An Unobtrusive and Non-Contact Method for Respiratory Measurement With Respiratory Region Detecting Algorithm Based on Depth Images

CHENGLU SUN[ID][1]**, WEI LI**[1]**, CHEN CHEN**[ID][1]**, ZEYU WANG**[1]**, AND WEI CHEN**[ID][1,2,3]**, (Senior Member, IEEE)**

[1]Center for Intelligent Medical Electronics, Department of Electronic Engineering, School of Information Science and Technology, Fudan University, Shanghai 200433, China
[2]Shanghai Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention, Shanghai 200000, China
[3]Human Phenome Institute, Fudan University, Shanghai 200000, China

Corresponding author: Wei Chen (w_chen@fudan.edu.cn)

ABSTRACT In order to obtain the respiratory condition unobtrusively and comfortably, a non-contact method based on the commercial depth camera Realsense SR300 was proposed to extract respiratory information from depth data. In this paper, a respiratory region detecting algorithm which is mainly based on the morphological method was proposed to obtain the region of interest (ROI) with the depth images. The proposed algorithm contains four steps: body edge extraction, noise reduction, ''image skeleton'' extraction, and respiratory region estimation. As a result, the respiratory waveform can be derived from the depth data in the ROI. For validation, experiments were carried out to verify the feasibility of obtaining the respiratory information with this approach. In consideration of different application scenarios, 20 kinds of conditions were designed and applied for the experiments. The respiratory rate extracted from the depth waveform can be calculated, and the accuracy achieved was 95.20% for all data while utilizing polysomnography thorax effort signal as gold standard. Through the Bland–Altman analysis, it represented that the proposed system had a good agreement ($r^2 = 0.88$) with the gold standard. In addition, the performances of the system in the 20 different conditions were analyzed by statistics, and the results showed that the system has good adaptability and robustness for different conditions. In conclusion, the proposed algorithm can fit different scenarios, and this paper provides a novel option for extracting the physiological information with depth data.

INDEX TERMS Unobtrusive monitoring, respiratory rate, depth imaging, Realsense SR300, respiratory region detecting algorithm.

I. INTRODUCTION

Respiratory activity is one of the most important physiological signals of human. In general, a healthy adult has 12-18 breaths in one minute and the infants can have more rapid and short respiration which has more than 40 times in one minute [1], [2]. In modern society, respiratory diseases which require more attention and long-term care have become the high incidence diseases due to the development of the industrialization process, such as the sleep apnea-hypopnea syndrome (SAHS) [3], chronic obstructive pulmonary diseases (COPD) [4], apnea and sudden infant death syndrome (SIDS) [5], [6]. If those respiratory diseases occurred during the sleep, it is more dangerous. Therefore, the measurements of respiration and relevant parameters such as respiratory rate, respiratory pattern and pulmonary functions are important, especially during the sleep. Until now, many respiratory measurement methods have been developed, such as spirometry, respiratory inductance plethysmography (RIP) [7], thoracic

impedance [8], impedance pneumography (IP) [9], photo-plethysmography (PPG) [10] and strain gauges [11]. During the sleep, subjects' respiratory condition can be monitored by the air flow and respiratory effort band which are two parts of the polysomnography (PSG) device [12] in specialized sleep laboratories. The PSG plays an important role in the SAHS, COPD, asthma and snore diagnosis. However, these techniques are inconvenient to hold or wear the devices and need to contact to subject while measuring which might interfere the subject's normal activity [13]. Furthermore, the airflow method which was considered as the most precise way in clinical is uncomfortable to patients [14]. Respiratory measurement for the infants in neonatal intensive care unit (NICU) is more harmful because the delicate skin and little body can not suffer the attached devices for a long time [15]. Hence in the current practice many health care staffs in NICU will observe the chest and abdomen fluctuations to count the respiratory rates of the infants, thus the accuracy and time efficiency will not be guaranteed.

Therefore, many researchers proposed unobtrusive techniques for measuring the respiratory conditions, such as the microwave-based method [16], [17], thermal imaging method [18], [19] and pressure detecting method [20]–[22]. These methods do not cause the discomfort and inconvenience to the user, but still have some shortcomings. For instance, microwave-based methods and thermal imaging methods which used the commercial products require relative large and specific instruments which need complicated setup procedure to measure the relevant signals. The pressure detecting method requires the subject to lie or sit at specific location to obtain the pressure signals caused by respiratory movements.

With the development of a variety of cameras and the Internet of Things (IoT), researchers began to use the image data to detect respiratory conditions, such as RGB cameras based method [23]–[25], RGB camera combined with depth sensors method [26], [27] and complete depth information method [28]–[34]. The camera based method is a novel way to measure the respiratory activity. Many reported studies extracted the respiratory information via the RGB images. Another part of those studies measured the respiratory condition by the RGB images combined with depth information, for example, built body model to tracking the respiration. However, the RGB video stream and the body model would invade individual privacy. In some relatively private application scenarios, the subjects do not want the device to collect their "real information" (the RGB video stream or the body model). The complete depth information method is a good choice for this case, but parts of studies only utilized complete depth information to conduct normal respiratory monitoring [28], [29]. The state of the art complete depth-based researches performed the stationary clinical respiratory assessment and pulmonary function testing [30]–[34]. However, those studies only performed the basic exploration without conducting the tests for a variety of environments

or conditions which are important in many application scenarios.

There are many advantages by using only the depth data to obtain the respiratory activity such as unobtrusiveness, high adaptability and sensitivity. In addition, the depth data does not invade the individual privacy because the depth information can only reflects the distance between the scene and the camera, unlike the RGB image which contains too much privacy information. Therefore, the research about the depth camera deserves further exploration.

This study aims to propose an unobtrusive and non-contact respiratory measurement based on the depth data only, which is simple and suitable for a variety of situations, such as the sleep monitoring and home health care. The main contributions of this study can be described as follows:

1) The designed system only collects the depth data, which can protect subjects' privacy. The equipment employed by this system is compact and small which is easy to set up, and the system is very cheap.

2) Based on the proposed respiratory region detecting algorithm, the system can capture the respiratory region and obtain the respiratory waveform for analysis only by the depth data. The method is easy to realize and implement, and does not need to collect a large amount of images for learning features.

3) This study provides complete experimental testing in different conditions to verify the performance and robustness of the system, including different measuring ranges, different clothes worn by the subjects and different postures acted by the subjects to verify the performance and stability of the system. The experimental results were analyzed in terms of accuracy, consistency and statistical differences, which showed that the system has promising performance.

This paper mainly describes an unobtrusive and non-contact method for detecting the respiratory activity with the depth camera by utilizing depth image processing technique, and experiments for verifying the performance on respiration monitoring. The context is organized as follows. In Section II, the principle of acquiring the depth information and the characteristic of the employed commercial depth camera Realsense SR300 are introduced. In addition, the algorithm focused on detecting the respiratory activity region and extracting the respiration condition from the depth data in the ROI is presented. In Section III, the experimental procedure is described by considering different experimental conditions, four external environments and five subject postures which combined 20 kinds in total. We calculated the accuracy of respiratory rate (RR) through depth waveform during a period of time under those conditions. In addition, we analyzed the consistency of results and statistical difference between different experimental conditions. In Section IV, the discussion with regard to the experimental results and the future work which might be the feasible research direction about this

study are put forward. In the last section, we summarize this paper.

II. MATERIALS

This section proposed a technical solution based on the Realsense SR300 depth camera to acquire the depth images and the process of generating the depth waveform for extracting respiratory condition after morphology operation.

A. PRINCIPLE OF ACQUIRING THE DEPTH DATA BASED ON STRUCTURED LIGHT

The applications based on depth information can be used in various fields, such as vision measurement, remote sensing and control, video or computer games, model reconstruction, and industrial automation [35]–[38]. There are many extensive and in-depth studies in capturing the depth information from scenes. Structured light method is one classical depth measurement technique which gained increasing attention due to its high speed and accuracy. This method can be demonstrated in Fig.1.
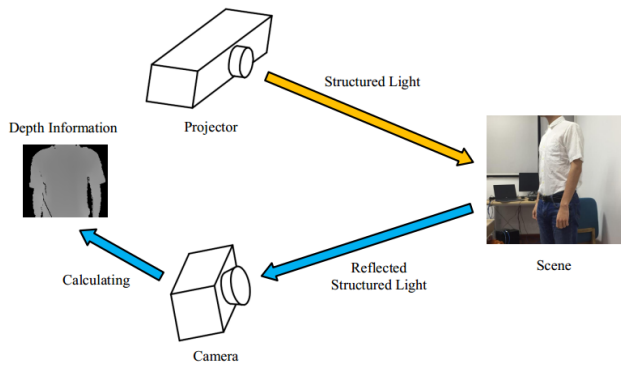

FIGURE 1. The process of obtaining the depth information based on the structured light method. The structured light method has a projector which can emit specific structured light to scenes or objects. Then the light is reflected by the objects, and the patterns in the light modulated by objects would be captured by a camera. Finally, the depth information can be calculated based on the reflected patterns and some prior information including space relationship of these devices.

When the reflected pattern is captured by the camera, it will be compared against a reference pattern. The reference pattern is obtained from a reference plane which is at a known distance from the projector. If the object whose distance to the projector is different with the reference plane, the corresponding position of its reflected pattern in imaging space would move. The move of the corresponding position in the pattern can be obtained by decoding the patterns in the imaging space [39]. Then the depth information of the object can be calculated based on the triangulation [40]. After obtaining the depth information of each point in the imaging space, one frame depth image can be obtained.

B. THE REALSENSE SR300 DEPTH CAMERA

For various applications, depth cameras with different specifications are produced, such as the Kinect [41],

TABLE 1. Basic parameters of the Kinect, Kinect V2, PrimeSensor and RealSense SR300.

