PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

# Large-volume NIR pattern projection sensor for continuous low-latency 3D measurements

Christoph Munkelt, Matthias Heinze, Christian Bräuer-Burchardt, Shantanu Kodgirwar, Peter Kühmstedt, et al.

Christoph Munkelt, Matthias Heinze, Christian Bräuer-Burchardt, Shantanu P. Kodgirwar, Peter Kühmstedt, Gunther Notni, "Large-volume NIR pattern projection sensor for continuous low-latency 3D measurements," Proc. SPIE 10991, Dimensional Optical Metrology and Inspection for Practical Applications VIII, 109910K (13 May 2019); doi: 10.1117/12.2518636

SPIE.

Event: SPIE Defense + Commercial Sensing, 2019, Baltimore, MD, United States

# Large-volume NIR pattern projection sensor for continuous low-latency 3D measurements

Christoph Munkelt*[a], Matthias Heinze[a], Christian Bräuer-Burchardt[a], Shantanu P. Kodgirwar[a], Peter Kühmstedt[a], Gunther Notni[a,b]
[a]Fraunhofer Institute for Applied Optics and Precision Engineering IOF, Albert-Einstein-Str. 7, 07745 Jena, Germany;
[b] Ilmenau University of Technology, Gustav-Kirchhoff-Platz 2, 98693 Ilmenau, Germany

## ABSTRACT

For continuous, low-latency, irritation-free 3D measurements in large-volumes, dot-pattern- or time-of-flight-based sensors have been traditionally used. However, measurement accuracy and temporal stability limits the application in demanding medical or industrial contexts. Practical solutions also need to remain cost-effective. To meet these requirements, we started from a simple GOBO-based, aperiodic sinusoidal pattern projection (using a near-infrared (NIR) LED) 3D sensor for medium-sized measurement volumes. By tuning the system for large-volume operation, we were able to obtain a reasonable combination of measurement accuracy and speed. The current realization covers a volume of up to 4.0 m x 2.2 m x 1.5 m (width x height x depth). The 3D data is acquired at > 20 fps at resolutions of > 1000 x 500 px and true end-to-end latencies of < 140 ms.

We present the system architecture consisting of GigE Vision cameras, a high-power LED-driven projection unit using a GOBO wheel, and the compute backend for the online GPU-based, temporal pattern correlation 3D calculation and filtering. To compensate for the low pattern intensity due to the short exposure time, we operate the cameras in 2x2 binning. Furthermore, the optics are tuned for large apertures to maximize light throughput. We characterize the sensor system with respect to measurement quality by quantitative evaluations including probing error, sphere-spacing error, and flatness measurement error. By comparison with another 3D sensor as a baseline, we show the benefits of our approach. Finally, we present measurement examples from human-machine interface (HMI).

**Keywords:** Large measurement volume, continuous, NIR, irritation-free, 3D scanner, low-latency, GOBO, aperiodic sinus

## 1. INTRODUCTION

Optical three-dimensional (3D) sensors based on structured light are increasingly used in different areas: industrial quality control, medical sciences, or entertainment industry. These systems for smaller to medium scale measurement volumes have been studied at length [1],[2]. By actively projecting *N* different patterns ("sequence length"), finding of identical points in the 2D images is robust, accurate, and independent of object texture. 3D coordinates of those points can be calculated by triangulation [3][4]. Recent advances allowed increasing the 3D data-recording rate vastly [5]. This enables to capture fast processes in three dimensions.

However, most high-resolution, high-speed 3D sensors typically operate in a burst regime. They separate the image capturing from the 3D reconstruction by storing their intermediate 2D images locally, for later offline processing [6]. Furthermore, to enable area-scan acquisition of a high number of points simultaneously, typically a high-power, visible light source is used. Both aspects are restricting the usage in scenarios, which require continuous, irritation-free measurements involving humans, e.g. human-machine interface (HMI) and patient monitoring.

