



US 20200178809A1

(19) **United States**(12) **Patent Application Publication**  
**WANG et al.**(10) **Pub. No.: US 2020/0178809 A1**(43) **Pub. Date: Jun. 11, 2020**(54) **DEVICE, SYSTEM AND METHOD FOR  
DETERMINING A PHYSIOLOGICAL  
PARAMETER OF A SUBJECT**(52) **U.S. CL.**  
CPC ..... *A61B 5/0077* (2013.01); *A61B 2576/00*  
(2013.01); *A61B 5/02416* (2013.01)(71) Applicant: **KONINKLIJKE PHILIPS N.V.**,  
Eindhoven (NL)(57) **ABSTRACT**(72) Inventors: **Wenjing WANG**, Utrecht (NL);  
**Albertus Cornelis DEN BRINKER**,  
Eindhoven (NL); **Gerard DE HAAN**,  
Eindhoven (NL)(21) Appl. No.: **16/637,284**(22) PCT Filed: **Aug. 1, 2018**(86) PCT No.: **PCT/EP2018/070845**

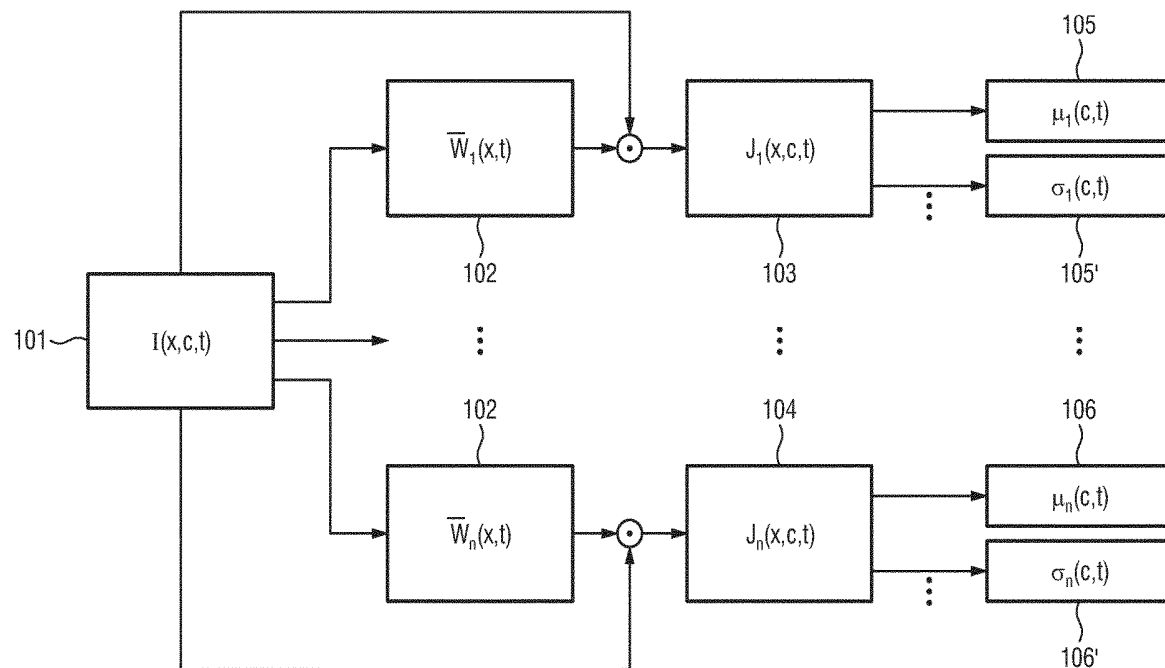
§ 371 (c)(1),

(2) Date: **Feb. 7, 2020**(30) **Foreign Application Priority Data**

Aug. 8, 2017 (EP) ..... 17185299.9

**Publication Classification**(51) **Int. Cl.**  
*A61B 5/00* (2006.01)  
*A61B 5/024* (2006.01)

The application relates to camera-based vital signs monitoring using remote PPG. A Device (10) for determining a physiological parameter comprises: an interface (11) for receiving image data comprising a time-sequence of image frames; a processor (12) for processing said data and configured to perform, for each frame, the steps of: generating a set of weighting maps comprising at least a first and second weighting map for spatially weighting pixels of the frame (102); determining a first weighted image frame by spatially weighting pixels of the image frame based on the first weighting map (103); determining a corresponding second weighted image frame; determining a first statistical parameter value based on the first weighted image frame (105); determining a corresponding second statistical parameter value; concatenating said first statistical parameter values over time based on the time-sequence of the corresponding image frames to obtain a first candidate signal (107); to obtain a corresponding second candidate signal (108); extracting a physiological parameter of the subject based on said first and/or said second candidate signal (109).



