# Method for counting objects in a predetermined spatial area

The present invention relates to a method for counting objects in a predetermined spatial area. The present invention also relates to an associated computer program product and information medium. The present invention also relates to a detection device able to carry out the counting method.

With the increasing use of video surveillance in public places to improve safety, many techniques for analyzing the videos provided by cameras have been developed.

Document EP 2,704,060 A2 in particular teaches a method for estimating the size of a crowd from an image, the method comprising the steps of determining a candidate to obtain the most informative data point by using an active learning algorithm. The method also includes a step for receiving an annotation from a user identifying the data from the most informative data point. The method also includes a step for applying a regression function to the determined most informative point as well as to the associated notation to minimize a loss on the annotated datum. The method lastly comprises a step for building a regression model by using the data annotated by the user and the deduced annotated data.

Document US 2000/0031005 A1 describes an optical system comprising a camera and a processor including video analysis software. The software is suitable for estimating a reference background frame representative of a background, estimating geometric parameters to represent an object variation scale in the given frame, obtaining a change detection map to distinguish the background from an object in the given frame and combining the change detection map with the geometric parameters to obtain a measurement of the congestion in the given frame.

However, when the spatial density of objects increases, the aforementioned methods provide unreliable results to determine the spatial density of objects in the provided frame.

There is therefore a need for a method for determining objects in a predetermined spatial area making it possible to obtain more reliable estimates, in particular when the spatial density of objects is high.

To that end, a method for counting objects in a predetermined spatial area is proposed. The method includes a first unsupervised learning stage comprising the steps of choosing a plurality of separate spatial densities of objects in the predetermined spatial area, providing parameters relative to a detector able to acquire at least one image, the image including the predetermined spatial area, the provided parameters comprising at

least a parameter relative to the focal length of the detector, a parameter relative to the altitude of the detector and a parameter relative to the incline of the detector with respect to the predetermined spatial area, generating a database including a plurality of images for each chosen spatial density, the generating step using the parameters provided in the provision step, and determining a law making it possible to estimate the spatial density in the predetermined spatial area by using the generated learning base. The method also includes a second stage of counting in the predetermined spatial area comprising the steps of acquiring an image including the predetermined spatial area via the detector, and estimating the spatial density of objects in the predetermined spatial area by applying the law to the acquired image.

According to specific embodiments, the counting method includes one or more of the following features, considered alone or according to any technically possible combinations:

- the generating step includes a step for correcting perspective defects related to the detector.
- the method includes a step for segmenting an image to determine the pixels containing an object.
- the determination step is carried out by learning, the learning including extracting characteristics, the extracted characteristics also being used in the estimating step.
- the determination step is carried out by minimizing a support vector regression.
- in the provision step, a parameter relative to the detector is a characteristic intrinsic to the detector.
- in the provision step, a position parameter of the detector and/or an orientation parameter of the detector is provided.

A computer program product is also described including software instructions, the software instructions implementing a method as previously described, when the software instructions are executed by a computer.

Also proposed is an information medium on which a computer program product as previously described is stored.

Also proposed is a surveillance device able to count the objects in a predetermined spatial area comprising a detector able to acquire at least one image, the image including the predetermined spatial area, the detector having parameters, one parameter being relative to the focal length of the detector, one parameter being relative to the altitude of detector and one parameter being relative to the incline of the detector relative to the predetermined spatial area. The device includes a processing module able to count the objects in the predetermined spatial area. The processing module is able to

choose a plurality of different spatial densities in the predetermined spatial area, receive parameters relative to the detector, the received parameters comprising at least a parameter relative to the focal length of the detector, a parameter relative to the altitude of the detector and a parameter relative to the incline of the detector with respect to the predetermined spatial area, generate a plurality of images for each chosen spatial density by using the received parameters, determine a law making it possible to estimate the spatial density in the predetermined spatial area by using the generated learning base as learning base and estimate the spatial density of objects in the predetermined area by applying the law to the acquired image.

Other features and advantages of the invention will appear upon reading the following description of embodiments of the invention, solely as an example and done in reference to the drawings, which are:

- figure 1, a schematic view of an example processing module making it possible to carry out a method for counting objects in a predetermined spatial area,
- figure 2, an example image generated to form a learning base, and
- figure 3, a flowchart of an example embodiment of a method for counting objects.,

A processing module 10 and a computer program product 12 are shown in figure 1.

