape measurement using fixed-pattern projection and temporal coding

Marcus Grosse
Martin Schaffer
Bastian Harendt
Richard Kowarschik



SPIE

Fast data acquisition for three-dimensional shape measurement using fixed-pattern projection and temporal coding

Marcus Grosse, Martin Schaffer, Bastian Harendt, and Richard Kowarschik

University of Jena, Institute of Applied Optics, Froebelstieg 1
07743 Jena, Germany
E-mail: marcus.g@uni-jena.de

Abstract. The field of image-based shape measurements using structured illumination has been an active research area for several decades now. A lot of different methods are widely used e.g. for industrial inspection, surveillance as well as multimedia purposes. Usually, short measurement times and high correspondence accuracy are mutually exclusive properties, due to the limitations of the used projection technology. Using a standard stereo-photogrammetry approach, as well as a simple, yet powerful structured light projection concept based on a slide-projector, we demonstrate an acquisition rate of more than 700 dense and accurate 3D measurements per second. © 2011 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.3646100]

Subject terms: photogrammetry; optical metrology; structured light; image processing.

Paper 110814LR received Jul. 12, 2011; revised manuscript received Sep. 6, 2011; accepted for publication Sep. 14, 2011; published online Oct. 11, 2011.

There are two distinct approaches to optical shape measurement systems employing structured light — profilometry^{1,2} and photogrammetry.^{3,4} The former methods exploit the known pattern structure and rely on a combined projector-camera calibration. The latter methods only induce a desired illumination structure onto the object and use multi-camera calibration as well as image correlation algorithms. Furthermore, both approaches can be subdivided into techniques using spatial or temporal image information,^{1,2} where spatial techniques require only one image per view for three-dimensional (3D) reconstruction and temporal techniques need a sequence of images with varying illumination for 3D reconstruction. Although spatial techniques are fast, they are comparably inaccurate and usually lead to sparse 3D reconstructions.² Single-image Fourier-transform fringe projection methods deliver dense reconstructions, but the phase unwrapping process, which is necessary to resolve ambiguous phases, still limits the complexity of scenes (height jumps larger than stripe pitch, steep slopes).^{5,6} To conclude, temporal coding methods can provide dense and accurate reconstruction of complex scenes, although Refs. 1 and 2 state that due to the temporal length of the data acquisition these techniques are limited to static scenes. Several groups showed

approaches to realize fast measuring temporal setups, either by reducing the temporal length of the sequence⁷⁻⁹ or by increasing the projection, as well as image acquisition rate.⁷⁻¹⁵ All of the known structured light temporal coding methods, which are almost exclusively based on modified fringe projection schemes, are in need of a specific set of distinct patterns to run a full measurement cycle. Therefore, they require a flexible projection unit able to project different structures, thus a digital light projector is usually used. In the case of profilometry methods the projector-camera system has to be calibrated, which has not been done yet for the fastest setups as they employ defocused binary patterns.^{9,14,15} Due to the limitations of these projection systems at high projection rates as well as the specially tailored image processing algorithms, which deal with the imposed restrictions, all fast measuring temporal methods suffer from worse measurement accuracy than their slower pendants.

A rarely used method for structured illumination is temporal coding employing statistical patterns as shown by Refs. 3 and 16 in a stereo-photogrammetry setup. Here a correspondence is found by assigning the most similar temporal gray value sequence to the reference sequence. It is possible to derive dense (more than 95% of the imaged object points are assigned), accurate, and outlier-free correspondences using $N \geq 9$ images. In contrast to stripe-projection schemes the applied temporal normalized cross-correlation can be fed with an arbitrary sequence-length, which makes the so-called sliding-window approach feasible, where every new image of a continuous measurement can be used to calculate an intermediate 3D reconstruction-state. Based on this probability codification scheme a high speed setup was demonstrated by Ref. 17, which generates objective speckle patterns in the measurement volume using a laser-illuminated rotating diffuser. A projection rate of 500 Hz and an image acquisition speed of 207 Hz enabling the recording of 17 3D reconstructions/s was demonstrated. Due to the coherent illumination subjective speckles appeared and reduced the measurement accuracy compared to bandlimited patterns proposed by Ref. 3. As the statistical nature of the applied patterns implies that a simple movement of the pattern will preserve the statistical temporal coding of an object point, it is our idea to create a temporal setup using only one pattern, which is shifted in time. This simple idea enables us to use a much simpler projection unit, which imposes no restriction on the image acquisition rate as long as the pattern is shifted fast enough between two successive images. As we will show, it is furthermore not necessary to stop the movement nor to know, calibrate, or estimate the movement direction of the pattern for the reconstruction.

