

DESCRIPTION

Dispositif de classification de scène de l'environnement d'un véhicule**Domaine technique**

[0001] L'invention concerne de manière générale les systèmes de traitement d'image et en particulier un dispositif et un procédé de classification de scène de l'environnement d'un véhicule automatisé.

[0002] Les technologies d'automatisation des systèmes de conduite ont connu un essor majeur au cours des dernières années, dans le domaine des véhicules automatisés, tels que les véhicules autonomes, qu'ils soient terrestres, marins ou sous-marins. La navigation autonome est rendue possible par la reconnaissance en temps réel de l'environnement du véhicule qui a l'avantage de ne nécessiter que très peu de matériel. En effet, généralement, seul un dispositif d'acquisition d'images associées à une unité de calcul, embarquée ou non, est nécessaire. Dans des solutions existantes, l'unité de calcul est configurée pour effectuer une classification sémantique des scènes de l'environnement. Pour cela, l'unité de calcul utilise un procédé de segmentation sémantique qui permet d'obtenir une information sur la classe de chacun des pixels d'une image. Il s'agit d'une tâche complexe de vision par ordinateur qui s'appuie classiquement sur des réseaux de neurones profonds entraînés pour la classification sémantique des scènes de l'environnement du véhicule.

[0003] Ainsi, lorsque le véhicule se déplace dans son environnement, il est en mesure d'analyser finement l'environnement pour détecter les obstacles, les éviter et optimiser son déplacement en cohérence avec le milieu dans lequel il évolue.

[0004] Dans les systèmes actuels, l'apprentissage de réseaux de neurones est effectué à partir d'une base de données regroupant des milliers d'images annotées sémantiquement. Les réseaux de neurones apprennent ainsi à reconnaître automatiquement la classe des pixels d'une image grâce à ces annotations.

[0005] Cependant, un frein important au déploiement de tels systèmes est le coût humain élevé lié à la mise en place de ces approches, dites de segmentation sémantique. En effet, de tels systèmes nécessitent une annotation manuelle, au niveau de chaque pixel d'une image, parmi les milliers d'images de la base de données pour permettre une classification robuste. Or, ces annotations sémantiques sont nécessaires pour pouvoir mettre en œuvre la phase d'apprentissage de l'intelligence artificielle. Ainsi, le déploiement de véhicules autonomes sur des scènes encore inexplorées nécessite un travail conséquent de ré-annotation manuelle, entraînant des retards importants dans l'exploitation de ces nouveaux environnements. De plus l'annotation précise de ces images est parfois subjective du fait du

milieu étudié où la frontière entre deux classes n'est pas clairement définie (par exemple, frontière entre le sable et la colonne d'eau dans un milieu sous-marin).

[0006] Il existe ainsi un besoin pour un dispositif et un procédé amélioré de classification des scènes de l'environnement d'un véhicule automatisé.

5 Définition générale de l'invention

[0007] L'invention vient améliorer la situation en proposant un dispositif de classification d'une scène de l'environnement d'un véhicule, comprenant une unité de classification utilisant un réseau de neurones configuré pour classer une scène en réponse à une image reçue représentant la scène et une unité d'apprentissage configurée pour entraîner le réseau de neurones, à partir d'un ensemble d'images stockées dans une base de données. Avantageusement, le dispositif de classification comprend une unité d'annotation d'image comprenant :

[0008] - un module de pré-traitement d'image configuré pour extraire un ensemble d'images d'une image reçue en fonction d'un critère de position dans l'image et d'au moins un critère relatif à un paramètre d'image,

[0009] - un module de classification configuré pour déterminer, pour chaque image, un ensemble de probabilités d'appartenance, chacune associée à une classe parmi un ensemble de classes prédéfinies, et

[0010] - un module d'annotation d'image configuré pour déterminer une annotation en correspondance avec l'image à partir des ensembles de probabilités déterminées pour toutes les images, l'annotation étant transmise à la base de données pour un stockage en association avec l'image dans la base de données.

[0011] Selon un aspect de l'invention le module d'annotation d'image peut être configuré pour fusionner les probabilités d'appartenance déterminées pour toutes les classes et pour toutes les images et pour reconstruire une annotation, pixel par pixel, pour l'image reçue, ce qui fournit l'annotation.

[0012] Dans un mode de réalisation, chaque ensemble de probabilités déterminé pour une image peut être représenté par un vecteur de probabilités ayant un ensemble de composantes correspondant aux probabilités d'appartenance, chacune associée à une des classes.

[0013] Dans un mode de réalisation, un paramètre d'image peut être une information de fréquence d'image, le critère relatif au paramètre d'image étant un niveau de fréquence.

[0014] Le critère d'image utilisé pour l'extraction peut être relatif à des plages de valeurs du paramètre d'image.

[0015] Dans un mode de réalisation, l'unité de prétraitement utilise un pas fixe ou variable pour mettre en œuvre une extraction des imagerie par fenêtre glissante.

[0016] Dans un mode de réalisation, le module d'annotation d'image peut être configuré pour sélectionner pour chaque pixel d'image, la classe ayant une probabilité d'appartenance maximale parmi les probabilités d'appartenance obtenues pour toutes les imagerie.

[0017] Dans un mode de réalisation, le module de pré-traitement peut être configuré pour mettre en œuvre, préalablement à la détermination de l'ensemble de probabilités pour une imagerie, une opération de redimensionnement de l'imagerie en fonction du niveau de son paramètre d'image.

[0018] Dans un mode de réalisation, le module de pré-traitement peut être configuré pour ajouter, pour chaque imagerie de taille $K \times K$ parmi les P imagerie extraites de l'image d'entrée de taille $M \times N$, les C probabilités d'appartenance obtenues pour l'imagerie dans un patch de taille $K \times K$, correspondant à l'imagerie, sur une matrice de taille $M \times N \times C$ correspondant à l'image d'entrée.