The interaction of the computer program product 12 with the processing module 10 makes it possible to implement a method for counting objects in a predetermined spatial area.

The processing module 10 is preferably a computer.

More generally, the processing module 10 is an electronic computer able to manipulate and/or transform data represented as electronic or physical quantities in registers of the processing module 10 and/or memories into other similar data corresponding to physical data in the memories, registers or other types of display, transmission or storage devices.

The processing module 10 includes a processor 14 comprising a data processing unit 16, memories 18 and an information medium reader 20. The processing module 10 also comprises a keyboard 22 and a display unit 24.

The computer program product 12 includes an information medium 20.

The information medium 20 is a readable medium. An information medium 20 is a medium readable by the processing module 10, usually by the data processing unit 14. The information medium 20 is a medium suitable for storing electronic instructions and able to be coupled with a bus of a computer system.

As an example, the readable information medium 20 is a floppy disk, an optical disc, a CD-ROM, a magnetic-optical disc, a ROM memory, a RAM memory, an EPROM memory, an EEPROM memory, a magnetic card or an optical card.

A computer program comprising program instructions is stored on the information medium 20.

The computer program can be loaded on the data processing unit 14 and is suitable for driving the implementation of the method for counting objects in the predetermined spatial area when the computer program is implemented on the processing module 10.

The operation of the processing module 10 interacting with the computer program product 12 is now described in reference to figures 2 and 3, which schematically illustrates one example embodiment of the method for counting objects.

"Object" refers to a countable element. In the context of the invention, the object is observable, i.e., a detector 30 as shown in figure 3 is able to observe the object in the predetermined area.

Such a detector 30 is, for example, a camera 30.

Such a camera 30 is characterized by parameters P.

The detector 30 thus has intrinsic parameters $P_{intrinsic}$, i.e., not depending on the environment, such as the focal length of the detector 30 or the size of the image acquired by the detector 30.

The intrinsic parameters $P_{intrinsic}$ are parameters specific to the camera 30.

The detector 30 also has extrinsic parameters $P_{extrinsic}$. For example, a parameter relative to the position $P_{position}$ of the camera 30 depending on the environment in question is an extrinsic parameter $P_{extrinsic}$.

Preferably, a position parameter $P_{position}$ of the camera 30 is the altitude of the camera 30.

According to another example, a position parameter $P_{position}$ of the camera 30 is the distance of the camera 30 with respect to the predetermined spatial area.

Typically, the distance of the camera 30 from the predetermined spatial area is defined as the distance between the plane of the detectors of the camera 30 and the point of the predetermined spatial area that is closest to the camera 30.

According to another example, assuming that an orthonormal plane of reference (X, Y, Z) is defined for the space to be monitored, other position parameters $P_{position}$ of the camera 30 are the position along the axis X and the position along the axis Y.

Likewise, a parameter related to the orientation $P_{orientation}$ of the camera 30 and depending on the mechanical system holding the camera 30 is an extrinsic parameter $P_{extrinsic}$.

Preferably, the orientation parameter $P_{orientation}$ of the camera 30 is the incline angle of the viewing axis of the camera 30 relative to the horizontal. Such an angle is generally called angle of pitch.

Additionally, other orientation parameters $P_{orientation}$ are taken into account.

Another orientation parameter $P_{orientation}$ of the camera 30 is for example the angular orientation of the direction of the optical axis of the camera 30 relative to an absolute frame of reference. The direction of the optical axis of the camera 30 is typically called heading angle.

The incline of the camera 30 around its viewing axis, also called angle of roll, is another example of an orientation parameter $P_{orientation}$ of the camera 30.

It should be noted that the orientation parameters P orientation of the camera 30 can also be expressed in the form of Euler angles, i.e., angles of precession, nutation and specific rotation. The angle of nutation is generally the favored parameter.

Hereinafter, the objects are considered to be persons and the detector 30 is considered to be a video surveillance camera. In such a case, the predetermined spatial area is an area to be monitored.

According to another example, the objects are objects of biological interest observed using a detector 30 and an optic performing a microscope function.

According to still another example, the objects are carbon nanotubes and in this case, the detector is an electron microscope.

More generally, it is possible to apply the considered method to any type of objects whose behavior can be realistically stimulated.

The predetermined spatial area ZS generally depends on the object.

As an illustration, when the object in question is an object of biological interest, the predetermined spatial area ZS is part of the sample observed using the microscope.