Specifications	Kinect	Kinect V2	Prime Sensor	SR300
Depth resolution	320×240	512×424	640×480	640×480
Depth range (m)	1.2 - 3.5	0.5 - 4.5	0.8 -3.5	0.2 - 1.5
Depth frame rate	30 Hz	30 Hz	30 Hz	60 Hz

Kinect V2 [42], [43], PrimeSensor [44] and RealSense camera SR300 [45], [46]. The basic parameters of those depth cameras are summarized in Table 1. From the Table 1, we can see that the SR300 is better than Kinect, Kinect V2 and PrimeSensor in the resolution and the frame rate. As for the effective depth range, the PrimeSensor seems have greater depth range, however, the minimum effective range is 0.8 m, which exceeds the practical range of possible scenarios such as home care. Hence, the SR300 depth camera is more suitable for this study. The SR300 can generate the depth images without the visible lights via 60 frames per second (FPS), and the resolution of depth images can achieve 640×480. This frame rate means that the depth image can capture the depth information within 60 Hz which is enough for the respiratory activities, the effective range of the SR300 is optimized from 0.2 to 1.5 m for using indoors [47]. The SR300 can use USB 3.0 port to transmit the depth data which ensures the stability of the system.
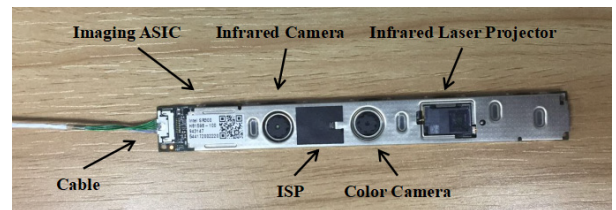

FIGURE 2. The SR300 camera implements an infrared laser projector, a fast VGA infrared camera, a color camera with integrated image signal processor (ISP) and imaging application specific integrated circuit (ASIC).

Therefore, the Intel RealSense camera SR300 which employed structured coded light technology is adopted in this study as shown in Fig.2 [45]. The SR300 enables to output synchronized color, depth and infrared image stream data to the processor, however, this study only utilizes the depth data stream. It should be noted that the emitted structured light is infrared light, the system can work properly in dark environments. In addition, the system can work in a normal operating temperature of 0-35 °C, and a humidity of 90% RH at 30 °C [45].

The process of generating one frame depth image is shown in Fig.3. The infrared laser projector emits a set of predefined, increasing spatial frequency coded infrared vertical bar patterns to the object [47]. When these coded lights contact the object, the vertical bar patterns would be warped
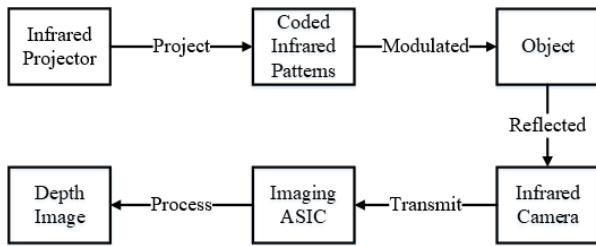
FIGURE 3. The process of generating depth image by the SR300.

by the objects. Then the infrared camera will capture the reflected, warped patterns to compare them with the reference patterns for decoding and calculating the depth information of the objects as described in above section. In the SR300 system, the infrared camera pixels are processed by the imaging ASIC to generate one frame depth image. Finally, the depth images are transmitted to the client system.

C. THE PROPOSED RESPIRATORY REGION DETECTING ALGORITHM

To extract respiratory activity, the respiratory region needs to be detected. Therefore, the respiratory region detection algorithm is developed to process all depth images, Then the depth waveform can be plotted by recording average depth change in the respiratory region. The proposed respiratory region detecting algorithm is mainly based on morphological method, which contains four steps: body edge extraction, noise reduction, "image skeleton" extraction and respiratory region estimation as shown in Fig.4.
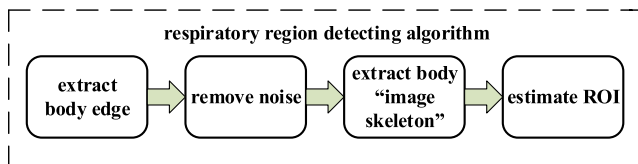

FIGURE 4. The architecture of the proposed respiratory region tracking algorithm.

1) BODY EDGE EXTRACTION

After obtaining the depth images, background should be segmented and removed. But in this system, the distant background and objects can't return the effective depth data because they are beyond the effective distance of the camera. As mentioned before, the effective range of this system is 0.2-1.5m, if the subject is within this range, the subject will occupy a relatively large area, or cover other non-human objects. When the subject leaves this range, the fixed non-human object will not cause a depth change. Considering our application scenario, the case that non-human objects occupy majority area can be artificially avoided. In other word, the depth camera can eliminate the distant background automatically during acquiring the depth data. So traditional

background removal step is not needed which would greatly reduce the processing time.

After the depth camera capturing the subject who is in the effective range, the image is transformed into a binary image by thresholding, and a contour-tracing algorithm based on the Moore-Neighbor tracing algorithm with Jacob's stopping condition is used to determine the regions in the thresholded binary image [48], [49]. This algorithm also obtains other regions which belong to the small parts of the body, like hands, patterns on the cloth, head and background near the subject. But those regions' contours comprise less pixels than the contour of main body region, which means the body contour would contain most pixels if the subject was in the effective range of the camera. Thus, the body edge can be extracted by selecting the contour which has most pixels.

2) NOISE REMOVAL PROCEDURE

The extracted body contour would have unsmoothed and noisy edge, such as clothes' fold, disturbance from other objects. Those abrupt parts would lead to useless features which influence the respiratory region estimation. Therefore, the body contour should be smoothed to retain the main edge.

The Fourier descriptors can be utilized to describe the contour and select main components [50]. Supposing that the boundary in an image comprised of K points with a two dimensional coordinate system. Starting at an arbitrary point (x_0, y_0), coordinate pairs $(x_0, y_0), (x_1, y_1), (x_2, y_2), \ldots, (x_{K-1}, y_{K-1})$ are the points in traversing the boundary with the counter-clockwise direction, and these coordinates can be expressed in the form $x(k) = x_k$ and $y(k) = y_k$. For $k = 0, 1, 2, \ldots, K - 1$, the boundary can be represented as the sequence of coordinates $s(k) = [x(k), y(k)]$. Moreover, each coordinate pair can be treated as a complex number

$$s(k) = x(k) + jy(k) \tag{1}$$

where $x(k)$ is considered as the real part and $y(k)$ is treated as the imaginary part of a sequence of complex numbers $s(k)$, for $k = 0, 1, 2, \ldots, K - 1$. The nature of the boundary did not change while transforming the 2-D boundary to a 1-D sequence.

From (2), the discrete Fourier transform (DFT) of $s(k)$ is

$$a(u) = \sum_{k=0}^{K-1} s(k)e^{-j2\pi uk/K} \tag{2}$$

for $u = 0, 1, 2, \ldots, K - 1$. The complex coefficients $a(u)$ is the Fourier descriptors of the boundary. Obviously, the inverse Fourier transform of these coefficients can be drawn as $s(k)$ in the following form:

$$s(k) = \frac{1}{K} \sum_{u=0}^{K-1} a(u)e^{j2\pi uk/K} \tag{3}$$

for $k = 0, 1, 2, \ldots, K - 1$. Suppose only the first P coefficients are used to describe the image instead of all coefficients, namely setting $a(u) = 0$ for $u > P - 1$ in (3).

The approximation expression can be drawn as the following form:

$$\hat{s}(k) = \frac{1}{K} \sum_{u=0}^{P-1} a(u) e^{j 2\pi u k / K} \qquad (4)$$

for $k = 0, 1, 2, \ldots, K - 1$. Although only P terms are used to recover each component of $\hat{s}(k)$, the same number of points still exists in the reconstructed boundary. The high-frequency components represent the detail information while the low-frequency components determine the global information. Hence many details of the boundary would be lost if fewer terms are used for reconstruction which means the small P. For instance, Fig.5 shows a boundary of the upper part of a body which consists of 2996 points. The Fourier descriptors of the boundary which contains 2996 items can be obtained using (2). As the number of descriptors gradually reduced, different boundaries reconstructed by partial Fourier descriptors can be obtained.
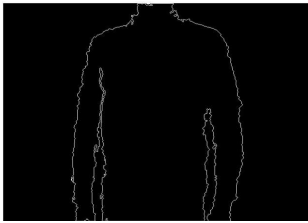

FIGURE 5. The boundary of the upper part of the body which consisting of 2996 points.
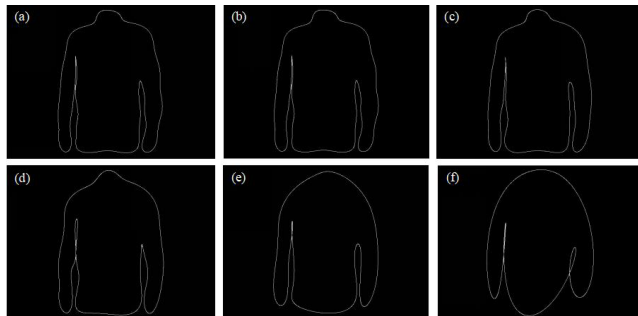

FIGURE 6. (a)-(f) The boundaries reconstructed from the Fourier descriptors whose number are 80%, 50%, 40%, 30%, 20% and 10% of original 2996 descriptors.

Fig.6(a) to Fig.6(f) show the boundaries reconstructed from the Fourier descriptors whose number are 80%, 50%, 40%, 30%, 20% and 10% of 2996, respectively. These percentages are approximate to 2397, 1498, 1198, 899, 599, and 300, respectively. Fig.6(b) shows the reconstructed boundary which only used half of the 2996 descriptors. It is noteworthy that there is no distinct difference between Fig.6(b) and Fig.6(a) which reconstructed by 80% of original descriptors, and the Fig.6(a) and Fig.6(d) are also very similar.

