

DESCRIPTION

Device for classifying a scene of the environment of a vehicle**Technical field**

[0001] The invention relates in general to image processing systems, and in particular to a device and a method for classifying a scene of the environment of an automated vehicle.

[0002] Driving system automation technologies have experienced significant growth in recent years in the field of automated vehicles, such as autonomous vehicles, be these land vehicles, marine vehicles or underwater vehicles. Autonomous navigation is made possible by real-time recognition of the environment of the vehicle, which has the advantage of requiring only very little hardware. Indeed, generally speaking, all that is needed is an image acquisition device associated with an on-board or remote computing unit. In some existing solutions, the computing unit is configured to carry out semantic classification of scenes of the environment. For this purpose, the computing unit uses a semantic segmentation method that makes it possible to obtain information about the class of each of the pixels of an image. This is a complex computer vision task that conventionally relies on deep neural networks that have been trained to carry out semantic classification of scenes of the environment of the vehicle.

[0003] Thus, when the vehicle is moving in its environment, it is able to analyze the environment finely in order to detect obstacles, avoid them and optimize its movement in accordance with the environment in which it is operating.

[0004] In current systems, neural networks are taught from a database grouping together thousands of semantically annotated images. The neural networks thus learn to automatically recognize the class of pixels of an image based on these annotations.

[0005] However, a major hindrance to the deployment of such systems is the high human cost associated with the implementation of these approaches, which are referred to as semantic segmentation approaches. Indeed, such systems require manual annotation, at the level of each pixel of an image, from among the thousands of images in the database in order to enable robust classification. However, these semantic annotations are necessary to be able to implement the learning phase for the artificial intelligence. The deployment of autonomous vehicles on scenes that are still unexplored thus requires a substantial amount of manual re-annotation work, resulting in significant delays in the exploitation of these new environments. Moreover, precise annotation of these images is sometimes subjective on account of the environment being studied where the boundary between two classes is not clearly defined (for example a boundary between sand and the water column in an underwater environment).

[0006] There is thus a need for an improved device and method for classifying scenes of the environment of an automated vehicle.

General definition of the invention

[0007] The invention aims to improve the situation by proposing a device for classifying a scene of the environment of a vehicle, comprising a classification unit that uses a neural network configured to classify a scene in response to a received image depicting the scene and a learning unit configured to train the neural network on a set of images stored in a database. Advantageously, the classification device comprises an image annotation unit comprising:

[0008] - an image preprocessing module configured to extract a set of thumbnails from a received image as a function of a criterion concerning position in the image and at least one criterion relating to an image parameter,

[0009] - a classification module configured to determine, for each thumbnail, a set of probabilities of belonging, each associated with a class from among a set of predefined classes, and

[0010] - an image annotation module configured to determine an annotation corresponding to the image based on the sets of probabilities determined for all of the thumbnails, the annotation being transmitted to the database for storage in association with the image in the database.

[0011] According to one aspect of the invention, the image annotation module may be configured to merge the probabilities of belonging determined for all of the classes and for all of the thumbnails, and to reconstruct an annotation, pixel by pixel, for the received image, thereby providing the annotation.

[0012] In one embodiment, each set of probabilities determined for a thumbnail may be represented by a probability vector having a set of components corresponding to the probabilities of belonging, each associated with one of the classes.

[0013] In one embodiment, an image parameter may be image frequency information, the criterion relating to the image parameter being a frequency level.

[0014] The image criterion used for the extraction may relate to ranges of values of the image parameter.

[0015] In one embodiment, the preprocessing unit uses a fixed or variable step to implement sliding window-based thumbnail extraction.

[0016] In one embodiment, the image annotation module may be configured to select, for each image pixel, the class having a maximum probability of belonging from among the probabilities of belonging obtained for all of the thumbnails.

5 [0017] In one embodiment, the preprocessing module may be configured, prior to determining the set of probabilities for a thumbnail, to implement an operation of resizing the thumbnail as a function of the level of its image parameter.

[0018] In one embodiment, the preprocessing module may be configured, for each thumbnail of size $K \times K$ from among the P thumbnails extracted from the input image of size $M \times N$, to add the C probabilities of belonging obtained for the thumbnail in a patch of size $K \times K$,
10 corresponding to the thumbnail, to a matrix of size $M \times N \times C$ corresponding to the input image.