Dans un mode de réalisation, la matrice de probabilité Mp peut être définie selon l'équation :

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

où P désigne le vecteur de probabilités P en sortie de l'unité de classification et D le vecteur des différentes dimensions d'extraction des imagerie.

Dans un mode de réalisation, la matrice de probabilité Mp peut être définie selon l'équation :

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

P désigne le vecteur de probabilités P en sortie de l'unité de classification, $G(i, j)$ désigne une fenêtre Gaussienne et D désigne le vecteur des différentes dimensions d'extraction des imagerie.

Selon un aspect de l'invention la fenêtre gaussienne $G(i, j)$ peut être définie par :

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

[0020] L'annotation de l'image reçue peut alors être déterminée en sélectionnant, en chaque point de la matrice de probabilité Mp , l'indice ayant la valeur maximale dans le vecteur de probabilités associé à chaque pixel.

[0021] Il est en outre proposé un procédé de classification d'une scène de l'environnement d'un véhicule, comprenant une étape de classification utilisant un réseau de neurones configuré pour classer une scène en réponse à une image reçue représentant la scène et

une étape d'apprentissage configurée pour entraîner le réseau de neurones, à partir d'un ensemble d'images stockées dans une base de données. Avantageusement, le procédé comprend :

5 [0022] - un pré-traitement de l'image consistant à extraire un ensemble d'imagettes d'une image reçue en fonction d'un critère de position dans l'image et d'au moins un critère relatif à un paramètre d'image,

[0023] - une étape de classification consistant à déterminer, pour chaque imagette, un ensemble de probabilités d'appartenance chacune associée à une classe parmi un ensemble de classes prédéfinies, et

10 [0024] - une étape d'annotation d'image consistant à déterminer une annotation en correspondance avec l'image à partir des ensembles de probabilités déterminées pour toutes les imagettes, l'annotation étant transmise à la base de données pour un stockage en association avec l'image dans la base de données.

15 [0025] Les modes de réalisation de l'invention permettent ainsi d'annoter les pixels des images de la base de données d'apprentissage, avec une complexité et un temps de traitement réduits.

Brève Description des Figures

20 [0026] D'autres caractéristiques, détails et avantages de l'invention ressortiront à la lecture de la description faite en référence aux dessins annexés donnés à titre d'exemple et qui représentent, respectivement :

[0027] [Fig.1] - La figure 1 représente schématiquement un véhicule implémentant un système de contrôle, selon les modes de réalisation de l'invention.

[0028] [Fig.2] - La figure 2 représente schématiquement la structure de l'unité d'annotation selon des modes de réalisation de l'invention.

25 [0029] [Fig.3] La figure 3 décrit le procédé d'annotation d'images mis en œuvre par l'unité d'annotation d'image selon des modes de réalisation de l'invention.

[0030] [Fig.4] La figure 4 représente un exemple de matrice de probabilité et de patches, selon un exemple de réalisation.

30 [0031] [Fig.5] La figure 5 illustre un changement de la taille de la fenêtre glissante en fonction de la classe.

[0032] [Fig.6] La figure 6 illustre le procédé de détermination d'annotations d'image, dans un mode de réalisation où une approche mono-échelle est utilisée.

[0033] [Fig.7] La figure 7 illustre le procédé de détermination d'annotations d'image, dans un mode de réalisation où une approche multi-échelle est utilisée.

[0034] [Fig.8] La figure 8 illustre le procédé de détermination d'annotations d'image, appliqué à une séquence d'images, utilisant le classifieur pré-entraîné.

5 Description détaillée de la demande

[0035] La Figure 1 représente schématiquement un véhicule 100 implémentant un système de contrôle 10 du véhicule, selon les modes de réalisation de l'invention.

10 [0036] Le système de contrôle 10 est configuré pour reconnaître en temps réel l'environnement du véhicule 100 et d'adapter la conduite du véhicule 100 dynamiquement en fonction de la reconnaissance effectuée afin de permettre une navigation automatisée, comme par exemple une navigation autonome.

[0037] Le système de contrôle 10 comprend un dispositif de classification de scènes de l'environnement du véhicule 1 (encore appelé plus simplement « système de classification ») et un système de perception 2.

15 [0038] Le dispositif de classification 1 est configuré pour effectuer une classification sémantique des scènes de l'environnement du véhicule 100 à partir des images détectées par le système de perception 2 du véhicule 100.

20 [0039] Le véhicule 100 peut être un véhicule automatisé ou semi-automatisé apte à se déplacer dans un environnement, tel qu'un véhicule terrestre, marin ou sous-marin. L'environnement peut donc être un environnement terrestre, marin ou sous-marin.

25 [0040] L'environnement du véhicule 100 est associé à un ensemble de classes prédéfinies, en fonction de l'application de l'invention, telles que par exemple l'ensemble de classes suivant : [chaussée, trottoir, végétation, véhicule]. Dans des modes de réalisation, les classes associées à l'environnement peuvent être modifiées (ajout, suppression de classes par exemple) en mettant en œuvre un réapprentissage du système.

[0041] Le système de contrôle 10 peut être un système d'aide à la conduite configuré pour optimiser la conduite du véhicule et assurer sa sécurité (détection d'obstacles par exemple) à partir des informations déterminées par le système de perception 2.

30 [0042] Le système de perception 2 comprend un ensemble de dispositifs de détection agencés sur le véhicule 100 et configurés pour fournir des signaux représentatifs d'une image de l'environnement dans lequel se déplace le véhicule 100. Les dispositifs de détection 20 peuvent comprendre une ou plusieurs caméras. Une caméra peut être équipée d'un capteur photosensible sensible au rayonnement d'un spectre de lumière compris dans l'infrarouge et/ou dans le visible. Les caméras peuvent en outre comprendre une ou

plusieurs caméras acoustiques et/ou laser (dans des applications de l'invention en milieu sous-marin par exemple).

