

A System And Method For Measuring Vital Body Signs

Field of the Invention

This invention relates to a system and method of estimating body vitals like Heart
5 Rate, Respiration Rate, Blood Oxygen Saturation and Blood Pressure of a person.
Specifically, this invention relates to a system and method of estimating body vitals like
Heart Rate, Respiration Rate, Blood Oxygen Saturation and Blood Pressure of a person
based on images from a colour camera and an active depth sensor simultaneously.

Prior Art

Photoplethysmography (PPG) is an optically obtained plethysmogram used to
detect blood volume changes in the microvascular bed of tissue. Remote
Photoplethysmography allows estimation of blood volume changes without skin contact by
using ambient light sources and video cameras. Under proper lighting conditions, minute
15 variations in skin colour and temperature due to blood volume changes can be observed.
This allows estimating blood volume changes without skin contact. However, due to the
use of ambient light sources, slight change in the surrounding would inevitably lead to
inaccurate results from the plethysmogram.

Therefore, those skilled in the art are striving to provide an improved method and
20 system of using PPG to remotely obtain vital signs of patients accurately.

Summary of the Invention

The above and other problems are solved and an advance in the art is made by a
system and method in accordance with this invention. A first advantage of the system and
25 method in accordance with this invention is that the system and method allows non-contact
measurement of vital signs. A second advantage of the system and method in accordance
with this invention is that the system and method is able to estimate vital signs within a

short period of time, at less than 5 seconds, with an accuracy of more than 90%. A third advantage of the system and method in accordance with this invention is that the system and method is independent of the ambient light. A fourth advantage of the system and method in accordance with this invention is that the system and method utilises devices
5 that are readily available. This allows remote monitoring of patient's vital signs with ease.

A first aspect of the disclosure describes a system for estimating vitals of a person. The system comprises an image sensor adapted to capture a sequence of color (RGB) and near infrared (NIR) images, a processing unit comprising a processor, memory and instructions stored on the memory and executable by the processor to: estimating vitals of
10 the person from the sequence of RGB and NIR images.

In an embodiment of the first aspect of the disclosure, the instruction to estimating vitals of the person from the sequence of RGB and NIR images comprises instructions to: time synchronising the sequence of RGB and NIR images; detecting region of interest within the sequence of RGB and NIR images; extracting raw signals from the detected
15 region of interest; extracting photoplethysmography (PPG) signals from the raw signals; and estimating vitals of the person from the extracted PPG signals.

In an embodiment of the first aspect of the disclosure, the instruction to detecting region of interest within the sequence of RGB and NIR images comprises instructions to: detecting a face of the person within each of the sequence of NIR images; in response to
20 detecting the face of the person, detecting at least one region of interest within the detected face of the person; and mapping the detected region of interest in each of the sequence of NIR images to the corresponding sequence of RGB images.

In an embodiment of the first aspect of the disclosure, the instruction to detecting at least one region of interest within the detected face of the person comprises instructions to:
25 applying facial landmark detection where 81 prominent feature points on the face the person are detected; and overlaying a box at a top of the detected 81 prominent feature points between facial edge and eyebrows, wherein the box is one region of interest.

In an embodiment of the first aspect of the disclosure, the instruction to detecting at least one region of interest within the detected face of the person further comprises instructions to: overlaying two boxes between facial edge and nose, wherein the two box are two other region of interests.

5 In an embodiment of the first aspect of the disclosure, the instruction to detecting region of interest within the sequence of RGB and NIR images comprises instructions to: computing an average pixel intensity of the detected region of interest for each of a plurality of frequency bands including Red, Green and Blue from the sequence of RGB images and NIR frequency band from the sequence of NIR images; and appending the average pixel
10 intensity value of each frequency band to the pixel intensity value of corresponding frequency band in respective RGB and NIR images to form a time series for each frequency band.

In an embodiment of the first aspect of the disclosure, the instruction to detecting region of interest within the sequence of RGB and NIR images further comprises
15 instructions to: determining if a length of the time series is equal to a predetermined length, L ; in response to the length of the time series being equal to the predetermined length, L , normalising the pixel intensity value in the time series for each frequency band.

In an embodiment of the first aspect of the disclosure, the instruction to extracting PPG signals from the raw signals comprises instructions to: applying a moving average
20 filter to the pixel intensity value in the time series for each frequency band, the moving average filter has a window length of $\alpha_1 * L$; up-sampling the pixel intensity value in the time series for each frequency band using cubic spline interpolation method to form a new length, L_2 , where $L_2 = \alpha_2 * L$; applying a detrending algorithm based on smoothness priors' formulation for the upsampled pixel intensity value in the time series for each frequency
25 band wherein a smoothness factor, λ , of the detrending algorithm is set based on the following expression, $\lambda = \alpha_3 * L_2$; and applying Independent Component Analysis (ICA) to extract PPG signal from the detrended pixel intensity value in the time series for each

frequency band, wherein values of $\alpha_1, \alpha_2, \alpha_3$ & L are calibrated based on the imaging sensor.

