Accurate Irritation-Free 3D Scanning of Human Face and Body Sequences

Christian BRÄUER-BURCHARDT*[1], Anika BRAHM[1], Christian RULFF[1],
Stefan HEIST[1], Peter KÜHMSTEDT[1], Gunther NOTNI[1,2]
[1] Fraunhofer Institute IOF, Jena, Germany;
[2] Technical University Ilmenau, Ilmenau, Germany

Abstract

Two new 3D scanning setups for the fast acquisition of sequences of moving persons based on near-infrared structured light projection are introduced. Both systems consist of a pair of stereo cameras and a projection unit and are working in the near-infrared range (illumination wavelength is over 780 nm) which is absolutely irritation-free for the observed person. The first system is based on the so-called GOBO projector and the second system uses a multi-aperture-projection unit consisting of ten projecting components. Measurements of 3D sequences of moving persons and examples of changing facial expressions are given in order to show the power of the setups. The results obtained by the use of the different projection systems are compared concerning measurement accuracy and robustness.

Keywords: 3D human motion scanning, near-infrared structured light projection, irritation-free 3D scanning

1. Introduction

Accurate 3D reconstruction of the human face and body provides many application fields especially in medicine, human body and face modeling, animation, applications in virtual life, or computer games and entertainment. New developments of gesture and facial expression detection and interpretation can be used for communication tasks or machine control. Whereas gestures and facial expressions can be detected in 2D image sequences yet, 3D scanning of faces may provide more detailed information which can help to increase the power of corresponding gesture analysis systems.

At our institute, we are developing 3D scanners based on active pattern projection for different types of applications. Recently, our main focus has been on the ultrafast realization of 3D object acquisition, up to a 3D data rate of more than 1 kHz [1]. This allows the investigation of dynamic processes, such as airbag inflation, car crash-tests, or human movements at certain activities like sports or in the working process. Not all applications where human beings are scanned require such a high temporal resolution. The majority of applications concerning moving humans need 3D data frequencies between 1 Hz and 25 Hz. The illumination by structured light may be a disturbing factor at the scanning of humans (glare effect) and often needs dark operation areas. These conditions may prevent certain applications. Hence, an irritation-free 3D scanning would be desirable.

In this work, we introduce two new 3D scanning setups which we developed for the acquisition of human facial or body activities. An overview of the performance parameters and the technical setup properties and some experimental measurement results and application examples are given which characterize the power of the two setups.

2. State of the art

The rapid technological development of hardware and software components for 3D scanning systems is the basis for many different fields of application of 3D body scanning. Regarding the scanning technology, we have to distinguish between scanning of static and moving objects. Whereas the 3D scanning of static objects or non-moving persons has its history mainly in medical applications such as computer-assisted tomography (CAT) [2] and nuclear magnetic resonance imaging (NMRI) [3], the 3D scanning of moving persons was strongly shifted into the focus of research in the last few years. Consequently, especially in the last few years, many applications of 3D body scanning of moving humans have arisen. However, the methodologies and techniques for 3D scanning of non-moving persons are also being refined, e.g. in the fields of medicine, forensics, apparel industry, and security.

*christian.braeuer-burchardt@iof.fraunhofer.de; +49-3641-807235
  Fraunhofer Institute IOF, Albert-Einstein-Str. 7, 07745 Jena, Germany, http://www.iof.fraunhofer.de/

A survey of 3D body and medical scanner techniques and methodologies is given by Rodriguez-Quinonez et al. [4]. Whereas the 3D scanning of static objects or non-moving persons should typically be very robust, accurate, and of high 3D data point resolution, the 3D registration of moving persons has higher challenges concerning storage capacity and speed of data processing. Recently, many methods for 3D scanning of human face and sequences of moving bodies have been introduced. These Methods include different kinds of sensors, e.g. video-based systems [5], small tracker sensors [6], or systems using structured illumination. Structured light techniques may work in the visible (VIS) or infrared (IR) spectral range. Many methods for the 3D reconstruction using visible structured light also for moving objects or body part have been published (e.g. [7-9]) and should not be explained in detail here. In a previous work we introduced a high-speed 3D scanning system based on aperiodic sinusoidal fringe pattern projection in the visible light with a 3D frame rate of up to 1.5 kHz [10]. Fu et al. [11] introduced a structured light scanning system for 3D reconstruction of dressed human models based on near-infrared illumination and recording of fringe patterns. Although a maximum 2D frame rate of 480 Hz is stated, measurement examples are given only from static objects. The average measurement error is given as less than 1 mm for a measurement field of approximately 200 mm x 150 mm. Hansen et al. [12] introduce 3D face reconstructions of non-moving persons from a photometric stereo system using near-infrared light. Sandbach et al. [13] give a survey of static and dynamic 3D facial expression recognition methods. Recently, movements of human bodies have been captured three-dimensionally using Microsoft Kinect devices (see [14]). However, these solutions do not provide the ability of very accurate 3D measurements.