In one embodiment, the probability matrix Mp may be defined according to the equation:

$$Mp(i, j, n) = Mp(i, j, n) + P(n) * \frac{\min(D)}{D(n)},$$

where P denotes the probability vector P at the output of the classification unit and D denotes the vector of the various extraction dimensions of the thumbnails.

15 In one embodiment, the probability matrix Mp may be defined according to the equation:

$$Mp(i, j, n) = Mp(i, j, n) + G(i, j) * P(n) * \frac{\min(D)}{D(n)}$$

P denotes the probability vector P at the output of the classification unit, $G(i, j)$ denotes a Gaussian window and D denotes the vector of the various extraction dimensions of the thumbnails.

20 According to one aspect of the invention, the Gaussian window $G(i, j)$ may be defined by:

$$[0019] \quad G(i, j) = e^{\left(\frac{1}{2}\right)\left(\frac{i}{\sigma}\right)^2} e^{\left(\frac{1}{2}\right)\left(\frac{j}{\sigma}\right)^2} \text{ for } i \in [0, Z], j \in [0, Z]$$

[0020] The annotation of the received image may then be determined by selecting, at each point of the probability matrix Mp , the index having the maximum value in the probability vector associated with each pixel.

25 [0021] Also proposed is a method for classifying a scene of the environment of a vehicle, comprising a classification step that uses a neural network configured to classify a scene in response to a received image depicting the scene and a learning step configured to train the neural network on a set of images stored in a database. Advantageously, the method comprises:

[0022] - preprocessing the image, consisting in extracting a set of thumbnails from a received image as a function of a criterion concerning position in the image and at least one criterion relating to an image parameter,

[0023] - a classification step consisting in determining, for each thumbnail, a set of probabilities of belonging, each associated with a class from among a set of predefined classes, and

[0024] - an image annotation step consisting in determining an annotation corresponding to the image based on the sets of probabilities determined for all of the thumbnails, the annotation being transmitted to the database for storage in association with the image in the database.

[0025] The embodiments of the invention thus make it possible to annotate the pixels of the images in the learning database with reduced complexity and processing time.

Brief description of the figures

[0026] Other features, details and advantages of the invention will become apparent on reading the description given with reference to the appended drawings, which are given by way of example and in which, respectively:

[0027] [Fig.1] Figure 1 schematically shows a vehicle implementing a control system, according to the embodiments of the invention.

[0028] [Fig.2] Figure 2 schematically shows the structure of the annotation unit according to some embodiments of the invention.

[0029] [Fig.3] Figure 3 depicts the image annotation method implemented by the image annotation unit according to some embodiments of the invention.

[0030] [Fig.4] Figure 4 shows one example of a probability matrix and patches, according to one exemplary embodiment.

[0031] [Fig.5] Figure 5 illustrates a change in the size of the sliding window as a function of the class.

[0032] [Fig.6] Figure 6 illustrates the method for determining image annotations, in one embodiment in which a single-scale approach is used.

[0033] [Fig.7] Figure 7 illustrates the method for determining image annotations, in one embodiment in which a multi-scale approach is used.

[0034] [Fig.8] Figure 8 illustrates the method for determining image annotations, applied to a sequence of images, using the pre-trained classifier.

Detailed description of the application

[0035] Figure 1 schematically shows a vehicle 100 implementing a control system 10 for controlling the vehicle, according to the embodiments of the invention.

5 [0036] The control system 10 is configured to recognize, in real time, the environment of the vehicle 100 and to adapt the driving of the vehicle 100 dynamically as a function of the recognition carried out in order to allow automated navigation, such as for example autonomous navigation.

[0037] The control system 10 comprises a device 1 for classifying scenes of the environment of the vehicle (also referred to more simply as a "classification system") and a perception system 2.

10 [0038] The classification device 1 is configured to carry out semantic classification of the scenes of the environment of the vehicle 100 based on the images detected by the perception system 2 of the vehicle 100.

15 [0039] The vehicle 100 may be an automated or semi-automated vehicle capable of moving in an environment, such as a land vehicle, marine vehicle or underwater vehicle. The environment may therefore be a land environment, marine environment or underwater environment.

[0040] The environment of the vehicle 100 is associated with a set of predefined classes, depending on the application of the invention, such as for example the following set of classes: [road, sidewalk, vegetation, vehicle]. In some embodiments, the classes associated with the environment may be modified (addition, deletion of classes for example) by re-teaching the system.

20 [0041] The control system 10 may be a driving assistance system configured to optimize the driving of the vehicle and safeguard the vehicle (obstacle detection for example) based on the information determined by the perception system 2.