In an embodiment of the first aspect of the disclosure, the instruction to applying Independent Component Analysis (ICA) to extract PPG signal from the detrended pixel intensity value in the time series for each frequency band comprises instructions to: grouping the detrended pixel intensity value in the time series for each frequency band into a set of Red PPG signal and a set of NIR PPG signals wherein the Red PPG signal is derived from the frequency bands of Red, Green and Blue and the NIR PPG signal is derived from the frequency bands of NIR, Green and Blue; applying a band pass filter with cutoff frequency of between 0.16 and 2 Hz to the Red PPG signal; applying a band pass filter with cutoff frequency of between 0.16 and 2 Hz to the NIR PPG signal.

In an embodiment of the first aspect of the disclosure, the instruction to estimating vitals of the person from the extracted PPG signals comprises instructions to: computing Ratio (R) of time-domain range for NIR PPG signal and Red PPG signal wherein the Ratio (R) can be expressed with the following expression,

$$\text{Ratio (R)} = \frac{\text{Amplitude range of Red PPG signal} / \text{Arithmetic mean of Red PPG signal}}{\text{Amplitude range of NIR PPG signal} / \text{Arithmetic mean of NIR PPG signal}}$$

where amplitude range of Red PPG signal relates to difference between minimum and maximum values of the amplitude of the Red PPG signal in time domain, arithmetic mean of Red PPG signal relates to arithmetic mean of the Red PPG signal in between minimum and maximum values of the amplitude of the Red PPG signal in time domain, amplitude range of NIR PPG signal relates to difference between minimum and maximum values of the amplitude of the NIR PPG signal in time domain and arithmetic mean of NIR PPG signal relates to arithmetic mean of the NIR PPG signal in between minimum and maximum values of the amplitude of the NIR PPG signal in time domain; estimating the Blood Oxygen Saturation with the following expression,

$$\text{Blood Oxygen Saturation (SPO}_2\text{)} = \beta_1 + \beta_2 * R$$

where β_1 refers to off-set parameter and β_2 refers to slope parameter of the linear regression curve which maps the Ratio (R) to the blood oxygen saturation (SPO₂).

In an embodiment of the first aspect of the disclosure, the instruction to estimating vitals of the person from the extracted PPG signals comprises instructions to: detecting
 5 Systolic Peak (P_s), Diastolic peak (P_d), and global minima (P_m) from the time-domain NIR PPG signal; estimating the Systolic and Diastolic Blood Pressure with the following expressions,

$$\text{Systolic Blood Pressure (SBP)} = e^{(\gamma_1 * \log(HR) + \gamma_2 * \log(P_s - P_m) + \gamma_3)}$$

$$\text{Diastolic Blood Pressure (DBP)} = e^{(\gamma_4 * \log(HR) + \gamma_5 * \log(P_d - P_m) + \gamma_6)}$$

10 where $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6$ are parameters of the regression curve which maps the Heart Rate and NIR PPG signal amplitude features to the systolic and diastolic blood pressure, γ_1 and γ_4 are frequency-slope parameters for systolic and diastolic blood pressures respectively, γ_2 and γ_5 are amplitude-slope parameters for systolic and diastolic blood pressures respectively, γ_3 and γ_6 are off-set parameters for systolic and diastolic
 15 blood pressures respectively.

In an embodiment of the first aspect of the disclosure, the instruction to detecting Systolic Peak (P_s), Diastolic peak (P_d), and global minima (P_m) from the time-domain NIR PPG signal comprises instructions to: applying a peak detection algorithm on the NIR PPG signal to obtain a second order differential; identifying a global maxima in the NIR PPG
 20 signal as the Systolic Peak (P_s), a peak in the NIR PPG signal corresponding to a minima following the Systolic Peak (P_s) in the second order derivative as Diastolic peak (P_d), a maxima immediately after the systolic peak in the second order derivative as the Diatric notch, a minima of the entire NIR PPG signal as the global minima (P_m).

In an embodiment of the first aspect of the disclosure, the instruction to estimating
 25 vitals of the person from the extracted PPG signals comprises instructions to: applying fast fourier transform to the NIR PPG signal to transform the NIR PPG signal from time-domain to frequency domain; computing a plurality of local maxima of a power spectral density

(PSD) of the NIR PPG signal in frequency domain; in response to a local maxima between the frequency range of 0.8 and 2 Hz, estimating a Heart Rate (HR) with the following expression,

$$\text{Heart Rate (HR)} = 60 * f_{hr}$$

5 where f_{hr} refers to the local maxima that is between the frequency range of 0.8 and 2 Hz; and in response to a local maxima between the frequency range of 0.1 and 0.5 Hz, estimating a Respiration Rate (RR) with the following expression,

$$\text{Respiration Rate (RR)} = 60 * f_{rr}$$

 where f_{rr} refers to the local maxima that is between the frequency range of 0.1 and
10 0.5 Hz.

In an embodiment of the first aspect of the disclosure, the instruction to estimating vitals of the person from the extracted PPG signals comprises instruction to applying a machine learning model based on a linear regression model.

A second aspect of the disclosure describes a method for estimating vitals of a
15 person. The method comprises estimating vitals of the person from a sequence of RGB and NIR images.