This phenomenon represented that when the coefficients whose amount between 30% and 80% of original descriptors

were used to reconstruct images, the differences between these reconstructed boundaries are very small. From Fig.6(d), we can clearly see that 30% of 2996 descriptors were sufficient enough to retain main shape of the original boundary, a torso with two arms can be clearly distinguished. However, Fig.6(e) and (f) recovered from only 20% and 10% of original descriptors respectively are not good enough due to lots of loss of human body edge information, for example, the neck shape and the details of the edge of the body are lost. If the descriptors were further reduced, the result would tend to a circle. The top Fourier descriptors which relates to low-frequency information control the global features of the image, such as size, shape and profile. The high-frequency information mainly control the detail information, such as edge and texture. The information of the body posture shape and size is controlled by the top Fourier descriptors, and 30% of Fourier descriptors is enough to preserve those information. As the preceding example demonstrates, most of the Fourier descriptors can be thrown out, and a few descriptors are significant enough to reflect the global feature of a body boundary. This main body shape without the high-frequency details can benefit a lot to the following process.

3) BODY "IMAGE SKELETON" EXTRACTION

After the body region being reconstructed by a few number of Fourier descriptors, the binary image is used for extracting "image skeleton" of the main body part with employment of morphology method. It's worth noting that we extract the "image skeleton" of the body part instead of true skeleton of the body for estimating the respiratory region. Because the "image skeleton" can concentrate the characteristic of the body edge and reflect the real body region which contains the respiratory region. However, the true skeleton cannot consider the physical characteristic of subject such as bodily form, which may result in misjudging the respiratory region. As shown in Fig.7, the "image skeleton" of Fig.5 can be utilized to roughly represent the body region, which is extracted by a morphological image processing method.
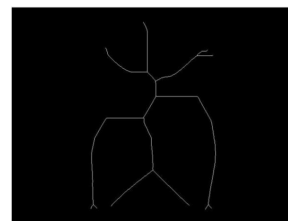

FIGURE 7. The "image skeleton" of the body region extracted by the morphological method.

Morphological image processing method is commonly considered as a tool for extracting some image features such as boundary, skeleton and convex hull, which could be useful in representation or description of image characteristic. It also can be used for pre-processing or post-processing the images, such as filtering, thinning and pruning operation. The adopted method utilizes some operating elements which have their

own structure on the binary image, then a new image can be generated by running these operating elements over original image to make them visit every point of the original image. In this study, the operation is a combination in terms of "erosions" and "openings" which are detailed in [51].

After the process, a lot of skeleton subsets can be obtained for constituting an integrated skeleton image, by this way, the skeleton is maximally thin which is one pixel wide. The immediate advantage of the morphological method is that the operation is simple and don't require the RGB image used for providing the information.

4) RESPIRATORY REGION OR RESPIRATORY INFLUENCED REGION OBTAINING

The respiratory region can be estimated with the help of "image skeleton" which can reflect the body region, such as thorax and abdomen region. If the location of the respiratory region is acquired from the "image skeleton", the respiration condition can be speculated based on the variation of depth data in the ROI. It's worth noting that people may have lateral or prone posture instead of supine posture during sleep, while it is difficult to extract the information of the thorax or abdomen region from the images with this condition. Nevertheless, the respiratory efforts always result in alternating movements on the whole upper body, in other words, although the depth camera can only capture the lateral posture or the prone posture of subjects, the respiratory information can also be extracted from the depth values in shoulder region or back region. Therefore, it's important to be able to locate the respiratory influenced region such as the shoulder, the back, and even upper arm based on the "image skeleton" when the depth images didn't contain the obvious respiratory region like the thorax and abdomen. In this case, the popular machine learning method for finding out the ROI is not suitable for this condition [52], [53], due to lack of the depth images for training and fewer features from those noisy depth images. Hence, the morphological processing method which can extract the ROI simply, would be more suitable for processing the depth image.
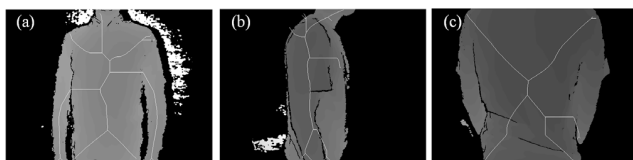

FIGURE 8. The "image skeleton" extracted from the corresponding depth images with the condition that the subject was in the front of the camera with the three different directions.

With the help of the "image skeleton", the ROI can be estimated in the following way. Fig.8 (a), (b), (c) show the depth images and their corresponding "image skeleton" when the subject was in the front of the camera with three different directions. It can be seen that the "image skeleton" reflect well the real region of the human body in those images. There always has one approximate vertical skeleton which play the

role of "supporting" the structure of "image skeleton" just like the spine in real skeleton structure, in addition, some skeletons with different direction extended from the vertical skeleton reflect the size of the body. So if the information of those skeletons were obtained, it would facilitate the positioning of the ROI.

Many methods can be utilized to obtain the morphological features like the skeletons and specific points in the binary image. For example, the Hough Transform (HT) and Look Up Table (LUT) are classic methods for detecting complex and specific patterns of points in binary image data [54], [55]. The HF achieves this by converting the original points into parameter space, then finding out specific values of parameters which can characterize the desired patterns. But practically, subjects are not always in front of the camera with a standard posture, and "image skeletons" would have various characteristics which make it complicated to extract useful information with limited preconditions. The LUT is a simple and efficient way to find out the specific patterns such as endpoints, junction points or other patterns with complex form. If the objective's patterns are not complicated, the LUT only needs a little memory to store the pre-computed patterns of object point. For example, the endpoints in the binary image can be represented as a 3×3 matrix with eight different styles when the endpoint is in the centre of the matrix.
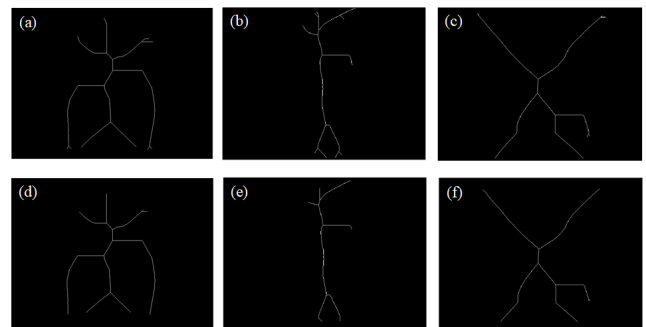

FIGURE 9. The (a), (b) and (c) represent the "image skeleton" extracted from the Fig.8, and (d), (e) and (f) represent the corresponding "image skeleton" after the twenty iterations pruning process.

In order to simplify the process of finding endpoints and further avoid useless information, an iterative process was conducted to obtain the final endpoints as shown in the Fig.9. Fig.9 (a), (b), (c) represent the "image skeleton" extracted from the Fig.8, the (d), (e), (f) is corresponding "image skeleton" after the twenty iterations pruning process. It can be seen that most of the minutiae have been pruned, some redundant endpoints are reduced, and new skeleton is simplified. The final endpoints in the processed "image skeleton" are good indicators for delimiting the respiratory and respiratory influenced region.

Then the ROI could be approximately obtained by connecting the endpoints in sequence to enclose a polygon region or constructing a rectangle ROI based on the maximum and minimum coordinates of those endpoints [32].
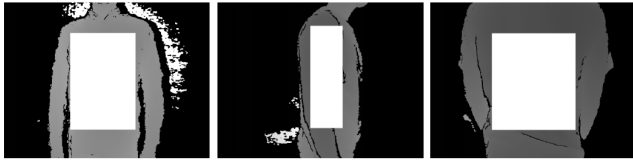
FIGURE 10. The ROIs of original depth images whose points were labeled with the white masks.

For example, let the final endpoints be represented by the coordinate pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, assume $x_{min} = \min\{x_1, x_2, \ldots, x_n\}$, $x_{max} = \max\{x_1, x_2, \ldots, x_n\}$, $y_{min} = \min\{y_1, y_2, ..., y_n\}$ and $y_{max} = \max\{y_1, y_2, \ldots, y_n\}$. Draw two vertical lines through point x_{min} and x_{max}, and draw two horizontal lines through point y_{min} and y_{max}, then enclosed the four lines to obtain a rectangle region, then scale the rectangular region down to the original 80% to obtain the rectangle ROI. To solve this problem with an easy solution, the rectangle region is adopted to represent the ROI. As shown in Fig.10, the coordinate positions of each points in the ROI of the depth images shown in the Fig.8 can be labeled with a white mask, then the depth data in the ROI can be extracted via the location information.

D. PLOT DEPTH WAVEFORM AND CALCULATE RESPIRATORY RATE FROM DEPTH CHANGE IN THE ROI

In order to obtain the respiratory state, the average depth value of all points in the ROI was calculated. The average depth value can reflect respiratory activity condition on the shooting moment of a depth image. As shown in Fig.11, along one 30-second epoch, the normalized average depth values and processed depth waveform can be plotted with black dots and red solid line, respectively. The trend of the average depth values (black dots) reflected the alternating respiratory movements, then the depth waveform could be depicted by connecting those average depth values. Although the primary trend can be observed without processing, it is hard to extract some respiratory features automatically such as the RR from the raw waves. Since the average depth value of ROI would be polluted by noises caused by slight body movements, the depth waveform should be processed for removing the noises as post-processing, as the red solid line drawn in the Fig.11. We applied the Butterworth low-pass filter with cut off frequency of 15 Hz for filtering the system error as a preliminary process, then utilized the piecewise polynomial model to process the filtered depth signals and obtain the desired information due to its powerful denoising ability.

