(54) Title of the Invention: Detecting physiological responses based on facial skin color changes

(51) INT CL: **A61B 5/024** (2006.01)

(21) Application No: 1906592.9

(22) Date of Filing: 02.10.2017

Date Lodged: 10.05.2019

(30) Priority Data:
(31) **62408677** (32) **14.10.2016** (33) **US**
(31) **62456105** (32) **07.02.2017** (33) **US**
(31) **62480496** (32) **02.04.2017** (33) **US**

(60) Parent of Application No(s)
2114500.8 under section 15(9) of the Patents Act 1977

(86) International Application Data:
PCT/IB2017/056069 En 02.10.2017

(87) International Publication Data:
WO2018/069791 En 19.04.2018

(43) Date of Reproduction by UK Office 17.07.2019

(56) Documents Cited:
WO 2016/025323 A1 **US 20170095157 A1**
US 20160360970 A1 **US 20160270656 A1**
US 20130124039 A1

(58) Field of Search:
As for published application 2570247 A viz:
INT CL **A61B**
Other: **Not Advised**
updated as appropriate

Additional Fields
Other: **None**
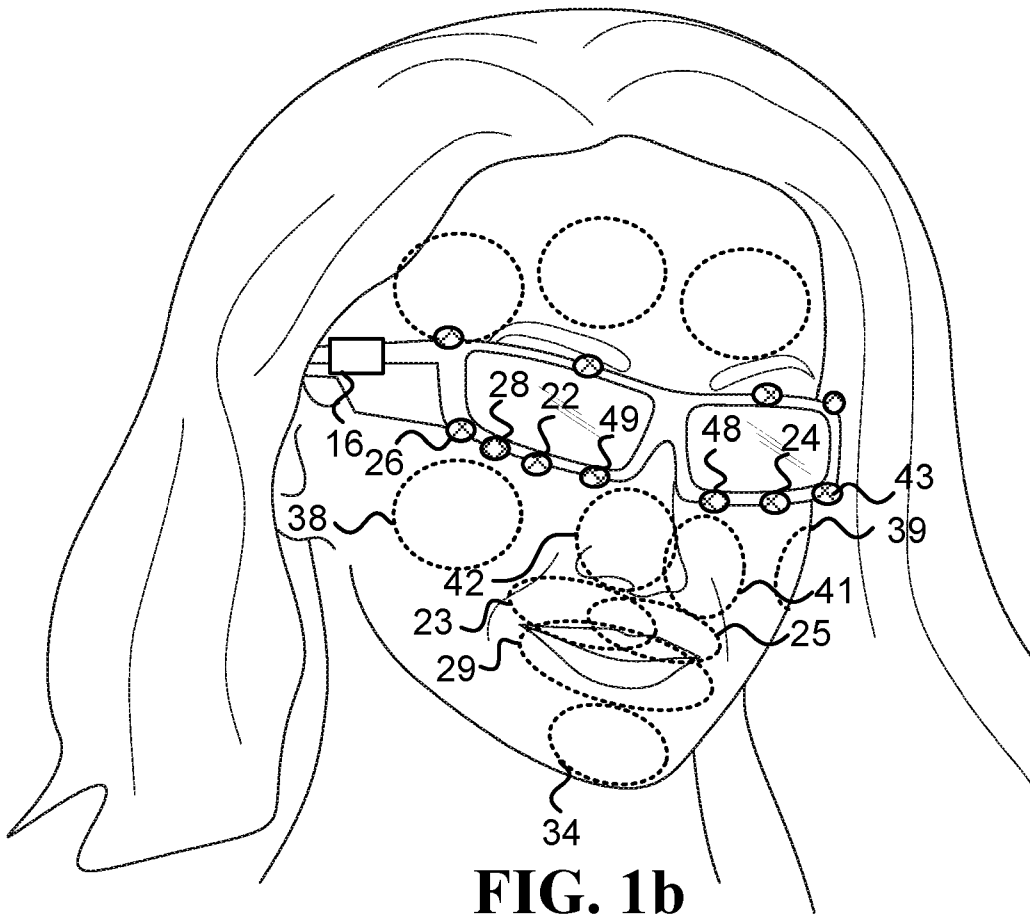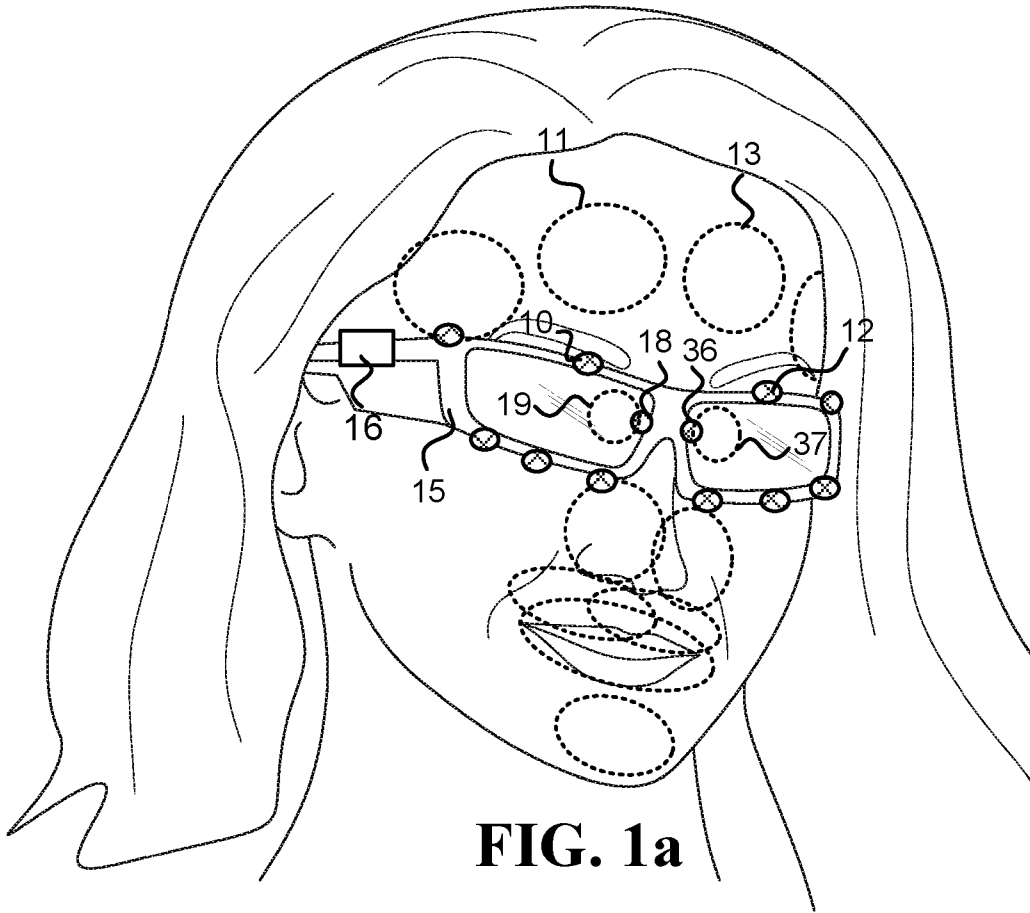
(72) Inventor(s):
Arie Tzvieli
Gil Thieberger
Ari M Frank

(73) Proprietor(s):
Facense Ltd.
(Incorporated in Israel)
POB 294, Kiryat Tivon 36010, Israel

(74) Agent and/or Address for Service:
Geoffrey Robert Dallimore
2 The Whitehouse, 17 Brunswick Hill, Reading,
RG1 7YT, United Kingdom

GB 2570247 B

FIG. 1a

FIG. 1b
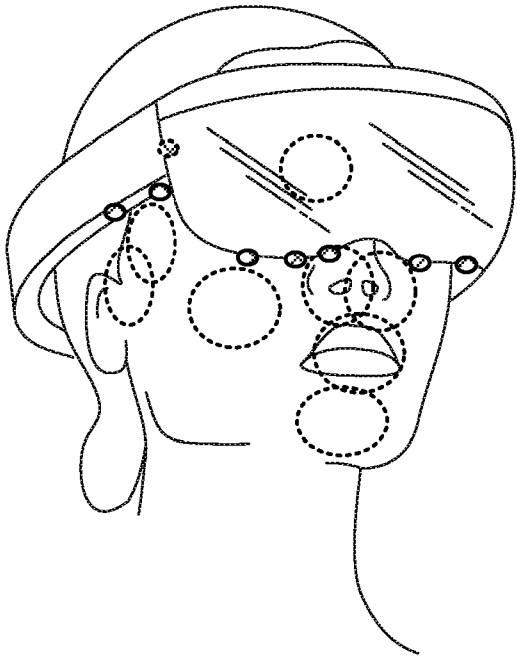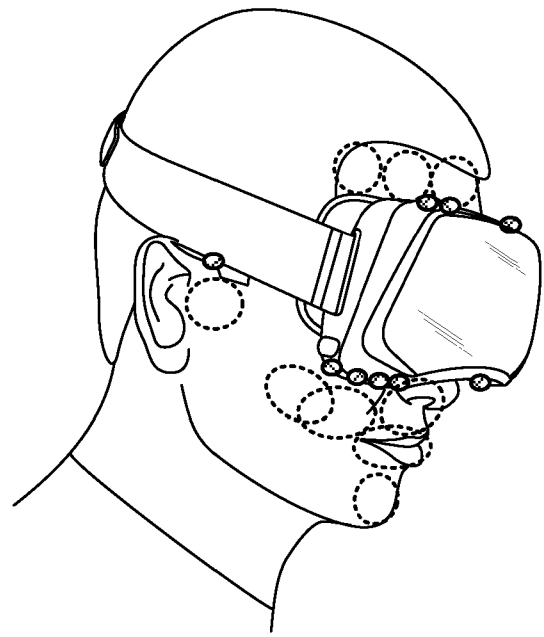
FIG. 2

