

## High-Speed Accurate 3D Scanning of Human Motion Sequences

Christian BRÄUER-BURCHARDT<sup>\*1</sup>, Stefan HEIST<sup>1</sup>, Ingo SCHMIDT<sup>1</sup>,  
Peter LUTZKE<sup>1</sup>, Peter KÜHMSTEDT<sup>1</sup>, Gunther NOTNI<sup>1,2</sup>

<sup>1</sup> Fraunhofer Institute Applied Optics and Precision Engineering Jena, Germany;

<sup>2</sup> Technical University Ilmenau, Germany

DOI: 10.15221/15.194 <http://dx.doi.org/10.15221/15.194>

### Abstract

In this work a new high-speed scanning technique based on pattern projection is introduced which allows acquiring human body 3D scans with a frame rate of up to 1.5 kHz. The measurement field and the resolution in the object space are variable. Thus one can either record whole-body scans with a resolution of 2 mm or highly resolved detailed resolutions (of e.g. certain muscle regions) of 300  $\mu\text{m}$ . The 3D scanning methodology is based on fringe projection using a self-made projection unit and a recording using a high-speed stereo camera pair. Movements with velocities of approximately 40 km/h using the highest resolution can be resolved with high accuracy of the 3D data without disturbances by the setup. Hence, an analysis of locally restricted body parts can be performed, e.g. muscle vibrations as well as complex motion sequences, e.g. at certain forms of sport such as baseball, golf, or tennis.

**Keywords:** 3D human motion scanning, fringe projection technique, high-speed-scanning

### 1. Introduction

The 3D scanning of the human body or parts of it is performed for different reasons. Medical applications are either concerned with looking into the body (computed tomography, nuclear magnetic resonance tomography) or scanning the surface in order to get the shape of the body (face scanning). For security access systems 3D face recognition may be important.

New developments in the computer technique allow more and more powerful scanning systems with increasing resolution and high 3D measurement data accuracy. This includes the recording of highly resolved 3D data sequences of moving human bodies in order to get 3D information in the temporal course. Such 3D tracking techniques may be advantageous (more robust, higher precision) over 2D tracking methods because of the additional dimensional information.

Body scanning of humans in motion became more and more important in the last few years. Whereas 2D scanning techniques from video streams can be used for face recognition or person tracking, 3D body scanning may be suitable e.g. in medical applications and physiologic analysis for sports. Medical applications of 3D body scanning in motions could be the observation of muscle activities under load or after stimulation. Quick body movements at sport activities are interesting for physiological analysis of the complex motion sequences.

Available body scanning technique can realize the acquisition of the body motion with a certain frequency. This frequency depends on the technical parameters of the scanning hardware.

Scanning systems which are suitable for 3D human motion recording are e.g. the ProCapture system by Xcitex Inc. [1], Xsens MVN [2] by Xsens Technologies B.V., or iPi Motion Capture [3] by iPi Soft LLC. ProCapture is a video based system which calculates 3D data using up to eight high-speed cameras where a frame rate of 250 Hz can be obtained for 0.83 Mpix image size. The MVN system is based on small inertial tracker sensors which can be placed on the body by wearable straps or a special suite. Output 3D rates between 60 Hz and 240 Hz are available. The iPi system is based on data obtained by Microsoft Kinect cameras [4], Sony PlayStation3 Eye camera, or ASUS Xtion [5].

Harendt et al. [6] introduced a setup with 200 Hz image recording frequency for moving object 3D measurement and showed examples of moving hands.

The biomechanical evaluation of movement in sport is treated by the work edited by Payton and Bartlett [7] who give a survey over several motion capture methods and possible applications in sport. Analysis of sports activities were performed e.g. by Escamilla and Andrews [8] and Seroyer et al [9] using electromyographic (see e.g. [10]) data.

The goal of our work was to develop a 3D scanning system which is able to realize a frame rate of the 3D data of at least 1 kHz using a chip size of 1 Mpix, i.e. about 2 mm spatial resolution of a whole body scan based on pattern projection technique. This would allow us to analyze movements with a maximal velocity of about 40 km/h.

\*christian.braeuer-burchardt@iof.fraunhofer.de; +49-3641-807235; <http://www.iof.fraunhofer.de/>

## 2. The methodology

### 2.1. Pattern projection technique

Fringe projection profilometry (also called phasogrammetry) is one of the established methods of structured light projection techniques [11], which is used in industrial quality management, rapid prototyping, archaeology, cultural heritage preservation, and other fields of application. It has been shown that this technique is suitable for the recording of 3D datasets of moving objects (see e.g. [12,13,14]) including body parts [6,13,15] or persons using different patterns for structured light illumination such as laser speckle patterns [15], band limited stochastic [6,15], or sinusoidal fringe patterns [12].