[0043] Les dispositifs de détection 20 peuvent être agencés sur le véhicule de manière à acquérir des images de l'environnement du véhicule situé à l'avant du véhicule 100, sur les
5 côtés, au-dessus et/ou au-dessous du véhicule 100 en fonction de l'application de l'invention. Par exemple, dans un environnement terrestre, les dispositifs de détection peuvent être fixés à l'avant du véhicule et/ou dans l'habitacle du véhicule et/ou sur le toit du véhicule. Les dispositifs de détection peuvent en outre comprendre un lidar (Laser Detection And Ranging), et/ou un radar.

10 [0044] D'autres types de capteurs peuvent être utilisés par le système de perception 2 tels qu'un capteur à ultrasons, un capteur d'angle de volant, un capteur de vitesse de roue, un capteur de pression de freinage, un capteur de vitesse de lacet et d'accélération transversales, ou une combinaison de ces derniers.

[0045] Le système de perception 2 est configuré pour déterminer des images de
15 l'environnement à partir des informations détectées par l'ensemble de dispositifs de détection. Dans des modes de réalisation, le système de perception 2 peut comprendre en outre un dispositif de traitement 22 configuré pour mettre en œuvre des algorithmes de fusion multicapteur aptes à combiner des informations issues des différents dispositifs de détection 20 pour déterminer les images de l'environnement.

20 [0046] Le dispositif de classification 1 comprend une unité d'apprentissage 11 pour entraîner au préalable le réseau de neurones utilisé sur des jeux de données réelles stockées dans une base de données d'images 3. La base de données contient pour chaque donnée un couple image/annotation.

[0047] Le dispositif de classification 1 comprend en outre une unité d'annotation d'image 12
25 configurée pour annoter automatiquement une ou plusieurs images de la base de données d'images 3 utilisée pour la phase d'apprentissage. L'unité d'annotation d'image 12 est ainsi configurée pour générer les annotations stockées dans la base de données d'images 3. Une annotation d'image est représentée par une matrice de taille identique à l'image où chaque élément a une valeur entre 0 et N-1 avec N désignant le nombre de classes (appartenance
30 d'un pixel à une classe).

[0048] De telles annotations sont utilisées par l'unité d'apprentissage 11 pour entraîner le ou les réseaux de neurones utilisés.

[0049] Le dispositif de classification 1 comprend une unité de classification 14 configurée pour segmenter sémantiquement chaque image détectée par le système de perception 2 en
35 temps réel, en utilisant le réseau de neurones préalablement entraîné.

[0050] Pour faciliter la compréhension des modes de réalisation de l'invention, des définitions ou notions relatives aux réseaux de neurones sont détaillées ci-après en relation avec les modes de réalisation de l'invention.

5 [0051] Un réseau de neurones comprend des neurones interconnectés entre eux par des synapses qui peuvent être implémentées sous la forme de mémoires numériques. Un réseau de neurones peut comprendre un ensemble de couches successives, comprenant une couche d'entrée portant le signal d'entrée et une couche de sortie portant le résultat de la prédiction effectuée par le réseau de neurones (résultat du réseau), et une ou plusieurs couches intermédiaires. Chaque couche d'un réseau de neurones prend ses entrées sur les
10 sorties de la couche précédente. Le nombre de neurones sur chaque couche est égal au nombre d'entrées des neurones de la couche suivante. Une couche donnée du réseau de neurones comprend ainsi un ensemble de neurones prenant leurs entrées sur les neurones de la couche précédente.

[0052] Les signaux propagés en entrée et en sortie des couches du réseau peuvent être des
15 valeurs numériques (information codée dans la valeur des signaux), ou des impulsions électriques dans le cas d'un codage impulsionnel (information codée temporellement selon l'ordre d'arrivée des impulsions ou selon la fréquence des impulsions).

[0053] Un réseau de neurones comprend un ensemble de données d'entrée (ou 'coefficients d'entrée') et des données de sortie (ou 'coefficients de sortie').

20 [0054] Les valeurs de sorties sont calculées à partir des entrées et des poids synaptiques en appliquant une fonction d'activation aux coefficients d'entrée.

[0055] Chaque neurone du réseau de neurones est configuré pour calculer une somme pondérée de ses coefficients d'entrée en utilisant une fonction de combinaison et les poids synaptiques puis en appliquant une fonction d'activation à la somme pondérée obtenue pour
25 produire sa sortie :

[0056] Les poids synaptiques d'un réseau de neurones sont déterminés par apprentissage dans une phase d'apprentissage mise en œuvre par l'unité d'apprentissage. Des valeurs aléatoires sont initialement assignées aux poids du réseau de neurones puis un ensemble de données réelles issues de la base de données 3 sont utilisées pour réaliser
30 l'apprentissage. L'apprentissage d'un réseau de neurones consiste à déterminer les valeurs optimales des poids synaptiques, pour chaque neurone du réseau de neurones, de la dernière couche du réseau à la première, en utilisant une fonction d'apprentissage.

[0057] La phase d'apprentissage peut mettre en œuvre une pluralité d'itérations de la fonction d'apprentissage, chaque itération comprenant une étape de propagation vers l'avant ('forward') et une étape de rétropropagation ('backward') pour corriger les erreurs entre les
35

sorties obtenues dans la phase de propagation avant et les sorties attendues pour l'échantillon d'entrée considéré.

[0058] La phase d'apprentissage permet ainsi de comparer la sortie obtenue par rapport à la sortie attendue (dans le cas d'une méthode supervisée), et en fonction de cette comparaison, de mettre à jour les liaisons entre les neurones représentés par les poids synaptiques pour améliorer le résultat final.