The piecewise polynomial model is defined by multiple sub-models, each sub-model applied to a certain sub-interval of the main domain (one epoch). The process of the sub-model can be expressed as follows:

Let $\boldsymbol{v}$ represents a sequence of depth values $\boldsymbol{v} = \{v_1, v_2, \ldots, v_n\}$ at their corresponding time indices $\mathbf{t} = \{t_1, t_2, \ldots, t_n\}$ (in one epoch), then a relation between them
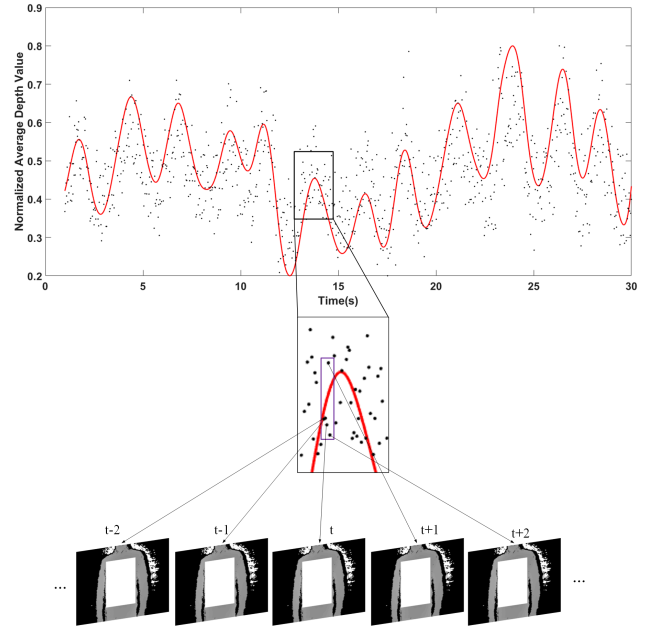

FIGURE 11. The normalized average depth values and the processed depth waveform are plotted with black dots and red solid line respectively, along one 30-second epoch. The abscissa axis and the vertical axis in the coordinate system represent the time and the normalized average depth value of the ROI, respectively. As mentioned before, each depth value is obtained by averaging all depth values in the ROI of the depth image at the corresponding frame. For example, five depth data points in the purple box represent the averaged values of the ROI in five images acquired at the moment of t−2, t−1, t, t+1, t+2.

can be modeled by

$$v_i = h(t_i) + \varepsilon_i \quad (i = 1, 2, ..., n) \tag{5}$$

where h is a smoothing spline function, ε_i are independent and identically distributed residuals. The smoothing function can be estimated by minimizing the objective function such that

$$\hat{h} = \arg\min_h \left[p \sum_{i=1}^{n} [v_i - h(t_i)]^2 + (1-p) \int_{t}^{t_n} h''(t)^2 dt \right] \tag{6}$$

where p is the smoothing parameter. The smoothing function can be expressed by cubic splines and least squares approximation [56].

This smoothing process was realized by the curve fitting tools in Matlab2016a, and it consist two steps, the first step was pre-smoothing and the second step was fine-tuning. In the first step, the curve was pre-smoothed by the piecewise polynomial model with a predefined smoothing parameter ($p = 0.999976823063204$) [57], and the larger the p value, the more high-frequency information in the model retains, on the contrary, more high-frequency information will be reduced. The parameter value was chosen by initial experimental exploration.

Once the smoothed depth waveform is generated, the RR can be calculated. In general, the RR is calculated from the time interval between adjacent end-inspiration points in the

respiratory waveform for a certain time called one epoch. The end-inspiration points can be easily obtained in the smoothed curve by a peak detection process. For this study, we set 2 minutes as one epoch and calculate its RR. Many preliminary explorations suggested that if the object was closer to the camera, the depth value extracted from the object was smaller. So the valleys in the depth waveform represented the end of inhalation, since the human body expanded and was closer to the camera during the end of inhalation. Thus, the depth waveform would be turned over and normalized before peak detection. Then the RR for each epoch could be given by

$$INT_i = \frac{\sum_{j=1}^{n-1}(t_{i,j+1} - t_{i,j})}{n-1} \qquad (7)$$

$$RR_i = \frac{60}{INT_i} \qquad (8)$$

in which $t_{i,j}$ is the moment of jth peak of the ith epoch, and INT_i is mean peak interval for ith epoch, n is the number of peaks, so the RR of ith epoch could be demonstrated by RR_i. The unit of the INT and RR is second (s) and times per minute, respectively.

Since different subject's respiratory rate is quite different and the depth camera can capture components of various frequencies, hence, the main component of the respiratory signal of one subject may be noise to other subjects. For example, for the subject whose RR is 23/min, the signal with relative higher frequency components may be the desired respiratory signal, but a component with the same frequency components may be noise for subject whose RR is 8/min. Using a uniform model for each subject to obtain the desired smooth respiratory waveform is not easy. In addition, the fitted curve would also lose some information, and a single smoothing parameter cannot adapt to the depth data of all cases. Therefore, after the preliminary filtering process, the RR calculated by the first step can be exploited as an indicator to reflect the respiratory condition. Then we can adjust the smooth model to optimize the morphology of the depth waveform based on respiratory condition. So once obtaining the RR of one epoch, the piecewise polynomial model would be fine-tuned as the second step of smoothing process.

The rule to select the smoothing parameter p of piecewise polynomial model in the second step according to the RR of the previous step is defined as follows:

$$p = \begin{cases} 0.999905936547889 & RR < 14 \\ 0.999976823063204 & 14 \leq RR < 17 \\ 0.999991823797345 & RR \geq 17 \end{cases} \qquad (9)$$

where p is the smoothing parameter. As mentioned above, p value can reflect the frequency components remained in the waveform, thus the p value would be turned up while the RR calculated from the first step is greater and equal than 17, which indicates the respiration may have more high-frequency information. Conversely, when the RR is less

than 14, which shows a respiration of relatively lower frequency, the p value will be turned down. And the p value will not be changed if the RR is between 14 and 17. Besides, the preliminary tests verified that the RR 14 and 17 as the threshold values can achieve promising results. After the smoothing model being adjusted, the RR would be recalculated from the adjusted depth waveform processed by the new model. The three different p values in the second step were selected based on empirical experiments, and it can effectively solve the problem that the predefined smoothing parameter was incapable to be applied in various respiratory patterns with different RRs. Then define the average accuracy of RR for estimating the performance of generating the depth waveform in the following form [21]:

$$\text{Accuracy of RR} = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \sqrt{\frac{(RR_i^D - RR_i^P)^2}{(RR_i^P)^2}}\right) \qquad (10)$$

in which RR_i^D denotes the RR of ith epoch calculated from depth waveform generated by the average depth values, and RR_i^P denotes the RR of ith epoch calculated from the thorax effort signals obtained by the PSG device which can be considered as the gold-standard system. The N represents the total number of epochs used for calculating the accuracy of RR with one certain condition.

III. EXPERIMENT AND RESULTS
In this section, the experiments of extracting the RR from the depth waveform were conducted to verify the performance of the algorithm introduced in Section II-C and Section II-D, while the RR calculated from the standard PSG thorax effort signal was applied for comparison.

A. EXPERIMENT SETUP
The experiments were conducted in the Home Lab at the Center for Intelligent Medical Electronics, Fudan University. In order to verify the feasibility of the system under different conditions. Twenty kinds of scenes which combined five basic postures of the subjects and four external conditions were carried out in the experiments. Ten healthy people (five males and five females, age 25.2 ± 2.44, BMI 20.53 ± 1.43) were employed to be the subjects as shown in the Table 2.

1) FIVE BASIC POSTURES OF THE SUBJECT
In this study, we consider five basic postures for measuring the RR. For indoor monitoring, subject might lay in bed with supine, lateral or prone posture, and the subject might also sit on a chair or stand, and these postures were shown in Fig.12.

The five basic postures can be derived and labeled with No. 1 to No. 5 which are described as follows:
1) No. 1-the subject lay in bed with supine posture, the camera was placed above the body.
2) No. 2-the subject lay in bed with lateral posture, the camera was placed above the body.
3) No. 3-the subject lay in bed with prone posture, the camera was placed above the body.

TABLE 2. The basic information of ten subjects.

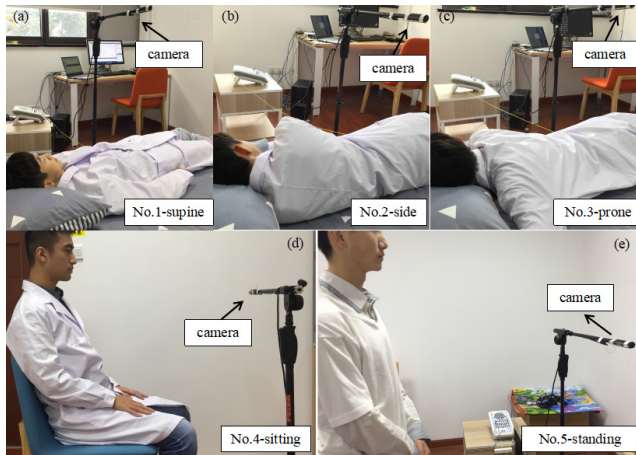| Test	Gender	Age	Height	BMI
1	male	27	182	19.62
2	male	24	170	22.49
3	male	24	177	22.98
4	male	22	175	20.08
5	male	24	168	21.61
6	female	28	162	20.4
7	female	25	163	19.75
8	female	29	166	19.28
9	female	22	173	18.5
10	female	27	161	20.58

FIGURE 12. (a)-(e) Five basic postures: (a) lay in the bed with supine posture (b) lay in the bed with lateral posture (c) lay in the bed with prone posture (d) sitting on the chair (e) standing in front of the camera.

4) No. 4-the subject was sitting on chair in front of the camera, and the height of seat and camera is 48 cm and 102 cm from the ground, respectively, which can also catch the thorax region.

5) No. 5- the subject was standing in front of the camera. The height of the camera is about 125 cm, in which case the thorax and adjacent region can be captured by the camera for most adults.

2) FOUR EXTERNAL CONDITIONS

Besides the five basic postures, four different external conditions were taken into account. People might wear different kinds of clothes and be in different distance with the camera, thus thickness of clothes and distances to the camera are important factors for depth detection. Hence the conditions that subjects wearing clothes with different thicknesses and being different distances from thorax surface of subject to the camera were considered as shown in the Fig.13.

Four external conditions were generalized as the application scenes and labeled with No. 1 to No. 4 which are described as follows:
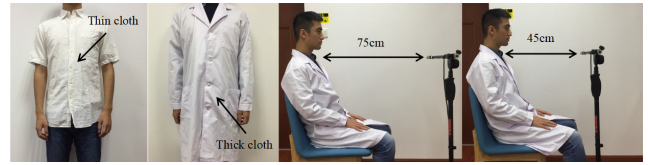

FIGURE 13. Two different clothes and two different distances employed by this experiment. The thin cloth and the thick cloth is common T-shirt and white gown with shirt inside, respectively.