In an embodiment of the second aspect of the disclosure, the step of estimating vitals of the person from the sequence of RGB and NIR images comprises: time synchronising the sequence of RGB and NIR images; detecting region of interest within the
20 sequence of RGB and NIR images; extracting raw signals from the detected region of interest; extracting photoplethysmography (PPG) signals from the raw signals; and estimating vitals of the person from the extracted PPG signals.

In an embodiment of the second aspect of the disclosure, the step of detecting region of interest within the sequence of RGB and NIR images comprises: detecting a face
25 of the person within each of the sequence of NIR images; in response to detecting the face of the person, detecting at least one region of interest within the detected face of the person;

and mapping the detected region of interest in each of the sequence of NIR images to the corresponding sequence of RGB images.

In an embodiment of the second aspect of the disclosure, the step of detecting at least one region of interest within the detected face of the person comprises instructions to:

- 5 applying facial landmark detection where 81 prominent feature points on the face the person are detected; and overlaying a box at a top of the detected 81 prominent feature points between facial edge and eyebrows, wherein the box is one region of interest.

In an embodiment of the second aspect of the disclosure, the step of detecting at least one region of interest within the detected face of the person further comprises

- 10 instructions to: overlaying two boxes between facial edge and nose, wherein the two box are two other region of interests.

In an embodiment of the second aspect of the disclosure, the step of detecting region of interest within the sequence of RGB and NIR images comprises instructions to:

- 15 computing an average pixel intensity of the detected region of interest for each of a plurality of frequency bands including Red, Green and Blue from the sequence of RGB images and NIR frequency band from the sequence of NIR images; and appending the average pixel intensity value of each frequency band to the pixel intensity value of corresponding frequency band in respective RGB and NIR images to form a time series for each frequency band.

20 In an embodiment of the second aspect of the disclosure, the step of detecting region of interest within the sequence of RGB and NIR images further comprises instructions to: determining if a length of the time series is equal to a predetermined length, L ; in response to the length of the time series being equal to the predetermined length, L , normalising the pixel intensity value in the time series for each frequency band.

25 In an embodiment of the second aspect of the disclosure, the step of extracting PPG signals from the raw signals comprises instructions to: applying a moving average filter to the pixel intensity value in the time series for each frequency band, the moving

average filter has a window length of $\alpha_1 * L$; up-sampling the pixel intensity value in the time series for each frequency band using cubic spline interpolation method to form a new length, L_2 , where $L_2 = \alpha_2 * L$; applying a detrending algorithm based on smoothness priors' formulation for the upsampled pixel intensity value in the time series for each frequency

5 band wherein a smoothness factor, λ , of the detrending algorithm is set based on the following expression, $\lambda = \alpha_3 * L_2$; and applying Independent Component Analysis (ICA) to extract PPG signal from the detrended pixel intensity value in the time series for each frequency band, wherein values of $\alpha_1, \alpha_2, \alpha_3$ & L are calibrated based on the imaging sensor.

10 In an embodiment of the second aspect of the disclosure, the step of applying Independent Component Analysis (ICA) to extract PPG signal from the detrended pixel intensity value in the time series for each frequency band comprises instructions to: grouping the detrended pixel intensity value in the time series for each frequency band into a set of Red PPG signal and a set of NIR PPG signals wherein the Red PPG signal is

15 derived from the frequency bands of Red, Green and Blue and the NIR PPG signal is derived from the frequency bands of NIR, Green and Blue; applying a band pass filter with cutoff frequency of between 0.16 and 2 Hz to the Red PPG signal; applying a band pass filter with cutoff frequency of between 0.16 and 2 Hz to the NIR PPG signal.

In an embodiment of the first aspect of the disclosure, the instruction to estimating

20 vitals of the person from the extracted PPG signals comprises instructions to: computing Ratio (R) of time-domain range for NIR PPG signal and Red PPG signal wherein the Ratio (R) can be expressed with the following expression,

$$\text{Ratio (R)} = \frac{\text{Amplitude range of Red PPG signal} / \text{Arithmetic mean of Red PPG signal}}{\text{Amplitude range of NIR PPG signal} / \text{Arithmetic mean of NIR PPG signal}}$$

where amplitude range of Red PPG signal relates to difference between minimum

25 and maximum values of the amplitude of the Red PPG signal in time domain, arithmetic

mean of Red PPG signal relates to arithmetic mean of the Red PPG signal in between minimum and maximum values of the amplitude of the Red PPG signal in time domain, amplitude range of NIR PPG signal relates to difference between minimum and maximum values of the amplitude of the NIR PPG signal in time domain and arithmetic mean of NIR PPG signal relates to arithmetic mean of the NIR PPG signal in between minimum and maximum values of the amplitude of the NIR PPG signal in time domain; estimating the Blood Oxygen Saturation with the following expression,

$$\text{Blood Oxygen Saturation (SPO}_2\text{)} = \beta_1 + \beta_2 * R$$

where β_1 refers to off-set parameter and β_2 refers to slope parameter of the linear regression curve which maps the Ratio (R) to the blood oxygen saturation (SPO₂).