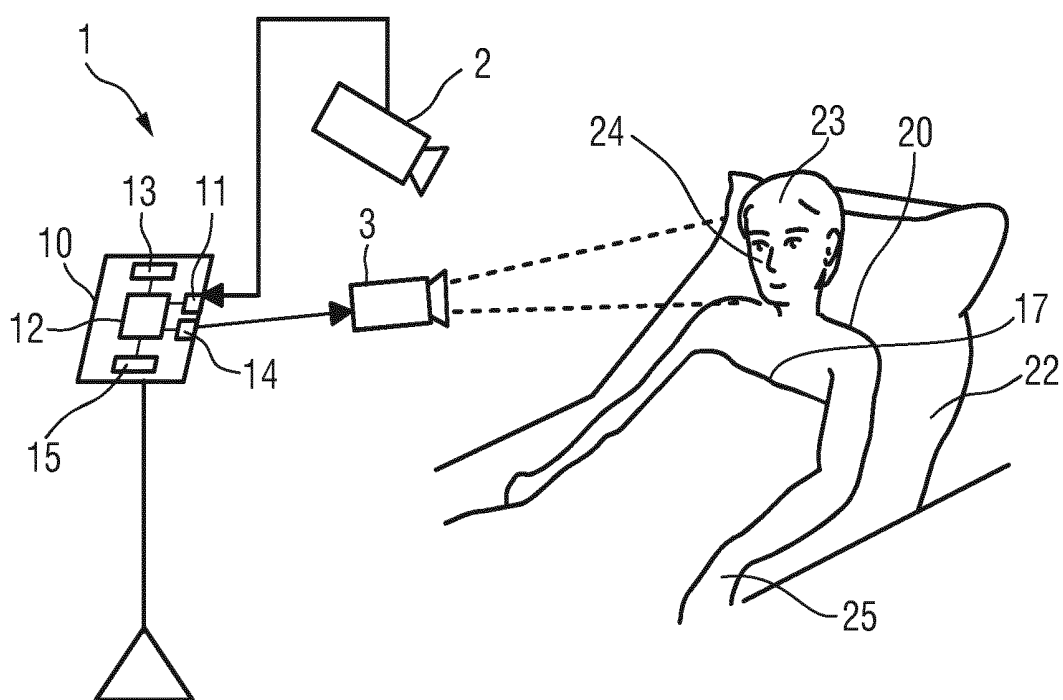


FIG.1

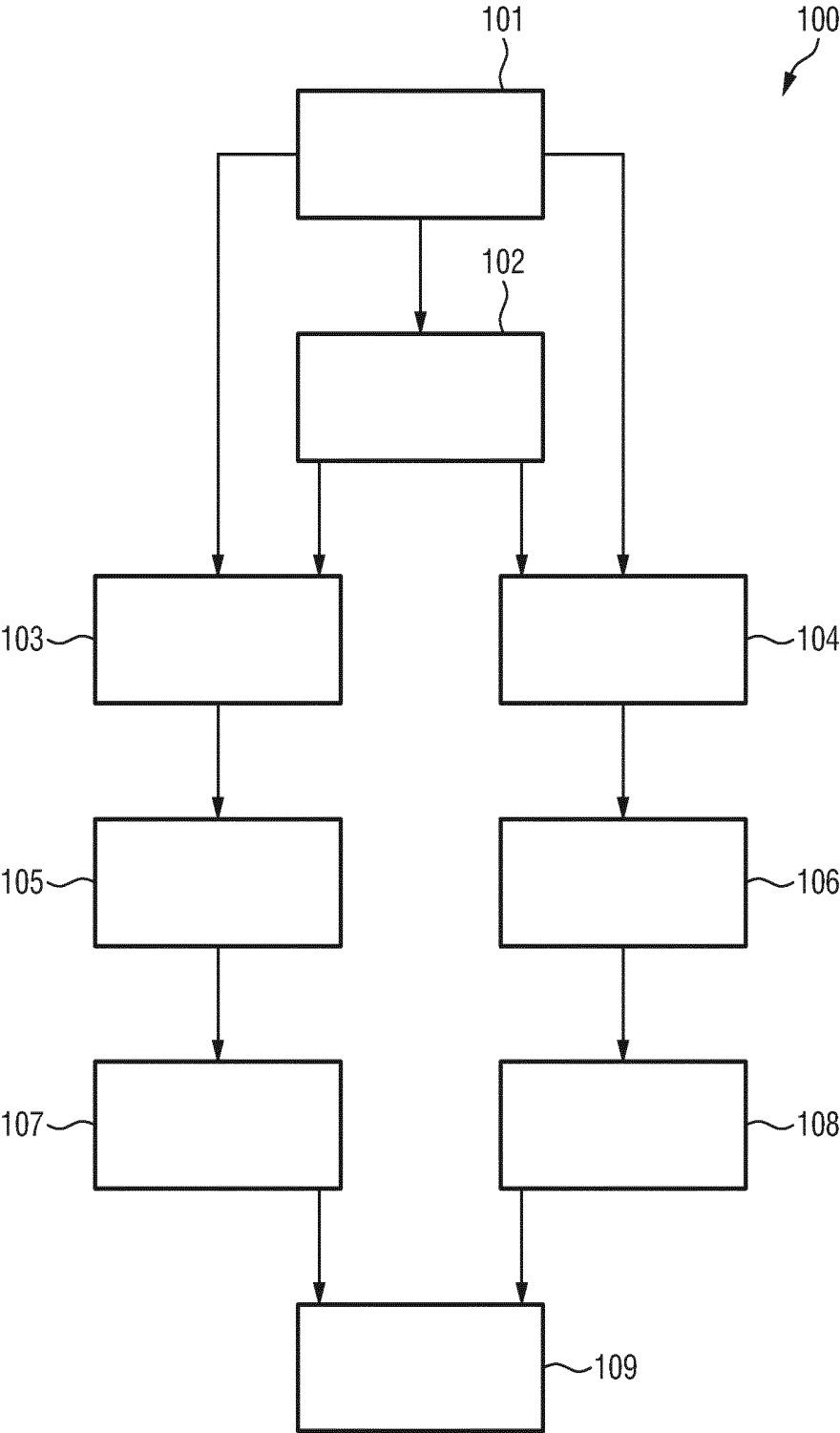


FIG.2

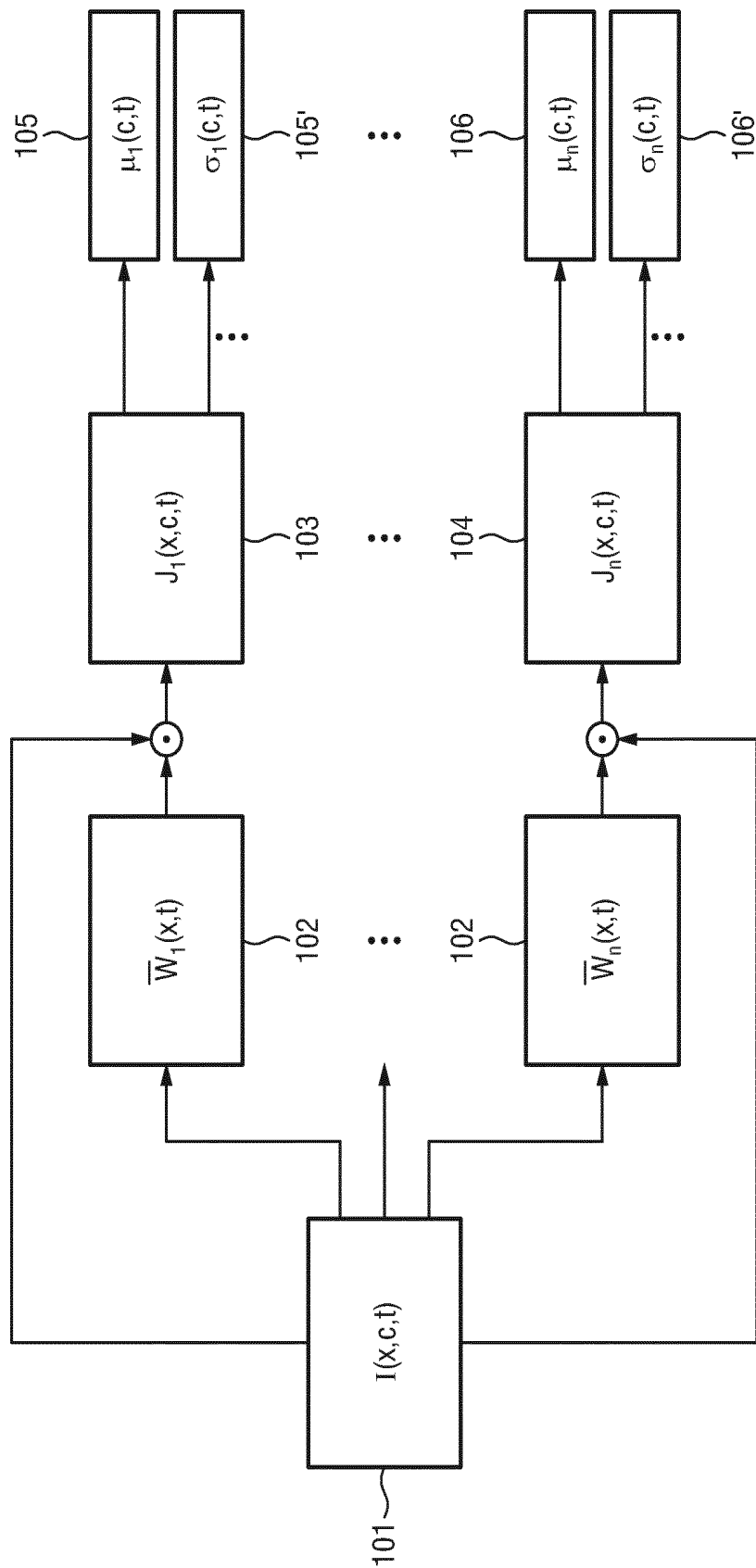


FIG.3

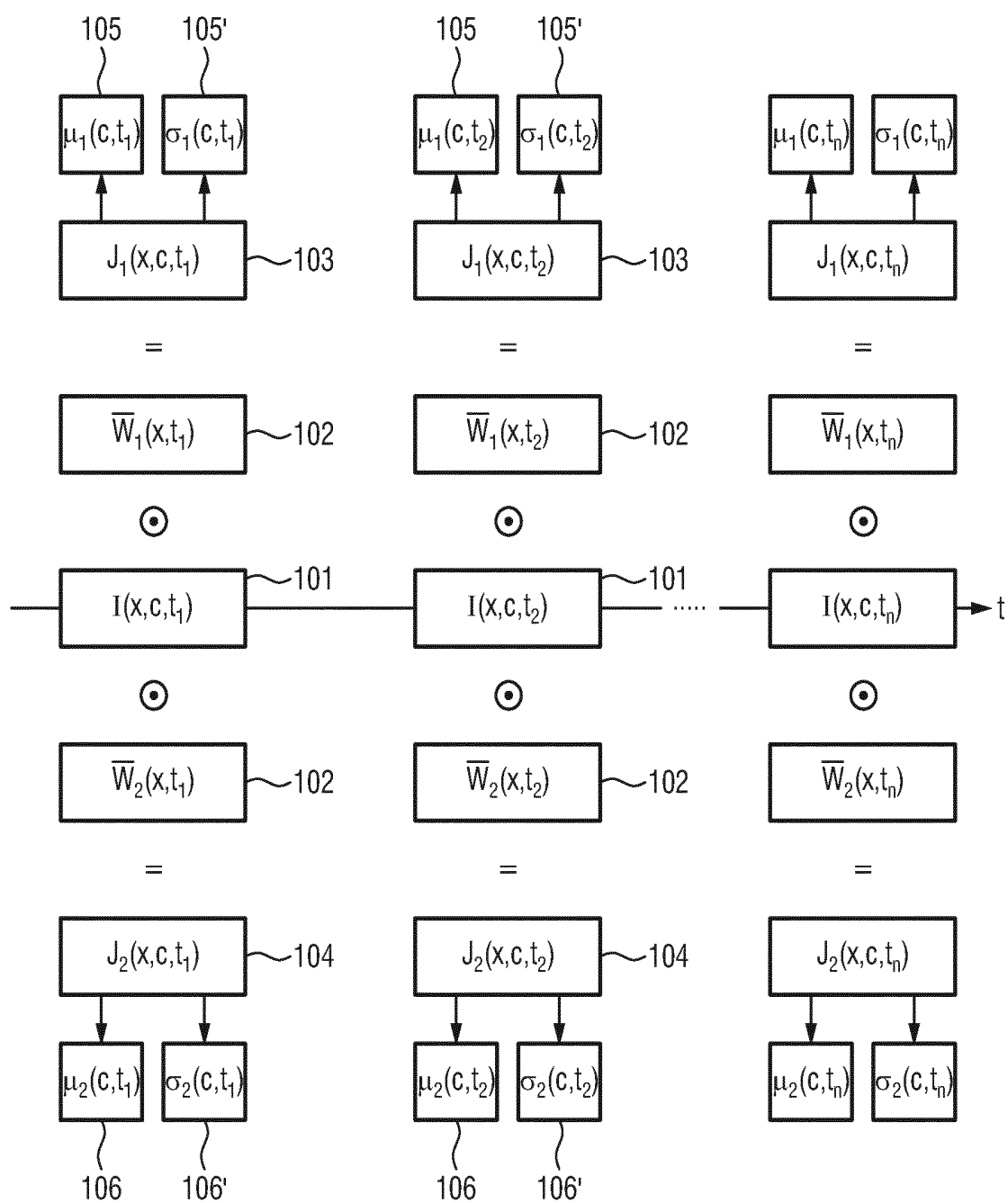


FIG.4

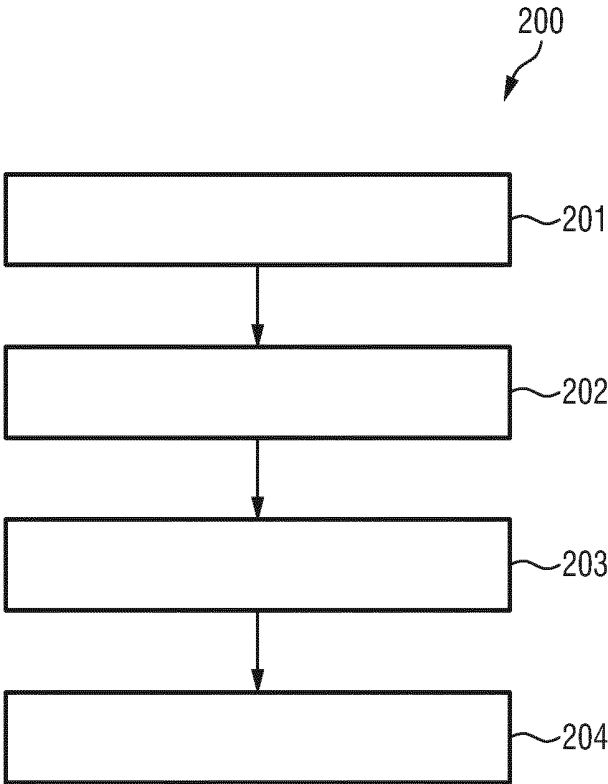


FIG.5

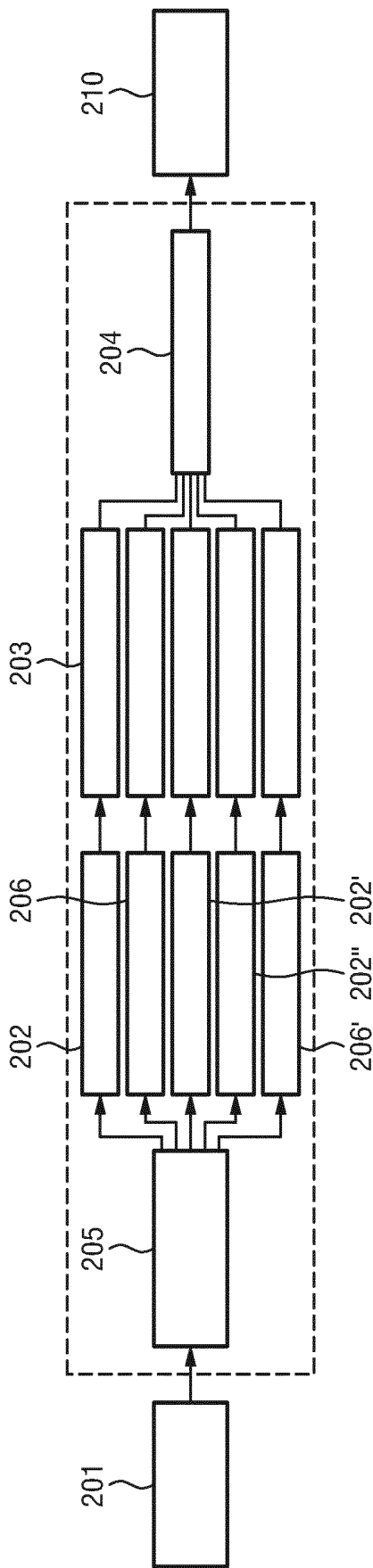


FIG.6

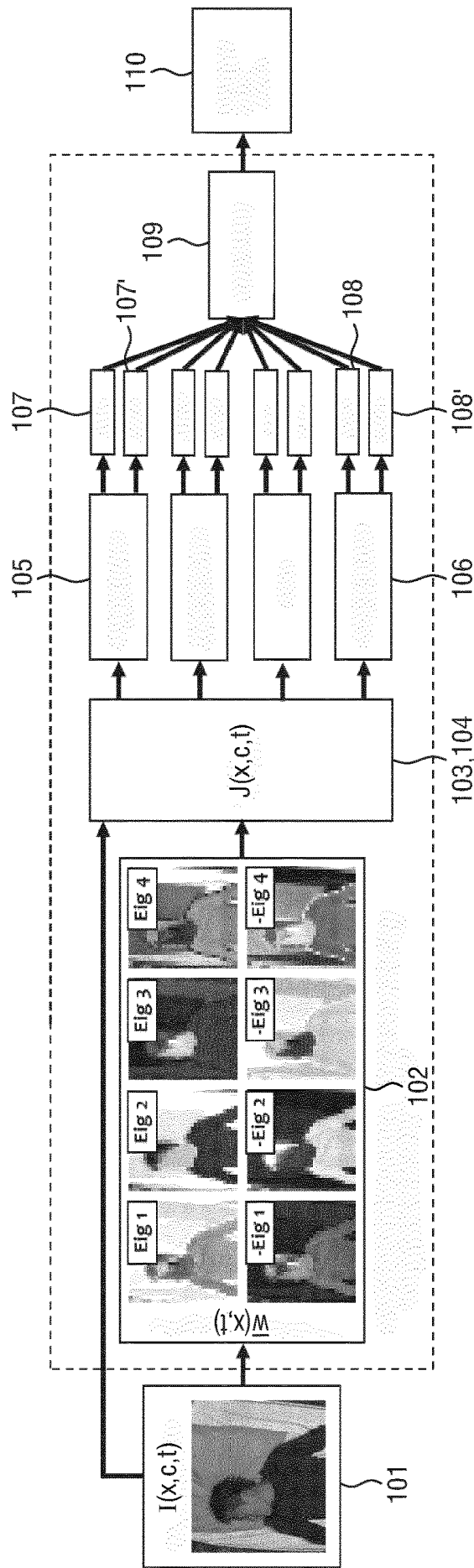


FIG.7



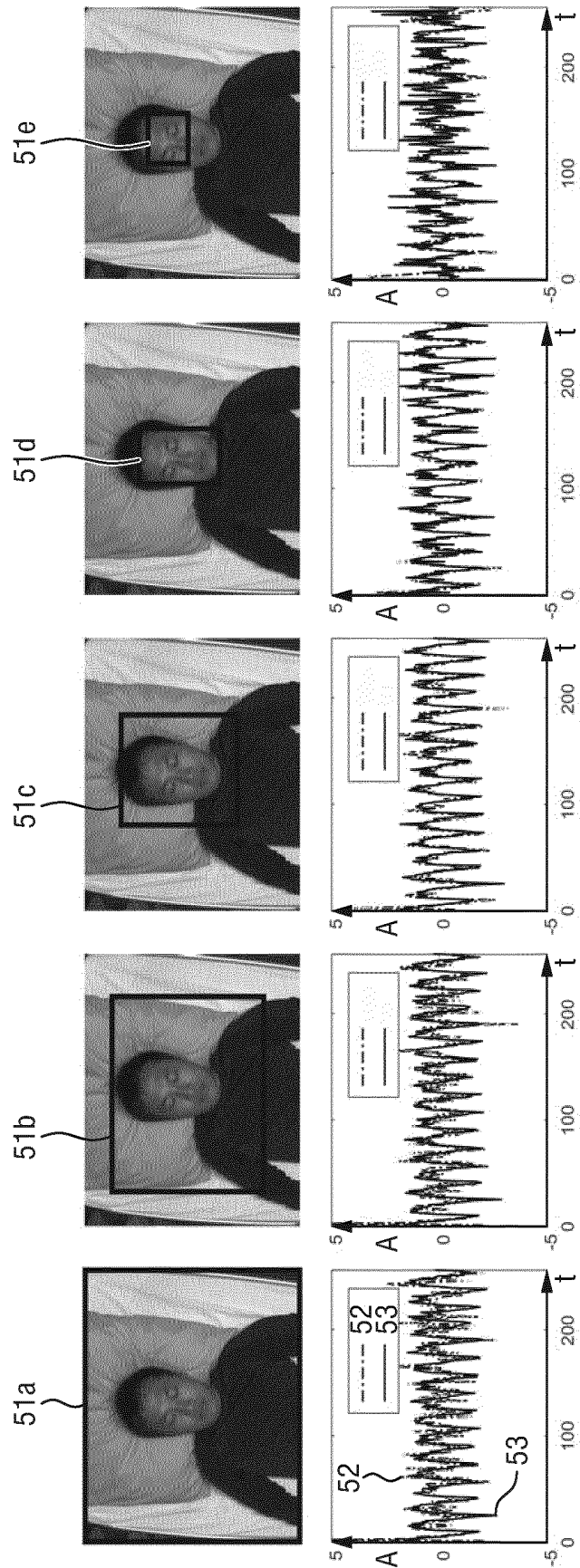


FIG.10

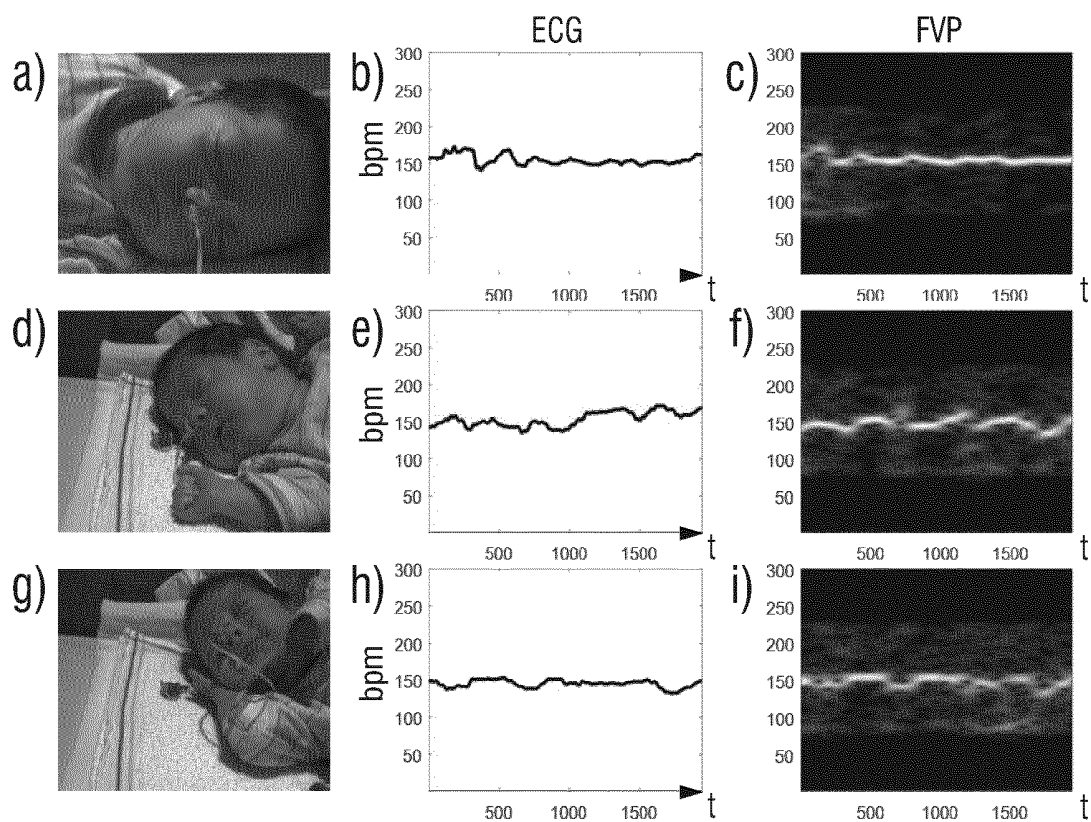


FIG.11

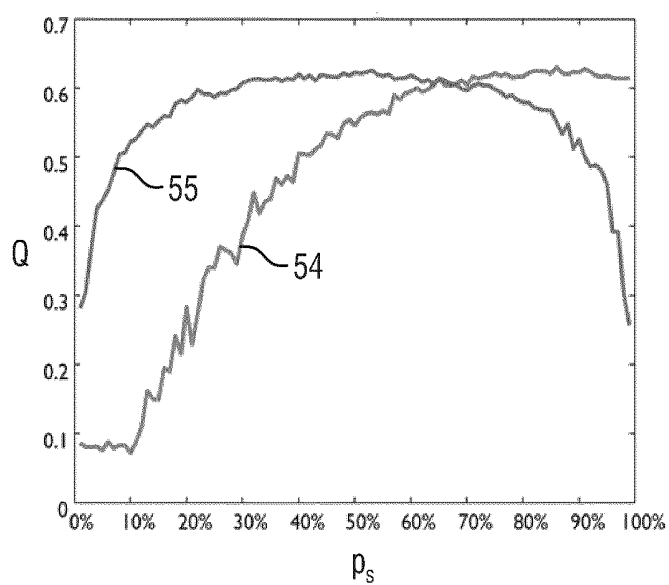


FIG.12

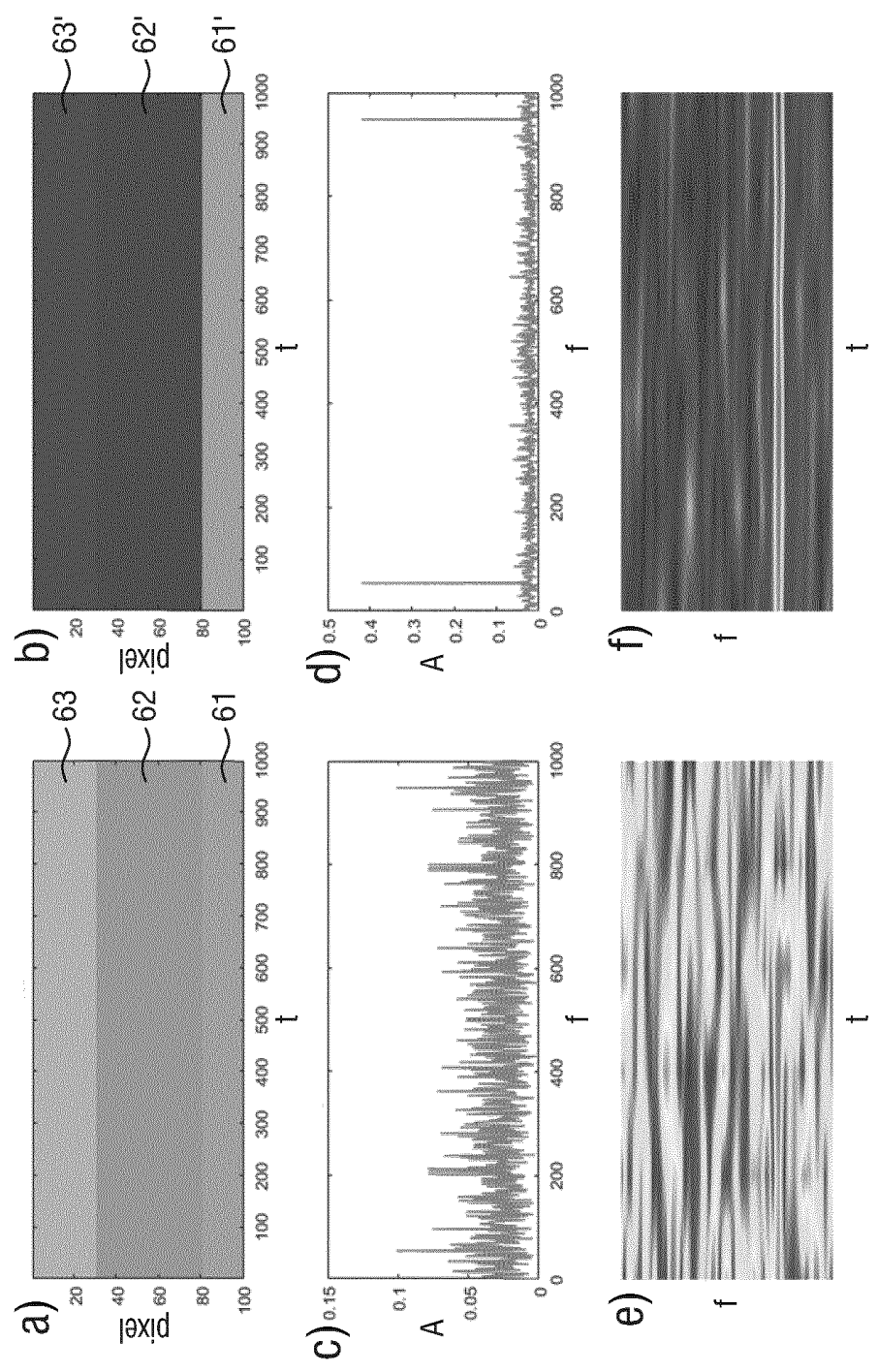


FIG.13

## DEVICE, SYSTEM AND METHOD FOR DETERMINING A PHYSIOLOGICAL PARAMETER OF A SUBJECT

### FIELD OF THE INVENTION

[0001] The present invention relates to the fields of medical technology and camera-based vital signs monitoring using remote photoplethysmography (rPPG). In particular, the present invention relates to a device and system for determining a physiological parameter of a subject. The present invention further relates to a corresponding method and a computer program for carrying out said method.

### BACKGROUND OF THE INVENTION

[0002] Vital signs of a person, for example the heart rate (HR), the respiration rate (RR) or the arterial blood oxygen saturation, serve as indicators of the current state of a person and as powerful predictors of serious medical events. For this reason, vital signs are extensively monitored in inpatient and outpatient care settings, at home or in further health, leisure and fitness settings.

[0003] One way of measuring vital signs is plethysmography. Plethysmography generally refers to the measurement of volume changes of an organ or a body part and in particular to the detection of volume changes due to a cardio-vascular pulse wave traveling through the body of a subject with every heartbeat.

[0004] Photoplethysmography (PPG) is an optical measurement technique that evaluates a time-variant change of light reflectance or transmission of an area or volume of interest. PPG is based on the principle that blood absorbs light more than surrounding tissue, so variations in blood volume with every heart beat affect transmission or reflectance correspondingly. Besides information about the heart rate, a PPG waveform can comprise information attributable to further physiological phenomena such as the respiration. By evaluating the transmittance and/or reflectivity at different wavelengths (typically red and infrared), the blood oxygen saturation can be determined.

[0005] Recently, non-contact, remote PPG (rPPG) devices (also called camera rPPG device herein) for unobtrusive measurements have been introduced. Remote PPG utilizes light sources or, in general radiation sources, disposed remotely from the subject of interest. Similarly, also a detector, e.g., a camera or a photo detector, can be disposed remotely from the subject of interest. Therefore, remote photoplethysmographic systems and devices are considered unobtrusive and well suited for medical as well as non-medical everyday applications. This technology particularly has distinct advantages for patients with extreme skin sensitivity requiring vital signs monitoring such as Neonatal Intensive Care Unit (NICU) patients with extremely fragile skin, premature babies, or patients with extensive burns.

[0006] Verkruysse et al., "Remote plethysmographic imaging using ambient light", *Optics Express*, 16(26), 22 Dec. 2008, pp. 21434-21445 demonstrates that photoplethysmographic signals can be measured remotely using ambient light and a conventional consumer level video camera, using red, green and blue (RGB) color channels.

[0007] A drawback of camera-based vital signs monitoring using remote photoplethysmography, is that often only limited region-of-interest (ROI) in the camera images provides valuable vital sign information. The region-of-interest

has to be selected in the image frames and tracked over time. For example, face detection and face tracking can be used to identify and track a region-of-interest such as the cheeks or forehead of a subject.

[0008] US 2015/0125051 A1 discloses a further improved remote PPG device wherein the computational effort for ROI selection is reduced. Instead of performing face detection for every single image frame, it is suggested that—after an initial face detection—an area which is to be chosen for vital signal extraction is moved to track the ROI by only estimating an image shift. Hence, by estimating the shift or displacement vector, it is no longer necessary to apply computationally expensive face detection for each subsequent image frame. However, a conventional face-detector may fail on side views of a face and thus provide inaccurate ROI selection.