1) No. 1-the subject wearing thin cloth with 45 cm from the camera.
2) No. 2-the subject wearing thin cloth with 75 cm from the camera.
3) No. 3-the subject wearing thick cloth with 45 cm from the camera.
4) No. 4-the subject wearing thick cloth with 75 cm from the camera.

The distances of 45 cm and 75 cm are chosen according to the possible future scenarios. For example, in sleep monitoring, it is suitable when the distance between the depth camera and the human body is 75 cm, and for the home health care, 45 cm is probably the distance from the body to the computer monitor or the laptop.

Some samples of the captured depth images within 45cm and 75cm distances are shown in the Fig.14.
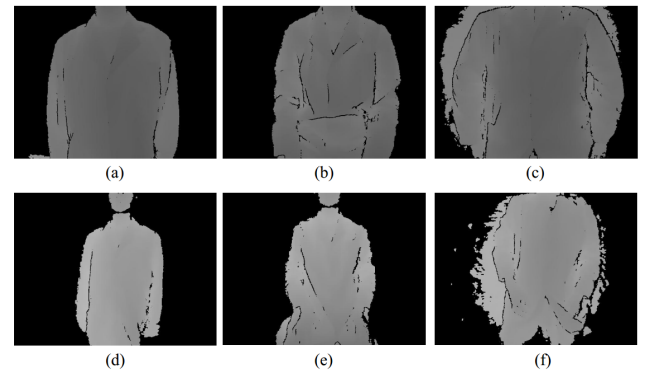

FIGURE 14. Six samples of the captured depth images of Subject-1. The images obtained with 45 cm and 75 cm distance are shown in the upper row and the lower row, respectively. (a) and (d) show that the subject is in the stand posture, (b) and (e) show the subject is in the sit posture, and (c) and (f) show the subject is in the supine posture.

B. THE PERFORMANCE OF EXTRACTING RESPIRATORATION RATE FROM DEPTH WAVEFORM WITH DIFFERENT CONDITIONS

The experiments were conducted on 20 different conditions. For each subject, three repeated trials were carried out for each condition, and each trial collected depth data lasting for more than two minutes. The ten subjects were allowed to have normal breathing and asked to keep one basic posture during one trial. Some studies let the subjects breathe with fixed RR, and extract the corresponding respiratory waveform [25], [58]. But if the subject is requested to breathe
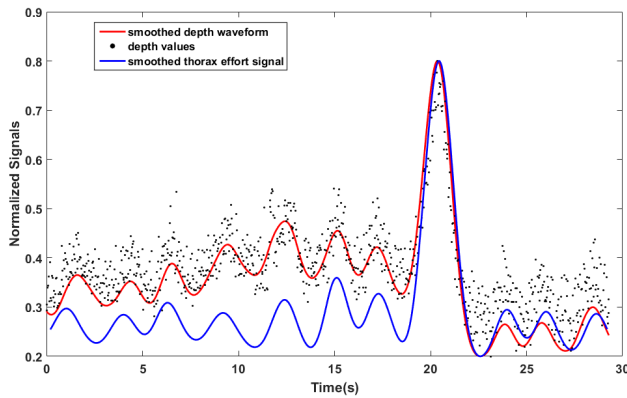
FIGURE 15. The depth waveform and the waveform obtained from thorax effort signal of PSG is drawn by the red solid line and blue solid line in the same 30-second epoch, respectively. About 11 peaks which represent 11 end-inspirations can be found in each wave, and the eighth respiratory movement is a deep breath depicted well by the depth wave.

at a fixed RR, the respiration may not meet the breathing habits of the subject. In addition, the respiratory waveform extracted from those fixed patterns will be more usual, which are more convenient for processing. But we expect the proposed system can measure the respiration without too many restrictions and obtain the less usual signals, thus our research attempts to allow the subjects to breathe in the most natural condition. To obtain the standard respiratory signal for comparison, the Compumedics Grael PSG system was applied to simultaneously record the thorax effort signal as the reference. The depth images were recorded with timestamp, and the signals in the PSG system have the time information during the recording, hence, the time synchronization process can be easily realized. Then both the depth wave and PSG thorax effort wave would be processed as discussed before. An example of comparison between the processed depth waves and PSG thorax effort waves for the same 30 second is shown in Fig.15. It can be seen that the two waves have similar trend. In this study, we calculate the RR with the help of peak detection approach whose performance is dependent on the smoothness degree of the waveform, or the similarity between the smoothed depth waveform and the thorax effort waveform. Therefore, the similarity between the two kinds of waveforms is very important. The correlation between the depth waveform and the thorax effort waveform which can reflect the similarity was calculated. The average correlation between two kinds of waveforms is 0.8587 ± 0.1263 via all subjects in 20 different conditions, and each condition had 30 epochs data (N = 30). The result represented that the depth waveform extracted from the depth data has a promising similarity to the thorax effort waveform, which indicates that the depth waveform has potential to extract more meaningful information.

The Bland-Altman analysis is an important tool to estimate the agreement between the designed method and gold standard. The Bland-Altman analysis for the RR of all 600 samples obtained by the proposed system and the PSG device is conducted as shown in Fig.16.
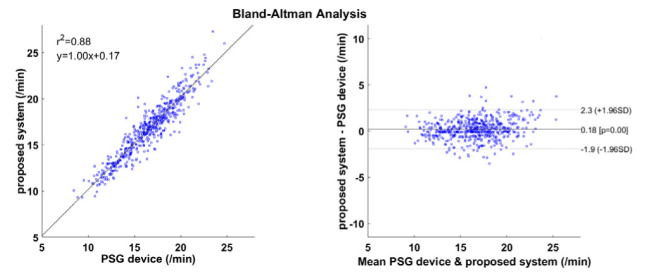

FIGURE 16. The Bland-Altman analysis for the RR of all 600 samples obtained by the proposed system and the PSG device.

From the regression line between hypothetical measurements done by the proposed system and the PSG device, it can be seen that the two methods have good agreement ($r^2 = 0.88$, y = 1.00x + 0.17). From the Bland and Altman plot, we can see that most of the differences between the results obtained from the proposed system and the PSG device are in an acceptable range. Besides, we also find that the subject's RR occupies the range of 10 to 23 /min, completely covering the healthy people's RR, and even contains some breaths with abnormal frequency.

Then the average accuracy of RR from depth waveform could be calculated as mentioned in Section II-D, while the thorax effort wave's RR would be extracted as the standard result. Table 3 showed the average accuracy of RR via all subjects with 20 different conditions, and each condition had 30 epochs data (N = 30).

TABLE 3. The average accuracy of RR calculated from depth waveform via all subjects with 20 different conditions.

Postures	Accuracy of RR (mean% ± SD%)				
	External conditions				Total
	No.1	No.2	No.3	No.4	
No.1	94.75 ± 3.52	94.36 ± 4.93	96.47 ± 3.06	94.86 ± 5.14	95.11 ± 4.28
No.2	95.36 ± 5.00	**94.05 ± 4.68**	95.31 ± 4.20	96.15± 4.74	95.22 ± 4.67
No.3	95.19 ± 3.06	94.88 ± 5.04	95.20 ± 4.35	94.29 ± 4.19	94.89 ± 4.19
No.4	**96.54 ± 3.96**	95.58 ± 4.53	95.81 ± 3.27	96.00 ± 3.26	95.98 ± 3.76
No.5	94.81 ± 4.06	94.13 ± 4.35	94.69 ± 4.36	95.48 ± 3.63	94.78 ± 4.09
Total	95.33 ± 3.97	94.60 ± 4.68	95.49 ± 3.88	95.36 ± 4.25	95.20 ± 4.21

From Table 3, it can be seen that the best performance of extracting the RR from depth waveform is 96.54% in the condition that subject was sitting on the chair with a thin cloth and the distance between the camera and subject is 45 cm. The worst performance is 94.05% in the condition that subject wore a thin cloth lying in bed with a lateral posture and the distance between the camera and subject is 75 cm.

The total average accuracy of RR calculated from depth waveform is 95.20%. If the posture is considered as the independent variable, the results obtained from five different postures are shown in the last column. Similarly, the results obtained from four different external conditions without regard to different postures are shown in the last row. The accuracy of the RR extracted from the depth waveform can achieve 95.11%, 95.22% and 95.98% while the subject was lying in the bed with supine posture, lateral posture and sitting in the chair, respectively, which suggested that the system was very suitable for the home health care. When subjects wore thick cloth with 45 cm distance from the camera, it can achieve the best performance of 95.49% in four different external conditions. Although most of the results with different conditions in Table 3 are promising (94.05%-96.54%), we cannot determine whether the samples extracted from different external conditions or different postures come from the same population based on those limited measurements. Thus, the results need to be tested for determining whether the differences between those results obtained from different conditions has statistical significance. In addition, the effect of age for the RR measurement is also worth studying, so we divided our subjects into two groups, five subjects in the first group are the older people whose age are 25, 27, 27, 28, and 29, and the other five subjects in the second group are the younger people whose age are 22, 22, 24, 24, and 24. Hence, the differences of the results between two age groups should also be investigated.

In this case, the Kruskal-Wallis one-way analysis of variance test was employed to analyze the results [59]. The Kruskal-Wallis test is a non-parametric version of the classical one-way analysis of variance, and an extension of the Wilcoxon rank sum test to more than two sample groups [60]. It compares the medians of the groups to determine whether the samples come from the same population, or equivalently from different populations with the same distribution. Hence three Kruskal-Wallis tests were used for estimating whether the results extracted from different postures, different external conditions and different age groups have significant differences.

The first test is to analyze whether the differences in the results obtained from four different external conditions are statistically significant. The Kruskal-Wallis ANOVA and box diagram of the results from the first test are shown in the first row of main area of Table 4 and Fig.17, respectively.