To check the limitations of fixed-pattern projection versus pattern-sequence projection we performed a simulation of a perfect stereo-photogrammetry setup consisting of two pinhole cameras as well as a projector via a ray-tracer. As the calibration parameters of this setup are perfectly known, only the employed structured light coding affects the solution of the correspondence problem and therefore the measurement accuracy. We compared the results of the method shown by Ref. 3 using 48 different band-limited patterns with the concept of shifting one band-limited pattern 48 times. The measurement accuracy of both approaches was identical as long as a minimum shift of six pixels of the fixed pattern between two successive images was reached within

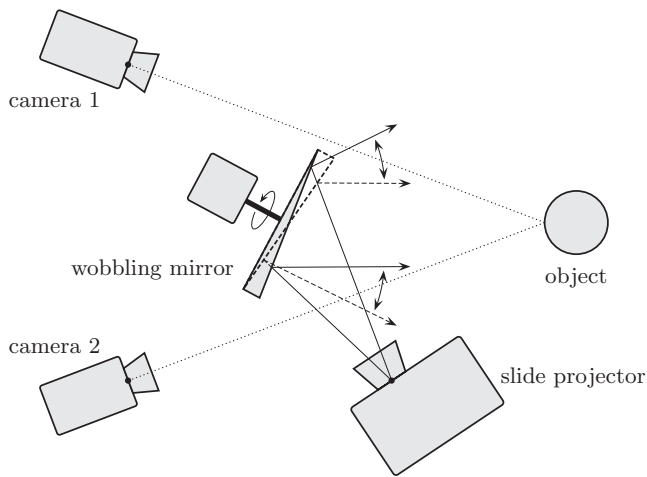


Fig. 1 Schematic of the stereo-photogrammetry setup and the proposed structured illumination unit.

an image (relative accuracy of 4.1×10^{-5} —ratio of noise of the derived 3D points to diagonal of the measurement volume—within a simulated measurement volume of $\approx 22.5 \text{ cm} \times 15 \text{ cm} \times 10 \text{ cm}$ using two simulated cameras with a resolution of 720×480 pixels). To show the feasibility of this approach for a real world measurement, we manufactured a conventional slide ($2.4 \text{ cm} \times 3.6 \text{ cm}$) based on a digital image of a band-limited pattern. This slide was projected using a regular slide projector with a thermal light source, which provides a constant brightness ($P_{\text{projector}} \approx 300 \text{ W}$) in contrast to most light-sources used in consumer digital light processings, which show brightness fluctuations at short exposure times. To continuously shift the pattern within the measurement volume we use a motor-driven wobbling mirror (60 to 1200 rpm, radius $r = 3.0 \text{ cm}$), which is positioned aside the slide projector to deflect the structured light into the measurement volume (see Fig. 1). For image acquisition two CCD cameras (camera type noted as C1) with a resolution of 640×480 pixels at 207 Hz are used (focal length $f = 17.0 \text{ mm}$, aperture stop of 1.4, pixel-size $\Delta x = 7.6 \mu\text{m}$). Alternatively, two highspeed CMOS-cameras (camera type noted as C2) with a resolution of 720×480 pixels at 11,000 Hz are used (focal length $f = 17.0 \text{ mm}$, aperture stop of 1.4, pixel-size $\Delta x = 11 \mu\text{m}$) and synchronously triggered via an external frequency generator.