3. Setup description

3.1. Basic principles

Fringe projection profilometry (also called phasogrammetry) is one of the established methods of structured light projection (see e.g. [15]), but other structured illumination patterns (binary dot patterns, speckle patterns, aperiodic sinusoidal patterns, see [16]) are also suitable for this technique. The restriction to wavelengths over 780 nm is no restriction concerning the measurement principle. However, illumination sources should be changed or spectral filters must be used in the case of projected pattern generation. The recording process by one or more cameras should be optimized to the finally used wavelengths.

3.1.1. Measurement principle
Both measurement setups we introduce are based on the measurement principle of stereo observation and structured light illumination. They both consist of a projection and an observation unit. The observation is realized by a stereo-camera arrangement with blocking filters for the visible light frequencies. The projection unit projects structured light patterns onto the measurement object which help to identify corresponding points in both cameras.

3.1.2. Structured light pattern projection and correspondence finding
There are several possibilities for the used pattern for the structured illumination. Typically, sinusoidal fringe patterns, binary dot patterns, laser speckle patterns, and others are used for such optical measurement systems (see e.g. [7-9]). We used aperiodic sinusoidal fringe patterns [16] as shown in Fig. 1 for both setups. For the first setup, the generation of the patterns was realized by a so-called GOBO projector. Here, the projection is realized by a rotating wheel with radially arranged fringes and illumination by a high power LED lamp (see Fig. 2). The principle is described in detail in [1].
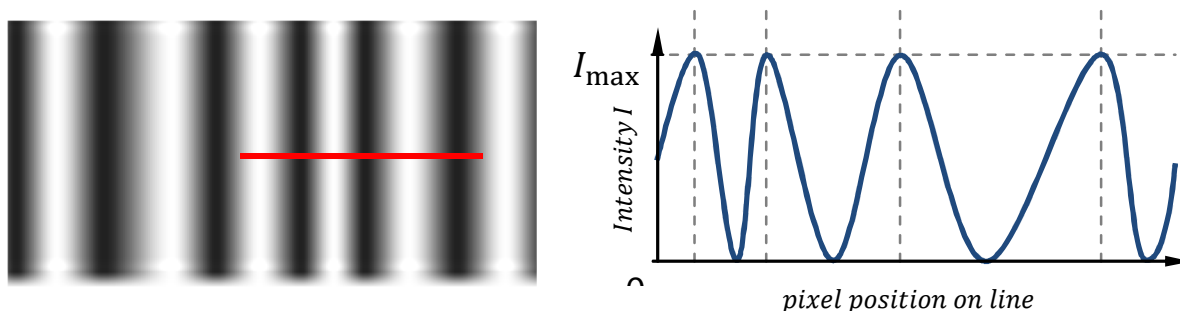

Fig. 1. Aperiodic sinusoidal fringe pattern: section of a camera observation (left)
and grey-value profile (right) of a part of a line (red line)

The second setup also uses aperiodic sinusoidal fringe patterns but under application of the special principle of multi-aperture projection (MAP). For more details concerning MAP see e.g. [16].

3.1.3. 3D point calculation
The 3D measurement data are obtained by triangulation between corresponding points in the two cameras according to the principle known from photogrammetry [17]. Corresponding points are found by comparison of the temporal patterns observed at certain camera pixels along epipolar lines, which are obtained by calibration of the stereo cameras. We applied the calibration according to Zhangs method [18] using a planar chessboard pattern. However, any arbitrary suitable calibration technique can be applied for the stereo cameras. It is not necessary to calibrate the projection unit because it is not directly included in the measurement process.