The results indicate that there is no significant difference between the four groups ($p > 0.05$), so it can be concluded that those four external environments didn't influence the performance for obtaining the RR from depth data. In other words, the depth camera has a good ability to adapt clothes with different thickness on subject and different distance between the camera and subject in an effective range.

The second test is to analyze whether the differences in the results obtained from the five different postures are statistically significant. The Kruskal-Wallis ANOVA of those results

TABLE 4. The ANOVA table for the results obtained from three experiments via Kruskal-Wallis test.

Test	SS	df	MS	Chi-sq	p
1	74583.1	3	24861.0	2.48	0.4786
2	249759.5	4	62439.9	8.31	0.0808
3	85490.4	1	85490.4	2.84	0.0917

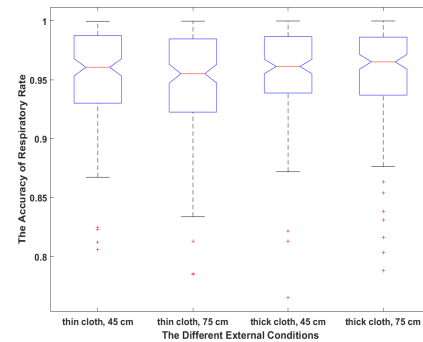SS = sum of squares of deviations, df = degree of freedom, MS = mean square.

FIGURE 17. The box figure of the RR accuracies for four external conditions.
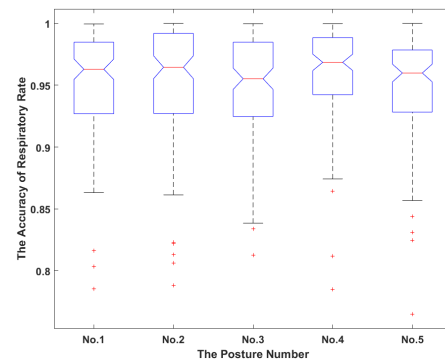

FIGURE 18. The box figure of the RR accuracies for five different postures.

is shown in the second row of the main area of Table 4, and the box diagram of the second test results is shown in Fig.18.

The results indicated that there is still no significant difference between the five groups ($p > 0.05$). Therefore, these results suggested that those five postures acted by the subjects didn't influence the method performance. In other words, the system has good ability to extract the RR with different postures on subjects.

The third test is to analyze whether the differences in the results obtained from the two age groups are statistically significant. The Kruskal-Wallis ANOVA of those results is shown in the third row of the main area of Table 4, and the box diagram of the third test results is shown in Fig.19.

The results indicated that there is still no significant difference between the two age groups ($p > 0.05$). Therefore, the
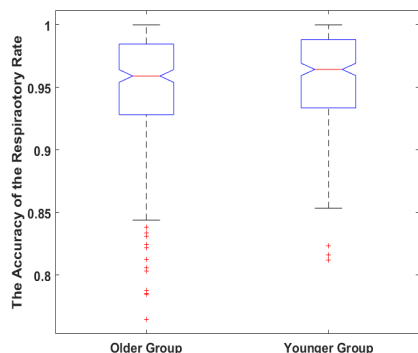
FIGURE 19. The box figure of the RR accuracies for the older and younger groups.

test suggested that the age of the subjects in the range of 22-29 didn't influence the method performance.

C. COMPARISON WITH THE STATE-OF-THE-ART APPROACHES

In the field of respiratory measurement, many studied based on the depth images have proved the feasibility of their systems from different perspectives and different measurement metrics [26], [27], [29], [32]–[34]. Those methods contribute to the state-of-art unobtrusive respiration monitoring performance in their directions and provide important guidance for related research. In order to demonstrate the validity of our proposed method and get a whole picture of the development of this field, we compared our proposed model with those related studies as the state-of-the-art approaches. We introduce each of them in the following paragraphs.

1) Wijenayake and Park [26] proposed a respiratory motion tracking approach by exploiting the advantages of RGB and depth images. A specific respiratory motion model is utilized for external respiratory motion measurement, and the respiratory pattern's correlation coefficient is $r^2 = 0.941$ comparing to the spirometry result as the groundtruth.

2) Yu *et al.* [27] developed a system to measure human chest wall motion for respiratory volume estimation based on depth image sensing technique, the respiratory volume is estimated by measuring morphological changes of the chest wall. The system was evaluated by comparing with a spirometry, and the agreement in respiratory volume measurement is $r^2 = 0.933$.

3) Benetazzo *et al.* [29] presented an RR measurement algorithm that uses the depth images, the RR is derived by measuring morphological changes of the chest wall by identifying the human chest, computes its distance from the camera. The results from five subjects have shown that the error between the measurements obtained by the algorithm and those obtained by the spirometry has a maximum value of M = 0.533 and the correlation coefficient is $r^2 = 0.863$.

4) Ostadabbas *et al.* [32] proposed a vision-based passive airway resistance estimation technology based on the depth

images, for obtaining the tidal volume of the subjects, an automatic chest bounding detection algorithm combined with the airflow calculation approach is developed. The average error of 0.07 ± 0.06 liter for estimating the tidal volume (TV) and patients with airway obstruction (AO) were detected with 80% accuracy.

5) Soleimani *et al.* [33] presented a vision-based trunk-motion tolerant approach which estimated lung volume-time data remotely in forced vital capacity (FVC) by using two opposing Kinects. After modelling of trunk shape, the chest-surface respiratory pattern can be computed on temporal geometrical features extracted from the chest and posterior shapes. Then the FVC is estimated by the TV waves, while the spirometry is used as the groundtruth. When the proposed method used only the TV scaling factor, the agreement in FVC measurement is $r^2 = 0.93$, and the average normalized L2 error of is 0.05.

6) Sharp *et al.* [34] developed a chest volume estimation method based on the Kinect V2, which reconstructed the 3D models of the subject's thorax to obtain the volume-time and flow-time curves. The FVC was used to estimate the performance of the method, and the spirometry was considered as the groundtruth. The agreement r^2 in FVC measurement can achieve 0.99 that is a very high value, but the agreement r^2 in forced expiratory volume (FEV) measurement did not demonstrate acceptable limits of agreement ($r^2 = 0.38$).

Table 5 shows the comparison between our method and the state-of-the-art approaches across number of subjects, material, groundtruth, respiratory indicator, and measurement metric. It can be seen that most related studies measure the respiratory volumes as the respiratory indicator, such as TV, FVC. Those studies show good results with a larger number of subjects which indicated that the respiratory volume measurement based on the depth images is reliable and promising. The respiratory volumes are commonly collected from single respiratory effect with the spirometry or the pulse oximeter. The experiment process of volume measurement is faster than the process of RR measurement because measuring RR needs a period of time, and the epochs with more than 2-min length was used to measure the RR in our study. It would occupy a long period of time of the subjects, which leads to the lack of RR experiments. Comparing with the study which also used the RR as the indicator [29], the proposed study collected a larger amount of experimental data and achieved better results. Besides, our proposed method can be a powerful complement to the related studies which used the depth images for acquiring the respiratory indicators. With the increasing number of respiratory indicators that can be extracted by the depth images, the unobtrusive respiratory measurement based on the depth camera will become more comprehensive. In addition, these experiments also proved that it is sufficient to extract the respiratory features for the measurement by using only the depth images without RGB images.

TABLE 5. The performance Comparison of the state-of-the-art methods.

Method	Num. of subjects	Material (images)	Groundtruth	Respiratory indicator	Measurement metric 1		Measurement metric 2	
					Metric	Value	Metric	Value
Ref [26]	10	RGB + Depth	Spirometry	RP	r^2	0.941	-	-
Ref [27]	12	RGB + Depth	Spirometry	Volume	r^2	0.933	-	-
Ref [29]	5	Depth	Spirometry	**RR**	r^2	0.863	Max error	0.533
Ref [32]	14	Depth	Pulse oximeter	TV	Average error	0.07 liter	Acc. of AO	80%
Ref [33]	35	Dual-Depth	Spirometry	FVC	r^2	0.93	L2 error	0.05
Ref [34]	80	Depth	Spirometry	FVC/FEV	r^2 (FVC)	0.99	r^2 (FEV)	0.38
Our method	**10**	**Depth**	**Throax effort**	**RR**	**r^2**	**0.88**	**Acc.**	**95.2%**

Num. = number, RP = respiratory pattern, TV = tidal volume, Acc. = accuracy, AO = airway obstruction, FVC = forced vital capacity, RR = respiratory rate.

IV. DISCUSSION AND FUTURE WORK

The proposed system is convenient for setup, completely non-contact and has good adaptability in various kinds of indoor environments. More importantly, the system is applicable in home healthcare, community healthcare and other scenes due to its low-cost and high accuracy. Instead, the thorax effort band would influence the subjects' normal breath since the band is tightly tied on chest or abdomen which results in unnatural activity of the chest or the abdomen. So the proposed approach may have a better performance in practical use without the respiratory effort band as reference and be more comfortable. In addition, the algorithm entirely did not use the RGB data, so there is no risk to expose the individual privacy of users, such as appearance, characteristic, habit or other information except for the depth and the body silhouette, which makes it easier to be accepted than common RGB camera monitoring by the users.

The main application scenario of this system is indoor, and the effective range is close (0.2-1.5m), for cooperating the range, the system should be placed where people often appear. If only the non-human object appears in the front of the camera (for example a bookshelf), the depth data would not have obvious change, thus it can be distinguished from a human. In addition, the system is designed for designated area monitoring instead of wide monitoring, correct placement can avoid an extreme situation that the human and non-human object with same size occur simultaneously in the effective range.

