

(51) International Patent Classification:
G06F 19/00 (2011.01)(21) International Application Number:
PCT/EP2015/067324(22) International Filing Date:
29 July 2015 (29.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14382294.8 30 July 2014 (30.07.2014) EP

(71) Applicants: UNIVERSITAT AUTONOMA DE BARCELONA [ES/ES]; Edifici A, Campus universitari s/n, E-08193 Bellaterra (Cerdanyola del Vallès) (ES). CENTRO DE VISIÓN POR COMPUTADOR (CVC) [ES/ES]; Edificio O, Campus de la UAB, E-08193 Bellaterra (Cerdanyola del Vallès) (ES). UNIVERSITAT DE BARCELONA [ES/ES]; Centre de Patents de la UB, Baldiri Raixac 4 - Torre D, E-08028 Barcelona (ES). ACCEPLAN ACCESIBILIDAD S.L. [ES/ES]; Riera Alta 10, 2º 1ª, E-08001 Barcelona (ES). FUNDACIÓ PER A LA UNIVERSITAT OBERTA DE CATALUNYA (UOC) [ES/ES]; Avenida del Tibidabo, 39-43, E-08035 Barcelona (ES).

(72) Inventors: BARÓ SOLÉ, Xavier; Avenida del Tibidabo, 39-43, E-08035 Barcelona (ES). ESCALERA GUERRERO, Sergio; Centre de Patents de la UB, Baldiri Reixac

4 - Torre D, E-08028 Barcelona (ES). GONZÁLEZ SABATÉ, Jordi; Campus de la UAB - Edifici O, E-08193 Bellaterra (Cerdanyola del Vallès) (ES). MACKAY JARQUE, Martha; Riera Alta 10, 2º 1ª, E-08001 Barcelona (ES).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: A COMPUTER-IMPLEMENTED METHOD AND A SYSTEM FOR REMOTELY MONITORING A USER, AND A COMPUTER PROGRAM PRODUCT IMPLEMENTING THE METHOD

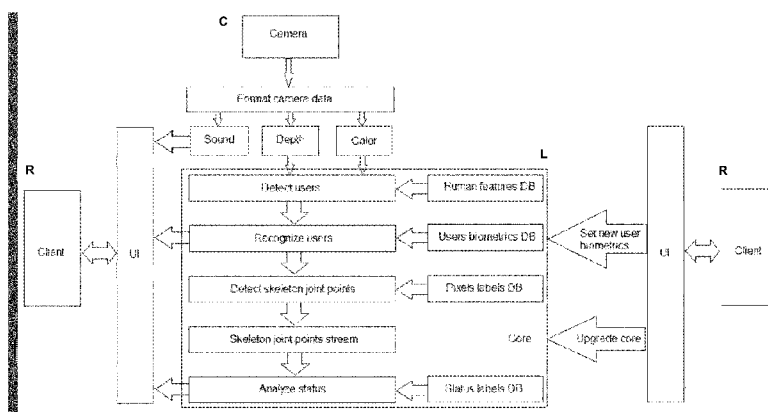


FIG. 1

(57) Abstract: The method is applied to a plurality of unidentified users and comprises: -automatically extracting biometrics for at least some of the plurality of unidentified users and storing the extracted biometrics in a biometrics database correlated to identity information of the correspondent user, -acquiring multi-modal user data for a monitored user, including at least depth data; -recognizing the identity of the monitored user by automatically extracting biometrics for the monitored user and searching the biometrics database to look for a match; -analysing the acquired multi-modal user data to identify a status of the monitored user; and -presenting the identified status and/or information associated thereto, to a remote computing means. The system and the computer program product are both adapted to implement the method of the invention.

**A computer-implemented method and a system for remotely monitoring a user,
and a computer program product implementing the method**

Field of the Invention

5 The present invention generally relates, in a first aspect, to a computer-implemented method for remotely monitoring a user, comprising acquiring and analysing multi-modal user data, including depth data, of a monitored user, and more particularly to a method applied to a plurality of unidentified users.

 A second aspect of the invention concerns to a system adapted to implement the
10 method of the first aspect.

 A third aspect of the invention relates to a computer program product adapted to implement the processing steps of the method of the first aspect of the invention.

Background of the Invention

15 Remote monitoring of users in indoor applications, such as patients of a hospital or elders of a residence, nowadays is a need which has not yet been completely satisfied.

 However, there are some known proposals focused on such remote user monitoring, although they are clearly susceptible of improving. Next, some patent
20 documents disclosing such proposals are identified and their relevant background briefly described.

 US2014/0052464 A1 discloses a method and a system for remotely monitoring a patient by acquiring information corresponding to a condition of a patient from a plurality of input sources, selecting an analytical model from a database which also stores
25 several analytical models and medical data, based on the information acquired and on the medical data for said patient, and determining a state of the patient and formulate a health prediction by performing another algorithm on the acquired information and medical data with the selected analytical model. Based on the determined state and on the health prediction, a recommendation is determined and transmitted to a remote
30 entity.

 The proposal made by US2014/0052464 A1 has several drawbacks, including the needing of working only with identified patients and of having access to the medical data of the patients.