Note that the cameras are not synchronized with the projection unit. The baseline of the C1 camera stereo setup is $b = 0.35 \text{ m}$ ($b = 0.25 \text{ m}$ for C2 camera stereo setup), and the test-object is placed about 0.7 m from the C1 cameras (0.5 m from the C2 cameras). The two different stereo-setup geometries were chosen to equalize the imaged object area per pixel in order to make the results most comparable. Therefore, the camera type C1 covers a measurement volume of $\approx 20.0 \text{ cm} \times 15 \text{ cm} \times 10 \text{ cm}$ and the camera type C2 a volume of $\approx 22.5 \text{ cm} \times 15 \text{ cm} \times 10 \text{ cm}$, respectively, due to the bigger horizontal resolution. The intrinsic parameters, which model the imaging properties of the cameras are calibrated *a priori* using Ref. 18. The extrinsic parameters, describing the relative orientation and position of the cameras, are determined by the eight point algorithm¹⁹ using point correspondences of the measured scene. For an actual measurement,

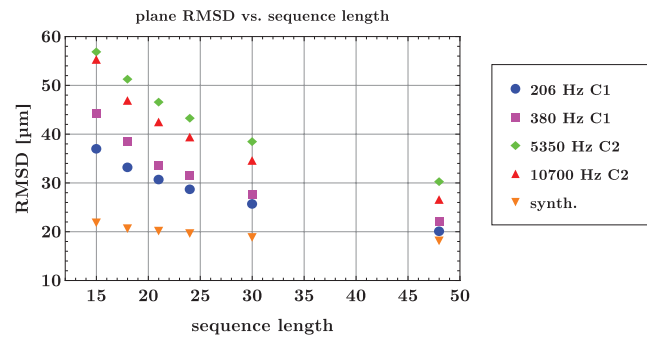


Fig. 2 Root mean square deviation of a plane normal at different image acquisition speeds; C1 denotes the CCD camera, C2 denotes the used CMOS camera.

the cameras are acquiring images at the specified speed, and the reconstruction of 3D data is done after the measurement. To show that the achievable correspondence assignment accuracy is not dependent on the measurement time t_M anymore, we conducted 3D reconstructions of a static plane normal (certified plane deviation of $\sigma_{\text{plane}} < 3.4 \mu\text{m}$) at 207 Hz, 380 Hz (camera type C1, 640×240 pixels@380 Hz), and 5350 Hz, as well as 10,700 Hz (camera type C2) image acquisition rates. Only the rotation speed of the mirror was adjusted between the measurements. The root-mean-square deviation σ of the points to a plane fit of the pointcloud was calculated, which is a direct measure for the quality of the correspondence assignment. An area of $2.5 \text{ cm} \times 2.5 \text{ cm}$ was cut out of the plane and used as the basis for the plane fit, as well as the calculation of the noise value σ (about 6000 points per noise value were used, no filtering or averaging of the points was conducted). The results depending on the chosen sequence length are shown in Fig. 2, which depicts an almost identical reconstruction quality of the measurements for each camera type. Therefore, even higher 3D capturing rates should be possible by employing a brighter illumination source, faster cameras, and an increased rotation speed according to the desired image acquisition speed. For a 3D acquisition rate of 222 independent (no sliding window is used) 3D measurements per second (3D-fps) a noise value of $\sigma \approx 26,7 \mu\text{m}$ is shown ($N = 48$ @ 10,700 Hz, $t_M = 4.5 \text{ ms}$). For a rate of 713 3D-fps the noise value is increased to $\sigma \approx 55.4 \mu\text{m}$ ($N = 15$ @ 10,700 Hz, $t_M = 1.4 \text{ ms}$) as a shorter image sequence is used. An intermediate sequence length yields intermediate correspondence accuracies. The discrepancy between the simulated noise values and the experimental values is due to actual noise sources and image aberrations of the camera-lens in the real world experiment. The difference of the results concerning the camera types C1 and C2 are caused by the lower noise and longer exposure time of the low-speed scheme with the CCD (C1). At the moment, the calculation of 3D data takes about 0.75 s per 3D reconstruction (no *a priori* information is used, e.g., depth restrictions or information of a previous 3D state). Therefore, the calculation time of 222 reconstructions takes about 166.5 s. Depending on the desired task, it should be possible to specifically tailor the image processing algorithms to decrease calculation time (e.g., by exploiting depth assumptions or steady movement assumptions).