In an embodiment of the second aspect of the disclosure, the step of estimating vitals of the person from the extracted PPG signals comprises instructions to: detecting Systolic Peak (P_s), Diastolic peak (P_d), and global minima (P_m) from the time-domain NIR PPG signal; estimating the Systolic and Diastolic Blood Pressure with the following expressions,

$$\text{Systolic Blood Pressure (SBP)} = e^{(\gamma_1 * \log(HR) + \gamma_2 * \log(P_s - P_m) + \gamma_3)}$$

$$\text{Diastolic Blood Pressure (DBP)} = e^{(\gamma_4 * \log(HR) + \gamma_5 * \log(P_d - P_m) + \gamma_6)}$$

where $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6$ are parameters of the regression curve which maps the Heart Rate and NIR PPG signal amplitude features to the systolic and diastolic blood pressure, γ_1 and γ_4 are frequency-slope parameters for systolic and diastolic blood pressures respectively, γ_2 and γ_5 are amplitude-slope parameters for systolic and diastolic blood pressures respectively, γ_3 and γ_6 are off-set parameters for systolic and diastolic blood pressures respectively.

In an embodiment of the second aspect of the disclosure, the step of detecting Systolic Peak (P_s), Diastolic peak (P_d), and global minima (P_m) from the time-domain NIR PPG signal comprises instructions to: applying a peak detection algorithm on the NIR PPG signal to obtain a second order differential; identifying a global maxima in the NIR PPG

signal as the Systolic Peak (P_s), a peak in the NIR PPG signal corresponding to a minima following the Systolic Peak (P_s) in the second order derivative as Diastolic peak (P_d), a maxima immediately after the systolic peak in the second order derivative as the Diatric notch, a minima of the entire NIR PPG signal as the global minima (P_m).

5 In an embodiment of the second aspect of the disclosure, the step of estimating vitals of the person from the extracted PPG signals comprises instructions to: applying fast fourier transform to the NIR PPG signal to transform the NIR PPG signal from time-domain to frequency domain; computing a plurality of local maxima of a power spectral density (PSD) of the NIR PPG signal in frequency domain; in response to a local maxima between
10 the frequency range of 0.8 and 2 Hz, estimating a Heart Rate (HR) with the following expression,

$$\text{Heart Rate (HR)} = 60 * f_{hr}$$

where f_{hr} refers to the local maxima that is between the frequency range of 0.8 and 2 Hz; and in response to a local maxima between the frequency range of 0.1 and 0.5 Hz,
15 estimating a Respiration Rate (RR) with the following expression,

$$\text{Respiration Rate (RR)} = 60 * f_{rr}$$

where f_{rr} refers to the local maxima that is between the frequency range of 0.1 and 0.5 Hz.

In an embodiment of the second aspect of the disclosure, the step of estimating
20 vitals of the person from the extracted PPG signals comprises applying a machine learning model based on a linear regression model.

Brief Description of the Drawings

The above and other features and advantages in accordance with this invention are
25 described in the following detailed description and are shown in the following drawings:

Figure 1 illustrating a system for estimating vitals of a person in accordance with an embodiment of this invention;

Figure 2 illustrating a process flow that is executable by a processing unit of the system in figure 1 in accordance with an embodiment of this invention;

Figure 3 illustrating a sample image of colour image and NIR image in accordance with an embodiment of this invention;

5 Figure 4 illustrating a process flow for extracting raw signal in accordance with an embodiment of this invention;

Figure 5 illustrating a process flow for extracting PPG signal in accordance with an embodiment of this invention;

10 Figure 6 illustrating a process flow for estimating blood oxygen saturation of a person from the extracted PPG signal in accordance of this invention;

Figure 7 illustrating a process flow for estimating Systolic and Diastolic Blood Pressure of a person from the extracted PPG signal in accordance of this invention;

Figure 8 illustrating a process flow for estimating Heart Rate and Respiration Rate of a person from the extracted PPG signal in accordance of this invention;

15 Figure 9 illustrating the NIR PPG signal and the second order differential of the NIR PPG signal in accordance of this invention;

Figure 10 illustrating a linear regression curve that can be used for estimation of SPO_2 from Ratio (R); and

20 Figure 11 illustrating a linear regression curve that can be used for estimation of blood pressure.

Detailed Description

This invention relates to a system and method of estimating body vitals like Heart Rate, Respiration Rate, Blood Oxygen Saturation and Blood Pressure of a person.

25 Specifically, this invention relates to a system and method of estimating body vitals like Heart Rate, Respiration Rate, Blood Oxygen Saturation and Blood Pressure of a person based on images from a Colour camera and depth sensor simultaneously.

Figure 1 illustrates a system 100 for estimating body vitals such as Heart Rate, Respiration Rate, Blood Oxygen Saturation and Blood Pressure of a person. The system 100 comprises a depth sensor 110, an image sensor 120 and a processing unit 130.