In the context of the described example, when the object is a person, the predetermined spatial area ZS is part of the environment, for example, a street portion. This clearly appears in the image of figure 2.

In such a case, the camera 30 is able to acquire an image of the predetermined spatial area ZS; typically, the camera 30 is able to acquire an image of the street. Furthermore, the counting method is a method for counting the crowd in an environment.

As shown schematically in figure 3, the method includes two successive phases P1 and P2. The first phase P1 corresponds to the upper part of figure 3 (delimited by the dotted lines 32), while the second phase P2 corresponds to the lower part of figure 3. The first phase P1 is an unsupervised learning phase, while the second phase P2 is a counting phase in the predetermined spatial area ZS.

The first unsupervised learning phase P1 seeks to obtain a law making it possible to estimate the spatial density of persons in the predetermined area.

The expression "unsupervised learning" means that the learning data are generated automatically.

In such a context, the law is a link between characteristics of an image and a value of the spatial density of persons in the predetermined spatial area ZS. The law thus makes it possible to determine the characteristics involved in the increased spatial density of persons.

During the first learning phase P1, parameters of the detector 30 are used to generate a learning base automatically and, from the learning base, a regression model is learned by the processing module 10.

The second phase P2 for counting in the predetermined spatial area ZS seeks to apply the law to actual cases.

An “actual case" refers to images acquired by the camera 30 and not generated automatically, like in the case of the first, unsupervised learning phase P1.

More specifically, the video stream from the camera 30 is collected, then characteristics are extracted from the video stream to obtain the spatial density over the entire image owing to the regression model learned during the implementation of the first learning phase P1.

The implementation of the method therefore makes it possible to obtain the number of persons in the predetermined spatial area ZS.

The first unsupervised learning phase P1 includes four steps: a selection step S100, a provision step S102, a generating step S104 and a determining step S106.

In the selection step S100, a plurality of different spatial densities of persons in the predetermined spatial area ZS is chosen.

According to one embodiment, each spatial density is evenly distributed between a zero spatial density and a high spatial density.

In general, a spatial density is considered to be high based on the context, and in particular the viewing conditions.

Typically, a spatial density greater than or equal to 50 persons in the predetermined area is considered a high spatial density.

At the end of the selection step S100, a set of spatial densities has therefore been obtained.

According to one embodiment, the selection step S100 is carried out for several separate areas of the image.

In the provision step S102, parameters relative to the camera 30 are provided.

The provided parameters comprise at least a parameter relative to the focal length of the detector 30, a parameter relative to the altitude of the detector 30, a parameter relative to the incline of the detector 30 with respect to the predetermined spatial area ZS and a parameter relative to the position of the main point.

A parameter relative to the position of the main point is for example the position of the optical center relative to the center of the image.

According to one example, a parameter relative to the camera 30 is an intrinsic parameter $P_{intrinsic}$.

As an example, the focal length of the camera 30 or the number of pixels of the camera 30 is provided.

According to another example, the parameter relative to the camera 30 is at least one from among the group made up of a position parameter $P_{position}$ and an orientation parameter $P_{orientation}$ of the camera 30 relative to the environment.

At the end of the provision step S102, parameters relative to the camera 30 are obtained. Among these parameters, there is at least a parameter relative to the focal length of the detector 30, a parameter relative to the altitude of the detector 30, a parameter relative to the incline of the detector 30 with respect to the predetermined spatial area ZS and a parameter relative to the position of the main point.

In the generating step S104, a learning base is generated.

To that end, in the generating step S104, both a plurality of images and an annotation element corresponding to the ground truth are generated. The term "ground truth" is most often referred to using the English term "ground truth".

The plurality of images is generated for each spatial density chosen in the selection step S100.

"Generation" means that the processing module 10 generates synthetic images by using the parameter relative to the camera 30. The synthetic images are not images from the camera 30. These are images generated as if they were taken by the camera 30, but that are synthetic images.

Preferably, the positions of pedestrians including the synthetic images are calculated to correspond to a spatial density field in a three-dimensional environment

comprised in an area comprising the viewing field of the camera 30 and its immediate vicinity to avoid edge effects.

As an example, to configure the spatial density field, a spatial density gradient is created going from a maximal spatial density to a minimal spatial density through the image. The gradient is oriented along the constant depth lines in the image, such that for a given depth, the spatial density intervals are from very dense to very rare. This corresponds to the hypothesis that for a given depth, the appearance of the crowd with a constant spatial density is approximately the same along a constant depth line.