25 [0042] The perception system 2 comprises a set of detection devices arranged on the vehicle 100 and configured to provide signals representative of an image of the environment in which the vehicle 100 is moving. The detection devices 20 may comprise one or more cameras. A camera may be equipped with a photosensitive sensor sensitive to radiation in a light spectrum contained in the infrared and/or in the visible. The cameras may furthermore comprise one or more acoustic and/or laser cameras (in applications of the invention in an underwater environment, for example).

30 [0043] The detection devices 20 may be arranged on the vehicle so as to acquire images of the environment of the vehicle located in front of the vehicle 100, to the sides, above and/or below the vehicle 100, depending on the application of the invention. For example, in a land-

based environment, the detection devices may be fixed to the front of the vehicle and/or in the passenger compartment of the vehicle and/or on the roof of the vehicle. The detection devices may furthermore comprise a lidar (laser detection and ranging) and/or a radar.

[0044] Other types of sensors may be used by the perception system 2, such as an ultrasonic sensor, a steering wheel angle sensor, a wheel speed sensor, a brake pressure sensor, a transverse yaw rate and acceleration sensor, or a combination thereof.

[0045] The perception system 2 is configured to determine images of the environment based on the information detected by the set of detection devices. In some embodiments, the perception system 2 may furthermore comprise a processing device 22 configured to implement multi-sensor fusion algorithms capable of combining information from the various detection devices 20 in order to determine the images of the environment.

[0046] The classification device 1 comprises a learning unit 11 for training, beforehand, the neural network used on sets of real data stored in an image database 3. The database contains an image/annotation pair for each datum.

[0047] The classification device 1 furthermore comprises an image annotation unit 12 configured to automatically annotate one or more images in the image database 3 used for the learning phase. The image annotation unit 12 is thus configured to generate the annotations stored in the image database 3. An image annotation is represented by a matrix of size identical to the image, in which each element has a value between 0 and N-1, with N denoting the number of classes (a pixel belonging to a class).

[0048] Such annotations are used by the learning unit 11 to train the one or more neural networks used.

[0049] The classification device 1 comprises a classification unit 14 configured to semantically segment each image detected by the perception system 2 in real time, using the pre-trained neural network.

[0050] For ease of understanding of the embodiments of the invention, definitions or concepts relating to neural networks are detailed below in relation to the embodiments of the invention.

[0051] A neural network comprises neurons that are interconnected with one another by synapses, which may be implemented in the form of digital memories. A neural network may comprise a set of successive layers, comprising an input layer carrying the input signal and an output layer carrying the result of the prediction made by the neural network (result of the network), and one or more intermediate layers. Each layer of a neural network takes its inputs from the outputs of the previous layer. The number of neurons in each layer is equal to the

number of inputs of the neurons of the following layer. A given layer of the neural network thus comprises a set of neurons that take their inputs from the neurons of the previous layer.

[0052] The signals propagated at the input and output of the layers of the network may be digital values (information coded in the value of the signals), or electrical pulses in the case of pulse coding (information that is coded temporally according to the order of arrival of the pulses or according to the frequency of the pulses).

[0053] A neural network comprises a set of input data (or 'input coefficients') and output data (or 'output coefficients').

[0054] The output values are calculated based on the inputs and synaptic weights by applying an activation function to the input coefficients.

[0055] Each neuron of the neural network is configured to calculate a weighted sum of its input coefficients using a combination function and the synaptic weights, and then by applying an activation function to the resulting weighted sum in order to produce its output:

[0056] The synaptic weights of a neural network are determined by learning in a learning phase implemented by the learning unit. Random values are initially assigned to the weights of the neural network, and then a set of real data taken from the database 3 are used to carry out the learning. Learning of a neural network consists in determining the optimum values of the synaptic weights, for each neuron of the neural network, from the last layer of the network to the first, using a learning function.

[0057] The learning phase may implement a plurality of iterations of the learning function, each iteration comprising a forward propagation step and a backpropagation step for correcting errors between the outputs obtained in the forward propagation phase and the outputs expected for the input sample under consideration.

[0058] The learning phase thus makes it possible to compare the obtained output with the expected output (in the case of a supervised method) and, depending on this comparison, to update the connections between the neurons represented by the synaptic weights to improve the final result.