[0059] Dans la phase de propagation vers l'avant, des ensembles de données d'entrée de la base de données 3 sont utilisés pour mettre en œuvre l'apprentissage. Chaque ensemble de données forme un échantillon x associé à des valeurs souhaitées (ou valeurs attendues). Le signal correspondant à l'échantillon d'entrée est propagé en avant dans les couches du réseau de neurones à partir de la première couche, d'une couche $(k-1)$ à la couche suivante (k) jusqu'à la dernière couche. Dans la phase de propagation vers l'avant, la fonction d'activation ϕ et les poids synaptiques connectant les neurones d'une couche précédente $(k-1)$ et d'une couche suivante (k) sont utilisés.

[0060] Lorsque la propagation vers l'avant est terminée, un résultat y est obtenu à la sortie.

[0061] Dans la phase de rétropropagation, les erreurs éventuelles obtenues par un neurone sont rétropropagées à ses synapses et aux neurones qui lui sont reliés. La rétropropagation peut être par exemple une rétropropagation par gradient pour modifier les poids synaptiques en tenant compte de leur impact dans les erreurs engendrées. Ainsi, les poids synaptiques qui contribuent à engendrer une erreur importante peuvent être modifiés de manière plus significative que les poids qui ont engendré une erreur moins importante.

[0062] La durée de la phase d'apprentissage peut dépendre de la taille de la base de données 3 stockant les échantillons utilisés pour l'apprentissage et de la taille du réseau.

[0063] Après la phase d'apprentissage, une phase dite de généralisation, plus rapide, est mise en œuvre par l'unité de classification 14. Dans la phase de généralisation, les poids appris de la phase d'apprentissage sont utilisés (réseau de neurones statique dans lequel les poids sont figés). Des données d'entrée correspondant aux images de la scène détectées en temps réel par le système de perception 2 sont présentées au réseau de neurones qui fournit une réponse du réseau de neurones représentant la probabilité d'appartenance de chaque pixel aux classes prédéfinies.

[0064] Le système de contrôle 2 peut mettre en œuvre un ensemble d'applications de conduite utilisant la classification déterminée par le dispositif de classification 1 telles qu'une application de changement de trajectoire pour aider le conducteur à changer de trajectoire ou une application d'avertissement de collision.

[0065] La figure 2 représente schématiquement la structure de l'unité d'annotation 12 selon des modes de réalisation de l'invention.

[0066] L'unité d'annotation d'image 12 est configuré pour déterminer une annotation à assigner à une image représentative d'une scène de l'environnement à analyser, en réponse à la réception de cette image déterminée par le système de perception 2. Le dispositif de classification 1 peut ensuite stocker l'image en association avec l'annotation déterminée dans la base de données 3.

[0067] Le dispositif de classification de scènes 1 peut ainsi effectuer une annotation sémantique des images permettant de produire une segmentation sémantique de l'environnement du véhicule 100 en un temps minimal, même dans le cas de scènes encore inexplorées où aucune annotation n'est disponible.

[0068] Dans un mode de réalisation, comme illustré sur la figure 2, l'unité d'annotation d'image 12 peut comprendre :

[0069] – un module de prétraitement d'image 120 configuré pour pré-traiter les images et extraire des sous-images (ou imagerie) de l'image reçue ('image d'entrée) ;

[0070] – un module de classification 122 (encore appelé « classifieur ») configuré pour effectuer une classification des imagerie extraites par l'unité de prétraitement 102, en utilisant le réseau de neurones, pour déterminer les probabilités d'appartenance de chaque imagerie ('thumbnail image' en langue anglo-saxonne) à chacune des classes de l'environnement du véhicule ;

[0071] – Un module d'annotation 124 (encore appelé « unité de reconstruction ») configuré pour fusionner les réponses du module de classification 122 obtenues pour chaque imagerie et pour reconstruire une annotation sémantique, pixel par pixel, pour l'image complète considérée, ce qui fournit l'annotation à assigner à l'image reçue.

[0072] Le module de prétraitement d'image 120 peut être configuré pour collecter les informations locales de l'image considérée, selon plusieurs niveaux et pour extraire un ensemble d'imagerie en fonction d'un critère de position d'image et d'au moins un critère d'image relatif à un paramètre d'image. Une imagerie peut ainsi correspondre à des informations de position dans l'image et à des critères d'image définis à partir des niveaux d'échelle du paramètre d'image considéré (i.e. de plages de valeurs du paramètre d'image).

[0073] Le critère d'image utilisé pour l'extraction des imagerie peut être relatif à différents types de paramètres d'image, tels que par exemple la fréquence d'image, la texture d'image, les contours d'image.

[0074] Les niveaux d'échelle du paramètre d'image considérés pour l'extraction peuvent être par exemple une pyramide d'échelle de facteur donné, comme par exemple de facteur 2. Dans un exemple d'application de l'invention où le critère d'image utilisé pour l'extraction est relatif au paramètre d'image correspondant à la résolution maximale d'image, pour une
5 résolution maximale de 224x224 pixels, et une résolution minimale de 14x14 pixels, la pyramide d'échelle de facteur 2 utilisée comprend les niveaux : 224x224 ; 112x112 ; 56x56 ; 28x28 ; et 14x14.

[0075] La résolution maximale peut dépendre de la résolution du système de perception (2). Les imagerie sont extraites pour chaque niveau d'échelle.

10 [0076] L'unité de prétraitement 120 peut prendre en compte les informations basses et hautes fréquences inhérentes à chacune des classes de la scène de l'environnement considéré. L'unité de prétraitement 120 peut utiliser un pas fixe pour mettre en œuvre une extraction par fenêtre glissante multi-échelles. Le pas peut être prédéfini. Le pas peut être appliqué sur la grille de pixels de l'image (verticalement et horizontalement). Le pas et les
15 niveaux d'échelles permettent de paramétrer la finesse des annotations sémantiques qui seront automatiquement générées.

[0077] Le classifieur 122 est configuré pour déterminer, pour chaque imagerie extraite, un ensemble de probabilités d'appartenance chacune associée à une classe parmi un ensemble de classes prédéfinies.