The basic principle of fringe projection technique is e.g. described by Schreiber and Notni [16]. The final results of this procedure are pairs of image points resulting from the same object point, the homologous or corresponding points, respectively. On the basis of the identified homologous points the calculation of the 3D co-ordinates is done by the triangulation technique (see e.g. [17]).

A modification of the phasogrammetric method [16] was performed in order to find the corresponding points in the second camera image. Finding point correspondences is now realized by temporal correlation of the fringe patterns at each pixel in the image of camera  $C_1$  as described by Heist et al [18] and briefly explained in section 2.3. In this case, aperiodic sinusoidal fringes are projected temporally with the same frame rate as the camera recording in order to realize a distinction of the pixels using temporal sequences. These sequences are typically between six and twelve fringe images long.

For our 3D motion body scanning setup we used a pair of high-speed stereo-cameras and a self-made projection unit which quickly projects aperiodic sinusoidal fringe patterns onto the measurement object. The advantages of this kind of projection over classical fringe projection profilometry are that no phase values are necessary, a much faster projection is possible, and disturbances due to movements can be compensated up to a certain degree. The disadvantage of aperiodic sinusoidal fringe projection in contrast to the classical fringe projection technique is a lower accuracy. However, high precision is not always necessary, especially in the case of moving bodies.

### 2.2. Calibration and image pre-processing

The calibration of the setup is performed the same way as for setups of classical photogrammetry. The projection unit is not calibrated and does not contribute to the calculation of the 3D data. A chessboard was used as calibration pattern and a series of at least 80 image pairs was used as the input of the calibration software. This software is self-made and uses the well-known OpenCV library [19]. The output of the calibration procedure is a set including the extrinsic and intrinsic camera parameters and distortion coefficients.

These data can be used to perform pre-processing before finding point correspondences. Each image of the recorded sequence will first be distortion corrected and second rectified [20] in such a manner that all image lines between camera  $C_1$  and  $C_2$  are corresponding due to epipolar geometry [17].

### 2.3. Detection of corresponding points

The calculation of the searched 3D data will be performed by triangulation [17] using the set of calibration parameters on the one hand and a set of co-ordinate pairs of corresponding image points on the other hand.

Finding corresponding points is performed on corresponding epipolar lines [17] which are equivalent to image rows with the same number in the case of rectified images [20]. For each pixel  $p_i$  in the image of camera  $C_1$  the temporal cross correlation coefficient  $cc(p_i, d)$  concerning a certain disparity  $d$  from the interval  $[d_{min}, d_{max}]$  in the image of camera  $C_2$  is calculated according to

$$cc(p_i, d) = \frac{\sum_{j=1}^n ((x_{i,j} - \bar{x}_i) \cdot (y_{i+d,j} - \bar{y}_{i+d}))}{\sqrt{\sum_{j=1}^n ((x_{i,j} - \bar{x}_i)^2) \cdot \sum_{j=1}^n ((y_{i+d,j} - \bar{y}_{i+d})^2)}} \quad (1)$$

The  $\bar{x}_i$  and  $\bar{y}_{i+d}$  are the temporal intensity mean values over the images  $j = 1, \dots, n$  at  $i$  or  $i+d$ , respectively. The  $x_i$  are the intensity values at  $i$ -th pixel of camera  $C_1$  image and  $y_{i+d}$  are the intensity values at  $i+d$ -th pixel of the camera  $C_2$  image in the corresponding row and  $j$  is the temporal index.

The point with co-ordinate  $q_{i+d}$  with the maximal correlation coefficient is the first candidate of being the corresponding pixel to  $p_i$ :

$$q_{i+d} = \arg \max_{d \in [d_{\min}, d_{\max}]} (cc(p_i, d)). \quad (2)$$

#### 2.4. 3D data calculation

Corresponding points and calibration parameters are the input for the 3D point calculation by triangulation (see e.g. [17]). The 3D point clouds initially contain some outliers. The number of outliers depends on the measurement conditions (e.g. disturbing additional light), surface properties (e.g. reflecting, smooth, light or dark), and the velocity of the object. The outliers are removed by a filtering operator in the 3D space. Afterwards the remaining 3D points can be used to generate a surface structure of the scanned object. See an example in figure 1 which shows the original 3D point cloud, with the point cloud removed outliers, and the 3D surface representation.

