(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 3 358 498 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
29.01.2020 Bulletin 2020/05

(21) Application number: **18154272.1**

(22) Date of filing: **30.01.2018**

(51) Int Cl.:
G06K 9/00 *(2006.01)* ***G06K 9/62*** *(2006.01)*
A61B 5/11 *(2006.01)* ***G06K 9/66*** *(2006.01)*
G06K 9/46 *(2006.01)* ***A61B 5/00*** *(2006.01)*

(54) **METHOD AND APPARATUS FOR AUTOMATIC EVENT PREDICTION**

VERFAHREN UND VORRICHTUNG ZUR AUTOMATISCHEN EREIGNISVORHERSAGE

PROCÉDÉ ET APPAREIL DE PRÉDICTION D'ÉVÉNEMENT AUTOMATIQUE

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **02.02.2017 US 201762453857 P**

(43) Date of publication of application:
08.08.2018 Bulletin 2018/32

(73) Proprietor: **Hill-Rom Services, Inc.**
Batesville, IN 47006-9167 (US)

(72) Inventors:
- **WONG, Alexander Sheung Lai**
  Batesville, IN 47006-9167 (US)
- **CHWYL, Brendan James**
  Batesville, IN 47006-9167 (US)
- **CHUNG, Audrey Gina**
  Batesville, IN 47006-9167 (US)
- **SHAFIEE, Mohammad Javad**
  Batesville, IN 47006-9167 (US)
- **FU, Yongji**
  Batesville, IN 47006-9167 (US)

(74) Representative: **Findlay, Alice Rosemary**
Reddie & Grose LLP
The White Chapel Building
10 Whitechapel High Street
London E1 8QS (GB)

(56) References cited:
- **GRIMM TIMO ET AL: "Sleep position classification from a depth camera using Bed Aligned Maps", 2016 23RD INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION (ICPR), IEEE, 4 December 2016 (2016-12-04), pages 319-324, XP033085604, DOI: 10.1109/ICPR.2016.7899653 [retrieved on 2017-04-13]**
- **MARTINEZ MANUEL ET AL: ""BAM!" Depth-Based Body Analysis in Criti", 27 August 2013 (2013-08-27), MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION - MICCAI 2015 : 18TH INTERNATIONAL CONFERENCE, MUNICH, GERMANY, OCTOBER 5-9, 2015; PROCEEDINGS; [LECTURE NOTES IN COMPUTER SCIENCE; LECT.NOTES COMPUTER], SPRINGER INTERNATIONAL PUBLISHING, CH, XP047038571, ISSN: 0302-9743 ISBN: 978-3-642-38287-1 * Section 3 ***
- **ACHILLES FELIX ET AL: "Patient MoCap: Human Pose Estimation Under Blanket Occlusion for Hospital Monitoring Applications", 2 October 2016 (2016-10-02), MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION - MICCAI 2015 : 18TH INTERNATIONAL CONFERENCE, MUNICH, GERMANY, OCTOBER 5-9, 2015; PROCEEDINGS; [LECTURE NOTES IN COMPUTER SCIENCE; LECT.NOTES COMPUTER], SPRINGER INTERNATIONAL PUBLISHING, CH, XP047359596, ISSN: 0302-9743 ISBN: 978-3-642-38287-1 [retrieved on 2016-10-02] * Section 3.1 Section 3.2; figure 1 * * Section 4 ***

Printed by Jouve, 75001 PARIS (FR)

(Cont. next page)

- **CHWYL BRENDAN ET AL: "DeepPredict: A deep predictive intelligence platform for patient monitoring", 2017 39TH ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (EMBC), IEEE, 11 July 2017 (2017-07-11), pages 4309-4312, XP033152978, DOI: 10.1109/EMBC.2017.8037809 [retrieved on 2017-09-13]**

Description

[0001] The present disclosure is related to the use of cameras in a patient environment to collect image data and predict adverse events. More specifically, the present disclosure is related to a method of establishing a likelihood of a bed exit event using real time image data that is modified by pre-trained deep features using a neural network.

[0002] A prominent concern in healthcare facilities is the limited number of available healthcare providers (e.g., nurses, doctors) relative to the growing number of patients. One area that can be automated to potentially improve the overall work flow is the monitoring of patient's bed exit events from hospital beds. Automatic bed exit detection systems have previously been proposed to assist healthcare providers. For example, see Grimm Timo et al.: "Sleep position classification from a depth camera using Bed Aligned Maps", 2016 23rd international conference on pattern recognition (ICPR), IEEE, 4 December 2016, pages 319-324.

[0003] U.S. Patent No. 5,276,432 discloses a system which uses four load cells to measure weight at each corner of a hospital bed. The patient's center of gravity with respect to the bed frame is calculated using the information captured by the load cells. If the center of gravity of the patient is not contained within a predetermined region, a bed exit event is said to be detected. Similarly, Scott [9] has patented a system in which the presence of a patient is detected by a multitude of sensors. Specifically, the presence or absence of a patient is determined by the dielectric constant, which is measured within a predefined detection space.

[0004] U.S. Published Patent Application No. 20160063846 discloses a system which, in addition to detecting the presence of a patient, aims to determine the location of a patient on a hospital bed. The proposed system uses a patient support assembly on a frame which consists of at least one deformation sensor, indicating the deformation of the frame. A location determination unit is connected to the deformation sensor to provide the lateral or longitudinal location of the patient, depending on the deformation sensor's location.

[0005] While methods exist for detecting bed exit events, detecting these exit events as they occur does not allow for a healthcare professional to provide timely assistance during the exit event. A system for automatic bed exit prediction would enable the notification of healthcare providers prior to a bed exit event so that prompt assistance can be provided to patients. This would significantly lower the potential for patient injuries (e.g., falls, strains) occurring during an unassisted bed exit event.

[0006] The present disclosure discloses one or more of the following features alone or in any combination. The present invention is defined in the appended claims.

[0007] According to a first aspect of the present disclosure, an apparatus for predicting that a patient is about to exit from a patient support apparatus comprises a camera positioned with the patient support apparatus in the field of view of the camera and a controller receiving signals representative of images from the camera. The controller is operable to capture time sequenced video images from the camera. The controller is also operable to input the time sequenced video images to a convolution layer and the time sequenced video images are convolved with a convolution kernel to produce a defined number of feature maps. The controller is further operable to input the feature maps into a context-aware pooling layer to extract relevant features of interest from the feature maps and generate feature vectors. The controller is still further operable to input a feature vector to a first fully connected layer such that each element of the feature vector is connected to a plurality of artificial neurons in the first fully connected layer and each combination outputs a first connected layer value. The controller is yet still further operable to input the values derived by each combination of first fully connected layer into a second fully connected layer such that each value is connected to a plurality of artificial neurons in the second fully connected layer such that each combination outputs a second connected layer value. The controller is yet still further operable to input the second connected layer values into an output layer which provides a non-exit probability which defines the likelihood that a patient exit event will not occur in a predetermined time and an exit probability which defines the likelihood that a patient exit event will occur in the predetermined time. The controller is yet still also operable to utilize the non-exit probability and exit probability to determine the likelihood of a patient exit event to generate a signal when the determine likelihood of a patient exit event exceeds a threshold value. The controller is also operable, if the signal is generated based on the determined likelihood of a patient exit event exceeds a threshold value, to generate a notification of the impending event.

[0008] The term an "artificial neuron" as used herein means a programmable element that is capable of storing information and working in conjunction with other artificial neurons to mimic biologic neural networks.

[0009] An artificial neural network is defined in Wiley Electrical and Electronics Engineering Dictionary, copyright 2004, John Wiley & Sons as: "A network composed of many processing modules that are interconnected by elements which are programmable and that store information, so that any number of them may work together. Its goal is to mimic the action of biologic neural networks, displaying skills such as the ability to solve problems, find patterns and learn.

[0010] In some embodiments of the first aspect, the convolution layer applies a rectifier when the feature maps are generated.

[0011] In some embodiments of the first aspect, the rectifier introduces non-saturating linearity to the features maps.

[0012] In some embodiments of the first aspect, the rectifier is an absolute value function.

[0013] In some embodiments of the first aspect, the context-aware pooling layer rectifies for a variation in the camera's field of view.

[0014] In some embodiments of the first aspect, the first fully connected layer includes 50 artificial neurons.

[0015] In some embodiments of the first aspect, the second fully connected layer comprises 10 artificial neurons.

[0016] In some embodiments of the first aspect, the first fully connected layer and second fully connected layer apply a transfer function.

[0017] In some embodiments of the first aspect, the transfer function is the tansig(x) function.

[0018] In some embodiments of the first aspect, the first fully connected layer and the second fully connected layer are developed by training via stochastic gradient descent to produce a set of deep features.