FIG. 3
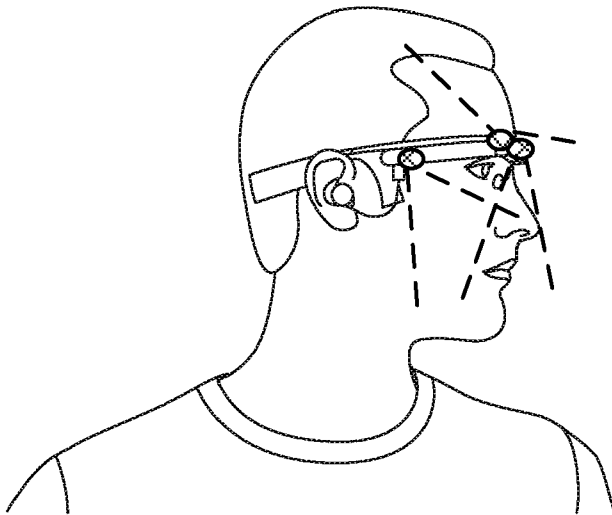
FIG. 4

FIG. 5
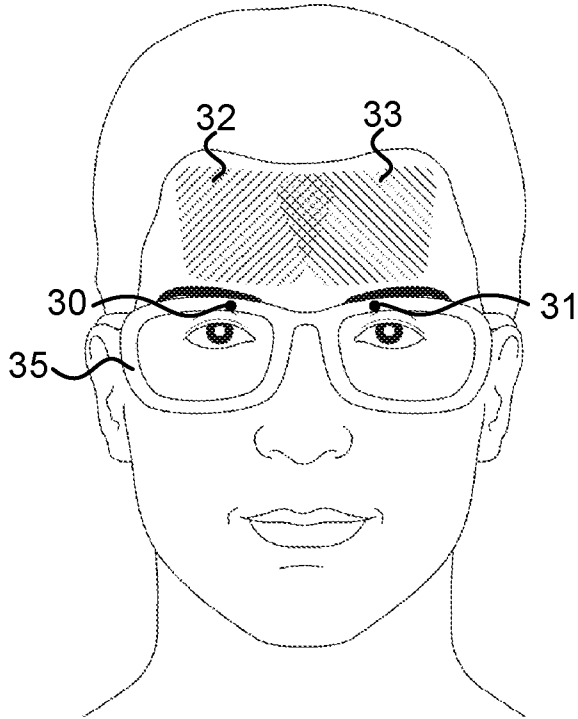
FIG. 6
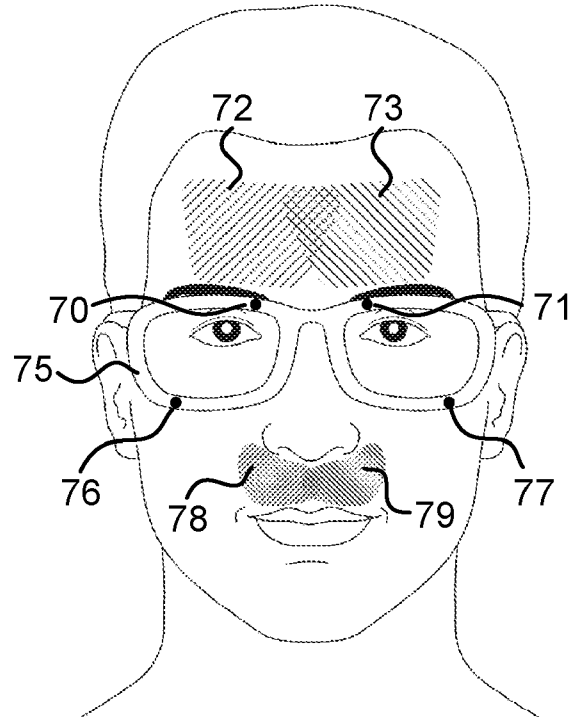
FIG. 7
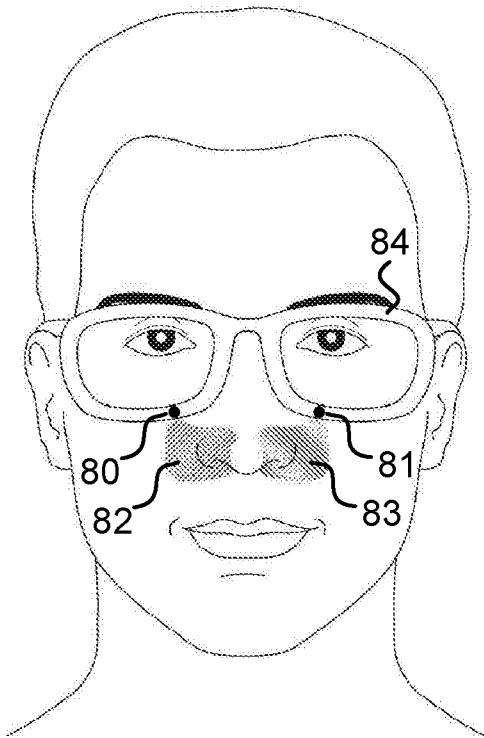
FIG. 8
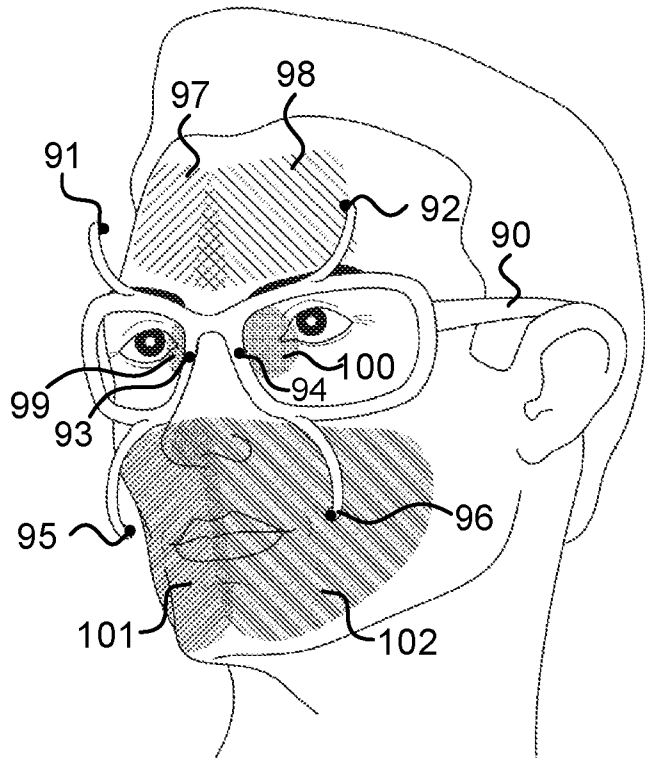
FIG. 9
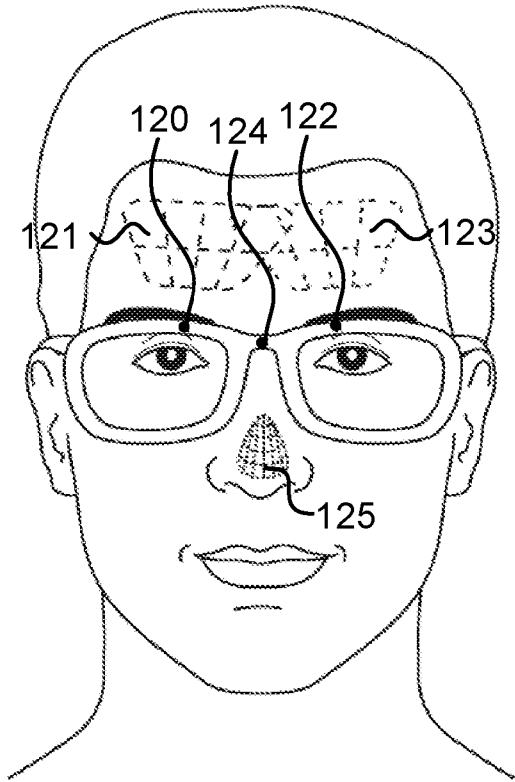
FIG. 10
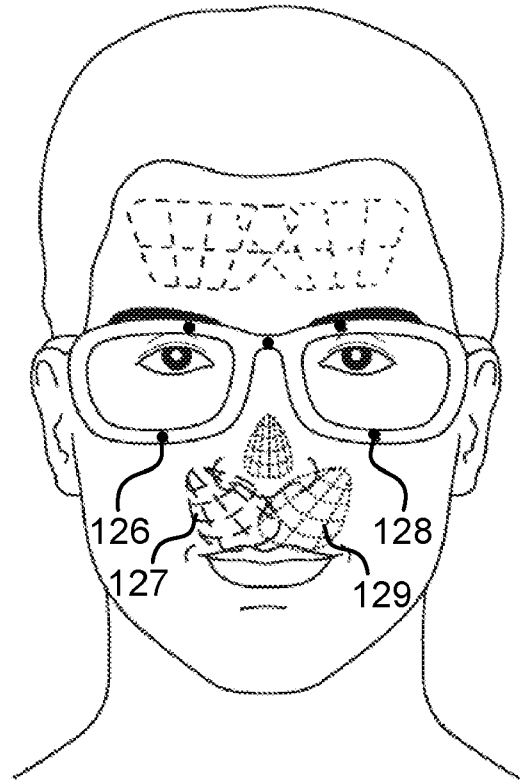
FIG. 11
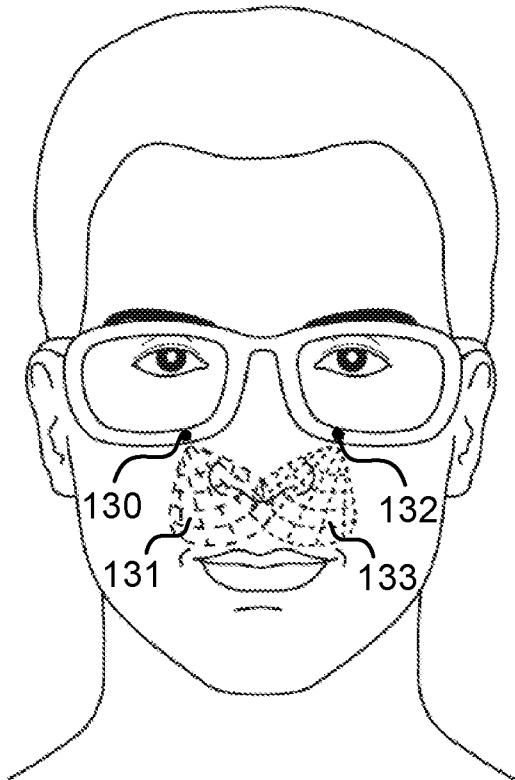
FIG. 12
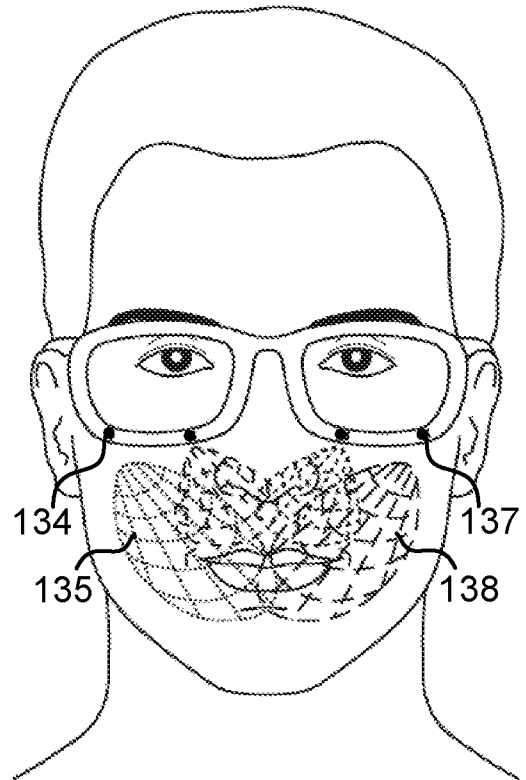
FIG. 13
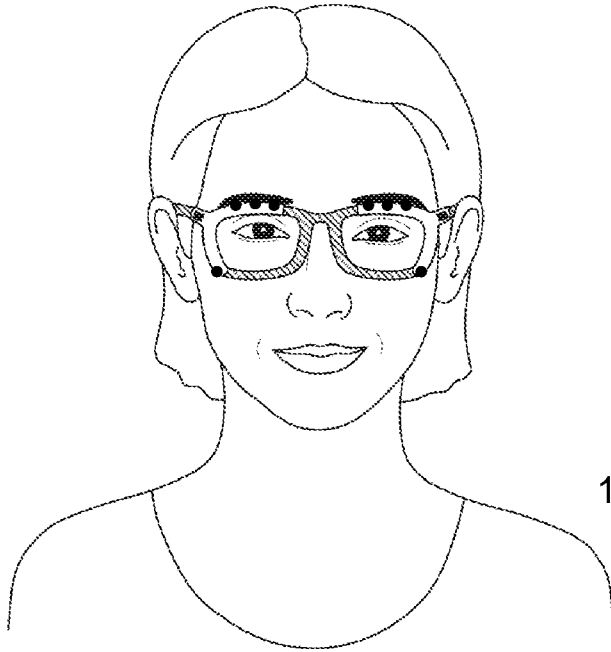
FIG. 14a
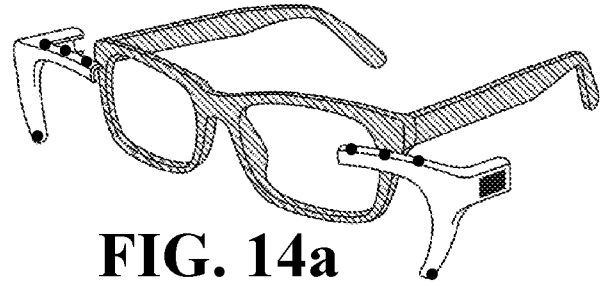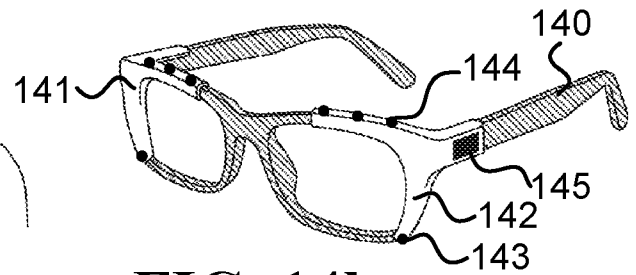
FIG. 14c

FIG. 14b

140
141
144
145
142
143

149
150
147
148

FIG. 15a

FIG. 15b

174

175

177
173

170

172

176

FIG. 17a

171

FIG. 17b

165

164
163

160

161

162

167

FIG. 16a

FIG. 16b

FIG. 18

Stroke alert

FIG. 19

FIG. 20

```
┌─────────────────┐              ▲              ┌─────────────────┐
│   Processor     │◄────────────►│◄────────────►│  User Interface │
│      401        │              │              │       404       │
└─────────────────┘              │              └─────────────────┘
                                 │
┌─────────────────┐            Bus 406           ┌─────────────────┐
│    Memory       │◄────────────►│◄────────────►│  Communication  │
│      402        │              │              │   interface     │
└─────────────────┘              │              │      405        │
                                 │              └─────────────────┘
┌─────────────────┐              │
│   Computer-     │◄────────────►│                    ⚡
│ Readable Medium │              │                 ╭───────╮
│      403        │              │                 │Network│
└─────────────────┘              ▼                 │  408  │
                                                   ╰───────╯
              400
```

FIG. 21a

```
┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
│  ┌─────────────────────────────────┐  │    ～410
│  │           Processor             │  │
│  │             411                 │  │
│  └─────────────────────────────────┘  │
│                                       │
│  ┌──────────────┐  ┌───────────────┐  │
│  │   Memory     │  │ Communication │  │
│  │    412       │  │   interface   │  │
│  │              │  │     413       │  │
│  └──────────────┘  └───────────────┘  │
└ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
```
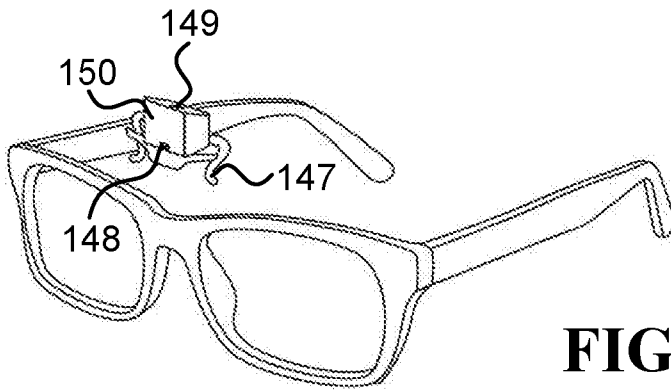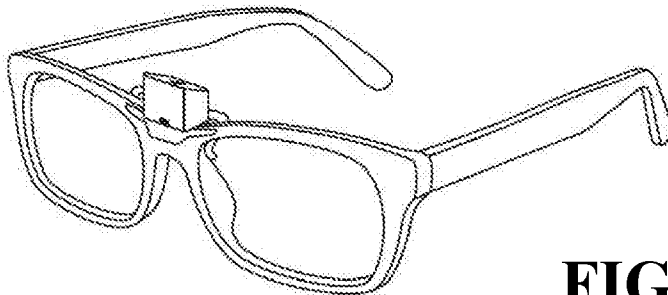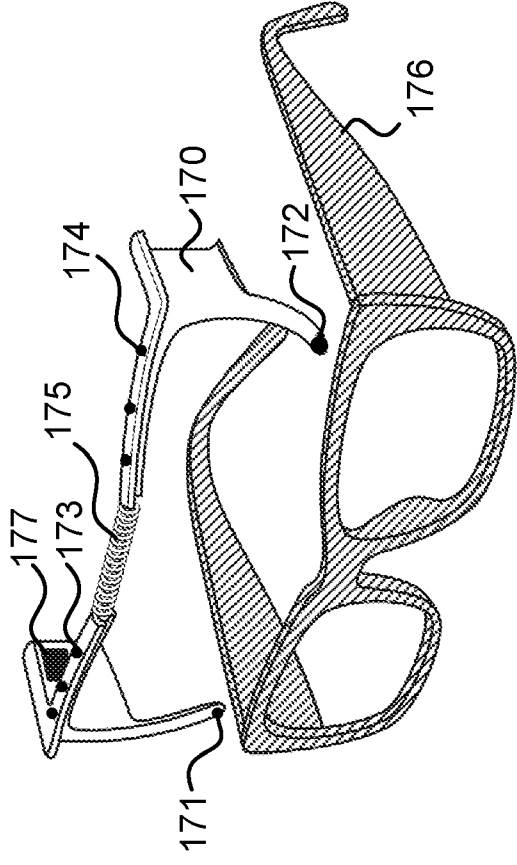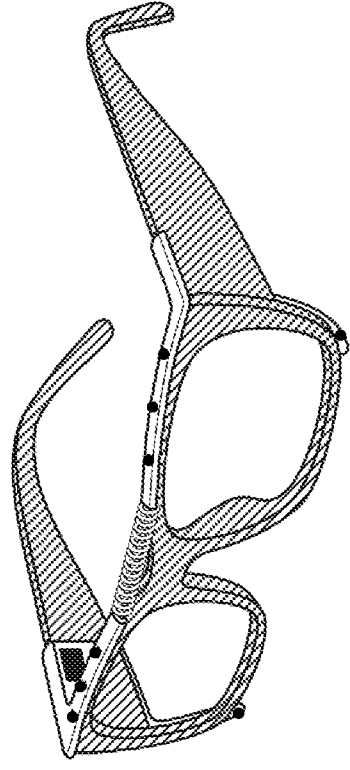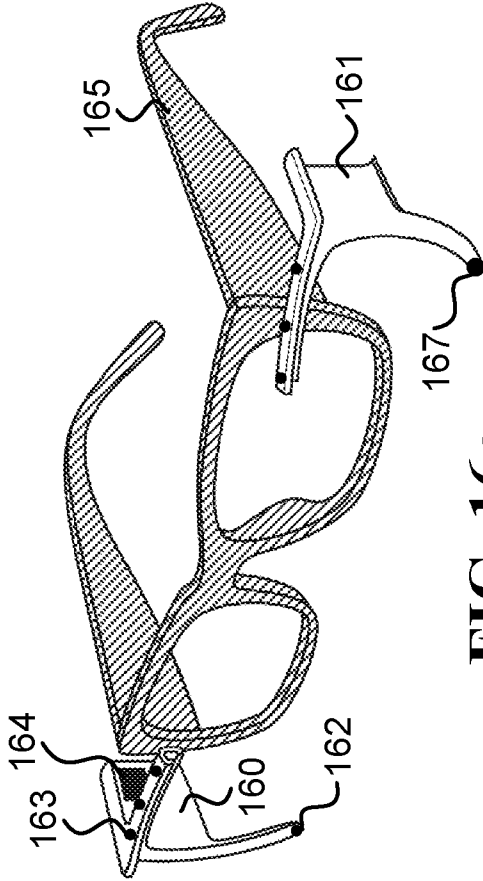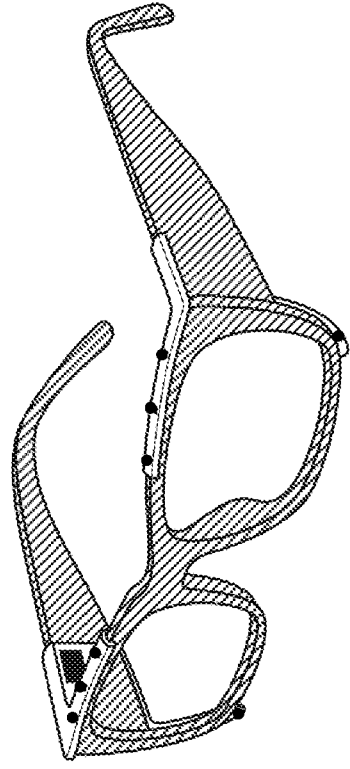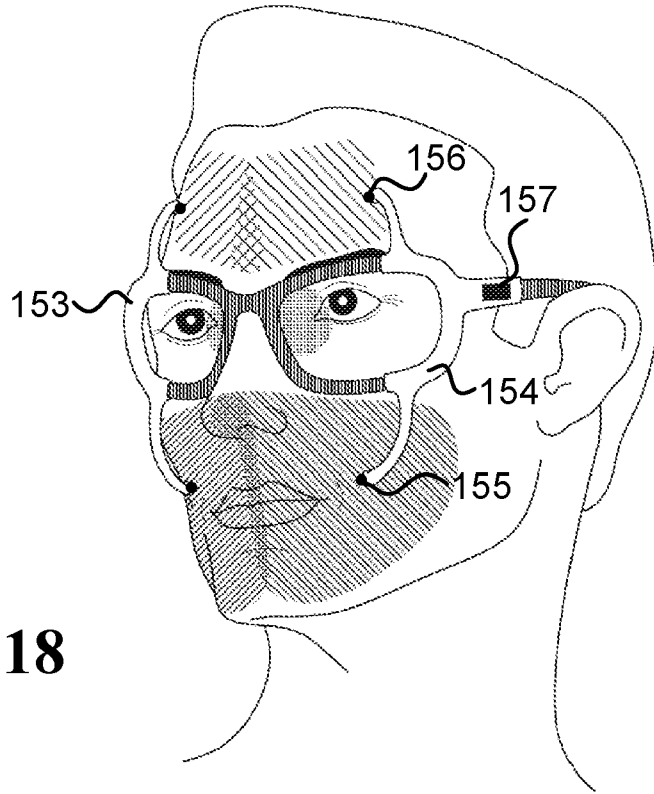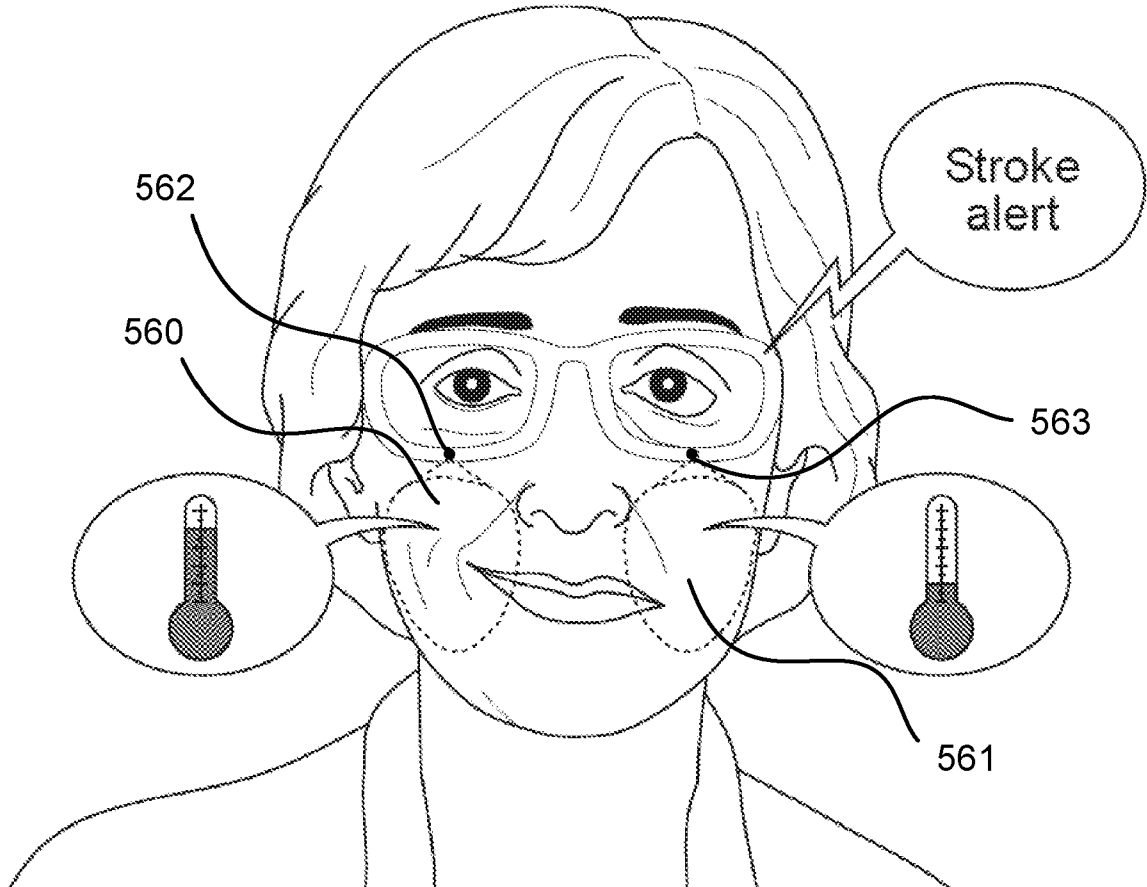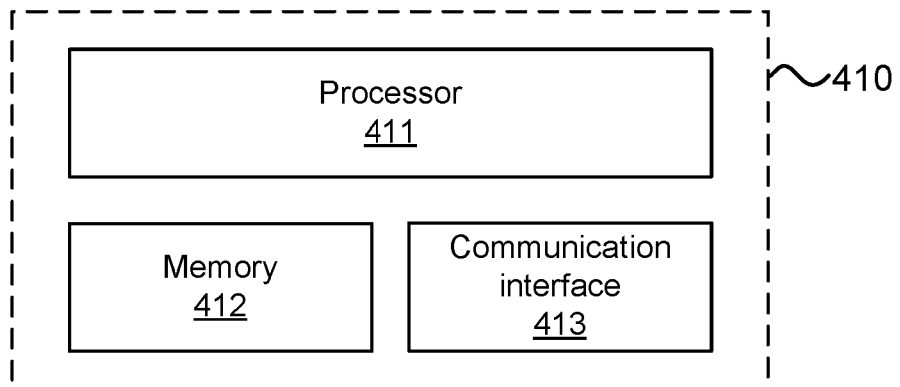
FIG. 21b