20 [0078] Dans un mode de réalisation, le classifieur 122 est configuré pour assigner, à chaque imagerie extraite par le module de prétraitement d'image 120, un vecteur caractéristique (encore appelé « vecteur de probabilités ») comprenant des composantes, chaque composante étant associée à une des classes de la scène de l'environnement considéré et ayant une valeur correspondant à la probabilité d'appartenance de l'imagerie à la classe
25 associée. Ainsi, le vecteur de probabilités assigné à une imagerie représente la probabilité d'appartenance de l'imagerie à chacune des classes de la scène de l'environnement considéré. Le classifieur 122 utilise le réseau de neurones, qui peut être notamment un réseau de neurones profonds, préentraîné au préalable sur un jeu de données adapté, pendant une phase d'apprentissage, pour permettre une classification robuste.

30 [0079] Le module d'annotations 124 est configuré pour fusionner les réponses du classifieur 122 (représentées par les vecteurs de probabilité obtenus pour chaque imagerie) de manière à agréger les probabilités d'appartenance obtenues pour chaque imagerie, correspondant à des positions d'image et à des critères d'image différents. Le module d'annotations 124 est en particulier configuré pour attribuer à chaque pixel de l'image initiale un vecteur de
35 probabilités indiquant son degré d'appartenance aux classes présentes dans la scène

considérée. Le module d'annotation 124 associe ainsi à l'image considérée une annotation sémantique en sélectionnant pour chaque pixel d'image la classe ayant une probabilité d'appartenance maximale. Tel qu'utilisé ici, le terme « annotation » est associé à une image et fait référence à un ensemble d'informations de classe, chaque information de classe étant associée à un pixel de l'image et correspondant à la classe de probabilité maximale déterminée.

[0080] L'unité d'annotation d'image 12 peut alors stocker les annotations déterminées en association avec l'image correspondante dans la base de données d'images 3.

[0081] L'unité d'apprentissage 11 peut alors mettre en œuvre un apprentissage du réseau de neurones de segmentation sémantique n'utilisant pas des annotations du même type.

[0082] Pour réaliser l'apprentissage, l'unité d'apprentissage 11 peut utiliser les annotations ainsi générées.

[0083] Il est à noter que les annotations déterminées par l'unité de reconstruction 124 peuvent ne pas être parfaites et former un ensemble d'annotations dites dégradées. Cependant, la construction du réseau de neurones des modules 11 et 14 étant de type encodeur/décodeur, ce type de réseau va naturellement lisser les annotations fournies en entrées et ainsi limiter l'impact de l'imperfection de telles annotations dégradées.

[0084] Le classifieur 122 peut être entraîné pour reconnaître les classes sur des imagerie de taille donnée, telle que par exemple des images de tailles 224x224. Le classifieur 122 peut être configuré pour prédire les probabilités d'appartenance aux classes prédéfinies sur chaque image extraite de l'image, à plusieurs niveaux d'échelles. Dans un mode de réalisation particulier, le classifieur 122 peut notamment être configuré pour prédire la classe majoritaire (i.e. ayant une probabilité d'appartenance maximale) sur chaque image extraite de l'image, à plusieurs niveaux d'échelles.

[0085] De telles prédictions peuvent être obtenues en appliquant un principe sous-jacent de fenêtre glissante multi-échelles. La fenêtre glissante est une technique qui permet d'analyser une image en divisant celle-ci en petites régions appelées fenêtres et en déplaçant progressivement cette fenêtre sur toute l'image pour effectuer une analyse locale. Le processus de fenêtre glissante comporte trois étapes :

[0086] – une étape d'initialisation de la fenêtre en position et en taille. La fenêtre peut être carrée et sa position initiale être située à l'origine de l'image. La taille de la fenêtre dépend de l'échelle fixée dans l'approche multi-échelle.

[0087] – une étape d'analyse des pixels de la fenêtre en effectuant une classification de la fenêtre pour en extraire ses probabilités d'appartenance aux classes du système.

- [0088] – une étape de déplacement de la fenêtre, après avoir analysé la région actuelle. Dans cette étape, la fenêtre est déplacée d'un certain nombre de pixels, appelé "pas". Cette étape se répète jusqu'à ce que la fenêtre ait couvert l'ensemble de l'image. Le pas peut être fixe et égal à la résolution de l'échelle minimale (par exemple pour une taille d'image de 14x14, le pas est égal à 14).
- [0089] Une approche multi-échelle permet d'avoir autant de fenêtres glissantes que de tailles souhaitées (pyramide d'échelle).
- [0090] La fusion des prédictions de classes disponibles pour chaque pixel, réalisée par l'unité de reconstruction 124, permet d'obtenir une annotation sémantique de l'image considérée qui peut être dégradée. Chaque annotation ainsi obtenue vient enrichir la base de données d'annotation sémantique 3 et peut être utilisée pour la mise en œuvre de l'apprentissage par l'unité d'apprentissage 11, ce qui permet d'utiliser un réseau de neurones de segmentation sémantique fort.
- [0091] La figure 3 décrit le procédé d'annotation d'images mis en œuvre par l'unité d'annotation d'image selon des modes de réalisation de l'invention.
- [0092] A l'étape 300, une image d'entrée de taille $M \times N$ est reçue.
- [0093] A l'étape 302, l'image d'entrée est pré-traitée. Cette étape comprend l'extraction de sous-images (ou imagerie) de l'image d'entrée. L'image peut notamment être une image couleur. Chaque imagerie peut être associée à une résolution d'image.
- [0094] Dans l'étape 302, les caractéristiques de l'image d'entrée reçue à l'étape 300 peuvent être stockées dans une mémoire d'imagerie (par exemple un catalogue d'imagerie), dans laquelle les différentes imagerie extraites de l'image sont associées à des positions et à des résolutions d'image fixes, qui peuvent être différentes. Les imagerie peuvent être avantageusement définies selon plusieurs niveaux d'échelle. La prise en compte des informations multi-échelles des imagerie permet la description de l'image à la fois au niveau des textures présentes dans celle-ci, et caractérisant chaque classe, mais aussi au niveau des structures et contours visibles dans l'image. Dans un mode de réalisation, le procédé peut par exemple mettre en œuvre une approche multi-fréquentielle permettant l'extraction des informations basses et hautes fréquences simultanément.
- [0095] A l'étape 304, les probabilités d'appartenance de chaque imagerie extraite à chacune des classes de l'environnement du véhicule sont déterminées en mettant en œuvre une classification (c'est-à-dire une prédiction de classe) de l'imagerie basée sur le réseau de neurones, fournissant un vecteur de probabilités pour chaque classe (cette étape peut être implémentée par le classifieur 122).

