We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,000
Open access books available

148,000
International authors and editors

185M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Usage of RGB-D Multi-Sensor Imaging System for Medical Applications

Libor Hargaš and Dušan Koniar

Abstract

This chapter presents an inclusion of 3D optical (RGB-D) sensors into medical clinical practice, as an alternative to the conventional imaging and diagnostic methods, which are expensive in many aspects. It focuses on obstructive sleep apnea, the respiratory syndrome that occurs in an increasing proportion of the population, including children. We introduce the novel application, a response to the request for an alternative pre-diagnostic method for obstructive sleep apnea in the region of Slovakia. The main objective of the proposed system is to obtain an extensive dataset of scans (head and face) from various views and add detailed information about patient. The application consists of the 3D craniofacial scanning system using multiple depth camera sensors. Several technologies are presented with the proposed methodology for their comprehensive comparison based on depth sensing and evaluation of their suitability for parallel multi-view scanning (mutual interference, noise parameters). The application also includes the assistance algorithm guaranteeing the patient's head positioning, graphical interface for scanning management, and standardized EU medical sleep questionnaire. Compared to polysomnography, which is the golden standard for this diagnostics, the needed data acquisition time is reduced significantly, the same with the price and accessibility.

Keywords: RGB-D sensors, multi-sensor system, 3D imaging, medical applications, obstructive sleep apnea, time of flight sensor, structured light sensor, stereo vision

1. Introduction

Obstructive sleep apnea syndrome (OSAS) is a common sleep disorder with arising prevalence. The course of the disease is followed by repeating breath interruptions during sleep. The reason is the collapse of soft upper airway tissue. This restriction of ventilation results in several breathing difficulties, such as snoring during sleep and hypoxemia. OSAS also leads to long-term changes in autonomous functions, hypertension, or reduced left ventricular function. As a result of the propagation and activation of inflammatory pathways, immune regulation is often disrupted in pediatric patients. Early diagnostics and relevant treatment are the keys for improve the health status of patients.

Polysomnography (PSG) is the golden standard for OSAS diagnostics [1]. PSG is performed for whole night in specialized sleep laboratories and the PSG device is continuously monitoring selected vital functions, such as ECG, EEG, thoraco-abdominal movements, nasal airflow, blood oxygen saturation, snoring, or body position. OSAS is confirmed if the number of apnea episodes is higher than 15 during a night and the episode takes more than 10 seconds [1, 2]. The apnea-hypopnea index (AHI) is calculated, which is used to indicate the severity of the disease.

Diagnostics and care of patients with obstructive sleep apnea vary by country and depend on the patient's symptoms. Available data suggest that most cases remain undiagnosed and untreated even in developed countries, which can increase the risk of cardiovascular, metabolic, and neural diseases and affect the quality of life. Generally, PSG is a time-consuming procedure carried out using specialized equipment, so there will be always a patient limit for undergoing the diagnostic test. As an example, there is only one specialized child and adolescent sleep laboratory in Slovakia for approximately one million children under 19 years.

Therefore, there is a reasoned demand for alternative or pre-diagnostic testing that will distinguish the patients with a high risk of OSA. Today's PSG testing can be also proved by telemedicine. Although it holds great promise to change health care delivery, it has not been proven to have the same accuracy as conventional PSG. Besides the conventional PSG, there are several supplementary testing methods. One of the best known is the sleep questionnaire [3–5] focused on the medical history of the patient and the physical examination. Another supportive diagnostic tool is the pulse oximetry or examination of a specific protein in blood serum. Mentioned questionnaires evaluate the subjective and objective symptoms, as well as the craniofacial and intraoral anatomy. These structures are considered an important indicator of the predisposition of the OSAS [6]. Nowadays there is an effort to make diagnostics more available, therefore, the emphasis is placed on the use of fast imaging techniques.

Many recent publications [7–9] focus on the face anatomy (craniofacial anthropometry, structure of the soft tissue in oral cavity, and anterior neck subcutaneous fat tissue thickness) and use advanced imaging techniques such as X-ray [10], MRI [11], and CT [12, 13]. Although, mentioned modalities do not match the criteria for cost reduction, faster procedure, and simplification of the clinical examination. Last but not least, the speed of scanning process is very important to avoid motion artifacts in resulting 3D models, especially if the system is dedicated to pediatric medicine.

As an alternative, we present an optical depth multi-sensor system that can be used excluding other emerging disadvantages with lower quality of an output model. Optical depth sensors allow capturing the nature of craniofacial anatomy needed for prediction of OSAS, such as shape and contour in a faster, cheaper, and more readily available way, compared with the other imaging techniques [14]. The geometrical precision of an output model is the key attribute for the desired application. It is also the goal of the proposed multi-camera parallel scanning system — to reconstruct a complete 3D model of the object from a collection of images taken from known camera viewpoints. Therefore, it is important to choose a suitable optical sensor with the least measurement error. For a real application, the main requirement is to obtain a complete 3D model without any noisy artifacts. In this work, we aim to evaluate each of the camera technologies: The Intel® RealSense™ Depth Camera D415 Series sensors, Stereolab's ZED Mini depth camera, Microsoft Kinect for Windows V2, and Intel® RealSense™ Camera SR300 and offer a comparison of individual operating technologies. With the selection of a suitable optical sensor, also the fact that the scanning object is a pediatric patient is taken into account.

Based on accuracy measurements, we prefer active stereo pair technology. The design, including the optimal topology of used cameras, user interface, and implementation of conventional OSAS screening questionnaire is introduced. Our effort is to predict the probability of OSAS occurrence without the need for traditional polysomnography testing. For this purpose, in the first stage of research, we use the scanning system primarily to obtain the 3D models of the patient's head, and subsequently, the database of point cloud models will be created for further research (automated extraction of key points in the face and head and automated measurement of geometric dependencies indicating the risk of OSAS). Currently, the absence of the database of 3D scans is the crucial limitation of OSAS data processing and assessment. Many studies dedicated to automated diagnostics of OSAS suffer from the datasets with small numbers of images and models. For this reason, building a huge datasets of 3D scans taken from various points of view with additive information about the patient is one of the main objectives of our research.

Additive information is a de facto electronic version of an internationally standardized sleep questionnaire. This dataset will be used for further automated diagnostics and research in this field. The result of our work is the system that consists of a fixing stand (that allows changing the camera layout) and software web-based application (includes the data annotation and the assistance support system) that helps the operator to set the patient's head into the normalized position. Using the advantage of machine learning it seems to be possible to evaluate the presence of OSAS from the point cloud representation of the patient's head and neck [15]. In the future, we assume the use of obtained dataset (composed of different views and facial expressions) with additive information in OSAS automated diagnostics. The experimental system is located in Martin University Hospital in Slovakia, Clinic of Children and Adolescents.