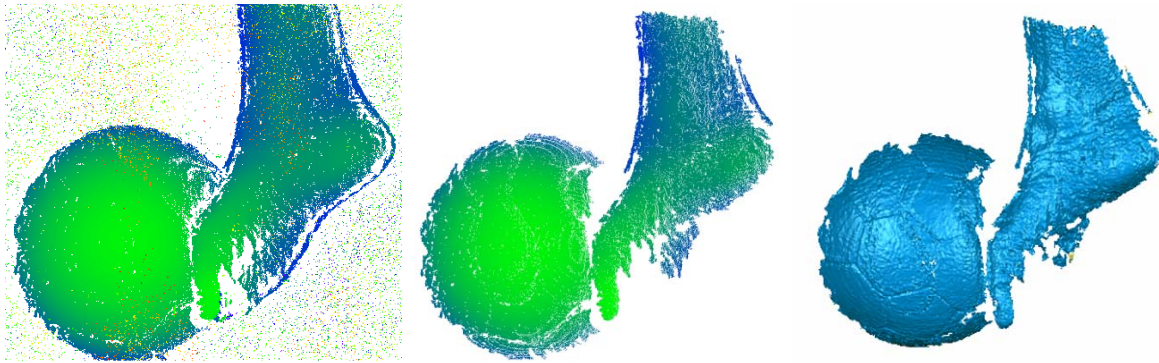


Fig. 1. Color coded 3D dataset with outliers (left), removed outliers (middle), and 3D surface representation (right)

### 3. The scanning setup

As already mentioned the laboratory setup consists of a stereo pair of high-speed cameras and a projection unit (see figure 2). The cameras are of the type Photron Fastcam SA-X2 and the lenses are AP-S Nikkor 35/1.8.

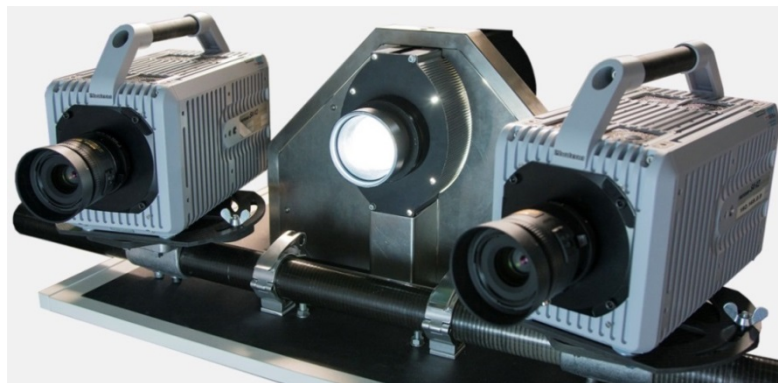


Fig. 2. 3D scanning setup including two cameras and the self-made projector

The maximal resolution of the cameras is 1 Mpix (1000 x 1000 pixel). Images with this resolution can be recorded with a maximum frame rate of 13.5 kHz. Images with lower resolution can be acquired with up to 60 kHz. The two cameras record the images absolutely the same time and with the same exposure time.

Projection is performed by a rotating wheel with slits which produces aperiodic sinusoidal fringes. Synchronization between the projector and the cameras is not necessary.

Measurement fields of sizes between 0.5 m x 0.5 m and 2.0 m x 2.0 m are possible. The basic length of the stereo rig can vary between 0.4 m and 1.0 m.

The storage recording capacity of a real-time measurement is currently 8 GB per camera. This equals for a full resolution recording at 13.5 kHz a length of the 3D point-cloud sequence of about 0.5 seconds.

This is, of course, very short. However, for the analysis of certain movements (throwing procedure, muscle vibrations, muscle constriction process) this might be sufficient.

In order to record the motion profile over a longer time period application of lower temporal resolution is possible. For instance, for 50 3D frames per second based on ten-image sequences for temporal correlation, a recording length of approximately 13 seconds can be achieved. However, with modified hardware components the recording capacity may be improved.

## 4. Measurement examples

### 4.1. Foot movement recording on a treadmill

The first experiment shows the foot of a runner on a treadmill. The scanner was placed in such a way that the angle between the projection direction and the running direction was approximately  $45^\circ$ .

The measurement field was about 0.5 m x 0.5 m and the maximal resolution of 1 Mpix was used. Image recording frame rate was 12 kHz and twelve temporal consecutive images were used to calculate one 3D scan. This corresponds to a 3D data frequency of 1 kHz.