The number of temporal patterns for correspondence is variable. Whereas a small number (short pattern sequence) leads to a high 3D frame rate but a higher risk of failing the correspondence, a longer pattern sequence increases the change to find the correct correspondence and improves the measurement accuracy. However more patterns decrease the 3D frame rate. Usually, a minimum of six patterns is necessary for correspondence finding. We use typically ten patterns.

3.2. The NIR GOBO system
The experimental setup of the NIR GOBO system is shown in Fig. 2. A high power LED at 850 nm emission wavelength was used and homogenized with a light funnel made of mirrors with a protected gold coating. The aperiodic sinusoidal patterns were generated using a glass GOBO wheel (according to Heist et al. [1]) on which the pattern was made of a chrome coating. Thus, only the near-infrared radiation is projected into the measurement plane for an irritation-free measurement. Two CMOS cameras, each with a maximum resolution of 4 MPix at a full frame rate of 180 Hz, were arranged in stereo-vision setup with a baseline of approximately 200 mm. To avoid overexposure and disruptive effects from the visible wavelength range, both cameras were equipped with an additional long-pass filter with a cut-on wavelength of 780 nm. The working distance of the system could be scaled from 0.5 m to 2 m in which the measurement area varied from 0.3 x 0.3 m^2 to about 1.5 x 1.5 m^2, depending on the optics. The image acquisition was done by a frame grabber via Camera Link connection.
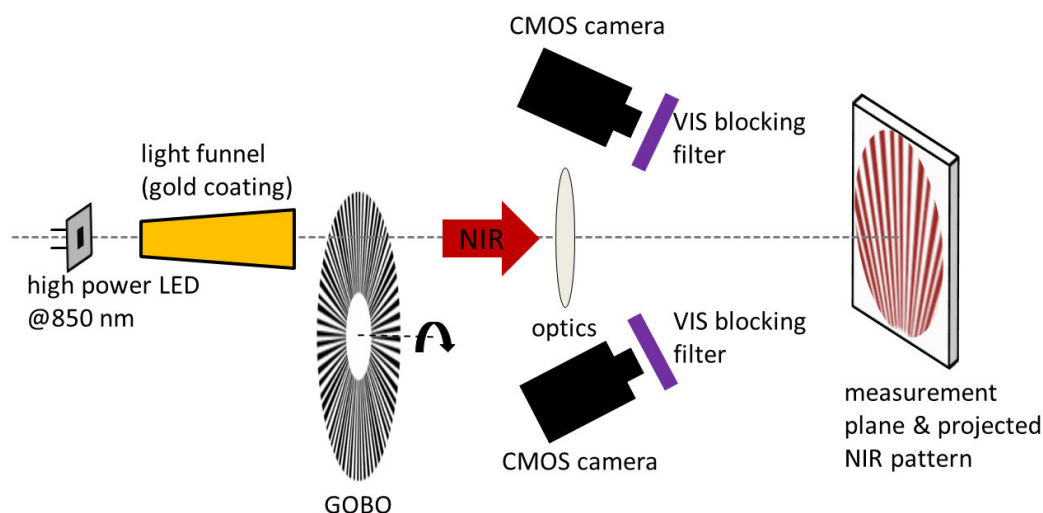

Fig. 2. Setup of the NIR GOBO system

3.3. The NIR multi-aperture system
A second system based on NIR multi-aperture projection is shown in Fig. 3. Overall, it consists of ten apertures for the projection and two NIR enhanced CMOS cameras, each with a maximum resolution of 4 MPix at a maximum frame rate of 90 Hz. Both cameras are equipped with a VIS blocking filter (cut-on wavelength 780 nm) for the detection of the near-infrared patterns. Additionally, the system has a third camera, also with 4 MPix at a maximum frame rate of 90 Hz, which is used to measure only the color information of objects. The triangulation angle between both NIR enhanced cameras is about 8° and the baseline between the two NIR cameras approximately 220 mm. The color camera is placed next to one NIR enhanced camera at a distance of about 40 mm.