The following experiment shows the feasibility of the setup measuring a dynamic and complex scene acquired at

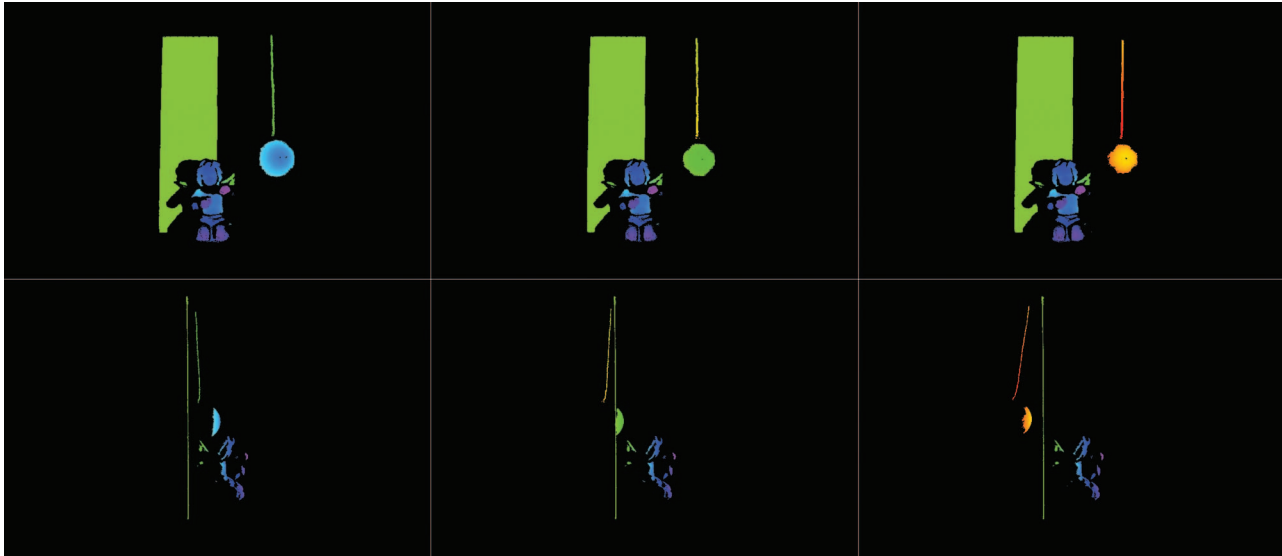


Fig. 3 Scene composed of a plane normal, an angel figure (height 6 cm), and a swinging ball (radius $r = 1.9$ cm); each successive 3D frame is separated about 89.7 ms; depth of the scene is color-coded; about 60,000 points are assigned per reconstruction; raw point data is shown — [Video 1](#): video of a scene as seen by a camera; [Video 2](#): video of reconstructed scene; [Video 3](#): shows the reconstruction of a human head derived with the same experimental parameters. Video 1, QuickTime, 3.3 MB, [URL: <http://dx.doi.org/10.1117/1.3646100.1>], (Video 2, QuickTime, 3.6 MB), [URL: <http://dx.doi.org/10.1117/1.3646100.2>], (Video 3, QuickTime, 2.5 MB), [URL: <http://dx.doi.org/10.1117/1.3646100.3>].

$R = 10, 700$ Hz, where three separate objects, one of them moving, are reconstructed. Figure 3 shows a series of three 3D reconstructions as seen from two arbitrary chosen virtual perspectives of the scene ($N = 24$, $t_M = 2.24$ ms) successively separated 89.7 ms in time.