[0059] In the forward propagation phase, input datasets from the database 3 are used to implement the learning. Each dataset forms a sample x associated with desired values (or expected values). The signal corresponding to the input sample is propagated forwards in the layers of the neural network from the first layer, from one layer $(k-1)$ to the next layer (k) , to the last layer. In the forward propagation phase, the activation function ϕ and the synaptic weights connecting the neurons of a previous layer $(k-1)$ and a following layer (k) are used.

[0060] When the forward propagation is complete, a result y is obtained at the output.

[0061] In the backpropagation phase, any errors obtained by a neuron are backpropagated to its synapses and to the neurons connected thereto. The backpropagation may be for example gradient backpropagation in order to modify the synaptic weights by taking into account their impact on the errors generated. Synaptic weights that contribute to generating a significant error may thus be modified to a greater extent than weights that generated a less significant error.

[0062] The duration of the learning phase may depend on the size of the database 3 storing the samples used for learning and the size of the network.

[0063] After the learning phase, what is referred to as a generalization phase, which is faster, is implemented by the classification unit 14. In the generalization phase, the weights learned in the learning phase are used (static neural network in which the weights are fixed). Input data corresponding to the images of the scene that are detected in real time by the perception system 2 are submitted to the neural network, which provides a response from the neural network representing the probability of each pixel belonging to the predefined classes.

[0064] The control system 2 may implement a set of driving applications using the classification determined by the classification device 1, such as a trajectory change application to help the driver to change trajectory, or a collision warning application.

[0065] Figure 2 schematically shows the structure of the annotation unit 12 according to some embodiments of the invention.

[0066] The image annotation unit 12 is configured to determine an annotation to be assigned to an image representative of a scene of the environment to be analyzed, in response to the receipt of this image determined by the perception system 2. The classification device 1 may then store the image in association with the determined annotation in the database 3.

[0067] The scene classification device 1 may thus carry out semantic annotation of the images, making it possible to produce a semantic segmentation of the environment of the vehicle 100 in a minimal amount of time, even in the case of scenes that are still unexplored, in which no annotation is available.

[0068] In one embodiment, as illustrated in figure 2, the image annotation unit 12 may comprise:

[0069] – an image preprocessing module 120 configured to preprocess the images and extract sub-images (or thumbnails) from the received image (input image);

[0070] – a classification module 122 (also called a "classifier") configured to classify the thumbnails extracted by the preprocessing unit 102, using the neural network, in order to

determine the probabilities of each thumbnail (thumbnail image) belonging to each of the classes of the environment of the vehicle;

[0071] – an annotation module 124 (also called a "reconstruction unit") configured to merge the responses from the classification module 122 obtained for each thumbnail and to reconstruct a semantic annotation, pixel by pixel, for the complete image under consideration, thereby providing the annotation to be assigned to the received image.

[0072] The image preprocessing module 120 may be configured to collect the local information concerning the image under consideration, according to multiple levels, and to extract a set of thumbnails as a function of a criterion concerning image position and at least one image criterion relating to an image parameter. A thumbnail may thus correspond to information concerning position in the image and to image criteria defined based on the scale levels of the image parameter under consideration (that is to say ranges of values of the image parameter).

[0073] The image criterion used to extract the thumbnails may relate to various types of image parameters, such as for example image frequency, image texture and image contours.

[0074] The scale levels of the image parameter under consideration for the extraction may be for example a scale pyramid of a given factor, such as for example a factor of 2. In one exemplary application of the invention in which the image criterion used for the extraction relates to the image parameter corresponding to the maximum image resolution, for a maximum resolution of 224x224 pixels, and a minimum resolution of 14x14 pixels, the scale pyramid of factor 2 that is used comprises the following levels: 224x224; 112x112; 56x56; 28x28; and 14x14.

[0075] The maximum resolution may depend on the resolution of the perception system (2). The thumbnails are extracted for each scale level.

[0076] The preprocessing unit 120 may take into account low-frequency and high-frequency information inherent to each of the classes of the scene of the environment under consideration. The preprocessing unit 120 may use a fixed step to implement multi-scale sliding window-based extraction. The step may be predefined. The step may be applied to the pixel grid of the image (vertically and horizontally). The step and the scale levels make it possible to parameterize the finesse of the semantic annotations that will be generated automatically.

[0077] The classifier 122 is configured to determine, for each extracted thumbnail, a set of probabilities of belonging, each associated with a class from among a set of predefined classes.