Alternatively, a mixture of Gaussian functions positioned randomly in the predetermined spatial area ZS with a random standard deviation is used. Such a mixture makes it possible to generate a close image of the crowds obtained in reality.

One approach making it possible to implement the calculation of the ground truth is to define a spatial density ground truth function, called function F. To calculate the function F for each pixel of the generated image, the spatial density ground truth function F is defined by the following equation:

$$\forall\, p \in I, F(p) = \sum_{P \in \mathbf{P}} N(p; P, \sigma^2 . 1_{2x2})$$

where:

- **P** is the list of positions in the image where pedestrians are found,

- $N(p, P, \sigma^2)$ is a two-dimensional Gaussian kernel evaluated at the pixel p with mean position $P$, and

- $\sigma$ is a coefficient.

The position $P$ of a pedestrian in the image is calculated by using a parameter of the camera 30 by projecting, in the image, a point situated in the position of the pedestrian in question and at the altitude corresponding to half the height of a human being. Typically, the average height of a human being is considered to be 1.70 meters (m).

Similarly, the coefficient $\sigma$ is calculated as one third of half of the apparent height of a human being at the position $P$ in the image. Consequently, the value assumed by the coefficient $\sigma$ for the pedestrian $P$ depends on its location in the image. Typically, the coefficient $\sigma$ is around several pixels.

The calculation of the ground truth is done automatically and avoids annotation by a human being. Such an operation is therefore more precise, since the operation is done automatically.

At the end of the generating step S104, a database is therefore obtained that may serve as learning database. The database includes a set of synthetic images with which, for each image, a ground truth is associated.

In the determining step S106, a law is determined.

The law makes it possible to estimate the spatial density in the predetermined spatial area.

To that end, the learning base generated in the generating step S104 is used as learning base. As previously explained, it should be noted that the learning base is generated without the implementation of a manual annotation by an operator.

According to the example of figure 3, the determining step S106 is carried out through four successive steps: a step S106a for correcting the perspective, a step S106b for segmenting the crowd, a step S106c for extracting characteristics and a learning step S106d.

In the correction step S106a, account is taken of the fact that in an image, the objects close to the detector 30 appear larger than the objects that are further away. Such a perspective effect must be taken into account when analyzing characteristics to be extracted from the image, since a characteristic extracted from an object of the foreground represents a smaller portion of the object than an extracted characteristic that is extracted further away in the image.

In the correction step S106a, two successive steps are carried out, i.e., a step for normalizing characteristics by using a perspective normalization map and a step for dividing the image into strips of quasi-constant depth.

The normalization of the characteristics is implemented by using a perspective normalization map in which a pixel weight is related to the provided depth for the object that generated the pixel. Weighting of the pixels is applied for distant objects. The weighting increases the considered weight if the object is further away than an object in the foreground.

The perspective normalization map is generated by using the camera parameters 30.

More specifically, the correction factor is obtained by differential rectification. Thus, instead of using two-dimensional parameters such as the height or width, it is preferable to account for the three-dimensional nature of a human being. To that end, the appearance variation of the three-dimensional volume occupied by the pedestrian. A perspective correction factor W(p) for the pixel p is deduced therefrom.

As an example, the perspective correction factor W(p) is obtained using the following formula:

$$W(p) = \frac{\max_{\forall p}(A_1(p) + A_2(p))}{A_1(p) + A_2(p)}$$

Where:

- the human being is modeled by a cylinder with a circular base,
- $A_1(p)$ is the apparent area of the lateral area, and
- $A_2(p)$ is the apparent area of the upper circular base.

The map thus obtained is used to normalize the characteristics by applying calculated weights to each of the pixels that are in the predetermined spatial area ZS if the characteristic is a surface type characteristic and by applying the square root of the weights to the characteristics related to the edge effects if the characteristic is a contour type characteristic.

For segmentation into strips of constant depth in the image, the viewing angle from which the pedestrian is viewed is embodied by a cylinder with height h and radius r for a camera 30 that is positioned at a height H.

It appears that this angle varies depending on the position of the pedestrian in the image. The strip considered in the image is considered to be extreme values that are not too far from one another. To that end, a maximum angular variation is determined in the strip corresponding to a pixel.