DETECTING PHYSIOLOGICAL RESPONSES BASED ON FACIAL SKIN COLOR CHANGES

TECHNICAL FIELD

[0001]   This application relates to head-mounted systems to measure facial temperature and to capture images of the face.

BACKGROUND

[0002]   Manifestation of various physiological responses involves temperature changes at various regions of the human face; measuring temperatures and/or temperature changes at the various regions on the face may help determine the level of stress a person is feeling, an extent of an allergic reaction the person has, or how a user feels, e.g., whether the user is nervous, calm, or happy. In some cases manifestation of physiological responses can involve facial skin color changes (FSCC), such as slight color changes due to a cardiac pulse or expression of an emotional response. FSCC can be detected from visible-light images of various regions of the face.

[0003]   Monitoring and analyzing measurements of the face (e.g., thermal measurements and/or the images described above) can be useful for many health-related and life-logging related applications. However, collecting such data over time when people are going through their daily activities can be very difficult, typically involves utilizing cameras that need to be continually pointed at a person's face, and often involves performing various complex image analysis procedures, such as procedures involving image registration and face tracking. Furthermore, due to the data being collected in uncontrolled settings, the measurements may be affected by various confounding factors such as thermal radiation directed at the face, touching the face, or consumption of certain substances (e.g., medication, alcohol, or caffeine). Therefore, due to the many applications they may enable, there is a need to be able to collect thermal measurements at various regions of a person's face. Preferably, the measurements are to be collected over a long period of time, while the person performs various day-to-day activities in uncontrolled settings.
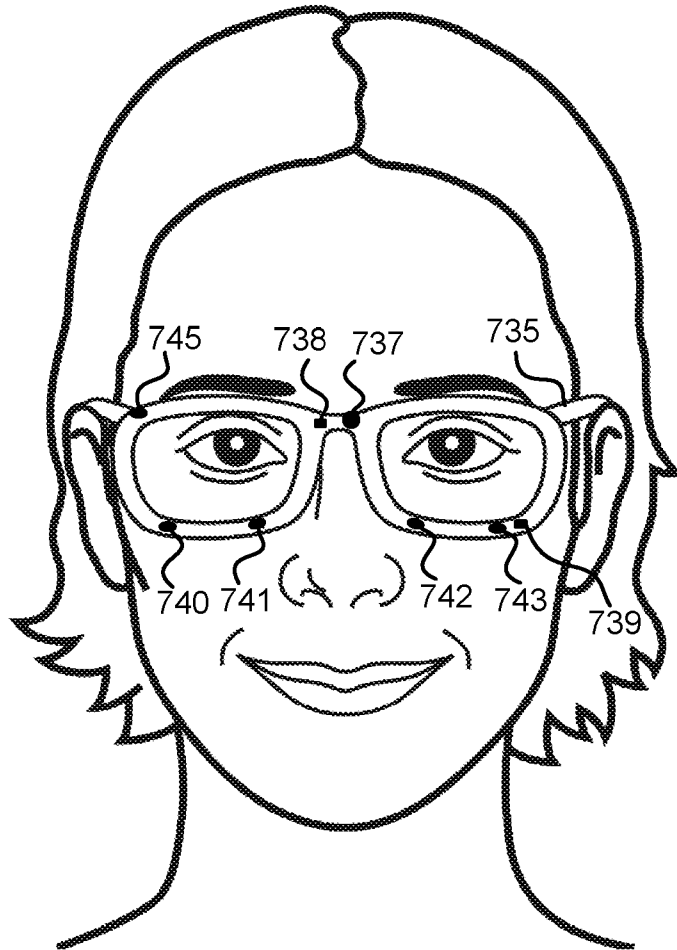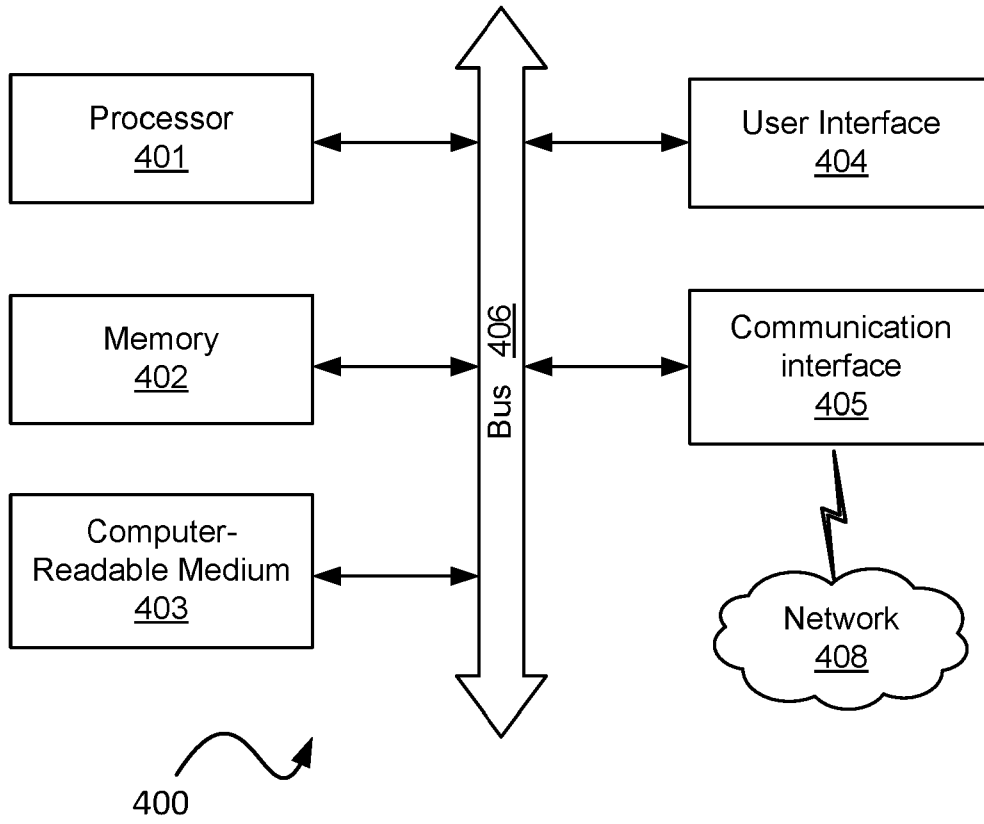
SUMMARY

[0004]   In one aspect, the present invention provides a system configured to detect a physiological response based on facial skin color changes (FSCC), comprising: an inward-facing head-mounted visible-light camera (VCAM$_{in}$) configured to take images of a region of interest (IM$_{ROI}$) on a user's face; wherein the region of interest (ROI) covers at least a portion of at least one of the following regions on the user's face: forehead, nose, and cheek; and wherein the ROI is illuminated by ambient light; and a computer configured to detect the physiological response based on a photoplethysmographic signal manifested by FSCC recognizable in the IM$_{ROI}$.

[0005] In another aspect, the present invention provides a method for detecting a physiological response, comprising: taking, utilizing an inward-facing head-mounted visible-light camera ($VCAM_{in}$), images of region of interest (IM_{ROI}) on a user's face; wherein the region of interest (ROI) covers at least a portion of at least one of the following regions on the user's face: forehead, nose, and cheek; and wherein the ROI is illuminated by ambient light; taking, utilizing an outward-facing head-mounted visible-light camera ($VCAM_{out}$), images of the environment (IM_{ENV}); wherein IM_{ENV} is indicative of illumination towards the face; and detecting the physiological response based on: a photoplethysmographic signal manifested by FSCC recognizable in IM_{ROI}, and IM_{ENV}; wherein utilizing IM_{ENV} accounts, at least in part, for variations in ambient light..

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] The embodiments are herein described by way of example only, with reference to the following drawings:

[0007] FIG. 1a and FIG. 1b illustrate various inward-facing head-mounted cameras coupled to an eyeglasses frame;

[0008] FIG. 2 illustrates inward-facing head-mounted cameras coupled to an augmented reality device;

[0009] FIG. 3 illustrates head-mounted cameras coupled to a virtual reality device;

[0010] FIG. 4 illustrates a side view of head-mounted cameras coupled to an augmented reality device;

[0011] FIG. 5 illustrates a side view of head-mounted cameras coupled to a sunglasses frame;

[0012] FIG. 6 to FIG. 9 illustrate head-mounted systems (HMSs) configured to measure various regions of interest (ROIs);

[0013] FIG. 10 to FIG. 13 illustrate various systems that include inward-facing head-mounted cameras having multi-pixel sensors (FPA sensors);

[0014] FIG. 14a, FIG. 14b, and FIG. 14c illustrate two right and left clip-on devices that are configured to attached/detached from an eyeglasses frame;

[0015] FIG. 15a and FIG. 15b illustrate a clip-on device that includes inward-facing head-mounted cameras pointed at the lower part of the face and the forehead;

[0016] FIG. 16a and FIG. 16b illustrate right and left clip-on devices that are configured to be attached behind an eyeglasses frame;

[0017] FIG. 17a and FIG. 17b illustrate a single-unit clip-on device that is configured to be attached behind an eyeglasses frame;

[0018] FIG. 18 illustrates right and left clip-on devices, which are configured to be attached/detached from an eyeglasses frame, and have protruding arms to hold inward-facing head-mounted cameras;

[0019] FIG. 19 illustrates a scenario in which an alert regarding a possible stroke is issued;

[0020] FIG. 20 illustrates a system configured to detect a physiological response based on facial skin color changes (FSCC); and

[0021]   FIG. 21a and FIG. 21b are schematic illustrations of computer systems.

DETAILED DESCRIPTION

[0022]   A "thermal camera" refers herein to a non-contact device that measures electromagnetic radiation having wavelengths longer than 2500 nanometer (nm) and does not touch its region of interest (ROI). A thermal camera may include one sensing element (pixel), or multiple sensing elements that are also referred to herein as "sensing pixels", "pixels", and/or focal-plane array (FPA). A thermal camera may be based on an uncooled thermal sensor, such as a thermopile sensor, a microbolometer sensor (where microbolometer refers to any type of a bolometer sensor and its equivalents), a pyroelectric sensor, or a ferroelectric sensor.

[0023]   Sentences in the form of "thermal measurements of an ROI" (usually denoted TH_{ROI} or some variant thereof) refer to at least one of: (i) temperature measurements of the ROI (T_{ROI}), such as when using thermopile or microbolometer sensors, and (ii) temperature change measurements of the ROI (ΔT_{ROI}), such as when using a pyroelectric sensor or when deriving the temperature changes from temperature measurements taken at different times by a thermopile sensor or a microbolometer sensor.

[0024]   A device, such as a thermal camera, may be positioned such that it occludes an ROI on the user's face, or the device may be positioned such that it does not occlude the ROI. Sentences in the form of "the system/camera does not occlude the ROI" indicate that the ROI can be observed by a third person located in front of the user and looking at the ROI, such as illustrated by all the ROIs in FIG. 7, FIG. 11 and FIG. 19. Sentences in the form of "the system/camera occludes the ROI" indicate that some of the ROIs cannot be observed directly by that third person, such as ROIs 19 and 37 that are occluded by the lenses in FIG. 1a, and ROIs 97 and 102 that are occluded by cameras 91 and 96, respectively, in FIG. 9.

[0025]   Although many of the disclosed examples can use occluding thermal cameras successfully, in certain scenarios, such as when using a head-mounted system (HMS) on a daily basis and/or in a normal day-to-day setting, using thermal cameras that do not occlude their ROIs on the face may provide one or more advantages to the user, to the HMS, and/or to the thermal cameras, which may relate to one or more of the following: esthetics, better ventilation of the face, reduced weight, simplicity to wear, and reduced likelihood to being tarnished.

[0026]   A "Visible-light camera" refers to a non-contact device designed to detect at least some of the visible spectrum, such as cameras with optical lenses and CMOS or CCD sensors.

[0027]   The term "inward-facing head-mounted camera" refers to a camera configured to be worn on a user's head and to remain pointed at its ROI, which is on the user's face, also when the user's head makes angular and lateral movements (such as movements with an angular velocity above 0.1 rad/sec, above 0.5 rad/sec, and/or above 1 rad/sec). A head-mounted camera (which may be inward-facing and/or outward-facing) may be physically coupled to a frame worn on the user's head, may be attached to eyeglass using a clip-on mechanism (configured to be attached to and detached from the eyeglasses), or may be mounted

to the user's head using any other known device that keeps the camera in a fixed position relative to the user's head also when the head moves. Sentences in the form of "camera physically coupled to the frame" mean that the camera moves with the frame, such as when the camera is fixed to (or integrated into) the frame, or when the camera is fixed to (or integrated into) an element that is physically coupled to the frame. The abbreviation "CAM" denotes "inward-facing head-mounted thermal camera", the abbreviation "CAM$_{out}$" denotes "outward-facing head-mounted thermal camera", the abbreviation "VCAM" denotes "inward-facing head-mounted visible-light camera", and the abbreviation "VCAM$_{out}$" denotes "outward-facing head-mounted visible-light camera".