Figure 3 shows 3D representations of selected time points from the beginning to the end of the recorded sequence with duration of 0.2 seconds. Figure 4 shows 3D representations of selected time points from time point  $t_{110} = 0.11$  s to time point  $t_{150} = 0.15$  s. The time step from view ( $t_i$ ) to view ( $t_{i+1}$ ) is 1.0 ms. Figure 5 shows a 3D surface representation of the same scans shown in figure 4.

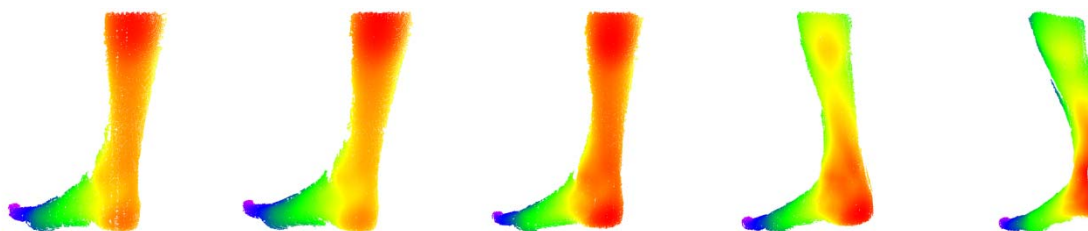


Fig. 3. Color-coded 3D representation of a 3D scan sequence from  $t_0 = 0.0$  s to  $t_{200} = 0.2$  s (from left to right)

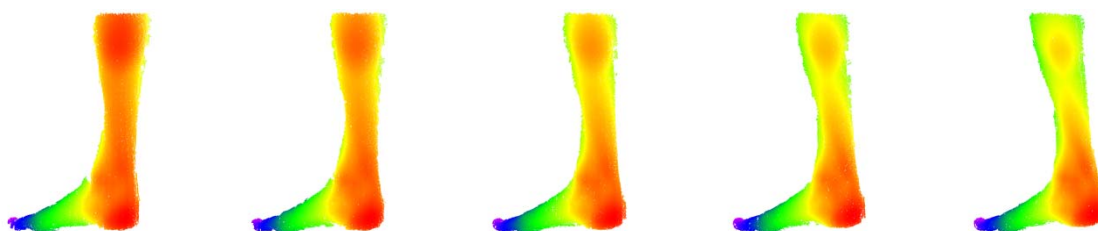


Fig. 4. Color-coded 3D representation of a 3D scan sequence from  $t_{110} = 0.11$  s to  $t_{150} = 0.15$  s (from left to right)

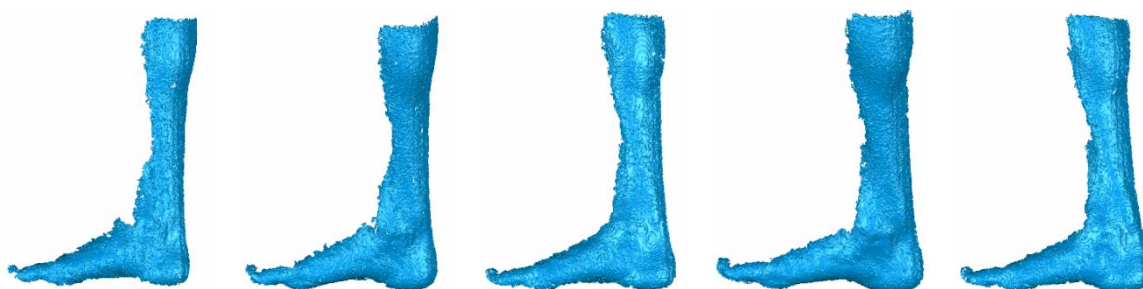


Fig. 5. 3D surface representation of a 3D scan sequence from  $t_{110} = 0.11$  s to  $t_{150} = 0.15$  s (from left to right)

### 4.2. Football kick recording

The second experiment shows the kicking of a ball by a naked foot. The same parameters as in the first experiment were used, except nine temporal consecutive images were used to calculate one 3D scan. This results in a 3D frame rate of about 1.3 kHz. A rough analysis of consecutive 3D datasets led to an estimated maximal velocity of the foot of about 30 km/h. The maximal ball velocity was approximately 45 km/h. In the following figures the moment of the kick can be watched in detail.