[0009] WO 2017/093379 A1 discloses a device, system and method for determining vital sign information of a subject. To provide an increased signal quality and an improved robustness of the obtained vital sign information with respect to motion and low SNR, the proposed device tries to find the linear combination of the color channels, which suppresses the distortions best in a frequency band including the pulse rate, and consequently use this same linear combination to extract the desired vital sign information (e.g. represented by a vital sign information signal such as a respiration signal or Mayer waves) in a lower frequency band.

[0010] US 2016/343135 A1 discloses an apparatus for determining a pulse signal from a video sequence, the apparatus comprising a processing unit configured to obtain a video sequence, the video sequence comprising a plurality of image frames; form a plurality of video sub-sequences, each video sub-sequence comprising a frame segment from each image frame in a subset of the image frames, wherein each image frame is divided into a plurality of frame segments; for a first video sub-sequence formed from frame segments from a first subset of image frames, comparing a representative value for the first video sub-sequence to representative values for video sub-sequences formed from frame segments from a second subset of image frames; concatenate the first video sub-sequence to a second video sub-sequence formed from frame segments from the second subset of image frames based on the comparison of representative values; and determine a pulse signal from the concatenated video sub-sequences.

[0011] WO 2017/121834 A1 discloses a device, system and method for generating a photoplethysmographic image carrying vital sign information of a subject.

### SUMMARY OF THE INVENTION

[0012] It would be advantageous to provide an improved device, system and method for determining a physiological parameter of a subject. In particular, it would be advantageous to provide a device system and method for determining a physiological parameter of a subject that is robust regarding body motion and/or posture. In particular, it would be advantageous to overcome drawbacks regarding face detection and region-of-interest tracking.

[0013] To better address one or more of these concerns, in a first aspect of the present invention a device for determining a physiological parameter of a subject is presented, the device comprising:

[0014] an interface for receiving image data of a scene, said image data comprising a time-sequence of image frames; and

[0015] a processor for processing said image data, wherein the processor is configured to perform, for each of said image frames, the steps of:

[0016] generating a set of weighting maps comprising at least a first weighting map and a second weighting map for spatially weighting the pixels of the image frame;

[0017] determining a first weighted image frame by spatially weighting the pixels of the image frame based on the first weighting map;

[0018] determining a second weighted image frame by spatially weighting the pixels of the image frame based on the second weighting map;

[0019] determining a first statistical parameter value based on the first weighted image frame;

[0020] determining a second statistical parameter value based on the second weighted image frame;

[0021] wherein the processor is further configured to perform the steps of:

[0022] concatenating said first statistical parameter values over time based on the time-sequence of the corresponding image frames to obtain a first candidate signal;

[0023] concatenating said second statistical parameter values over time based on the time-sequence of the corresponding image frames to obtain a second candidate signal;

[0024] extracting a physiological parameter of the subject based on said first and/or said second candidate signal.

[0025] In a further aspect of the present invention a system for determining a physiological parameter of a subject is presented, the system comprising:

[0026] an imaging unit configured to acquire image data of a scene, said image data comprising a time-sequence of image frames; and

[0027] a device for determining a physiological parameter of a subject as described above based on the acquired image data.

[0028] In yet further aspects of the present invention, there are provided a corresponding method, a computer program which comprises program code means for causing a computer to perform the steps of the method disclosed herein when said computer program is carried out on a computer as well as a non-transitory computer-readable recording medium that stores therein a computer program product, which, when executed by a processor, causes the method disclosed herein to be performed.

[0029] Preferred embodiments of the invention are defined in the dependent claims. It shall be understood that the claimed method, system, computer program and medium can have similar and/or identical preferred embodiments as the claimed device, in particular as defined in the dependent claims and as disclosed herein.