[0078] In one embodiment, the classifier 122 is configured to assign, to each thumbnail extracted by the image preprocessing module 120, a feature vector (also called a "probability vector") comprising components, each component being associated with one of the classes of

the scene of the environment under consideration and having a value corresponding to the probability of the thumbnail belonging to the associated class. The probability vector assigned to a thumbnail thus represents the probability of the thumbnail belonging to each of the classes of the scene of the environment under consideration. The classifier 122 uses the neural network, which may notably be a deep neural network, pre-trained beforehand on an appropriate dataset, during a learning phase, to enable robust classification.

[0079] The annotation module 124 is configured to merge the responses from the classifier 122 (represented by the probability vectors obtained for each thumbnail) so as to aggregate the probabilities of belonging obtained for each thumbnail, corresponding to different image positions and image criteria. The annotation module 124 is in particular configured to assign, to each pixel of the initial image, a probability vector indicating the degree to which it belongs to the classes present in the scene under consideration. The annotation module 124 thus associates a semantic annotation with the image under consideration by selecting, for each image pixel, the class having a maximum probability of belonging. As used here, the term "annotation" is associated with an image and refers to a set of class information, each item of class information being associated with a pixel of the image and corresponding to the determined maximum probability class.

[0080] The image annotation unit 12 may then store the determined annotations in association with the corresponding image in the image database 3.

[0081] The learning unit 11 may then implement learning of the semantic segmentation neural network not using annotations of the same type.

[0082] To carry out the learning, the learning unit 11 may use the annotations thus generated.

[0083] It should be noted that the annotations determined by the reconstruction unit 124 might not be perfect, and may form a set of what are referred to as degraded annotations. However, since the neural network of the modules 11 and 14 is designed as an encoder/decoder, this type of network will naturally smooth the annotations provided as inputs and thus limit the impact of the imperfection of such degraded annotations.

[0084] The classifier 122 may be trained to recognize the classes on thumbnails of a given size, such as for example images of sizes 224×224. The classifier 122 may be configured to predict the probabilities of belonging to the predefined classes on each thumbnail extracted from the image, at multiple scale levels. In one particular embodiment, the classifier 122 may notably be configured to predict the majority class (that is to say the class having a maximum probability of belonging) on each thumbnail extracted from the image, at multiple scale levels.

[0085] Such predictions may be obtained by applying an underlying multi-scale sliding window principle. Sliding window is a technique that makes it possible to analyze an image by dividing

it into small regions called windows and by gradually moving this window across the whole image in order to carry out a local analysis. The sliding window process comprises three steps:

[0086] – a step of initializing the window in terms of position and size. The window may be square and its initial position may be located at the origin of the image. The size of the window depends on the scale set in the multi-scale approach.

[0087] – a step of analyzing the pixels of the window by classifying the window in order to extract therefrom its probabilities of belonging to the classes of the system.

[0088] – a step of moving the window, after having analyzed the current region. In this step, the window is moved by a certain number of pixels, called a "step". This process step is repeated until the window has covered the entire image. The step may be fixed and equal to the minimum scale resolution (for example, for an image size of 14x14, the step is equal to 14).

[0089] A multi-scale approach makes it possible to have as many sliding windows as there are desired sizes (scale pyramid).

[0090] Merging the class predictions available for each pixel, carried out by the reconstruction unit 124, makes it possible to obtain a semantic annotation of the image under consideration that may be degraded. Each annotation thus obtained enriches the semantic annotation database 3 and may be used to implement learning by the learning unit 11, thereby making it possible to use a strong semantic segmentation neural network.

[0091] Figure 3 depicts the image annotation method implemented by the image annotation unit according to some embodiments of the invention.

[0092] In step 300, an input image of size $M \times N$ is received.

[0093] In step 302, the input image is preprocessed. This step comprises extracting sub-images (or thumbnails) from the input image. The image may notably be a color image. Each thumbnail may be associated with an image resolution.

[0094] In step 302, the features of the input image received in step 300 may be stored in a thumbnail memory (for example a thumbnail catalog), in which the various thumbnails extracted from the image are associated with fixed image positions and resolutions, which may be different. The thumbnails may advantageously be defined according to multiple scale levels.

Taking into account the multi-scale information of the thumbnails makes it possible to describe the image not only in terms of the textures present therein and characterizing each class, but also in terms of the structures and contours visible in the image. In one embodiment, the method may for example implement a multi-frequency approach that makes it possible to extract low-frequency and high-frequency information simultaneously.

[0095] In step 304, the probabilities of each extracted thumbnail belonging to each of the classes of the environment of the vehicle are determined by implementing a classification (that is to say a class prediction) of the thumbnail based on the neural network, providing a probability vector for each class (this step may be implemented by the classifier 122).