To conclude, the presented method enables dense and precise 3D shape measurements of complex scenes within a very short measurement time. Due to the temporal coding concept the correspondence accuracy of moving objects is reduced, yet it is still possible to derive 3D information up to a certain object speed depending on the acquisition rate and sequence length. As the temporal sequence length is not fixed with the used coding scheme, improved image processing algorithms may enable even higher 3D rates, especially if the spatial image information is used in combination with the temporal image information. The vision, to be hopefully reached in future work, is to use the special properties of statistical coding, namely its gorgeous characteristics to be statistical in space and time, to automatically choose the spatiotemporal image information on a pixel-basis to match the imaged object points speed best.

1. S. S. Gorthi and P. Rastogi, "Fringe projection techniques: Whither we are?" *Opt. Lasers Eng.* **48**, 133–140 (2010).
2. J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recogn.* **43**(8), 2666–2680 (2010).
3. A. Wiegmann, H. Wagner, and R. Kowarschik, "Human face measurement by projecting bandlimited random patterns," *Opt. Express* **14**(17), 7692–7698 (2006).
4. P. Kühmstedt, C. Munkelt, M. Heinze, C. Bräuer-Burchardt, and G. Notni, "3D shape measurement with phase correlation based fringe projection," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 6616 (2007).
5. E. Lally, J. Gong, and A. Wang, "Method of multiple references for 3d imaging with Fourier transform interferometry," *Opt. Express* **18**(17), 17591–17596 (2010).
6. L. C. Chen and X. L. Nguyen, "Dynamic 3d surface profilometry using a novel colour pattern encoded with a multiple triangular model," *Meas. Sci. Technol.* **21**(5), 054009 (2010).
7. Z. Wang, H. Du, S. Park, and H. Xie, "Three-dimensional shape measurement with a fast and accurate approach," *Appl. Opt.* **48**(6), 1052–1061 (2009).
8. L. Kinell, "Spatiotemporal approach for real-time absolute shape measurements by use of projected fringes," *Appl. Opt.* **43**(15), 3018–3027 (2004).
9. Y. Gong and S. Zhang, "Ultrafast 3-d shape measurement with an off-the-shelf dlp projector," *Opt. Express* **18**(19), 19743–19754 (2010).
10. S. Zhang and S.-T. Yau, "High-resolution real-time 3d absolute coordinate measurement based on a phase-shifting method," *Opt. Express* **14**(7), 2644–2649 (2006).
11. S. Knig and S. Gumhold, "Image-based motion compensation for structured light scanning of dynamic surfaces," in *EG Workshop on Dynamic 3D Imaging*, pp. 173–181 (2007).
12. S. G. Narasimhan, S. J. Koppal, and S. Yamazaki, "Temporal dithering of illumination for fast active vision," in *Proceeding of the European Conference on Computer Vision*, pp. 830–844 (2008).
13. S. Zhang, "Recent progresses on real-time 3d shape measurement using digital fringe projection techniques," *Opt. Lasers Eng.* **48**, 149–158 (2010).
14. Y. Wang and S. Zhang, "Superfast multifrequency phase-shifting technique with optimal pulse width modulation," *Opt. Express* **19**, 5149–5155 (2011).
15. Y. Wang, S. Zhang, and J. H. Oliver, "3d shape measurement technique for multiple rapidly moving objects," *Opt. Express* **19**(9), 8539–8545 (2011).
16. P. Albrecht and B. Michaelis, "Stereo photogrammetry with improved spatial resolution," in *IEEE Pattern Recognition* **14**, 845–849 (1998).
17. M. Schaffer, M. Große, and R. Kowarschik, "High-speed pattern projection for three-dimensional shape measurement using laser speckles," *Appl. Opt.* **49**(18), 3622–3629 (2010).
18. Z. Zhang, "A flexible new technique for camera calibration," Technical report, Microsoft Research (1998).
19. R. I. Hartley, "In defense of the 8-point algorithm," in *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 580–593 (1997).