[0030] The present invention is based on the idea to use a set of different weighting maps to replace the conventional approach of ROI detection and to eliminate the need for motion tracking of the ROI. With the use of different weighting maps, different spatially weighted versions of the image data (e.g. of an input video signal) can be computed (optionally in parallel) from which a number of candidate signals can be extracted directly, based on a statistical parameter, i.e., without further pixel selection. Generating a set of weighting maps thus comprises generating a first weighting map for spatially weighting the pixels of the

image frame and generating a second (different) weighting map for spatially weighting the same pixels of the image frame. The candidate signals are determined by concatenating statistical parameter values indicative of the respective weighted image frames. The statistical parameter values are concatenated over time based on the time-sequence of the corresponding image frames.

[0031] Hence, a set of different weighting maps comprising at least a first and a second weighting map are computed from the input image data or input video (e.g. focusing on different color clusters). Next the input video can be weighted with all these maps, wherein the individual image frames are weighted by their corresponding weighting maps. A weighting map indicates how much weight should be given to a particular pixel or region of the corresponding image frame. Thereby, the importance of the respective pixels or regions is set for the subsequent processing steps. For each weighted image one or more statistical parameters (e.g. mean, standard deviation or variance) can be computed. Hence, each weighted image can be condensed into a statistical parameter value per weighted image frame. Thereafter, these statistical parameters of successive frames can be concatenated. For example, a sequence of the mean pixel values per image frame can be generated. Thereby, a sequence of the first statistical parameter values (a value for each of the first weighted images) forms the first candidate signal and a sequence of the second statistical parameter values (a value for each of the second weighted images) forms the second candidate signal. In other words, the candidate pulse signal can be extracted for each of the weighted videos. Finally, an optional selection or combination amongst the candidate signals can be made and the physiological parameter value such as a pulse of the subject can be extracted therefrom e.g. using known (pulse) extraction techniques.

[0032] Hence, one core aspect is that different weighting maps are defined and it can be decided afterwards which map(s) and corresponding candidate signals are most suitable to extract the physiological parameter of the subject from, i.e., without prior knowledge. In particular, no identification or tracking of facial features is required and also non-facial skin areas can contribute valuable information.

[0033] In the following, some terms which are used throughout the application, shall be briefly explained and defined:

[0034] As used herein, a physiological parameter of the subject can refer to a physiological parameter indicative of a vital sign of the subject, such as a pulse, respiration rate or blood oxygen saturation of the subject.

[0035] As used herein, the image data of the scene can in particular refer to video data acquired by a (RGB) video camera. The images can represent a scene comprising at least portions of the subject's skin. The image data can refer to at least some of the image frames forming a video signal such as an RGB video signal, monochrome video signal, range imaging data or IR (infrared) video, optionally comprising one or more channels, in particular channels corresponding to different wavelengths.

[0036] As used herein, a statistical parameter value can refer to a statistical measure indicative of values of pixels of an image frame. For example, the statistical parameter value can refer to a central tendency such as a mean or median value of the pixels of an image frame, or a parameter indicative of a statistical dispersion such as standard deviation.

tion, variance, quartile range or other parameter. Each of the candidate signals can be determined by concatenating statistical parameter values of the corresponding image frames over time. For example, for each of the first weighted image frames a mean value of the pixel values of this weighted image frame is determined and the mean values of subsequent image frames form the candidate signal over time. The same applies to the sequence of a second weighted image frames. The first statistical parameter value and the second statistical parameter value can be obtained using the same or a different statistical operation, for example taking the mean of the weighted image frames. Hence, the 'first' statistical parameter indicates that it refers to the sequence of first weighted image frames, whereas the 'second' statistical parameter values indicate that they correspond to the second weighted image frames.

**[0037]** The extraction of the physiological parameter of the subject based on the first and/or second candidate signal can be performed using known techniques such as evaluating a fixed weighted sum over candidate signals of different wavelength channels (RGB, IR), blind source separation techniques advantageously involving both candidate signals such as blind source separation (based on selecting the most pulsatile independent signal), principal component analysis (PCA), independent component analysis (ICA), CHROM (chrominance-based pulse extraction), POS (wherein the pulse is extracted in a plane orthogonal to a skin-tone vector), PBV method (which uses a predetermined signature of a blood volume pulse vector), or APBV (an adaptive version of the PBV-method, also allowing blood oxygen saturation monitoring).

**[0038]** The processor can optionally be configured to perform additional steps such as normalizing the candidate signals, e.g. by dividing them by their (moving window) temporal mean, taking their logarithm and/or removing an offset. Moreover, the processor can optionally perform image (pre-)processing steps such as resizing, rescaling, resampling, or cropping of the (weighted) image frames. A (weighted) image frame as used herein may thus also refer to such a (pre-)processed (weighted) image frame. Optionally, the solution proposed herein of using a set of weighting maps can also be applied to only a portion of the image frame. Optionally, said portion of the image may be identified by a (coarse) region-of-interest selection and/or tracking (pre-)processing step. An advantage of this combination is that the accuracy and/or robustness can be further improved.

**[0039]** According to an embodiment, the processor is configured to generate the first and second weighting map for spatially weighting the pixels of the image frame (102) such that a non-zero weight is assigned to at least one same pixel of the image frame with both the first and the second weighting map. In consequence, at least one same pixel of the first and the second weighted image frame may have a non-zero value. In other words, the processor is configured to determine a set of weighting maps wherein at least a non-zero portion of the first weighting map and a non-zero portion of the second weighting map overlap. Hence, different weights can be assigned to the same pixel or plurality of pixels. Based thereon, the respective first and second weighted image frames, first and second statistical parameter values as well as first and second candidate signals can be obtained. Optionally, the processor is configured to generate the first and second weighting map for spatially

weighting the pixels of the image frame (102) such that a non-zero weight is assigned to at least 5%, at least 10% or at least 20% of the pixels of the image frame.

**[0040]** According to an embodiment, the processor can be configured to determine said set of weighting maps for each image based on a local property of the image. In other words, the weighting maps are preferably not indicative of a temporal property regarding a change from frame to frame but refer to a property of the corresponding image frame indicative of a spatial or spatially varying feature. A weighting map can thus be determined based on the content of the corresponding image frame.

**[0041]** In a refinement, said local property of the image can comprises at least one of a brightness, color, texture, depth and/or temperature. A color can refer to a relative value of a pixel in different wavelength channels. For example, a color may refer to a pixel value in RGB (red, blue, green) representation but may also refer to at least two channels indicative of different wavelength ranges such as different (near) infrared wavelength intervals.

**[0042]** Optionally, the processor can be configured to determine said weighting maps based on a similarity of pixel values of the image frame to a target category. Said target category can be derived based on the content of the image frame, in particular based on spectral clustering. For example, colors in the image can be fed to some form of clustering, for example, adapted to determine principal components from a similarity matrix, e.g. using K-means. Clustering can be based on one or more properties such as color, texture, temperature and/or depth. An advantage of clustering is that a target category can be selected automatically. Alternatively, one or more predetermined target categories can be used such as commonly occurring skin colors or components of skin colors that are different from a background such as hospital bedding. It should be highlighted that previous knowledge of the skin type of the subject is not necessarily required. Instead, different weighting maps can be generated to obtain corresponding candidate signals (without prior knowledge) and the physiological parameter can be extracted based thereon.

**[0043]** Optionally, the processor can be configured to determine said weighting maps based on an affinity matrix indicative of spectral clustering. An advantage of this approach is that an affinity matrix can be generated based on the image content, in particular without requiring external input or any predetermined target category.

**[0044]** Optionally, the processor can be configured to generate normalized weighting maps. An advantage of normalized weighting maps is that normalized weighting maps can avoid variations of an average value over time caused by the weighting. Such variations could induce noise or modulations in the physiological parameter that is to be extracted from the image data. For example, each weighting map may be normalized by dividing individual pixel values or weights by a sum or average over all weights or pixel values in that weighting map.

**[0045]** Optionally, the step of generating said set of weighting maps comprises at least one of zeroing weights below a predetermined threshold, offset removal and/or thresholding. An advantage of this embodiment is that the weighting process can be simplified. For example, negative weights can be prevented, an offset can be removed, e.g. by subtracting a smallest weight or an average. Thresholding can be used to reduce the number weighting levels of the

weighting map. Zeroing smaller weights can be used to suppress values below a predetermined threshold.

**[0046]** Optionally, the step of extracting the physiological parameter of the subject can further comprise the step of selecting at least one of said first and said second candidate signals based on a quality metric. For example, the candidate signal having the higher signal-to-noise ratio (SNR) may form the basis for extracting the physiological parameter of the subject. In addition or in the alternative, a flatness of a spectrum, a height of a spectral peak in a (normalized) spectrum or an entropy of the spectra of the first and second candidate signals may be evaluated. The evaluation can include a comparison with a predetermined threshold.

**[0047]** Optionally, the processor can be further configured to apply a blind source separation (BSS) technique to the candidate signals to obtain independent signals and to select at least one of said independent signals based on a quality metric. Examples of blind source separation techniques are principal component analysis (PCA) and independent component analysis (ICA). Quality metrics can be used as described above.

**[0048]** Optionally, the processor can be configured to combine at least two of the candidate signals, in particular in the frequency domain in particular wherein different relative weights may be given to individual spectral components of said candidate signals, in particular to combine different spectral components of said candidate signals in individually optimal ways. Various options exist to determine the optimal weights per spectral components, e.g. they may be based on the amplitude of the spectral components. Different relative weights may be given to individual spectral components of said candidate signals. Optionally, a combination signal may be synthesized based on the candidate signals. In an embodiment, blind-source-separation may work on complete candidate signals in the time domain (i.e. providing a linear combination), whereas the spectral weighting may work on the spectral components in the frequency domain (thereby providing a non-linear combination). In other words, the processor can be configured to combine at least two of the candidate signals by means of a linear combination or a non-linear combination to obtain a combination signal, wherein the processor is further configured to extract the physiological parameter of the subject based on the combination signal. An advantage of such a combination is that the use of the information content of the candidate signals can be further improved.

**[0049]** Optionally, said image data can comprise at least two channels, in particular a first channel indicative of a first wavelength interval and a second channel indicative of a second wavelength interval. An advantage of this embodiment is that this enables even more precise weighting than using a single channel, since more parameters are available. For example, first different channels, using the same statistical metrics, may be combined to candidate signals for each of the different statistical metrics, and next combined, either in the time domain (e.g. using BSS), or in the frequency domain allowing different combination weights per spectral component. Alternatively, spectral weighting and combining of different wavelength channels may be reversed. As an example of multiple wavelength channels, different color channels of an RGB can be evaluated. In addition or in the alternative at least one of said channels may be an infrared (IR), near infrared (NIR) or thermal imaging channel. For example, one or more different infrared channels may be

evaluated, for example indicative wavelength intervals comprising a wavelength of 775 nm, 800 nm and/or 905 nm, respectively. In addition or in the alternative, at least one of said channels can be indicative of a depth or range acquired by a range imaging technique such as using time of flight (TOF) camera, using structured light or a stereo camera. For example, the image may be segmented taking into account a temperature (warm skin) or a distance (expected distance of skin or background to camera).

**[0050]** Hence, referring again to the system for determining a physiological parameter of the subject, the imaging unit configured to acquire image data of the scene can comprise one or more of an RGB (video) camera, a temperature camera and/or a depth camera.

**[0051]** It shall be understood that the processor (also referred to as processing unit) can be implemented as a single entity such as a microcontroller, field programmable gate array (FPGA) or may also be implemented as a distributed processing device comprising a plurality of separate processing entities or even a cloud-based solution. The processor may also be shared with other applications. The interface can be wired or wireless interface for receiving said image data. The interface can be an interface of the processor. The proposed device may also refer to a signal processing device. Advantageously, the proposed device or system may be co-integrated in a patient monitor or hospital information system.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**[0052]** These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments described hereinafter. In the following drawings

**[0053]** FIG. 1 shows a schematic diagram of a system in accordance with an embodiment of the present invention;

**[0054]** FIG. 2 shows a flow chart of a method according to a first aspect of the present disclosure;

**[0055]** FIG. 3 shows a flow chart regarding processing of an individual image frame in accordance with said first aspect;

**[0056]** FIG. 4 shows a diagram regarding the processing of a sequence of image frames in accordance with said first aspect;

**[0057]** FIG. 5 shows a flow chart of a method according to a second aspect of the present disclosure;

**[0058]** FIG. 6 shows a second flow chart in accordance with said second aspect of the present disclosure;

**[0059]** FIG. 7 shows a block diagram combining advantageous aspects of the first and the second aspect of the present disclosure;

**[0060]** FIG. 8 shows a diagram of images corresponding to different processing steps;

**[0061]** FIG. 9 shows a diagram of images and weighting maps over time;

**[0062]** FIG. 10 shows a diagram regarding extraction of a vital sign parameter based on different regions of an image;

**[0063]** FIG. 11 shows a comparison of a conventional ECG-based measurement and camera-based vital signs measurement according to the present disclosure;

**[0064]** FIG. 12 shows a graph of a signal quality of obtained candidate signals based on a statistical parameter indicative of a central tendency or statistical parameter indicative of a statistical dispersion versus the skin percentage in a region-of-interest; and

[0065] FIG. 13 shows a diagram regarding weighting of image frames;

#### DETAILED DESCRIPTION OF THE EMBODIMENTS

[0066] FIG. 1 shows a schematic diagram of a system 1 for determining a physiological parameter of a subject. The system 1 comprises an imaging unit 2 configured to acquire image data of a scene, said image data comprising a time-sequence of image frames; and a device 10 for determining a physiological parameter of a subject 20.

[0067] In this example shown in FIG. 1, the subject 20 is a patient lying in bed 22, for instance in a hospital or other healthcare facility, but may also be a neonate or premature infant, e.g. lying in an incubator, or a person at home or in a different environment, such as an athlete doing sports.

[0068] The imaging unit 2 may include a camera (also referred to as detection unit or remote PPG sensor) for acquiring an image data, said image data comprising a time-sequence of image frames. The image data can be indicative of reflected electromagnetic radiation, in particular in a wavelength range of visual and/or infrared light. The image data represents a scene, preferably comprising skin areas 23, 24, 25 of the subject 20 from which a physiological parameter of the subject 20 can be derived. Exemplary skin areas that are usually not covered by a blanket 26 or clothing are the forehead 23, the cheeks 24 or the hands or arms 25.