As an illustration, two strips B1 and B2 are shown in figure 2, the two strips B1 and B2 being relative to the predetermined spatial area ZS.

At the end of the correction step S106a, the defects related to the particular shot of the camera 30 have been taken into account in analyzing the synthetic images.

In the step S106b for segmenting the crowd, the pixels belonging to the crowd are determined in the entire image. To that end, in the learning base, ground truth data are used.

In the extraction step S106c, the characteristics are extracted.

The characteristics are connected to a particular pixel p; the pixel p is a pixel segmented in the segmenting step S106b. This ensures that the pixel p is related to the presence of a crowd.

The characteristic attached to a pixel p is calculated over a sliding window $W^p_{rw}$ with radius $r_w$ centered on the pixel p.

The extracted characteristics are of two types, contour characteristics and texture characteristics.

As an example, the contour characteristics are obtained by using a Canny edge detector.

The contour characteristics are for example the orientation of the contours, or their length.

Alternatively, the step S106c for extracting contour characteristics is carried out by using a Minkowski fractal dimension. The step is also denoted using the acronym MFD.

The step S106c for extracting texture characteristics is for example carried out by gray level dependency matrices. Such matrices are typically referred to using the acronym GLDM.

The extraction step S106c also includes a step for the presentation of the extracted characteristics in the form of a vector having a plurality of dimensions by differentiation of the radius values $r_w$.

More specifically, a reference value is considered for the radius $r_w$ corresponding to one third of half of the average apparent height in the image of a human being as previously described for the coefficient $\sigma$.

Such reference values, denoted $r_{ref}$, are used to obtain six different radius values $r_w$, which are, according to the example of figure 3: 1/4. $r_{ref}$, 1/2. $r_{ref}$, $r_{ref}$, 3/2. $r_{ref}$, 2. $r_{ref}$ et 3. $r_{ref}$.

As a result, according to one embodiment by the applicant, for each pixel of the image, a vector of characteristics with dimensions 1*144 is obtained.

In the learning step S106d, according to the illustrated example, a support vector regression is used. Such a regression is also denoted SVR. Such a regression makes it possible to link the characteristics vector with the values assumed by the spatial density in the density function F.

The performance of the SVR technique in depends greatly on a proper adjustment of parameters that are chosen both in the kernel and the technique itself.

In the illustrated example, the kernel is a RBF Gaussian kernel, RBF being the acronym for "Radial Basis Function". Hereinafter, the kernel is denoted kernel k.

To characterize such a kernel, three parameters must be characterized: a parameter of the kernel γ, a penalty parameter of the error term C and a loss function parameter ε.

The kernel parameter γ is chosen such that the kernel k verifies the following equation: $k(\mathbf{x_p}, \mathbf{x_q}) = exp\left(-\gamma . \|\mathbf{x_p} - \mathbf{x_q}\|^2\right)$

Where $\mathbf{x_p}$ and $\mathbf{x_q}$ are the $p^{th}$ and $q^{th}$ characteristics vectors.

Optimizing the choice of the three parameters makes it possible to obtain good performance levels. A typical method for obtaining good parameters is dividing a cross-validation with the learning set.

Since a cross-validation is time-consuming, the following constraints were used for the penalty parameter of C and the parameter of the loss function ε.

$$C = max(|\bar{y} - 3.\sigma_y|, |\bar{y} + 3.\sigma_y|)$$

$$\varepsilon = \tau.\sigma_b\sqrt{\frac{\log n}{n}}$$

Where:

- $\bar{y}$ is the mean value of the ground truth values,
- $\sigma_y$ is the standard deviation of the ground truth values,
- $\sigma_b$ is the standard deviation of the noise of the ground truth values,
- $n$ is the number of learning data, and
- $\tau$ is a constant value.

The applicant chose to set the constant value $\tau$ at three and for the kernel parameter γ to set the parameter at the value of the median from the list of all of the values of the squares of the norm between $\mathbf{x_p}$ and $\mathbf{x_q}$ for all possible p and q.

At the end of the learning step S106d, the law is obtained making it possible to estimate the spatial density in the predetermined spatial area.

In the specific case that is described here, the law is a regression model, since the learning technique used is a SVR technique.

The second phase P2 for counting in the area includes three steps: an acquisition step S202, an obtaining step S204 and an estimating step S206.

In the acquisition step S202, an image is acquired by the detector 30.

The image includes at least the predetermined spatial area ZS.