- [0096] L'étape 304 (étape de classification) consiste à déterminer, à partir du réseau de neurones préalablement entraîné sur des imagerie annotées et ayant une taille fixe (par exemple 224x224), les probabilités d'appartenance aux différentes classes localement présentes dans chaque imagerie extraite à l'étape 302, et le vecteur de probabilités regroupant ces probabilités d'appartenance pour les différentes classes. Le réseau de neurones peut être un réseau de neurones profonds, tel que par exemple un réseau du type AlexNet ou SqueezeNet. Le réseau de neurones est utilisé pour déterminer les probabilités d'appartenance de chaque imagerie extraite à chacune des C classes, C étant le nombre de classes présentes dans la scène considérée de l'environnement. Ces classes peuvent être prédéfinies pendant l'entraînement (i.e. l'apprentissage) du réseau de classification d'images. Dans des modes de réalisation, l'étape 304 peut comprendre, avant la détermination du vecteur de probabilités d'une imagerie, une opération de redimensionnement d'une imagerie en fonction de son niveau d'échelle, afin de rendre compatible l'entrée du classifieur 122 avec la taille de l'imagerie.
- 15 [0097] L'étape 304 fournit ainsi, pour chaque imagerie, C probabilités d'appartenance qui décrivent localement le contenu de l'image originale à une position et à une échelle donnée de l'image correspondant à une imagerie (les C probabilités d'appartenance pouvant constituer les C composantes du vecteur de probabilités qui est donc alors de taille C).
- 20 [0098] A l'étape 306, une annotation sémantique est déterminée pour chaque pixel de l'image en utilisant les vecteurs de probabilités obtenues pour les différentes images issues de l'image d'entrée. L'étape 306 peut être mise en œuvre par l'unité de reconstruction d'annotation d'image 124.
- 25 [0099] A l'étape 306, une annotation sémantique est déterminée pour chaque image à partir des probabilités d'appartenance aux classes déterminées à l'étape 304, pour toutes les images issues de l'image d'entrée. Ainsi en considérant P images extraites à l'étape 302, l'étape 306 utilise les P vecteurs de probabilités déterminées pour les P images.
- 30 [0100] L'étape 306 comprend, pour chaque imagerie de taille $K \times K$ parmi les P images extraites de l'image d'entrée de taille $M \times N$, un ajout des C probabilités d'appartenance obtenues pour l'imagerie dans un patch de taille $K \times K$ (un patch est un carré centré sur un pixel d'image), correspondant à l'imagerie, sur une matrice de taille $M \times N \times C$ correspondant à l'image d'entrée. Une telle matrice représente donc les probabilités que figure chacune des classes en tout point de l'image d'origine. L'étape 302 d'extraction des images selon des critères de position et de niveaux d'échelle permet de construire entièrement la matrice de probabilités.

[0101] La figure 4 représente un exemple de matrice de probabilité 40 et de patches 400, selon un exemple de réalisation. La figure 4 illustre un changement du pas de la fenêtre glissante qui extrait les imagerie, le décalage entre deux imagerie étant fixé au début du procédé et restant inchangé tout au long de son application. Il est à noter qu'il est possible
 5 de descendre à un pas de 1 pixel. Cependant, plus le pas est faible, plus le nombre d'opérations augmente. Par exemple, pour un pas de $224/x$ pixels, le nombre d'opérations est 4^{x-1} plus important que pour un pas de 224 pixels.

[0102] Dans un mode de réalisation, une classe peut être contrainte à une résolution d'imagerie donnée pour tenir compte de la variété des environnements dans lesquels les
 10 véhicules autonomes évoluent. En effet, dans de tels environnements, il est fréquent que certaines classes représentent plutôt des informations de textures que des informations structurelles ou de contours et inversement. Selon l'application de l'invention, il est possible d'observer les résultats à différentes dimensions d'extraction (i.e. échelles) et de choisir pour chaque classe la dimension la plus cohérente.

[0103] La figure 5 illustre un changement de la taille de la fenêtre glissante en fonction de la classe. Comme illustré sur la figure 5, un facteur différent peut être appliqué pour chaque classe pour tenir compte de la taille variable des patches ajoutés à la matrice de probabilités.

[0104] Sur une même image, des imagerie de tailles différentes sont extraites et permettent d'établir les probabilités d'appartenance à des classes spécifiques en fonction des
 20 informations qu'elles contiennent (par exemple informations de niveau de fréquences, basses ou hautes fréquences).

[0105] En considérant une matrice de probabilités Mp , un vecteur de probabilités P en sortie du classifieur 122 à base de réseau de neurones et le vecteur D des différentes dimensions d'extraction (i.e. niveaux d'échelles), la matrice de probabilité peut être définie selon
 25 l'équation (1) suivante :

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

[0107] Pour une dimension d'imagerie donnée, seules les probabilités sur les classes concernées par cette dimension sont ajoutées à l'étape 302.