The depth camera can capture slight changes of space depth caused by alternating respiratory movement. However, depth image does not have as many image features as the RGB images, such as texture, color and intensity, and the depth values in the images change inconspicuously. Besides, the depth image is not good at providing elaborate edges due to the principle of acquiring the depth data as shown in the Fig.10 and Fig.14. In addition, there is not a large amount of depth images to learn the image characteristics, neither a lot of labor resources to manually segment or label

the ROI of the depth images for providing the groundtruth. Those inconveniences make many popular methods difficult to be implemented, such as the machine learning approach. So the proposed algorithm utilizes morphological operation to extract the ROI, and calculate the RR. This process is similar to the approach that many medical staffs observe the movements of the chest, abdomen or face to determine the respiration condition by the naked eye, which has solid physical meaning. It is very easy to be understood and accepted by the medical staffs. Although the proposed method is relatively simple compared with the machine learning methods, it has achieved promising results in the RR measurement. However, the machine learning has outstanding effects in the field of image processing which has potential to generate more accurate ROI and respiratory waveform. So gradually establish the databases of depth images related with physiological signal monitoring to enable the machine learning to extract different parts of the ROI in the depth image, obtain more respiratory indicators such as tidal volume, will proceed in the future.

The experiment results demonstrated that all five postures and four external environments will lead to the promising accuracy of RR. The average accuracy of RR extracted from all epochs obtained in our study is 95.20% which showed a good performance for general application. In addition, the postures and the external environments didn't influence the performance of estimating RR from depth data via the Kruskal-Wallis test. It represented that the system has good adaptive capacity in different indoor applications, such as the sleep monitoring, home health management and community medical service. From the Bland-Altman analysis, it can be seen that the proposed system and the PSG device have a good agreement whose r^2 is 0.88. Besides, the subject's RR occupies the range of 10 to 23 per min, which shows that the proposed system can have promising performance and adaptability for a relatively wide respiratory range. Besides, the subjects were not requested to breathe at a fixed RR, however, metronome breathing respiratory maneuvers which

can test the performance of the system at different frequencies are also important. In future work, a metronome breathing experiment should be added.

However, there still has the difference between the standard thorax effort signal and the depth signal in this system on the accuracy and waveform. Except for system error, another factor to create the difference is that the respiratory region would generate involuntary mild movements caused by heartbeat and body shake which cannot be captured by thorax effort band or nasal airflow approach. In essence, the raw depth data may contain much other physiological information. Except for the RR, other features are likely to be extracted from the depth waveform, like the respiratory pattern or habit. So the system has potential to generate more accurate respiratory information and have more extensive applications, if we figure out what are the factors to make the depth data change.

In the process of collecting depth data, the request for the subject is to try not to have large body shake, but the experiment allows the subject to use the smartphone, the first reason is to make the subject not boring during the experiment, the second reason is that the use of a smartphone frequently causes the swings in the arm or head, which is close to the real scene and can further verify the system stability. In some preliminary tests, we attempted to collect data while the subject was stepping. The results showed that the body's front and rear swings were captured by the depth camera, which results in larger body motion artifacts while the PSG respiratory effort band was less affected. So the impact of body motion artifacts is huge for extracting the respiratory condition. Soleimani *et al.* [33] extracted the chest-surface respiratory pattern by performing a principal component analysis (PCA) on temporal 3-D geometrical features extracted from the chest and posterior shape models to eliminate the body motion artifacts. But the artifacts elimination which completely based on the original signals still requires further exploration in our study.

In addition to the terms to be improved as mentioned above, this study still has some limitations. For example, our experiments mainly aim at the young people test (age 22-29), and the experiment environment is the Home Lab. It is not conducted on patients with respiratory diseases or in practical environments such as in hospital. Because the experiment has a tedious procedure, which needs to conduct 60 trials for each subject. Furthermore, we did not control the age in the process of collecting volunteers which leads to the phenomenon that the self-enrolled subjects are in the age range of twenties who are highly active people. This study utilized the consistency and the accuracy of RR between the proposed method and PSG device. In fact, there are many other indicators in the respiratory medicine field that can be analyzed, such as tidal volume. In addition, the off-the-shelf commercial product SR300 was utilized to constitute the system, but the color camera in it didn't provide any information which is a resource waste. The future work will strive to overcome these limitations, such as conduct the clinical experiment on the patients, simplify the experimental process, expand the age range of the subjects, analyze more respiratory medical indicators and find a pure depth information extraction system.

V. CONCLUSION

This study proposed an unobtrusive and non-contact method for respiratory measurement based on the respiratory region detecting algorithm via depth images. By utilizing this method, the system can extract accurate RR from the depth waveform. The accuracy of RR has no significant statistical difference between five different postures and four kinds of external environments. The average accuracy of RR extracted from depth waveform for all 20 different conditions can achieve 95.20% and the proposed system had a good agreement ($r^2 = 0.88$) with the gold standard, the analysis results demonstrated that this system has promising performance, good adaptability and robustness, which is very suitable for home health care. Therefore, this study can provide a novel non-contact way to monitor the respiratory condition which is more convenient, simple, comfortable, accurate and unobtrusive.

REFERENCES

[1] K. E. Barrett *et al.*, *Ganong's Review of Medical Physiology*, vol. 23. New York, NY, USA: McGraw-Hill, 2009.

[2] W. Q. Lindh *et al.*, "*Delmar's Comprehensive Medical Assisting: Administrative and Clinical Competencies*. New York, NY, USA: Delmer, 2013.

[3] H. M. Engleman, R. N. Kingshott, S. E. Martin, and N. J. Douglas, "Cognitive function in the sleep apnea/hypopnea syndrome (SAHS)," *Sleep*, vol. 23, pp. 102–108, Jun. 2000.

[4] T. F. Murphy and S. Sethi, "Chronic obstructive pulmonary disease," *Drugs Aging*, vol. 19, no. 10, pp. 761–775, 2002.

[5] T. J. Matthews and M. F. MacDorman, "Infant mortality statistics from the 2010 period linked birth/infant death data set," *Nat. Vital Statist. Rep.*, vol. 62, no. 8, pp. 1–26, 2013.

[6] H. C. Kinney and B. T. Thach, "The sudden infant death syndrome," *New England J. Med.*, vol. 361, no. 8, pp. 795–805, 2009.

[7] G. K. Wolf and J. H. Arnold, "Noninvasive assessment of lung volume: Respiratory inductance plethysmography and electrical impedance tomography," *Crit. Care Med.*, vol. 33, no. 3, pp. S163–S169, 2005.

[8] J. H. Houtveen, P. F. Groot, and E. J. de Geus, "Validation of the thoracic impedance derived respiratory signal using multilevel analysis," *Int. J. Psychophysiol*, vol. 59, pp. 97–106, Feb. 2006.

[9] V. P. Seppa, J. Viik, and J. Hyttinen, "Assessment of pulmonary flow using impedance pneumography," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2277–2285, Sep. 2010.

[10] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ECGs," *Comput. Cardiol.*, vol. 12, pp. 113–116, 1985.

[11] A. de Groote, Y. Verbandt, M. Paiva, and P. Mathys, "Measurement of thoracoabdominal asynchrony: Importance of sensor sensitivity to cross section deformations," *J. Appl. Physiol*, vol. 88, no. 4, pp. 1295–1302, Apr. 2000.

[12] M. G. Gregório, M. Jacomelli, D. Inoue, P. R. Genta, A. C. de Figueiredo, and G. Lorenzi-Filho, "Comparison of full versus short induced-sleep polysomnography for the diagnosis of sleep apnea," *Laryngoscope*, vol. 121, no. 5, pp. 1098–1103, 2011.

[13] T. Wang, D. Zhang, Y. Zheng, T. Gu, X. Zhou, and B. Dorizzi, "C-FMCW based contactless respiration detection using acoustic signal," in *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, 2018, vol. 1, no. 4, p. 170.

[14] L. von Seidlein *et al.*, "Airflow attenuation and bed net utilization: Observations from Africa and Asia," *Malaria J.*, vol. 11, no. 1, p. 200, 2012.

[15] S. R. Sardesai, M. K. Kornacka, W. Walas, and R. Ramanathan, "Iatrogenic skin injury in the neonatal intensive care unit," *J. Maternal-Fetal Neonatal Med.*, vol. 24, no. 2, pp. 197–203, 2011.

[16] A. Singh, V. Lubecke, and O. Boric-Lubecke, "Pulse pressure monitoring through non-contact cardiac motion detection using 2.45 GHz microwave Doppler radar," in *Proc. Eng. Med. Biol. Soc.*, Boston, MA, USA, 2011, pp. 4336–4339.

[17] N. Celik, R. Gagarin, G. C. Huang, M. F. Iskander, and B. W. Berg, "Microwave stethoscope: Development and benchmarking of a vital signs sensor using computer-controlled phantoms and human studies," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 8, pp. 2341–2349, Aug. 2014.

[18] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, and S. Leonhardt, "Neonatal non-contact respiratory monitoring based on real-time infrared thermography," *Biomed. Eng. Online*, vol. 10, no. 1, p. 93, 2011.

[19] Y. Nakayama, G. Sun, S. Abe, and T. Matsui, "Non-contact measurement of respiratory and heart rates using a CMOS camera-equipped infrared camera for prompt infection screening at airport quarantine stations," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. (CIVEMSA)*, Shenzhen, China, Jun. 2015, pp. 1–4.

[20] G. G. Mora, J. M. Kortelainen, E. R. P. Hernández, M. Tenhunen, A. M. Bianchi, and M. O. Méndez, "Evaluation of pressure bed sensor for automatic SAHS screening," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 7, pp. 1935–1943, Jul. 2015.

[21] C. Sun, W. Li, and W. Chen, "A compressed sensing based method for reducing the sampling time of a high resolution pressure sensor array system," *Sensors*, vol. 17, no. 8, p. 1848, 2017.

[22] H. Uesawa, T. Takehara, and A. Ueno, "Non-contact measurements of diaphragm electromyogram, electrocardiogram and respiratory variations with sheet-type fabric electrodes for neonatal monitoring," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*, Las Vegas, NV, USA, Mar. 2018, pp. 25–28.

[23] B. A. Reyes, N. Reljin, Y. Kong, Y. Nam, and K. H. Chon, "Tidal volume and instantaneous respiration rate estimation using a volumetric surrogate signal acquired via a smartphone camera," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 764–777, May 2017.