5 [0096] Step 304 (classification step) consists in determining, using the neural network pre-trained on annotated thumbnails having a fixed size (for example 224x224), the probabilities of belonging to the various classes locally present in each thumbnail extracted in step 302, and the probability vector grouping together these probabilities of belonging for the various classes. The neural network may be a deep neural network, such as for example an AlexNet
10 or SqueezeNet network. The neural network is used to determine the probabilities of each extracted thumbnail belonging to each of the C classes, C being the number of classes present in the scene under consideration of the environment. These classes may be predefined during training (that is to say learning) of the image classification network. In some embodiments, step 304 may comprise, before determining the probability vector of a thumbnail, an operation
15 of resizing a thumbnail as a function of its scale level, in order to make the input of the classifier 122 compatible with the size of the thumbnail.

[0097] Step 304 thus provides, for each thumbnail, C probabilities of belonging that locally describe the content of the original image at a given position and scale of the image corresponding to a thumbnail (the C probabilities of belonging being able to constitute the C
20 components of the probability vector, which is therefore then of size C).

[0098] In step 306, a semantic annotation is determined for each pixel of the image using the probability vectors obtained for the various thumbnails derived from the input image. Step 306 may be implemented by the image annotation reconstruction unit 124.

[0099] In step 306, a semantic annotation is determined for each image based on the
25 probabilities of belonging to the classes determined in step 304, for all of the thumbnails derived from the input image. Thus, considering P thumbnails extracted in step 302, step 306 uses the P probability vectors determined for the P thumbnails.

[0100] Step 306 comprises, for each thumbnail of size $K \times K$ from among the P thumbnails extracted from the input image of size $M \times N$, adding the C probabilities of belonging obtained
30 for the thumbnail in a patch of size $K \times K$ (a patch is a square centred on an image pixel), corresponding to the thumbnail, to a matrix of size $M \times N \times C$ corresponding to the input image. Such a matrix therefore represents the probabilities of each of the classes appearing at any point of the original image. Step 302 of extracting the thumbnails according to position and scale level criteria makes it possible to construct the probability matrix in full.

[0101] Figure 4 shows one example of a probability matrix 40 and patches 400, according to one exemplary embodiment. Figure 4 illustrates a change in the step of the sliding window that extracts the thumbnails, the offset between two thumbnails being set at the start of the method and remaining unchanged throughout the application thereof. It should be noted that it is possible to drop to a step of 1 pixel. However, the smaller the step, the more the number of operations increases. For example, for a step of $224/x$ pixels, the number of operations is 4^{x-1} times greater than for a step of 224 pixels.

[0102] In one embodiment, a class may be constrained to a given thumbnail resolution so as to take account of the variety of environments in which autonomous vehicles operate. Indeed, in such environments, it is common for some classes to represent texture information rather than structural or contour information, and vice versa. Depending on the application of the invention, it is possible to observe the results at various extraction dimensions (that is to say scales) and to choose the most coherent dimension for each class.

[0103] Figure 5 illustrates a change in the size of the sliding window as a function of the class. As illustrated in figure 5, a different factor may be applied for each class in order to take account of the variable size of patches added to the probability matrix.

[0104] Thumbnails of different sizes are extracted from one and the same image and make it possible to establish the probabilities of belonging to specific classes as a function of the information that they contain (for example low-frequency or high-frequency frequency level information).

[0105] Considering a probability matrix Mp , a probability vector P at the output of the neural network-based classifier 122 and the vector D of the various extraction dimensions (that is to say scale levels), the probability matrix may be defined according to the following equation (1):

$$[0106] \quad Mp(i, j, n) = Mp(i, j, n) + P(n) * \frac{\min(D)}{D(n)} \quad (1)$$

[0107] For a given thumbnail dimension, only the probabilities on the classes concerned by this dimension are added in step 302.