2. Related research in the field

Nowadays, the research of the new predictive diagnostic method of OSAS based on the 2D or 3D craniofacial image of the patient uses the advantage of machine learning, artificial intelligence, or statistical analysis. This research is based on the fact that OSAS occurrence is correlated with many diseases and syndromes (obesity, Down syndrome, adenotonsillar hypertrophy …) manifesting on the head and face. Many modern diagnostic approaches for OSAS use automated detection of selected points on the head and face (e.g. eye corners, lips, earlobes, chin …) and measure distance between selected points (**Figure 1**). Description of head and face with given distances serves as basic for classification process to compare with normal (physiological) model. Craniofacial points and their distances correspond with metrics obtained from paper sleep questionnaires dedicated to computing the score of OSAS risk.

Frontal and profile 2D facial photographic images of the control and experimental group are used in Ref. [16], and the features and landmarks were identified. The features were processed by support vector machine (SVM) classifier and get the resulting accuracy of 80% correct OSAS detections. In Ref [17], the authors use the training set of 3D images of 400 patients. All of them were identified with PSG, AHI index and were divided into 4 groups. The landmarks were identified manually and were expressed as the Euclidian and geodetic distance between them. The distances were considered as the features for further OSAS classification. In Ref [18], for OSAS distinction geometrical morphometry is used, also via 3D photography; in Ref [14]
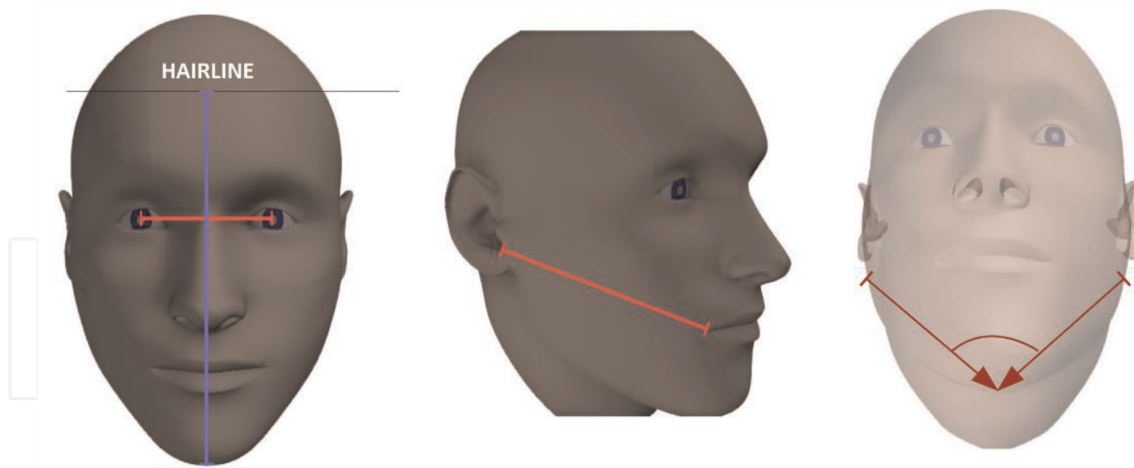
Figure 1.
Selected examples of craniofacial measurement.

convolutional neural network is a tool for OSAS prediction. Although the network was not trained on the depth data, the pretrained network achieved the 67% accuracy. Taking into account the information from previous research in this area, we can say that 3D model of the face and neck of the patient contains sufficient shape and structural features to determine the OSAS prediction. In most cases, the studies work with limited datasets, small groups of patients, or use only the frontal 3D scan of patients. As an alternative, we offer a standardized and well-described 3D model acquisition scanning system, applied in the clinical environment.

3. RGB-D sensors

Solving image processing tasks in various fields of research [19–21] is often helpful to obtain depth information for a better description of the scene in addition to color information. The goal is to capture the geometrical nature of the real-world object and convert it to the digital format with the highest possible accuracy. For obtaining mentioned information the depth sensors (RGB-D sensors) are widely used. They can convert the scene to the 2D plane called depth map. The depth map can be converted back to 3D space using reversed reconstruction. The depth map is usually represented by a monochromatic image, where the intensity value of the pixel represents the distance of the corresponding point from the imaging sensor. Using a combination of depth maps and color RGB images we can create a textured 3D model of scene. One of the novel application areas is the reconstruction of the 3D surface in medical research (scanning of the human heads, faces, or other body parts), taking the advantage of the noninvasive nature of digital imaging. A geometrically accurate model of head is applicable in medicine for predicting various diseases, e.g., respiratory syndromes, where the 3D representation of the patient's head and neck offers detailed visualiza-tion of craniofacial parameters with a given accuracy. In the field of 3D imaging, we know many principles and methods, e.g., photogrammetry or laser scanning devices. These methods provide high-quality 3D information; on the other side, their applica-tion is limited by the size of scanned object, size of scanner, or both object and scanner, these devices are often expensive and scanning time is too high. Also in most cases, laser scanners are not eye-safe. In the next sections, the basic principles of RGB-D sensors will be described.

3.1 Time of flight RGB-D sensor

The time of flight (ToF) RGB-D sensors are optical sensors that measure the depth of scene using an active light source. This light source emits an amplitude modulated signal. The emitted signal can be continuous or impulse. Most ToF cameras generate amplitude-modulated continuous waves (AMCW) with a frequency near IR for illuminating the scene [22]. The depth of scene is based on measurement of the amplitude of phase shift between received and transmitted modulated signal. The depth information for each pixel can be calculated by the synchronous demodulation of the received modulated light in the detector. The demodulation can be performed by interleaving with the original modulated signal.

3.2 Stereo RGB-D sensor

3.2.1 Passive stereo sensor

Passive stereo vision RGB-D sensors reproduce the depth of the scene the same way as the binocular vision of a human. The scene must be captured from different points of view. This is done by using two RGB sensors (corresponding to human eyes) horizontally separated by known distance. This distance is called baseline. For example, the depth sensor ZED MINI used in our experiments has two RGB sensors separated by a 12 cm baseline. The depth of scene is then computed based on disparity of corresponding points in single views. Solving the correspondence problem means giving a point in the image and finding the same point in another image [23]. Stereo sensors use computationally intensive algorithms to search for point matches and for computation of depth. These sensors are suitable for environments with good lighting conditions including outdoors.

3.2.2 Active stereo sensor

If the scene contains fewer color and intensity variations, or lighting conditions are not good, the passive stereo vision system can be less effective and accurate. The typical example of such environment is the texture-less surface like indoor dimly lit white walls. Active stereo vision relies on the addition of an optical projector that overlays the observed scene with a semi-random texture that facilitates finding correspondences. The current generation of RealSense D4xx sensors working in bright environments captures the texture of objects in really slight details and they are applicable also outdoors. In the case of scanning dynamic objects using the multi-sensor system, there is no limitation on how many sensors are used in a given physical layout. It does not decrease the quality of scanning process if several sensors project their light patterns to the same part of scene. All additional projectors actually improve the overall performance by adding more light and more texture [24].

