TITLE: Method and device for locating individuals in a three-dimensional frame of reference from two-dimensional images captured by a fixed camera

The present invention relates to a method for locating individuals in a three-dimensional frame of reference from two-dimensional images captured by a fixed camera. The present invention further relates to a location device associated with such method.

In different fields such as security, robotics or navigation, it is necessary to be able to locate an individual located in different environments. Airports, train stations or warehouses are sensitive places and need to be monitored by means of a three-dimensional location of the individuals who pass therethrough despite disruptions due to a plurality of constraints such as crowds or blocking.

To effectively locate individuals in such places, different methods are generally employed depending on needs.

The location methods of the prior art include all types of technologies amongst which: beacons, GPSs, cameras, LIDARs or RFID chips.

All such technologies are used to a variable extent in the field of security and have advantages and disadvantages always according to needs.

Overall, beacons, GPSs, cameras, LIDARs and all other types of sensors still raise many problems of cost and applicability in complex environments whereas RFID chips and beacons as well, involve intrusion and accuracy that are still too limiting.

Cameras remain the most widely used technology to locate individuals in specific environments.

However, it is known that cameras are often limited to two-dimensional surveillance and have still limited performance in complex environments such as train stations or airports.

There is thus a need for a location of individuals in a three-dimensional frame of reference, with good precision and which would be less expensive, less intrusive and more suited to complex environments.

To this end, the subject matter of the present description relates to a method for locating individuals in a three-dimensional frame of reference, comprising a ground plane, from two-dimensional images captured by a fixed camera, the camera having a field of view of a three-dimensional surveillance space, each captured image being represented by a matrix of pixels, each pixel having a position in a two-dimensional frame of reference of the image. The method comprises the following steps:

- geometric calibration of the camera associating each pixel of the two-dimensional images captured by the camera with a Cartesian position in the three-dimensional frame of reference of the surveillance space;
- processing of two-dimensional images, captured by the camera, for the detection of at least one identifiable element of the body of an individual, each identifiable element having an associated position in the two-dimensional frame of reference;

estimation of an absolute position of the individual on the ground plane of the three-dimensional frame of reference as a function of the at least one identifiable element, a predetermined average height of the individual and said geometric calibration. Advantageously, the estimation of the absolute position of the individual as a function of at least one identifiable element of the body of an individual and of the predetermined average height of the individual serves to improve the positioning accuracy in the three-dimensional frame of reference, starting from two-dimensional images.

According to other advantageous aspects of the invention, the method comprises one or a plurality of the following features, taken individually or according to all possible combinations.

The method further includes a reconstruction step including a determination of a silhouette of the individual in three dimensions, starting from the estimation of an absolute position of the individual on the ground plane of the three-dimensional frame of reference and of said predetermined average height of the individual.

The processing step includes a detection of at least 4 identifiable elements.

The processing step includes a detection of a configurable number of identifiable elements.

The step of estimating the absolute position of the individual on the ground plane of the three-dimensional frame of reference includes a calculation of an average of estimated positions of the individual in the three-dimensional frame of reference.

The two-dimensional image processing step is implemented by a neural network.

The identifiable elements of the body of an individual in a two-dimensional image are part of a set comprising the feet, head, hips, shoulders, the eyes or the nose.

The step of estimating an absolute position of the individual on the ground plane of the three-dimensional frame of reference implements a calibration matrix.

A further subject matter of the invention also relates to a device for locating individuals in a three-dimensional frame of reference comprising a ground plane, from two-dimensional images captured by a fixed camera, the device being connected or integrated into said camera, the camera having a field of view of a three-dimensional surveillance

space, each captured image being represented by a matrix of pixels, each pixel having a position in a two-dimensional frame of reference of the image. The device comprises a processor configured for implementing:

- a geometric calibration module of the camera associating each pixel of the two-dimensional images captured by the camera with a Cartesian position in the three-dimensional frame of reference of the surveillance space;
- a processing module for two-dimensional images captured by the camera, [for the detection of at least one identifiable element of the body of an individual, each identifiable element having an associated position in the two-dimensional frame of reference;
- a module for estimating an absolute position of the individual on the ground plane of the three-dimensional frame of reference as a function of the at least one identifiable element, of a predetermined average height of the individual and of said geometric calibration.