Another widely popular approach is the usage of time-of-flight (ToF) sensors, e.g. the Kinect v2 [7]. Despite the low lateral resolution of its depth sensor, a frame rate of 30 Hz with its irritation-free operation at the wavelength of 860 nm, and the low price led to its adoption in numerous applications. For domains requiring reliable measurements however, the typical

*christoph.munkelt@iof.fraunhofer.de; phone 49 3641 807-245; fax 49 3641 807-602; iof.fraunhofer.de

Dimensional Optical Metrology and Inspection for Practical Applications VIII, edited by Kevin G. Harding, Song Zhang, Proc. of SPIE Vol. 10991, 109910K · © 2019 SPIE
CCC code: 0277-786X/19/$18 · doi: 10.1117/12.2518636

ToF characteristics (influence of thermal instability, depth variations depending on surface reflection properties, optics- and IR-illumination-related distortions, distance-dependent systematic depth error, and a generally high standard deviation of depth readings [7]) limit its possible application.

Against this background, our goal was to expand an existing medium-volume scanner for continuous 3D measurements into a large volume scanner in order to estimate the adaptability of structured light 3D scanners for such a volume.

## 2. METHODOLOGY

For the intended application areas of HMI and patient monitoring, the following key characteristics are desirable:

1. Continuous 3D reconstruction with low latency,
2. Ability to reconstruct a measurement volume large enough to contain two people (e.g. $\geq 3 \times 2 \times 1$ m³),
3. Acquisition of medium-speed movements (e.g. pointing, presenting of objects to measure …).
4. Reasonable measurement uncertainty (both absolute error and noise, e.g. $\leq 0.1$ % of body diagonal of measurement volume), and
5. Usage of standard sensor (e.g. no expensive high-speed) and infrastructure equipment.

Ideally, multiple of such sensors should be able to capture different viewpoints of the scene simultaneously to allow for complete reconstruction even under occlusion. The challenge arising from this requirement is further discussed in section 3.

Within the remainder of this section, details of the general measurement principle, as well as the chosen large volume calibration approach are described.

### 2.1 3D shape measurement by GOBO projection of aperiodic sinusoidal fringes

The basic 3D reconstruction principle used is triangulation of corresponding points recorded from a stereo camera arrangement. To identify those point pairs accurately and robustly in the 2D camera images, a series of pattern is projected onto the measurement object. Specifically, a series of aperiodic sinusoidal fringes is used [8]. To achieve a high projection rate, the GOBO (GOes Before Optics) approach is used. By utilizing a rotating slide, bright and fast pattern modulation without high technical complexity can be achieved [5].

A series of images with different of such fringe patterns is synchronously captured by two cameras. For each pixel, a gray value stack is obtained. 2D image pixels observing identical object points should show identical variations in intensity. Thus, corresponding pixels are determined by finding pairs with the highest correlation value from normalized temporal cross-correlation [8]. To speed things up, the search is restricted to the epipolar line.

### 2.2 Sensor calibration for large measurement volumes

One necessity for accurate 3D reconstruction within large volumes is a good calibration of the required wide-angle optics. Especially, since those lenses typically exhibit rather high optical distortion. Inexpediently, to calibrate the sensor optimally, one has to capture data over the entire sensor area, including the sensor corners. This requires spacious and numerous calibration patterns, favorably including some basic topography. Since it is impractical to move such a large calibration board, a large calibration scene was mounted to the wall and the sensor was moved instead.

We show our calibration setup in Figure 1a). To calibrate the sensor, images from approximately 14 different sensor poses in front of the homogenous illuminated marker board are captured. For accurate results, recording of markers up to the sensors' corners is important. Figure 1b) shows a camera view with the detection results of the ArUco markers [9]. As thousands of them can be robustly detected across multiple sensor poses and within both cameras, a bundle-block adjustment (using the software suite BINGO [10]) can be used to accurately determine the intrinsic parameters and relative sensor poses of the stereo camera system. One resulting calibration solution is shown in Figure 1c).

To summarize, the calibration approach performs well even within a large measurement volume [11]. This is achieved using bundle-block adjustment with marker detection over a large scene at the intended measuring distance.

