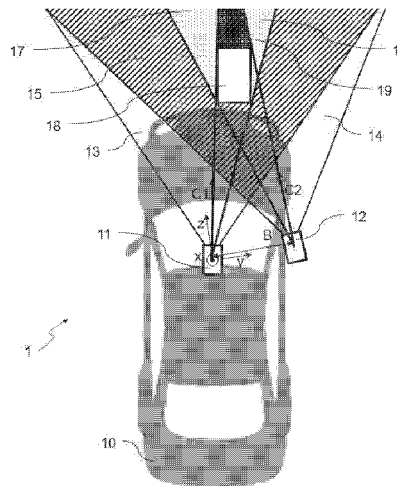


12 DEMANDE DE BREVET D'INVENTION A1

22 Date de dépôt : 09.11.23.	71 Demandeur(s) : STELLANTIS AUTO SAS Société par actions simplifiée (SAS) — FR.
30 Priorité :	72 Inventeur(s) : LI HAI.
43 Date de mise à la disposition du public de la demande : 16.05.25 Bulletin 25/20.	73 Titulaire(s) : STELLANTIS AUTO SAS Société par actions simplifiée.
56 Liste des documents cités dans le rapport de recherche préliminaire : Se reporter à la fin du présent fascicule	54 Mandataires : Procédé ou dispositif d'acquisition de données d'apprentissage d'un système de vision.
60 Références à d'autres documents nationaux apparentés :	57 Procédé ou dispositif d'acquisition d'une base de données d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule, le système de vision comprenant une première caméra (11) et une deuxième caméra (12) disposées de manière à acquérir
Demande(s) d'extension :	

chacune une image d'une scène selon un point de vue différent. En effet, le procédé comprend la réception de données représentatives d'un ensemble de quatre images, acquises par la première et la deuxième caméras à deux instants temporels distincts, la détermination d'une profondeur associée à l'objet courant par le modèle de prédiction de profondeur à partir de boîtes englobantes et l'enregistrement dans la base de donnée d'apprentissage d'un élément courant comprenant des données représentatives de l'objet courant et de la profondeur et les quatre images.

Figure pour l'abrégé : Figure 1



Description

Titre de l'invention : Procédé et dispositif d'acquisition de données d'apprentissage d'un système de vision.

Domaine technique

- [0001] La présente invention concerne les procédés et dispositifs d'acquisition de données d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule, par exemple dans un véhicule automobile. La présente invention concerne également un procédé d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule à partir de données d'apprentissage acquises par le système de vision embarqué.

Arrière-plan technologique

- [0002] De nombreux véhicules modernes sont équipés de systèmes d'aide à la conduite dits ADAS (de l'anglais « Advanced Driver-Assistance System » ou en français « Système d'aide à la conduite avancé »). De tels systèmes ADAS sont des systèmes de sécurité passifs et actifs conçus pour éliminer la part d'erreur humaine dans la conduite de véhicules de tous types. Les ADAS utilisent des technologies avancées pour assister le conducteur pendant la conduite et améliorer ainsi ses performances. Les ADAS utilisent une combinaison de technologies de capteurs pour percevoir l'environnement autour d'un véhicule, puis fournissent des informations au conducteur ou agissent sur certains systèmes du véhicule.
- [0003] Il existe plusieurs niveaux d'ADAS, tels que les caméras de recul et les capteurs d'angle mort, les systèmes d'alerte de franchissement de ligne, les régulateurs de vitesse adaptatifs ou encore les systèmes de stationnement automatique.
- [0004] Les ADAS embarqués dans un véhicule sont alimentés par des données obtenues d'un ou plusieurs capteurs embarqués tels que, par exemple, des caméras. Ces caméras permettent notamment de détecter et de situer d'autres usagers de la route ou d'éventuels obstacles présents autour d'un véhicule afin, par exemple :
- d'adapter l'éclairage du véhicule en fonction de la présence d'autres usagers ;
 - de réguler de façon automatique la vitesse du véhicule ;
 - d'agir sur le système de freinage en cas de risque d'impact avec un objet.
- [0005] Une position d'un autre usager ou d'un obstacle est par exemple déterminée par un système de vision comprenant un modèle de prédiction d'une profondeur ou d'une distance. Un tel modèle est par exemple appris à l'aide d'images, ces images étant obtenues d'une base de données universelle, par exemple Kitti® ou Sceneflow®. Kitti® par exemple fournit des images d'un environnement routier en centre-ville, mais une telle base de données ne comprend cependant pas l'intégralité des envi-

ronnements routiers dans lequel peut évoluer un véhicule. Les données d'apprentissage sont alors inadaptées pour entraîner le modèle de prédiction pour un véhicule circulant dans d'autres environnements routiers.