According to another aspect, the invention relates to a system for locating individuals in a three-dimensional frame of reference, including a fixed camera and a device for locating individuals in a three-dimensional frame of reference comprising a ground plane, starting from two-dimensional images captured by said fixed camera, the location being such as briefly described hereinabove.

According to another aspect, the invention relates to a computer program including software instructions which, when implemented by a computer system, implement a method of locating individuals in a three-dimensional frame of reference as briefly defined hereinabove.

The invention will be clearer upon reading the following description, given only as an example, but not limited to, and making reference to the drawings wherein:

[Fig. 1] Figure 1 is a schematic representation of a system for locating an individual in a three-dimensional surveillance space comprising a ground plane, said surveillance being implemented by a location associated with a fixed camera;

[Fig.2] Figure 2 is a flowchart of the steps of a method of locating by the location device shown in Figure 1;

[Fig. 3] Figure 3 is a schematic representation of a determination of a silhouette of the individual in the three-dimensional surveillance space.

In Figure 1, a three-dimensional surveillance space 5 comprising an individual 10 is monitored by means of two-dimensional images in a two-dimensional frame of reference

with perpendicular axes (U, V) provided with a frame of reference with a transverse axis V and a longitudinal axis U, the frame of reference being centered on the lower left edge of the images captured by a camera 15, e.g. an optical camera.

The monitoring space 5 is included in a three-dimensional frame of reference comprising a longitudinal axis X, a transverse axis Y and an anterior-posterior axis Z. For example, the three-dimensional frame of reference is the terrestrial frame of reference (or world frame of reference).

The camera 15 is e.g. a video surveillance camera configured to capture successions of images at a predetermined capture frequency. The camera 15 is positioned spatially at a predetermined position. In other words, the camera 15 has a fixed position in space.

Advantageously, as a result, it is possible to calculate a reliable geometric calibration between the three-dimensional frame of reference and the two-dimensional frame of reference.

The three-dimensional surveillance space 5 is in a direct field of view of the camera 15.

Each captured two-dimensional image is represented by a matrix of pixels, each pixel having a position in the two-dimensional frame of reference associated with the image.

Hereinafter in the description, it is considered that the set of pixels composing the two-dimensional images corresponds to Cartesian positions in the three-dimensional frame of reference.

The camera 15 comprises or is connected to a location device 40. The location device 40 and the camera 15 form a location system 60 for locating an individual in the three-dimensional frame of reference.

The location device 40 is a programmable electronic device, e.g. a computer, configured to communicate via a wired or wireless communication link with the camera 15.

According to a variant, the location device 40 is a processing unit integrated into the camera 15.

The location device 40 for locating individuals in a three-dimensional frame of reference implements a detection of a plurality of identifiable elements 20A, 20B, 20C, 20D of the body of said individual 10, and an estimation of a plurality of estimated positions 22A, 22B, 22C, 22D associated with each of the identifiable elements 20A, 20B, 20C, 20D and of an absolute position 24.

The plurality of identifiable elements 20A, 20B, 20C, 20D is a set of physical parts characteristic of an individual 10 being chosen e.g. from the group consisting of: feet, head, hips, shoulders, eyes or nose.

Moreover, each identifiable element 20A, 20B, 20C, 20D then has a two-dimensional position in the two-dimensional frame of reference of the image.

The plurality of estimated positions 22A, 22B, 22C, 22D are projections in the three-dimensional frame of reference, on the ground plane, along the axis Y of the two-dimensional positions of the plurality of identifiable elements 20A, 20B, 20C, 20D in the two-dimensional frame of reference, the projections using a predetermined average height of the individual.

The location device 40 is configured to estimate an absolute position 24 of the individual 10 considered as a point of the feet of said individual 10 in the three-dimensional frame of reference.

The location device 40 comprises an information processing unit 48 consisting e.g. of a processor 50 and an electronic memory 52 associated with the processor 50.

The location device 40 comprises a geometric calibration module 42, an image processing module 44 and a position estimation module 46.