[0069] The image frames captured by the imaging may particularly correspond to a video sequence captured by means of an analog or digital photo sensor, e.g. in a (digital) camera. Such a camera usually includes a photo sensor-array, such as a CMOS or CCD image-sensor, which may also operate in a specific spectral range (visible, NIR) or provide information for different spectral ranges, particularly enabling the extraction of PPG signals. The camera may provide an analog or digital signal. The image frames include a plurality of image pixels having associated pixel values. Particularly, the image frames may include pixels representing light intensity values captured with different photosensitive elements of a photo sensor. These photosensitive elements may be sensitive in a specific spectral range (i.e. representing a specific color). The image frames include at least some image pixels being representative of a pulsatile region such as a skin portion of the person. Thereby, an image pixel may correspond to one photosensitive element of a photo-detector and its (analog or digital) output or may be determined based on a combination (e.g. through binning) of a plurality of the photosensitive elements.

[0070] The system 1 may further comprise an optional light source 3 (also called illumination source or light source or electromagnetic radiator), such as a lamp or LED, for illuminating/irradiating a region-of-interest, such as the skin areas 23, 24 of the patient's face (e.g. part of the cheek or forehead), with light, for instance in a predetermined wavelength range or ranges (e.g. in the red, green and/or infrared wavelength range(s)). The light reflected from the scene in response to said illumination is detected by the imaging unit or camera 2. In another embodiment no dedicated light source is provided, but ambient light is used for illumination of the subject scene. From the reflected light only light in a desired wavelength range (e.g. green and red or infrared light, or light in a sufficiently large wavelength range covering at least two wavelength channels) may be detected

and/or evaluated. In case of using a thermal camera, radiation of the human body may be used directly.

[0071] The device 10 comprises an interface 11 for receiving image data of the scene, the image data comprising a time-sequence of image frames; and a processor 12.

[0072] According to a first aspect of the present disclosure, the processor 12 is configured to perform at least some of the steps as described with reference to FIG. 2 to FIG. 4.

[0073] According to a second aspect of the present disclosure, the processor is configured to perform at least some of the steps as described with reference to FIG. 5 and FIG. 6.

[0074] The device 10 can comprise a memory or storage 13 that stores therein a computer program product or program code which causes at least one of said methods to be performed when carried out by the processor 12. The device 10 can further comprise an interface 14 for controlling another entity such as an external light source 3.

[0075] The device 10 can further comprise an interface 15 for displaying the extracted physiological parameter such as the pulse of the subject 20 and/or for providing medical personnel with an interface to change settings of the device 10, the camera 2, the light source 3 and/or any other parameter of the system 1. Such an interface 20 may comprise different displays, buttons, touchscreens, keyboards or other human machine interface means.

[0076] A system 1 as illustrated in FIG. 1 may, e.g., be located in a hospital, healthcare facility, elderly care facility or the like. Apart from the monitoring of patients, the solutions proposed herein may also be applied in other fields such as neonate monitoring, general surveillance applications, security monitoring or so-called lifestyle environments, such as fitness equipment, a wearable, a handheld device like a smartphone, or the like. The uni- or bidirectional communication between the device 10 and the imaging unit 2 may work via a wireless or wired communication interface. Other embodiments of the present invention may include a device 10, which is not provided stand-alone, but integrated into another entity such as for example the camera 2, a patient monitor, a hospital information system (HIS), cloud-based solution or other entity.

[0077] Remote photoplethysmography (rPPG) enables contactless monitoring of a physiological parameter of a subject such as monitoring of a cardiac activity by detecting pulse-induced subtle color changes of human skin surface for example using a regular RGB camera. In recent years, algorithms used for pulse extraction have matured but the additionally required means for full automation are much less developed, especially for the long-term monitoring. There are two conventional ways to automate an rPPG system. The first one (most commonly used) uses face detection, face tracking and skin selection. The second one uses dense video segmentation with local pulse estimation to find living skin pixels to initialize the measurement. However, neither is designed for long-term monitoring in real clinical applications such as sleep monitoring and neonatal monitoring. For instance the first approach is ipso facto not applicable to general body-parts (e.g., palm) or newborns. Furthermore, face detection may fail when the subject changes posture during sleep, when the camera registers the face under an unfavorable angle or when part of the face is covered by a blanket. The second approach needs spatio-temporally coherent local segments to create long-term time-tubes for pulse extraction and living-skin detection.

Hence, this method is sensitive to local motion and computationally expensive. Essentially, living-skin detection and pulse extraction depend on each other. Regarding these conventional approaches, the common feature is that both include the region-of-interest (ROI) identification as an essential step prior to pulse extraction.

**[0078]** The inventors recognize that for vital signs monitoring, it is only required to extract a target signal indicative of a physiological parameter of the subject such as the pulse of the subject, as the output but it is not necessary to provide specifics of a location of a region-of-interest (ROI location). Hence, according to the first aspect of the present disclosure, there is proposed a method to directly extract a physiological parameter of the subject such as the pulse, even from a full video, eliminating the step of ROI localization and tracking.

**[0079]** The approach described with reference to the first aspect of the present disclosure is based on assumption that the DC-colors of skin and background are usually quite stable over time. Even though the spatial location of the subject may vary from image to image, as the subject can be anywhere in the image, the DC-colors of surfaces in the scene (including skin and background such as hospital bedding) will hardly change. Therefore, it is proposed to use e.g. the DC-color as a feature to automate the pulse extraction, rather than an ROI location. Hence, the proposal builds on the hypothesis that the background color and light source color remain stable at least in the (relatively short/sliding) time-window required for extracting the physiological parameter. At least in restricted applications such as a clinical setup it is further possible to manage the illumination by light sources and to control background color to have at least a short-term stable emission spectrum. Hence, it is suggested to exploit e.g. the DC-color as a spatial feature to differentiate objects in an image for pulse extraction, ignoring their locations or motions.

**[0080]** FIG. 2 shows a flow chart of a method for determining a physiological parameter of a subject according to the first aspect of the present disclosure. The method is denoted in its entirety by reference numeral 100.

**[0081]** In a first step 101, image data of a scene is received, said image data comprising a time-sequence of image frames. In the non-limiting embodiment described herein, the image data can be video data acquired by an RGB camera.

**[0082]** In a second step 102, a set of weighting maps is generated for each of said image frames. Each set of weighting maps comprises at least a first weighting map and a second weighting map for weighting pixels of the corresponding image frame.

**[0083]** In a third step 103, a first weighted image frame is determined by weighting pixels of the image frame (received in step 101) based on the first weighting map (determined in step 102). Correspondingly, in step 104, a second weighted image frame is determined by weighting pixels of the image frame (acquired in step 101) by weighting pixels of the image frame based on the second weighting map (also determined in step 102).

**[0084]** In step 105 a first statistical parameter value is determined based on the first weighted image frame as determined in step 103. Correspondingly in step 106 a second statistical parameter value is determined based on the second weighted image frame as determined in step 104.

**[0085]** The aforementioned steps 102-106 are repeated for each of the image frames comprised in the image data such

that a sequence of first and second statistical parameter values is provided over time. The first statistical parameter values over time provided by step 105 are concatenated in step 107 based on the time-sequence of the corresponding image frames to obtain a first candidate signal. Correspondingly, the second statistical parameter values over time provided by step 108 are concatenated based on the time-sequence of the corresponding image frames to obtain a second candidate signal. It should be noted that the image data may comprise a plurality but not necessarily all image frames provided by the imaging unit.

**[0086]** In step 109, a physiological parameter of the subject is extracted based on said first and/or said second candidate signal. In the given non-limiting example, the physiological parameter can be a pulse of the subject. In particular, the first and/or second candidate signals may be selected based on a quality metric. For example the candidate signal providing the better signal to noise ratio in a frequency range of interest for pulse extraction, e.g., between 40 bpm and 240 bpm, may be selected and the physiological parameter extracted based thereon. Alternative quality metrics can be used.

**[0087]** FIG. 3 illustrates the processing for a single image frame in more detail. The same reference numerals as in FIG. 2 will be used so that a correspondence between the figures can be established. As described above, in step 101 image data is received. In the following examples, the image data may refer to a video sequence registered by an RGB camera viewing a scene including a living skin. As used herein  $I(x, c, t)$  denotes the intensity of a pixel at the location  $x$  of an image in the (color) channel  $c$  recorded at time  $t$ . Herein, the pixel location is given by an index  $x$ . Alternatively, reference could also be made to vertical and horizontal pixel coordinates within the image frame corresponding to said index. In a typical setup there can be  $c=1, 2, 3$  corresponding to the R-G-B (red, green blue) channels of a standard RGB camera. Optionally, down-sampling can be applied and the pixel  $x$  can be created from a down-sampled version of the image to reduce both the quantization noise and the computational complexity. For example, a standard  $640 \times 480$  or  $1920 \times 1080$  pixel image can be down-sampled to e.g.  $20 \times 20$  down-sampled pixels. The time  $t$  can also denote a frame index which is related to the time by a recording rate of e.g. 20 frames per second (fps).

**[0088]** For each image frame, i.e. for each time  $t$ , a set of weighting maps  $W_1(x, t) \dots W_n(x, t)$  is determined in step 102. Optionally, different weighting maps can be created for the different channels. However, as shown in FIG. 3, it is also possible to create a common weighting map taking into consideration the pixel values in two or more, preferably all, of the channels to fully exploit the information given in the channels of the image.

**[0089]** In step 103, a first weighted image frame  $J_1(x, c, t)$  is determined by weighting pixels of the image frame  $I(x, c, t)$  based on the first weighting map  $W_1(x, t)$ . Correspondingly, the second weighted image frame is generated by weighting pixels of the (same) image frame  $I(x, c, t)$  based on the second weighting map  $W_2(x, t)$  in step 104. It will be appreciated that the set of weighting maps can comprise a plurality of weighting maps  $1 \dots n$ . Hence, a plurality of corresponding weighted image frames  $J_1 \dots J_n$  can be determined accordingly.

**[0090]** For each of said weighted image frames, one or more statistical parameter values are extracted in a next

stage. For example, for the first weighted image frame a first statistical parameter value can be extracted in step 105. In the given example, a mean or average value  $\mu_1(c, t)$  is determined. Optionally, further statistical parameter values such as the standard deviation of the pixel values in the corresponding weighted image frame  $J_1(x, c, t)$  can be determined as or  $(c, t)$  as indicated by reference numeral 105'. Correspondingly, one or more second statistical parameter values can be determined based on the second weighted image frame in steps 106 and 106'. The same applies to further optional weighted image frames. The sequence of first statistical parameter values obtained over time from a sequence of first weighted images over time can be concatenated over time to obtain a first candidate signal. Correspondingly, the sequence of second statistical parameter values obtained over time from the sequence of second weighted images over time can be concatenated over time to obtain a first candidate signal.

[0091] FIG. 4 illustrates the process over time. Therein, each 'column' of processing steps denotes one point in time  $t_1 \dots t_n$ . In the given example four candidate signals are generated. In the given example, the first candidate signal can be obtained by concatenating the first statistical parameter values over time, here  $g_1(c, t_1) \dots \mu_1(c, t_n)$  based on the time-sequence of the corresponding image frames  $I(x, c, t_1) \dots I(x, c, t_n)$ . Correspondingly, a second candidate signal can be obtained by concatenating said second statistical parameter values over time, here  $\mu_2(c, t_1) \dots \mu_2(c, t_n)$  based on the time-sequence of the corresponding image frames  $I(x, c, t_1) \dots I(x, c, t_n)$ . Hence, in the given example, the first (and second) statistical parameter values refer to an average of the image frames that have been weighted by the first weighting map  $\bar{W}_1$  (and the second weighting map  $\bar{W}_2$ ), respectively.

[0092] Optionally, additional statistical parameter values can be obtained based on the first weighted image frame and/or the second image frame. In the given example, the standard deviation  $\sigma$  as a dispersion related statistical parameter value is determined. Hence, a further candidate signal can be obtained by concatenating the statistical parameter values  $\sigma_1(c, t_1) \dots \sigma_1(c, t_n)$  based on the time-sequence of the corresponding image frames  $I(x, c, t_1) \dots I(x, c, t_n)$  as obtained in the processing steps 105. The same can be applied to obtain a fourth candidate signal based on the processing steps 106' accordingly. Hence, the physiological parameter of the subject can be extracted based on one or more of said first, second, third and fourth candidate signals.

[0093] FIGS. 5 and 6 refer to a second aspect of the present disclosure. Features described in conjunction with the second embodiment can be implemented in combination with features of the first aspect but may also be implemented independent thereof.

[0094] Referring to FIG. 5, in a first step 201, image data of a scene is received, said image data comprising a time-sequence of image frames.

[0095] In step 202, for each of said image frames, a first statistical parameter value indicative of a statistical dispersion of pixel values of said image frame is determined.

[0096] In step 203, said first statistical parameter values of the individual image frames are concatenated over time, based on the time-sequence of the corresponding image frame, to obtain a first candidate signal.

[0097] In step 204 a physiological parameter of the subject can be extracted based on said first candidate signal.

[0098] The inventors have found that a signal quality of a PPG or candidate signal can be improved in particular in case where many non-skin pixels pollute the image frame or a region-of-interest therein or if a weighting map (according to the first aspect) is not fully accurate. In other words, a typical region from which the candidate signal is extracted comprises skin pixels as well as non-skin pixels that may corrupt an extracted physiological parameter when they are combined with skin-pixels. The inventors have recognized that problems and disadvantages of the prior art relating to an averaging of pixels can be mitigated by computing a statistical property characterizing or indicative of a statistical dispersion, such as the variance or standard deviation, of pixels of the image frame and using this statistical parameter value in determining a candidate signal based on which a physiological parameter value of the subject is extracted. Hence, it has been found that evaluating a statistical parameter value indicative of a statistical dispersion of pixel values of said image frame can provide a superior result in case of pollution due to non-skin pixels.