- [0006] La qualité de l'entraînement du modèle de prédiction est cependant très importante, en effet, les profondeurs ou distances prédites par le modèle de prédiction représentent par exemple des distances auxquelles se situent les autres usagers ou obstacles présents dans l'environnement routier du véhicule équipé du système de vision et des ADAS. De la qualité des données émises par le système de vision dépend donc le bon fonctionnement des périphériques d'aides à la conduite utilisant ces données.

Résumé de la présente invention

- [0007] Un objet de la présente invention est de résoudre au moins l'un des problèmes de l'arrière-plan technologique décrit précédemment.
- [0008] Un autre objet de la présente invention est d'améliorer la pertinence des données d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule
- [0009] Un autre objet de la présente invention est d'améliorer la sécurité routière, notamment par l'amélioration de la sûreté de fonctionnement des système ADAS alimentés par des données obtenues d'un système de vision embarqué dans le véhicule.
- [0010] Selon un premier aspect, la présente invention concerne un procédé d'acquisition d'une base de données d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule, le système de vision comprenant un ensemble de caméras comprenant une première caméra et une deuxième caméra disposées de manière à acquérir chacune une image d'une scène tridimensionnelle selon un point de vue différent, la base de données d'apprentissage comprenant un premier ensemble d'éléments, chaque élément du premier ensemble d'éléments comprenant des premières données représentatives de :
- un type d'objet,
 - une première profondeur, et
 - un premier ensemble de quatre images,
- le procédé étant caractérisé en ce qu'il comprend les étapes suivantes :
- détermination d'une première distribution de profondeurs associée à un type d'objet cible à partir des premières profondeurs d'un deuxième ensemble d'éléments appartenant au premier ensemble d'éléments de la base de données d'apprentissage, chaque type d'objet d'un élément du deuxième ensemble d'éléments correspondant au type d'objet cible ;
 - réception de deuxièmes données représentatives d'un deuxième ensemble de quatre images, le deuxième ensemble de quatre images comprenant :
 - une première et une deuxième images acquises respectivement par la première

caméra et la deuxième caméra à un premier instant temporel,

- une troisième et une quatrième images acquises respectivement par la première caméra et la deuxième caméra à un deuxième instant temporel différent du premier instant temporel ;

- détection d'un objet courant, dans chaque image du deuxième ensemble de quatre images, un type de l'objet courant correspondant au type d'objet cible ;

- détermination, dans chaque image du deuxième ensemble de quatre images, d'une boîte englobante comprenant des pixels associés à l'objet courant ;

- détermination d'une deuxième profondeur associée à l'objet courant par le modèle de prédiction de profondeur à partir des boîtes englobantes ;

- ajout de la deuxième profondeur à la première distribution de profondeurs pour générer une deuxième distribution de profondeurs ; et

- enregistrement dans la base de donnée d'apprentissage d'un élément courant comprenant des troisièmes données représentatives de l'objet courant et de la deuxième profondeur et les deuxièmes données, l'enregistrement étant fonction de la deuxième distribution de profondeurs.

[0011] Un tel procédé permet ainsi d'obtenir une base de données adaptée à l'environnement routier rencontré par le véhicule. En effet, les types d'objet et les images associées à ces objets sont acquises par le système de vision d'un véhicule et sont donc représentatives de scènes réelles se déroulant autour d'un véhicule. L'enregistrement étant fonction de la deuxième distribution de profondeurs, la base de données est alimentée de manière sélective afin de limiter sa taille tout en gardant des éléments pertinents.

[0012] Selon une variante du procédé, chaque élément de la base de données d'apprentissage comprenant en outre des données représentatives d'un type d'environnement routier, le procédé comprend en outre une étape de détermination d'un type d'environnement routier courant associé à l'élément courant à partir des deuxièmes données, des données représentatives du type d'environnement courant étant enregistrées dans la base de données d'apprentissage.