Each aperture of the multi-aperture projector is identical and consists of a high power LED with a center wavelength of 850 nm and a maximum total radiant flux of about 1 W (at 1 Ampere and t_{pulse} of 10 ms). In front of each LED, we placed a 40-mm-long light funnel made of glass with a protected gold coating to realize a uniform illumination. Furthermore, each aperture was equipped with its own aperiodic sinusoidal pattern slide made of a borosilicate glass and a chrome coating which was placed directly in front of the light funnel. Using optics consisting of two lenses the pattern was projected onto a measurement plane with a size of about 0.3 x 0.3 m^2 at a distance of about 1.5 m. The LEDs were connected to each other to realize a sequentially and synchronized switching regarding to the cameras. Due to the fact that the RGB pixels of the color camera are not sensitive for 850 nm, the color camera does not see the NIR pattern during the measurements.
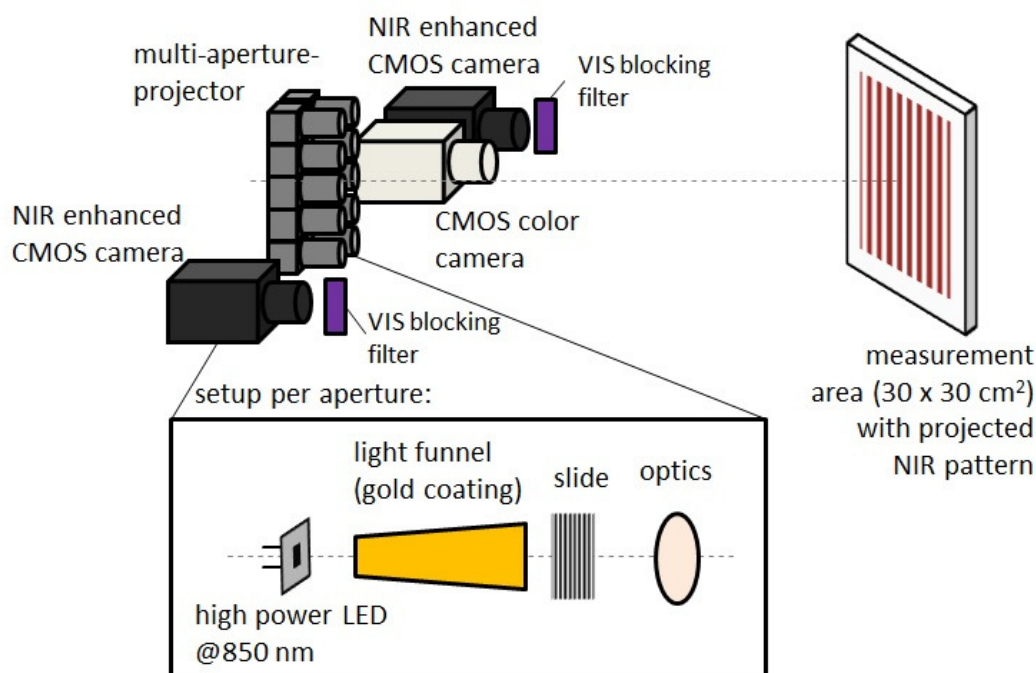

Fig. 3. Experimental setup of the NIR multi-aperture system

4. Measurement examples and accuracy experiments

4.1. Measurement examples using the GOBO system

Several measurements of persons with changing facial expressions were performed. The following example (Fig. 4) shows a sequence over eight seconds showing different emotional expression of a male person. At the margins of the face measurement, artifacts can be observed which should be detected as outliers and removed in a post-processing step.
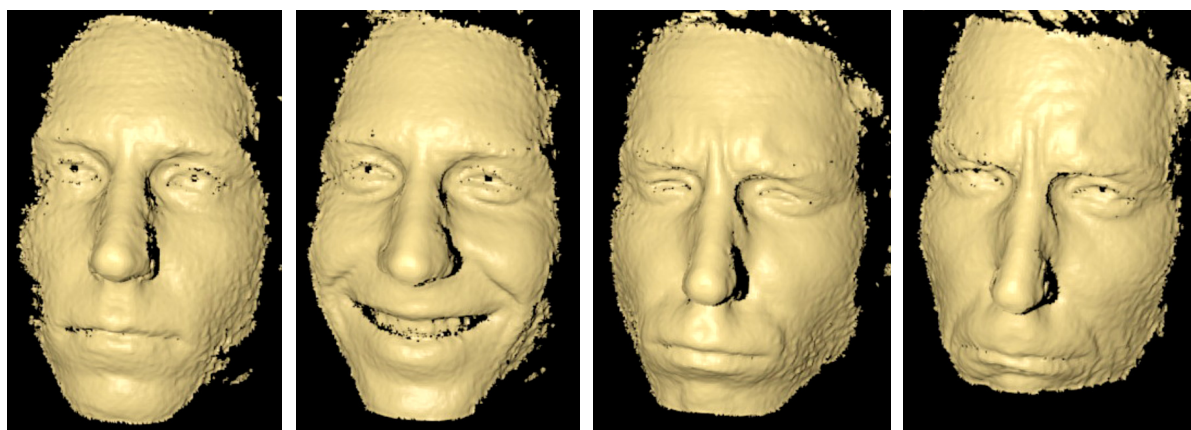