3.3 Structured light RGB-D sensor

Depth sensors based on structured light (SLS) need additive light source. This source projects the regular patterns to the scene. The surface of the object distorts this regular pattern. If the structure of light pattern is known, the depth image of the scene can be easily computed based on its distortion [23]. Light patterns are emitted in infrared band, so the entire process is invisible to the user. If stripes are used as regular

patterns, the optical resolution of the depth map can be increased by the reduction of strips width.

4. Measuring of accuracy of RGB-D sensors

In this section, the methods and basic measurements are described leading to proper RGB-D sensor selection, that allows capturing objects in the parallel multi-view system. Parallel multi-sensor (multi-view) system is required to reduce the scanning time (because object of interest is a pediatric patient and its potential motion can cause artifacts in resulting 3D model). Parallel means that all sensors in a given topology are capturing the object at the same time. The partial views (depth maps or point clouds produced by single sensors) must be then registered (aligned) and joined into final 3D model.

In our previous study [25], we described interference artifacts, which occur while scanning using ToF sensors. The interference is present also using several SLS sensors: projected patterns are overlaid on the surface of objects. A passive or active stereo camera pair is the technology that, in principle, does not suffer from interference in parallel multi-view systems. The depth scanning precision of sensors is compared in several recent works [26]. Known methodologies for error estimation of sensors often use a precise object and its digital model as ground truth, which is difficult to obtain. The main benefit of versatile methods described in this section is a comprehensive comparison of all sensor technologies. The measurement is based on capturing testing patterns on surfaces at small distances. The ToF sensors seem to be more accurate against the passive or active stereo pairs, according to the recent works [27, 28]. The next contribution of this research to the practice is the evaluation of the differences between these technologies in terms of accuracy. The results should show if the stereo pairs can achieve similar depth-sensing accuracy as ToF or structured light sensors at small distances.

4.1 The noise measurement

The noise measurement is a simple method based on evaluation of time variability of single points in the depth map. The depth is measured against a flat surface at several given distances. Depth variability (or standard deviation of the depth) in given points is represented as the noise of a depth sensor. Obviously, the noise increases with the sensor-to-object distance. All sensors were placed at distances of 0.5 m, 0.7 m, and 1 m from the flat surface. The scene was captured in 10 seconds.

4.2 Ideal cloud fitting

Another metric for depth error estimation is a simplified technique based on the methodology of the study [29]. Study [30] brings another technique that can be used as a generalized method for depth error estimation for any device.

The depth error estimation is based on comparing two point clouds. First point cloud is created from captured depth map and the second is the ideal software model, as shown in **Figure 2**. The results of the following measurements are extracted from our study [31]. The testing pattern is the chessboard 9 × 7 squares (square side length is 36 mm). The ideal reference point cloud was generated with the same dimensions.
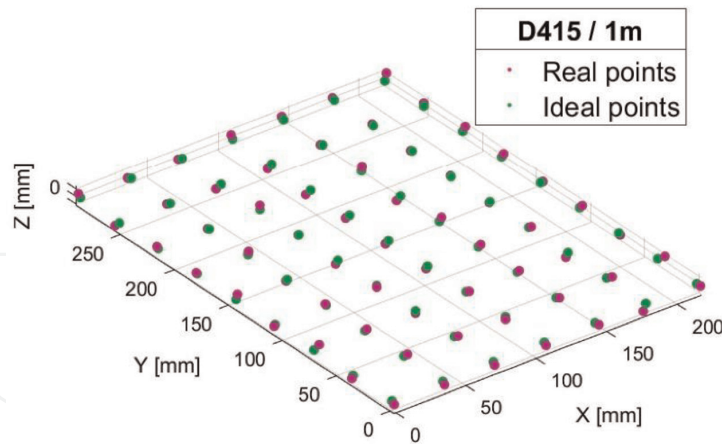
Figure 2.
Real point cloud captured by RealSense D415 sensor compared with ideal one.

The corners of a chessboard in RGB image were detected using the OpenCV algorithm. Based on equations of the pinhole camera model for projection from image coordinate system to world coordinate system (X, Y, Z) we obtain the real point cloud. The Z-coordinate is obtained from the depth map at pixel position (u, v). Following the Eq. (1) and (2), it is needed to know the intrinsic parameters Cx, Cy, fx, and fy of cameras for getting world coordinates X and Y:

$$X = \frac{u - C_x}{f_x} Z, \tag{1}$$

$$Y = \frac{v - C_y}{f_y} Z, \tag{2}$$

Captured (real) point cloud is fitted to the ideal one using translation and rotation estimation. As the global registration technique, Coherent Point Drift was used and final precise registration was Iterative Closest Point (ICP). The Root Mean Square Error (RMSE) of Euclidean distance was used as a relevant metric for sensor accuracy assessment:

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (p_i - p_i')^2}, \tag{3}$$

where the p_i and p_i', are the sets of coordinates of the real and ideal points, respectively.

Since we are not able to construct the precise 3D object with chessboard patterns and its ideal software model, we decide to use a flat surface. To simulate 3D scene, we captured the flat chessboard from 3 different views (**Figure 3**). The resulting error is represented as the mean value of errors obtained from views A, B, and C.