[0099] Advantageously, this concept can be combined with the first aspect of the present disclosure. For example, referring again to FIG. 2, the statistical parameter value that is determined based on the first weighted image frame can be a statistical parameter value indicative of a statistical dispersion of pixel values of said weighted image frame. Correspondingly, the aspect of using a weighted image frame as an input can be used in the method as described with reference to FIG. 5 by providing for example a weighted image frame as an input in step 201.

[0100] FIG. 6 shows a second embodiment of processing steps according to the second aspect of the present disclosure. The same reference numerals as in FIG. 5 will be used so that a correspondence between the figures can be established.

[0101] In step 201, the image data of a scene is received via an interface, said image data comprising a time-sequence of image frames. In the embodiment shown in FIG. 6, there is provided an optional pre-processing step 205, wherein a region-of-interest can be selected. The region-of-interest can be selected using conventional processing techniques such as face detection and tracking. In the alternative, the aspect of using weighting maps can be implemented as a preprocessing step 205. The output of said preprocessing step may then be provided as an image frame to the subsequent processing steps.

[0102] In step 202, for each of said image frames (optionally after the preprocessing), a first statistical parameter value indicative of a statistical dispersion of the pixel values of such image frame is determined. The first statistical parameter value can be indicative of at least one of a standard deviation, a variance, mean absolute difference, median absolute difference and/or an interquartile range.

[0103] In a further optional step 206, for each of said image frames as obtained from step 205 or directly from step 201, a second statistical parameter value indicative of a central tendency of pixel values of said image frame can be determined. It has been found that by determining a first statistical parameter value over time indicative of a statistical dispersion of pixel values provides an advantageous candidate signal in case of pollution by non-skin pixels in the image frame (or a region-of-interest selected therein). On the other hand, in case of non-skin pixels below a predetermined threshold, the evaluation of a central tendency of

pixel values of the image frame, such as a mean or average of the pixel values, can provide improved performance.

[0104] In step 203 said first statistical parameter values are concatenated over time based on the time-sequence of the corresponding image frames to obtain a first candidate signal. Correspondingly in step 207 said second statistical parameter values can be concatenated over time based on the time-sequence of the corresponding image frames to obtain a second candidate signal.

[0105] Optionally, one or more additional statistical parameter values indicative of a statistical dispersion of pixel values can be evaluated, for example, evaluating a standard deviation or variance as the first statistical parameter value and further evaluating an interquartile range as an additional processing step 202'. Correspondingly, additional second statistical parameter values indicative of other central tendency parameters can be evaluated. Correspondingly, the respective statistical parameter values can be concatenated over time to obtain additional candidate signals.

[0106] In step 204 a physiological parameter of the subject, such as the pulse of the subject, can be extracted based on said plurality of candidate signals. The same extraction methods as for the first aspect can be applied. The physiological parameter of the subject can be provided as an output in step 210.

[0107] FIG. 7 now refers to an embodiment combining advantageous features of both the first and the second aspect of the present disclosure. In a given non-limiting example an RGB video is provided as the image data of a scene comprising a time-sequence image frames. Given a video registered by an RGB camera viewing the scene including a living-skin,  $I(x, c, t)$  is used to denote the intensity of a pixel at a location  $x$  of an image, in the channel  $c$ , recorded at time  $t$ . Herein, the channels  $c=1, 2, 3$  again correspond to the R-G-B channels of a standard RGB camera. The pixel  $x$  can optionally be created from a down-sampled version of the image to reduce both quantization noise and computational complexity, e.g., 20x20 down-sampled pixels or patches from an input image frame comprising 640x480 pixels. As used herein, the time  $t$  can again denote the frame index which corresponds to the time by the frame rate of herein 20 frames per second (fps).

[0108] In step 102 a set of weighting maps is created. Optionally a normalization step can be performed. Since it is aimed to combine the (down-sampled) pixels sharing the same color feature for pulse extraction through a set of weighting maps, the color-features can be normalized to be independent of the light intensity. The intensity of each pixel  $x$  can be normalized by:

$$I_n(x, c, t) = \frac{I(x, c, t)}{\sum_{c=1}^3 I(x, c, t)}, \quad (1)$$

where  $I_n(x, c, t)$  denotes the locally normalized color values.

[0109] Next,  $I_n(t)$  can be used to generate multiple weighting maps wherein the pixels sharing similar normalized values are assigned to a similar weight. For example, spectral clustering as described in A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances In Neural Information Processing Systems*, MIT Press, 2001, pp. 849-856, can be used to build a fully connected affinity/similarity graph for all the patches

using  $I_n(t)$  and decompose it into uncorrelated subspaces, where a subspace can be used as an independent weighting mask to discriminate patches or pixels with different colors.

[0110] In line with this disclosure, the affinity matrix for all patch pixels in the  $t$ -frame can be built as:

$$A_{x,y}(t) = \|I_n(x, t) - I_n(y, t)\|_2, \quad (2)$$

where  $\|\cdot\|_2$  denotes the L2-norm (e.g. Euclidean distance),  $A$  denotes the symmetric and non-negative affinity matrix, where (down-sampled) pixels are pairwise connected. In a next step  $A$  can be decomposed into orthogonal (uncorrelated) subspaces using singular value decomposition (SVD):

$$A(t) = u(t) \cdot s(t) \cdot u^T(t), \quad (3)$$

where  $u(t)$  and  $s(t)$  denote the eigenvectors ( $u(t) \cdot u(t)^T = 1$ ) and eigenvalues, respectively. Since each eigenvector describes a group of patches having a similar color feature, a number of  $K$ , preferably  $K$  top-ranked, eigenvectors can be used to create the weighting maps, where  $K$  can be defined, either automatically (using  $s(t)$ ), or manually. Optionally, to fully exploit the eigenvectors, both the  $u(t)$  and  $-u(t)$  (i.e., the opposite direction) can be used to create the weighting vectors:

$$w(t) = [u_1(t), \dots, u_K(t), -u_1(t), \dots, -u_K(t)], \quad (4)$$

where  $u_i(t)$  denotes the  $i$ -th (column) eigenvector and each column of  $w(t)$  represents an image weighting vector. A number of  $2K$  weighting vectors can be generated by using the top- $K$  eigenvectors. Optionally, since the weights in  $w(t)$  should preferably be non-negative and temporally stable, i.e., not driven by pulse, it can first be shifted by:

$$\hat{w}_i(t) = w_i(t) - \min(w_i(t)), \quad (5)$$

where  $w_i(t)$  denotes the  $i$ -th (column) weighting vector and  $\min(\cdot)$  denotes the minimum operator, and then optionally normalized by:

$$\bar{w}_i(t) = \frac{\hat{w}_i(t)}{\text{sum}(\hat{w}_i(t))}, \quad (6)$$

where  $\text{sum}(\cdot)$  denotes the summation operator. This step is advantageous as it guarantees that the total weight for each frame can be the same, i.e. a normalization is applied. Hence, the total weight is time-stable and not driven/modulated by pulse. Each  $\bar{w}_i(t)$  can be reshaped into a weighting map (with the same dimension as the image), denoted as  $\bar{W}_i(t)$ , and used to weight each channel of  $I(t)$  as:

$$J_i(x, c, t) = \bar{W}_i(x, t) \odot I(x, c, t), \quad (7)$$

where  $J_i(x, c, t)$  denotes the  $x$ -th pixel in the  $c$ -th channel weighted by the  $i$ -th weighting map at time  $t$ . In FIG. 7 this step is denoted by 103, 104. The same reference numerals as in FIG. 2 to 4 are used so that a correspondence between the figures can be established.

[0111] In a next step 105, 106 each weighted image can be condensed into one or more spatial color representations given by a statistical parameter value. The respective statistical parameter values can be concatenated over time to provide a plurality candidate signals 107, 107', 108, 108'.

[0112] FIG. 8 exemplifies exemplary RGB images in FIGS. 8(a) and (b) and Near Infrared (NIR) images in FIGS. 8(c) and (d) at a point-in time  $t$  along with their corresponding weighting maps based on the top-4 eigenvectors.

The first image in FIG. 8(a) to (d) shows the intensity as described by  $I(x, c, t)$ . The second image in FIG. 8(a) to (d) shows the normalized intensity  $I_n(x, c, t)$ , as determined by equation (1) above. Since positive and negative eigenvectors are considered, eight weighting maps  $W_i$  are provided for each image.

[0113] It has been found that, in particular regarding the RGB images, both a visible light source color and background color can influence the weighting map. Also, from the normalized pixel values, as shown in the respective second images of FIG. 8(a) to (d) it can be seen that normalized pixel values of the skin can be very different in different lighting conditions. Such different conditions may be hardly modeled/learned off-line. Hence, an advantage of the approach disclosed herein is that no previous knowledge may be required. For the NIR images (in this example obtained at 675 nm, 800 nm and 905 nm), it has been found the visible light color does not influence the weighting maps substantially. It has been found that the skin can have similar reflections in the used NIR wavelengths, as opposed to the RGB wavelengths. This may also be the case for typical bedding materials. However, advantageously a difference between the skin reflection and bed reflection may start to occur at around 905 nm, which can be attributed to water absorption of the skin which is typically absent in the bedding. Hence, even in the challenging background with a white pillow (i.e., the color of the white pillow may be similar to that of skin in NIR), the weighting maps may still be used to discriminate skin and pillow using water absorption contrast at least at 905 nm.

[0114] FIG. 9 shows a sequence of images and generated weighting maps from the first four eigenvectors (Fig. 1 to Fig. 4). It has been found that regardless of the posture and position of the subject, the weighting maps are consistently biasing the image by attenuating similar parts of the scene. The first row in FIG. 9 shows a sequence of RGB images over time. The second to fifth row indicate the generated weighting maps corresponding to the respective image in the first row.

[0115] During a first period denoted by P1 the subject is lying in bed facing the camera. In such a situation, also conventional face detection may provide good results. In the second phase indicated by P2 the subject performs a motion by turning to the side. Hence, transitioning into a position where conventional face detection may fail. It should further be noted that the weighting map in the penultimate row not only correctly identifies a face region but also correctly detects the hand 25 of the subject which can therefore provide additional valuable input for determining the physiological parameter of the subject. During period P3 the subject assumes a position lying on the side. During period P4 the subject leaves the bed. In such a situation, conventional approaches that rely on tracking a region-of-interest may fail once the subject has left the field of view of the camera. However, as shown in period P5, the approach proposed herein correctly resumes operation and can again correctly weight the skin portions of the subject as e.g. indicated by the weighting map based on the third eigenvector Fig. 3.

[0116] Reference is now made to an advantageous combination with the second aspect of the present disclosure wherein a statistical parameter value indicative of a statistical dispersion of pixel values of an image frame is evaluated. The conventional way of combining image pixel values

into a single value is spatial averaging, i.e., the calculation of a central tendency metric. However, it only works when the weighted RGB image where the correct skin-pixels are provided with an increased weight. In case the weighted image frame is polluted by non-skin pixels, such averaging may provide inaccurate results. The inventors have recognized that problems and disadvantages of the prior art relating to an averaging of pixels can be mitigated by computing a statistical property characterizing or indicative of a statistical dispersion, such as the variance or standard deviation, of pixels of the weighted image frame to combine the weighted pixel values and using this statistical parameter value in determining a candidate signal based on which a physiological parameter value of the subject is extracted. As mentioned above, a pixel as used herein can also refer to the down-sampled image pixels or patches.

[0117] The rationale of using a statistical parameter value indicative of a statistical dispersion of the pixel values such as the variance is the following: When the non-skin pixels dominate the weighted image, they dominate the mean. Subtracting the mean and measuring the additional variation may thus reduce the impact of the non-skin pixels. Further details will be explained further below. Based on this understanding it had been recognized that the variance will be less effective when the skin and non-skin pixels have similar DC-color. Nevertheless, it can be seen as an alternative or back-up for the mean, especially in the case of imperfect weighting maps, where both the skin and known skin regions are emphasized.

[0118] FIG. 10 illustrates that mean and variance can have complementary strength when combining pixels from different weighted RGB images using different regions. The top row illustrates the same image frame, wherein different regions that form the basis for the combination of pixel values are indicated by the frames 51a-51e. The graphs in the bottom row provide the extracted candidate signals that are extracted from the regions as indicated by the frames 51a-51e, respectively. In each of the graphs in the bottom row, the horizontal axis denotes the time  $t$ , given by the frame number, wherein the vertical axis denotes an amplitude of the extracted candidate signal.

[0119] For each of said image frames, two statistical parameter values are determined from the regions indicated by frames 51a to 51e. The first statistical parameter value is indicative of a statistical dispersion of pixel values within said frame. The second statistical parameter value is indicative of a central tendency of pixel values within said frame. The respective statistical parameter values are concatenated over time based on the time-sequence of the corresponding image frames to obtain a first and a second candidate signal. In the given example, the first candidate signal 53 corresponds to the standard deviation as a parameter indicative of the statistical dispersion, whereas the second candidate signal 52 corresponds to the mean as a central tendency metric.

[0120] More precisely, in the given example, the mean-based candidate signal 52 and the variance-based signal 53 are generated by mean (R/G) and var (R/G) respectively, wherein R/G is a ratio of the pixel values in the red and green color channels of an RGB image. For a visualization purpose, both the mean signal and variant signal have low frequency components (<40 bpm) removed, mean subtracted and standard deviation normalized. It has been found that the mean-based candidate signal 52 and the variance-

based candidate signal **53** show complementary strength. When non-skin pixels dominate, as in **51a**, the variance-based candidate signal **53** is advantageous, whereas when skin pixels dominate, as in **51e**, the mean-based candidate signal **52** performs better. Hence, it has been found that evaluating a statistical parameter value indicative of statistical dispersion can provide better performance in case of a polluted image frame with non-skin pixels.

[0121] Referring again to FIG. 7, step **105** this can be used to determine, for each weighted image frame  $J_i$  weighted by a corresponding weighting map  $W_i$ , a first statistical parameter value indicative of a statistical dispersion of pixel values of said weighted image frame and a second statistical parameter value indicative of a central tendency of pixel values of said weighted image frame. The output of step **105** are thus two statistical parameter values. The respective statistical parameter values can be concatenated over time based on the time-sequence of the corresponding weighted image frames to obtain a first candidate signal **107** based on said first statistical parameter values, a second candidate signal **107'** based on said second statistical parameter values. Correspondingly, additional weighted image frames can be processed accordingly, as indicated by step **106**, to obtain further candidate signals **108**, **108'**.