Fig. 4. 3D reconstructed face showing different kinds of emotion: neutral, funny, thoughtful, and angry (left to right)

Additionally, measurements of non-moving human bodies were performed. Figure 5 illustrates the reconstruction process of a back of a person. Here, spatial resolution was about 0.3 mm and the measurement volume was about 600 mm x 600 mm x 200 mm. In the 3D representation measurement errors correlating with the fringe structure of the illumination can be seen. These errors are associated to the actual pattern realized by the GOBO wheel and should also be reduced in the future.

An example of a front view body scan is shown in Fig. 6. Spatial resolution was about 0.5 mm and measurement volume was about 1000 mm x 1000 mm x 400 mm. Here both skin and dressed body parts were captured.
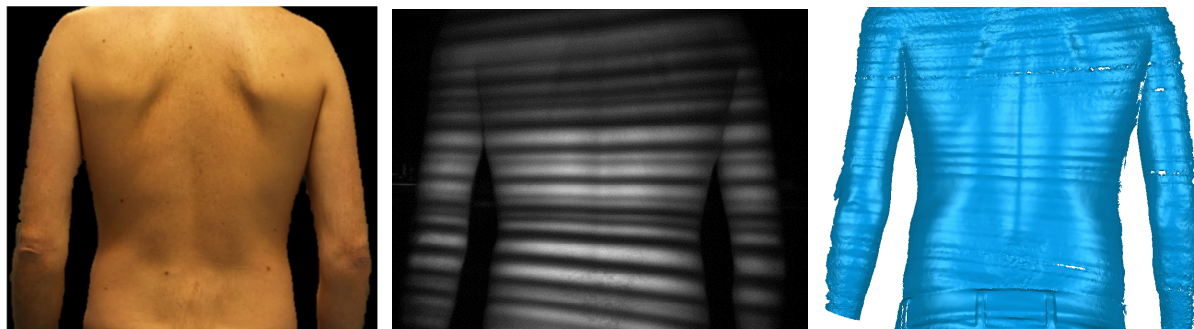

Fig. 5. 3D reconstructed back of a male person: VIS color image (left), IR fringe image (middle), 3D representation (right) of the measurement result
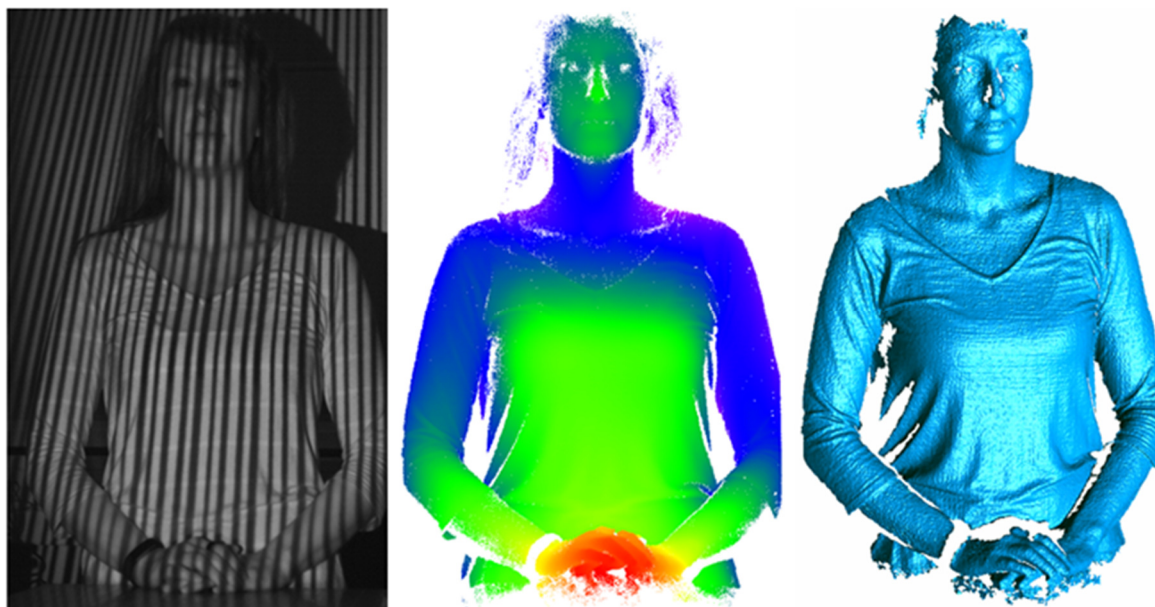