Figure 6 shows 3D color coded representations of selected time points from  $t_0 = 0$  ms to  $t_{50} = 37.5$  ms of the recorded sequence. Even the deformation of the football can be recognized. The time step from view to view is 0.75 ms. Figure 7 shows 3D color-coded representations of selected time points from time point  $t_{10} = 7.5$  ms to time point  $t_{20} = 15$  ms. Figure 8 shows 3D surface representations of selected time points from  $t_{10} = 7.5$  ms to  $t_{45} = 33.75$  ms. The time step from view ( $t_j$ ) to view ( $t_{j+1}$ ) is always 0.75 ms.

The touch of the ball is between time stamp  $t_{12}$  and  $t_{18}$  (see figure 7). In the 3D surface representation (figure 8) it can be seen that the unmoved ball is reconstructed with only small disturbances (pentagon and hexagon structures are well recognizable). Even shortly after the shot ( $t_{20}$ ) the structures are to be seen in figure 8. This means that the ball velocity is still low. At time stamp  $t_{30}$  the ball is already in faster motion and probably flexibly deformed. The 3D reconstruction quality of the ball with high velocity (time stamp  $t_{40}$  and  $t_{45}$ ) is poor due to the velocity, which was estimated to be approximately 45 km/h. The acceleration of the ball is documented by the estimated velocities between time stamp  $t_{12} = 9$  ms and  $t_{44} = 33$  ms (see table 1). Table 1 also contains the results of the diameter measurements of the ball and the estimated foot velocities which go down after the kick to about 20 km/h.

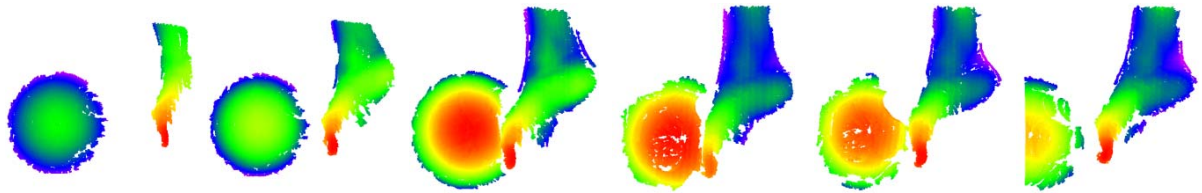


Fig. 6. Color coded 3D representation of a sequence of 3D scans from  $t_0 = 0$  ms to  $t_{50} = 37.5$  ms (from left to right)

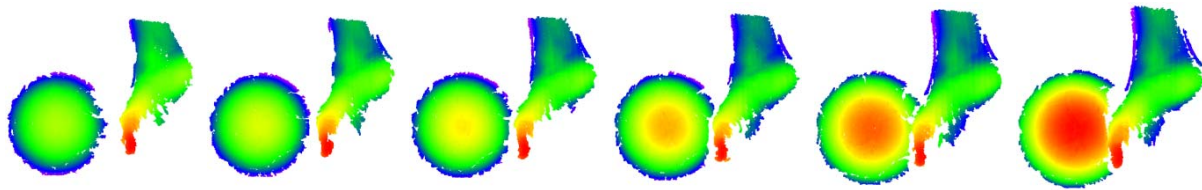


Fig. 7. Color coded 3D representation of a sequence of 3D scans from  $t_{10} = 7.5$  ms to  $t_{20} = 15$  ms (from left to right)

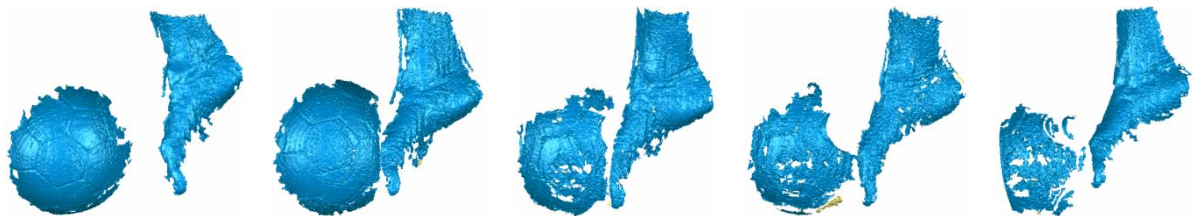


Fig. 8. 3D surface representation of a sequence of 3D scans from  $t_{10} = 7.5$  ms to  $t_{45} = 33.75$  ms (from left to right)

Table 1. Measurement values for ball diameter, ball velocity, and foot velocity.