[0028]   Sentences in the form of "a frame configured to be worn on a user's head" or "a frame worn on a user's head" refer to a mechanical structure that loads more than 50% of its weight on the user's head. For example, an eyeglasses frame may include two temples connected to two rims connected by a bridge; the frame in Oculus Rift™ includes the foam placed on the user's face and the straps; and the frames in Google Glass™ and Spectacles by Snap Inc. are similar to eyeglasses frames. Additionally or alternatively, the frame may connect to, be affixed within, and/or be integrated with, a helmet (e.g., sports, motorcycle, bicycle, and/or combat helmets) and/or a brainwave-measuring headset.

[0029]   When a thermal camera is inward-facing and head-mounted, challenges faced by systems known in the art that are used to acquire thermal measurements, which include non-head-mounted thermal cameras, may be simplified and even eliminated with some of the examples described herein. Some of these challenges may involve dealing with complications caused by movements of the user, image registration, ROI alignment, tracking based on hot spots or markers, and motion compensation in the IR domain.

[0030]   In various examples, cameras are located close to a user's face, such as at most 2cm, 5cm, 10cm, 15cm, or 20cm from the face. The distance from the face/head in sentences such as "a camera located less than 15 cm from the face/head" refers to the shortest possible distance between the camera and the face/head. The head-mounted cameras may be lightweight, such that each camera weighs below 10g, 5g, 1g, and/or 0.5g.

[0031]   The following figures show various examples of HMSs equipped with head-mounted cameras. FIG. 1a illustrates various inward-facing head-mounted cameras coupled to an eyeglasses frame 15. Cameras 10 and 12 measure regions 11 and 13 on the forehead, respectively. Cameras 18 and 36 measure regions on the periorbital areas 19 and 37, respectively. The HMS further includes an optional computer 16, which may include a processor, memory, a battery and/or a communication module. FIG. 1b illustrates a similar HMS in which inward-facing head-mounted cameras 48 and 49 measure regions 41 and 41, respectively. Cameras 22 and 24 measure regions 23 and 25, respectively. Camera 28 measures region 29. And cameras 26 and 43 measure regions 38 and 39, respectively.

[0032]   FIG. 2 illustrates inward-facing head-mounted cameras coupled to an augmented reality device such as Microsoft HoloLens™. FIG. 3 illustrates head-mounted cameras coupled to a virtual reality device such as Facebook's Oculus Rift™. FIG. 4 is a side view illustration of head-mounted cameras

coupled to an augmented reality device such as Google Glass™. FIG. 5 is another side view illustration of head-mounted cameras coupled to a sunglasses frame.

[0033] FIG. 6 to FIG. 9 illustrate HMSs configured to measure various regions of interest (ROIs). FIG. 6 illustrates a frame 35 that mounts inward-facing head-mounted cameras 30 and 31 that measure regions 32 and 33 on the forehead, respectively. FIG. 7 illustrates a frame 75 that mounts inward-facing head-mounted cameras 70 and 71 that measure regions 72 and 73 on the forehead, respectively, and inward-facing head-mounted cameras 76 and 77 that measure regions 78 and 79 on the upper lip, respectively. FIG. 8 illustrates a frame 84 that mounts inward-facing head-mounted cameras 80 and 81 that measure regions 82 and 83 on the sides of the nose, respectively. And FIG. 9 illustrates a frame 90 that includes (i) inward-facing head-mounted cameras 91 and 92 that are mounted to protruding arms and measure regions 97 and 98 on the forehead, respectively, (ii) inward-facing head-mounted cameras 95 and 96, which are also mounted to protruding arms, which measure regions 101 and 102 on the lower part of the face, respectively, and (iii) head-mounted cameras 93 and 94 that measure regions on the periorbital areas 99 and 100, respectively.

[0034] FIG. 10 to FIG. 13 illustrate various inward-facing head-mounted cameras having multi-pixel sensors (FPA sensors), configured to measure various ROIs. FIG. 10 illustrates head-mounted cameras 120 and 122 that measure regions 121 and 123 on the forehead, respectively, and mounts head-mounted camera 124 that measure region 125 on the nose. FIG. 11 illustrates head-mounted cameras 126 and 128 that measure regions 127 and 129 on the upper lip, respectively, in addition to the head-mounted cameras already described in FIG. 10. FIG. 12 illustrates head-mounted cameras 130 and 132 that measure larger regions 131 and 133 on the upper lip and the sides of the nose, respectively. And FIG. 13 illustrates head-mounted cameras 134 and 137 that measure regions 135 and 138 on the right and left cheeks and right and left sides of the mouth, respectively, in addition to the head-mounted cameras already described in FIG. 12.

[0035] The head-mounted cameras may be physically coupled to the frame using a clip-on device configured to be attached/detached from a pair of eyeglasses in order to secure/release the device to/from the eyeglasses, multiple times. The clip-on device holds at least an inward-facing camera, a processor, a battery, and a wireless communication module. Most of the clip-on device may be located in front of the frame (as illustrated in FIG. 14b, FIG. 15b, and FIG. 18), or alternatively, most of the clip-on device may be located behind the frame, as illustrated in FIG. 16b and FIG. 17b.

[0036] FIG. 14a, FIG. 14b, and FIG. 14c illustrate two right and left clip-on devices 141 and 142, respectively, configured to attached/detached from an eyeglasses frame 140. The clip-on device 142 includes an inward-facing head-mounted camera 143 pointed at a region on the lower part of the face (such as the upper lip, mouth, nose, and/or cheek), an inward-facing head-mounted camera 144 pointed at the forehead, and other electronics 145 (such as a processor, a battery, and/or a wireless communication module). The clip-on devices 141 and 142 may include additional cameras illustrated in the drawings as black circles.

[0037] FIG. 15a and FIG. 15b illustrate a clip-on device 147 that includes an inward-facing head-mounted camera 148 pointed at a region on the lower part of the face (such as the nose), and an inward-facing head-mounted camera 149 pointed at the forehead. The other electronics (such as a processor, a battery, and/or a wireless communication module) is located inside the box 150, which also holds the cameras 148 and 149.

[0038] FIG. 16a and FIG. 16b illustrate two right and left clip-on devices 160 and 161, respectively, configured to be attached behind an eyeglasses frame 165. The clip-on device 160 includes an inward-facing head-mounted camera 162 pointed at a region on the lower part of the face (such as the upper lip, mouth, nose, and/or cheek), an inward-facing head-mounted camera 163 pointed at the forehead, and other electronics 164 (such as a processor, a battery, and/or a wireless communication module). The clip-on devices 160 and 161 may include additional cameras illustrated in the drawings as black circles.

[0039] FIG. 17a and FIG. 17b illustrate a single-unit clip-on device 170, configured to be attached behind an eyeglasses frame 176. The single-unit clip-on device 170 includes inward-facing head-mounted cameras 171 and 172 pointed at regions on the lower part of the face (such as the upper lip, mouth, nose, and/or cheek), inward-facing head-mounted cameras 173 and 174 pointed at the forehead, a spring 175 configured to apply force that holds the clip-on device 170 to the frame 176, and other electronics 177 (such as a processor, a battery, and/or a wireless communication module). The clip-on device 170 may include additional cameras illustrated in the drawings as black circles.

[0040] FIG. 18 illustrates two right and left clip-on devices 153 and 154, respectively, configured to attached/detached from an eyeglasses frame, and having protruding arms to hold the inward-facing head-mounted cameras. Head-mounted camera 155 measures a region on the lower part of the face, head-mounted camera 156 measures regions on the forehead, and the left clip-on device 154 further includes other electronics 157 (such as a processor, a battery, and/or a wireless communication module). The clip-on devices 153 and 154 may include additional cameras illustrated in the drawings as black circles.

[0041] It is noted that the elliptic and other shapes of the ROIs in some of the drawings are just for illustration purposes, and the actual shapes of the ROIs are usually not as illustrated. It is possible to calculate the accurate shape of an ROI using various methods, such as a computerized simulation using a 3D model of the face and a model of a head-mounted system (HMS) to which a thermal camera is physically coupled, or by placing a LED instead of the sensor (while maintaining the same field of view) and observing the illumination pattern on the face. Furthermore, illustrations and discussions of a camera represent one or more cameras, where each camera may have the same FOV and/or different FOVs. Unless indicated to the contrary, the cameras may include one or more sensing elements (pixels), even when multiple sensing elements do not explicitly appear in the figures; when a camera includes multiple sensing elements then the illustrated ROI usually refers to the total ROI captured by the camera, which is made of multiple regions that are respectively captured by the different sensing elements. The positions of the cameras in the figures are just for illustration, and the cameras may be placed at other positions on the HMS.

[0042] Sentences in the form of an "ROI on an area", such as ROI on the forehead or an ROI on the nose, refer to at least a portion of the area. Depending on the context, and especially when using a CAM having just one pixel or a small number of pixels, the ROI may cover another area (in addition to the area). For example, a sentence in the form of "an ROI on the nose" may refer to either: 100% of the ROI is on the nose, or some of the ROI is on the nose and some of the ROI is on the upper lip.

[0043] Various examples described herein involve detection of physiological responses based on user measurements. Some examples of physiological responses include stress, an allergic reaction, an asthma attack, a stroke, dehydration, intoxication, or a headache (which includes a migraine). Other examples of physiological responses include manifestations of fear, startle, sexual arousal, anxiety, joy, pain or guilt. Still other examples of physiological responses include physiological signals such as a heart rate or a value of a respiratory parameter of the user. Optionally, detecting a physiological response may involve one or more of the following: determining whether the user has/had the physiological response, identifying an imminent attack associated with the physiological response, and/or calculating the extent of the physiological response.

[0044] Detection of the physiological response may be done by processing thermal measurements that fall within a certain window of time that characterizes the physiological response. For example, depending on the physiological response, the window may be five seconds long, thirty seconds long, two minutes long, five minutes long, fifteen minutes long, or one hour long. Detecting the physiological response may involve analysis of thermal measurements taken during multiple of the above-described windows, such as measurements taken during different days. A computer may receive a stream of thermal measurements, taken while the user wears an HMS with coupled thermal cameras during the day, and periodically evaluate measurements that fall within a sliding window of a certain size.

[0045] Models may be generated based on measurements taken over long periods. Sentences of the form of "measurements taken during different days" or "measurements taken over more than a week" are not limited to continuous measurements spanning the different days or over the week, respectively. For example, "measurements taken over more than a week" may be taken by eyeglasses equipped with thermal cameras, which are worn for more than a week, 8 hours a day. In this example, the user is not required to wear the eyeglasses while sleeping in order to take measurements over more than a week. Similarly, sentences of the form of "measurements taken over more than 5 days, at least 2 hours a day" refer to a set comprising at least 10 measurements taken over 5 different days, where at least two measurements are taken each day at times separated by at least two hours.

[0046] Utilizing measurements taken of a long period (e.g., measurements taken on "different days") may have an advantage of contributing to the generalizability of a trained model. Measurements taken over the long period likely include measurements taken in different environments and/or measurements taken while the measured user was in various physiological and/or mental states (e.g., before/after meals and/or while the measured user was sleepy/energetic/happy/depressed, etc.). Training a model on such data can improve the performance of systems that utilize the model in the diverse settings often

encountered in real-world use (as opposed to controlled laboratory-like settings). Additionally, taking the measurements over the long period may have the advantage of enabling collection of a large amount of training data that is required for some machine learning approaches (e.g., "deep learning").

[0047] Detecting the physiological response may involve performing various types of calculations by a computer. Optionally, detecting the physiological response may involve performing one or more of the following operations: comparing thermal measurements to a threshold (when the threshold is reached that may be indicative of an occurrence of the physiological response), comparing thermal measurements to a reference time series, and/or by performing calculations that involve a model trained using machine learning methods. Optionally, the thermal measurements upon which the one or more operations are performed are taken during a window of time of a certain length, which may optionally depend on the type of physiological response being detected. In one example, the window may be shorter than one or more of the following durations: five seconds, fifteen seconds, one minute, five minutes, thirty minute, one hour, four hours, one day, or one week. In another example, the window may be longer than one or more of the aforementioned durations. Thus, when measurements are taken over a long period, such as measurements taken over a period of more than a week, detection of the physiological response at a certain time may be done based on a subset of the measurements that falls within a certain window near the certain time; the detection at the certain time does not necessarily involve utilizing all values collected throughout the long period.

[0048] Detecting the physiological response of a user may involve utilizing baseline thermal measurement values, most of which were taken when the user was not experiencing the physiological response. Optionally, detecting the physiological response may rely on observing a change to typical temperatures at one or more ROIs (the baseline), where different users might have different typical temperatures at the ROIs (i.e., different baselines). Optionally, detecting the physiological response may rely on observing a change to a baseline level, which is determined based on previous measurements taken during the preceding minutes and/or hours.

[0049] Detecting a physiological response may involve determining the extent of the physiological response, which may be expressed in various ways that are indicative of the extent of the physiological response, such as: (i) a binary value indicative of whether the user experienced, and/or is experiencing, the physiological response, (ii) a numerical value indicative of the magnitude of the physiological response, (iii) a categorial value indicative of the severity/extent of the physiological response, (iv) an expected change in thermal measurements of an ROI (denoted TH_{ROI} or some variation thereof), and/or (v) rate of change in TH_{ROI}. Optionally, when the physiological response corresponds to a physiological signal (e.g., a heart rate, a breathing rate, and an extent of frontal lobe brain activity), the extent of the physiological response may be interpreted as the value of the physiological signal.

[0050] Herein, "machine learning" methods refers to learning from examples using one or more approaches. Optionally, the approaches may be considered supervised, semi-supervised, and/or unsupervised methods. Examples of machine learning approaches include: decision tree learning,

association rule learning, regression models, nearest neighbors classifiers, artificial neural networks, deep learning, inductive logic programming, support vector machines, clustering, Bayesian networks, reinforcement learning, representation learning, similarity and metric learning, sparse dictionary learning, genetic algorithms, rule-based machine learning, and/or learning classifier systems.

[0051] Herein, a "machine learning-based model" is a model trained using machine learning methods. For brevity's sake, at times, a "machine learning-based model" may simply be called a "model". Referring to a model as being "machine learning-based" is intended to indicate that the model is trained using machine learning methods (otherwise, "model" may also refer to a model generated by methods other than machine learning).

[0052] In some examples, which involve utilizing a machine learning-based model, a computer is configured to detect the physiological response by generating feature values based on the thermal measurements (and possibly other values), and/or based on values derived therefrom (e.g., statistics of the measurements). The computer then utilizes the machine learning-based model to calculate, based on the feature values, a value that is indicative of whether, and/or to what extent, the user is experiencing (and/or is about to experience) the physiological response. Optionally, calculating said value is considered "detecting the physiological response". Optionally, the value calculated by the computer is indicative of the probability that the user has/had the physiological response.

[0053] Herein, feature values may be considered input to a computer that utilizes a model to perform the calculation of a value, such as the value indicative of the extent of the physiological response mentioned above. It is to be noted that the terms "feature" and "feature value" may be used interchangeably when the context of their use is clear. However, a "feature" typically refers to a certain type of value, and represents a property, while "feature value" is the value of the property with a certain instance (sample). For example, a feature may be temperature at a certain ROI, while the feature value corresponding to that feature may be 36.9 °C in one instance and 37.3 °C in another instance.

[0054] A machine learning-based model may be used to detect a physiological response is trained based on data that includes samples. Each sample includes feature values and a label. The feature values may include various types of values. At least some of the feature values of a sample are generated based on measurements of a user taken during a certain period of time (e.g., thermal measurements taken during the certain period of time). Optionally, some of the feature values may be based on various other sources of information described herein. The label is indicative of a physiological response of the user corresponding to the certain period of time. Optionally, the label may be indicative of whether the physiological response occurred during the certain period and/or the extent of the physiological response during the certain period. Additionally or alternatively, the label may be indicative of how long the physiological response lasted. Labels of samples may be generated using various approaches, such as self-report by users, annotation by experts that analyze the training data, automatic annotation by a computer that analyzes the training data and/or analyzes additional data related to the training data, and/or utilizing additional sensors that provide data useful for generating the labels. It is to be noted that herein

when it is stated that a model is trained based on certain measurements (e.g., "a model trained based on TH_{ROI} taken on different days"), it means that the model was trained on samples comprising feature values generated based on the certain measurements and labels corresponding to the certain measurements. Optionally, a label corresponding to a measurement is indicative of the physiological response at the time the measurement was taken.

[0055] Various types of feature values may be generated based on thermal measurements. In one example, some feature values are indicative of temperatures at certain ROIs. In another example, other feature values may represent a temperature change at certain ROIs. The temperature changes may be with respect to a certain time and/or with respect to a different ROI. In order to better detect physiological responses that take some time to manifest, some feature values may describe temperatures (or temperature changes) at a certain ROI at different points of time. Optionally, these feature values may include various functions and/or statistics of the thermal measurements such as minimum/maximum measurement values and/or average values during certain windows of time.