The depth sensor 110 is an active depth sensor that works in the near infrared (NIR) region. By using active depth sensor 110, we can eliminate error due to variations in ambient light. It is also observed that absorption of the light by Hemoglobin is lowest in the NIR region. Specifically, the skin cells and Hemoglobin in blood absorbs less NIR light (800-1000 nm) compared to visible light and hence the blood flow changes underneath the skin are sensed with greater contrast compared to visible light. Due to low absorption, the PPG signal will have the higher signal to noise ratio. Since the active depth sensor works in the NIR region, we can use active depth sensor 110 to shine NIR light on the skin of a person and measure its absorption to assist in estimating the vitals of the person.

The image sensor 120 is a typical colour camera for capturing sequence of images. Such image sensor 120 is commonly known in the art and a detailed description of the image sensor 120 would be omitted for brevity.

The processing unit 130 is a typical computing system that comprises a processor, memory and instructions stored on the memory and executable by the processor. The processor may be a processor, microprocessor, microcontroller, application specific integrated circuit, digital signal processor (DSP), programmable logic circuit, or other data processing device that executes instructions to perform the processes in accordance with the present invention. The processor has the capability to execute various applications that are stored in the memory. The memory may include read-only memory (ROM), random-access memory (RAM), electrically erasable programmable ROM (EEPROM), flash cards, or any storage medium. Instructions are computing codes, software applications that are stored on the memory and executable by the processor to perform the processes in accordance with this invention. Such computing system is well known in the art and hence only briefly described herein. The instructions can developed in C++ language (or any other

known programming language) and can be run on System on Chip (SoC) like Raspberry Pi or/and mobile devices like cell phones or tablet PCs.

Figure 2 illustrates a process flow 200 that is executed by the instructions of the processing unit to estimate vital signs of a person in accordance with this invention. To use
5 the system 100, both active depth sensor 110 and image sensor 120 would pointing in the same direction. Images captured by both sensors 110 and 120 would be transmitted to the processing unit 130.

Process 200 begins with step 205 where the images from the active depth sensor 110 and image sensor 120 are time synchronised. Essentially, there should not be any time
10 shift between the images from both sensors.

In step 210, process 200 detects the region of interest. The region of interest (RoI) includes, forehead, cheek and any other exposed skin. In one embodiment, when the system is arranged in a manner to capture the face of the person, the region of interest would include the forehead and cheek. Figure 3 illustrates an example of an image taken
15 from the active depth sensor 110 and image sensor 120. Specifically, the left image 310 is taken from the active depth sensor 110 and the right image 320 is taken from the image sensor 120. In step 210, the process first detects the face 311 and thereafter detects whether the person is wearing a mask. If the person is not wearing a mask, the process would detect the forehead region 312 and the cheek regions 313. If the person is wearing
20 a mask, the process would detect the forehead region 312 only. One method of detecting the RoI is by using facial landmark detection where 81 prominent feature points on the face are detected. The 81 prominent feature points include eyebrow, eye, nose, mouth and facial edge. Based on the 81 prominent detected points, the process would be able to determine if the person of interest is facing the camera (i.e. frontal, profile, or semi-frontal) and whether
25 the person is wearing a mask (i.e. mouth and part of nose and bottom facial edges would be obscured). Using the 81 prominent feature points detected, we can simply overlay a box on the top of the detected 81 prominent feature points between the facial edge and

eyebrows. If the person is not wearing a mask, another two boxes would be overlay at the cheeks which are between the facial edge and nose.

Based on the detected region, similar region would be marked on the image from the image sensor. The region of interest is computed for all the images from the active
5 depth sensor 110 and mapped to the images from image sensor 120 using the calibration. The RoI detection is executed on the captured images to form a set of time synchronised images. The set of time synchronised images is a length of no more than 5 seconds. In one embodiment, both active depth sensor 110 and image sensor 120 are calibrated using standard Tsai camera calibration technique to know the pixel mapping between them so
10 that when we detect the region of interest in one image, it can be mapped to the other image without additional computations. For example, if we detect forehead and cheek regions in the image from the active depth camera, we can map the regions corresponding to the forehead and cheek regions in the image from the image sensor by using the calibration.

15 In step 215, process 200 extracts raw signals within the detected region(s) of the images from both the active depth sensor 110 and image sensor 120. Further details will be described below with reference to figure 4.

In step 220, process 200 extracts PPG signal from the raw signals. Further details will be described below with reference to figure 5.

20 In step 225, process 200 estimates the vital signs including Respiration Rate (RR), Heart Rate (HR), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP) and Blood Oxygen Saturation (SPO₂). Further details will be described below with reference to figure 6.

Process 200 ends after step 225.

25 Figure 4 illustrates a process flow 400 that is executed by the instructions of the processing unit to extract raw signals within the detected region(s) of the images from both the active depth sensor of step 215 in process 200 in accordance with this invention. The

raw signals are pixel intensity values of each frequency band in the RGB images and NIR images. Process 400 begins with step 405 to compute the average pixel intensity of the detected region(s) for each of the frequency bands including Red, Green and Blue from the image of the image sensor 220 and NIR from the depth sensor 210. Each detected Region of Interest (RoI) is a rectangular matrix of the pixel intensity values as identified as 312 and 313 in Figure 3. For each of the frequency bands in the images (Red, Green, Blue and NIR), the arithmetic mean of all the pixels (average pixel intensity) within the region of interest 312 and 313 is computed in step 405. The average pixel intensity value for each frequency band computed in step 405 is then appended to the time series of the corresponding frequency band in step 410.