 Another of said proposals is disclosed by EP2560141 A1, which describes an
35 interactive virtual care system which includes sensors to acquire multi-modal user data related to user movement, including depth image data, and analyses the acquired multi-

modal user data to identify an anomaly in the user movement. Said analysis is performed by comparing the acquired multi-modal user data to predetermined historical user data (such as medical history) and/or statistical norm data for users to identify an anomaly in the user movement. The user movement, a real version and a computer-generated version, and the identified anomaly can be displayed in a responder interface which can be placed at a remote location.

The system disclosed by EP2560141 A1 shares some of the drawbacks of the proposal made by US2014/0052464 A1, as it also needs accessing historic data for the user, such as medical history, and thus can only work with previously identified users.

10

Summary of the invention

It is an object of the present invention to offer an alternative to the prior state of the art, with the purpose of providing a method and a system for remotely monitoring a user which lacks of the drawbacks of the above cited proposals, and is able of working with unidentified users, and thus without the need of using users historic data.

15

To that end, the present invention relates, in a first aspect, to a computer-implemented method for remotely monitoring a user, the method comprising:

20

- acquiring multi-modal user data related to a monitored user, including at least depth data, using at least a depth camera in a local surveillance area;
- analysing, by processing means, the acquired multi-modal user data to identify a status of the monitored user; and
- presenting said identified status of the monitored user and/or information associated thereto, to a remote computing means.

25

Contrary to the known methods, the method of the first aspect of the present invention is applied to a plurality of unidentified users and comprises:

30

- automatically extracting biometrics for at least some of said plurality of unidentified users and storing the extracted biometrics in a biometrics database correlated to identity information of the correspondent user, and
- recognizing the identity of the monitored user by automatically extracting biometrics for the monitored user and searching the biometrics database to look for a match.

Said information associated to the identified status of the monitored user presented to the remote computing means includes at least, for a preferred embodiment, information indicative of an alarm.

For another embodiment, the information associated to the identified status of the monitored user presented to the remote computing means includes at least, for a preferred embodiment, information indicative of an alarm

For an embodiment, the method further comprises presenting part or all of the
5 acquired multi-modal user data to the remote computing means.

For an embodiment, said multi-modal user data comprises at least audio data, RGB video data and infrared depth video data.

According to a preferred embodiment, the mentioned biometrics are extracted from multi-modal user data including at least depth video or depth images data acquired
10 using at least a depth camera, whether said depth camera and/or other depth cameras depending on the local surveillance area in which the user which biometrics are being extracted is placed.

Optionally, the biometrics are extracted also from audio data and/or RGB video data of said multi-modal user data and/or from other suitable data sources.

15 According to an embodiment, the method of the first aspect of the invention comprises analysing, by said and/or another processing means, the acquired multi-modal user data to perform the next steps:

- i) detecting the monitored user,
- ii) performing said recognizing of the monitored user,
- 20 iii) detecting skeleton point joints of the monitored user, and
- iv) performing, using said processing means, said identifying of the status of the monitored user analysing at least the detected skeleton point joints of step iii) and also movement and 3D geometry, regarding the monitored user, using at least depth information of said acquired multi-modal user data.

25 The above mentioned step iv) is performed, for an embodiment, by analysing also data from sensors carried by the monitored user (for example in the form of a bracelet, helmet, chest band, etc.), including an accelerometer and/or a gyroscope and/or a temperature sensor and/or a blood pressure sensor and/or an oximeter sensor and/or an ECG sensor assembly and/or an EEG sensor assembly and/or an EMG
30 sensor assembly.

For an embodiment, said processing means are implemented in local computing means comprising and/or connected with said depth camera, all of said steps i) to iv) being performed by said local computing means, the local computing means being bidirectionally connected to said remote computing means and delivering the latter with
35 at least said identified status of the monitored user and/or said information associated thereto.

For an alternative embodiment, a local portion of said processing means are implemented in local computing means and a remote portion of the processing means are implemented in said remote computing means, some of said steps i) to iv) being performed by said local computing means and the rest of the steps i) to iv) being performed by said remote portion of the processing means implemented the remote computing means.

For another alternative embodiment, the processing means are implemented only in said remote computing means, all of said steps i) to iv) being performed by the remote computing means.

10 The detection of step i) is performed, for a preferred embodiment, using colour image data and depth information acquired by the depth camera, applying HOG features in RGBD space and learning those features on a predefined database of users and nonusers, and, afterwards, extracting body pixels by depth thresholding and surface normal analysing in an area within a defined bounding box around the monitored user.

15 Regarding the skeleton joint points, they are extracted at step iii), for a preferred embodiment, applying a model-based Random forest algorithm, using at least depth information acquired by said depth camera, comprising dividing the depth image data corresponding to the body of the monitored user into body parts, attaching an identification pixel label to each body part, and building a gestures database with at least
20 part of the attached identification pixel labels corresponding to different gestures, such as gestures associated to getting up, bending over, falling, picking up an object, etc. Optionally, if necessary, the model-based Random forest algorithm can be applied to colour image data to complement the results obtained from the depth image data.