[0108] In one embodiment, the step set in step 302 of extracting the thumbnails (step of the sliding window) may make it possible to control the finesse of the semantic annotation of the final image. In some embodiments, Gaussian filtering-type smoothing may also be implemented in order to improve the readability of the information present at the boundaries of the classes, which are defined by the size of the extraction step, according to equation (3). Indeed, with a given step, it is not generally possible to obtain various predicted classes on an area smaller than such a pixel step. In some embodiments, each probability matrix resulting from step 304 may be weighted by a two-dimensional Gaussian window, for each scale level

used. Thus, rather than directly adding the probability of the class under consideration to an area $Z \times Z$, the results of a two-dimensional Gaussian window of size $Z \times Z$ multiplied by the probability of the class are added according to the following equation (2):

$$[0109] \quad Mp(i, j, n) = Mp(i, j, n) + G(i, j) * P(n) * \frac{\min(D)}{D(n)} \quad (2)$$

5 [0110] In equation (2), the Gaussian window $G(i, j)$ is defined by:

$$[0111] \quad G(i, j) = e^{\left(\frac{1}{2}\right)\left(\frac{i}{\sigma}\right)^2} e^{\left(\frac{1}{2}\right)\left(\frac{j}{\sigma}\right)^2} \text{ for } i \in [0, Z], j \in [0, Z] \quad (3)$$

[0112] The semantic annotation of the initial image may then be determined by selecting, at each point of the probability matrix Mp , the index between 0 and C , the class having the highest probability. Each pixel of the input image is thus assigned the class to which it is most likely to belong.

10

[0113] Figure 6 illustrates the method for determining image annotations, in one embodiment in which a single-scale approach is used.

[0114] In the example of figure 6, the method according to the embodiments of the invention determines deteriorated annotations for C classes. The original image is divided for example into thumbnails of 224×224 , and the classifier 122 is applied to each thumbnail in order to form a vector of probabilities of belonging to each of the classes. The vector thus obtained is added to the part corresponding to the extracted thumbnail in a probability matrix of the same size as that of the input image, but having a depth equal to C , rather than the depth of the input image (which may be an RGB (red green blue) image) equal to 3. In one embodiment, in step 308,

15

20 each pixel of the probability matrix Mp may have a color associated with it in order to determine annotations suitable for being displayed in a human-machine interface. Moreover, the index having the maximum probability (that is to say index having the maximum value in the probability vector associated with each pixel) may be extracted in order to automatically generate the semantic annotation of the initial image.

25 [0115] Figure 7 illustrates the method for determining image annotations, in one embodiment in which a multi-scale approach is used. In such an embodiment, thumbnails of various resolutions are extracted from the initial image in order to encapsulate the scale level information (for example high-frequency and low-frequency information) and thus extract not only the image contours, but also the image texture. The various thumbnails may then be

30 adjusted in terms of size in order to be classified in step 304, before being merged in step 306 in order to determine the final probability matrix used to automatically construct the semantic annotation of the input image under consideration.

[0116] Figure 8 illustrates the method for determining image annotations, applied to a sequence of images, using the pre-trained classifier 122 (block 81) based on a classifier model 82 and annotated images 80A for each class making up the scene being studied, a sequence of input images 80B for which a semantic annotation is to be determined. The method
5 implements a weighted sliding window-based multi-scale classification approach (block 83) in order to establish a probability matrix (block 85) of each pixel belonging to a given class in order to determine (block 84) the semantic annotation of the initial image, thereby making it possible to obtain annotated images 85.

[0117] The embodiments of the invention are thus particularly suitable for any application
10 using a semantic segmentation neural network, and more particularly for re-teaching it based on new data, such as for example a seabed segmentation application. They allow the deployment of AUV in such environments with a reduced semantic annotation cost. In such an exemplary application, the embodiments of the invention may be used for risk detection. Indeed, they make it possible to determine, automatically or dynamically, semantically
15 annotated images on a predetermined number of classes representative of seabeds, to highlight the proportion of each class in an image and to analyze whether a portion of the image does not belong to these classes, this indicating the probability of being faced with an unknown element. According to another example, the system and method according to the
20 embodiments of the invention may be used to calculate the proportion of the representation of the class "water" in the image (that is to say the class representing water in the images) in order to determine whether an AUV (autonomous underwater vehicle) is at an acceptable depth to navigate its environment using SLAM (simultaneous localization and mapping) technologies, for example.

[0118] A person skilled in the art will understand that the system or subsystems according to
25 the embodiments of the invention may be implemented in various ways by hardware, software or a combination of hardware and software, notably in the form of program code that may be distributed in the form of a program product, in various forms. In particular, the program code may be distributed using computer-readable media, which may include computer-readable storage media and communication media. The methods described in the present
30 description may notably be implemented in the form of computer program instructions able to be executed by one or more processors in a computer device. These computer program instructions may also be stored in a computer-readable medium.