[0122] The respective candidate signals over time wherein e.g. mean and variance may have complementary strengths as indicated in FIG. 10, can be written as:

$$\begin{cases} T_{2i-1}(c, t) = \text{mean}(J_i(x, c, t)) \\ T_{2i}(c, t) = \text{var}(J_i(x, c, t)) \end{cases} \quad (8)$$

where  $T_i(c, t)$  denotes a candidate signal;  $\text{mean}(\cdot)$  denotes an averaging operator;  $\text{var}(\cdot)$  denotes the variance operator. Hence, in view of the two different ways to determining and concatenating different statistical parameter values **107** and **107'** as well as **108** and **108'** in the example of FIG. 7, the number of temporal traces is double the number of weighting maps.

[0123] Referring again to FIG. 7, in the next step **109**, a physiological parameter can be extracted based on said candidate signals obtained in the previous step. To extract the physiological parameter from the candidate signals  $T_i$ , known algorithms used in photoplethysmography can be applied. Exemplary algorithms are: HUE (G. R. Tsouri and Z. Li, "On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras", *J. Biomed. Opt.*, vol. 20, no. 4, p. 048002, 2015); PCA (M. Lewandowska et al., "Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity", in *Proc. Federated Conf. Comput. Sci. Inform. Syst. (FedCSIS)*, pp. 405-410, 2011); ICA (M.-Z. Poh et al., "Advancements in noncontact, multiparameter physiological measurements using a webcam", *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7-11, 2011); CHROM (G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG", *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878-2886, 2013); PBV (G. de Haan and A. van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature", *Physiol. Meas.*, vol. 35, no. 9, pp. 1913-1922, 2014); ABPV (M. van Gastel, S. Stuijk and G. de Haan, "New principle for measuring arterial blood oxygenation, enabling motion-

robust remote monitoring", *Nature Scientific Reports*, p. 1-16 19 2016); and POS (W. Wang et al., "Algorithmic principles of remote-PPG", *IEEE Trans. Biomed. Eng.*, vol. PP, no. 99, 2016).

[0124] The extraction of a remote-PPG signal or physiological parameter from each of the candidate signals  $T_i$  can generally be expressed as:

$$P_i = \text{rPPG}(T_i), \quad (9)$$

where  $\text{rPPG}(\cdot)$  denotes a core rPPG function, i.e. an algorithm for extracting a physiological parameter of the subject, such as the pulse, from the input candidate signal.

[0125] Optionally, further processing steps can be applied in order to determine a most likely pulse-signal from the candidate pulse signals  $P_i$ . For example, due to the use of both a central tendency (e.g. the mean) and a dispersion-related measure (e.g. the variance) for the spatial pixel combination, multiple  $T_i$  (and thus  $P_i$ ) may contain useful pulsatile content. Therefore, it would be advantageous to combine several candidate signals instead of selecting just one of them. Advantageously, for the case of extracting a pulse of the subject as the physiological parameter of the subject, only pulsatile frequency components in  $P_i$  may be of interest such that it is proposed to combine frequency components from a different candidate signals instead of directly combining (time) signals.

[0126] Advantageously, in order to arrive at a clean output, higher weights can be given to components that are more likely to be pulse-related during the combination. However it has to be considered that the frequency amplitude may not directly be used to determine the weights or selected the components, because a large amplitude may not be due to pulse but due to motion artifacts. In view of a relation of pulsatile energy and motion energy it is thus proposed to estimate an intensity signal of each  $T_i$  and use an energy ratio between pulsatile components and intensity components as weights. The rationale is: if a frequency component in  $P_i$  is caused by pulse, it should have larger pulsatile energy with respect to the total intensity energy. If a component has balanced pulsatile and intensity energies, its "pulsatile energy" is more likely to be noise/motion induced. It should be mentioned that the use of intensity signal here is mainly for suppressing the background components although it may suppress motion artifacts as well.

[0127] The extraction of an intensity-signal from each  $T_i$  can be expressed as:

$$Z_i = \sum_{c=1}^3 T_i(c), \quad (10)$$

which is basically the summation of the R, G and B feature signals in case of an RGB signal. Since a local spectral energy contrast between the components in  $P_i$  and  $Z_i$  is used to derive their combining weights, their total energy (i.e., standard deviation) is first normalized and then they are transformed into the frequency domain e.g. using the Discrete Fourier Transform (DFT):

$$\begin{cases} F_{P_i} = \text{DFT}\left(\frac{P_i - \mu(P_i)}{\sigma(P_i)}\right) \\ F_{Z_i} = \text{DFT}\left(\frac{Z_i - \mu(Z_i)}{\sigma(Z_i)}\right) \end{cases} \quad (11)$$

where DFT ( $\cdot$ ) denotes the DFT operator. A weight for  $b$ -th frequency component in  $Fp_i$  can be derived by:

$$W_{i,b} = \begin{cases} \frac{\text{abs}(Fp_{i,b})}{1 + \text{abs}(Fz_{i,b})}, & \text{if } b \in B = [b_1, b_2], \\ 0, & \text{elsewhere,} \end{cases} \quad (12)$$

where  $\text{abs}(\cdot)$  takes the absolute value (i.e. amplitude) of a complex value;  $B$  optionally denotes a band for filtering such as a heart-rate and for eliminating clearly non-pulsatile components which can e.g. be defined as [40, 240] beats per minute (bpm) according to the resolution of  $Fp_i$ . Optionally an additive component in the denominator, here +1, is provided which prevents boosting of noise when dividing a very small value, i.e., the total energy is 1 after the normalization in (11). Afterwards, the weighting vector  $W_i = [W_{i,1}, W_{i,2}, \dots, W_{i,n}]$  can be used to weight and combine  $Fp_i$  as:

$$Fh = \sum_{i=1}^{AK} W_i \odot Fp_i, \quad (13)$$

**[0128]** The combined frequency spectrum  $Fh$  can further be transformed back to the time domain, e.g. using the Inverse Discrete Fourier Transform (IDFT):

$$h = \text{IDFT}(Fh), \quad (14)$$

where IDFT ( $\cdot$ ) denotes the Inverse Discrete Fourier Transform operator. Consequently, a long-term pulse-signal or physiological parameter signal  $H$  can be derived by concatenating sections  $h$ , e.g. by overlap-adding  $h$ , (preferably after removing its mean and normalizing its standard deviation) estimated in different time windows or short sequences, e.g., using a sliding window. The physiological parameter of the subject, such as the pulse rate, can then be extracted based thereon and provided as an output in step 110 of FIG. 7.

**[0129]** FIG. 11 shows a comparison of the performance of the system provided herein with a contact-based conventional ECG-based system. The graphs in the first column denoted by (a), (d), (g) show exemplary image frames acquired by an RGB camera in a neonatal intensive care unit (NICU). As can be seen, the baby shows significant movement between the image frames. Supply hoses render the detection of facial features difficult. The second row with graphs (b), (e), (h) shows the pulse of the baby acquired by a conventional contact-based electrocardiogram (ECG). The horizontal axis denotes the time  $t$  in frames of the imaging unit whereas the vertical axis denotes the pulse rate in beats per minute (bpm). The third column of graphs indicated by (c), (f), (i) shows graphs wherein the pulse rate is determined using the method as proposed herein, in particular as described in detail with reference to FIG. 7. The method is referred to as full video pulse-extraction (FVP). As can be seen from the comparison of the second and third column, the contactless method disclosed herein shows very good correspondence with the ground truth (ECG) contact-based measurement.

**[0130]** FIG. 12 shows a graph relating to a performance improvement that can be achieved by the second aspect of the present disclosure. The horizontal axis denotes a per-

centage of skin pixels  $p_s$  of an image frame from which the physiological parameter value is extracted; whereas the vertical axis denotes a quality measure of the extracted signal.

**[0131]** Referring to FIG. 10, the frame 51a comprises a low percentage of skin pixels whereas the frame 51e provides a high percentage of skin pixels. In FIG. 12, the curve 55 denotes a quality regarding a candidate signal based on a first statistical parameter indicative of a statistical dispersion of pixel values of said frame, here the variance (cf. trace 53 in FIG. 10). The curve 54, on the other hand, denotes a quality regarding a candidate signal based on a second statistical parameter value indicative of a central tendency of pixel values of said image frames, here in a mean of the pixel values of the respective frame (cf. trace 52 in FIG. 10). As can be seen from FIG. 12 and also explained with reference to FIG. 10, the two curves 54 and 55 have complementary strength, wherein the mean-based signal 54 is advantageous when a high percentage of skin pixels is provided (corresponding to good ROI selection or accurate weighting according to the first aspect of the present disclosure), whereas the evaluation based on the variance denoted by curve 55 shows advantageous performance in case of a polluted signal comprising a high percentage of non-skin pixels. Hence, dispersion can be particularly helpful in case the image frame is polluted by a significant number of non-skin pixels. However, since it cannot be assumed that non-skin pixels will always be present, in an advantageous embodiment both a first dispersion-based candidate signal and a second candidate signal based on a central tendency can be evaluated.

**[0132]** By extracting candidate signals using both techniques, or both statistical metrics, during the extraction step the candidate signal providing the best signal quality can be evaluated. In case multiple wavelength channels are available a more robust approach for extracting a physiological signal can optionally be used. Examples have been shown in the literature and include: a fixed weighted sum over candidate signals of different wavelength channels (RGB, NIR), CHROM, POS, PBV-method, ABPV-method, blind source separation (PCA, ICA) preferably after normalizing the candidate signals, in particular when the candidate signals are based on the central tendency of pixels, e.g. by dividing by their temporal mean, or taking their logarithm and offset removal.

**[0133]** In case a dispersion metric is used to provide the (concatenated) candidate signal, relative pulsabilities in different wavelength channels may be affected by a distribution of pixels. For blind source separation methods this may not be a problem, but for methods that make assumption on, either a direction in color-space of the distortions, or of the PPG-signal (or other physiological parameter signal) itself it is advantageous to correct the candidate signals with respect to gain or sign, prior to extraction. More particularly, said correction may be based on a first principal component of pixel values. For example this can be based on a vector that points from a color point of skin color to a color point of background of pixel values in particular in a region-of-interest or highly weighted region.

**[0134]** It has been found that, in case of using multiple color channels, the use of statistical dispersion metrics may change a relative amplitude between different concatenated color traces. Hence, it may render the rPPG extraction methods such as PBV, ABPV, CHROM, or POS, using blood

volume pulse priors invalid or sub-optimal. Several different ways of correction can be used, optionally even in parallel: (i) Use of principal component analysis (PCA) to correct a relative amplitude of different color channels (or candidate signals obtained based on different color channels). Alternatively, a PBV vector of PBV, a set of PBV-vectors in ABPV, or a projection-axes of CHROM/POS can be corrected. Thereby, prior-related rPPG methods can be used. (b) Use of weighting maps, in particular in accordance with the first aspect of the present disclosure, to suppress non-skin pixels (similar to “pulling non-skin pixels to black”). Thereby, prior-related rPPG methods can be used. (c) Use the blind source separation (BSS) extraction. Thereby, amplitude-correction or pixel-weighting are not necessarily needed. (d) Combine the multi-wavelength images into a single-wavelength image (for example by  $2 \log(G) - \log(R) - \log(B)$ ) in a preprocessing step and then use a dispersion metric to combine spatial pixels.

**[0135]** Advantageously such a correction that does not have to be measured every frame, but only e.g. once in a predetermined analysis window. Sliding window processing can optionally be applied. Such a vector may be temporarily filtered e.g. recursively, to stabilize the correction. Gain correction may comprise dividing the color channels by their respective component of said principal component. For example, such a connection based on (a first) principal component of pixel values may assume that next to the skin pixels a single background color occurs in a region-of-interest.

**[0136]** In a complex situation, however, several different background colors may occur. In such a case such a correction may not be correct, since there is no single different background color. Optionally, a weighting can thus be applied, wherein non-skin pixels in the background are attenuated based on a likelihood of being skin pixels. Preferably pixels can be pulled towards black or white the more likely they do not belong to the skin. This causes multiple colored background patches to concentrate near a single color point, such as white or black in the given example, which can make the aforementioned correction valid again.

**[0137]** An exemplary embodiment of this process is illustrated in FIG. 13. The left column in FIG. 13 refers to an exemplary situation wherein a region-of-interest (ROI) phantom is provided having skin pixels **61** as well as differently colored non-skin regions **62** and **63**. The right column refers to the same ROI phantom however with correction applied, wherein non-skin pixels in regions **62'** and **63'** are attenuated. As can be seen from the second row, the ROI phantom without weighting applied shows a distorted spectrum having a low quality factor in FIG. 13 (c), whereas the spectrum with the weighting applied shows very distinct frequency peaks in FIG. 13 (d). The horizontal axis in FIGS. 13(c) and (d) denotes the frequency, whereas the vertical axis denotes the amplitude A.

**[0138]** FIGS. 13 (e) and (f) show the corresponding spectrograms wherein the horizontal axis denotes the time t and the vertical axis denotes the frequency f. As can be seen from the comparison of FIGS. 13 (e) and (f), the weighting by attenuating non-skin pixels leads to a very clean spectrogram wherein the pulsatile signal component can be clearly distinguished.

**[0139]** Referring again to the advantageous combination of the first and the second aspect of the present disclosure,

it should be noticed that in many practical situations it can be difficult to make a clear cut distinction between skin and non-skin parts in an image. Typically, setting a high specificity results in loss of many pixels or areas that could have provided valuable information, but setting a high sensitivity for skin typically results in many non-skin areas thereby diluting the information of interest. This incapacity of striking a correct balance is an underlying notion in the map generation as it simply refutes the idea of hard boundaries. A consequence is that the signal processing needs to handle both relatively clean signals in particular due to proper weighting maps on the one hand and rather polluted situations on the other hand. This ability can be attained by the second aspect of the present disclosure by advantageously evaluating both a first statistical parameter value indicative of a statistical dispersion of pixel values of said image frame and a second statistical parameter value indicative of a central tendency of pixel values. In the non-limiting examples described herein, the mean as a central tendency measure and variance as a dispersion-related measure will be evaluated. The commonly used approaches only use the mean as source of information only. As will be shown below in more detail below and as also indicated by FIG. 12, the mean and variance have complementary strength. To simplify the illustration reference will be made to a single channel case. However, conclusions carry over to multi-channel situations.