Preferably, the method further comprises teaching the model-based Random
25 forest algorithm through feature vectors on body pixels of said body parts and their identification pixel labels, and, later, using the model-based Random forest algorithm to predict each identification pixel label of a new detected body using its associated feature vector, wherein said feature vectors are build using relative depth information between different voxels of said body parts, such that the feature vectors comprise depth
30 relations between said voxels.

According to an embodiment, the method of the first aspect of the invention comprises adding the detected skeleton joint points to a continuous stream, including at least a depth video or a sequence of depth images of the acquired multi-modal user data, and performing the identifying of the status of the monitored user at step iv)
35 analysing said continuous stream, regarding the movement and 3D geometry of the

monitored user, on temporal sequences of the continuous stream, together with the detected skeleton point joints.

For an embodiment, the method further comprises presenting said continuous stream to the remote computing means.

5 Preferably, the method comprises performing the biometrics extraction used for the recognizing of step ii) after the skeleton joints points detection of step iii) has been performed, and performing said biometrics extraction by at least analysing the spatial relations between the detected skeleton joint points (for example by correlating body limbs).

10 For an embodiment, the method of the first aspect of the invention comprises building a status database with information regarding different styles of status of users, and using the information included in said status database to teach an algorithm in charge of performing step iv) to define the action the monitored user is doing based on the continuous stream to identify the status of the monitored user, which, generally, is a
15 behavioural status, such as: walking, sitting, standing, laying down, watching TV, sleeping and/or falling.

Preferably, the method comprises performing in real time at least the identification of the status of the monitored user.

According to an embodiment, the method of the first aspect of the invention
20 comprises using a single user multi camera system comprising a plurality of depth cameras aiming to different local surveillance areas through which the user can circulate, and performing the acquiring of the multi-modal user data and processing thereof by using only the depth camera aimed to the area where the user is placed, and stop processing for other cameras until the user is available within their respective
25 surveillance areas, in order to reduce processing such that the method operates in real time.

In order also to further reduce processing, the method comprises, for an embodiment, performing the biometrics extraction from the above mentioned continuous stream, generally, regarding the monitored user, from an initial portion of the continuous
30 stream and performing the analysis to identify a status of the monitored user on next portions of the same continuous stream.

A second aspect of the present invention relates to a system for remotely monitoring a user, comprising:

- local acquiring means for acquiring multi-modal user data related to a
35 monitored user, said local acquiring means comprising at least one depth camera to acquire the multi-modal user data in a local surveillance area;

- processing means for analysing, by means of data processing, the acquired multi-modal user data to identify a status of the monitored user;

- remote computing means; and

- means for presenting said identified status of the monitored user and/or information associated thereto, to said remote computing means.

Contrary to the known systems, the system of the second aspect of the invention is a multi-user system and comprises:

- means for automatically extracting biometrics for at least some of a plurality of unidentified users and storing the extracted biometrics in a biometrics database correlated to identity information of the correspondent user, said biometrics database being included in the system, and

- means for recognizing the identity of the monitored user by automatically extracting biometrics for the monitored user and searching the biometrics database to look for a match.

The system of the second aspect of the invention is adapted for performing the method of the first aspect, where all the steps of the method, except for the step of acquiring multi-modal user data (which is performed by at least a depth camera), are performed by data processing by means of said processing means, the processing means being implemented in the remote computing means and/or in local computing means included in the system.

According to an embodiment, the remote computing means are adapted for establishing bidirectional communication with the local computing means, to allow, to an authorized person, to control, configure and upgrade operation settings of the local computing means.

For an embodiment, the remote computing means are implemented in a portable computing device, such as a smartphone or a tablet, and has installed therein a software application for performing at least part of the above mentioned authorized control, configuration and upgrade of operation settings of the local computing means, and/or to manage received identified status and alarms and/or to perform the identification of said status and the generation of said alarms, and/or to receive or build the continuous stream and show the latter in a display of the portable computing device and/or to present part or all of the acquired multi-modal user data.

A third aspect of the invention relates to a computer program product, which includes code instructions that when executed in a computer implement all the steps of the method of the first aspect of the invention, except for the step of acquiring multi-modal user data.

According to an embodiment, the computer program product comprises a first software application running on the above mentioned local computing means and a second software application constituting the above mentioned software application installed in the portable computing device, said first and second software application
5 having capabilities for exchanging data with each other.

The previous and other advantages and features will be better understood from the following detailed description of embodiments, with reference to the attached drawings, which must be considered in an illustrative and non-limiting manner.

10 Brief description of the drawings

FIG. 1 schematically shows the system of the present invention, for an embodiment; and

FIG. 2 depicts different examples of use case scenarios of the system of the present invention, showing some possible identified user status of the monitored user.

15

Detailed description of particular embodiments

Although the system of the present invention can be used for remotely monitoring different kind of users, it has been particularly thought for elders monitoring, to help people protecting elders easier and in a more comfortable way besides high
20 safety in indoor applications. It provides the tools, both hardware and software, in a package needed for such a monitoring. For a preferred embodiment, it is a portable system that can be put in any place in the room where it can see the user.

For an embodiment or basic version of the system, it uses a single camera for monitoring. While this is a client-side specification, for another embodiment, a server
25 can be involved for a multi-camera system in more elaborated versions to bring the framework to the client in a cheaper, robust and wider usage. This multi-camera system allows monitoring users automatically everywhere in indoor environments.