4.3 Ideal plane fitting

As described in the study [23], another way of depth error estimation is fitting the captured real point cloud to an ideal surface. We used the plane without chessboard captured from different positions similarly in previous method. The mean Euclidean
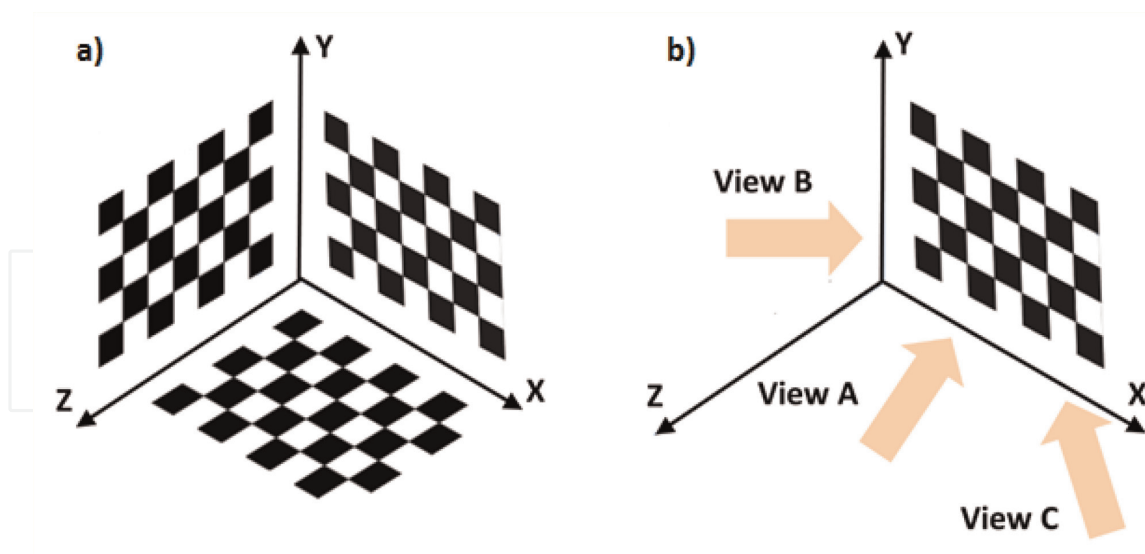
Figure 3.
The possible approach on how to capture the test chessboard pattern in 3D: (a) precise 3D construction used in studies [29, 30] (b) our approach: Capturing test chessboard from several views.
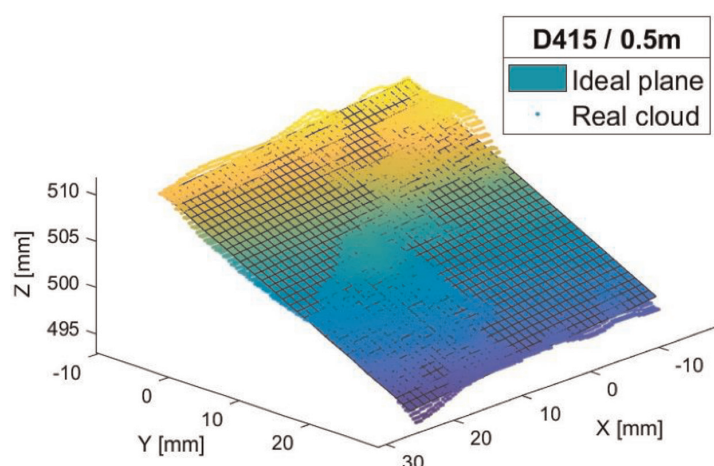

Figure 4.
Real point cloud captured by RealSense D415 and fitted to ideal plane.

distance between the ideal plane and real point cloud represent the estimated error. The fitting of real and ideal point clouds is shown in **Figure 4** [31].

4.4 Comparison of selected RGB-D sensors

In accuracy measurement of sensors described above, we used 4 sensors of different principles. We also assessed the suitability of these sensors in multi-sensor parallel configurations. The ToF sensor in KinectV2 and the structured light sensor in RealSense SR300 use infrared light, so their usage in parallel multi-view system is complicated due to mutual interference. The good candidates for desired imaging system are ZED MINI and RealSense D415. These sensors represent passive and active stereo pairs technology, respectively. The main depth sensor parameters of each camera are summarized in **Table 1** and available on product websites [32, 33] and comparison table in [23].

Sensor	Technology	DFOV*	Max. resolution	FR* [fps]	Range [m]
RealSense D415	Active stereo	72	1280 × 720	90	0.3–10
Kinect V2	ToF	70 × 60	512 × 424	30	0.5–4.5
ZED MINI	Stereo	110	4416 × 1242	100	0.15–12
RealSense SR300	Structured light	90	640 × 480	60	0.2–1.5

*Maximal FR value might depend on resolution used/ *DFOV − Diagonal Field of View.*

Table 1.
Depth sensors parameters comparison.

Sensor	σ for distance [mm]		
	500	**700**	**1000**
RealSense D415	0.307	0.639	1.303
ZED MINI	1.499	1.343	2.180
Kinect V2	1.151	1.267	1.375
RealSense SR300	0.124	0.253	0.716

Table 2.
Standard deviation of depth error for different distances.

Table 1 brings the comparison of key parameters of each sensor, such as resolution, diagonal field of view (DFOV), frame rate, and range for an optimal distance of sensor from object.

4.4.1 Comparison based on noise measurement

The noise measurement methodology is described above and results are taken from our study [31]. The standard deviation of depth error for different distances represents amount of sensor noise. The comparison of sensors is in **Table 2**.

Figure 5 shows the statistical comparison for a distance of 0.5 m. The amount of noise increases with the distance between the sensor and the surface. In this comparison, the offset of sensor is ignored, and also variable part of signal is taken into the account.

As seen in **Table 2**, SLS technology (RealSense RS 300) achieved the best results. As expected, there is an evident difference between ZED MINI and RealSense D415. As expected the accuracy of active stereo sensor is more accurate than passive sensor.

4.4.2 Comparison based on ideal cloud fitting

In this experiment, the same sensor parameters were set, as in the previous measurement. In the case of ZED MINI and RealSense D415, the chessboard pattern was captured from distances of 0.5 m and 1 m. The results including **Table 3** are obtained from our study [31].

In this experiment, the results only for two sensors were shown, because of several negative facts. Because the RGB resolutions of ZED MINI and RealSense D415 sensors
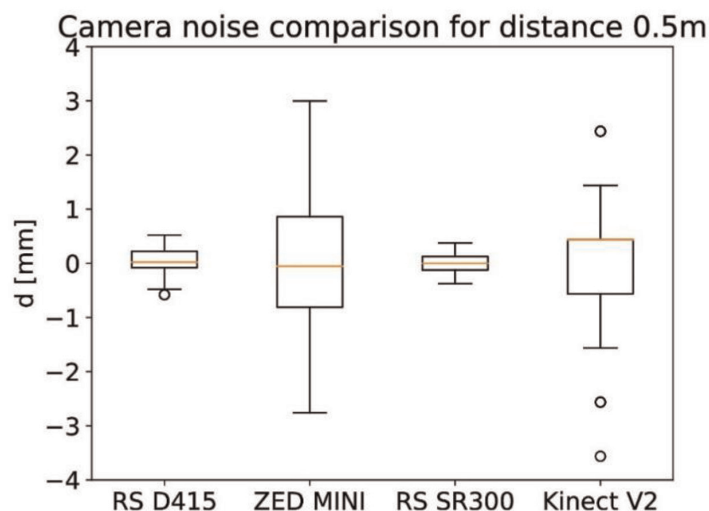
Figure 5.
Noise of depth sensors comparison for distance 0.5 m.

Sensor	RMS error for distance [mm]	
	500	**1000**
RealSense D415	1.382	3.172
ZED MINI	1.803	4.582

Table 3.
RMS error for multiple distances and positions using ideal cloud fitting.

are the same, we expect the same corner detection error. For this reason, the comparison of passive and active stereo sensors in this way we consider as precise. Some problems occurred when capturing the chessboard pattern with SR300 and KinectV2 sensors. The sensor SR300 produces the depth map that contains "empty areas" of unknown depth demonstrated as black holes in depth images. Due to this fact it is not able to compute the depth of the chessboard corner point. We can say that the real point of cloud construction is impossible. Also, while using KinectV2 in a distance of more than 0.7 m, the depth map contains the black regions of unknown depth. The depth deviation is caused by different object surface reflections. Such a phenomenon, associated with ToF camera calibration is described in the study [34]. To avoid this, the color version of chessboard instead of black and white can be used. Also, the precisely constructed cube covered by the chessboard pattern is a potential way, as described in [30]. The small resolution of Kinect V2 RGB image does not allow to detect chessboard corners correctly. Due to this fact, the comparison of all 4 technologies in this way we consider inadequate.