[0019] According to a second aspect of the present disclosure, a method of predicting that a patient is about to exit from a patient support apparatus that is in the field of view of a camera comprises receiving signals representative of images from the camera and capturing time sequenced video images from the camera. The method also comprises inputting the time sequenced video images to a convolution layer and convolving the time sequenced video images with a convolution kernel to produce a defined number of feature maps. The method further comprises inputting the feature maps into a context-aware pooling layer to extract relevant features of interest from the feature maps and generate feature vectors. The method still also comprises inputting a feature vector to a first fully connected layer such that each element of the feature vector is connected to a plurality of artificial neurons in the first fully connected layer and each combination outputs a first connected layer value. The method still further comprises inputting the values derived by each combination of first fully connected layer into a second fully connected layer such that each value is connected to a plurality of artificial neurons in the second fully connected layer such that each combination outputs a second connected layer value. The method yet still further comprises inputting the second connected layer values into an output layer which provides a non-exit probability which defines the likelihood that a patient exit event will not occur in a predetermined time and an exit probability which defines the likelihood that a patient exit event will occur in the predetermined time. The method also further comprises utilizing the non-exit probability and exit probability to determine the likelihood of a patient exit event to generate a signal when the determined likelihood of a patient exit event exceeds a threshold value. The method also still further comprises, if the signal is generated based on the determined likelihood of a patient exit event exceeds a threshold value, generating a notification of the impending event.

[0020] In some embodiments of the second aspect, the convolution layer applies a rectifier when the feature maps are generated.

[0021] In some embodiments of the second aspect, the rectifier introduces non-saturating linearity to the features maps.

[0022] In some embodiments of the second aspect, the rectifier is an absolute value function.

[0023] In some embodiments of the second aspect, the context-aware pooling layer rectifies for a variation in the camera's field of view.

[0024] In some embodiments of the second aspect, the first fully connected layer includes 50 artificial neurons.

[0025] In some embodiments of the second aspect, the second fully connected layer comprises 10 artificial neurons.

[0026] In some embodiments of the second aspect, the first fully connected layer and second fully connected layer apply a transfer function.

[0027] In some embodiments of the second aspect, the transfer function is the tansig(x) function.

[0028] In some embodiments of the second aspect, the first fully connected layer and the second fully connected layer are developed by training via stochastic gradient descent to produce a set of deep features.

[0029] The invention will now be further described by way of example with reference to the accompanying drawings, in which:

Fig. 1 is a perspective view of a camera mounted in a room and a patient support apparatus positioned in a field of view of the camera;

Fig. 2 is a process flow chart for a controller to monitor images from the camera and predict, based on information derived from the images, that a patient is likely to exit the bed in some predetermined future time interval;

Fig. 3 is a model of the prediction system presently disclosed;

Fig. 4 is an illustration of a first segmentation approach used to segment video data when the patient support apparatus is arranged in a first configuration;

Fig. 5 is an illustration of a first segmentation approach used to segment video data when the patient support apparatus is arranged in a second configuration;

Fig. 6 is a plot of the relationship of specificity to sensitivity for a particular dataset analyzed by the prediction system of the present disclosure.

[0030] Referring to Fig. 1, a patient 10 is shown to be positioned on a patient support apparatus 12, illustratively embodied as a hospital bed. In other embodiments, the patient support apparatus may be embodied as a chair, a wheelchair, a table, a gurney, a stretcher, or the like. The hospital bed 12 has a base 14, a lift system 16, an upper frame

18, and a deck 20 with at least one articulated deck section 22. The articulated deck section 22 is movable relative to the upper frame 18 to a plurality of positions. A camera 24 is positioned on a ceiling (not shown) of a room 26 in a fixed location having coordinates X, Y, Z relative to an origin or datum 28 of the room 26. The camera 24 is connected to a controller 25 which receives signals form the camera 24 indicative of the images captured by the camera 24. The controller 25 is operable to perform use the signal to establish the likelihood that a patient will exit the patient support apparatus in a predetermined time period, as will be discussed in further detail below. If such a signal is generated, the signal is acted upon by an alarm system 82 to provide either an electronic record of the alarm condition, an auditory alarm, or a visual alarm, or any combination of the three.

[0031] The relative location of the bed 12 is established by identifying the position of two components of the bed 12. For example, the location of a first caster 30 is defined by coordinates X', Y', Z' and the location of a second caster 32 is defined by coordinates X", Y", Z". The relative position of the bed 10 to the camera 24 is used by an algorithm to rectify the position of the bed 10 in predicting bed exits from the video data from the camera 24. The presently disclosed analysis platform is a deep convolutional neural network trained to predict patient exits from the bed 10 from video data.

[0032] To achieve consistent temporal resolution from a wide variety of video data, all video sequences are subsampled to a frame rate of one frame per second. It should be understood that while a frame rate of one frame per second is used in the illustrative embodiment, other frame rates may be used. Referring to the process 42 shown in Fig. 2 which applies a model shown in Fig. 3, windowed sequences of video image frames 34 are captured at a step 44 and first processed by a 1×1 3D convolutional layer 36 at step 46 to extract feature maps. In some embodiments, the feature maps are rectified as suggested by optional step 60. These feature maps are then subjected to a context-aware pooling layer 38 at step 48 where meaningful feature vectors are extracted from each feature map, independent of the video camera's view point. At step 50, the feature map is processed at a first fully connected layer (FC1). The output of the first fully connected layer is then processed by a second fully connected layer (FC2) at step 52. The two fully connected layers FC1 and FC2 impose non-linearity into the extracted feature set. These values proceed to an output layer 40 analysis at step 54 with two values being output: one corresponding to the likelihood of a bed exit event occurring in the near future 56, and the other corresponding to the likelihood of a bed exit event not occurring in the near future 58.

[0033] The lxl 3D convolutional layer 36 is applied at step 44 to the windowed sequence of image frames 34, defined as an $n \times m \times (3 \times N)$ matrix where n and m correspond to the width and height of each frame, N corresponds to the number of frames contained in the windowed sequence, and 3 represents the number of color channels for each image. Convolution in image processing is the process of adding each element of the image to its local neighbors, weighted by a kernel. A kernel is a signal processing value, generalized as a matrix, with the kernel being developed based on empirical data. For example, in a particular case where two three-by-three matrices, one a kernel, and the other an image piece, is convolved by flipping both the rows and columns of the kernel and then multiplying locationally similar image values and summing the products. The [2,2] element of the resulting image would be a weighted combination of all the entries of the image matrix, with weights given by the kernel:

$$\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} * \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\right)[2,2] = (i*1)+(h*2)+(g*3)+(f*4)+(e*5)+(d*6)+(c*7)+(b*8)+(a*9).$$

The other entries would be similarly weighted, where the kernel is positioned on the center of each of the boundary points of the image, and the weighted values are summed.

[0034] In the present disclosure, the windowed sequence image frame 34 matrix is inputted into the 3D convolutional layer 36, where a $1 \times 1 \times (N \times 3)$ 3D convolutional kernel is convolved with each set of windowed video frames 34 to produce a set of N feature maps. This kernel is convolved along the first two dimensions (xy-plane) per pixel and convolved along the third dimension with a temporal stride of 3 to account for the three color channels of each frame. Again, the kernel is determined empirically.

[0035] At step 60 of the process 42, an absolute value rectifier is applied to introduce non-saturating nonlinearity to the layer. It should be understood that other functions may be applied to rectify the convolution layer in other embodiments and that the absolute value rectification is but one example of a function that drives non-saturating nonlinearity to the output from the convolution layer 36.

[0036] To achieve context-aware pooling, meaningful feature vectors must be extracted from the feature maps produced in the previous layer to accurately predict a bed exit event. The feature maps contain information from the video camera's entire field of view 62 (See Fig. 1); however, to rectify data collected from cameras with different fields of view, only the region of interest (i.e., the hospital bed) is considered. Context-aware pooling is used to extract relevant features from the region of interest.

[0037] To determine the region of interest, a model of the hospital bed consisting of six regions 70a, 72a, 74a, 76a, 78a, and 80a is constructed as suggested in Fig. 4. The hospital bed 10 remains stationary relative to the video camera 24 throughout the duration of a video sequence, and a model of the hospital bed is built only once per video sequence. To facilitate context-aware pooling, each feature map is considered as a lattice of nodes where each node, x, represents a single feature within the map. For each feature map, nodes within each region, generalized as *h,* are subject to pooling and then concatenated to create a $1 \times (j \times N)$ feature vector as shown in Figs. 3 and 4. This is expressed mathematically as

$$y = pool\ (x_i \mid i \in h_j)\ for\ all\ j; \qquad (1)$$

where x_i represents the i^{th} node in the region h_j,j is in the range [1, 6], and *pool(.)* is the pooling function. It should be understood that additional granularity may be developed by increasing the number of regions from 6 to some larger value.

[0038] In situations where the position or orientation of the hospital bed 10 is in a different relationship to the camera 24 or if the bed 10 is in a different configuration, a different arrangement of six regions, 70b, 72b, 74b, 76b, 78b, and 80b may be applied to create the feature vector as suggested in Fig. 5. Using this approach, a rectification factor may be applied to correct for the differences between the developed kernel and the appropriate kernel to apply in a particular application.