In step 410, process 400 appends the average pixel intensity value to form a time series corresponding to the frequency band. Specifically, process 400 appends the average pixel intensity value of each frequency band to the pixel intensity value of corresponding frequency band in respective RGB and NIR images to form a time series for each frequency band, the frequency band includes R, G B and NIR. The time series relate to the time synchronised video for estimating the vital signs. Each set of time series is no longer than 5 seconds of video clip. The average pixel intensity values are appended in the following manner:

1. For Red time series, the average pixel intensity values of the Red frequency band computed in step 405 is appended to the pixel intensity values of Red frequency band in each image from the video clip to form the red time series;

2. For Green time series, the average pixel intensity values of the Green frequency band computed in step 405 is appended to the pixel intensity values of Green frequency band in each image from the video clip to form the green time series;

3. For Blue time series, the average pixel intensity values of the Blue frequency band computed in step 405 is appended to the pixel intensity values of Blue frequency band in each image from the video clip to form the blue time series; and

4. For NIR time series, the average pixel intensity values of the NIR frequency band computed in step 405 is appended to the pixel intensity values of NIR frequency band in each image from the video clip to form the NIR time series.

At the end of the step 410, process 400 generates a table as illustrated below. Each column of Red, Blue, Green and NIR corresponds to a time series where the average pixel intensity value has been appended to respective frequency band. Below table only illustrates four time stamps, at 0 second, 0.036 second, 0.072 second and 0.108 second. One skilled in the art will recognise that more time stamps are required for each successful estimation of the vital signs. As mentioned previously, the time series should contain no more than 5 seconds worth of data. Preferably, if the camera frame capture rate is less than or equal to 30 FPS, the minimum would be 104 frames; and if camera frame capture rate is more than 30 FPS, the minimum would be 128 frames.

Time stamp (s)	Red	Blue	Green	NIR
0.000	94.1	120.8	96.6	100.9
0.036	90.1	116.8	103.6	99.9
0.072	92.2	118.6	99.4	101.3
0.108	91.0	121.4	101.3	98.6

In step 415, process 400 determines if the time series length is equal to a predetermined length, L . The predetermined length, L , is no more than 5 seconds, which is computed based on the frame rate of the image sensor. If the length of the time series is equal to the predetermined length, L , process 400 proceeds to step 425. If the time series is not equal to the predetermined length, L , process 400 repeats from step 205. Essentially, in order to accurately estimate the vital signs, the system requires a minimum length of data and this length is no more than 5 seconds. One skilled in the art will recognise that other conditions may be applied in step 415 without departing from the invention. For

example, step 415 may be modified to determine if the time series length is less than or not more than the predetermined length, L .

In step 425, process 400 normalises the pixel intensity value in the time series for each frequency band. Specifically, the pixel intensity value from each frequency band in the time series are normalised to restrict the amplitude to be within $[0, 1]$. One example of normalising the pixel intensity value in each frequency band in the time series is by subtracting the pixel intensity value in each frequency band by their arithmetic mean of the pixel intensity value and divided by their standard deviation the pixel intensity value.

Process 400 ends after step 425.

Figure 5 illustrates a process flow 500 that is executed by the instructions of the processing unit to extract PPG signal from the raw signals of step 220 in process 200 in accordance with this invention. Process 500 begins with step 505 by applying a moving average filter for each frequency band. Specifically, process 505 applies a moving average filter to the pixel intensity value in the time series for each frequency band. The moving average filter has a window length of $\alpha_1 * L$, to remove sudden amplitude changes in the raw signal caused due to movement of the person across the time series. The choice of α_1 and L are shown in the table below.

In step 510, process 500 up-samples the raw signal using cubic spline interpolation method to a new length, L_2 , where $L_2 = \alpha_2 * L$. Specifically, process 500 up-samples the pixel intensity value in the time series for each frequency band using cubic spline interpolation method to form the new length. Up-sampling would involve adding zero samples between the raw signal to increase the sampling rate. By using cubic spline interpolation method, the added zero samples would be reconstructed to non-zero samples which are typically approximate to adjacent samples of the raw signal. This increases the resolution of the respiration rate and heart rate and thus the accuracy of the estimate.

In step 515, process 500 applies a detrending algorithm to the up-sampled pixel intensity value in the time series for each frequency band. For example, an advanced

detrending algorithm based on smoothness priors' formulation is applied to remove noise in the up-sampled raw signal while preserving the frequency bands corresponding to respiration rate and heart rate. Further details of the detrending algorithm can be found in Tarvainen, M.P., Ranta-Aho, P.O. and Karjalainen, P.A., 2002. An advanced detrending method with application to HRV analysis. *IEEE Transactions on Biomedical Engineering*, 49(2), pp.172-175. The smoothness factor, λ , of the detrending algorithm is set based on the following expression,

$$\lambda = \alpha_3 * L_2$$

The above expression automates the tuning of the smoothness factor, λ , and eliminates the need for re-tuning the algorithm when the length of raw signal is varied.