4.4.3 Comparison based on ideal plane fitting

Table 4 taken from Ref [31] shows the comparison of all tested sensor technologies.

The estimated depth error is independent of the corners detection error, so the corresponding results in **Tables 3** and **4** are different. The difference in

Sensor	RMS error for distance [mm]	
	500	700
RealSense D415	0.646	1.026
ZED MINI	0.782	1.052
Kinect V2	1.515	1.588
RealSense SR300	0.321	0.918

Table 4.
RMS error for different distances and positions using ideal plane fitting.

RMSE between structured light and active stereo pair for a distance of 0.5 m is only 0.3 mm.

5. Development of multi-sensor system

Based on previous measurements we decided to use active stereo sensor RealSense D415 as a key element of our imaging system. The scanning accuracy is comparable to other sensor technologies and it is absolutely sufficient for given medical use. On the other hand, active stereo pair does not suffer mutual interference in parallel mode of scanning. From the previous results, we can determine the optimal distance of head from the sensors and this distance is approximately 0.5 m. This distance is respected in a physical model of the sensor stand. For mounting 3 sensors in the fixed position, we use the constructed stand, shown in **Figure 6a**. The spatial configuration schema is shown in **Figure 6b**.

The distance d is set approximately to 0.5 m. In our application, the frontal side of the object (the face) is the most important for scanning, so the layout needs to be set to obtain a high fill rate in this area. For future automated processing of captured 3D models and their normalization, the patient must sit in a normalized position. To avoid covering the important parts of the head we had to replace physical head fixation by software assistance tool mentioned later. This assistance algorithm places the head to the center of frontal sensor because the features obtained from this view (facial view) are the most important.
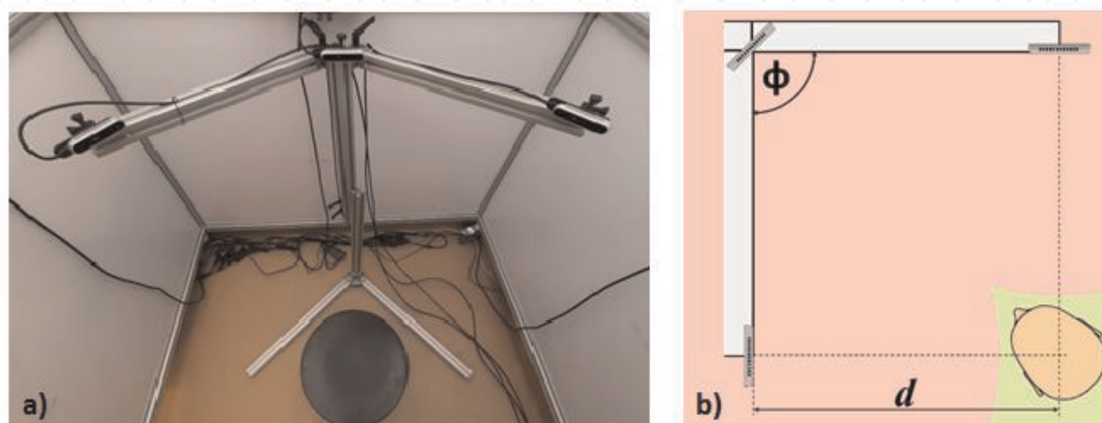

Figure 6.
(a) Adjustable sensor stand. (b) Layout of sensor positions (top view).

5.1 Data capturing

All the sensors are connected via the USB 3 interface to the acquisition computer, which allows capturing the color and depth frames with a 30 fps frame rate at Full HD spatial resolution. The system for the capturing and data processing is designed as GUI running as a web application. The server application is created using Python. This application acquires the depth and RGB color image frames from all sensors and streams the image data to user interface. The RealSense SDK tool is used for controlling the sensors, flask for server operation, and OpenCV framework for image processing. Image acquisition and streaming of the image data run in separated threads. The data flow is reduced due to JPEG encoding of images (depth and also RGB). When the server application receives a request for saving from the user interface, the saving procedure is triggered. Acquired images are stored locally in a temporary folder. When image acquisition process is finalized, the content of the temporary folder is zipped, encrypted, and named by patient identifier and actual time. The script then copies the ZIP to external server storage to collect and backup the data. When the external server is not accessible, the file is queued and sent in the next time [35].

5.2 Graphical user interface of system

The graphical user interface enables to control the scanning of the patient's head and neck and also to annotate the captured data. For annotating, the digital version of standardized sleep questionnaire is part of application and it is described later. Web-based design of this interface enables to use of any portable device in the network to provide scanning and annotating the data. Also, the interface can be accessed locally on the PC where the server runs. The main window of the interface is shown in **Figure** 7. Menu on the left side contains several settings:

- # of frames – the number of color and depth frames saved by one shot,
- ID – personal identifier of patient,
- Expression – the facial expression of the patient.

Number of frames taken by "one shot" helps to provide temporal filtering of data. Considering our study [25], resulting depth image from given view is a product of averaging the stack of images in buffer. This averaging helps to eliminate noise artifacts in 3D reconstruction.
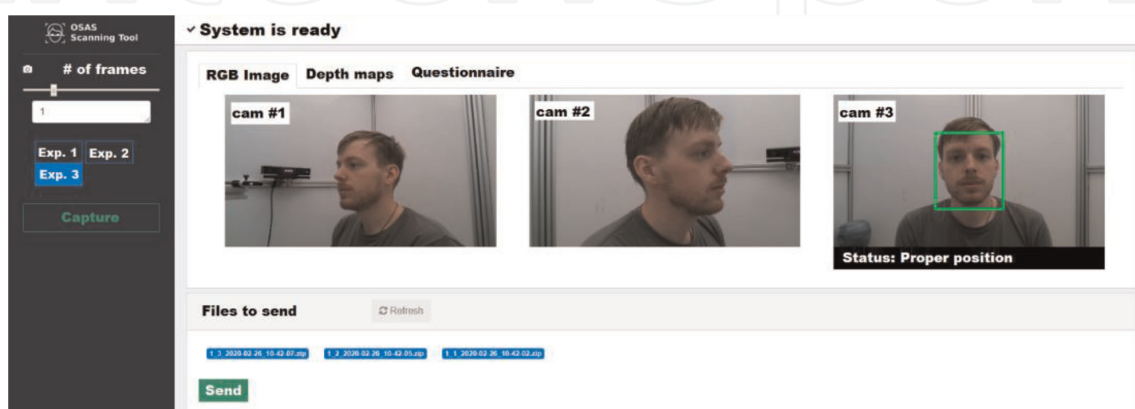

Figure 7.
Graphical user interface – Main window.