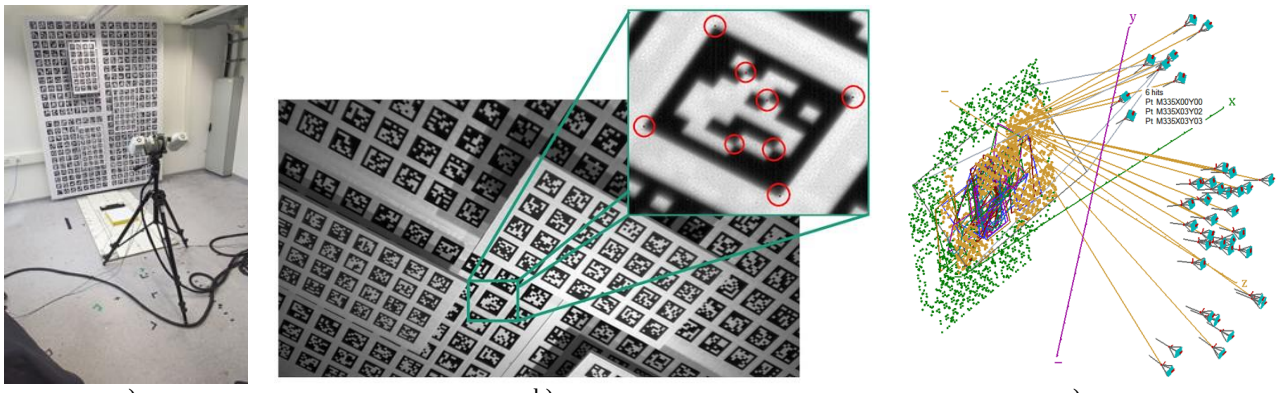


Figure 1: Sensor calibration environment for accurate intrinsic calibration: a) Sensor calibration pose in front of wall-mounted marker board; b) Enlarged section of 2D camera image with detected ArUco marker points; c) Resulting bundle block adjustment result of chosen sensor positions and orientations.

## 3. EXPERIMENTS

To evaluate the effectiveness of the proposed approach, we prototypically built an NIR 3D scanner (see Figure 2) using GOBO projected aperiodic sinusoidal fringes, capable of capturing a measurement volume of approximately 4.0 m x 2.2 m x 2.0 m. Within that volume, two people can interact. Alternatively, one human can interact with a robot, e.g. an autonomous scanning station. The GOBO projector uses two high-power NIR LEDs (Osram SFH 4761) at 850 nm. The sensor captures the scene with two synchronized, NIR enhanced monochrome GigE cameras (Baumer LXG-20NIR). They continuously deliver up to 111 fps at their full resolution of 2048 x 1088 px with a dual Gigabit Ethernet connection. Alternatively, they can operate at up to 336 fps at 2 x 2 binning with respective resolution of 1024 x 544 px. Each camera uses a Schneider Kreuznach Cinegon 1.9/10 lens, while the GOBO projector uses a Schneider Kreuznach Cinegon 1.8/16. To render the scanner mostly immune to ambient lighting, each lens is equipped with a narrow optical bandpass filter. Table 1 summarizes the technical features of our 3D scanner.

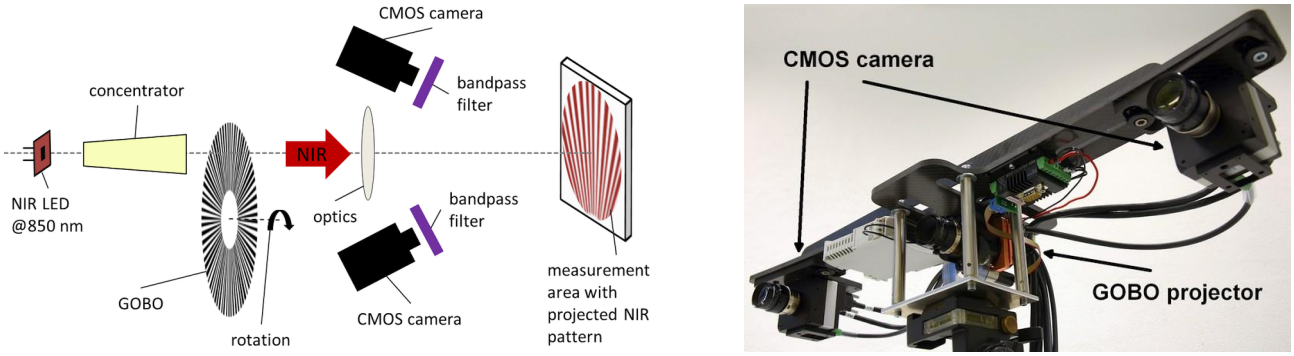


Figure 2: Prototype large-volume NIR pattern projection 3D scanner: a) GOBO 3D scanner schematics (figure from [12]); b) Photo of an early, prototypical realization of the scanner for continuous low-latency 3D reconstruction.