In the basic version, the system uses a depth camera, such as the Microsoft Kinect® camera, to capture the environment in colour and depth, along with a specific
30 usage computer to process the raw information from the camera. This is the basics of the system. Then, based on the client usages, an alarm system is needed. Here, the client interacts with the system through an internet line using a Smartphone/PC based application. The client can see the user instantly through the depth camera (for personal rights, just the depth image can be shown to the client) or view the messages of the
35 system sent to his/her Smartphone/PC based application. The messages are sent

based on the predefined user status. A ringtone alarm would occur on the client Smartphone/PC if the user status was urgent.

At the first time, the client puts the depth camera where the user can be seen from the camera viewpoint, connects the usb cable of the depth camera to the processor
5 computer (called in a previous section as local computing means), connects the power cables of the depth camera and computer, and finally runs the system. Now, the client can install the Smartphone/PC based application on his/her device (called in a previous section as remote computing means) easily and communicate with the system, set the settings, upgrade the software, etc. The security protocols do not allow anonymous
10 persons reach the system illegally. Therefore, the system is easy to work and safe enough.

In FIG. 1 a technical implementation of the system of the present invention is shown, including the next elements:

- a depth camera C to acquire the multi-modal user data in a local surveillance
15 area, including sound, depth images and colour images which are conditioned in a suitable format;
- a local processor computer L, implementing a core, for analysing the acquired multi-modal user data to identify a status of the monitored user; and
- remote computing means R, or client (two of them are illustrated in Figure 1),
20 bidirectionally communicated with the local computing means by a user interface UI.

Two kinds of software applications are used:

1. A core implementation to analyse the raw information achieved by the depth camera C, which is installed on the local processor computer L and provides the
25 feedbacks to the client R. This part is responsible to detect the user and monitor it. The architecture of this part can be seen in the Figure 1.

2. And a GUI to interact with the core L. This is the client-side software and is responsible to receive messages (such as the results of the recognizing of users) and alarms (and any result of the identified status, if needed, even if not associated to an
30 alarm status) from the core and show them to the client R, and upgrade new versions of the core and set new user biometrics (if needed) on the processor computer L. Although two clients R (i.e. remote computing means) have been depicted in Figure 1 for performing different functions, each client R can perform the described functions involving the shown and described bidirectional data flow.

35 Particularly, according to the illustrated embodiment, local processor computer L implements the steps of the method already described in a previous section, by means

of the illustrated functional modules, regarding the detection of users using a Human features Database (DB), the recognizing of the detected users using an users biometrics database (DB), the detection of skeleton joint points using a pixels labels Database (DB), the generation of a continuous stream including the detected skeleton joint points and also a depth video and the analysis of said continuous stream to identify the status of the patient.

For an implementation of the system/method of the present invention, the depth camera C gives the scene data in 30 fps. This data must be processed by the core in real time which is a critical task. The first task is detecting users pixels and separate them from the background. This can be done in real time using background subtraction (see T. Bouwmans, F. Porikli, B. Horferlin, A. Vacavant, Handbook on "Background Modeling and Foreground Detection for Video Surveillance: Traditional and Recent Approaches, Implementations, Benchmarking and Evaluation", CRC Press, Taylor and Francis Group, June 2014) and depth thresholding method then tracking them (see Lyudmila Mihaylova, Paul Brasnett, Nishan Canagarajan and David Bull (2007). Object Tracking by Particle Filtering Techniques in Video Sequences; In: Advances and Challenges in Multisensor Data and Information. NATO Security Through Science Series, 8. Netherlands: IOS Press. pp. 260–268), which is inaccurate mostly when the user is near to other stuffs in the room, and needs an initial gesture or movement. Instead, a superfast approach is used according to the present invention, thanks to the GPU programming, applying HOG features (see V. Prisacariu, I. Reid, "fastHOG - a real-time GPU implementation of HOG", Technical report, University of Oxford, 2012) in RGBD space and learning those features on a predefined database of users and nonusers. Afterwards, while a bounding box around the user is placed, body pixels are extracted by depth thresholding and surface normal analysing (see [http://en.wikipedia.org/wiki/Normal_\(geometry\)](http://en.wikipedia.org/wiki/Normal_(geometry))).

In multi-user systems, it is important to know who the detected user is. Each individual has different biometrics like hair colour, face shape, height, weight, body size, etc., that combination of them makes a unique feature allows us to recognize him/her (see <http://en.wikipedia.org/wiki/Surveillance>). The system and method of the present invention automatically extracts these biometrics for each user for the first time and keep them in a database which can be upgraded. Users are recognized by searching this database and matching the parameters. Then the user can be tracked after recognizing him/her and do this process whenever the system lost the user from tracking. Although in FIG. 1 the functional module "Recognize users" has been placed before the functional module "Detect skeleton joint points", the recognizing of users is preferably done after

detecting skeleton joint points, because then biometrics are easier to extract and there is a correlation among body limbs.