[0039] Two fully connected layers, FC1 and FC2, exist in this platform to provide non-linearity. Each element of the feature vector is connected to each of the 50 artificial neurons contained in FC1. In turn, each artificial neuron in FC1 is connected to each of the 10 artificial neurons contained in FC2, which are then connected to the output layer. These layers are trained via stochastic gradient descent in order to produce a set of deep features that enable prediction of whether a patient exit event will occur in the near future. Both fully connected layers use the tansig transfer function, defined as:

$$tansig(x) = \frac{2}{1+\exp(-2x)} - 1 \qquad (2)$$

[0040] The output layer employs soft-max pooling and provides an interpretable likelihood of a bed exit event, with one artificial neuron representing the likelihood of a bed exit event occurring and the other artificial neuron representing the likelihood that a bed exit event will not occur. The likelihood is provided as a decimal value ranging from 0 to 1, where 0 suggests that a bed exit event is very unlikely and 1 suggests that a bed exit event is very likely.

EXAMPLE

[0041] To verify the efficacy of the system, a dataset consisting of 187 hours of 3 channel red-green-blue (RGB) video data was used. The dataset was segmented into 1009 windowed sequences of data, each 40 seconds in length. These windows were manually selected such that, for positive samples, a bed exit event occurred at 30 seconds, while no bed exit event occurred within the windowed sequence for negative samples.

[0042] A correct positive prediction was said to have occurred if the predicted likelihood of a bed exit event exceeded a confidence threshold, τ, at any point in the seven seconds preceding the bed exit event whereas a correct negative prediction was said to have occurred if the predicted likelihood of a bed exit event remained below τ in the same time interval. The number of frames per window, N, was set to 7. The pooling function in the context-aware pooling layer was defined as the average of all nodes within a given region.

[0043] Using the segmented dataset, two experiments were run. First, the value of τ was altered to determine its optimal value with respect to accuracy. Second, the platform's fully connected layers were trained using different amounts of data to observe its effect on accuracy, sensitivity, and specificity.

[0044] While a maximum accuracy of 90% was determined to occur when $\tau = 0.89$, it is important to consider the sensitivity and specificity values in the context of the application. In the case of bed exit prediction, it is preferable to decrease τ in order to increase sensitivity, thereby missing fewer bed exit events. However, there is a sensitivity-specificity tradeoff, and this results in decreased specificity. Fig. 6 shows the relationship between specificity and sensitivity as τ is varied from 0 to 1. A value of $\tau = 0.73$ was empirically determined to provide a good balance between specificity and sensitivity for the application of bed exit prediction.

[0045] Using an empirically determined τ of 0.73, the values of accuracy, sensitivity, and specificity for three different platforms are tabulated in Table I. The platforms vary only in their fully connected layer, where each platform was trained using a different amount of data. Sensitivity decreases before slightly increasing and the specificity is initially improved before slightly dropping. Sensitivity and specificity are dependent on the number of positive and negative samples and

their fluctuation is attributed to a varying ratio of positive and negative samples as the amount of training data was increased. Accuracy takes into account both positive and negative samples, and its steady increase indicates that the platform generally performs better when trained on more data.

TABLE I: A confidence interval of $\tau = 0.73$ was used for these experiments.

Hours of Data	Accuracy (%)	Sensitivity (%)	Specificity (%)
180	86.47	78.87	94.07
140	85.13	74.65	95.60
19	83.54	81.18	85.89

[0046] Using the model trained on 180 hours of data, a confidence threshold of $\tau = 0.73$, and the entire unsegmented dataset, the average prediction time and relative between-error intervals were calculated (as shown in Table II). It can be seen that with this threshold, missed bed exit events are expected to occur only once every 3.1 hours while false alarms are expected to occur only once every 11 hours. The missed bed exit interval is thought to be lower than the expected false alarm interval due to the high proportion of negative samples to positive samples within this dataset. It is noted that many of the false positives can be attributed to motion within the bed area. This was most commonly caused by participants moving a tray table or laptop computers on a tray table positioned over top of the bed and patient, but was also sometimes caused by moving shadows. Because all videos were subsampled to a frame rate of one frame per second, false negatives are likely due to rapid bed exits. In this case, the frame rate may be insufficient to capture the bed exit event in detail.

TABLE II: Mean prediction time and predicted time-between error intervals for the platform when trained on 180 hours of data and a confidence threshold of $\tau = 0.73$.

Mean Prediction Time (s)	Expected Missed Bed Exit Interval	Expected False Alarm Interval
10.96 ± 9.01	3.1 hours	11 hours

[0047] Results indicate that the approach is capable of predicting patient bed exit events of up to seven seconds in advance with a high degree of accuracy. It is contemplated that other embodiments may include automating the manual point selection process for generating the hospital bed model in the context-aware pooling layer to enable a completely automated system, as well as adding the capability for the system to automatically adapt in the event the hospital bed moves in relation to the camera view point. In addition, the platform can be further improved by incorporating data from any pre-existing bed sensors into the deep convolutional neural network architecture.

Claims

1. An apparatus for predicting that a patient is about to exit from a patient support apparatus (12), the apparatus comprising
a camera (24) positioned with the patient support apparatus (12) in the field of view of the camera (24),
a controller (25) receiving signals representative of images from the camera (24), the controller (25) operable to:

capture (44) time sequenced video images from the camera (24),
process the video images in a convolution neural network by:

inputting (46) the time sequenced video images to a convolution layer (36) and the time sequenced video images are convolved with a convolution kernel to produce a defined number of feature maps, each feature map corresponding to each video image, each feature map comprising a lattice of nodes, each node representing a single feature within the map, each feature map being segmented into a region of interest corresponding to a model of the patient support apparatus, the region of interest consisting of a plurality of regions,
inputting (48) the feature maps into a context-aware pooling layer (38) to pool the nodes within each of the plurality of regions to extract features of interest from the feature maps and generate feature vectors,
inputting (50) a feature vector to a first fully connected layer (FC1) such that each element of the feature vector is connected to a plurality of artificial neurons in the first fully connected layer (FC1) and each

combination outputs a first connected layer value,
inputting (52) the values derived by each combination of first fully connected layer (FC1) into a second fully connected layer (FC2) such that each value is connected to a plurality of artificial neurons in the second fully connected layer (FC2) such that each combination outputs a second connected layer value,
inputting (54) the second connected layer values into an output layer (40) which provides a non-exit probability (58) which defines the likelihood that a patient exit event will not occur in a predetermined time and an exit probability (56) which defines the likelihood that a patient exit event will occur in the predetermined time,

then utilizing the non-exit probability (58) and exit probability (56) provided by the convolutional neural network to determine the likelihood of a patient exit event to generate a signal when the determined likelihood of a patient exit event exceeds a threshold value, and
if the signal is generated based on the determined likelihood of a patient exit event exceeds a threshold value, generating a notification of the impending event.

2. The apparatus of claim 1, wherein the convolution layer (36) applies a rectifier when the feature maps are generated.

3. The apparatus of claim 2, wherein the rectifier introduces non-saturating linearity to the features maps.

4. The apparatus of claim 3, wherein the rectifier is an absolute value function.

5. The apparatus of any preceding claim, wherein the context-aware pooling layer (38) rectifies for a variation in the camera's field of view by applying a rectification factor to a developed kernel to adjust for the difference in the camera's current field of view as compared to the camera's field of view at the time the developed kernel was determined.

6. The apparatus of any preceding claim, wherein the first fully connected layer (FC1) includes 50 artificial neurons.

7. The apparatus of any preceding claim, wherein the second fully connected layer (FC2) comprises 10 artificial neurons.

8. The apparatus of any preceding claim, wherein first fully connected layer (FC1) and second fully connected layer (FC2) apply a transfer function.

9. The apparatus of claim 8, wherein the transfer function is the tansig(x) function.

10. The apparatus of any of any preceding claim, wherein the first fully connected layer (FC1) and the second fully connected layer (FC2) are developed by training via stochastic gradient descent to produce a set of deep features.

11. A method of predicting that a patient is about to exit from a patient support apparatus that is in the field of view of a camera (24), the method comprising:

receiving signals representative of images from the camera (24),
capturing (44) time sequenced video images from the camera (24),
processing the video images in a convolution neural network by:

inputting (46) the time sequenced video images to a convolution layer (36) and convolving the time sequenced video images with a convolution kernel to produce a defined number of feature maps, each feature map corresponding to each video image, each feature map comprising a lattice of nodes, each node representing a single feature within the map, each feature map being segmented into a region of interest corresponding to a model of the patient support apparatus, the region of interest consisting of a plurality of regions,
inputting (48) the feature maps into a context-aware pooling layer (38) to pool the nodes within each of the plurality of regions to extract features of interest from the feature maps and generate feature vectors,
inputting (50) a feature vector to a first fully connected layer (FC1) such that each element of the feature vector is connected to a plurality of artificial neurons in the first fully connected layer (FC1) and each combination outputs a first connected layer value,
inputting (52) the values derived by each combination of first fully connected layer (FC1) into a second fully connected layer (FC2) such that each value is connected to a plurality of artificial neurons in the second

fully connected layer (FC2) such that each combination outputs a second connected layer value,
inputting (54) the second connected layer values into an output layer which provides a non-exit probability (58) which defines the likelihood that a patient exit event will not occur in a predetermined time and an exit probability (56) which defines the likelihood that a patient exit event will occur in the predetermined time,

then utilizing the non-exit probability (58) and exit probability (56) provided by the convolutional neural network to determine the likelihood of a patient exit event to generate a signal when the determined likelihood of a patient exit event exceeds a threshold value, and
if the signal is generated based on the determined likelihood of a patient exit event exceeds a threshold value, generating a notification of the impending event.