The data transfer and computing architecture remains the same as in [11]: Dual Gigabit Ethernet links for fast camera data transfer, 10 Gigabit Ethernet switch as uplink concentrator, and online 3D calculation using GPU-based OpenCL code. To fulfill the low-latency requirements, we chose a maximum exposure time of 10 ms. Since we calculate one 3D frame from 10 2D images, the end-to-end latency, consisting of 2D image acquisition, data transfer and GPU-based 3D calculation, is approximately 140 ms. However, using overlapping sequences of 2D images for 3D data calculation, we achieve 3D frame rates of 20 Hz ("partial overlap": every 5 2D images, a new 3D data frame from 10 2D images is calculated) to 100 Hz

("full overlap": a 3D frame at every new 2D image). The end-user has to choose the right balance between required GPU computing power (to calculate the 3D data) and required CPU computing power (to analyze and process the 3D data).

Because of the large measurement volume and the low-power, LED-based NIR illumination, the perceived brightness of the projected patterns is quite low. The camera exposure time could be increased accordingly. Alternatively, to fulfil our continuous low-latency 3D reconstruction requirement, camera operation in 2x2 binning mode was chosen. While this reduces the lateral resolution of the scanner, it quadruples the sensitivity of the pixels, allowing for 100 Hz 2D image rate even with low pattern brightness. If higher latencies (through increased exposure times) are acceptable in trade for higher sensor resolution, the scanner can also be operated in full resolution, yielding object pixel sizes as low as 2 mm.

Typically, we use multiple 3D scanners to capture multiple viewpoints of the same scene. This allows for complete 3D object reconstruction even under occlusion [11]. For continuous reconstruction, however, we typically synchronize the projected patterns of the involved scanners in a time-multiplexing fashion. We do this in order to avoid pattern interference and resulting distortions. For this prototype, the chosen light source was not yet powerful enough to split e.g. the available 10 ms exposure time into 2 x 5 ms. Therefore, within this paper we show only reconstructions from one active 3D sensor.

Table 1: Technical features of prototype large measurement volume NIR 3D scanner.

| | |
|---|---|
| **Operating wave length** | 850 nm |
| **Pattern projector optical power** | ~ 9 W |
| **Camera resolution** | 2048 x 1088 *(1024 x 544 with 2x2 binning)* |
| **Field of view (FoV)** | ~ 60 x 30 degree |
| **2D frame rate** | 100 Hz |
| **Sequence length** | 10 images |
| **3D frame rate** | 100 Hz (full overlap) … 20 Hz (partial overlap) |
| **End-to-end latency** | < 140 ms |
| **Measuring distance** | ~ 4 m |
| **Triangulation angle** | ~ 9 degree |
| **Measuring volume** | ~ 4.0 m (W) x 2.2 m (H) x 1.5 m (D) |
| **Object pixel size (GSD)** | 2 mm *(4 mm with 2x2 binning)* |

The following sections highlight results from two different measurement series.

### 3.1 Scanner characterization measurements

As a baseline, we use the 3D scanner in its original configuration ("baseline"). Using the same cameras, but a different set of optics, it covers a measurement volume of approximately 2 m (W) x 1 m (H) x 0.8 m (D) for continuous 3D reconstruction with up to three simultaneous active 3D scanners [13] at 10 Hz 3D frame rate. As reference values, the following is used:

- The standard deviation of best-fit plane fitting results ("$\sigma_{plane}$") against a 400 mm x 400 mm flatness normal,
- The standard deviation of best-fit sphere fitting results ("$\sigma_{sphere}$") against 50 mm radius spheres, as well as
- The length deviation ("SD") from the calibrated sphere center distance of 300 mm of a ball bed normal.