no	time [ms]	$D_{ball}$ [mm] $\pm$ SD	$v_{ball}$ [km/h]	$v_{foot}$ [km/h]
12	9	221.0 $\pm$ 1.1	0.2	30,0
16	12	221.1 $\pm$ 1.2	0.1	29.6
20	15	220.5 $\pm$ 1.7	1.2	29.9
24	18	224.5 $\pm$ 2.0	9.9	32.6
28	21	223.5 $\pm$ 2.0	19.0	24.5
32	24	222.6 $\pm$ 2.2	32.9	25.5
36	27	221.2 $\pm$ 2.4	43.7	20.3
40	30	219.9 $\pm$ 2.1	45.8	21.5
44	33	219.3 $\pm$ 2.4	46.1	20.6

Figure 9 shows the velocities of the ball and the foot depending on time and a graphic of the ball diameter depending on time. The acceleration phase of the ball can be well recognized being approximately between 15 and 27 ms. The same time a slowdown of the foot velocity can be detected.



The ball diameter measurement is quite stable in the time of rest (constant diameter, low standard deviation). In the dynamic phase the measurement values are changing due to deformation and motion.

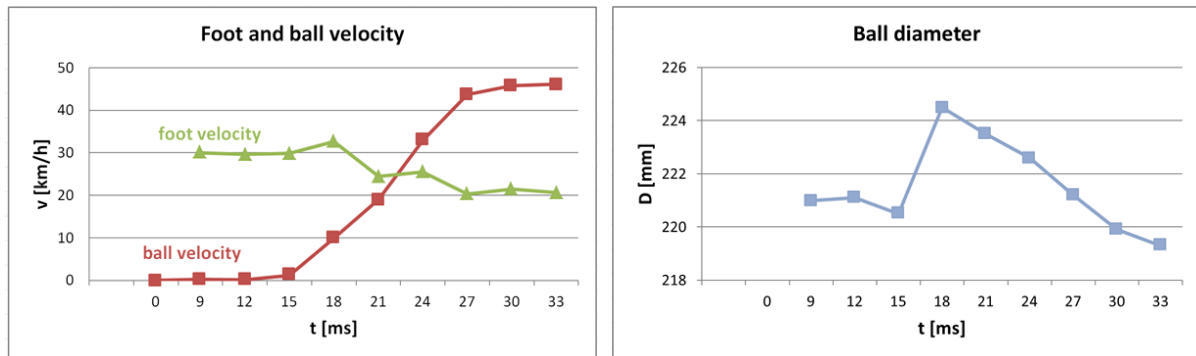


Fig. 9. Velocity of the ball (red squares) and the foot (green triangles) from  $t_{12} = 9$  ms to  $t_{44} = 33$  ms (left) and diameter measurement of the ball (blue squares) from  $t_{12} = 9$  ms to  $t_{44} = 33$  ms (right)

#### 4.3. Upper body motion acquisition of a thrower

The next experiment shows a whole-body measurement of a ball thrower. The measurement field was about 2.0 m x 2.0 m and the maximal resolution of 1 Mpix was used. Image recording frame rate was 12 kHz and twelve temporal consecutive images were used to calculate one 3D scan. This means a 3D data frequency of 1 kHz.

Figures 10 and 11 show color-coded and surface 3D representation of the upper body of the thrower in the initial phase of the throw over a period of about 0.12 seconds. It is to be seen how the throwing hand is moving back.

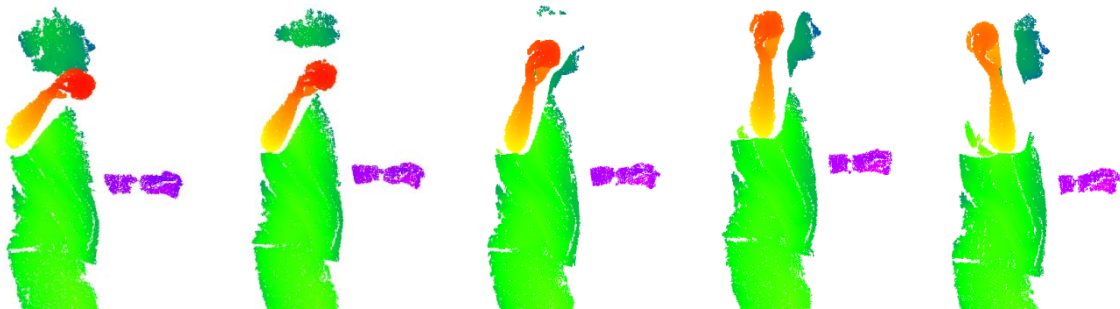


Fig. 10. Color coded 3D representation of a 3D scan sequence from  $t_0 = 0$  ms to  $t_{160} = 120$  ms (from left to right)