12. The method of claim 11, wherein the convolution layer (36) applies a rectifier function when the feature maps are generated.

13. The method of claim 12, wherein the rectifier function introduces non-saturating linearity to the features maps.

14. The method of claim 13, wherein the rectifier function is an absolute value function.

15. The method of any one of claims 11 to 14, wherein the context-aware pooling layer (38) rectifies for a variation in the camera's field of view by applying a rectification factor to a developed kernel to adjust for the difference in the camera's current field of view as compared to the camera's field of view at the time the developed kernel was determined.

Patentansprüche

1. Eine Vorrichtung zum Vorhersagen, dass ein Patient gleich aus einer Patiententragevorrichtung (12) aussteigen wird, wobei die Vorrichtung Folgendes umfasst:

eine Kamera (24), die so positioniert ist, dass die Patiententragevorrichtung (12) in dem Sichtfeld der Kamera (24) liegt,
eine Steuerung (25), die Signale empfängt, welche repräsentativ für Bilder von der Kamera (24) sind, wobei die Steuerung (25) wirksam ist, um:

zeitlich aufeinanderfolgende Videobilder von der Kamera (24) zu erfassen (44),
die Videobilder in einem neuronalen Faltungsnetzwerk durch folgende Schritte zu verarbeiten:

Eingeben (46) der zeitlich aufeinanderfolgenden Videobilder in eine Faltungsschicht (36), wobei die zeitlich aufeinanderfolgenden Videobilder mit einem Faltungskern gefaltet werden, um eine definierte Anzahl an Merkmalskarten zu erstellen, wobei jede Merkmalskarte jedem Videobild entspricht, wobei jede Merkmalskarte ein Gitter aus Knoten umfasst, wobei jeder Knoten ein einzelnes Merkmal auf der Karte darstellt, wobei jede Merkmalskarte in einen Bereich von Interesse segmentiert ist, der einem Modell der Patiententragevorrichtung entspricht, wobei der Bereich von Interesse aus einer Mehrzahl von Bereichen besteht,
Eingeben (48) der Merkmalskarten in eine kontext-bewusste Aggregationsschicht (38), um die Knoten in jedem der Mehrzahl von Bereichen zu aggregieren, um Merkmale von Interesse aus den Merkmalskarten zu extrahieren und Merkmalsvektoren zu erzeugen,
Eingeben (50) eines Merkmalsvektors in eine erste vollständig verbundene Schicht (FC1), so dass jedes Element des Merkmalsvektors mit einer Mehrzahl von künstlichen Neuronen in der ersten vollständig verbundenen Schicht (FC1) verbunden ist und jede Kombination einen Wert der ersten verbundenen Schicht ausgibt,
Eingeben (52) der durch jede Kombination der ersten vollständig verbundenen Schicht (FC1) abgeleiteten Werte in eine zweite vollständig verbundene Schicht (FC2), so dass jeder Wert mit einer Mehrzahl von künstlichen Neuronen in der zweiten vollständig verbundenen Schicht (FC2) verbunden ist, so dass jede Kombination einen Wert der zweiten verbundenen Schicht ausgibt,
Eingeben (54) der Werte der zweiten verbundenen Schicht in eine Ausgabeschicht (40), die eine Wahrscheinlichkeit eines Nicht-Ausstieges (58), welche die Wahrscheinlichkeit definiert, dass ein Patientenausstiegsereignis nicht in einer vorbestimmten Zeit auftritt, und eine Wahrscheinlichkeit eines

Ausstieges (56) bereitstellt, welche die Wahrscheinlichkeit definiert, dass ein Patientenausstiegsereignis in der vorbestimmten Zeit auftritt,

darauffolgendes Nutzen der von dem neuronalen Faltungsnetzwerk bereitgestellten Wahrscheinlichkeit eines Nicht-Ausstieges (58) und der Wahrscheinlichkeit eines Ausstieges (56), um die Wahrscheinlichkeit eines Patientenausstiegsereignisses zu bestimmen, um ein Signal zu erzeugen, wenn die bestimmte Wahrscheinlichkeit eines Patientenausstiegsereignisses einen Schwellwert überschreitet, und
wenn das Signal basierend darauf erzeugt wird, dass die Wahrscheinlichkeit eines Patientenausstiegsereignisses einen Schwellwert überschreitet, Erzeugen einer Benachrichtigung über das bevorstehende Ereignis.

2. Die Vorrichtung nach Anspruch 1, wobei die Faltungsschicht (36) einen Gleichrichter anwendet, wenn die Merkmalskarten erzeugt werden.

3. Die Vorrichtung nach Anspruch 2, wobei der Gleichrichter eine nicht-sättigende Linearität in die Merkmalskarten einfügt.

4. Die Vorrichtung nach Anspruch 3, wobei der Gleichrichter eine Betragsfunktion ist.

5. Die Vorrichtung nach einem der vorstehenden Ansprüche, wobei die kontext-bewusste Aggregationsschicht (38) eine Gleichrichtung in Bezug auf eine Änderung im Sichtfeld der Kamera ausführt, indem ein Gleichrichtungsfaktor auf einen entwickelten Kern angewendet wird, um die Differenz in dem aktuellen Sichtfeld der Kamera im Vergleich zu dem Sichtfeld der Kamera zum Zeitpunkt der Bestimmung des entwickelten Kerns auszugleichen.

6. Die Vorrichtung nach einem der vorstehenden Ansprüche, wobei die erste vollständig verbundene Schicht (FC1) 50 künstliche Neuronen enthält.

7. Die Vorrichtung nach einem der vorstehenden Ansprüche, wobei die zweite vollständig verbundene Schicht (FC2) 10 künstliche Neuronen enthält.

8. Die Vorrichtung nach einem der vorstehenden Ansprüche, wobei die erste vollständig verbundene Schicht (FC1) und die zweite vollständig verbundene Schicht (FC2) eine Übertragungsfunktion anwenden.

9. Die Vorrichtung nach Anspruch 8, wobei die Übertragungsfunktion eine tansig(x)-Funktion ist.

10. Die Vorrichtung nach einem der vorstehenden Ansprüche, wobei die erste vollständig verbundene Schicht (FC1) und die zweite vollständig verbundene Schicht (FC2) durch Training über das stochastische Gradientenabstiegsverfahren entwickelt werden, um einen Satz von tiefen Merkmalen zu erstellen.

11. Ein Verfahren zum Vorhersagen, dass ein Patient gleich aus einer Patiententragevorrichtung aussteigen wird, die in dem Sichtfeld einer Kamera (24) liegt, wobei das Verfahren folgende Schritte umfasst:

Empfangen von Signalen, welche repräsentativ für Bilder von der Kamera (24) sind,
Erfassen (44) von zeitlich aufeinanderfolgenden Videobildern von der Kamera (24),
Verarbeiten der Videobilder in einem neuronalen Faltungsnetzwerk durch folgende Schritte:

Eingeben (46) der zeitlich aufeinanderfolgenden Videobilder in eine Faltungsschicht (36) und Falten der zeitlich aufeinanderfolgenden Videobilder mit einem Faltungskern, um eine definierte Anzahl an Merkmalskarten zu erstellen, wobei jede Merkmalskarte jedem Videobild entspricht, wobei jede Merkmalskarte ein Gitter aus Knoten umfasst, wobei jeder Knoten ein einzelnes Merkmal auf der Karte darstellt, wobei jede Merkmalskarte in einen Bereich von Interesse segmentiert ist, der einem Modell der Patiententragevorrichtung entspricht, wobei der Bereich von Interesse aus einer Mehrzahl von Bereichen besteht,
Eingeben (48) der Merkmalskarten in eine kontext-bewusste Aggregationsschicht (38), um die Knoten in jedem der Mehrzahl von Bereichen zu aggregieren, um Merkmale von Interesse aus den Merkmalskarten zu extrahieren und Merkmalsvektoren zu erzeugen,
Eingeben (50) eines Merkmalsvektors in eine erste vollständig verbundene Schicht (FC1), so dass jedes Element des Merkmalsvektors mit einer Mehrzahl von künstlichen Neuronen in der ersten vollständig verbundenen Schicht (FC1) verbunden ist und jede Kombination einen Wert der ersten verbundenen Schicht