The geometric calibration module 42 is configured to associate each of the pixels of the two-dimensional images captured by the camera 15 with a Cartesian position in the three-dimensional frame of reference, the pixels being projected onto the ground of the surveillance space 5.

The image processing module 44 is configured to detect at least one identifiable element 20A, 20B, 20C, 20D of the body of an individual 10, each identifiable element 20A, 20B, 20C, 20D having an associated position in the two-dimensional frame of reference.

The position estimation module 46 is configured to estimate an absolute position 24 of the individual on the ground plane of the three-dimensional frame of reference as a function of at least one identifiable element 20A, 20B, 20C, 20D, of a predetermined average height of the individual and of the geometric calibration.

In one embodiment, the geometric calibration module 42, the image processing module 44 and the position estimation module 46 are each produced in the form of a software, or a software brick which can be executed by the processor 50. The memory 52 is then apt to store a calibration, image processing and estimation software.

In a variant (not shown), the geometric calibration module 42, the image processing module 44, and the position estimation module 46 are each produced in the form of a programmable logic component, such as an FPGA (*Field Programmable Gate Array*), or further of a dedicated integrated circuit, such as an ASIC (*Application Specific Integrated Circuit*).

In one embodiment, the image processing module 44 implements a previously trained neural network to recognize, from two-dimensional images, and to position in the

two-dimensional frame of reference of each image, at least one identifiable element of the body of the individual.

In general, the neural network includes an ordered succession of neuron layers, each of which takes the inputs thereof from the outputs of the preceding layer.

More precisely, each layer comprises neurons taking the inputs thereof from the outputs of the neurons of the preceding layer, or from the input variables for the first layer.

In a variant, more complex neural network structures can be envisaged with a layer which can be connected to a layer farther away than the immediately preceding layer.

Each neuron is also associated with an operation, i.e. a type of treatment, to be performed by said neuron within the corresponding processing layer.

Each layer is linked to the other layers by a plurality of synapses. A synaptic weight is associated with each synapse, and each synapse forms a link between two neurons. Same is often a real number that can take positive as well as negative values. In certain cases, synaptic weight is a complex number.

Each neuron is apt to perform a weighted sum of the value(s) received from the neurons of the preceding layer, each value then being multiplied by the respective synaptic weight of each synapse, or link between said neuron and the neurons of the preceding layer, then to apply an activation function, typically a non-linear function, to said weighted sum, and to deliver at the output of said neuron, more particularly to the neurons of the next layer which are connected thereto, the value resulting from the application of the activation function. The activation function is used for introducing a non-linearity in the processing performed by each neuron. The sigmoid function, the hyperbolic tangent function, the Heaviside function are examples of activation functions.

As an optional addition, each neuron is also apt to apply, in addition, a multiplicative factor, also called bias, to the output of the activation function, and the value delivered at the output of said neuron is then the product of the bias value and of the value derived from the activation function.

Such a neural network is trained on a database including images of individuals having identifiable elements 20A, 20B, 20C, 20D as defined hereinabove.

The operation of the location device 40 will now be described with reference to Figure 2 which schematically illustrates, in an embodiment, the main steps of a location method.

The location method includes a first geometric calibration step 100 associating each of the pixels of the two-dimensional images captured by the video device 15 with a Cartesian position in the three-dimensional frame of reference.

For example, the geometric calibration module 42 performs a geometric calibration from calibration matrices, according to the following relation for a point P of Cartesian coordinates (X, Y, Z) in the three-dimensional frame of reference the image of which has coordinates (u.v) in the 2D image:

[MATH 1]

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = M_{int} M_{ext} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

With: $M_{int}M_{ext}$: the projection matrix associating each pixel with a Cartesian position in the three-dimensional frame of reference,

- M_{int}: the so-called intrinsic matrix, the coefficients of which are the intrinsic parameters of the camera
- M_{ext}: the so-called extrinsic matrix, the coefficients of which are extrinsic parameters which can vary according to the position of the camera, the parameters representing a translation and a rotation

Thereby, as a function of a matrix of pixels of the captured image, a calibration matrix is determined associating each point of the three-dimensional frame of reference with a pixel of the matrix of pixels.