The results for the baseline were taken directly from the continuous 3D data stream, recorded only at the front / middle / back positions of its measurement volume at approximate distances of 2650 mm / 2800 mm / 2900 mm. The baseline plane fit standard deviations $\sigma_{plane}$ at the respective positions are 0.33 mm / 0.16 mm / 0.35 mm. The baseline sphere fit standard deviations $\sigma_{sphere}$ for the two spheres at the respective positions are (0.38 mm, 0.25 mm) / (0.26 mm, 0.30 mm) / (0.45 mm, 0,65 mm). The length deviations SD at the respective positions are 0.01 mm / -0.02 / -0.12

For comparison, both the ball bed and the flatness normal were placed at different positions of the large measurement volume (see Figure 3a). The 3D data was taken directly from the live 3D data stream. This represents our continuous, low-latency measurement approach best. At the subfigure b), the color-coded deviation against a fitted sphere is shown. Some

noise is visible, as well as some "bulge" at the border of the reconstructed part of the sphere. But the best-fit standard deviation $\sigma_{sphere\ 2}$ = 0.40 mm is comparable to "baseline" results of similar sensor distance. In subfigure c) the color-coded deviation against a fitted plane is depicted. Again, some noise is visible, but no significant bending, which indicates sufficient calibration quality.

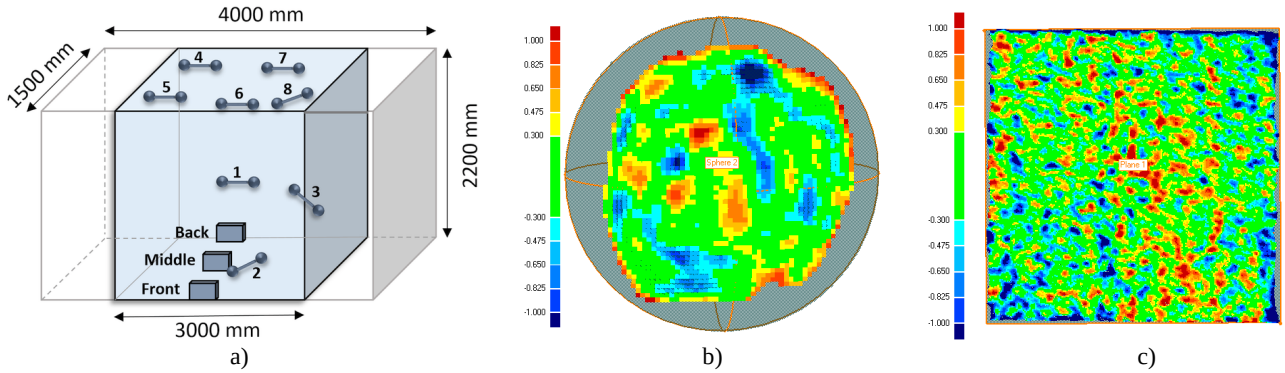


Figure 3: Characterization of 3D scanning results: a) Positions of ball bed normal (50 mm radius, 300 mm sphere distance, depicted as "1"-"8") and flatness normal (400 mm x 400 mm, depicted as "Front" – "Back") within the measurement volume; b) Color-coded sphere fit deviations of "Sphere 2" in position "1"; c) Color-coded plane fit deviations in "Middle" position (legend: ± 0.3 mm nominal [green], ± 1.0 mm critical).

Quantitative results are presented in Table 2. The fit results of the flatness normal and the ball bed are shown with respect to their measurement positions and respective measurement distance. Two trends are easily noticeable: the farther away from the sensor (see e.g. the various $\sigma_{plane}$ values), the higher the measurement uncertainty. This is caused mainly from the decrease in perceived light intensity, if the object is far away from the pattern projector. Other factors are the limited depth of field and the decreasing triangulation angle with increasing measuring distance. Compared to the baseline, we could not increase the camera baseline (due to usage of an existing scanner), which results in increased measurement uncertainty.

The other trend to be observed is that the results are poorer at the edges of the volume, e.g. ball bed positions 4 or 7. Not only increased $\sigma_{sphere}$ values are incurred, also the sphere distance error SD is increased. This results mainly from the inhomogeneous pattern projection and inhomogeneous relative illumination of the lens. Other factors are residual distortion effects of the camera lenses.