Skeleton joint points are extracted applying a superfast (using GPU programming) model-based Random forest approach (see Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake, Real-Time Human Pose Recognition in Parts from a Single Depth Image, in CVPR, IEEE, June 2011). Here, body is divided into parts attached a label to each part, and then a database of pixel labels of very different gestures is composed. Random forest is a learning algorithm and it is taught, according to the present invention, through feature vectors on body pixels and their labels. Later, the algorithm predicts each pixel label of the new detected body using its associated feature vector. Feature vectors are using depth information of other pixels to be computed, particularly using relative depth information between different voxels of the body parts. Body parts are defined such that mean point of each part shows the joint point of it. Working with depth enables us to continue without light. In working with RGB images in presence of light, some works extracted joint points instantly from body bounding box but the accuracy is less than working with depth.

Skeleton joint points are added to the stream and using this continuous stream the status of the user can be analysed, i.e. identified from analysis the continuous stream. Depending on the client, different status can be defined, walking, sitting, standing, laying down, watching TV, sleeping, falling, etc. A database for different styles of each status is made and an algorithm to define the action the user is doing based on the current stream is taught. A solution can be hierarchical aligned cluster analysis (see F. Zhou, F. De la Torre and J. K. Hodgins. Hierarchical Aligned Cluster Analysis for Temporal Clustering of Human Motion. IEEE Transactions Pattern Analysis and Machine Intelligence (PAMI), vol. 35, no. 3, pp. 582-596, 2013).

Each part of the core needs heavy processing. The system can be tuned to reduce processing and bring the system into real time. For instance, using the core for initial recognition and applying tracking in the next frames, or in a single user multi camera system, focusing the system on the camera sees the user and stop processing for other cameras until he is available.

Examples of Use Case Scenarios:

Examples of use case scenarios of the system of the present invention are depicted in Figure 2, which shows several depth cameras C connected to their respective local computing means L in three different situations of monitored users:

- Scenario 1 (User walking at home): Users with a certain degree of mobility but who can move freely in their homes, for example a dining room.
- Scenario 2 (User watching TV): Many carers leave home at those moments when the person they are taking care of is performing an activity that does not involve any movement, for example watching TV or also performing resting activities (reading, eating, etc.). In these cases, the user is monitored in a small space like a room or a hall.
- Scenario 3 (User sleeping): Another situation that is considered is to remotely monitor users while they are in bed because of a disease or because they are sleeping.

In these three scenarios, each local computing means L communicate with the remote computing means R, sending to the latter the above described data, including biometric information about the user and, generally, the above described continuous stream and identified user status (although depending on the embodiment, the building of said stream and/or the identification of the user status can be performed whether by the local computing means L or by the remote computing means R). The remote computing means R analyses the received data and is responsible for communicating with the devices of authorized persons, like caregivers to the elderly or relatives, through the network with information indicative of an alarm.

In these scenarios, such authorized persons can perform other activities, since they know that if something happens they will receive a warning, they will see what the user is doing, and they will be able to communicate at a distance with his/her voice with the user.

A person skilled in the art could introduce changes and modifications in the embodiments described without departing from the scope of the invention as it is defined in the attached claims.

Claims

- 1.- A computer-implemented method for remotely monitoring a user, comprising:
- acquiring multi-modal user data related to a monitored user, including at least
5 depth data, using at least a depth camera in a local surveillance area;
 - analysing, by processing means, the acquired multi-modal user data to identify a status of the monitored user; and
 - presenting said identified status of the monitored user and/or information associated thereto, to a remote computing means;
- 10 **characterised in that** the method comprises:
- automatically extracting biometrics for at least some of a plurality of unidentified users, and storing the extracted biometrics in a biometrics database correlated to identity information of the correspondent user; and
 - recognizing the identity of the monitored user by automatically extracting
15 biometrics for the monitored user and searching the biometrics database to look for a match.
- 2.- The method according to claim 1, wherein said multi-modal user data comprises at least audio data, RGB video data and infrared depth video data.
- 3.- The method according to any of the claims 1 or 2, wherein said biometrics are
20 extracted from multi-modal user data including depth video or depth images data acquired using at least a depth camera.
- 4.- The method according to claim 3, wherein said biometrics are extracted also from audio data and/or RGB video data of said multi-modal user data.
- 5.- The method according to any of the claims 1 to 4, comprising analysing, by
25 said processing means, the acquired multi-modal user data to perform the next steps:
- i) detecting the monitored user,
 - ii) performing said recognizing of the monitored user,
 - iii) detecting skeleton point joints of the monitored user, and
 - iv) performing, using said processing means, said identifying of the status of the
30 monitored user analysing at least the detected skeleton point joints of step iii) and also movement and 3D geometry, regarding the monitored user, using at least depth information of said acquired multi-modal user data.
- 6.- The method according to claim 5, wherein said step iv) is performed by analysing also data from sensors carried by the monitored user, said sensors
35 comprising an accelerometer and/or a gyroscope and/or a temperature sensor and/or a

blood pressure sensor and/or an oximeter sensor and/or an ECG sensor assembly and/or an EEG sensor assembly and/or an EMG sensor assembly.

7.- The method according to any of the claims 5 or 6, wherein said processing means are implemented in local computing means comprising and/or connected with said depth camera, all of said steps i) to iv) being performed by said local computing means, the local computing means being bidirectionally connected to said remote computing means and delivering the latter with at least said identified status of the monitored user and/or said information associated thereto.