For greater accuracy, the geometric calibration is implemented manually by an operator

In addition, during the geometric calibration, possible geometric distortions are taken into account using various methods such as the checkered pattern method.

In a variant, the geometric calibration is implemented semi-automatically or automatically for easy deployment, the automatic calibration method including a search for the horizon line or of vanishing points, starting from structuring elements of the image or from a checkerboard.

In a two-dimensional image processing step 200, the image processing module 44 implements a detection of at least one identifiable element 20A, 20B, 20C, 20D each identifiable element 20A, 20B, 20C, 20D having an associated estimated position 22A, 22B, 22C, 22D.

As explained hereinabove, the image processing step 200 preferably implements a neural network trained beforehand to detect said identifiable elements 20A, 20B, 20C, 20D.

For example, the image processing step 200 implements a deep learning algorithm, among the following algorithms: CMU open-pose, Mask RCNN, AlphaPose.

Preferably, the image processing step 200 includes a detection of at least 4 identifiable elements 20A, 20B, 20C, 20D.

In a variant, the image processing step 200 includes a detection of a configurable number of identifiable elements 20A, 20B, 20C, 20D.

Finally, in a position estimation step 300, the position estimation module 46 calculates an absolute position 24 of the individual 10 in the three-dimensional surveillance space 5, starting from the geometric calibration of the at least one identifiable element 20A, 20B, 20C, 20D of the body of the individual 10 and from a predetermined average height of the individual 10.

The average height of the individual is configurable and is e.g. set at 1m75.

From the predetermined average height of the individual, a predetermined average height of each identifiable element of the individual is estimated.

Thereby, considering an average height of the individual set at 1m75, the predetermined height of the hips of the individual 10 is estimated at 88 cm, the predetermined height of the shoulders and neck of the individual 10 is estimated at 1m52.

Knowing the position of each identifiable element of the body of the individual 10 and the average height of each element Y_e, using the projection matrix, the position:

[MATH 2]

$$Pos_{est(i)} = f(Mint, Mext, Pu, Pv, Ye)$$

With:

- $Pos_{est(i)}$: estimated position of the identifiable element i;
- P_u: two-dimensional position of the identifiable element i along the axis u;
- P_v: two-dimensional position of the identifiable element i along the axis v;
- Y_e: average height of the element i
- f: projection function, which is used to solve [Math1], and which estimates X,Z starting from the projection matrices M_{int} and M_{ext}, of the position of the pixel (P_u, P_v) and of the height Ye.

In one embodiment, the absolute position 24 of the individual in the ground plane of the three-dimensional frame of reference is deduced from the average of the estimated positions on the ground plane 22A, 22B, 22C, 22D associated with each of the identifiable elements 20A, 20B, 20C, 20D and from the geometric calibration such as:

[MATH 3]

$$Pos_{abs} = \frac{Pos_{est(1)} + \cdots + Pos_{est(E)}}{E}$$

With:

- Pos_{abs}: absolute position 24 in the three-dimensional frame of reference,
- E: number of identifiable elements,

In a variant, each identifiable element 20A, 20B, 20C, 20D is associated with a reliability weight serving to determine the absolute position 24 according to a weighted average of the estimated positions 22A, 22B, 22C, 22D of each of the identifiable elements 20A, 20B, 20C, 20D associated with said reliability weight such that:

[MATH 4]

$$Pos_{abs} = \frac{A_1 Pos_{est(1)} + \cdots + A_E Pos_{est(E)}}{A_1 + \cdots + A_E}$$

With:

- A_1: reliability weight of the identifiable element i.

In a variant and with reference to Figure 3, during a reconstruction step 400, a digital silhouette of the individual 10, in the three-dimensional frame of reference, is determined from the estimation of the absolute position 24 of the individual 10 and of at least one predetermined average height of the individual.

In addition, an identifier is associated with each silhouette thereby reconstructed and serves to follow an individual 10 in a crowd.

It should then be understood that the location method, and the associated location device 40, serve to accurately locate an individual 10 from images captured by conventional cameras, making the location economical, flexible and non-intrusive.