[24] M. H. Li, A. Yadollahi, and B. Taati, "Noncontact vision-based cardiopulmonary monitoring in different sleeping positions," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1367–1375, Sep. 2017.

[25] R. Janssen, W. Wang, A. Moço, and G. de Haan, "Video-based respiration monitoring with automatic region of interest detection," *Physiol. Meas.*, vol. 37, no. 1, p. 100, 2016.

[26] U. Wijenayake and S.-Y. Park, "Real-time external respiratory motion measuring technique using an RGB-D camera and principal component analysis," *Sensors*, vol. 17, no. 8, p. 1840, 2017.

[27] M.-C. Yu, J.-L. Liou, S.-W. Kuo, M.-S. Lee, and Y.-P. Hung, "Noncontact respiratory measurement of volume change using depth camera," in *Proc. 34th Int. Conf. IEEE Eng. Med. Biol. Soc.*, San Diego, CA, USA, Aug./Sep. 2012, pp. 2371–2374.

[28] M. Alnowami, B. Alnwaimi, F. Tahavori, M. Copland, and K. Wells, "A quantitative assessment of using the Kinect for Xbox 360 for respiratory surface motion tracking," *Proc. SPIE*, vol. 8316, p. 83161T, Feb. 2012.

[29] F. Benetazzo, A. Freddi, A. Monteriù, and S. Longhi, "Respiratory rate detection algorithm based on RGB-D camera: Theoretical background and experimental results," *Healthcare Technol. Lett.*, vol. 1, no. 3, pp. 81–86, 2014.

[30] V. Soleimani *et al.*, "Depth-based whole body photoplethysmography in remote pulmonary function testing," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 6, pp. 1421–1431, Jun. 2018.

[31] V. Soleimani *et al.*, "Remote, depth-based lung function assessment," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1943–1958, Aug. 2017.

[32] S. Ostadabbas *et al.*, "A vision-based respiration monitoring system for passive airway resistance estimation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 9, pp. 1904–1913, Sep. 2016.

[33] V. Soleimani, M. Mirmehdi, D. Damen, and J. Dodd, "Markerless active trunk shape modelling for motion tolerant remote respiratory assessment," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 2077–2081.

[34] C. Sharp *et al.*, "Toward respiratory assessment using depth measurements from a time-of-flight sensor," *Frontiers Physiol.*, vol. 8, p. 65, Feb. 2017.

[35] R. Vandenhouten and O. Wendlandt, "High-precision optical position measurement in indoor environments," *Optik Photon.*, vol. 10, no. 3, pp. 24–27, 2015.

[36] C.-H. Wu, Y.-N. Sun, and C.-C. Chang, "Three-dimensional modeling from endoscopic video using geometric constraints via feature positioning," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 7, pp. 1199–1211, Jul. 2007.

[37] B. Jones *et al.*, "RoomAlive: Magical experiences enabled by scalable, adaptive projector-camera units," in *Proc. 27th Annu. ACM Symp. User Interface Softw. Technol.*, Honolulu, HI, USA, 2014, pp. 637–644.

[38] J. Li *et al.*, "Large depth-of-view portable three-dimensional laser scanner and its segmental calibration for robot vision," *Opt. Lasers Eng.*, vol. 45, no. 11, pp. 1077–1087, 2007.

[39] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of Kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, Feb. 2012.

[40] C. Zhang and Z. Zhang, "Calibration between depth and color sensors for commodity depth cameras," in *Computer Vision and Machine Learning With RGB-D Sensors*. Cham, Switzerland: Springer, 2014, pp. 47–64.

[41] T. Dutta, "Evaluation of the Kinect sensor for 3-D kinematic measurement in the workplace," *Appl. Ergonom.*, vol. 43, pp. 645–649, Jul. 2012.

[42] F. L. Siena, B. Byrom, P. Watts, and P. Breedon, "Utilising the intel realsense camera for measuring health outcomes in clinical research," *J. Med. Syst.*, vol. 42, no. 3, p. 53, 2018.

[43] A. Procházka, H. Charvátová, O. Vyšata, J. Kopal, and J. Chambers, "Breathing analysis using thermal and depth imaging camera video records," *Sensors*, vol. 17, no. 6, p. 1408, 2017.

[44] PrimeSensor. [Online]. Available: http://www.i3du.gr/pdf/primesense.pdf

[45] *Intel RealSense Camera SR300*. Accessed: Jun. 2016. [Online]. Available: https://software.intel.com/en-us/realsense/sr300

[46] M. Carfagni, R. Furferi, L. Governi, M. Servi, F. Uccheddu, and Y. Volpe, "On the performance of the Intel SR300 depth camera: Metrological and critical characterization," *IEEE Sensors J.*, vol. 17, no. 14, pp. 4508–4519, Jul. 2017.

[47] Intel Corporation, Santa Clara, CA, USA. (Jun. 2016). *Embedded Coded Light 3D Imaging System With Full High Definition Color Camera, Revision 1.0.* [Online]. Available: https://software.intel.com/sites/default/files/managed/0c/ec/realsense-sr300-product-datasheet-rev-1-0.pdf

[48] Y. Li, "Hand gesture recognition using Kinect," in *Proc. IEEE 3rd Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Beijing, China, Jun. 2012, pp. 196–199.

[49] S. R. Huddar *et al.*, "Novel algorithm for segmentation and automatic identification of pests on plants using image processing," in *Proc. IEEE 3rd Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Coimbatore, India, Jul. 2012, pp. 1–5.

[50] C. T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Comput.*, vol. C-100, no. 3, pp. 269–281, Mar. 1972.

[51] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2007, pp. 684–685.

[52] B. Hernández, G. Olague, R. Hammoud, L. Trujillo, and E. Romero, "Visual learning of texture descriptors for facial expression recognition in thermal imagery," *Comput. Vis. Image Understand.*, vol. 106, nos. 2–3, pp. 258–269, 2007.

[53] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, pp. 259–275, Jan. 2003.

[54] P. Mukhopadhyay and B. B. Chaudhuri, "A survey of Hough transform" *Pattern Recognit.*, vol. 48, no. 3, pp. 993–1010, 2015.

[55] T. Shimobaba, H. Nakayama, N. Masuda, and T. Ito, "Rapid calculation algorithm of Fresnel computer-generated-hologram using look-up table and wavefront-recording plane methods for three-dimensional display," *Opt. Express*, vol. 18, no. 19, pp. 19504–19509, 2010.

[56] X. Long, J. Foussier, P. Fonseca, R. Haakma, and R. M. Aarts, "Analyzing respiratory effort amplitude for automated sleep stage classification," *Biomed. Signal Process. Control*, vol. 14, pp. 197–205, Nov. 2014.

[57] M. D. Mckay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 14, no. 1, pp. 55–61, 2000.

[58] K. Y. Lin, D. Y. Chen, and W. J. Tsai, "Image-based motion-tolerant remote respiratory rate evaluation," *IEEE Sensors J.*, vol. 16, no. 9, pp. 3263–3271, May 2016.

[59] Y. Chan and R. P. Walmsley, "Learning and understanding the Kruskal–Wallis one-way analysis-of-variance-by-ranks test for differences among three or more independent groups," *Phys. Therapy*, vol. 77, no. 12, pp. 1755–1761, 1997.

[60] S. Siegel and N. J. Castellan, Jr., *Nonparametric Statistics for the Behavioral Sciences*, vol. 33, no. 1, 2nd ed. New York, NY, USA: McGraw-Hill, 1956, pp. 99–100.

CHENGLU SUN received the B.S. degree in mechatronic engineering from the Shanghai Normal University Tianhua College, Shanghai, in 2013, and the M.S. degree in mechanical manufacture and automation from the University of Shanghai for Science and Technology, Shanghai, in 2016. He is currently pursuing the Ph.D. degree with the Center for Intelligent Medical Electronics, Department of Electronic Engineering, School of Information Science and Technology, Fudan University, Shanghai, China. His research interests include signal processing, machine learning, and computer vision with an emphasis on unobtrusive physiological signal monitoring.
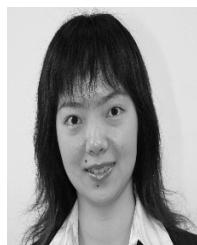
WEI LI received the B.S. degree from the Department of Electronic Engineering, Fudan University, Shanghai, China, in 2016, where he is currently pursuing the M.S. degree with the Center for Intelligent Medical Electronics, Department of Electronic Engineering. His research interests include physiological signal measurement and sleep monitoring, biomedical signal processing, sensor technology, and unobtrusive physiological measurement techniques.

CHEN CHEN received the M.S. degree in embedded system from the Institut Supérieur d'Electronique de Paris, in 2013, and the Ph.D. degree in computer science from Université Pierre et Marie Curie, in 2016. She is currently a Postdoctoral Researcher with the Centre of Intelligent Medical Electronics, School of Information Science and Technology, Fudan University. Her research interests include biomedical engineering, focusing on biomedical signal processing, wearable sensor systems, sleep analysis, and personalized health monitoring.

ZEYU WANG received the B.S. degree from the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China, in 2017. He is currently pursuing the M.S. degree with the Center for Intelligent Medical Electronics, Department of Electronic Engineering, Fudan University. His research interests include physiological signal measurement such as EEG, ECG and EMG signal. His research interests include biomedical signal processing, sensor technology, and unobtrusive physiological measurement techniques.

WEI CHEN received the B.S. and M.S. degrees from Xi'an Jiaotong University, in 1999 and 2002, respectively, and the Ph.D. degree from The University of Melbourne, in 2007. From 2008 to 2015, she was an Assistant Professor with the Department of Industrial Design, Technical University of Eindhoven. She is currently a Full Professor with the Department of Electronic Engineering, Fudan University. Her research interests include wearable sensor system, biomedical signal processing, and health monitoring. She is a Vice Chair of the IEEE Sensors and Systems Council. She is also an Associate Editor of the IEEE JOURNAL OF BIOMEDICAL HEALTH INFORMATICS. She is also a Regional Representative of the IEEE EMBS Technical Committee on Wearable Biomedical Sensors and Systems.

. . .