[0056] It is to be noted that when it is stated that feature values are generated based on data comprising multiple sources, it means that for each source, there is at least one feature value that is generated based on that source (and possibly other data). For example, stating that feature values are generated from thermal measurements of first and second ROIs (TH_{ROI1} and TH_{ROI2}, respectively) means that the feature values may include a first feature value generated based on TH_{ROI1} and a second feature value generated based on TH_{ROI2}. Optionally, a sample is considered generated based on measurements of a user (e.g., measurements comprising TH_{ROI1} and TH_{ROI2}) when it includes feature values generated based on the measurements of the user.

[0057] In addition to feature values that are generated based on thermal measurements at least some feature values utilized by a computer (e.g., to detect a physiological response or train a mode) may be generated based on additional sources of data that may affect temperatures measured at various facial ROIs. Some examples of the additional sources include: (i) measurements of the environment such as temperature, humidity level, noise level, elevation, air quality, a wind speed, precipitation, and infrared radiation; (ii) contextual information such as the time of day (e.g., to account for effects of the circadian rhythm), day of month (e.g., to account for effects of the lunar rhythm), day in the year (e.g., to account for seasonal effects), and/or stage in a menstrual cycle; (iii) information about the user being measured such as sex, age, weight, height, and/or body build. Alternatively or additionally, at least some feature values may be generated based on physiological signals of the user obtained by sensors that are not thermal cameras, such as a visible-light camera, a photoplethysmogram (PPG) sensor, an electrocardiogram (ECG) sensor, an electroencephalography (EEG) sensor, a galvanic skin response (GSR) sensor, or a thermistor.

[0058] The machine learning-based model used to detect a physiological response may be trained based on data collected in day-to-day, real world scenarios. As such, the data may be collected at different times of the day, while users perform various activities, and in various environmental conditions. Utilizing such

diverse training data may enable a trained model to be more resilient to the various effects different conditions can have on the values of thermal measurements, and consequently, be able to achieve better detection of the physiological response in real world day-to-day scenarios.

[0059] Since real world day-to-day conditions are not the same all the time, sometimes detection of the physiological response may be hampered by what is referred to herein as "confounding factors". A confounding factor can be a cause of warming and/or cooling of certain regions of the face, which is unrelated to a physiological response being detected, and as such, may reduce the accuracy of the detection of the physiological response. Some examples of confounding factors include: (i) environmental phenomena such as direct sunlight, air conditioning, and/or wind; (ii) things that are on the user's face, which are not typically there and/or do not characterize the faces of most users (e.g., cosmetics, ointments, sweat, hair, facial hair, skin blemishes, acne, inflammation, piercings, body paint, and food leftovers); (iii) physical activity that may affect the user's heart rate, blood circulation, and/or blood distribution (e.g., walking, running, jumping, and/or bending over); (iv) consumption of substances to which the body has a physiological response that may involve changes to temperatures at various facial ROIs, such as various medications, alcohol, caffeine, tobacco, and/or certain types of food; and/or (v) disruptive facial movements (e.g., frowning, talking, eating, drinking, sneezing, and coughing).

[0060] Occurrences of confounding factors may not always be easily identified in thermal measurements. Thus, systems may incorporate measures designed to accommodate for the confounding factors. These measures may involve generating feature values that are based on additional sensors, other than the thermal cameras. These measures may involve refraining from detecting the physiological response, which should be interpreted as refraining from providing an indication that the user has the physiological response. For example, if an occurrence of a certain confounding factor is identified, such as strong directional sunlight that heats one side of the face, the system may refrain from detecting that the user had a stroke. In this example, the user may not be alerted even though a temperature difference between symmetric ROIs on both sides of the face reaches a threshold that, under other circumstances, would warrant alerting the user.

[0061] Training data used to train a model for detecting a physiological response may include a diverse set of samples corresponding to various conditions, some of which involve occurrence of confounding factors (when there is no physiological response and/or when there is a physiological response). Having samples in which a confounding factor occurs (e.g., the user is in direct sunlight or touches the face) can lead to a model that is less susceptible to wrongfully detect the physiological response (which may be considered an occurrence of a false positive) in real world situations.

[0062] After a model is trained, the model may be provided for use by a system that detects the physiological response. Providing the model may involve performing different operations, such as forwarding the model to the system via a computer network and/or a shared computer storage medium, storing the model in a location from which the system can retrieve the model (such as a database and/or cloud-based storage), and/or notifying the system regarding the existence of the model and/or regarding

an update to the model.

[0063]   A model for detecting a physiological response may include different types of parameters. Following are some examples of various possibilities for the model and the type of calculations that may be accordingly performed by a computer in order to detect the physiological response: (a) the model comprises parameters of a decision tree. Optionally, the computer simulates a traversal along a path in the decision tree, determining which branches to take based on the feature values. A value indicative of the physiological response may be obtained at the leaf node and/or based on calculations involving values on nodes and/or edges along the path; (b) the model comprises parameters of a regression model (e.g., regression coefficients in a linear regression model or a logistic regression model). Optionally, the computer multiplies the feature values (which may be considered a regressor) with the parameters of the regression model in order to obtain the value indicative of the physiological response; and/or (c) the model comprises parameters of a neural network. For example, the parameters may include values defining at least the following: (i) an interconnection pattern between different layers of neurons, (ii) weights of the interconnections, and (iii) activation functions that convert each neuron's weighted input to its output activation. Optionally, the computer provides the feature values as inputs to the neural network, computes the values of the various activation functions and propagates values between layers, and obtains an output from the network, which is the value indicative of the physiological response.

[0064]   A user interface (UI) may be utilized to notify the user and/or some other entity, such as a caregiver, about the physiological response and/or present an alert responsive to an indication that the extent of the physiological response reaches a threshold. The UI may include a screen to display the notification and/or alert, a speaker to play an audio notification, a tactile UI, and/or a vibrating UI. "Alerting" about a physiological response of a user may refer to informing about one or more of the following: the occurrence of a physiological response that the user does not usually have (e.g., a stroke, intoxication, and/or dehydration), an imminent physiological response (e.g., an allergic reaction, an epilepsy attack, and/or a migraine), and an extent of the physiological response reaching a threshold (e.g., stress and/or anger reaching a predetermined level).

[0065]   Many physiological responses are manifested through changes at various regions of the human face. For example, measuring temperatures and/or temperature changes may help determine the amount of stress a person is feeling, or extent of an allergic reaction the person has. In another example, measuring temperatures at regions of the face can help determine how a user feels, e.g., whether the user is nervous, calm, or happy. Similarly, visible-light images of the face can be analyzed to determine emotional responses and various physiological signals.

[0066]   Thus, monitoring and analyzing the face can be useful for many health-related and life-logging related applications. However, collecting such data over time, when people are going through their daily activities, can be very difficult. Often, collection of such data involves utilizing cameras that may be bulky, unaesthetic, and/or expensive, which need to be continually pointed at a person's face. Additionally, due to the people's movements in their day-to-day activities, collecting the required

measurements often involves performing various complex image analysis procedures, such as procedures involving image registration and face tracking.

[0067]   Therefore, due to the many applications they may enable, there is a need to be able to collect images (e.g., visible-light images and/or thermal measurements) of various regions of a person's face. Preferably, these images should be collected without requiring extensive effort or discomforting the person.

[0068]   Many people wear eyeglasses throughout their daily lives for various reasons, such as vision correction or to protect from excessive sunlight. Eyeglasses typically do not include sensors that measure the wearer, such as cameras that take images of regions of the face. In order to enable collection of such images, which may be used for various applications, such as detection of physiological responses, a clip-on device may be attached to the eyeglasses.

[0069]   One way in which a user may wear a head-mounted camera (such as CAM or VCAM) involves attaching a clip-on device that houses the camera onto a frame worn by the user, such as an eyeglasses frame. This may enable the user to be selective regarding when to use the head-mounted camera and take advantage of eyeglasses that he or she owns, which may be comfortable and/or esthetically pleasing.

[0070]   The clip-on device may include a body that can be attached and detached, multiple times, from a pair of eyeglasses in order to secure and release the clip-on device from the eyeglasses. The body is a structure that has one or more components fixed to it. For example, the body may have one or more inward-facing camera fixed to it. Additionally, the body may have a wireless communication module fixed to it. Some additional components that may each be optionally fixed to the body include a processor, a battery, and one or more outward-facing cameras.

[0071]   In one example, "eyeglasses" are limited to prescription eyeglasses, prescription sunglasses, plano sunglasses, and/or augmented reality eyeglasses. This means that "eyeglasses" do not refer to helmets, hats, virtual reality devices, and goggles designed to be worn over eyeglasses. Additionally or alternatively, neither attaching the clip-on device to the eyeglasses nor detaching the clip-on device from the eyeglasses should take more than 10 seconds for an average user. This means that manipulating the clip-on device is not a complicated task. Optionally, the body is configured to be detached from the eyeglasses by the user who wears the eyeglasses, who is not a technician, and without using a tool such as a screwdriver or a knife. Thus, the clip-on device may be attached and detached as needed, e.g., enabling the user to attach the clip-on when there is a need to take measurements, and otherwise have it detached.

[0072]   In order to be warn comfortably, possibly for long durations, the clip-on device is a lightweight device, weighing less than 40 g (i.e., the total weight of the body and the components fixed to it is less than 40 g). Optionally, the clip-on device weighs below 20 g and/or below 10 g.

[0073]   The body is a structure to which components (e.g., an inward-facing camera) may be fixed such that the various components do not fall off while the clip-on device is attached to the eyeglasses. Optionally, at least some of the various components that are fixed to the body remain in the same location and/or orientation when the body is attached to the eyeglasses. Herein, stating that a component is "fixed"

to the body is intended to indicate that, during normal use (e.g., involving securing/releasing the clip-on device), the components are typically not detached from the body. This is opposed to the body itself, which in normal use is separated from the eyeglasses frame, and as such, is not considered "fixed" to the eyeglasses frame.

[0074] The body may be a rigid structure made of a material such as plastic, metal, and/or an alloy (e.g., carbon alloy). Optionally, the rigid structure is shaped such that it fits the contours of at least a portion of the frame of the eyeglasses in order to enable a secure and stable attachment to the eyeglasses. Alternatively, the body may be made of a flexible material, such as rubber. Optionally, the flexible body is shaped such that it fits the contours of at least a portion of the frame of the eyeglasses in order to enable a secure and stable attachment to the eyeglasses. Additionally or alternatively, the flexible body may assume the shape of a portion of the frame when it is attached to the eyeglasses.

[0075] The body may utilize various mechanisms in order to stay attached to the eyeglasses. The body may include a clip member configured to being clipped on the eyeglasses. Alternatively, the body may include a magnet configured to attach to a magnet connected to the eyeglasses and/or to a metallic portion of the eyeglasses. In yet another alternative, the body may include a resting tab configured to secure the clip-on to the eyeglasses. In still another alternative, the body may include a retention member (e.g., a clasp, buckle, clamp, fastener, hook, or latch) configured to impermanently couple the clip-on to the eyeglasses. For example, clasp 147 is utilized to secure the clip-on device illustrated in FIG. 15a to the frame of the eyeglasses. And in yet another alternative, the body may include a spring configured to apply force that presses the body towards the eyeglasses. An example of this type of mechanism is illustrated in FIG. 17a where spring 175 is used to apply force that pushes body 170 and secures it in place to frame 176.

[0076] Herein, to "impermanently couple" something means to attach in a way that is easily detached without excessive effort. For example, coupling something by clipping it on or closing a latch is considered impermanently coupling it. Coupling by screwing a screw with a screwdriver, gluing, or welding is not considered impermanently coupling. The latter would be examples of what may be considered to "fix" a component to the body.

[0077] The inward-facing camera is fixed to the body. It takes images of a region of interest on the face of a user who wears the eyeglasses. Optionally, the inward-facing camera remains pointed at the region of interest even when the user's head makes lateral and/or angular movements. The inward-facing camera may be any of the CAMs and/or VCAMs described in this disclosure. Optionally, the inward-facing camera weighs less than 10 g, 5 g or 1 g. Optionally, the inward-facing camera is a thermal camera based on a thermopile sensor, a pyroelectric sensor, or a microbolometer sensor, which may be a FPA sensor.

[0078] The inward-facing camera may include a multi-pixel sensor and a lens, and the sensor plane is tilted by more than 2° relative to the lens plane according to the Scheimpflug principle in order to capture sharper images when the body is attached to the eyeglasses that are worn by a user.

[0079] The clip-one device may include additional components that are fixed to it. For example, the

clip-on device include a wireless communication module fixed to the body which transmits measurements (e.g., images and/or thermal measurements) taken by one or more of the cameras that are fixed to the body. Optionally, the clip-on device may include a battery fixed to the body, which provides power to one or more components fixed to the body. Optionally, the clip-on device may include a processor that controls the operation of one or more of the components fixed to the body and/or processes measurements taken by the camera fixed to the body.

[0080]   A computer may receive measurements taken by the inward-facing camera (and possibly other cameras fixed to the body), and utilizes the measurements to detect a physiological response. Optionally, the computer is not fixed to the body. For example, the computer may belong to a device of the user (e.g., a smartphone or a smartwatch), or the computer may be a cloud-based server. Optionally, the computer receives, over a wireless channel, the measurements, which are sent by the wireless communication module.

[0081]   The following are various examples of using different types of inward- and outward-facing cameras that are fixed to the body, which may be used to take images of various regions of interest on the face of the user who wears the eyeglasses. It is to be noted that while the discussion below generally refers to a single "inward-facing camera" and/or a single "outward-facing camera", the clip-on device may include multiple inward- and/or outward-facing cameras.

[0082]   In some examples, the inward-facing camera is a thermal camera. Optionally, when the body is attached to the eyeglasses, the thermal camera is located less than 5 cm from the user's face. Optionally, measurements taken by the thermal camera are transmitted by the wireless communication module and are received by a computer that uses them to detect a physiological response of the user. In one example, when the body is attached to the eyeglasses, the optical axis of the thermal camera is above 20^0 from the Frankfort horizontal plane, and the thermal camera takes thermal measurements of a region on the user's forehead. In another example, when the body is attached to the eyeglasses, the thermal camera takes thermal measurements of a region on the user's nose. In yet another example, when the body is attached to the eyeglasses, the thermal camera takes thermal measurements of a region on a periorbital area of the user.

[0083]   Alternatively, when the body is attached to the eyeglasses, the thermal camera is located below eye-level of a user who wears the eyeglasses and at least 2 cm from the vertical symmetry axis that divides the user's face (i.e., the axis the goes down the center of the user's forehead and nose). Additionally, when the body is attached to the eyeglasses, the inward-facing thermal camera takes thermal measurements of a region on at least one of the following parts of the user's face: upper lip, lips, and a cheek. Optionally, measurements taken by the thermal camera are transmitted by the wireless communication module and are received by a computer that uses them to detect a physiological response of the user.

[0084]   In another example, the inward-facing camera is a visible-light camera. Optionally, when the body is attached to the eyeglasses, the visible-light camera is located less than 10 cm from the user's face.

Optionally, images taken by the visible-light camera are transmitted by the wireless communication module and are received by a computer that uses them to detect a physiological response of the user. Optionally, the computer detects the physiological response based on facial skin color changes (FSCC) that are recognizable in the images. In one example, when the body is attached to the eyeglasses, the optical axis of the visible-light camera is above 20° from the Frankfort horizontal plane, and the visible-light camera takes images of a region located above the user's eyes. In another example, when the body is attached to the eyeglasses, the visible-light camera takes images of a region on the nose of a user who wears the eyeglasses. In still another example, the computer detects the physiological response based on facial expressions, and when the body is attached to the eyeglasses, the visible-light camera takes images of a region above or below the user's eyes.

[0085]   Alternatively, when the body is attached to the eyeglasses, the visible-light camera takes images of a region on an eye (IM_E) of a user who wears the eyeglasses, and is located less than 10 cm from the user's face. Optionally, the images are transmitted by the wireless communication module and are received by a computer that detects a physiological response based in IM_E.

[0086]   In one example, the computer detects the physiological response based on color changes to certain parts of the eye, such as the sclera and/or the iris. Due to the many blood vessels that are close to the surface of the eye, physiological responses that are manifested through changes to the blood flow (e.g., a cardiac pulse and certain emotional responses), may cause recognizable changes to the color of the certain parts of the eye. The various techniques described in this disclosure for detecting a physiological response based on FSCC that is recognizable in images can be applied by one skilled in the art to detect a physiological response based on color changes to the sclera and/or iris; while the sclera and iris are not the same color as a person's skin, they too exhibit blood flow-related color changes that are qualitatively similar to FSCC, and thus may be analyzed using similar techniques to the techniques used to analyze FSCC involving the forehead, nose, and/or cheeks.

[0087]   In another example, IM_E may be utilized to determine the size of the pupil, which may be utilized by the computer to detect certain emotional responses (such as based on the assumption that the pupil's response reflects emotional arousal associated with increased sympathetic activity).