[0119] The invention is not limited to the embodiments described above by way of non-limiting
35 example. It encompasses all variant embodiments that may be envisaged by a person skilled in the art.

CLAIMS

1. A device for classifying a scene of the environment of a vehicle, comprising a classification unit (14) that uses a neural network configured to classify a scene in response to a received image depicting the scene and a learning unit (11) configured to train said neural network on a set of images stored in a database (3), characterized in that it comprises an image annotation unit comprising:
- an image preprocessing module (120) configured to extract a set of thumbnails from a received image as a function of a criterion concerning position in the image and at least one criterion relating to an image parameter,
 - a classification module (122) configured to determine, for each thumbnail, a set of probabilities of belonging, each associated with a class from among a set of predefined classes, and
 - an image annotation module (124) configured to determine an annotation corresponding to said image based on the sets of probabilities determined for all of said thumbnails, said annotation being transmitted to said database (3) for storage in association with said image in the database.
2. The classification device as claimed in claim 1, wherein the image annotation module (124) is configured to merge the probabilities of belonging determined for all of the classes and for all of the thumbnails, and to reconstruct an annotation, pixel by pixel, for said received image, thereby providing said annotation.
3. The classification device as claimed in either of claims 1 and 2, wherein each set of probabilities determined for a thumbnail is represented by a probability vector having a set of components corresponding to said probabilities of belonging, each associated with one of said classes.
4. The classification device as claimed in one of the preceding claims, wherein an image parameter is image frequency information, the criterion relating to said image parameter being a frequency level.
5. The classification device as claimed in one of the preceding claims, wherein the preprocessing unit uses a fixed or variable step to implement sliding window-based thumbnail extraction.
6. The classification device as claimed in one of the preceding claims, wherein the image annotation module (124) is configured to select, for each image pixel, the class having a maximum probability of belonging from among the probabilities of belonging obtained for all of the thumbnails.

7. The classification device as claimed in one of the preceding claims, wherein the preprocessing module is configured, prior to determining the set of probabilities for a thumbnail, to implement an operation of resizing the thumbnail as a function of the level of its image parameter.

5 8. The classification device as claimed in one of the preceding claims, wherein the preprocessing module is configured, for each thumbnail of size $K \times K$ from among the P thumbnails extracted from the input image of size $M \times N$, to add the C probabilities of belonging obtained for the thumbnail in a patch of size $K \times K$, corresponding to the thumbnail, to a matrix of size $M \times N \times C$ corresponding to the input image.

10 9. The classification device as claimed in claims 3 and 8, wherein the probability matrix Mp is defined according to the equation:

$$Mp(i, j, n) = Mp(i, j, n) + P(n) * \frac{\min(D)}{D(n)},$$

where P denotes the probability vector P at the output of the classification unit and D denotes the vector of the various extraction dimensions of the thumbnails.

15 10. The classification device as claimed in claims 3 and 8, wherein the probability matrix Mp is defined according to the equation:

$$Mp(i, j, n) = Mp(i, j, n) + G(i, j) * P(n) * \frac{\min(D)}{D(n)}$$

20 where P denotes the probability vector P at the output of the classification unit, $G(i, j)$ denotes a Gaussian window and D denotes the vector of the various extraction dimensions of the thumbnails.

11. The classification device as claimed in claim 10, wherein the Gaussian window $G(i, j)$ is defined by:

$$G(i, j) = e^{\left(\frac{1}{2}\right)\left(\frac{i}{\sigma}\right)^2} e^{\left(\frac{1}{2}\right)\left(\frac{j}{\sigma}\right)^2} \text{ for } i \in [0, Z], j \in [0, Z],$$

25 the annotation of the received image then being determined by selecting, at each point of the probability matrix Mp , the index having the maximum value in the probability vector associated with each pixel.

30 12. A method for classifying a scene of the environment of a vehicle, comprising a classification step that uses a neural network configured to classify a scene in response to a received image depicting the scene and a learning step configured to train said neural network on a set of images stored in a database (3), characterized in that the method comprises:

- preprocessing the image (120), consisting in extracting a set of thumbnails from a received image as a function of a criterion concerning position in the image and at least one criterion relating to an image parameter,

- a classification step (122) consisting in determining, for each thumbnail, a set of probabilities of belonging, each associated with a class from among a set of predefined classes, and

- an image annotation step (124) consisting in determining an annotation corresponding to said image based on the sets of probabilities determined for all of said thumbnails, said annotation being transmitted to said database (3) for storage in association with said image in the database.