ausgibt,
Eingeben (52) der durch jede Kombination der ersten vollständig verbundenen Schicht (FC1) abgeleiteten Werte in eine zweite vollständig verbundene Schicht (FC2), so dass jeder Wert mit einer Mehrzahl von künstlichen Neuronen in der zweiten vollständig verbundenen Schicht (FC2) verbunden ist, so dass jede Kombination einen Wert der zweiten verbundenen Schicht ausgibt,
Eingeben (54) der Werte der zweiten verbundenen Schicht in eine Ausgabeschicht, die eine Wahrscheinlichkeit eines Nicht-Ausstieges (58), welche die Wahrscheinlichkeit definiert, dass ein Patientenausstiegsereignis nicht in einer vorbestimmten Zeit auftritt, und eine Wahrscheinlichkeit eines Ausstieges (56) bereitstellt, welche die Wahrscheinlichkeit definiert, dass ein Patientenausstiegsereignis in der vorbestimmten Zeit auftritt, darauffolgendes Nutzen der von dem neuronalen Faltungsnetzwerk bereitgestellten Wahrscheinlichkeit eines Nicht-Ausstieges (58) und der Wahrscheinlichkeit eines Ausstieges (56), um die Wahrscheinlichkeit eines Patientenausstiegsereignisses zu bestimmen, um ein Signal zu erzeugen, wenn die bestimmte Wahrscheinlichkeit eines Patientenausstiegsereignisses einen Schwellwert überschreitet, und wenn das Signal basierend darauf erzeugt wird, dass die Wahrscheinlichkeit eines Patientenausstiegsereignisses einen Schwellwert überschreitet, Erzeugen einer Benachrichtigung über das bevorstehende Ereignis.

12. Das Verfahren nach Anspruch 11, wobei die Faltungsschicht (36) eine Gleichrichterfunktion anwendet, wenn die Merkmalskarten erzeugt werden.

13. Das Verfahren nach Anspruch 12, wobei die Gleichrichterfunktion eine nicht-sättigende Linearität in die Merkmalskarten einfügt.

14. Das Verfahren nach Anspruch 13, wobei die Gleichrichterfunktion eine Betragsfunktion ist.

15. Das Verfahren nach einem der Ansprüche 11 bis 14, wobei die kontext-bewusste Aggregationsschicht (38) eine Gleichrichtung in Bezug auf eine Änderung im Sichtfeld der Kamera ausführt, indem ein Gleichrichtungsfaktor auf einen entwickelten Kern angewendet wird, um die Differenz in dem aktuellen Sichtfeld der Kamera im Vergleich zu dem Sichtfeld der Kamera zum Zeitpunkt der Bestimmung des entwickelten Kerns auszugleichen.

Revendications

1. Appareil destiné à prédire qu'un patient est sur le point de sortir d'un appareil de support de patient (12), l'appareil comprenant
une caméra (24) positionnée de sorte que l'appareil de support de patient (12) soit dans le champ de vision de la caméra (24),
un contrôleur (25) recevant des signaux représentant des images provenant de la caméra (24), le contrôleur (25) pouvant être utilisé pour :

capturer (44) des séquences d'images vidéo provenant de la caméra (24),
traiter les images vidéo dans un réseau neuronal convolutif :

en entrant (46) les séquences d'images vidéo dans une couche convolutive (36) et en convoluant les séquences d'images vidéo avec un noyau convolutif afin de produire un nombre défini de cartes de caractéristiques, chaque carte de caractéristiques correspondant à chaque image vidéo, chaque carte de caractéristiques comprenant un réseau de nœuds, chaque nœud représentant une caractéristique unique au sein de la carte, chaque carte de caractéristiques étant segmentée en une région d'intérêt correspondant à un modèle de l'appareil de support de patient, la région d'intérêt étant constituée d'une pluralité de régions,
en entrant (48) les cartes de caractéristiques dans une couche de regroupement sensible au contexte (38) pour regrouper les nœuds au sein de chacune de la pluralité de régions afin d'extraire des caractéristiques d'intérêt à partir des cartes de caractéristiques et de générer des vecteurs de caractéristiques,
en entrant (50) un vecteur de caractéristiques dans une première couche entièrement connectée (FC1) telle que chaque élément du vecteur de caractéristiques soit connecté à une pluralité de neurones artificiels dans la première couche entièrement connectée (FC1) et chaque combinaison donne en sortie une valeur de la première couche connectée,
en entrant (52) les valeurs obtenues par chaque combinaison de la première couche entièrement connectée (FC1) dans une deuxième couche entièrement connectée (FC2) telle que chaque valeur est connectée à

une pluralité de neurones artificiels dans la deuxième couche entièrement connectée (FC2) telle que chaque combinaison donne en sortie une valeur de la deuxième couche connectée,
en entrant (54) les valeurs de la deuxième couche connectée dans une couche de sortie (40) qui fournit une probabilité de non-sortie (58) qui définit la probabilité qu'un événement de sortie du patient ne se produise pas dans un délai prédéfini ainsi qu'une probabilité de sortie (56) qui définit la probabilité qu'un événement de sortie du patient se produise dans un délai prédéfini,

puis en utilisant la probabilité de non-sortie (58) et la probabilité de sortie (56) fournies par le réseau neuronal convolutif pour déterminer la probabilité d'un événement de sortie du patient afin de générer un signal lorsque la probabilité déterminée d'un événement de sortie du patient est supérieure à une valeur seuil, et
si le signal est généré parce que la probabilité déterminée d'un événement de sortie du patient est supérieure à une valeur seuil, en générant un avertissement concernant l'événement imminent.

2. Appareil selon la revendication 1, dans lequel la couche convolutive (36) applique un redresseur lorsque les cartes de caractéristiques sont générées.

3. Appareil selon la revendication 2, dans lequel le redresseur introduit une non-linéarité saturante dans les cartes de caractéristiques.

4. Appareil selon la revendication 3, dans lequel le redresseur est une fonction valeur absolue.

5. Appareil selon l'une quelconque des revendications précédentes, dans lequel la couche de regroupement sensible au contexte (38) corrige une variation du champ de vision de la caméra en appliquant un facteur de redressement à un noyau développé afin de s'adapter à la différence du champ de vision actuel de la caméra par rapport au champ de vision de la caméra au moment où le noyau développé a été déterminé.

6. Appareil selon l'une quelconque des revendications précédentes, dans lequel la première couche entièrement connectée (FC1) comprend 50 neurones artificiels.

7. Appareil selon l'une quelconque des revendications précédentes, dans lequel la deuxième couche entièrement connectée (FC2) comprend 10 neurones artificiels.

8. Appareil selon l'une quelconque des revendications précédentes, dans lequel une première couche entièrement connectée (FC1) et une deuxième couche entièrement connectée (FC2) appliquent une fonction de transfert.

9. Appareil selon la revendication 8, dans lequel la fonction de transfert est la fonction tansig(x).

10. Appareil selon l'une quelconque des revendications précédentes, dans lequel la première couche entièrement connectée (FC1) et la deuxième couche entièrement connectée (FC2) sont développées par entraînement via une descente stochastique de gradient afin de produire un ensemble de caractéristiques profondes.

11. Procédé destiné à prédire qu'un patient est sur le point de sortir d'un appareil de support de patient qui est dans le champ de vision d'une caméra (24), le procédé comprenant :

la réception de signaux représentant des images provenant de la caméra (24),
la capture (44) de séquences d'images vidéo provenant de la caméra (24),
le traitement des images vidéo dans un réseau neuronal convolutif :

en entrant (46) les séquences d'images vidéo dans une couche convolutive (36) et en convoluant les séquences d'images vidéo avec un noyau convolutif afin de produire un nombre défini de cartes de caractéristiques, chaque carte de caractéristiques correspondant à chaque image vidéo, chaque carte de caractéristiques comprenant un réseau de nœuds, chaque nœud représentant une caractéristique unique au sein de la carte, chaque carte de caractéristiques étant segmentée en une région d'intérêt correspondant à un modèle de l'appareil de support de patient, la région d'intérêt étant constituée d'une pluralité de régions,
en entrant (48) les cartes de caractéristiques dans une couche de regroupement sensible au contexte (38) pour regrouper les nœuds au sein de chacune de la pluralité de régions afin d'extraire des caractéristiques d'intérêt à partir des cartes de caractéristiques et de générer des vecteurs de caractéristiques,
en entrant (50) un vecteur de caractéristiques dans une première couche entièrement connectée (FC1)

telle que chaque élément du vecteur de caractéristiques soit connecté à une pluralité de neurones artificiels dans la première couche entièrement connectée (FC1) et chaque combinaison donne en sortie une valeur de la première couche connectée,

en entrant (52) les valeurs obtenues par chaque combinaison de la première couche entièrement connectée (FC1) dans une deuxième couche entièrement connectée (FC2) telle que chaque valeur est connectée à une pluralité de neurones artificiels dans la deuxième couche entièrement connectée (FC2) telle que chaque combinaison donne en sortie une valeur de la deuxième couche connectée,

en entrant (54) les valeurs de la deuxième couche connectée dans une couche de sortie qui fournit une probabilité de non-sortie (58) qui définit la probabilité qu'un événement de sortie du patient ne se produise pas dans un délai prédéfini et une probabilité de sortie (56) qui définit la probabilité qu'un événement de sortie du patient se produise dans le délai prédéfini, puis en utilisant la probabilité de non-sortie (58) et la probabilité de sortie (56) fournies par le réseau neuronal convolutif pour déterminer la probabilité d'un événement de sortie du patient afin de générer un signal lorsque la probabilité déterminée d'un événement de sortie du patient est supérieure à une valeur seuil, et

si le signal est généré parce que la probabilité déterminée d'un événement de sortie du patient est supérieure à une valeur seuil, en générant un avertissement concernant l'événement imminent.