8.- The method according to any of the claims 5 or 6, wherein a local portion of said processing means are implemented in local computing means and a remote portion of the processing means are implemented in said remote computing means, some of said steps i) to iv) being performed by said local computing means and the rest of the steps i) to iv) being performed by said remote portion of the processing means implemented the remote computing means.

9.- The method according to any of the claims claims 5 or 6, wherein said processing means are implemented in said remote computing means, all of said steps i) to iv) being performed by said remote computing means.

10.- The method according to any of the claims 5 to 9, wherein said detection of step i) is performed, using colour image data and depth information acquired by said depth camera, applying HOG features in RGBD space and learning those features on a predefined database of users and nonusers, and, afterwards, extracting body pixels by depth thresholding and surface normal analysing in an area within a defined bounding box around the monitored user.

11.- The method according to any of the claims 5 to 10, wherein said skeleton joint points are extracted, at step iii), applying a model-based Random forest algorithm, using at least depth information acquired by said depth camera, comprising dividing the depth image data corresponding to the body of the monitored user into body parts, attaching an identification pixel label to each body part, and building a gestures database with at least part of the attached identification pixel labels corresponding to different gestures.

12.- The method according to claim 11, further comprising teaching said model-based random forest algorithm through feature vectors on body pixels of said body parts and their identification pixel labels, and, later, using the model-based random forest algorithm to predict each identification pixel label of a new detected body using its associated feature vector, wherein said feature vectors are build using relative depth

information between different voxels of said body parts, such that the feature vectors comprise depth relations between said voxels.

13.- The method according to claim 12, comprising adding the detected skeleton joint points to a continuous stream, including at least a depth video or a sequence of
5 depth images of the acquired multi-modal user data, and performing the identifying of the status of the monitored user at step iv) analysing said continuous stream, regarding the movement and 3D geometry of the monitored user, on temporal sequences of said continuous stream, together with the detected skeleton point joints.

14.- The method according to claim 5 or according to any of claims 6 to 13 when
10 depending on claim 5, comprising performing the biometrics extraction used for the recognizing of step ii) after the skeleton joints points detection of step iii) has been performed, and performing said biometrics extraction by at least analysing the spatial relations between the detected skeleton joint points.

15.- The method of claim 14 when depending on claim 13, comprising performing
15 the biometrics extraction from said continuous stream.

16.- The method according to claim 15, comprising performing the biometrics extraction for the monitored user from an initial portion of the continuous stream and performing the analysis to identify a status of the monitored user on next portions of the continuous stream.

20 17.- The method according to claim 13, comprising building a status database with information regarding different styles of status of users, and using the information included in said status database to teach an algorithm in charge of performing step iv) to define the action the monitored user is doing based on the continuous stream to identify the status of the monitored user.

25 18.- The method according to any of the previous claims, wherein said status of the monitored user is a behavioural status.

19.- The method according to any of the previous claims, wherein the behavioural status is at least one of walking, sitting, standing, laying down, watching TV, sleeping and falling.

30 20.- The method according to any of the previous claims, comprising using a single user multi camera system comprising a plurality of depth cameras aiming to different local surveillance areas through which the user can circulate, and performing the acquiring of the multi-modal user data by using only the depth camera aimed to the area where the user is placed.

35 21.- The method according to any of the previous claims, comprising performing in real time at least the identification of the status of the monitored user.

22.- The method according to any of the previous claims, wherein said information associated to the identified status of the monitored user presented to the remote computing means is indicative of an alarm.

23.- A system for remotely monitoring a user, comprising:

- 5 - local acquiring means for acquiring multi-modal user data related to a monitored user, said local acquiring means comprising at least one depth camera (C) to acquire the multi-modal user data in a local surveillance area;
- processing means for analysing, by means of data processing, the acquired multi-modal user data to identify a status of the monitored user;
- 10 - remote computing means (R); and
- means for presenting said identified status of the monitored user and/or information associated thereto, to said remote computing means (R);

characterised in that the system is a multi-user system and comprises:

- means for automatically extracting biometrics for at least some of a plurality of
- 15 unidentified users and storing the extracted biometrics in a biometrics database correlated to identity information of the correspondent user, said biometrics database being included in the system, and
- means for recognizing the identity of the monitored user by automatically extracting biometrics for the monitored user and searching the biometrics database to
- 20 look for a match.

24.- The system according to claim 23, adapted for performing the method of any of claims 1 to 22, where all the steps of the method, except for the step of acquiring multi-modal user data, are performed by data processing by means of said processing means, the processing means being implemented in the remote computing means (R)

25 and/or in local computing means (L) included in the system.

25.- The system according to claim 24, wherein said remote computing means (R) are adapted for establishing bidirectional communication with the local computing means (L), to allow, to an authorized person, to control, configure and upgrade operation settings of the local computing means (L).

30 26.- The system according to claim 25, wherein said remote computing means (R) are implemented in a portable computing device.

27.- A computer program product, which includes code instructions that when executed in a computer implement all the steps of the method according to any of the claims 1 to 22, except for the step of acquiring multi-modal user data.