Fig. 6. 3D reconstructed front of a female person: IR fringe image (left), false color (middle) and relief 3D representation (right) of the measurement result

4.2. Measurement examples using the NIR multi-aperture system

Using the NIR multi-aperture setup measurements of persons showing different emotional facial expression were also performed. Additionally, color information was recorded which can be projected onto the 3D shape results. Here the spatial resolution in the object space was about 0.3 mm, the exposure time 3 ms, the sequence length ten images, and the image frame rate 170 Hz. Because of the use of all subsequent images as starting image for the sequences, the 3D frame rate is 170 Hz, too. However, subsequent sequences are not independent from the nine previous and nine following ones.

The following examples (Figs. 7 and 8) show sequence with different emotional facial expressions. Color information was mapped onto the 3D data-points using the additional color camera of the system.

Several points did not show a correct result, especially in the regions with strong gradients and at the hair. These points were left white.
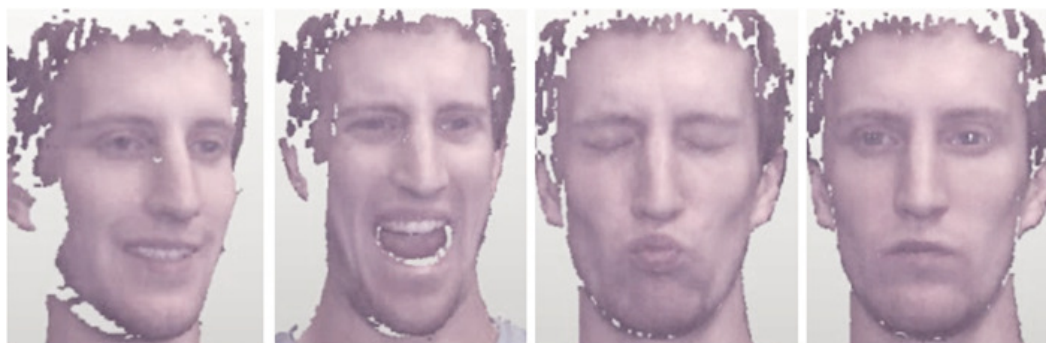
Fig. 7. 3D reconstructed face of a male person with different expression

Fig. 8. 3D reconstructed face of a female person with angry expression

Five different views of always the same 3D dataset are shown in Figs. 9 and 10, respectively. Here, the color mapped 3D point cloud was artificially placed into different directions regarding the observer. In the views deviating from the front view are naturally fewer 3D points to be seen, namely near the neck and at the hair.
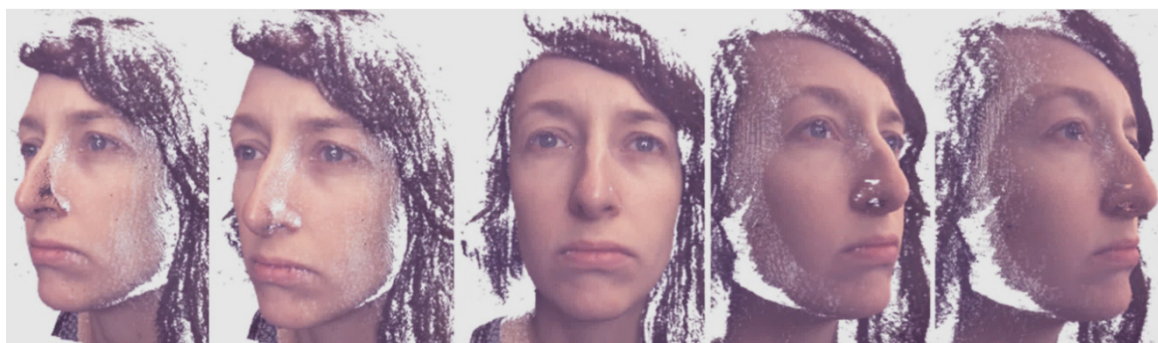