12. Procédé selon la revendication 11, dans lequel la couche convolutive (36) applique une fonction de redresseur lorsque les cartes de caractéristiques sont générées.

13. Procédé selon la revendication 12, dans lequel la fonction de redresseur introduit une non-linéarité saturante aux cartes de caractéristiques.

14. Procédé selon la revendication 13, dans lequel la fonction de redresseur est une fonction valeur absolue.

15. Procédé selon l'une quelconque des revendications 11 à 14, dans lequel la couche de regroupement sensible au contexte (38) corrige une variation dans le champ de vision de la caméra en appliquant un facteur de redressement à un noyau développé afin de s'adapter à la différence de champ de vision actuel de la caméra par rapport au champ de vision de la caméra au moment où le noyau développé a été déterminé.

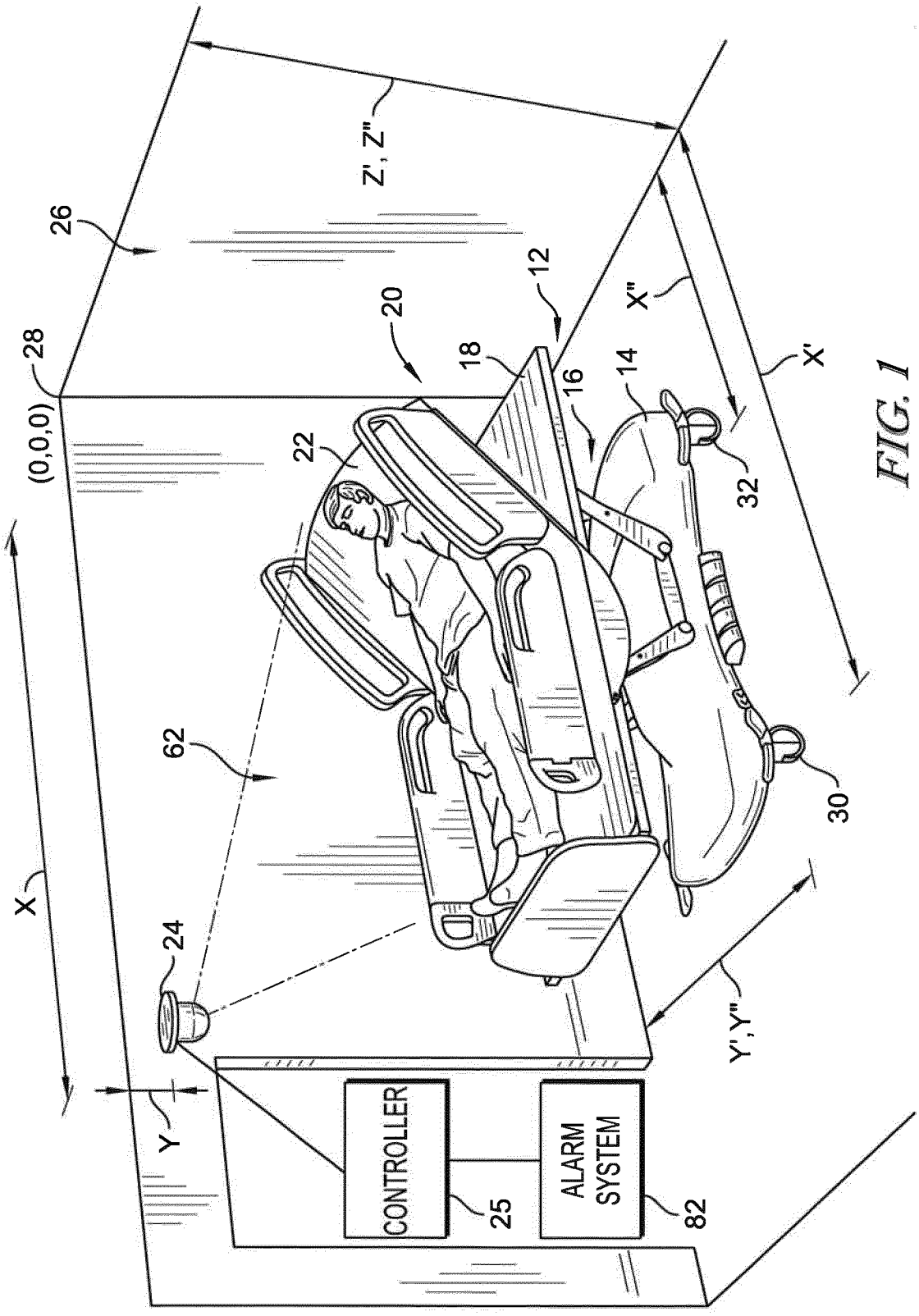
Z', Z"
26
12
X"
20
28
18
16
14
X'
(0,0,0)
22
32
62
30
X
24
Y', Y"
Y
CONTROLLER
25
ALARM SYSTEM
82

FIG. 1

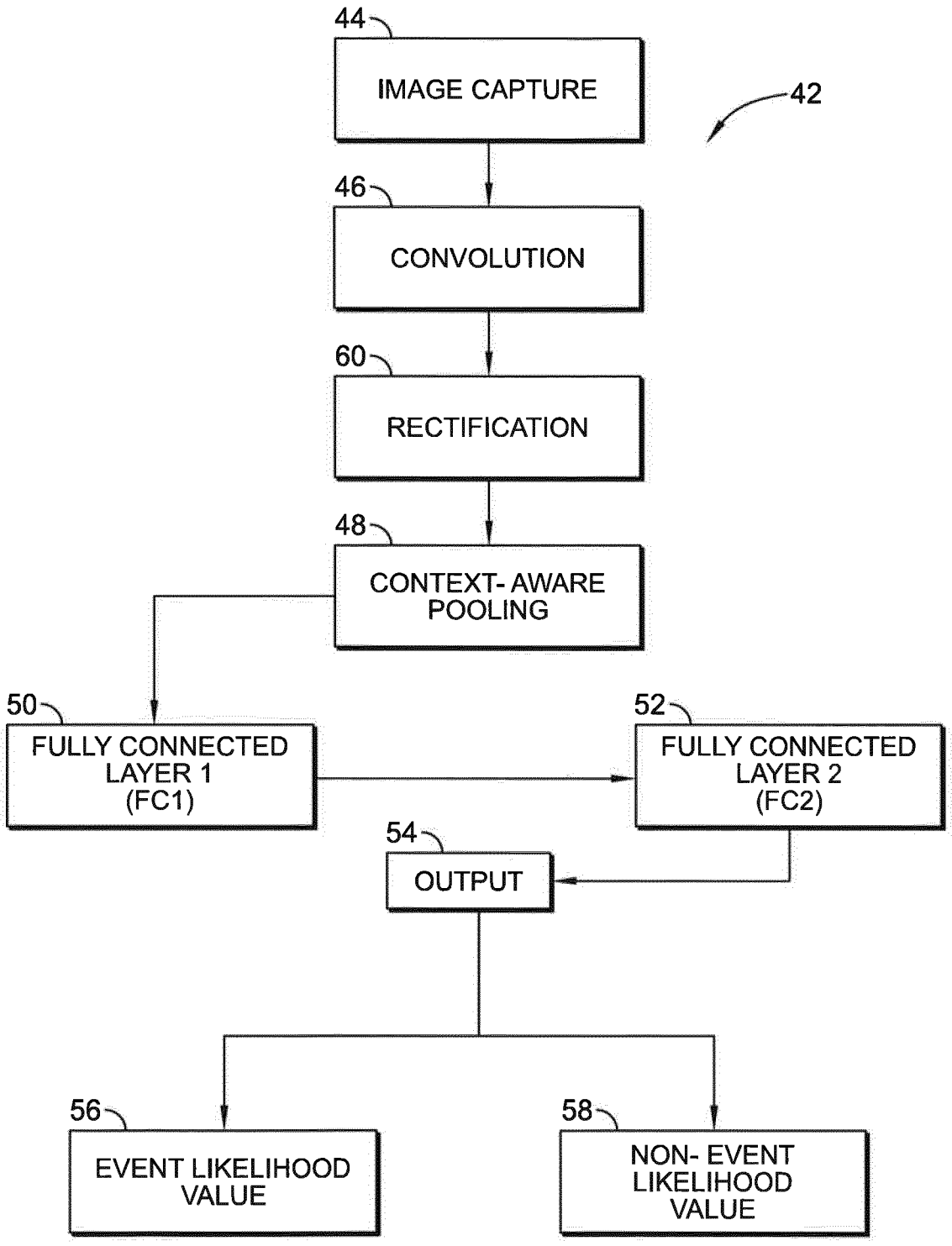
44
IMAGE CAPTURE
42
46
CONVOLUTION
60
RECTIFICATION
48
CONTEXT- AWARE POOLING
50
FULLY CONNECTED LAYER 1 (FC1)
52
FULLY CONNECTED LAYER 2 (FC2)
54
OUTPUT
56
EVENT LIKELIHOOD VALUE
58
NON- EVENT LIKELIHOOD VALUE