Facial expression is functionality prepared for further research. It could be interesting to correlate the OSAS detection based on normalized face expression (e.g., smile, neutral expression).

After capturing request, the data are zipped and sent to the acquisition server. If any error, the data are saved locally and resent to the acquisition server in next capturing request. If the webserver is unavailable, the warning message is displayed. The data saved on server can be identified by personal ID of patient. The user can switch between color and depth view.

5.3 Normalized head position assistance algorithm

For obtaining the most precise, accurate, and normalized 3D scans, we need to set the patient's head to a defined position (as equal as possible in all patients). This task may be very difficult and stressful for young patients. Based on this fact, we created the head position assistance tool based on the algorithm of eyes and head detection. Detected eye position is used to compute the difference from ideal eye position. The depth map is also available so, the algorithm can get the difference in eye positions in Z-axis. This information is obtained only from the central sensor images. In **Figure 8**, we can see the angle offsets of detected eyes from ideal position. The angles α and β are used for determining how much is the head rotated or tilted. The limits were chosen empirically and they can be improved during further research. Based on the depth map, we are also able to compute the distance of the head from sensors (if it is too close or too far from sensor). In other words, head positioning assistance tool helps to keep the head in red highlighted area according to **Figure 6b** [35].

When the head position is inside the optimal range, imaging can be provided. If the position of head is outside the tolerance, the moving, rotation, and tilting commands are shown for the user on the screen to adjust the head position. For detection of face in actual sensor view, the Viola-Jones algorithm is used. This algorithm is a frequently used tool for given object detection. The original algorithm is used to detect and classify the objects into several classes. In our case, it was trained for human faces. In comparison with other algorithms, the training time is relatively high, but detection is very fast. The algorithm uses Haar basis feature filters and it does not use multiplications [36]. The computation time is minimized by placing the classifiers with fewest
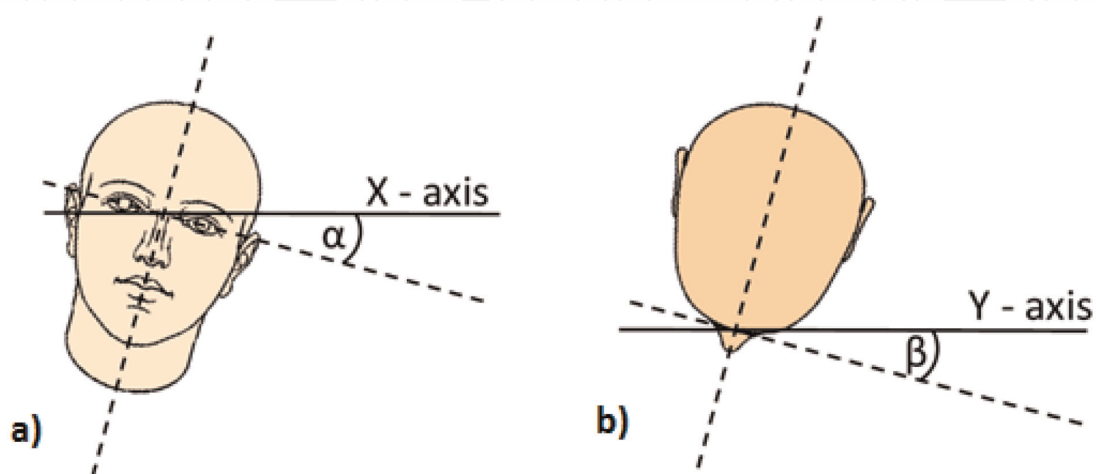

Figure 8.
Difference of eye positions: (a) X-axis offset. (b) Y-axis offset.

features at the beginning of cascade. The features are most commonly trained using Ada-Boost algorithm. This method selects only those features that improve the detection accuracy and potentially decrease the execution time.

5.4 The annotation questionnaire

In addition to mentioned functions, the application includes an online questionnaire, which is the digital format of the EU questionnaire [5]. Besides the 3D imaging of patient heads, the specialist (user) is able to insert additional information about the patient (age, weight, subjective rating of intraoral anatomy …) [35]. This additive information can be used to extend features for machine learning methods and automated diagnostics based on artificial intelligence. Implemented electronic questionnaire is shown in **Figure 9**.

5.5 Experimental results

After pilot testing, the system was placed in an experimental workplace inside the sleep laboratory of Clinic of Children and Adolescents in University Hospital Martin, Slovakia.

Application is designed in a simple layout and also assists medical staff (as we can see as an example in **Figure 10**) to obtain the best results without any sophisticated manipulation.

Figure 9.
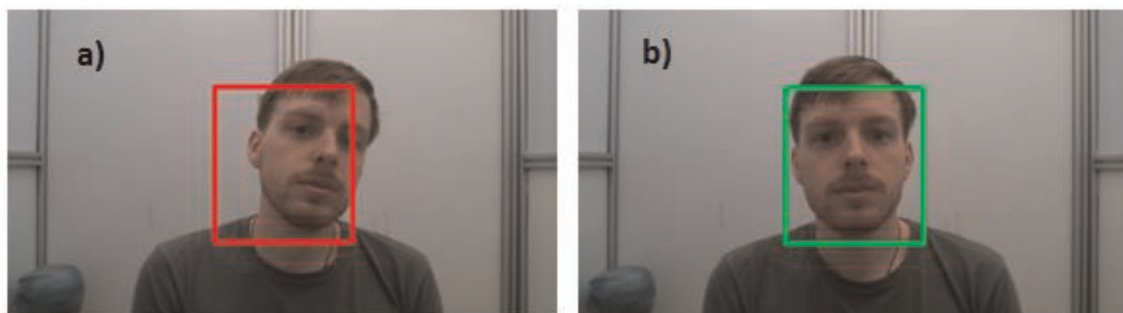*Graphical user interface — online EU questionnaire (excerpt).*