1/2

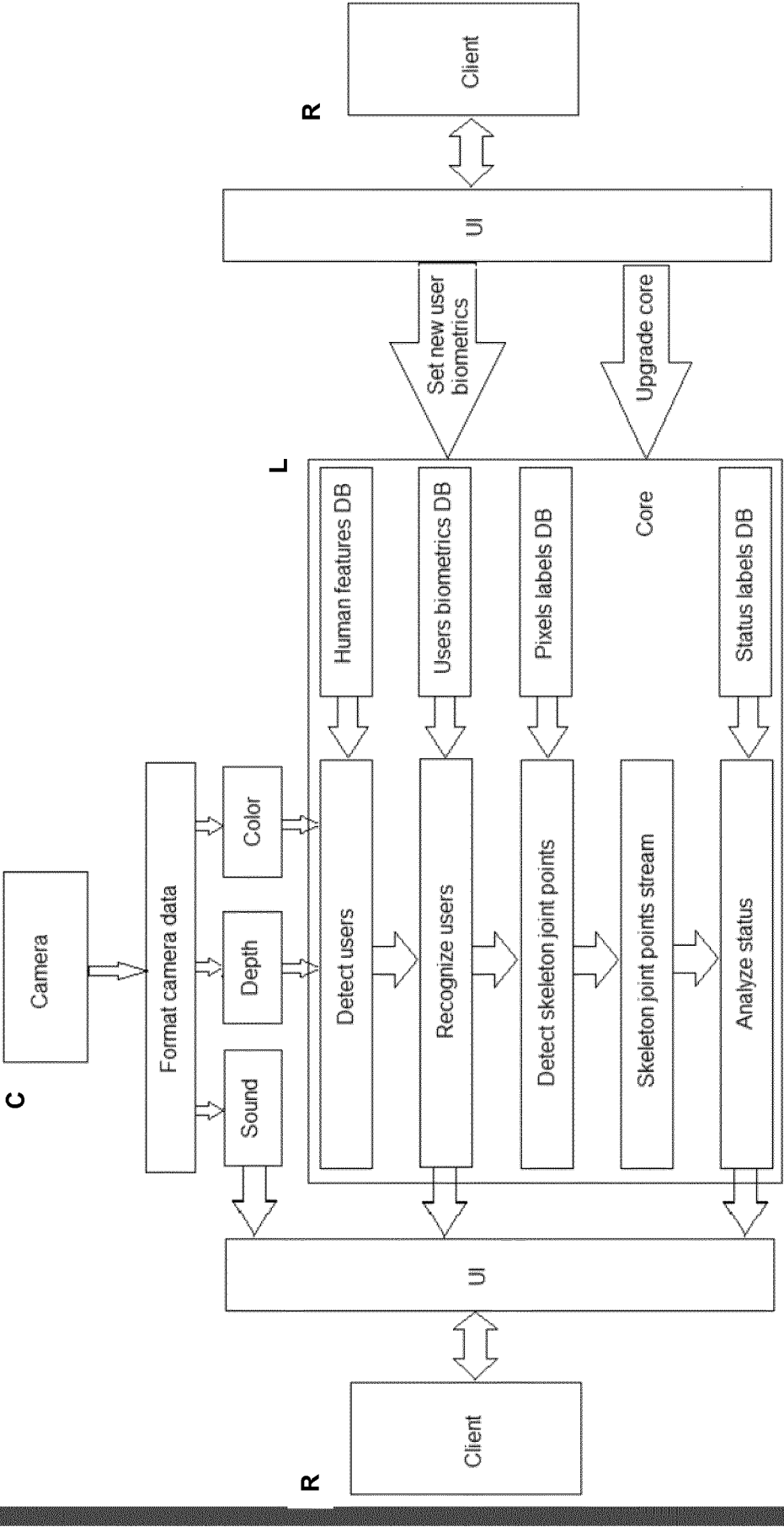


FIG. 1

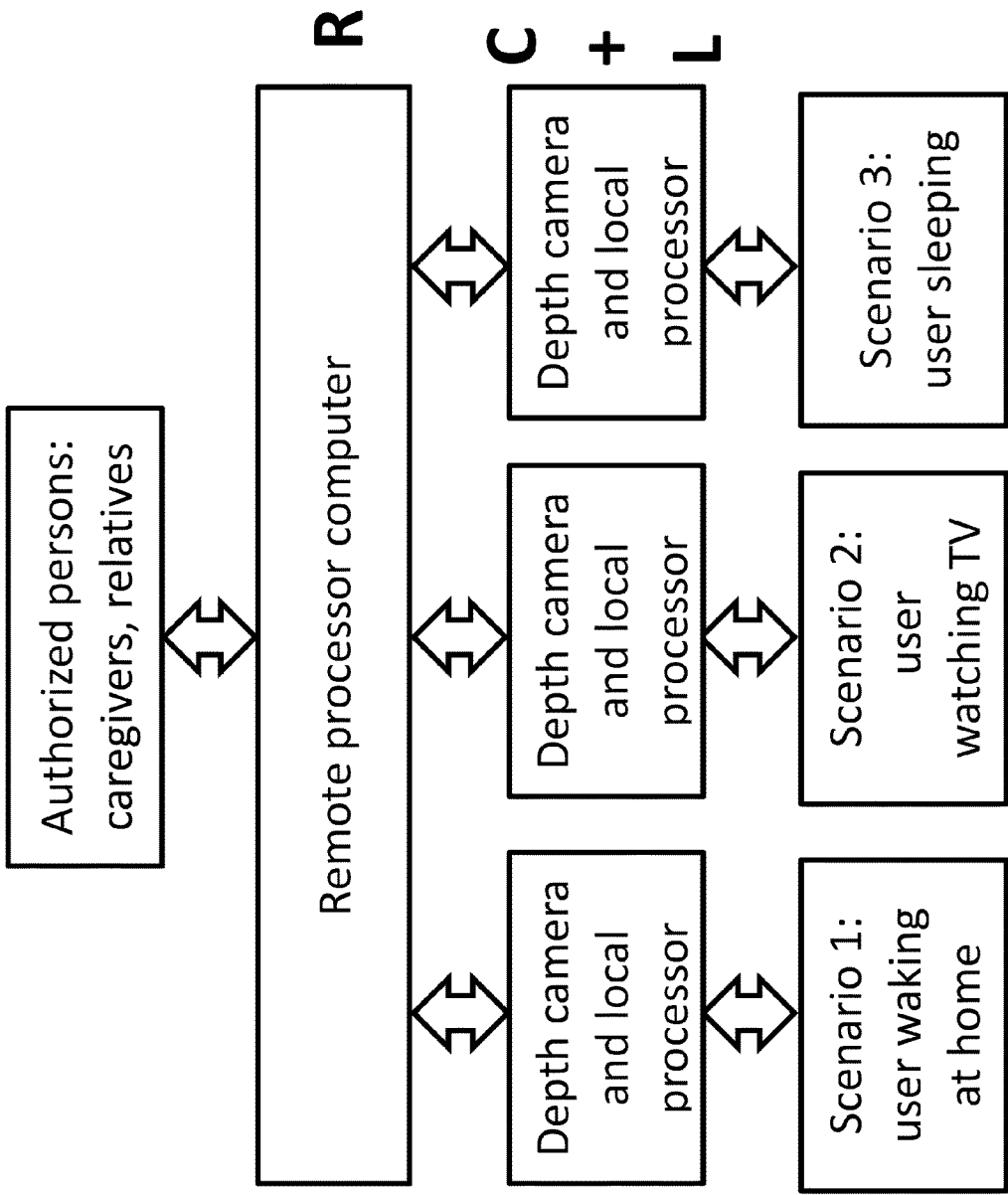


FIG. 2

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/067324

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F19/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012/075464 A1 (DERENNE RICHARD A [US] ET AL) 29 March 2012 (2012-03-29) abstract paragraph [0038] - paragraph [0044] paragraph [0063] paragraph [0068] - paragraph [0070] paragraph [0093] paragraph [0150] paragraph [0153] figures 1, 14, 15 ----- -/--	1-27



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

15 October 2015

Date of mailing of the international search report

23/10/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Schechner-Resom, M

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2015/067324

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	T D RÄTY: "Survey on Contemporary Remote Surveillance Systems for Public Safety", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: PART C:APPLICATIONS AND REVIEWS., vol. 40, no. 5, 1 September 2010 (2010-09-01), pages 493-515, XP055219318, US ISSN: 1094-6977, DOI: 10.1109/TSMCC.2010.2042446 the whole document -----	1-27
A	Matthew Fitzpatrick ET AL: "WORCESTER POLYTECHNIC INSTITUTE Real Time Person Tracking and Identification using the Kinect sensor Major Qualifying Project in Electrical & Computer Engineering", 25 April 2013 (2013-04-25), XP055218198, Retrieved from the Internet: URL:https://www.wpi.edu/Pubs/E-project/Available/E-project-042513-081838/unrestricted/Real_Time_Person_Tracking_and_Identification_using_the_Kinect_sensor.pdf [retrieved on 2015-10-05] the whole document -----	1-27
A	V. PRISACARIU; I. REID: "fastHOG - a real-time GPU implementation of HOG", TECHNICAL REPORT, 4 July 2012 (2012-07-04), XP055219207, cited in the application the whole document -----	1-27
A	JAMIE SHOTTON ET AL: "Real-time human pose recognition in parts from single depth images", COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2011 IEEE CONFERENCE ON, IEEE, 20 June 2011 (2011-06-20), pages 1297-1304, XP032037818, DOI: 10.1109/CVPR.2011.5995316 ISBN: 978-1-4577-0394-2 cited in the application the whole document ----- -/--	1-27

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2015/067324

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	FENG ZHOU ET AL: "Hierarchical Aligned Cluster Analysis for Temporal Clustering of Human Motion", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE COMPUTER SOCIETY, USA, vol. 35, no. 3, 1 March 2013 (2013-03-01), pages 582-596, XP011494042, ISSN: 0162-8828, DOI: 10.1109/TPAMI.2012.137 cited in the application the whole document -----	1-27

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2015/067324

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012075464 A1	29-03-2012	EP 2619724 A2	31-07-2013
		US 2012075464 A1	29-03-2012
		WO 2012040554 A2	29-03-2012