FIG. 2

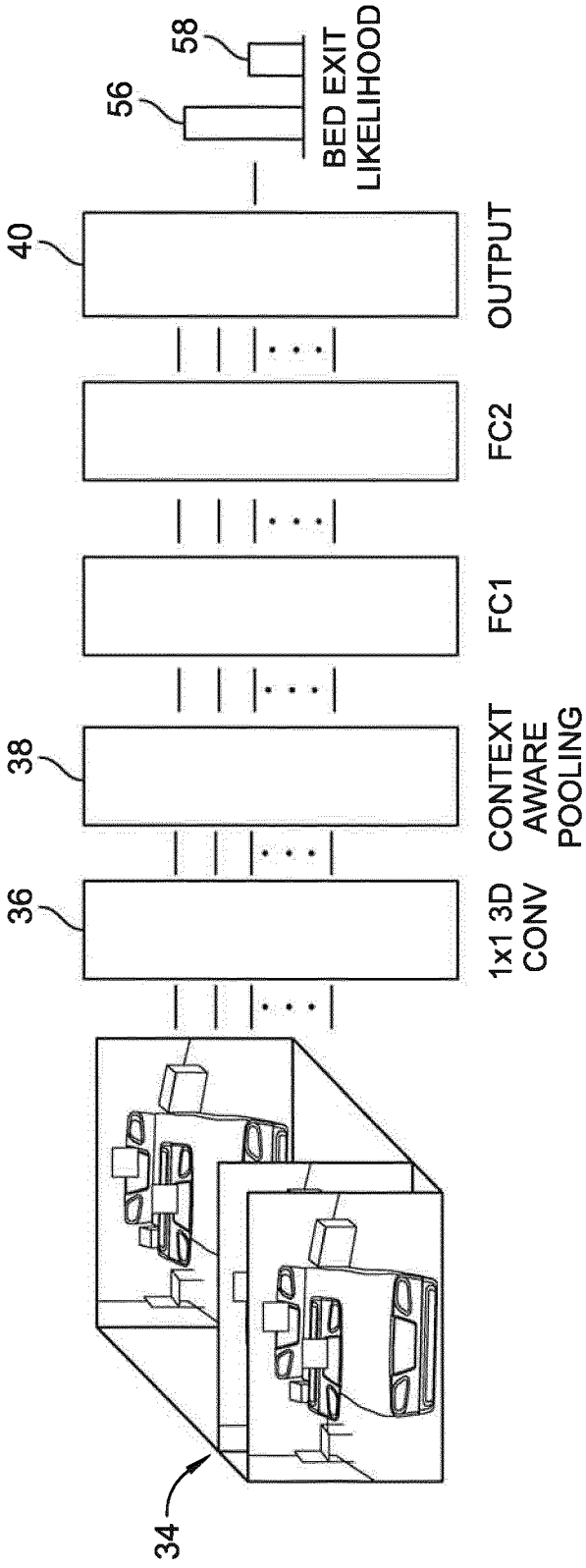
58
56
BED EXIT LIKELIHOOD
40
OUTPUT
FC2
FC1
38
CONTEXT AWARE POOLING
36
1x1 3D CONV
34