Table 2: Measuring results of a 400 mm x 400 mm flatness normal and a ball bed (two spheres of radius 50 mm in 300 mm distance). We have placed both reference objects at various positions throughout the measurement volume (see Figure 3).

| | Measurement position | Distance from sensor in mm | Standard deviation in mm | | | Sphere radius in mm | | Sphere distance error SD in mm |
|---|---|---|---|---|---|---|---|---|
| | | | $\sigma_{plane}$ | $\sigma_{sphere\ 1}$ | $\sigma_{sphere\ 2}$ | $r_1$ | $r_2$ | |
| **Flatness normal** | **Front** | 2500 | 0.37 | | | | | |
| | **Middle** | 3080 | 0.58 | | | | | |
| | **Back** | 3550 | 0.79 | | | | | |
| **Ball bed** | **1** | 2960 | | 0.47 | 0.40 | 51.25 | 50.77 | -0.25 |
| | **2** | 3250 | | 0.55 | 0.64 | 49.95 | 50.42 | 0.96 |
| | **3** | 3360 | | 0.75 | 0.76 | 50.74 | 50.99 | 1.08 |
| | **4** | 3600 | | 0.72 | 0.78 | 50.08 | 50.59 | -2.31 |
| | **5** | 2800 | | 0.51 | 0.47 | 50.58 | 50.35 | -1.18 |
| | **6** | 2730 | | 0.58 | 0.60 | 50.21 | 50.66 | 0.08 |
| | **7** | 3420 | | 0.66 | 0.75 | 51.28 | 49.79 | 1.73 |
| | **8** | 2700 | | 0.36 | 0.52 | 50.49 | 48.97 | 0.62 |

Compared to our baseline, the accuracy results are within the expected margins. The 3D data is available within ≤ 140 ms from the 3D scanner (from start of exposure of the first underlying 2D image). Further processing steps (Kinect-like data

fusion) could be applied at the expense of latency, if lower measurement uncertainty is required. Further improvements on the pattern projection by utilization of brighter LEDs, and usage of modern CMOS sensors with less noise, will enable shorter latency (by shorter exposure times) and simultaneous usage of multiple, time-multiplexed, 3D sensors.

### 3.2 Usage examples

We tested the scanning system in scenarios within human-machine interface (HMI) applications. The first, depicted in Figure 4a, represents a worker, who collaboratively interacts with a separate robot-mounted scanning station. Several task have to be achieved: Human and robot shall not collide. The scan station should recognize a "scan this!" gesture and the direction the human approaches it and orient itself accordingly. The scan station itself has a high resolution, yet narrow FoV 3D scanner – it needs external 3D pose orientation to recognize and scan the presented part from multiple views. Finally, the sensor network of multiple 3D scanners (for occlusion-free measurement volume coverage) needs to operate irritation-free, continuously, and with low latency. Figure 4b shows such a multiple-scanner result from a smaller volume.

a) b)

Figure 4: a) Human-machine interface (HMI) usage scenario: A technician presents a part to a robot-based scanning station. b) Three views of a 3D model, simultaneously acquired by three full-resolution 10 Hz 3D sensors (1.0 m x 2.0 m x 0.5 m).

The second usage scenario, shown in Figure 5, originates from a project to mobilize people with the help of movement-sound-interactions for rehabilitation purposes. To that end, one to two people can move freely within a large volume and a computer generates harmonic, chordal music in relation to the movement of the human. For mobilization aspects, the software focusses on hand and feet movements. It distinguishes various types of movements (waving, kicking, and beating) to different directions (front, side, up) to get clues about pitch and beat of the music. Figure 5 shows two actors, with the left playing "strings" through slower movements, and the right playing "drums" through faster beating hand movements. Again, low-latency, irritation-free, continuous 3D monitoring in a large measurement volume is required to fulfil the respective task.

Figure 5 3D data recording parameters:
- ≤ **140 ms latency**
- Raw data (no filter)
- Single 3D frame from **continuous 20 fps** data stream
- **Large measurement volume** (4 m x 2.2 m x 1.5 m)

Figure 5: Measurement results from a ≥ 10 seconds dynamic performance scene: Two actors play simultaneously on virtual music instruments. The 3D data was recorded with the frontal scanner at 100 fps 2D (20 fps 3D) rate using 2x2 binning.