**[0140]** In considering the task of pulse extraction, each pixel in an image may be described as either skin or non-skin. Thus, there can be skin and non-skin distributions in an image. Two statistical models can be assumed, for either case, with a probability density function, PDF  $p_o(x)$ , and associated mean  $\mu_o$  and standard deviation  $\sigma_o$  where x denotes the signal strength (color intensity) and o is either skin s or background b. It can furthermore be supposed that the full image has a fraction  $f_o$  of either pixels (implying  $f_s + f_b = 1$ ). The composite image pixel PDF  $p(x)$  can be written as:

$$p(x) = f_s p_s(x) + f_b p_b(x). \quad (15)$$

The mean of  $p(x)$  is:

$$\begin{aligned} \mu &= \mathbb{E}[p(x)] \\ &= f_s \cdot \mathbb{E}[p_s(x)] + f_b \cdot \mathbb{E}[p_b(x)] \\ &= f_s \cdot \mu_s + f_b \cdot \mu_b, \end{aligned} \quad (16)$$

where  $\mathbb{E}[\cdot]$  denotes the expectation. The variance of  $p(x)$  is:

$$\begin{aligned} \sigma^2 &= \mathbb{E}[p^2(x)] - (\mathbb{E}[p(x)])^2 \\ &= f_s \cdot \mathbb{E}[p_s^2(x)] + f_b \cdot \mathbb{E}[p_b^2(x)] - (f_s \cdot \mu_s + f_b \cdot \mu_b)^2. \end{aligned} \quad (17)$$

**[0141]** It is known that for  $p_o(x)$  (including the skin and non-skin),  $\mathbb{E}[p_o(x)] = \mu_o$  and  $\mathbb{E}[(p_o(x) - \mu_o)^2] = \sigma_o^2$ . Thus  $\mathbb{E}[p_o^2(x)]$  can be expressed as

$$\mathbb{E}[p_o^2(x)] = \mathbb{E}[(p_o(x) - \mu_o + \mu_o)^2] \quad (18)$$

-continued

$$\begin{aligned} &= \mathbb{E}[(p_o(x) - \mu_o)^2] + 2\mu_o \cdot \mathbb{E}[p_o(x) - \mu_o] + \mathbb{E}[\mu_o^2] \\ &= \sigma_o^2 + \mu_o^2. \end{aligned}$$

Therefore, [17] can be rewritten as:

$$\begin{aligned} \sigma^2 &= f_s \cdot (\sigma_s^2 + \mu_s^2) + f_b \cdot (\sigma_b^2 + \mu_b^2) - (f_s \cdot \mu_s + f_b \cdot \mu_b)^2 \\ &= f_s \cdot \sigma_s^2 + f_b \cdot \sigma_b^2 + f_s \cdot \mu_s^2 + f_b \cdot \mu_b^2 - (f_s \cdot \mu_s + f_b \cdot \mu_b)^2 \\ &= f_s \cdot \sigma_s^2 + f_b \cdot \sigma_b^2 + f_s \cdot f_b \cdot (\mu_s - \mu_b)^2. \end{aligned} \quad (19)$$

[0142] Now it can be assumed that the mean skin-level is modulated for example by the blood perfusion.  $\mu_s$  can be expressed as a combination of a steady DC-component and a time-dependent AC-component:

$$\mu_s = \bar{\mu}_s + \tilde{\mu}(t), \quad (20)$$

where  $\bar{\mu}_s$  is the steady DC component and  $\tilde{\mu}$  is the time-varying AC component. Furthermore it can be assumed that the background statistics are constant (i.e. assuming means such as a weighting mask to attenuate the background) and neglecting all modulations in the variance of the skin. Therefore, the full image mean in (16) can be rewritten as:

$$\begin{aligned} \mu &= f_s \cdot (\bar{\mu}_s + \tilde{\mu}(t)) + f_b \cdot \mu_b \\ &= f_s \cdot \bar{\mu}_s + f_b \cdot \mu_b + f_s \cdot \tilde{\mu}(t), \end{aligned} \quad (21)$$

and the full (weighted) image variance in (19) can be rewritten as:

$$\begin{aligned} \sigma^2 &= f_s \cdot \sigma_s^2 + f_b \cdot \sigma_b^2 + f_s \cdot f_b \cdot (\bar{\mu}_s + \tilde{\mu}(t) - \mu_b)^2 \\ &= f_s \cdot \sigma_s^2 + f_b \cdot \sigma_b^2 + f_s \cdot f_b \cdot \left( (\bar{\mu}_s - \mu_b)^2 + 2(\bar{\mu}_s - \mu_b) \cdot \tilde{\mu}(t) + \tilde{\mu}^2(t) \right) \\ &\approx f_s \cdot \sigma_s^2 + f_b \cdot \sigma_b^2 + f_s \cdot f_b \cdot (\bar{\mu}_s - \mu_b)^2 + \\ &\quad 2f_s \cdot f_b \cdot (\bar{\mu}_s - \mu_b) \cdot \tilde{\mu}(t) \end{aligned} \quad (22)$$

where  $\tilde{\mu}^2$  can be ignored in the approximation, as the squared pulsatile changes are orders of magnitude smaller than other DC-related components. Consequently, it can be found that the pulsatile component in the full (weighted) image mean and full (weighted) image variance are respectively provided by:

$$\begin{cases} \hat{\mu} = f_s \cdot \tilde{\mu}(t) \\ \hat{\sigma}^2 = 2f_s \cdot (1 - f_s) \cdot (\bar{\mu}_s - \mu_b) \cdot \tilde{\mu}(t) \end{cases}, \quad (23)$$

[0143] As expected, if  $f_s=0$  (no skin), there is no pulsatile component in either statistical variable. Furthermore it can be observed that the pulse-contribution to the mean is a linearly decreasing function of  $f_s$ , i.e., the fraction of skin-pixels.

[0144] In other words, with less skin-pixels also less pulsatile amplitude is contained in the mean. The variance

shows another behavior as the function of the skin fraction. It contains no pulsatile component in both extreme cases (all skin or all background) but peaks in the middle assuming at least some contrast between skin and background:  $\bar{\mu}_s - \mu_b \neq 0$ .

[0145] This is also reflected by FIG. 12 by trace 55. The previous indicates that dependent on the fraction  $f_s$  and contrast, there may be more pulsatile information in the variance than in the mean. This is actually the underlying explanation of the findings illustrated in FIG. 12. When the region-of-interest is dominated by skin-pixels, the mean signal reflects the blood volume changes in a better way (i.e. the signal is less noisy). When the region-of-interest contains certain amount of non-skin pixels, the variance signal shows much clear pulsatile variations. Therefore, the use of variance as a dispersion-related measure, in particular in addition to the mean, can be valuable since it cannot be safely assumed that the image frame only contains skin.

[0146] In conclusion, the present disclosure provides an advantageous device system and method for determining a physiological parameter of a subject. In particular, according to the first aspect of the present disclosure the need for a region-of-interest initialization/detection and tracking can be eliminated. The second aspect of the present disclosure further provides an improved signal in case of a polluted input signal wherein the image frames or portions selected thereof or highly weighted therein comprise a combination of skin pixels and non-skin pixels.

[0147] While the invention has been illustrated and described in detail in the drawings and foregoing description, such illustration and description are to be considered illustrative or exemplary and not restrictive; the invention is not limited to the disclosed embodiments. Other variations to the disclosed embodiments can be understood and effected by those skilled in the art in practicing the claimed invention, from a study of the drawings, the disclosure, and the appended claims.

[0148] In the claims, the word “comprising” does not exclude other elements or steps, and the indefinite article “a” or “an” does not exclude a plurality. A single processor, element or other unit may fulfill the functions of several items recited in the claims. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage.

[0149] A computer program may be stored/distributed on a suitable non-transitory medium, such as an optical storage medium or a solid-state medium supplied together with or as part of other hardware, but may also be distributed in other forms, such as via the Internet or other wired or wireless telecommunication systems.

[0150] Any reference signs in the claims should not be construed as limiting the scope.

1. A device for determining a physiological parameter of a subject, the device comprising:

an interface for receiving image data of a scene, said image data comprising a time-sequence of image frames; and

a processor for processing said image data,

wherein the processor is configured to perform, for each of said image frames, the steps of:

generating a set of weighting maps comprising at least a first weighting map and a second weighting map for spatially weighting the pixels of the image frame;

determining a first weighted image frame by spatially weighting the pixels of the image frame based on the first weighting map;  
 determining a second weighted image frame by spatially weighting the pixels of the image frame based on the second weighting map;  
 determining a first statistical parameter value based on the first weighted image frame;  
 determining a second statistical parameter value based on the second weighted image frame;

wherein the processor is further configured to perform the steps of:

concatenating said first statistical parameter values over time based on the time-sequence of the corresponding image frames to obtain a first candidate signal;  
 concatenating said second statistical parameter values over time based on the time-sequence of the corresponding image frames to obtain a second candidate signal; and  
 extracting a physiological parameter of the subject based on said first and/or said second candidate signal.

2. The device as claimed in claim 1, wherein the first and second weighting map are generated for spatially weighting the pixels of the image frame such that a non-zero weight is assigned to at least one same pixel of the image frame with both the first and the second weighting map.

3. The device as claimed in claim 1, wherein the processor is configured to determine said set of weighting maps for each image based on a local property of the image.

4. The device as claimed in claim 3, wherein said local property of the image comprises at least one of a brightness, color, texture, depth and/or temperature.

5. The device as claimed in claim 1, wherein the processor is configured to determine said weighting maps based on a similarity of pixels of the image frame to a target category; in particular, wherein said target category is derived based on the content of the image frame, in particular based on spectral clustering.

6. The device as claimed in claim 1, wherein the processor is configured to determine said weighting maps based on an affinity matrix indicative of spectral clustering.

7. The device as claimed in claim 1, wherein the processor is further configured to generate normalized weighting maps.

8. The device as claimed in claim 1, wherein the step of generating said set of weighting maps comprises at least one of zeroing weights below a predetermined threshold, offset removal and or thresholding.

9. The device as claimed in claim 1, wherein the step of extracting the physiological parameter of the subject further comprises the step of selecting at least one of said first and said second candidate signals based on a quality metric.

10. The device as claimed in claim 1, wherein the processor is further configured to apply a blind source separation technique to the candidate signals to obtain independent signals and to select at least one of said independent signals based on a quality metric.

11. The device as claimed in claim 1, wherein the processor is configured to combine at least two of said candidate signals in the frequency domain.

12. The device as claimed in claim 1, wherein said image data comprises at least two channels, in particular a first channel indicative of a first wavelength interval and a second channel indicative of a second wavelength interval.

13. A system for determining a physiological parameter of a subject; the system comprising:

an imaging unit configured to acquire image data of a scene, said image data comprising a time-sequence of image frames; and

a device for determining a physiological parameter of a subject as claimed in claim 1 based on the acquired image data.

14. A method for determining a physiological parameter of a subject, the method comprising the steps of:

receiving image data of a scene, said image data comprising a time-sequence of image frames;

for each of said image frames, performing the steps of:

generating a set of weighting maps comprising at least a first weighting map and a second weighting map for spatially weighting the pixels of the image frame;

determining a first weighted image frame by spatially weighting the pixels of the image frame based on the first weighting map;

determining a second weighted image frame by spatially weighting the pixels of the image frame based on the second weighting map;

determining a first statistical parameter value based on the first weighted image frame;

determining a second statistical parameter value based on the second weighted image frame;

wherein the method further comprises the steps of:

concatenating said first statistical parameter values over time based on the time-sequence of the corresponding image frames to obtain a first candidate signal;

concatenating said second statistical parameter values over time based on the time-sequence of the corresponding image frames to obtain a second candidate signal;

extracting a physiological parameter of the subject based on said first and/or said second candidate signal.

15. A non-transitory computer readable medium comprising program code means for causing a computer to carry out the steps of the method as claimed in claim 14 when said computer program is carried out on the computer.

\* \* \* \* \*

专利名称(译)	用于确定受试者的生理参数的设备，系统和方法		
公开(公告)号	<a href="#">US20200178809A1</a>	公开(公告)日	2020-06-11
申请号	US16/637284	申请日	2018-08-01
[标]申请(专利权)人(译)	皇家飞利浦电子股份有限公司		
申请(专利权)人(译)	皇家飞利浦N.V.		
当前申请(专利权)人(译)	皇家飞利浦N.V.		
[标]发明人	WANG WENJING DEN BRINKER ALBERTUS CORNELIS DE HAAN GERARD		
发明人	WANG, WENJING DEN BRINKER, ALBERTUS CORNELIS DE HAAN, GERARD		
IPC分类号	A61B5/00 A61B5/024		
CPC分类号	A61B2576/00 A61B5/02416 A61B5/0077 G16H30/40		
优先权	2017185299 2017-08-08 EP		
外部链接	<a href="#">Espacenet</a> <a href="#">USPTO</a>		

# 摘要(译)

该应用涉及使用远程PPG的基于摄像机的生命体征监测。一种用于确定生理参数的装置(10)，包括：接口(11)，用于接收包括图像帧的时间序列的图像数据。处理器(12)，用于处理所述数据并配置为针对每个帧执行以下步骤：生成一组加权图，该加权图包括至少一个第一加权图和第二加权图，用于对帧的像素进行空间加权(102)；根据第一加权图(103)，通过对图像帧的像素进行空间加权，确定第一加权图像帧。确定对应的第二加权图像帧；基于所述第一加权图像帧(105)确定第一统计参数值；确定相应的第二统计参数值；根据对应的图像帧的时间顺序将所述第一统计参数值随时间进行级联，以获得第一候选信号(107)。获得对应的第二候选信号(108)。基于所述第一和/或所述第二候选信号(109)提取受试者的生理参数。