In the obtaining step S204, the characteristics determined in the extraction step S106c are obtained for the image provided in the acquisition step S202.

The obtaining step S204 includes a segmenting step S204a and an extraction step S204b.

The segmenting step S204a is carried out owing to the use of an algorithm for detecting the crowd in the image.

The segmenting operation makes it possible to extract the characteristics relative to the crowd and not the characteristics relative to locations where there is no crowd. Improving the segmentation enhances the overall performance of the counting method.

The extraction step S204b is similar to the extraction step S106c described for the first phase P1.

Since the extraction is only done for the pixels for which a crowd has been detected, at the end of the obtaining step S204, the characteristics vectors are obtained in each pixel of the image for which a crowd has been detected.

In the estimating step S206, the spatial density of objects in the predetermined area is estimated. To that end, the law obtained at the end of the first phase P1 is applied to the acquired image. This link is embodied by arrow 50 in figure 3.

More specifically, the regression model obtained at the end of the first phase P1 is applied to obtain a spatial density value. In other words, the model is applied to each characteristics vector to deduce an estimate therefrom of the value assumed by the function F at the corresponding pixel p.

Furthermore, the estimate only being done in the segmented areas and the function F being the result of a Gaussian convolution, it is possible for the function F in some cases to assume non-negligible values also outside the segmented area.

To estimate the values assumed by the function F outside the segmented area, the obtained spatial density values are extrapolated by using a convolution of images by using a normalized kernel κ such that:

$$\forall\ (k_1, k_2), \kappa(k_1, k_2) = \exp(-\frac{(k_1 - u_0)^2 + (k_2 - v_0)^2}{\sigma^2})$$

with

- the constant $u_0$ defined by the equation$u_0 = u_c - \beta.(k_1 - u_c)$;
- the constant $v_0$ is defined by the equation $v_0 = v_c - \beta.(k_2 - v_c)$, and
- the constant $\beta$ is defined by the equation $\beta = \alpha.\frac{\sigma}{\sqrt{(k_1-u_c)^2+(k_2-v_c)^2}}$
- $u_C$, $v_C$ are the coordinates of the center of the kernel, these coordinates being adjustable by the processing module 10.

The parameter α is determined by using the images obtained by the learning parameter.

The parameter α is for example found by a dichotomy technique.

At the end of the estimating step S206, an estimate of the spatial density of the number of pedestrians in the predetermined spatial area ZS is obtained.

Optionally, the second phase P2 includes a deducing step during which the number of pedestrians is deduced from the spatial density estimated in the estimating step.

For example, the number of pedestrians in the predetermined area is obtained by integration over the entire predetermined area. The method makes it possible both to obtain analysis elements and to behave like a self-learning system.

Relative to the method of the state of the art, the proposed method uses a learning method in fact involving a database that is easier to obtain.

Indeed, in the methods of the state of the art, the method by learning is done using actual data, which involves a manual annotation by an operator of the data to obtain a learning base.

Furthermore, the learning base of the proposed method is of better quality than the learning base of the method of the state of the art. Indeed, the variability and the representativeness of the learning data are debatable for the methods of the state of the art. In particular, in practice, in the bases of the state of the art, the extreme situations, i.e., a spatial density greater than 50 persons in the area, are rarely present in the actual data, since this situation occurs too rarely in practice. Furthermore, the annotation by the operator is in reality unreliable, in particular when the crowd is very dense or the resolution is limited, these problems being able to add to one another. On the contrary, the base is generated to be representative of future usage cases of the detector 30, i.e., the various generated images make it possible to scan very different spatial densities ranging from not very dense to very dense.

Furthermore, the method has the advantage of being robust with respect to a change of position and/or orientation of the detector. Indeed, the viewing parameters are respected. The learning data are created virtually with the same camera 30 as that which will subsequently be used. To that end, data relative to the camera 30 are provided, such as intrinsic focal length or image size parameters $P_{intrinsic}$ or extrinsic position $P_{position}$ and orientation $P_{orientation}$ parameters. Owing to the parameters given by the camera 30, the learning database is built automatically, reliably and statistically representatively.

The robustness is also ensured by the fact that the change made on the camera 30 only assumes re-implementing the first phase P1 without acquiring ground images, which would be indispensable in a method of the state of the art.

The method is also advantageous in that the ground truth is immediately available, since the annotation is automatic and any change to the camera 30 can be taken into account immediately via learning of a regression model not assuming the provision of a new learning set.