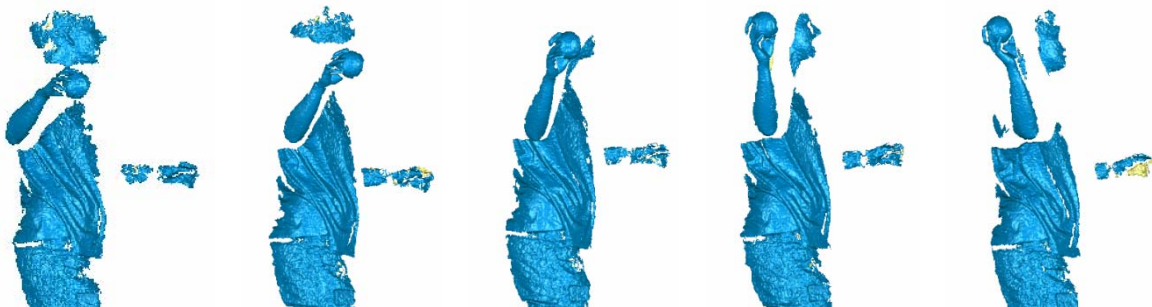


Fig. 11. 3D surface 3D representation of a 3D scan sequence from  $t_0 = 0$  ms to  $t_{160} = 120$  ms (from left to right)

#### 4.4. Whole body recording of a rope skipper

The last experiment shows a rope skipper.

Figure 12 shows a color-coded 3D representation of the rope skipper over a period of about 0.04 seconds. Because of the short recording time only a small body movement is visible which is best seen at the height change of the arms.

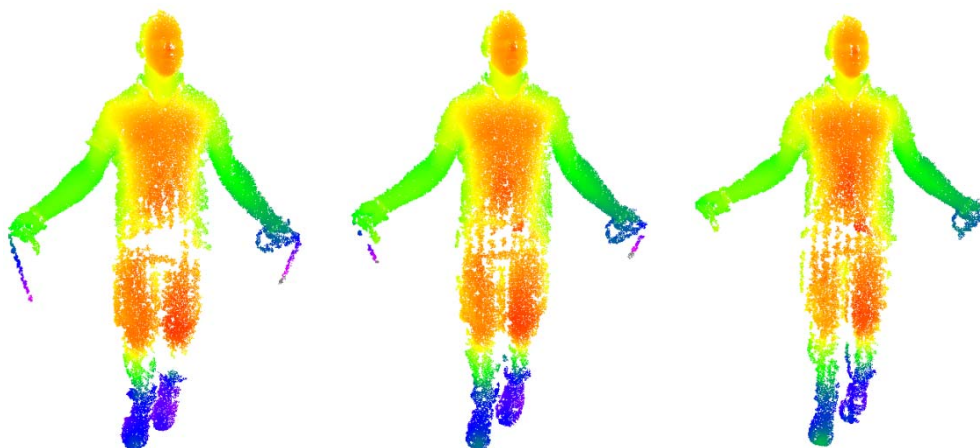


Fig. 12. Color coded 3D representation of 3D scans at  $t_0 = 0$  ms (left),  $t_{28} = 21$  ms (middle), and  $t_{56} = 42$  ms (right)

## 5. Discussion

The results show that 3D motion recording using structured light projection technique is possible. High 3D rates of  $>1$  kHz can be obtained with the presented setup under the limiting conditions of recording duration ( $<0.5$  sec) and maximal object velocity of approximately 40 km/h. This makes the analysis of 3D data in a very short time period possible. For example, as the recording of the football kick shows, there is a very short time period where the ball is touched and moved by the foot, where the ball is deformed. A possible application of the 3D analysis would be the volume change of the ball or the foot. This cannot be performed using a lower 3D data rate of e.g. 50 Hz.

However, as the results also show, the 3D data quality is not yet high enough for a precise 3D-measurement, e.g. of the ball volume or for the analysis of small body regions, e.g. muscles. This is either due to smearing effects of the objects velocity (30 km/h means an objects movement in the time of the nine-image sequence-recording of about twelve pixels!) and otherwise due to measurement errors. Here better recording conditions (stronger illumination, smoother and lighter object surfaces) and refined algorithms may lead to improvements in the 3D data quality. Perhaps, a combination of spatial and temporal correlation (see [6]) may lead to better 3D reconstruction results at reduced sequence length for one 3D scan. For extremely fast movements single-shot techniques (see [21]) may be applied, if the demands concerning the accuracy of the 3D reconstruction are not too high.