## 4. SUMMARY

We presented an adapted 3D scanning system for continuous low-latency 3D measurements within large volumes. It utilizes a GOBO projection-based NIR projector for high-speed pattern projection. The existing sensor platform, as well as infrastructure for calibration, proved sufficient for increasing the measurement volume from the baseline to the target 4 m (W) x 2.2 m (H) x 1.5 m (D). Since the pattern projection light intensity could not be increased, usage of 2x2 binning was chosen to increase camera sensitivity as a compromise. It was shown that the system is able to continually measure within (and monitor) a large volume with sufficient accuracy. Possible application areas are generally usage scenarios, where a lower limit for measurement accuracy, both in terms of noise and dimensional accuracy, are required. Traditionally, ToF sensors cannot satisfy these constraints for low-latency applications.

Recent advances with high-power NIR LEDs and Vertical-cavity surface-emitting lasers (VCSEL) might allow us to increase the optical power of the pattern projection, thus increasing the available light. This will enable lower latency due to decreased exposure times, simultaneous reconstruction from multiple 3D sensors, and higher lateral resolutions due to the option to refrain from binning.

## ACKNOWLEDGEMENTS

This study is supported by the Free State of Thuringia, the European Social Fund (ESF) of the European Union and the Thüringer Aufbaubank (TAB) within the research group DIADEM (2016 FGR 0044). Further support comes from the Free State of Thuringia, the European Social Fund (ESF) of the European Union and the Thüringer Aufbaubank (TAB) within the project “Bewegungsaktivierung und Musikgenerierung mittels 3D-Scanning” (2015 VF 0032).

## REFERENCES

[1] Pagès, J., Salvi, J., Garcia, R., Matabosch, C., “Overview of coded light projection techniques for automatic 3D profiling,” IEEE Trans. Robot. Autom. (1), 133–8 (2003).
[2] Chen, F., Brown, G. M., Song, M. “Overview of three-dimensional shape measurement using optical methods,” Opt. Eng. 39 (1), 10–22 (2000).
[3] Luhmann, T., Robson, S., Kyle, S., Harley, I., “Close range photogrammetry: principles, techniques and applications,” Wiley (2007).
[4] Hartley, R., Zisserman, A., “Multiple view geometry in computer vision,” Cambridge University Press (2004).
[5] Heist, S., Lutzke, P., Schmidt, I., Dietrich, P., Kühmstedt, P., Tünnermann, A., Notni, G., "High-speed three-dimensional shape measurement using GOBO projection," Optics and Lasers in Engineering (87), 90-96 (2016).
[6] An, Y.,, Zhang, S., "High-speed, high-accuracy large range 3D measurement," Proc. SPIE 10220, Dimensional Optical Metrology and Inspection for Practical Applications VI, 1022005 (2017).
[7] Lachat, E., Macher, H., Landes, T., Grussenmeyer, P., “Assessment and Calibration of a RGB-D Camera (Kinect v2 Sensor) Towards a Potential Use for Close-Range 3D Modeling,” Remote Sens., 7, 13070-13097 (2015).
[8] Heist, S., Kühmstedt, P., Tünnermann, A., Notni, G., “Experimental comparison of aperiodic sinusoidal fringes and phase-shifted sinusoidal fringes for high-speed three-dimensional shape measurement,” Opt. Eng. 55 (2), 024105 (2016).
[9] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, M. Marín-Jiménez, “Automatic generation and detection of highly reliable fiducial markers under occlusion,” Pattern Recognition. 47, 2280–2292 (2014).
[10] Kruck, E., “Bundle Adjustment for Engineering Applications - A Program System for Close Range Photogrammetry and Aerial Triangulation,” http://www.gip-aalen.de (2018).
[11] Munkelt, C., Heinze, M., Zimmermann, P. Kühmstedt, P. and Notni, G., “High Performance, low latency 3D sensor network for live full object reconstruction,” Proc. SPIE 10667, 1066706 (2018).
[12] Brahm, A., Ramm, R., Heist, S., Rulff, C., Kühmstedt, P., Notni, G., "Fast 3D NIR systems for facial measurement and lip-reading," Proc. SPIE 10220, Dimensional Optical Metrology and Inspection for Practical Applications VI, 102200P (2017).
[13] Diehm, V., “Charakterisierung eines neuartigen NIR-3D Sensornetzwerkes für große Messvolumen und hohe Bildwiederholraten,“ Bachelor Thesis, EAH Jena (2018).