The method in particular applies to the supervision of crowds through video surveillance for safety or transport purposes. In such an application, the method makes it possible to estimate the spatial density of a crowd and to count the number of pedestrians in the crowd.

Frequently, in such a situation, the acquisition step is done continuously such that the second phase P2 may be considered to be carried out iteratively on a continuous video stream.

The proposed method relates to computer-assisted viewing, the computer implementing the learning. The method can therefore be implemented using any computer or any other device. Multiple systems can be used with programs carrying out the preceding method, but it is also possible to consider using devices dedicated to carrying out the preceding method, which may be inserted into devices able to measure the supplied data. In particular, the method may be implemented by a surveillance device able to count the objects in a predetermined spatial area ZS comprising a detector 30 able to acquire at least one image, the image including the predetermined spatial area ZS and a processing module 10. The processing module 10 is suitable for counting objects in the predetermined spatial area ZS and is able to carry out the selection, provision S102, generating S104, determining S106, obtaining S204 and estimating S206 steps. The processing module 10 is also suitable for receiving parameters relative to the detector 30, the received parameters comprising at least a parameter relative to the focal length of the detector 30, a parameter relative to the altitude of the detector 30, a parameter relative to the incline of the detector 30 with respect to the predetermined spatial area ZS.

Furthermore, the proposed embodiments are not tied to a particular programming language. Incidentally, this means that multiple programming languages can be used to carry out the method previously outlined.

The method and embodiments described above may be combined with one another, in whole or in part, to create other embodiments of the invention.

CLAIMS

1.- A method for counting objects in a predetermined spatial area (ZS), the method being carried out by a processing module (10), the method including:

- a first unsupervised learning phase, comprising the following steps:
  - choosing a plurality of different spatial densities of objects in the predetermined spatial area (ZS),
  - providing parameters relative to a detector (30) able to acquire at least one image, the image including the predetermined spatial area (ZS), the provided parameters comprising at least a parameter relative to the focal length of the detector (30), a parameter relative to the altitude of the detector (30) and a parameter relative to the incline of the detector (30) with respect to the predetermined spatial area (ZS),
  - generating a database including a plurality of images for each chosen spatial density, the images being synthetic images, the generating step using the parameters provided in the providing step, and
  - determining a law making it possible to estimate the spatial density in the predetermined spatial area by using the generated learning base,
- a second phase for counting in the predetermined spatial area (ZS) comprising the following steps:
  - acquiring an image including the predetermined spatial area (ZS) via the detector (30), and
  - estimating the spatial density of objects in the predetermined spatial area (ZS) by applying the law to the acquired image.

2.- The method according to claim 1, wherein the generating step includes a step for correcting perspective defects related to the detector (30).

3.- The method according to any one of claims 1 to 2, wherein the determination step is carried out by learning, the learning including extracting characteristics, the extracted characteristics also being used in the estimating step.

4.- The method according to any one of claims 1 to 3, wherein the determination step is carried out by minimizing a support vector regression.

5.- A computer program product is also described including software instructions, the software instructions implementing a method according to any one of claims 1 to 4, when the software instructions are executed by a computer.

6.- An information medium on which a computer program product according to claim 5 is stored.

7.- A surveillance device able to count the objects in a predetermined spatial area (ZS) comprising:

- a detector (30) able to acquire at least one image, the image including the predetermined spatial area (ZS), the detector (30) having parameters, one parameter being relative to the focal length of the detector (30), one parameter being relative to the altitude of the detector (30) and one parameter being relative to the incline of the detector (30) with respect to the predetermined spatial area (ZS), and
- a processing module (10) able to count the objects in the predetermined spatial area (ZS),

the processing module (10) being able to:

- choose a plurality of different spatial densities in the predetermined spatial area (ZS),
- receive parameters relative to the detector (30), the received parameters comprising at least the parameter relative to the focal length of the detector (30), the parameter relative to the altitude of the detector (30), the parameter relative to the incline of the detector (30) with respect to the predetermined spatial area (ZS),
- generate a plurality of images for each chosen spatial density, the images being synthetic images, by using the received parameters, and
- determine a law making it possible to estimate the spatial density in the predetermined spatial area (ZS) by using the generated learning base as learning base, and
- estimate the spatial density of objects in the predetermined spatial area (ZS) by applying the law to the acquired image.