FIG. 3

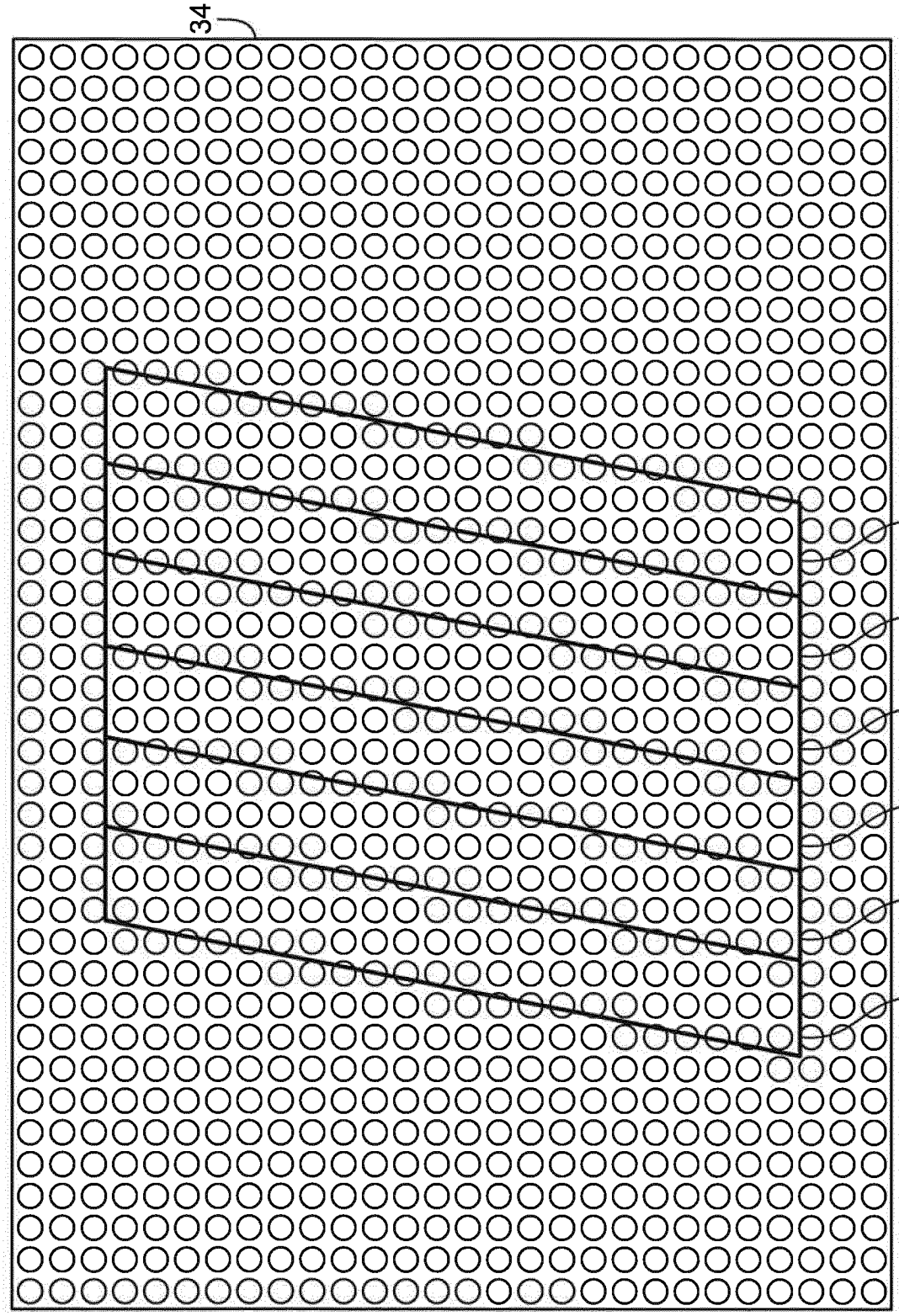

FIG. 4

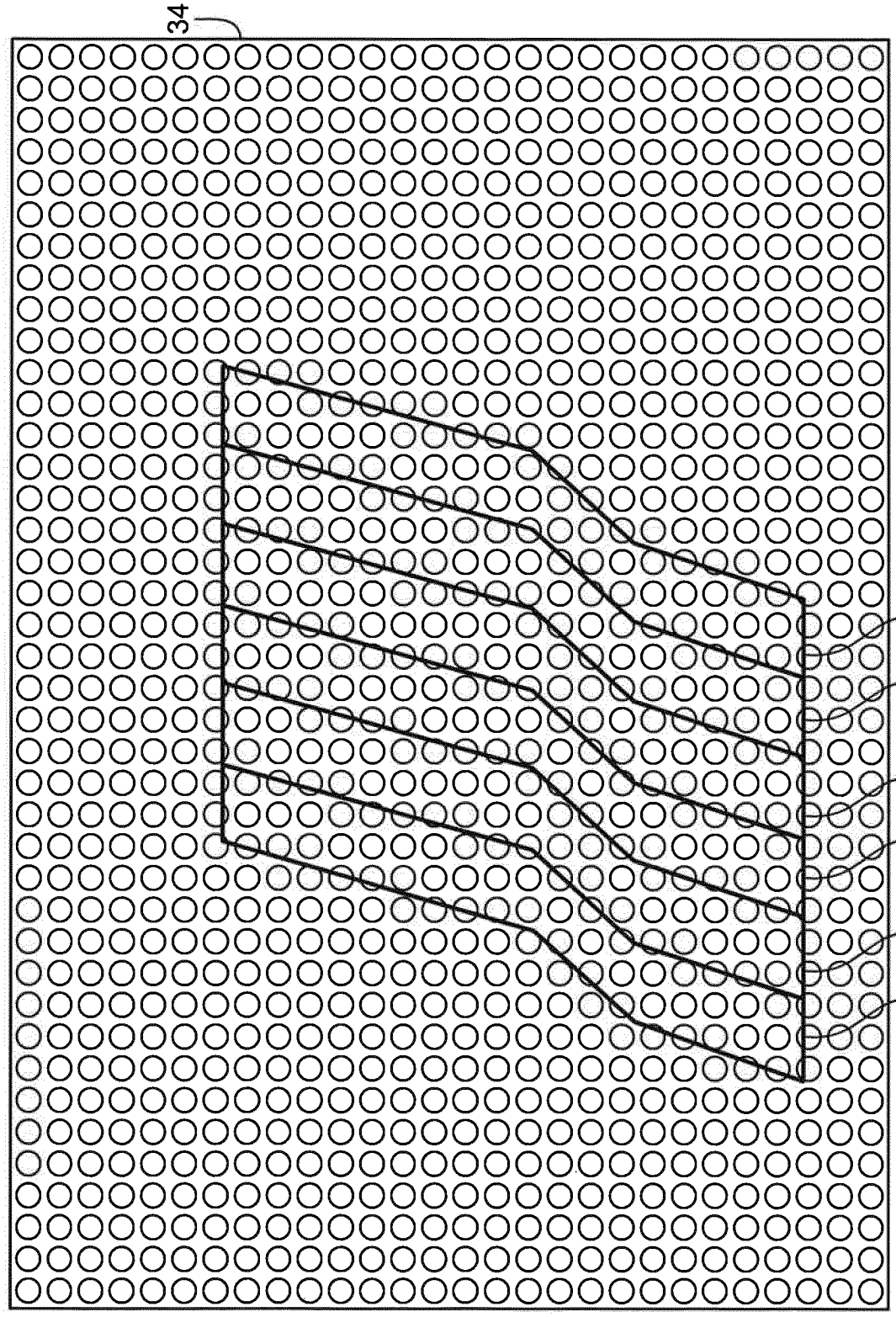

FIG. 5

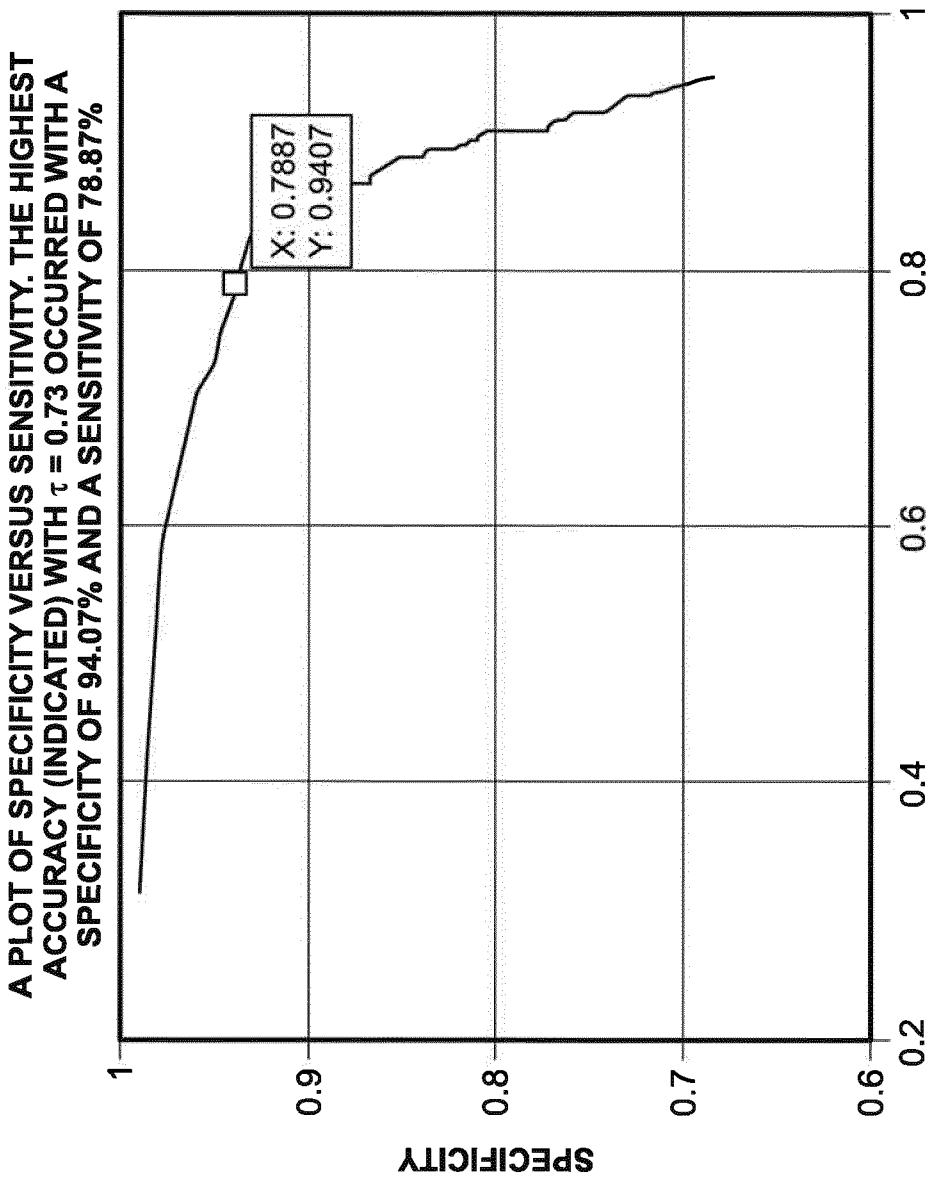
A PLOT OF SPECIFICITY VERSUS SENSITIVITY. THE HIGHEST ACCURACY (INDICATED) WITH τ = 0.73 OCCURRED WITH A SPECIFICITY OF 94.07% AND A SENSITIVITY OF 78.87%
X: 0.7887
Y: 0.9407
SPECIFICITY
1
0.9
0.8
0.7
0.6
SENSITIVITY
0.2
0.4
0.6
0.8
1

FIG. 6

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 5276432 A **[0003]**
- US 20160063846 A **[0004]**

Non-patent literature cited in the description

- Sleep position classification from a depth camera using Bed Aligned Maps. **GRIMM TIMO et al.** 2016 23rd international conference on pattern recognition (ICPR). IEEE, 04 December 2016, 319-324 **[0002]**
- Wiley Electrical and Electronics Engineering Dictionary. John Wiley & Sons, 2004 **[0009]**

专利名称(译)	自动事件预测的程序和设备		
公开(公告)号	EP3358498B1	公开(公告)日	2020-01-29
申请号	EP2018154272	申请日	2018-01-30
[标]申请(专利权)人(译)	希尔 - 罗姆服务股份有限公司		
申请(专利权)人(译)	HILL-ROM SERVICES，INC.		
当前申请(专利权)人(译)	HILL-ROM SERVICES，INC.		
[标]发明人	WONG ALEXANDER SHEUNG LAI CHWYL BRENDAN JAMES CHUNG AUDREY GINA SHAFIEE MOHAMMAD JAVAD FU YONGJI		
发明人	WONG, ALEXANDER SHEUNG LAI CHWYL, BRENDAN JAMES CHUNG, AUDREY GINA SHAFIEE, MOHAMMAD JAVAD FU, YONGJI		
IPC分类号	G06K9/00 G06K9/62 A61B5/11 G06K9/66 G06K9/46 A61B5/00		
CPC分类号	A61B5/0077 A61B5/1115 A61B5/7267 G06K9/00771 G06K9/4628 G06K9/6273 G06K9/66 G16H50/70 G06K9/00342 G06K9/6277 G06N3/0472 G08B21/22		
优先权	62/453857 2017-02-02 US		
其他公开文献	EP3358498A1		
外部链接	Espacenet		

摘要(译)

公开了一种用于预测来自摄像机系统的医院病床离开事件的方法和设备。 该系统使用由五个主要层组成的深度卷积神经网络处理视频数据：一个用于从原始视频数据生成特征图的1×1 3D卷积层，一个用于从不同摄像机角度校正数据的上下文感知池层，两个 完全连接的层用于施加预训练的深部特征，输出层用于提供床退出事件的可能性。