[0013] Les différents éléments de la base de données peuvent ainsi être rangés par type d'environnement routier.

[0014] Selon encore une variante du procédé, le modèle de prédiction de profondeur est entraîné à partir des deuxièmes données lorsque le type d'objet courant correspond au type d'objet cible.

[0015] Le modèle de prédiction de profondeur est ainsi entraîné de manière à améliorer la précision de prédiction de profondeur pour ce type d'objet cible.

[0016] Selon une variante supplémentaire du procédé, la deuxième profondeur est égale à une profondeur moyenne d'un ensemble de pixels de la boîte englobante de la première image.

- [0017] Selon encore une variante du procédé, l'élément courant est enregistré lorsque l'objet courant est entièrement visible dans chacune des images du deuxième ensemble de quatre images.
- [0018] Selon une autre variante du procédé, l'élément courant est enregistré si un nombre total d'éléments du deuxième ensemble d'éléments est inférieur à un seuil total.
- [0019] Selon encore une autre variante du procédé, l'élément courant est enregistré si un nombre d'éléments d'une partie du deuxième ensemble d'éléments comprenant une première profondeur appartenant à une même classe de profondeur que la deuxième profondeur est compris entre un seuil minimal et un seuil maximal, les seuils minimal et maximal étant définis en fonction d'un nombre total d'éléments du deuxième ensemble d'éléments et d'un nombre total de classes de profondeur.
- [0020] Selon une variante supplémentaire du procédé, le seuil maximal est déterminé par la fonction suivante :
- $$S_{\max} = 4 \times N_T / (3 \times I_T)$$
- avec :
- $S_{\max}$ le seuil maximal,
 - N_T le nombre total d'éléments du deuxième ensemble d'éléments, et
 - I_T le nombre d'intervalles.
- [0021] Selon un deuxième aspect, la présente invention concerne un dispositif d'acquisition d'une base de données d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule, le dispositif comprenant une mémoire associée à au moins un processeur configuré pour la mise en œuvre des étapes du procédé selon le premier aspect de la présente invention.
- [0022] Selon un troisième aspect, la présente invention concerne un véhicule, par exemple de type automobile, comprenant un dispositif tel que décrit ci-dessus selon le deuxième aspect de la présente invention.
- [0023] Selon un quatrième aspect, la présente invention concerne un programme d'ordinateur qui comporte des instructions adaptées pour l'exécution des étapes du procédé selon le premier aspect de la présente invention, ceci notamment lorsque le programme d'ordinateur est exécuté par au moins un processeur.
- [0024] Un tel programme d'ordinateur peut utiliser n'importe quel langage de programmation et être sous la forme d'un code source, d'un code objet, ou d'un code intermédiaire entre un code source et un code objet, tel que dans une forme partiellement compilée, ou dans n'importe quelle autre forme souhaitable.
- [0025] Selon un cinquième aspect, la présente invention concerne un support d'enregistrement lisible par un ordinateur sur lequel est enregistré un programme d'ordinateur comprenant des instructions pour l'exécution des étapes du procédé selon le premier aspect de la présente invention.

- [0026] D'une part, le support d'enregistrement peut être n'importe quel entité ou dispositif capable de stocker le programme. Par exemple, le support peut comporter un moyen de stockage, tel qu'une mémoire ROM, un CD-ROM ou une mémoire ROM de type circuit microélectronique, ou encore un moyen d'enregistrement magnétique ou un disque dur.
- [0027] D'autre part, ce support d'enregistrement peut également être un support transmissible tel qu'un signal électrique ou optique, un tel signal pouvant être acheminé via un câble électrique ou optique, par radio classique ou hertzienne ou par faisceau laser autodirigé ou par d'autres moyens. Le programme d'ordinateur selon la présente invention peut être en particulier téléchargé sur un réseau de type Internet.
- [0028] Alternativement, le support d'enregistrement peut être un circuit intégré dans lequel le programme d'ordinateur est incorporé, le circuit intégré étant adapté pour exécuter ou pour être utilisé dans l'exécution du procédé en question.