In step 520, process 500 applies Independent Component Analysis (ICA) to extract PPG signal from the detrended raw signal (i.e. detrended pixel intensity value in the time series for each frequency band). The detrended raw signal is grouped into two sets based on the frequency bands to extract Red and NIR PPG signals using ICA. Specifically, the frequency bands of Red, Green and Blue are used to extract Red PPG signal in step 530 and the frequency bands of NIR, Green and Blue are used to extract NIR PPG signal in step 540. ICA is a computational method for separating a multivariate signal into additive subcomponents. It is a standard technique well known in the field of signal processing and hence details of ICA is omitted for brevity.

In step 535, a band pass filter with cutoff frequency between 0.16 and 2 Hz is applied to the Red PPG signal to remove noise. Similarly in step 535, a band pass filter with cutoff frequency between 0.16 and 2 Hz is applied to the NIR PPG signal to remove noise. The cutoff frequency is set to 0.16 and 2 Hz which covers the frequency range of average human respiration rate and heart rate. The band pass filter is a Finite Impulse Response (FIR) filter which is a standard time domain filtering technique. The cutoff frequency can be inversed to translate to time domain.

The values of α_1 , α_2 , α_3 & L are calibrated based on the imaging devices used to capture video. The imaging device's video capture frame rate plays a key role in the calibration process, and this is dependent on the device and the processing unit 130. These calibrated values are used to automate the up-sampling and detrending algorithms. Below

5 are two examples of known devices with the calibrated values.

Device	Length (L)	Achievable Frame Rate (FPS)	Moving Average Filter Length parameter (α_1)	Up-sampling length parameter (α_2)	Detrending window length parameter (α_3)
Surface Go	104	20-24	FPS/200	256/FPS	$\alpha_2 * L/128$
Raspberry Pi 4 + Intel Real Sense Camera	128	38-42	FPS/200	256/FPS	$\alpha_2 * L/128$

Process 500 ends after step 535 and 545.

Figure 6 illustrates a process flow 600 that is executed by the instructions of the processing unit to estimate Blood Oxygen Saturation (SPO₂) in step 225 in process 200 in accordance with this invention. Process 600 begins with step 605 by computing Ratio (R)

10 of time-domain range for NIR PPG signal and Red PPG signal. Ratio (R) can be computed with the following expression,

$$\text{Ratio (R)} = \frac{\text{Amplitude range of Red PPG signal} / \text{Arithmetic mean of Red PPG signal}}{\text{Amplitude range of NIR PPG signal} / \text{Arithmetic mean of NIR PPG signal}}$$

Where amplitude range of Red PPG signal refers to difference between minimum

15 and maximum values of the amplitude of the Red PPG signal in time domain, arithmetic mean of Red PPG signal refers to arithmetic mean of the Red PPG signal in between minimum and maximum values of the amplitude of the Red PPG signal in time domain, amplitude range of NIR PPG signal refers to difference between minimum and maximum values of the amplitude of the NIR PPG signal in time domain and arithmetic mean of NIR

PPG signal refers to arithmetic mean of the NIR PPG signal in between minimum and maximum values of the amplitude of the NIR PPG signal in time domain.

In step 610, process 600 estimates the Blood Oxygen Saturation with the following expression.

$$5 \quad \text{Blood Oxygen Saturation (SPO}_2\text{)} = \beta_1 + \beta_2 * R$$

Where β_1 refers to off-set parameter and β_2 refers to slope parameter of the linear regression curve which maps the Ratio (R) to the blood oxygen saturation (SPO₂). β_1 and β_2 are variables that can be derived from the linear regression curve. Figure 10 illustrates a linear regression curve that can be used for estimation of SPO₂ from Ratio (R).

10 Process 600 ends after step 610.

Figure 7 illustrates a process flow 700 that is executed by the instructions of the processing unit to estimate Systolic and Diastolic Blood Pressure in step 225 in process 200 in accordance with this invention. Process 700 begins with step 705 by detecting Systolic Peak (P_s), Diastolic peak (P_d), and global minima (P_m) from the time-domain NIR PPG signal. From figure 9, the Systolic peak times can be detected accurately from the blood volume pulse (BVP) waveform as they are maxima within the signal. Here BVP waveform refers to the NIR PPG signal. Maxima refers to global maxima which is the maxima of the entire NIR PPG signal. The amplitude of the global maxima is identified as the Systolic Peak (P_s) 910. From figure 9, the largest minima within the second order derivative correspond to the Systolic Peaks (P_s) 910 and the minima following these typically correspond to the diastolic peaks/inflections. Based on the location of the diastolic peak in the second order derivative, the amplitude of the corresponding point on the NIR PPG signal is identified as Diastolic peak (P_d) 930. The location corresponding to the Diastolic notch 920 is identified as the maxima immediately after the location identified as the systolic peak in the second order derivative. The minima of the entire NIR PPG signal would be the global minima (P_m) 940.