[0108] Dans un mode de réalisation, le pas fixé à l'étape 302 d'extraction des imagerie (pas de la fenêtre glissante) peut permettre de contrôler la finesse de l'annotation sémantique de
 30 l'image finale. Dans des modes de réalisation, un lissage de type filtrage gaussien peut être en outre mis en œuvre pour améliorer la lisibilité des informations présentes aux frontières des classes, qui sont définies par la taille du pas d'extraction, selon l'équation (3). En effet, avec un pas donné, il n'est pas possible généralement d'obtenir différentes classes prédites

sur une surface inférieure à un tel pas de pixels. Dans des modes de réalisation, chaque matrice de probabilité issue de l'étape 304 peut être pondérée par une fenêtre gaussienne en deux dimensions, pour chaque niveau d'échelle utilisé. Ainsi, au lieu d'ajouter directement la probabilité de la classe considérée sur une surface $Z \times Z$, les résultats d'une fenêtre gaussienne en deux dimensions de taille $Z \times Z$ multipliée par la probabilité de la classe sont ajoutés selon l'équation (2) suivante :

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

[0110] Dans l'équation (2), la fenêtre gaussienne $G(i, j)$ est définie par :

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

[0112] L'annotation sémantique de l'image initiale peut alors être déterminée en sélectionnant, en chaque point de la matrice de probabilité Mp , l'indice entre 0 et C , de la classe ayant la probabilité la plus élevée. Ainsi, à chaque pixel de l'image d'entrée, est attribué la classe pour laquelle son appartenance est la plus probable.

[0113] La figure 6 illustre le procédé de détermination d'annotations d'image, dans un mode de réalisation où une approche mono-échelle est utilisée.

[0114] Dans l'exemple de la figure 6, le procédé selon les modes de réalisation de l'invention détermine des annotations détériorées pour C classes. L'image d'origine est divisée, par exemple, en imagerie de 224×224 , et le classifieur 122 est appliqué sur chaque imagerie pour former un vecteur de probabilités d'appartenance à chacune des classes. Le vecteur ainsi obtenu est ajouté sur la partie correspondant à l'imagerie extraite dans une matrice de probabilité de même taille que celle de l'image d'entrée, mais ayant une profondeur égale à C , au lieu de la profondeur de l'image d'entrée (qui peut être une image RGB pour 'Red Green Blue', signifiant 'Rouge Vert Bleu') égale à 3. Dans un mode de réalisation, à l'étape 308, à chaque pixel de la matrice de probabilité Mp , il peut être associé une couleur pour déterminer des annotations adaptées à une visualisation dans une interface homme machine. Par ailleurs, l'indice présentant la probabilité maximale (i.e. indice ayant la valeur maximale dans le vecteur de probabilités associé à chaque pixel) peut être extrait pour générer automatiquement l'annotation sémantique de l'image initiale.

[0115] La figure 7 illustre le procédé de détermination d'annotations d'image, dans un mode de réalisation où une approche multi-échelle est utilisée. Dans un tel mode de réalisation, des imagerie de différentes résolutions sont extraites de l'image initiale pour encapsuler les informations de niveau d'échelle (par exemple informations hautes et basses fréquences) et ainsi extraire non seulement les contours de l'image mais aussi la texture d'image. Les différentes imagerie peuvent être ensuite ajustées en taille pour être classifiées à l'étape

304 avant d'être fusionnées à l'étape 306 afin de déterminer la matrice de probabilité finale permettant de construire automatiquement l'annotation sémantique de l'image d'entrée considérée.

[0116] La figure 8 illustre le procédé de détermination d'annotations d'image, appliqué à une
5 séquence d'images, utilisant le classifieur 122 pré-entraîné (bloc 81) à partir d'un modèle de classifieur 82 et d'images annotées 80A pour chaque classe composant la scène étudiée, une séquence d'images d'entrée 80B pour lesquelles une annotation sémantique est à déterminer. Le procédé met en œuvre une approche de classification multi-échelles par fenêtre glissante pondérée (bloc 83) pour établir une matrice de probabilité d'appartenance
10 (bloc 85) de chaque pixel à une classe donnée pour déterminer (bloc 84) l'annotation sémantique de l'image initiale, ce qui permet d'obtenir des images annotées 85.

[0117] Les modes de réalisation de l'invention sont ainsi particulièrement adaptées à toute application utilisant un réseau de neurones de segmentation sémantique et plus particulièrement son réapprentissage à partir de données nouvelles, comme par exemple
15 une application de segmentation des fonds marins. Ils permettent le déploiement d'AUV dans des tels milieux avec un coût d'annotation sémantique réduit. Dans un tel exemple d'application, les modes de réalisation de l'invention peuvent être utilisés pour la détection de risques. En effet, ils permettent de déterminer automatiquement ou dynamiquement des images annotées de manière sémantique sur un nombre prédéterminé de classes
20 représentatives des fonds marins, de faire ressortir la proportion de chaque classe dans une image et d'analyser si une partie de l'image n'appartient pas à ces classes, ce qui indique la probabilité d'être face à un élément inconnu. Selon un autre exemple, le système et le procédé selon les modes de réalisation de l'invention peuvent être utilisés pour calculer la proportion de la représentation de la classe "eau" dans l'image (i.e. la classe représentant
25 l'eau dans les images) afin de déterminer si un AUV (acronyme pour « Autonomous Underwater System » signifiant Système Sous-marin Autonome) est à une profondeur acceptable pour naviguer dans son milieu en utilisant des technologies de type SLAM (Acronyme pour « simultaneous localization and mapping » signifiant 'cartographie et Localisation simultanée') par exemple.