[0088]   If needed as part of the computer's detection of the physiological response, identifying which portions of IM_E correspond to certain parts of the eye (e.g., the sclera or iris) can be done utilizing various image processing techniques known in the art. For example, identifying the iris and pupil size may be done using the techniques described in US patent application US20060147094, or in Hayes, Taylor R., and Alexander A. Petrov. "Mapping and correcting the influence of gaze position on pupil size measurements." Behavior Research Methods 48.2 (2016): 510-527. Additionally, due to the distinct color differences between the skin, the iris, and the sclera, identification of the iris and/or the white sclera can be easily done by image processing methods known in the art.

[0089]   In one example, the inward-facing camera is a visible-light camera; when the body is attached to the eyeglasses, the visible-light camera is located below eye-level of a user who wears the eyeglasses, and

at least 2 cm from the vertical symmetry axis that divides the user's face. The visible-light camera takes images (IM_{ROI}) of a region on the upper lip, lips, and/or a cheek. Optionally, IM_{ROI} are transmitted by the wireless communication module and are received by a computer that uses them to detect a physiological response. In one example, the physiological response is an emotional response, which is detected based on extracting facial expressions from IM_{ROI}. In another example, the physiological response is an emotional response, which is detected based on FSCC recognizable in IM_{ROI}. In still another example, the physiological response, which is detected based FSCC recognizable in IM_{ROI}, is heart rate and/or breathing rate.

[0090]   The body may include an outward-facing camera that may be utilized to provide measurements that may be used to account for various environmental interferences that can decrease detections of the physiological response of a user who wears the eyeglasses. Optionally, the outward-facing camera is a head-mounted camera. Optionally, the outward-facing camera is fixed to the body.

[0091]   In one example, the inward-facing camera is a thermal camera, and when the body is attached to the eyeglasses, the thermal camera is located less than 10 cm from the face of the user who wears the eyeglasses, and takes thermal measurements of a region of interest (TH_{ROI}) on the face of the user. In this example, an outward-facing head-mounted thermal camera takes thermal measurements of the environment (TH_{ENV}). The wireless communication module transmits TH_{ROI} and TH_{ENV} to a computer that detects an emotional response of the user based on TH_{ROI} and TH_{ENV}. Optionally, the computer utilizes TH_{ENV} to account for thermal interferences from the environment, as discussed elsewhere herein.

[0092]   In another example, the inward-facing camera is a visible-light camera, and when the body is attached to the eyeglasses, the visible-light camera is located less than 10 cm from the face of the user who wears the eyeglasses and takes images of a region of interest (IM_{ROI}) on the face of the user. In this example, an outward-facing head-mounted visible-light camera takes images of the environment (IM_{ENV}). The wireless communication module transmits IM_{ROI} and IM_{ENV} to a computer that detects an emotional response of the user based on IM_{ROI} and IM_{ENV}. Optionally, the computer detects the physiological response based on FSCC recognizable in IM_{ROI}, and utilizes IM_{ENV} to account for variations in ambient light, as discussed elsewhere herein.

[0093]   Inward-facing cameras attached to the body may be utilized for additional purposes, beyond detection of physiological responses. In one example, the inward-facing camera is a visible-light camera, and the clip-on device includes a second visible-light camera that is also fixed to the body. Optionally, the visible-light camera and/or the second visible-light camera are light field cameras. Optionally, when the body is attached to the eyeglasses, the first and second visible-light cameras are located less than 10 cm from the user's face, and take images of a first region above eye-level and a second region on the upper lip (IM_{ROI} and IM_{ROI2}, respectively). Optionally, the wireless communication module transmits IM_{ROI} and IM_{ROI2} to a computer that generates an avatar of the user based on IM_{ROI} and IM_{ROI2}. Some of the various approaches that may be utilized to generate the avatar based on IM_{ROI} and IM_{ROI2} are described in co-pending US patent publication 2016/0360970.

[0094]   The clip-on device may involve devices of various shapes, sizes, and/or locations of attachment to the eyeglasses. FIG. 14a to FIG. 18 illustrate some examples of clip-on devices. When the body is attached to the eyeglasses, most of the clip-on device may be located in front of the frame of the eyeglasses, as illustrated in FIG. 14b, FIG. 15b, and FIG. 18, or alternatively, most of the clip-on device may be located behind the frame, as illustrated in FIG. 16b and FIG. 17b. Some clip-on devices may include a single unit, such as illustrated in FIG. 15a and FIG. 17a. While other clip-on devices may include multiple units (which each may optionally be considered a separate clip-on device). Examples of multiple units being attached to the frame are illustrated in FIG. 14b, FIG. 16b, and FIG. 18. The following is a more detailed discussion regarding examples illustrated in the figures mentioned above.

[0095]   FIG. 14a, FIG. 14b, and FIG. 14c illustrate two right and left clip-on devices comprising bodies 141 and 142, respectively, which are configured to attached/detached from an eyeglasses frame 140. The body 142 has multiple inward-facing cameras fixed to it, such as camera 143 that points at a region on the lower part of the face (such as the upper lip, mouth, nose, and/or cheek), and camera 144 that points at the forehead. The body 142 may include other electronics 145, such as a processor, a battery, and/or a wireless communication module. The bodies 141 and 142 of the left and right clip-on devices may include additional cameras illustrated in the drawings as black circles.

[0096]   In one another example, the eyeglasses include left and right lenses, and when the body is attached to the eyeglasses, most of the volume of the clip-on device is located to the left of the left lens or to the right of the right lens. Optionally, the inward-facing camera takes images of at least one of: a region on the nose of a user wearing the eyeglasses, and a region on the mouth of the user. Optionally, a portion of the clip-on device that is located to the left of the left lens or to the right of the right lens does not obstruct the sight of the user when looking forward.

[0097]   FIG. 15a and FIG. 15b illustrate a clip-on device that includes a body 150, to which two head-mounted cameras are fixed: a head-mounted camera 148 that points at a region on the lower part of the face (such as the nose), and a head-mounted camera 149 that points at the forehead. The other electronics (such as a processor, a battery, and/or a wireless communication module) are located inside the body 150. The clip-on device is attached and detached from the frame of the eyeglasses with the clasp 147.

[0098]   In one example, when the body is attached to the eyeglasses, most of the volume of the clip-on device is located above the lenses of the eyeglasses, and the inward-facing camera takes images of a region on the forehead of a user who wears the eyeglasses. Optionally, a portion of the clip-on device that is located above the lenses of the eyeglasses does not obstruct the sight of the user when looking forward.

[0099]   While the clip-on device may often have a design intended to reduce the extent to which it sticks out beyond the frame, the clip-on device may include various protruding arms. Optionally, these arms may be utilized in order to position one or more cameras in a position suitable for taking images of certain regions of the face. FIG. 18 illustrates right and left clip-on devices that include bodies 153 and 154, respectively, which are configured to attached/detached from an eyeglasses frame. These bodies have protruding arms that hold the head-mounted cameras. Head-mounted camera 155 measures a region

on the lower part of the face, head-mounted camera 156 measures regions on the forehead. The left clip-on device also includes other electronics 157 (such as a processor, a battery, and/or a wireless communication module). The clip-on devices illustrated in this figure may include additional cameras illustrated in the drawings as black circles.

[0100]   In other examples, at least a certain portion of the clip-on device is located behind the eyeglasses' frame. Thus, when the clip-on device is attached to the eyeglasses, they may remain aesthetically pleasing, and attaching the clip-on device may cause little or no blocking of the user's vision. FIG. 16b and FIG. 17b illustrate two examples of clip-on devices that are mostly attached behind the frame. The following are some additional examples in which a portion of the clip-on device may be located behind the frame.

[0101]   FIG. 16a and FIG. 16b illustrate two, right and left, clip-on devices with bodies 160 and 161, respectively, configured to be attached behind an eyeglasses frame 165. The body 160 has various components fixed to it which include: an inward-facing head-mounted camera 162 pointed at a region below eye-level (such as the upper lip, mouth, nose, and/or cheek), an inward-facing head-mounted camera 163 pointed at a region above eye-level (such as the forehead), and other electronics 164 (such as a processor, a battery, and/or a wireless communication module). The right and left clip-on devices may include additional cameras illustrated in the drawings as black circles.

[0102]   FIG. 17a and FIG. 17b illustrate a single-unit clip-on device that includes the body 170, which is configured to be attached behind the eyeglasses frame 176. The body 170 has various cameras fixed to it, such as head-mounted cameras 171 and 172 that are pointed at regions on the lower part of the face (such as the upper lip, mouth, nose, and/or cheek), and head-mounted cameras 173 and 174 that are pointed at the forehead. The spring 175 is configured to apply force that holds the body 170 to the frame 176. Other electronics 177 (such as a processor, a battery, and/or a wireless communication module), may also be fixed to the body 170. The clip-on device may include additional cameras illustrated in the drawings as black circles.

[0103]   When the body is attached to the eyeglasses, more than 50% of the out-facing surface of the clip-on device may be located behind the eyeglasses frame. Optionally, a portion of the clip-on device that is located behind the eyeglasses frame is occluded from a viewer positioned directly opposite to the eyeglasses, at the same height as the eyeglasses. Thus, a portion of the clip-on device that is behind the frame might not be visible to other people from many angles, which can make the clip-on device less conspicuous and/or more aesthetically pleasing. Optionally, a larger portion of the clip-on device is behind the frame when the body is attached to the eyeglasses, such as more than 75% or 90% of the out-facing surface.

[0104]   Various biological processes cause facial skin color changes (FSCC). FSCC are typically a result of changes in the concentration levels of hemoglobin and blood oxygenation under a user's facial skin due to a physiological response that involves changes in the user's emotional state and/or changes in the user's physical state, and/or due to normal biological processes. These changes in the concentration

levels of hemoglobin and blood oxygenation can cause subtle changes in the hue and saturation components of the user's facial skin color.

[0105] There are well known methods to infer the emotional state and various physiological parameters (such as heart rate and breathing rate) of a person based on FSCC. For example, US patent application 20160098592 describes extracting emotions based on hemoglobin concentration changes (HCC) from red, green and blue (RGB) video. As another example, US patents number 8768438, 8977347, 8855384, 9020185, 8617081 and US patent application number 20130215244 describe extracting heart rate and related parameters from RGB video, near-IR video, and multi-spectral video streams. As still another example, the following three publications explain how FSCC (resulting from concentration changes of hemoglobin and/or oxygenation) are related to emotions: (i) Ramirez, Geovany A., et al. "Color analysis of facial skin: Detection of emotional state" in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014; (ii) Wang, Su-Jing, et al. "Micro-expression recognition using color spaces", in IEEE Transactions on Image Processing 24.12 (2015): 6034-6047; and (iii) Jimenez, Jorge, et al. "A practical appearance model for dynamic facial color", in ACM Transactions on Graphics (TOG). Vol. 29. No. 6. ACM, 2010.

[0106] All the prior art methods for detecting physiological responses based on FSCC analyze video of the face taken from a camera that is remote of the user, and thus these methods typically face challenges involving face tracking and image registration. Additionally, the prior art methods do not utilize measurements of illumination interferences (such as variations in ambient light), which may, in some cases, affect the accuracy of detections of physiological responses based on FSCC. Consequently, accuracy of the prior art may be susceptible to errors that degrade the performance of detections of physiological responses based on FSCC, especially in real world, less controlled settings where the user moves and the ambient light varies.

[0107] Some aspects of this disclosure involve detection of a physiological response based on facial skin color changes (FSCC) recognizable in images taken with an inward-facing head-mounted visible-light camera (VCAM$_{in}$). The prior art approaches mentioned above receive images from visible-light cameras that are not head-mounted, which means that they must use extensive image registration (in order to align the images) and are affected by non-uniformity of the visible-light sensor (because the same point on the target may be measured in consecutive images by different pixels). The image stability obtained from VCAM$_{in}$ is better than the image stability obtained from a visible-light camera that is not head-mounted. Mounting the camera to the face reduces the systematic error and enables better filtering of random errors, such as by averaging multiple measurements of the same pixel or pixels, and/or summing multiple measurements of the same pixel to improve the signal to noise ratio. Additionally, when there is a need to perform image registration on images obtained from VCAM$_{in}$, the transformation model (to relate one image to another) may be restricted to the maximum possible relative movement between VCAM$_{in}$ and the ROI, which is confined as a result of coupling the camera to the frame. This restricted transformation model is much less computational intensive than a full transformation model as in the

prior art configurations where the camera is not head-mounted.

[0108] Various physiological responses may be detected based on Facial skin color changes (FSCC) that occur on a user's face. In one embodiment, a system configured to detect a physiological response based on FSCC includes at least an inward-facing head-mounted visible-light camera ($VCAM_{in}$) and a computer. The system may optionally include additional elements such as a frame and additional inward-facing camera(s) and/or outward-facing camera(s).

[0109] FIG. 20 illustrates one embodiment of the system configured to detect a physiological response based on FSCC. The system includes a frame 735 (e.g., an eyeglasses frame) to which various cameras are physically coupled. These cameras include visible-light cameras 740, 741, 742, and 743, which may each take images of regions on the user's cheeks and/or nose. Each of these cameras may possibly be $VCAM_{in}$, which is discussed in more detail below. Another possibility for $VCAM_{in}$ is camera 745 that takes images of a region on the user's forehead and is coupled to the upper portion of the frame. Visible-light camera 737, which takes images of the environment (IM_{ENV}), is an example of $VCAM_{out}$ discussed below, which may optionally be included. Additional cameras that may optionally be included are outward-facing thermal camera 738 (which may be used to take TH_{ENV} mentioned below) and inward-facing thermal camera 739 (which may be used to take TH_{ROI2} mentioned below).

[0110] $VCAM_{in}$ is worn on the user's head and takes images of a region of interest (IM_{ROI}) on the user's face. Depending on the physiological response being detected, the ROI may cover various regions on the user's face. In one example, the ROI is on a cheek of the user, a region on the user's nose, and/or a region on the user's forehead. Optionally, $VCAM_{in}$ does not occlude the ROI, is located less than 10 cm from the user's face, and weighs below 10 g. The ROI is illuminated by ambient light. Optionally, the system does not occlude the ROI, and the ROI is not illuminated by a head-mounted light source. Alternatively, the ROI may be illuminated by a head-mounted light source that is weaker than the ambient light.

[0111] The computer detects the physiological response based on IM_{ROI} by relying on effects of FSCC that are recognizable in IM_{ROI}. Herein, sentences of the form "FSCC recognizable in IM_{ROI}" refer to effects of FSCC that may be identified and/or utilized by the computer, which are usually not recognized by the naked eye. The FSCC phenomenon may be utilized to detect various types of physiological responses. The physiological response that is detected may involve an expression of emotional response of the user. For example, the computer may detect whether the user's emotional response is neutral, positive, or negative. In another example, the computer may detect an emotional response that falls into a more specific category such as distress, happiness, anxiousness, sadness, frustration, intrigue, joy, disgust, anger, etc. Optionally, the expression of the emotional response may involve the user making a facial expression and/or a microexpression (whose occurrence may optionally be detected based on IM_{ROI}). Alternatively, detecting the physiological response involves determining one or more physiological signals of the user, such as a heart rate (which may also be referred to as "cardiac pulse"), heart rate variability, and/or a breathing rate.

[0112] IM$_{ROI}$ are images generated based on ambient light illumination that is reflected from the user's face. Variations in the reflected ambient light may cause FSCC that are unrelated to the physiological response being detected, and thus possibly lead to errors in the detection of the physiological response. In some embodiments, the system includes an outward-facing head-mounted visible-light camera (VCAM$_{out}$), which is worn on the user's head, and takes images of the environment (IM$_{ENV}$). Optionally, VCAM$_{out}$ is located less than 10 cm from the user's face and weighs below 10 g. Optionally, VCAM$_{out}$ may include optics that provide it with a wide field of view. Optionally, the computer detects the physiological response based on both IM$_{ROI}$ and IM$_{ENV}$. Given that IM$_{ENV}$ is indicative of illumination towards the face and IM$_{ROI}$ is indicative of reflections from the face, utilizing IM$_{ENV}$ in the detection of the physiological response can account, at least in part, for variations in ambient light that, when left unaccounted, may possibly lead to errors in detection of the physiological response.

[0113] It is noted that the system may include multiple VCAM$_{in}$ configured to take images of various ROIs on the face, IM$_{ROI}$ may include images taken from the multiple VCAM$_{in}$, and multiple VCAM$_{out}$ located at different locations and/or orientation relative to the face may be used to take images of the environment.

[0114] VCAM$_{in}$ and/or VCAM$_{out}$ may be physically coupled to a frame, such as an eyeglasses frame or an augmented realty device frame. Optionally, the angle between the optical axes of VCAM$_{in}$ and VCAM$_{out}$ is known to the computer, and may be utilized in the detection of the physiological response. Optionally, the angle between the optical axes of VCAM$_{in}$ and VCAM$_{out}$ is fixed.