In step 610, process 600 estimates the Systolic and Diastolic Blood Pressure with the following expressions,

$$\text{Systolic Blood Pressure (SBP)} = e^{(\gamma_1 * \log(HR) + \gamma_2 * \log(P_s - P_m) + \gamma_3)}$$

$$\text{Diastolic Blood Pressure (DBP)} = e^{(\gamma_4 * \log(HR) + \gamma_5 * \log(P_d - P_m) + \gamma_6)}$$

5 Where $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6$ are parameters of the regression curve which maps the Heart Rate and NIR PPG signal amplitude features to the systolic and diastolic blood pressure, γ_1 and γ_4 are frequency-slope parameters for systolic and diastolic blood pressures respectively, γ_2 and γ_5 are amplitude-slope parameters for systolic and diastolic blood pressures respectively, γ_3 and γ_6 are off-set parameters for systolic and diastolic
10 blood pressures respectively. Figure 11 illustrates a linear regression curve that can be used for estimation of blood pressure. The log in figure 11 refers to logarithm base-10 and ln refers to logarithm base-e (natural log).

Process 700 ends after step 710.

Figure 8 illustrates a process flow 800 that is executed by the instructions of the
15 processing unit to estimate Heart Rate (HR) and Respiration Rate (RR) in step 225 in process 200 in accordance with this invention. Process 800 begins with step 805 by applying fast fourier transform to the NIR PPG signal to transform the NIR PPG signal from time-domain to frequency domain.

In step 810, process 800 compute the local maxima of power spectral density (PSD).
20 Specifically, PSD is computed by dividing the NIR PPG in frequency domain by the length of the NIR PPG signal (L). The local maxima of PSD is then computed by using any generic peak finder algorithms.

In step 820, process 800 determines the frequency range of the local maxima to use for computing the heart rate and respiration rate. For Heart Rate computations, the
25 local maxima within the frequency range of between 0.8 and 2 Hz is used. For Respiration Rate computations, the local maxima within the frequency range of between 0.1 and 0.5 Hz is used. Hence, process 800 proceeds to step 830 to estimate the Heart Rate (HR)

using the local maxima that is that is within the frequency range of between 0.8 and 2 Hz and to step 840 to estimate the Respiration Rate (RR) using the local maxima that is within the frequency range of between 0.1 and 0.5 Hz.

In step 830, process 800 estimates the Heart Rate (HR) with the following
5 expression,

$$\text{Heart Rate (HR)} = 60 * f_{hr}$$

Where f_{hr} refers to the local maxima that is within the frequency range of between 0.8 and 2 Hz.

In step 840, process 800 estimates the Respiration Rate (RR) with the following
10 expression,

$$\text{Respiration Rate (RR)} = 60 * f_{rr}$$

Where f_{rr} refers to the local maxima that is within the frequency range of between 0.1 and 0.5 Hz.

Process 800 ends after steps 830 and 840.

15 The Heart rate, Respiration rate and Blood Pressure are estimated using NIR PPG signal derived from data captured by active depth camera. These estimates do not vary with the ambient lighting conditions. The blood pressure model works on the premise that the systolic and diastolic peaks identified which correspond to the respective heart activities. This method gives greater accuracy compared to just using the global maxima in the model.
20 The estimation of Blood Oxygen Saturation is based on the ratio of range of Red PPG signal to range of NIR PPG signal.

In short, the method according to this invention involves using an active depth camera to capture simultaneous, time synchronized, video frames of both Color (RGB) and NIR wavelengths. From the video frames, a person's face is detected and the region
25 corresponding to their forehead is extracted (region of interest) and the pixel intensity values are averaged to form a time domain raw signal across multiple video frames. The time domain raw signal is detrended to remove higher order harmonics and resampled and

filtered using Independent Component Analysis (ICA) and FIR bandpass filter to obtain the PPG signal. Vitals such as Pulse rate, Respiration Rate, Blood Oxygen saturation and Blood Pressure are estimated from the PPG signal using frequency analysis using FFT for the former two vitals and mathematical models for the latter two vitals.

5 Processes 600-800 illustrate one embodiment of estimating Pulse rate, Respiration Rate, Blood Oxygen saturation and Blood Pressure based on the Red PPG signal and NIR PPG signal. Another embodiment of estimating Pulse rate, Respiration Rate, Blood Oxygen saturation and Blood Pressure based on the Red PPG signal and NIR PPG signal is the use of a machine learning model to estimate vitals of the person from the the Red PPG
10 signal and NIR PPG signal. One example of such machine learning model is the linear regression model. In another embodiment, processes 600-800 would be used for estimating Pulse rate, Respiration Rate, Blood Oxygen saturation and Blood Pressure based on the Red PPG signal and NIR PPG signal for at least a predetermined number of datasets is obtained in order to train the model for the machine learning model before the
15 machine learning model takes over from processes 600-800.

The above is a description of exemplary embodiments of a system and method in accordance with this disclosure. It is foreseeable that those skilled in the art can and will design alternative system and method based on this disclosure.