[0118] L'homme du métier comprendra que le système ou les sous-systèmes selon les modes de réalisation de l'invention peuvent être mis en œuvre de diverses manières par matériel (« hardware »), logiciel, ou une combinaison de matériel(s) et de logiciels, notamment sous la forme de code de programme pouvant être distribué sous la forme d'un produit de programme, sous diverses formes. En particulier, le code de programme peut être
30 distribué à l'aide de supports lisibles par ordinateur, qui peuvent inclure des supports de stockage lisibles par ordinateur et des supports de communication. Les procédés décrits
35

dans la présente description peuvent être notamment implémentés sous la forme d'instructions de programme d'ordinateur exécutables par un ou plusieurs processeurs dans un dispositif informatique d'ordinateur. Ces instructions de programme d'ordinateur peuvent également être stockées dans un support lisible par ordinateur.

- 5 [0119] L'invention n'est pas limitée aux modes de réalisation décrits ci-avant à titre d'exemple non limitatif. Elle englobe toutes les variantes de réalisation qui pourront être envisagées par l'homme du métier.

REVENDICATIONS

1. Dispositif de classification d'une scène de l'environnement d'un véhicule, comprenant une unité de classification (14) utilisant un réseau de neurones configuré pour classer une scène en réponse à une image reçue représentant la scène et une unité d'apprentissage
5 (11) configurée pour entraîner ledit réseau de neurones, à partir d'un ensemble d'images stockées dans une base de données (3), caractérisé en ce qu'il comprend une unité d'annotation d'image comprenant :
 - un module de pré-traitement d'image (120) configuré pour extraire un ensemble d'images d'une image reçue en fonction d'un critère de position dans l'image et d'au
10 moins un critère relatif à un paramètre d'image,
 - un module de classification (122) configuré pour déterminer, pour chaque image, un ensemble de probabilités d'appartenance chacune associée à une classe parmi un ensemble de classes prédéfinies, et
 - un module d'annotation d'image (124) configuré pour déterminer une annotation en
15 correspondance avec ladite image à partir des ensembles de probabilités déterminées pour toutes lesdites images, ladite annotation étant transmise à ladite base de données (3) pour un stockage en association avec ladite image dans la base de données.
2. Dispositif de classification selon la revendication 1, dans lequel le module d'annotation d'image (124) est configuré pour fusionner les probabilités d'appartenance
20 déterminées pour toutes les classes et pour toutes les images et pour reconstruire une annotation, pixel par pixel, pour ladite image reçue, ce qui fournit ladite annotation.
3. Dispositif de classification selon l'une des revendications 1 et 2, dans lequel chaque ensemble de probabilités déterminées pour une image est représenté par un vecteur de probabilités ayant un ensemble de composantes correspondant auxdites probabilités
25 d'appartenance, chacune associée à une desdites classes.
4. Dispositif de classification selon l'une des revendications précédentes, dans lequel un paramètre d'image est une information de fréquence d'image, le critère relatif audit paramètre d'image étant un niveau de fréquence.
5. Dispositif de classification selon l'une des revendications précédentes, dans lequel
30 l'unité de prétraitement utilise un pas fixe ou variable pour mettre en œuvre une extraction des images par fenêtre glissante.
6. Dispositif de classification selon l'une des revendications précédentes, dans lequel le module d'annotation d'image (124) est configuré pour sélectionner pour chaque pixel

d'image, la classe ayant une probabilité d'appartenance maximale parmi les probabilités d'appartenance obtenues pour toutes les imagerie.

7. Dispositif de classification selon l'une des revendications précédentes, dans lequel le module de pré-traitement est configuré pour mettre en œuvre, préalablement à la
5 détermination de l'ensemble de probabilités pour une imagerie, une opération de redimensionnement de l'imagerie en fonction du niveau de son paramètre d'image.

8. Dispositif de classification selon l'une des revendications précédentes, dans lequel le module de pré-traitement est configuré pour ajouter, pour chaque imagerie de taille $K \times K$ parmi les P imagerie extraites de l'image d'entrée de taille $M \times N$, les C probabilités
10 d'appartenance obtenues pour l'imagerie dans un patch de taille $K \times K$, correspondant à l'imagerie, sur une matrice de taille $M \times N \times C$ correspondant à l'image d'entrée.

9. Dispositif de classification selon les revendications 3 et 8, dans lequel la matrice de probabilité Mp est définie selon l'équation :

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

15 où P désigne le vecteur de probabilités P en sortie de l'unité de classification et D le vecteur des différentes dimensions d'extraction des imagerie.

10. Dispositif de classification selon les revendications 3 et 8, dans lequel la matrice de probabilité Mp est définie selon l'équation :

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

20 où P désigne le vecteur de probabilités P en sortie de l'unité de classification, $G(i, j)$ une fenêtre Gaussienne et D le vecteur des différentes dimensions d'extraction des imagerie.

11. Dispositif de classification selon la revendication 10, dans lequel la fenêtre gaussienne $G(i, j)$ est définie par :

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

25 l'annotation de l'image reçue étant alors déterminée en sélectionnant, en chaque point de la matrice de probabilité Mp , l'indice ayant la valeur maximale dans le vecteur de probabilités associé à chaque pixel.

12. Procédé de classification d'une scène de l'environnement d'un véhicule, comprenant une étape de classification utilisant un réseau de neurones configuré pour classer une scène
30 en réponse à une image reçue représentant la scène et une étape d'apprentissage

configurée pour entraîner ledit réseau de neurones, à partir d'un ensemble d'images stockées dans une base de données (3), caractérisé en ce que le procédé comprend :

- un pré-traitement de l'image (120) consistant à extraire un ensemble d'imagettes d'une image reçue en fonction d'un critère de position dans l'image et d'au moins un critère relatif à
5 un paramètre d'image,
- une étape de classification (122) consistant à déterminer, pour chaque imagette, un ensemble de probabilités d'appartenance chacune associée à une classe parmi un ensemble de classes prédéfinies, et
- une étape d'annotation d'image (124) consistant à déterminer une annotation en
10 correspondance avec ladite image à partir des ensembles de probabilités déterminées pour toutes lesdites imagettes, ladite annotation étant transmise à ladite base de données (3) pour un stockage en association avec ladite image dans la base de données.