The proposed method is particularly useful for obtaining a more precise location of individuals in the presence of occlusions in the field of view of the camera, since the absolute position takes into account an estimated average height as a function of the predetermined average height of the individual for each identified element of the body of the individual.

CLAIMS

1. A method of locating individuals in a three-dimensional frame of reference, comprising a ground plane, from two-dimensional images captured by a fixed camera, the camera having a field of view of a three-dimensional surveillance space (5), each captured image being represented by a matrix of pixels, each pixel having a position in a two-dimensional frame of reference of the image, the method being characterized in that same comprises the following steps:

- geometric calibration (100) of the camera associating each pixel of the two-dimensional images captured by the camera with a Cartesian position in the three-dimensional frame of reference of the surveillance space (5);
- processing (200) of two-dimensional images, captured by the camera, for the detection of at least one identifiable element (20A, 20B, 20C, 20D) of the body of an individual (10), each identifiable element (20A, 20B, 20C, 20D) having an associated position in the two-dimensional frame of reference;
- estimation (300) of an absolute position (24) of the individual on the ground plane of the three-dimensional frame of reference as a function of the at least one identifiable element (20A, 20B, 20C, 20D), of a predetermined average height of the individual and of said geometric calibration.

2. The method according to claim 1, further including a reconstruction step (400) including a determination of a silhouette of the individual (10) in three dimensions, starting from the estimation of an absolute position (24) of the individual (10) on the ground plane of the three-dimensional frame of reference and of said predetermined average height of the individual.

3. The method according to any of claims 1 to 2, wherein the processing step (200) includes at least 4 identifiable elements (20A, 20B, 20C, 20D).

4. The method according to any of claims 1 to 3, wherein the processing step (200) includes a detection of a configurable number of identifiable elements (20A, 20B, 20C, 20D).

5. The method according to any of claims 1 to 4, wherein the step of estimation (300) of the absolute position (24) of the individual on the ground plane of the three-dimensional frame of reference comprises a calculation of an average of estimated

positions (22A, 22B, 22C, 22D) of the individual (10) in the three-dimensional frame of reference.

6. The method according to any of claims 1 to 5, wherein the two-dimensional image processing step (200) is implemented by neural network.

7. The method according to any one of claims 1 to 6, wherein the identifiable elements (20A, 20B, 20C, 20D) of the body of an individual (10) in a two-dimensional image are part of a set comprising the feet, head, hips, shoulders, the eyes or the nose.

8. The method according to any of claims 1 to 7, wherein the step of estimation (300) of an absolute position (24) of the individual (10) on the ground plane of the three-dimensional frame of reference uses a calibration matrix.

9. A computer program including software instructions which, when executed by a computer, implement a location method according to any of claims 1 to 8.

10. A device for locating (40) individuals in a three-dimensional frame of reference comprising a ground plane, from two-dimensional images captured by a fixed camera, the device being connected or integrated into said camera, the camera having a field of view of a three-dimensional surveillance space (5), each captured image being represented by a matrix of pixels, each pixel having a position in a two-dimensional frame of the image, the device being characterized in that same comprises a processor configured to implement:

- a geometric calibration module (42) of the camera associating each pixel of the two-dimensional images captured by the camera with a Cartesian position in the three-dimensional frame of reference of the surveillance space (5);
- a module (44) for processing two-dimensional images, captured by the camera, [for the detection of at least one identifiable element (20A, 20B, 20C, 20D) of the body of an individual (10), each identifiable element (20A, 20B, 20C, 20D) having an associated position in the two-dimensional frame of reference;
- a module (46) for estimating an absolute position (24) of the individual on the ground plane of the three-dimensional frame of reference as a function of at least one identifiable element (20A, 20B, 20C, 20D), of a predetermined average height of the individual and of said geometric calibration.

11. A location system (60) for locating individuals in a three-dimensional frame of reference, characterized in that same comprises a fixed camera (15) and a location device (40) for locating individuals in a three-dimensional frame of reference comprising a ground plane, from two-dimensional images captured by said fixed camera (15), the location (40) being in accordance with claim 10.