Figure 10.
Head positioning assistance: (a) incorrect head position; the command is tilt head right. (b) Correct position of head.
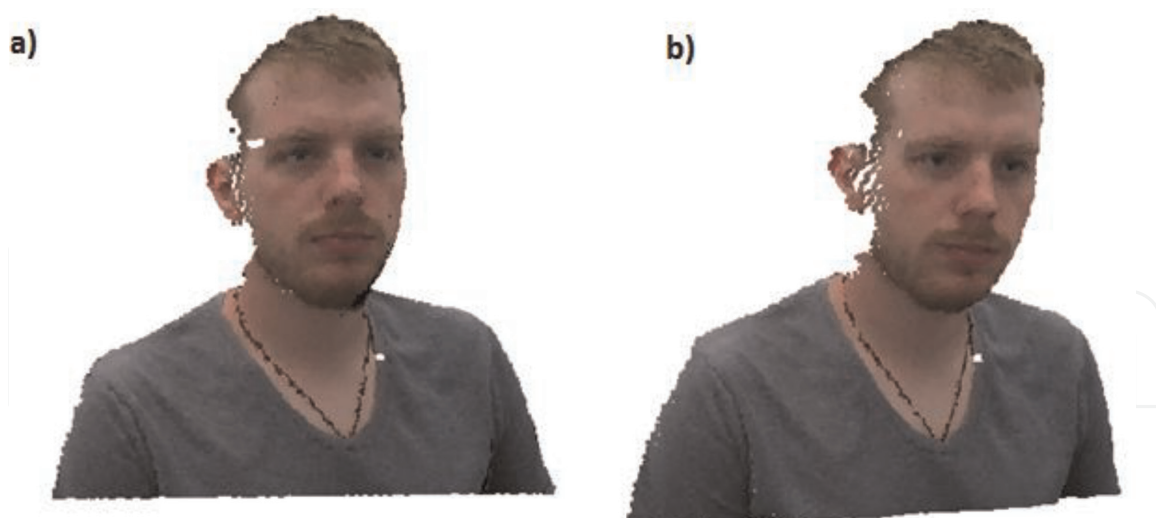
Figure 11.
The resulting 3D model of patient: (a) frontal view. (b) Rotated view.

After capturing the stack of images from single sensors, the 3D reconstruction of data can be provided. Using intrinsic camera parameters, we can compute the colored point cloud from depth and color frame. The point clouds of individual views must be also denoised and transformed into a common coordinate system. As a denoising method, the statistical outlier filter was used. The resulting 3D model with the possibility of rotation is shown in **Figure 11**.

The partial point clouds from single sensors are registered by the global registration RANSAC method and local refinement is done by ICP (Iterative Closest Point) algorithm. For further scientific research, it is very interesting to implement and compare different filtration algorithms for depth maps or point clouds, registration methods, and calibration algorithms that can improve the accuracy of models. Nowadays, it exists a lot of machine learning methods that can obtain the relevant features from the head (from depth maps, RGB images, or directly from 3D models) and will finalize the feature vector for automated diagnostics.

6. Conclusion

Our study focuses on the development of the multi-sensor scanning system, that aims to be the future pre-diagnostic tool for obstructive sleep apnea diagnostics. Because the obstructive sleep apnea syndrome can correspond with some abnormalities in cranio-facial parameters on the head, 3D scanning of head can be a promising procedure to obtain an automated method for OSAS screening. A system for early screening can easily prioritize the patients for complex diagnostics and then for early therapy. Especially in Slovakia, the waiting periods for conventional OSAS diagnostic can be several months.

RGB-D sensors are relatively non-expensive sensors with increasing popularity used in many fields: from entertainment to mechanical engineering or medical applications. To complete the 3D scanning system for biomedical use, the main research was focused on the selection of suitable RGB-D sensor for obtaining the accurate model of the head and neck. This model can be used for noninvasive automated procedures. After selecting the representative for all technologies of RGB-D sensors

we used some metrics, which can compare their accuracy. The noise measurement, the ideal point cloud fitting, and ideal plane fitting were selected for this assessment.

After the series of experiments, we can say that the difference in accuracy between the all sensors is not so significant and all of them could be used for our implementation. On the other hand, considering the second condition – multi-sensor parallel system – the mutual interference of sensors must be taken into the account. Because ToF sensors and also SLS sensors interfere and can generate interference artifacts, we focused on stereo pair technology of RGB-D sensors. Finally, we selected the active stereo pair Intel RealSense D415. Based on depth error, the optimal distance of the sensor from object is set to 0.5 m. The system with 3 sensors respects this distance.

The scanning system is driven by a web-based application with simple graphical user interface. 3D scans can be extended by information from a digitized EU sleep questionnaire. The database of 3D models with information form questionnaire is strictly needed to build an automated diagnostic system based on machine learning or artificial intelligence. These methods are now state of the art in many imaging and signal processing tasks. System is now implemented in clinical environment to obtain first elements of the dataset.

Further research can be oriented for selecting and implementing the filtration methods for obtained data, registration methods for partial models from single sensors, and calibration algorithms for the case of changes in sensor layout.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Author details

Libor Hargaš* and Dušan Koniar
Faculty of Electrical Engineering and Information Technology, Department of
Mechatronics and Electronics, University of Zilina, Zilina, Slovakia

*Address all correspondence to: libor.hargas@feit.uniza.sk

IntechOpen

References

[1] Ceska R. Interna. Triton; 2015. ISBN: 978-80-7387-885-6

[2] Villa MP, Pietropaoli N, Supino MC, Vitelli O, Rabasco J, Evangelisti M, et al. Diagnosis of pediatric obstructive sleep apnea syndrome in settings with limited resources. JAMA Otolaryngology–Head & Neck. 2015;**141**:990. DOI: 10.1001/jamaoto.2015.2354

[3] Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin questionnaire to identify patients at risk for the sleep apnea syndrome. Annals of Internal Medicine. 1999;**131**:485. DOI: 10.7326/0003-4819-131-7-199910050-00002

[4] Johns MW. A new method for measuring daytime sleepiness: The Epworth sleepiness scale. Sleep. 1991;**14**:540-545. DOI: 10.1093/sleep/14.6.540

[5] Feketeová E, Mucska I, Klobučníková K, Grešová S, Stimmelová J, Paraničová I, et al. EU questionnaire to screen for obstructive sleep Apnoea validated in Slovakia. Central European Journal of Public Health. 2018;**26**:S32-S36. DOI: 10.21101/cejph.a5278

[6] Myers KA, Mrkobrada M, Simel DL. Does this patient have obstructive sleep apnea?: The rational clinical Examina-tion systematic review. JAMA. 2013;**310**:731. DOI: 10.1001/jama.2013.276185

[7] Miles PG, Vig PS, Weyant RJ, Forrest TD, Rockette HE. Craniofacial structure and obstructive sleep apnea syndrome — A qualitative analysis and meta-analysis of the literature. American Journal of Orthodontics and Dentofacial Orthopedics. 1996;**109**:163-172. DOI: 10.1016/S0889-5406(96)70177-4

[8] Capistrano A, Cordeiro A, Capelozza Filho L, Almeida VC, de Silva PIC, Martinez S, de Almeida- Pedrin RR. Facial morphology and obstructive sleep apnea. Dentofacial Press Journal of Orthodontics. 2015;**20**:60–67. DOI: 10.1590/2177-6709.20.6.060-067.oar

[9] Hoekema A, Hovinga B, Stegenga B, De Bont LGM. Craniofacial morphology and obstructive sleep Apnoea: A Ceph-alometric analysis. Journal of Oral Rehabilitation. 2003;**30**:690-696. DOI: 10.1046/j.1365-2842.2003.01130.x

[10] de Mello Junior CF, Guimarães Filho HA, de Gomes CAB, de Paiva CCA. Radiological findings in patients with obstructive sleep apnea. Jornal Brasileiro de Pneumologia Publicacao Officials Society Brasilian Pneumology E Tisilogia. 2013;**39**:98–101. DOI:10.1590/s1806-37132013000100014