Fig. 9. 3D reconstructed face of a female person (one dataset) represented from five different points of view

Fig. 10. 3D reconstructed face of a female person (one dataset) represented from five different points of view

4.3. Accuracy evaluation

Additional experiments were performed concerning the measurement accuracy. Here, statements concerning random and unknown systematic measurement deviations are desired. The experiments were guided by the VDI/VDE guidelines [19] concerning optical measurement systems. First, the measurement volume MV, in which the accuracy measurements were performed, was selected. We determined the noise of the measurement values at certain regions of the measurement volume. Additionally, length measurement deviation L_d and flatness deviation F_d were determined in order to estimate the unknown systematic error of the setups. We define the length measurement error L_e as the span of the length measurement deviation values out of at least six measurements in different regions of the MV. Flatness deviation was defined (in contrast to the VDI/VDE guidelines) as the range of the signed distances of the measurement points from the best-fit plane calculated according to the least-squares method minus the noise represented by the local standard deviation.

Length measurement deviation was estimated by measurement of a ball bar in six different positions in the measurement volume and flatness deviation was estimated using a plane stone slab. It was placed into the measurement volume at different positions (at least five) covering the whole MV. A plane surface was fitted to the 3D measurement data using the self-made 3D analysis software "ARGUS". Several local measurements of the noise were performed by local fitting of a small plane surface (using at least 1000 measurement points). The noise is represented by the standard deviation σ of the Euclidean distances of the 3D points to the fitted local plane element. Average (weighted) noise values were determined and documented for central and marginal positions in the MV.

Accuracy of the Gobo setup

We set the measurement volume to a cuboid of 250 x 250 x 100 mm³. The spatial resolution was between 140 μm and 155 μm in the measurement volume, the exposure time of the recordings was 10 ms, and the number of images per sequence was ten. The following results of the accuracy measurements were obtained (see table 1). Additionally to the plane stone slab a ball-bar with calibrated length of 199.926 mm between the sphere center points was measured, and the measurement deviation was determined.

Accuracy of the multi-aperture setup

Experiments were performed analogously to those of the Gobo system. We set the measurement volume to a cuboid of 250 x 250 x 200 mm³. The spatial resolution was approximately 135 μm in the object space, the exposure time of the recordings was 10 ms, and the number of images per sequence was ten. The following results of the accuracy measurements (table 1) were obtained.

Table 1: Results of the accuracy measurements of the GOBO and the NIR multi aperture setup

quantity	L_e	F_d	σ_c (central position)	σ_m (marginal position)
Setup				
GOBO	2.1 mm	1.2 mm	40 μm	110 μm
Multi-aperture	0.5 mm	0.4 mm	60 μm	120 μm

5. Summary, discussion, and outlook

In this paper two new 3D scanning setups for the fast acquisition of sequences of moving persons based on structured light infrared projection were introduced. Both setups are suitable for the irritation-free 3D data acquisition of e.g. facial expressions, gestures, or certain body parts. The technical and performance parameters of both systems and measurement examples were given.

First experiments showed the measurement ability of both setups for human motion sequence scanning. Both systems have approximately the same noise. The amount of the noise is in the expected range. Compared to measurement systems in the visible light range, the noise is slightly higher. It would be desirable to reduce it by a suitable method in the future. The systematic error estimation represented by the flatness measurement of the plane and the length deviation measurement of the ball bar led to sufficient results. However, the results of the multi-array tube projector setup were significantly better. This is probably due to a better stereo camera calibration. If higher accuracy is necessary, calibration should be improved.