[0115] Due to the proximity of VCAM$_{in}$ to the face, there may be an acute angle between the optical axis of VCAM$_{in}$ and the ROI (e.g., when the ROI includes a region on the forehead). In order to improve the sharpness of IM$_{ROI}$, VCAM$_{in}$ may be configured to operate in a way that takes advantage of the Scheimpflug principle. For example, VCAM$_{in}$ may include a sensor and a lens; the sensor plane is tilted by a fixed angle greater than 2^0 relative to the lens plane according to the Scheimpflug principle in order to capture a sharper image when VCAM$_{in}$ is worn by the user (where the lens plane refers to a plane that is perpendicular to the optical axis of the lens, which may include one or more lenses). Optionally, VCAM$_{in}$ does not occlude the ROI. In another arrangement, VCAM$_{in}$ includes a sensor, a lens, and a motor; the motor tilts the lens relative to the sensor according to the Scheimpflug principle. The tilt improves the sharpness of IM$_{ROI}$ when VCAM$_{in}$ is worn by the user.

[0116] In addition to capturing images in the visible spectrum, light may be captured in the near infrared spectrum (NIR). VCAM$_{in}$ and/or VCAM$_{out}$ may include optics and sensors that capture light rays in at least one of the following NIR spectrum intervals: 700-800 nm, 700-900 nm, 700-1,000 nm. Optionally, the computer may utilize data obtained in a NIR spectrum interval to detect the physiological response (in addition to or instead of data obtained from the visible spectrum). Optionally, the sensors may be CCD sensors designed to be sensitive in the NIR spectrum and/or CMOS sensors designed to be sensitive in the NIR spectrum.

[0117] The computer may utilize various approaches in order to detect the physiological response

based on IM_{ROI}. Some examples of how such a detection may be implemented are provided in the prior art references mentioned above, which rely on FSCC to detect the physiological response. It is to be noted that while the prior art approaches involve analysis of video obtained from cameras that are not head-mounted, are typically more distant from the ROI than $VCAM_{in}$, and are possibly at different orientations relative to the ROI, the computational approaches described in the prior art used to detect physiological responses can be readily adapted by one skilled in the art to handle IM_{ROI}. In some cases, systems described herein may provide video in which a desired signal is more easily detectable compared to some of the prior art approaches. For example, given the short distance from $VCAM_{in}$ to the ROI, the ROI is expected to cover a larger portion of the images in IM_{ROI} compared to images obtained by video cameras in some of the prior art references. Additionally, due to the proximity of $VCAM_{in}$ to the ROI, additional illumination that is required in some prior art approaches, such as illuminating the skin for a pulse oximeter to obtain a photoplethysmographic (PPG) signal, may not be needed. Furthermore, given $VCAM_{in}$'s fixed location and orientation relative to the ROI (even when the user makes lateral and/or angular movements), many pre-processing steps that need to be implemented by the prior art approaches, such as image registration and/or face tracking, are extremely simplified or may be foregone altogether.

[0118] IM_{ROI} may undergo various preprocessing steps prior to being used by the computer to detect the physiological response and/or as part of the process of the detection of the physiological response. Some non-limiting examples of the preprocessing include: normalization of pixel intensities (e.g., to obtain a zero-mean unit variance time series signal), and conditioning a time series signal by constructing a square wave, a sine wave, or a user defined shape, such as that obtained from an ECG signal or a PPG signal as described in US patent number 8617081. Additionally or alternatively, feature values may be generated based on a single image or a sequence of images. In some examples, generation of feature values from one or more images may involve utilization of some of the various approaches described in this disclosure for generation of high-level and/or low-level image-based features.

[0119]   The following is a discussion of some approaches that may be utilized by the computer to detect the physiological response based on IM_{ROI}. Additionally, implementation-related details may be found in the provided references and the references cited therein. Optionally, IM_{ENV} may also be utilized by the computer to detect the physiological response (in addition to IM_{ROI}), as explained in more detail below.

[0120]   The physiological response may be detected using signal processing and/or analytical approaches. Optionally, these approaches may be used for detecting repetitive physiological signals (e.g., a heart rate, heart rate variability, or a breathing rate) in IM_{ROI} taken during a certain period. Optionally, the detected physiological response represents the value of the physiological signal of the user during the certain period.

[0121]   In one example, US patent number 8768438, titled "Determining cardiac arrhythmia from a video of a subject being monitored for cardiac function", describes how a heart rate may be determined based on FSCC, which are represented in a PPG signal obtained from video of the user. In this example, a time series signal is generated from video images of a subject's exposed skin, and a reference signal is

used to perform a constrained source separation (which is a variant of ICA) on the time series signals to obtain the PPG signal. Peak-to-peak pulse points are detected in the PPG signal, which may be analyzed to determine parameters such as heart rate, heart rate variability, and/or to obtain peak-to-peak pulse dynamics that can be indicative of conditions such as cardiac arrhythmia.

[0122] In another example, US patent number 8977347, titled "Video-based estimation of heart rate variability", describes how a times-series signal similar to the one described above may be subjected to a different type of analysis to detect the heart rate variability. In this example, the time series data are de-trended to remove slow non-stationary trends from the signal and filtered (e.g., using bandpass filtering). Following that, low frequency and high frequency components of the integrated power spectrum within the time series signal are extracted using Fast Fourier Transform (FFT). A ratio of the low and high frequency of the integrated power spectrum within these components is computed. And analysis of the dynamics of this ratio over time is used to estimate heart rate variability.

[0123] In yet another example, US patent number 9020185, titled "Systems and methods for non-contact heart rate sensing", describes how a times-series signals obtained from video of a user can be filtered and processed to separate an underlying pulsing signal by, for example, using an ICA algorithm. The separated pulsing signal from the algorithm can be transformed into frequency spacing data using FFT, in which the heart rate can be extracted or estimated.

[0124] The physiological response may be detected using machine learning-based methods. Optionally, these approaches may be used for detecting expressions of emotions and/or values of physiological signals.

[0125] Generally, machine learning-based approaches involve training a model on samples, with each sample including: feature values generated based on IM_{ROI} taken during a certain period, and a label indicative of the physiological response during the certain period. Optionally, the model may be personalized for a user by training the model on samples including: feature values generated based on IM_{ROI} of the user, and corresponding labels indicative of the user's respective physiological responses. Some of the feature values in a sample may be generated based on other sources of data (besides IM_{ROI}), such as measurements of the user generated using thermal cameras, movement sensors, and/or other physiological sensors, and/or measurements of the environment. Optionally, IM_{ROI} of the user taken during an earlier period may serve as a baseline to which to compare. Optionally, some of the feature values may include indications of confounding factors, which may affect FSCC, but are unrelated to the physiological response being detected. Some examples of confounding factors include touching the face, thermal radiation directed at the face, and consuming certain substances such as a medication, alcohol, caffeine, or nicotine.

[0126] Training the model may involve utilization of various training algorithms known in the art (e.g., algorithms for training neural networks and/or other approaches described herein). After the model is trained, feature values may be generated for IM_{ROI} for which the label (physiological response) is unknown, and the computer can utilize the model to detect the physiological response based on these

feature values.

[0127] The model may be trained based on data that includes measurements of the user, in which case it may be considered a personalized model of the user. Alternatively, the model may be trained based on data that includes measurements of one or more other users, in which case it may be considered a general model.

[0128] In order to achieve a robust model, which may be useful for detecting the physiological response in various conditions, the samples used in the training may include samples based on IM_{ROI} taken in different conditions and include samples with various labels (e.g., expressing or not expressing certain emotions, or different values of physiological signals). Optionally, the samples are generated based on IM_{ROI} taken on different days.

[0129] The following are four examples of different compositions of samples that may be used when training the model. The "measured user" in the four examples below may be "the user" who is mentioned above (e.g., when the model is a personalized model that was trained on data that includes measurements of the user), or a user from among one or more other users (e.g., when the model is a general model that was trained on data that includes measurements of the other users). In a first example, the system does not occlude the ROI, and the model is trained on samples generated from a first set of IM_{ROI} taken while the measured user was indoors and not in direct sunlight, and is also trained on other samples generated from a second set of IM_{ROI} taken while the measured user was outdoors, in direct sunlight. In a second example, the model is trained on samples generated from a first set of IM_{ROI} taken during daytime, and is also trained on other samples generated from a second set of IM_{ROI} taken during nighttime. In a third example, the model is trained on samples generated from a first set of IM_{ROI} taken while the measured user was exercising and moving, and is also trained on other samples generated from a second set of IM_{ROI} taken while the measured user was sitting and not exercising. And a fourth example, the model is trained on samples generated from a first set of IM_{ROI} taken less than 30 minutes after the measured user had an alcoholic beverage, and is also trained on other samples generated from a second set of IM_{ROI} taken on a day in which the measured user did not have an alcoholic beverage.

[0130] Labels for the samples may be obtained from various sources. The labels may be obtained utilizing one or more sensors that are not $VCAM_{in}$. In one example, a heart rate and/or heart rate variability may be measured using an ECG sensor. In another example, the breathing rate may be determined using a smart shirt with sensors attached to the chest (e.g., a smart shirt by Hexoskin®). In yet another example, a type emotional response of the user may be determined based on analysis of a facial expression made by the user, analysis of the user's voice, analysis of thermal measurements of regions of the face of the user, and/or analysis of one or more of the following sensor-measured physiological signals of the user: a heart rate, heart rate variability, breathing rate, and galvanic skin response.

[0131] Alternatively, a label describing an emotional response of the user may be inferred. In one example, the label may be based on semantic analysis of a communication of the user, which is indicative of the user's emotional state at the time IM_{ROI} were taken. In another example, the label may be generated

in a process in which the user is exposed to certain content, and a label is determined based on an expected emotional response corresponding to the certain content (e.g., happiness is an expected response to a nice image while distress is an expected response to a disturbing image).

[0132] Due to the nature of the physiological responses being detected and the type of data (video images), a machine learning approach that may be applied is "deep learning". The model may include parameters describing multiple hidden layers of a neural network. Optionally, the model may include a convolution neural network (CNN). In one example, the CNN may be utilized to identify certain patterns in the video images, such as the patterns of the reflected FSCC due to the physiological response. Optionally, detecting the physiological response may be done based on multiple, possibly successive, images that display a certain pattern of change over time (i.e., across multiple frames), which characterizes the physiological response being detected. Thus, detecting the physiological response may involve retaining state information that is based on previous images. Optionally, the model may include parameters that describe an architecture that supports such a capability. In one example, the model may include parameters of a recurrent neural network (RNN), which is a connectionist model that captures the dynamics of sequences of samples via cycles in the network's nodes. This enables RNNs to retain a state that can represent information from an arbitrarily long context window. In one example, the RNN may be implemented using a long short-term memory (LSTM) architecture. In another example, the RNN may be implemented using a bidirectional recurrent neural network architecture (BRNN).

[0133] Some of the prior art references mentioned herein provide additional detailed examples of machine learning-based approaches that may be utilized to detect the physiological response (especially in the case in which it corresponds to an emotional response). In one example, Ramirez, et al. ("Color analysis of facial skin: Detection of emotional state") describe detection of an emotional state using various machine learning algorithms including decision trees, multinomial logistic regression, and latent-dynamic conditional random fields. In another example, Wang, et al. ("Micro-expression recognition using color spaces") describe various feature extraction methods and pixel color value transformations, which are used to generate inputs for a support vector machine (SVM) classifier trained to identify microexpressions.

[0134] As mentioned above, IM_{ENV} may be utilized in the detection of the physiological response to account, at least in part, for illumination interferences that may lead to errors in the detection of the physiological response. There are different ways in which IM_{ENV} may be utilized for this purpose.

[0135] When variations in IM_{ENV} reach a certain threshold (e.g., which may correspond to ambient light variations above a certain extent), the computer may refrain from detecting the physiological response.

[0136] Alternatively, IM_{ENV} may be utilized to normalize IM_{ROI} with respect to the ambient light. For example, the intensity of pixels in IM_{ROI} may be adjusted based on the intensity of pixels in IM_{ENV} when IM_{ROI} were taken. US patent application number 20130215244 describes a method of normalization in which values of pixels from a region that does not contain a signal (e.g., background regions that include a different body part of the user or an object behind the user) are subtracted from regions of the image

that contain the signal of the physiological response. While the computational approach described therein may be applied here, the exact setup described therein may not work well in some cases due to the close proximity of $VCAM_{in}$ to the face and the fact that $VCAM_{in}$ is head-mounted. Thus, it may be advantageous to subtract a signal from the environment (IM_{ENV}) that is obtained from $VCAM_{out}$, which may more accurately represent the ambient light illuminating the face.

[0137] It is to be noted that training data that includes a ground-truth signal (i.e., values of the true physiological response corresponding to IM_{ROI} and IM_{ENV}) may be utilized to optimize the normalization procedure used to correct IM_{ROI} with respect to the ambient light measured in IM_{ENV}. For example, such optimization may be used to determine parameter values of a function that performs the subtraction above, which lead to the most accurate detections of the physiological response.

[0138] In still another alternative, IM_{ENV} may be utilized to generate feature values in addition to IM_{ROI}. Optionally, at least some of the same types of feature values generated based on IM_{ROI} may also be generated based on IM_{ENV}. Optionally, at least some of the feature values generated based on IM_{ENV} may relate to portions of images, such as average intensity of patches of pixels in IM_{ENV}.

[0139] By utilizing IM_{ENV} as inputs used for the detection of the physiological response, a machine learning-based model may be trained to be robust, and less susceptible, to environmental interferences such as ambient light variations. For example, if the training data used to train the model includes samples in which no physiological response was present (e.g., no measured emotional response or microexpression was made), but some ambient light variations might have introduced some FSCC-related signal, the model will be trained such that feature values based on IM_{ENV} are used to account for such cases. This can enable the computer to negate, at least in part, the effects of such environmental interferences, and possibly make more accurate detections of the physiological response.

[0140] The computer may receive an indication indicative of the user consuming a confounding substance that is expected to affect FSCC (e.g., alcohol, drugs, certain medications, and/or cigarettes). The computer detects the physiological response, while the consumed confounding substance affects FSCC, based on: IM_{ROI}, the indication, and a model that was trained on: a first set of IM_{ROI} taken while the confounding substance affected FSCC, and a second set of IM_{ROI} taken while the confounding substance did not affect FSCC.

[0141] Prior art FSCC systems are sensitive to user movements and do not operate well while the user is running. This is because state-of-the-art FSCC systems use hardware and automatic image trackers that are not accurate enough to crop correctly the ROI from the entire image while running, and the large errors in cropping the ROI are detrimental to the performances of the FSCC algorithms. Contrary to the prior art FSCC systems, the disclosed $VCAM_{in}$ remains pointed at its ROI also when the user's head makes angular and lateral movements, and thus the complicated challenges related to image registration and ROI tracking are much simplified or even eliminated. Therefore, systems based on $VCAM_{in}$ (such as the one illustrated in FIG. 20) may detect the physiological response (based on FSCC) also while the user is running.

[0142] VCAM$_{in}$ may be pointed at different regions on the face. In a first arrangement, the ROI is on the forehead, VCAM$_{in}$ is located less than 10 cm from the user's face, and optionally the optical axis of VCAM$_{in}$ is above 20° from the Frankfort horizontal plane. In a second arrangement, the ROI is on the nose, and VCAM$_{in}$ is located less than 10 cm from the user's face. Because VCAM$_{in}$ is located close to the face, it is possible to calculate the FSCC based on a small ROI, which is irrelevant to the non-head-mounted prior arts that are limited by the accuracy of their automatic image tracker. In a third arrangement, VCAM$_{in}$ is pointed at an eye of the user. The computer selects the sclera as the ROI and detects the physiological response based on color changes recognizable in IM$_{ROI}$ of the sclera. In a fourth arrangement, VCAM$_{in}$ is pointed at an eye of the user. The computer selects the iris as the ROI and detects the physiological response based on color changes recognizable in IM$_{ROI}$ of the iris. Optionally, the computer further calculates changes to the pupil diameter based on the IM$_{ROI}$ of the iris, and detects an emotional response of the user based on the changes to the pupil diameter.

[0143] In order to improve the detection accuracy, and in some cases in order to better account for interferences, the computer may utilize measurements of one or more head-mounted thermal cameras in the detection of the physiological response. The system may include an inward-facing head-mounted thermal camera that takes thermal measurements of a second ROI (TH$_{ROI2}$) on the user's face. Optionally, ROI and ROI$_2$ overlap, and the computer utilizes TH$_{ROI2}$ to detect the physiological response. Optionally, on average, detecting the physiological response based on both FSCC recognizable in IM$_{ROI}$ and TH$_{ROI2}$ is more accurate than detecting the physiological response based on the FSCC without TH$_{ROI2}$. Optionally, the computer utilizes TH$_{ROI2}$ to account, at least in part, for temperature changes, which may occur due to physical activity and/or consumption of certain medications that affect the blood flow. Optionally, the computer utilizes TH$_{ROI2}$ by generating feature values based on TH$_{ROI2}$, and utilizing a model that was trained on data comprising TH$_{ROI2}$ in order to detect the physiological response.