[11] Butorova E, Elfimova E, Shariya M, Litvin A. MRI measurement of airway soft tissues parameters in patients with obstructive sleep Apnoe. Journal of Hypertension. 2016;**34**:e331. DOI: 10.1097/01.hjh.0000492316.06821.77

[12] Chousangsuntorn K, Bhongmakapat T, Apirakkittikul N, Sungkarat W, Supakul N, Laothamatas J. Computed to-mography characterization and comparison with polysomnography for obstructive sleep apnea evaluation. Journal of Oral and Maxillofacial Surgery. 2018;**76**:854-872. DOI: 10.1016/j.joms.2017.09.006

[13] Barkdull GC, Kohl CA, Patel M, Davidson TM. Computed tomography imaging of patients with obstructive sleep apnea. The Laryngoscope. 2008;

118:1486-1492. DOI: 10.1097/ MLG.0b013e3181782706

[14] Islam SMS, Mahmood H, Al-Jumaily AA, Claxton S. Deep learning of facial depth maps for obstructive sleep apnea prediction. In: Proceedings of the 2018 International Conference on Machine Learning and Data Engineering, iCMLDE, 3-7 December 2018. Sydney, NSW, Australia: IEEE; 2018. DOI: 10.1109/iCMLDE.2018.00036

[15] Sutherland K, Lee RWW, Petocz P, Chan TO, Ng S, Hui DS, et al. Craniofacial phenotyping for prediction of obstructive sleep Apnoea in a Chinese population. Respirology Carlton Vic. 2016;**21**:1118-1125. DOI: 10.1111/ resp.12792

[16] de Chazal P, Tabatabaei Balaei A, Nosrati H. Screening patients for risk of sleep apnea using facial photographs. In: Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Seogwipo: IEEE; 2017. pp. 2006-2009

[17] Eastwood P, Gilani SZ, McArdle N, Hillman D, Walsh J, Maddison K, et al. Predicting sleep apnea from three-dimensional face photography. Journal of Clinical Sleep Medicine. 2020;**16**:493-502. DOI: 10.5664/ jcsm.8246

[18] Ozdemir ST, Ercan I, Can FE, Ocakoglu G, Cetinoglu ED, Ursavas A. Three-dimensional analysis of craniofacial shape in obstructive sleep apnea syndrome using geometric Morphometrics. International Journal of Morphology. 2019;**37**:338-343. DOI: 10.4067/S0717-95022019000 100338

[19] Lin T, Liu X. An intelligent recognition system for insulator string defects based on dimension correction and Opti-mized faster R-CNN. Electrical Engineering. 2020:1-9. DOI: 10.1007/ s00202-020-01099-z

[20] Uribe FA, Flores J. Parameter estimation of arbitrary-shape electrical cables through an image processing technique. Electrical Engineering. 2018; **100**:1749-1759. DOI: 10.1007/ s00202-017-0651-y

[21] Yan Z, Shi B, Sun L, Xiao J. Surface defect detection of aluminum alloy welds with 3D depth image and 2D gray image. International Journal of Advanced Manufacturing Technology. 2020;**110**:741-752. DOI: 10.1007/ s00170-020-05882-x

[22] Bulczak D, Lambers M, Kolb A. Quantified, interactive simulation of AMCW ToF camera including multipath effects. Sensors. 2018;**18**:13. DOI: 10.3390/s18010013

[23] Giancola S, Valenti M, Sala R. A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies. Switzerland AG: Springer Nature; 2018. DOI: 10.1007/978-3-319-91761-0. ISBN 978-3-319-91760-3

[24] Grunnet-Jepsen A, Winer P, Takagi A, Sweetser J, Zhao K, Khuong T, et al. Using the Real Sense D4xx Depth Sensors in Multi-Camera Configurations. Available from: https://simplecore.intel. com/realsensehub/wp-content/uploads/ sites/63/Multiple_Camera_WhitePaper04. pdf [Accessed: August 28, 2022]

[25] Volak J, Koniar D, Jabloncik F, Hargas L, Janisova S. Interference artifacts suppression in systems with multiple depth cameras. In: Proceedings of the 2019 42nd International Conference on Telecommunications and Signal Processing, TSP, 01-03 July 2019.

Budapest, Hungary: IEEE; 2019. DOI: 10.1109/TSP.2019.8768877

[26] Langmann B, Hartmann K, Loffeld O. Depth camera technology comparison and performance evaluation. In: Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods. 2012. DOI: 10.5220/0003778304380444

[27] Vit A, Shani G. Comparing RGB-D sensors for close range outdoor agricultural phenotyping. Sensors. 2018; **18**:4413. DOI: 10.3390/s18124413

[28] Chiu C-Y, Thelwell M, Senior T, Choppin S, Hart J, Wheat J. Comparison of depth cameras for three-dimensional reconstruction in medicine. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine. 2019;**233**:938-947. DOI: 10.1177/0954411919859922

[29] Ortiz LE, Cabrera VE, Goncalves LMG. Depth data error modeling of the ZED 3D vision sensor from Stereolabs. ELCVIA: Electronic Letters on Computer Vision and Image. 2018;**17**:1-15. DOI: 10.5565/rev/elcvia.1084

[30] Fernandez L, Avila V, Goncalves L. A generic approach for error estimation of depth data from (stereo and RGB-D) 3D Sensors. Preprints. 2017. DOI: 10.20944/preprints201705.0170.v1

[31] Bajzik J, Koniar D, Hargas L, Volak J, Janisova S. Depth sensor selection for specific application. In: Proceedings of the 2020 ELEKTRO, 25-28 May 2020. Taormina, Italy: IEEE; 2020. DOI: 10.1109/ELEKTRO49696.2020.9130293

[32] Intel® RealSenseTM Computer Vision - Depth and Tracking Cameras Available online: https://www.intelrealsense.com/ [Accessed: January 19, 2021]

[33] Stereolabs - Capture the World in 3D Available online: https://www.stereolabs.com/ [Accessed: January 19, 2021]

[34] Lindner M, Kolb A. Calibration of the intensity-related distance error of the PMD TOF-camera. In: Proceedings Volume 6764, Intelligent Robots and Computer Vision XXV: Algorithms, Techniques, and Active Vision, Event: Optics East. Boston, MA, United States. 2007. DOI: 10.1117/12.752808

[35] Stefunova S, Koniar D, Hargas L, Bulava J. Multi-camera scanning system for collecting and annotating 3D models of the head and neck. In: Proceedings of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME, 07-08 October 2021, Mauritius. 2021. DOI: 10.1109/55ICECCME52200.2021.9590878

[36] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, 08-14 December 2001. Kauai, HI, USA: IEEE; 2001. DOI: 10.1109/CVPR.2001.990517