Facial expression can be detected well in the 3D datasets. Hence, these data may be input of automatic facial expression detection based on 3D data, which may be developed in the future. Remaining random errors leading to noise can be reduced either by application of smoothing filters or by using more input data (i.e. longer sequences of aperiodic sinusoidal patterns). However, the latter

leads to longer processing times and lower 3D frame rates. This may be meaningful if the movement is slow but high measurement accuracy is required.

Remaining problems are the measurement artifacts in the 3D data especially in the data obtained by the GOBO setup. The fringe-like measurement errors are associated with the actual fringe patterns generated by the GOBO wheel and should be compensated by a suitable method. This could be an intelligent filter-operator and should be one of the future tasks. Similar errors are weaker in the results of the multi-aperture setup.

The improvement of the data quality of the 3D results is one of the main tasks for the future work as well as the establishment of a robust and accurate calibration procedure. Additionally, both setups should be tested in different application scenarios, such as gesture or facial expression recognition and human movement analysis in sports or medicine (e.g. in rehabilitation therapy). Another task for future activities is the data recording using larger measurement fields in order to realize 3D whole body acquisition without stitching.

References

[1] S. Heist et al., "High-speed three-dimensional shape measurement using GOBO projection", Optics and Lasers In Engineering 87, 2016, pp. 90–96

[2] R.A. Brooks and G. Di Chiro, "Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging", Physics in Medicine and Biology, Volume 21, Number 5, 1976, pp. 689-732

[3] W.P. Rothwell: Nuclear magnetic resonance imaging. Applied Optics, Vol. 24(23), 1985, pp. 3958-3968

[4] J. Rodriguez-Quinonez et al.: 3D Body & Medical Scanners' Technologies: "Methodology and Spatial Discriminations", in: Prof. Oleg Sergiyenko (Ed.): Optoelectronic Devices and Properties, Intech, 2011, pp. 307-322

[5] F. Bogo et al., "Detailed full-body reconstructions of moving people from monocular RGB-D sequences," in Proc. Int. Conf. Comput. Vis., 2015, pp. 2300–2308

[6] D. Roetenberg et al.: cXSENS TECHNOLOGIES - VERSION APRIL 3, 2013 1Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors, 2013

[7] M. Schaffer et al, "High-speed pattern projection for three-dimensional shape measurement using laser speckles", APPLIED OPTICS   Vol. 49, No. 18, 2010, pp. 3622-3629

[8] M. Große et al, "Fast data acquisition for three-dimensional shape measurement using fixed-pattern projection and temporal coding", Optical Engineering, Vol. 50(10), 2011, pp. 100503-1-3

[9] C. Zuo et al, "High-speed three-dimensional shape measurement for dynamic scenes using bi-frequency tripolar pulse-width-modulation fringe projection," Optics and Lasers in Engineering 51, 2013, pp. 953–960.

[10] C. Bräuer-Burchardt et al.: "High-Speed Accurate 3D Scanning of Human Motion Sequences", Proc. of 6th Int. Conf. on 3D Body Scanning Technologies, 2015, pp. 194-201

[11] B. Fu et al.: Robust near-infrared structured light scanning for 3D human model reconstruction, Proc. SPIE 8979, Emerging Digital Micromirror Device Based Systems and Applications VI, 89790A, 2014, pp. 1-6

[12] M. Hansen et al., "3D face reconstructions from photometric stereo using near infrared and visible light", Computer Vision and Image Understanding 114, 2010, 942–951

[13] G. Sandbachet al., "Static and dynamic 3D facial expression recognition: A comprehensive survey", Image and Vision Computing 30, 2012, pp. 683–697

[14] J. Smisek et al., "3D with Kinect", in: *Consumer Depth Cameras for Computer Vision*. Springer London, 2013, pp. 3-25.

[15] J. Salvi et al, "A state of the art in structured light patterns for surface profilometry," Pattern Recognition 43, 2010, pp. 2666–2680

[16] S. Heist et al, "Array-projection of aperiodic sinusoidal fringes for high-speed three-dimensional shape measurement," Opt. Eng. 53(11), 2014, pp. 112208-1-12.

[17] T. Luhmann et al, Close range photogrammetry. Wiley Whittles Publishing, 2006.

[18] Z. Zhang, "A flexible new technique for camera calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11), 2000, pp. 1330-1334

[19] VDI/VDE 2634. Optical 3D-measuring systems. VDI/VDE guidelines, Parts 1-3, 2008