[0144] Alternatively, the system may include an outward-facing head-mounted thermal camera that takes thermal measurements of the environment (TH$_{ENV}$). Optionally, the computer may utilize TH$_{ENV}$ to detect the physiological response (e.g., by generating feature values based on TH$_{ENV}$ and utilizing a model trained on data comprising TH$_{ENV}$). Optionally, on average, detecting the physiological response based on both FSCC recognizable in IM$_{ROI}$ and TH$_{ENV}$ is more accurate than detecting the physiological response based on the FSCC without TH$_{ENV}$. Optionally, the computer utilizes TH$_{ENV}$ to account, at least in part, for thermal interferences from the environment, such as direct sunlight and/or a nearby heater.

[0145] In addition to detecting a physiological response, the computer may utilize IM$_{ROI}$ to generate an avatar of the user (e.g., in order to represent the user in a virtual environment). Optionally, the avatar may express emotional responses of the user, which are detected based on IM$_{ROI}$. Optionally, the computer may modify the avatar of the user to show synthesized facial expressions that are not manifested in the user's actual facial expressions. The synthesized facial expressions may correspond to emotional responses detected based on FSCC that are recognizable in IM$_{ROI}$. Alternatively, the synthesized facial expressions correspond to emotional responses detected based on thermal measurements taken by CAM.

Some of the various approaches that may be utilized to generate the avatar based on IM_{ROI} are described in co-pending US patent publication 2016/0360970.

[0146] Various examples described herein involve an HMS that may be connected, using wires and/or wirelessly, with a device carried by the user and/or a non-wearable device. The HMS may include a battery, a computer, sensors, and a transceiver.

[0147] FIG. 21a and FIG. 21b are schematic illustrations of possible computers (400, 410) that are able to realize one or more of the examples discussed herein that include a "computer". The computer (400, 410) may be implemented in various ways, such as, but not limited to, a server, a client, a personal computer, a network device, a handheld device (e.g., a smartphone), an HMS (such as smart glasses, an augmented reality system, and/or a virtual reality system), a computing device embedded in a wearable device (e.g., a smartwatch or a computer embedded in clothing), a computing device implanted in the human body, and/or any other computer form capable of executing a set of computer instructions. Herein, an augmented reality system refers also to a mixed reality system. Further, references to a computer or processor include any collection of one or more computers and/or processors (which may be at different locations) that individually or jointly execute one or more sets of computer instructions. For example, a first computer may be embedded in the HMS that communicates with a second computer embedded in the user's smartphone that communicates over the Internet with a cloud computer.

[0148] The computer 400 includes one or more of the following components: processor 401, memory 402, computer readable medium 403, user interface 404, communication interface 405, and bus 406. The computer 410 includes one or more of the following components: processor 411, memory 412, and communication interface 413.

[0149] Thermal measurements that are forwarded to a processor/computer may include "raw" values that are essentially the same as the values measured by thermal cameras, and/or processed values that are the result of applying some form of preprocessing and/or analysis to the raw values. Examples of methods that may be used to process the raw values include analog signal processing, digital signal processing, and various forms of normalization, noise cancellation, and/or feature extraction.

[0150] At least some of the methods described herein are "computer-implemented methods" that are implemented on a computer, such as the computer (400, 410), by executing instructions on the processor (401, 411). Optionally, the instructions may be stored on a computer-readable medium, which may optionally be a non-transitory computer-readable medium. In response to execution by a system including a processor and memory, the instructions cause the system to perform the method steps.

[0151] Herein, a direction of the optical axis of a VCAM or a CAM that has focusing optics is determined by the focusing optics, while the direction of the optical axis of a CAM without focusing optics (such as a single pixel thermopile) is determined by the angle of maximum responsivity of its sensor. When optics are utilized to take measurements with a CAM, then the term CAM includes the optics (e.g., one or more lenses). The optics of a CAM may include one or more lenses made of a material suitable for the required wavelength, such as one or more of the following materials: Calcium Fluoride,

Gallium Arsenide, Germanium, Potassium Bromide, Sapphire, Silicon, Sodium Chloride, and Zinc Sulfide.Alternatively, the CAM optics may include one or more diffractive optical elements, and/or or a combination of one or more diffractive optical elements and one or more refractive optical elements.

[0152]   When CAM includes an optical limiter/ field limiter/ FOV limiter (such as a thermopile sensor inside a standard TO-39 package with a window, or a thermopile sensor with a polished metal field limiter), then the term CAM may also refer to the optical limiter. Depending on the context, the term CAM may also refer to a readout circuit adjacent to CAM, and/or to the housing that holds CAM.

[0153]   Herein, references to thermal measurements in the context of calculating values based on thermal measurements, generating feature values based on thermal measurements, or comparison of thermal measurements, relate to the values of the thermal measurements (which are values of temperature or values of temperature changes). Thus, a sentence in the form of "calculating based on TH_{ROI}" may be interpreted as "calculating based on the values of TH_{ROI}", and a sentence in the form of "comparing TH_{ROI1} and TH_{ROI2}" may be interpreted as "comparing values of TH_{ROI1} and values of TH_{ROI2}".

[0154]   Thermal measurements of an ROI (usually denoted TH_{ROI} or using a similar notation) may have various forms, such as time series, measurements taken according to a varying sampling frequency, and/or measurements taken at irregular intervals. In some arrangements, thermal measurements may include various statistics of the temperature measurements (T) and/or the changes to temperature measurements (ΔT), such as minimum, maximum, and/or average values. Thermal measurements may be raw and/or processed values. When a thermal camera has multiple sensing elements (pixels), the thermal measurements may include values corresponding to each of the pixels, and/or include values representing processing of the values of the pixels. The thermal measurements may be normalized, such as normalized with respect to a baseline (which is based on earlier thermal measurements), time of day, day in the month, type of activity being conducted by the user, and/or various environmental parameters (e.g., the environment's temperature, humidity, radiation level, etc.).Herein, sentences in the form of "X is indicative of Y" (and/or using variations thereof) mean that X includes information correlated with Y, up to the case where X equals Y. For example, sentences in the form of "thermal measurements indicative of a physiological response" mean that the thermal measurements include information from which it is possible to infer the physiological response. Stating that "X indicates Y" or "X indicating Y" may be interpreted as "X being indicative of Y". Additionally, sentences in the form of "provide/receive an indication indicating whether X happened" may refer herein to any indication method, including but not limited to: sending/receiving a signal when X happened and not sending/receiving a signal when X did not happen, not sending/receiving a signal when X happened and sending/receiving a signal when X did not happen, and/or sending/receiving a first signal when X happened and sending/receiving a second signal X did not happen.

[0155]   Herein, "most" of something is defined as above 51% of the something (including 100% of the something). Both a "portion" of something and a "region" of something refer herein to a value between a fraction of the something and 100% of the something. For example, sentences in the form of a "portion of

an area" may cover between 0.1% and 100% of the area. As another example, sentences in the form of a "region on the user's forehead" may cover between the smallest area captured by a single pixel (such as 0.1% or 5% of the forehead) and 100% of the forehead. The word "region" is open-ended, and a camera said to capture a specific region on the face may capture just a small part of the specific region, the entire specific region, and/or a portion of the specific region together with additional region(s).

[0156] Sentences in the form of "angle greater than 20°" refer to absolute values (which may be +20° or -20° in this example), unless specifically indicated, such as in a phrase having the form of "the optical axis of CAM is 20° above/below the Frankfort horizontal plane" where it is clearly indicated that the CAM is pointed upwards/downwards. The Frankfort horizontal plane is created by two lines from the superior aspects of the right/left external auditory canal to the most inferior point of the right/left orbital rims.

[0157] The terms "comprises," "comprising," "includes," "including," "has," "having", or any other variation thereof, indicate open-ended language that does not exclude additional limitations. The "a" or "an" is employed to describe one or more, and the singular also includes the plural unless it is obvious that it is meant otherwise; for example, sentences in the form of "a CAM configured to take thermal measurements of a region (TH_{ROI})" refers to one or more CAMs that take thermal measurements of one or more regions, including one CAM that takes thermal measurements of multiple regions; as another example, "a computer" refers to one or more computers, such as a combination of a wearable computer that operates together with a cloud computer.

[0158] The phrase "based on" is intended to mean "based, at least in part, on".

[0159] The terms "first", "second" and so forth are to be interpreted merely as ordinal designations, and shall not be limited in themselves. A predetermined value is a fixed value and/or a value determined any time before performing a calculation that compares a certain value with the predetermined value. A value is also considered to be a predetermined value when the logic, used to determine whether a threshold that utilizes the value is reached, is known before start performing computations to determine whether the threshold is reached.

[0160] Embodiments of the invention may include any variety of combinations and/or integrations of the features described herein. Although some examples may depict serial operations, embodiments may perform certain operations in parallel and/or in different orders from those depicted. Moreover, the use of repeated reference numerals and/or letters in the text and/or drawings is for the purpose of simplicity and clarity and does not in itself dictate a relationship between the various examples and/or configurations discussed. Embodiments are not limited to the order of steps of the methods, or to details of implementation of the devices, set in the description, drawings, or examples. Moreover, individual blocks illustrated in the figures may be functional in nature and therefore may not necessarily correspond to discrete hardware elements.

[0161] Certain features which may have been, for clarity, described in the context of separate examples, may also be provided in various combinations in a single embodiment. Conversely, features

which may have been, for brevity, described in the context of a single example, may also be provided separately or in any suitable sub-combination. Specific examples are presented by way of example, and not limitation. Moreover, it is evident that many alternatives, modifications, and variations will be apparent to those skilled in the art. It is to be understood that other embodiments may be utilized and structural changes may be made without departing from the scope of the invention.

CLAIMS:

1.   A system configured to detect a physiological response based on facial skin color changes (FSCC), comprising:

an inward-facing head-mounted visible-light camera (VCAM$_{in}$) configured to take images of a region of interest (IM$_{ROI}$) on a user's face; wherein the region of interest (ROI) covers at least a portion of at least one of the following regions on the user's face: forehead, nose, and cheek; and wherein the ROI is illuminated by ambient light; and

a computer configured to detect the physiological response based on a photoplethysmographic signal manifested by FSCC recognizable in IM$_{ROI}$.

2.   The system of claim 1, further comprising an outward-facing head-mounted visible-light camera (VCAM$_{out}$) configured to take images of the environment (IM$_{ENV}$); wherein the computer is further configured to utilize a model to detect the physiological response also based on IM$_{ENV}$.

3.   The system of claim 2, wherein IM$_{ENV}$ is indicative of illumination towards the face, IM$_{ROI}$ is indicative of reflections from the face, and the computer is configured to utilize IM$_{ENV}$ to account, at least in part, for variations in ambient light that cause errors in detections of the physiological response.

4.   The system of claim 2, wherein the physiological response comprises at least one of: a heart rate of the user, heart rate variability of the user, a breathing rate of the user, and an expression of emotional response of the user.

5.   The system of claim 2, wherein at least one of VCAM$_{in}$ and VCAM$_{out}$ comprise sensors configured to capture light rays also in at least one of the following near infrared (NIR) spectrum intervals: 700-800 nm, 700-900 nm, 700-1,000 nm; and the computer is further configured to detect the physiological response also based on data obtained in the NIR spectrum interval.

6.   The system of claim 2, wherein the model was trained on: a first set of IM$_{ROI}$ and IM$_{ENV}$ taken while being indoors and not in direct sunlight, and a second set of IM$_{ROI}$ and IM$_{ENV}$ taken while being outdoors in direct sunlight.

7.   The system of claim 2, wherein the model was trained on: a first set of IM$_{ROI}$ and IM$_{ENV}$ taken while exercising and moving, and a second set of IM$_{ROI}$ and IM$_{ENV}$ taken while sitting and not exercising.

8. The system of any one of claims 1 to 7, wherein $VCAM_{in}$ does not occlude the ROI, and the computer is further configured to utilize a model to detect the physiological response; and wherein the model was trained on: a first set of IM_{ROI} taken while being indoors and not in direct sunlight, and a second set of IM_{ROI} taken while being outdoors in direct sunlight.

9. The system of any one of claims 1 to 7, wherein the computer is further configured to utilize a model to detect the physiological response; and wherein the model was trained on: a first set of IM_{ROI} taken while exercising and moving, and a second set of IM_{ROI} taken while sitting and not exercising; whereby the physiological response is still detected by the computer based on FSCC recognizable in IM_{ROI} that are taken while the user is running.

10. The system of any one of claims 1 to 7, wherein the computer is further configured to utilize a model to detect the physiological response; and wherein the model was trained on: a first set of IM_{ROI} taken less than 30 minutes after the user drank an alcoholic beverage, and a second set of IM_{ROI} taken on a day in which the user did not drink an alcoholic beverage.

11. The system of any one of claims 1 to 7, wherein the computer is further configured to utilize a model to detect the physiological response; wherein the model was trained on samples, each sample comprising: feature values generated based on IM_{ROI} of the user, and a label indicative of an emotional response of the user; and wherein the label is generated based on one or more of the following: semantic analysis of a communication of the user, analysis of a facial expression made by the user, thermal measurements of regions of the face of the user, and one or more of the following physiological signals of the user: heart rate, heart rate variability, breathing rate, and galvanic skin response.

12. The system of any one of claims 1 to 7, wherein the computer is further configured to receive an indication indicative of the user consuming a confounding substance that affects FSCC, and to detect the physiological response, while the consumed confounding substance affects FSCC, based on: IM_{ROI}, the indication, and a model; wherein the model was trained on: a first set of IM_{ROI} taken while the confounding substance affected FSCC, and a second set of IM_{ROI} taken while the confounding substance did not affect FSCC.

13. The system of any one of claims 1 to 7, wherein the ROI is on the forehead, and further comprising a frame configured to be worn on the user's head and to hold $VCAM_{in}$ less than 10 cm from the user's face such that the optical axis of $VCAM_{in}$ is above 20° from the Frankfort horizontal plane.

14. The system of any one of claims 1 to 7, wherein the ROI is on the nose, and further comprising a frame configured to be worn on the user's head and to hold $VCAM_{in}$ less than 10 cm from the user's face.

15. The system of any one of claims 1 to 7, further comprising another $VCAM_{in}$ configured to be pointed at an eye of the user; and the computer is further configured to perform at least one of the following: (i) select the sclera as a second region of interest (ROI2), and detect the physiological response based on color changes recognizable in images taken by the another $VCAM_{in}$, and (ii) select the iris as ROI2, and to detect the physiological response based on color changes recognizable in the iris.

16. The system of any one of claims 1 to 7, further comprising an inward-facing head-mounted thermal camera configured to take thermal measurements of a second ROI (TH_{ROI2}) on the user's face; wherein ROI and ROI_2 overlap, and the computer is further configured to utilize TH_{ROI2} to detect the physiological response; whereby, on average, detecting the physiological response based on both FSCC recognizable in IM_{ROI} and TH_{ROI2} is more accurate than detecting the physiological response based on FSCC recognizable in IM_{ROI} without TH_{ROI2}.

17. The system of any one of claims 1 to 7, wherein $VCAM_{in}$ does not occlude the ROI, and $VCAM_{in}$ comprises a sensor and a lens; the sensor plane is tilted by a fixed angle greater than 2° relative to the lens plane according to the Scheimpflug principle in order to capture a sharper image when $VCAM_{in}$ is worn by the user.

18. The system of any one of claims 1 to 7, wherein $VCAM_{in}$ comprises a sensor, a lens, and a motor; the motor is configured to tilt the lens relative to the sensor according to the Scheimpflug principle; whereby the tilt improves the sharpness of IM_{ROI} when $VCAM_{in}$ is worn by the user.

19. The system of any one of claims 1 to 7, wherein the computer is further configured to generate an avatar of the user based on IM_{ROI}, and to modify the avatar to show a synthesized facial expression that corresponds to an emotional response detected based on FSCC recognizable in IM_{ROI}, which is not manifested in the user's facial expression.

20. A method for detecting a physiological response, comprising:
taking, utilizing an inward-facing head-mounted visible-light camera ($VCAM_{in}$), images of region of interest (IM_{ROI}) on a user's face; wherein the region of interest (ROI) covers at least a portion

of at least one of the following regions on the user's face: forehead, nose, and cheek; and wherein the ROI is illuminated by ambient light;

taking, utilizing an outward-facing head-mounted visible-light camera ($VCAM_{out}$), images of the environment (IM_{ENV}); wherein IM_{ENV} is indicative of illumination towards the face; and

detecting the physiological response based on: a photoplethysmographic signal manifested by FSCC recognizable in IM_{ROI}, and IM_{ENV}; wherein utilizing IM_{ENV} accounts, at least in part, for variations in ambient light.

21.   the method of claim 20, further comprising generating an avatar of the user based on IM_{ROI}, and modifying the avatar to show a synthesized facial expression that corresponds to an emotional response detected based on FSCC recognizable in IM_{ROI}, which is not manifested in the user's facial expression.