However, calculation effort is rather high and restricts such application to short period events such as single movements in sport (throw, kick, beat). Additionally, short dynamic processes involving other measurement objects such as airbag unfolding can be also analyzed by this technique.

## 6. Summary and outlook

A setup for high-speed accurate 3D scanning of human motion sequences was presented. We have shown selected examples of recording of 3D sequences of human body motions, such as runner on a treadmill, kicking feet, ball throw, and rope skipping. These examples give an impression of the possibilities of the presented technique concerning human motion 3D measurements in the fields of medical application, physiological analysis of different kinds of sports, injury prevention in sports, and others.

Future work should be addressed first on an easier handling of the data storage procedure in order to be able to record longer 3D data sequences within a reasonable time. Secondly, a solution for the fast change of the measurement field size should be developed in order to achieve optimal results in completeness, resolution, as well as measurement accuracy. This would be the precondition for the best analysis and interpretation of the 3D results of human body motion measurements.

A third goal for the future work would be the possibility of the 3D data acquisition of faster moving objects, i.e. velocities faster than 20 km/h for high resolution measurements and faster than 50 km/h for large measurement fields. This can be obtained either by enhancement of the frame rate of the recording or by reduction of the sequence length for the 3D point calculation. Sequence reduction may be obtained e.g. by combination of spatial and temporal correlation techniques (see [6]) or the use of single-shot methods (see e.g. [21]).

If an improvement of the 3D measurements concerning amount of data and measurement accuracy is achieved, then a series of further experimental measurements should be performed and analyzed.

## References

- [1] [www.xcitex.com](http://www.xcitex.com), accessed 2015/07/16, 9:31.
- [2] [www.xsens.com](http://www.xsens.com), accessed 2015/07/16, 9:32.
- [3] [www.ipisoft.com](http://www.ipisoft.com), accessed 2015/07/16, 9:38.
- [4] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia Mag.*, vol. 19, no. 2, 2012, pp. 4–10.
- [5] [https://www.asus.com/de/Multimedia/Xtion\\_PRO\\_LIVE/](https://www.asus.com/de/Multimedia/Xtion_PRO_LIVE/), accessed 2015/07/16, 9:52.
- [6] B Harendt et al, "3D shape measurement of static and moving objects with adaptive spatiotemporal correlation", *Applied Optics*, Vol. 53, No. 31/1, 2014, pp. 7507-7516.
- [7] C.J. Payton and R.M. Bartlett, *Biomechanical Evaluation of Movement in Sport and Exercise*, Routledge, New York and London, 2008.
- [8] R.F. Escamilla and J.R. Andrews, "Shoulder Muscle Recruitment Patterns and Related Biomechanics during Upper Extremity Sports", *Sports Med* 2009; 39 (7), 2009, pp. 569-590.
- [9] S.T. Seroyer et al, "The Kinetic Chain in Overhand Pitching: Its Potential Role for Performance Enhancement and Injury Prevention", *Sports Health* vol. 2, no. 2, 2010, pp. 135-146
- [10] G. Kamen, *Electromyographic Kinesiology*, in Robertson, DGE et al. *Research Methods in Biomechanics*. Champaign, IL, Human Kinetics Publ., 2004.
- [11] J. Salvi et al, "A state of the art in structured light patterns for surface profilometry," *Pattern Recognition* 43, 2010, pp. 2666–2680.
- [12] 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.
- [13] C. Zuo et al, "High-speed three-dimensional profilometry for multiple objects with complex shapes," *OPTICS EXPRESS*, Vol. 20, No. 17, 2012, pp. 19493-19510
- [14] 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
- [15] 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
- [16] W. Schreiber, and G. Notni, "Theory and arrangements of self-calibrating whole-body three-dimensional measurement systems using fringe projection techniques", *Optical Engineering* 39, 2000, pp. 159-69
- [17] T. Luhmann, et al, *Close range photogrammetry*. Wiley Whittles Publishing, 2006.
- [18] 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.
- [19] [http://docs.opencv.org/modules/calib3d/doc/camera\\_calibration\\_and\\_3d\\_reconstruction.html](http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html), accessed 2015/07/16, 11:15.
- [20] D. Oram, "Rectification for any epipolar geometry," *Proc BMVC*, 2001, pp. 653 -662.
- [21] C. Bräuer-Burchardt et al, "3D reconstruction with single image pairs and structured light projection for short-term ultra-high-speed applications", *Proc SPIE OM*, Vol. 9528, 2015, pp. 952808-1-10.