Brève description des figures

- [0029] D'autres caractéristiques et avantages de la présente invention ressortiront de la description des exemples de réalisation particuliers et non limitatifs de la présente invention ci-après, en référence aux figures 1 à 4 annexées, sur lesquelles :
- [0030] [Fig.1] illustre schématiquement un système de vision équipant un véhicule, selon un exemple de réalisation particulier et non limitatif de la présente invention ;
- [0031] [Fig.2] illustre un organigramme des différentes étapes d'un procédé d'acquisition d'une base de données d'apprentissage d'un modèle de vision embarqué dans le véhicule de la [Fig.1], selon un exemple de réalisation particulier et non limitatif de la présente invention ;
- [0032] [Fig.3] illustre un diagramme représentant la répartition statistique d'un nombre d'éléments de la base de données d'apprentissage par classe de profondeur, selon un exemple de réalisation particulier et non limitatif de la présente invention ;
- [0033] [Fig.4] illustre schématiquement un dispositif configuré pour l'acquisition d'une base de données d'apprentissage d'un modèle de vision embarqué dans le véhicule de la [Fig.1], selon un exemple de réalisation particulier et non limitatif de la présente invention.

Description des exemples de réalisation

- [0034] Un procédé et un dispositif d'acquisition d'une base de données d'apprentissage d'un modèle de vision embarqué dans un véhicule vont maintenant être décrits dans ce qui va suivre en référence conjointement aux figures 1 à 4. Des mêmes éléments sont identifiés avec des mêmes signes de référence tout au long de la description qui va suivre.
- [0035] Les termes « premier(s) », « deuxième(s) » (ou « première(s) », « deuxième(s) »),

etc. sont utilisés dans ce document par convention arbitraire pour permettre d'identifier et de distinguer différents éléments (tels que des opérations, des moyens, etc.) mis en œuvre dans les modes de réalisation décrits ci-après. De tels éléments peuvent être distincts ou correspondre à un seul et unique élément, selon le mode de réalisation.

- [0036] Selon un exemple particulier et non limitatif de réalisation de la présente invention, un procédé d'acquisition d'une base de données d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule est par exemple mis en œuvre par un calculateur du système embarqué du véhicule contrôlant ce système de vision.
- [0037] Le système de vision comprend un ensemble de caméras d'au moins deux caméras comprenant une première caméra et une deuxième caméra disposées de manière à acquérir chacune une image d'une scène tridimensionnelle selon un point de vue différent.
- [0038] La base de données d'apprentissage comprend quant à elle un premier ensemble d'éléments, chaque élément du premier ensemble d'éléments comprenant des premières données représentatives de :
- un type d'objet,
 - une première profondeur, et
 - un premier ensemble de quatre images.
- [0039] En effet, le procédé comprend la détermination d'une première distribution de profondeurs associée à un type d'objet cible à partir des premières profondeurs d'un deuxième ensemble d'éléments appartenant au premier ensemble d'éléments de la base de données d'apprentissage, chaque type d'objet d'un élément du deuxième ensemble d'éléments correspondant au type d'objet cible.
- [0040] Le procédé comprend également la réception de deuxièmes données représentatives d'un deuxième ensemble de quatre images, acquises par la première et la deuxième caméras à deux instants temporels distincts et la détection d'un objet courant dans chaque image du deuxième ensemble de quatre images, un type de l'objet courant correspondant au type d'objet cible.
- [0041] Une boîte englobante comprenant des pixels associés à l'objet courant est alors déterminée dans chaque image du deuxième ensemble de quatre images et une deuxième profondeur associée à l'objet courant est déterminée par le modèle de prédiction de profondeur à partir des boîtes englobantes.
- [0042] La deuxième profondeur est ajoutée à la première distribution de profondeurs pour générer une deuxième distribution de profondeurs et un élément courant comprenant des troisièmes données représentatives de l'objet courant et de la deuxième profondeur et les deuxièmes données est enregistré dans la base de données d'apprentissage, l'enregistrement étant fonction de la deuxième distribution de profondeurs.

- [0043] Un tel procédé permet ainsi d'obtenir une base de données adaptée à l'environnement routier et aux objets rencontrés par le véhicule. En effet, les types d'objet et les images associées à ces objets sont acquises par le système de vision d'un véhicule et sont donc représentatives de scènes réelles se déroulant autour d'un véhicule et telles que perçues par un tel système de vision embarqué.
- [0044] L'enregistrement étant fonction de la deuxième distribution de profondeurs, la base de données est alimentée de manière sélective afin de limiter sa taille tout en gardant des éléments pertinents, les éléments pertinents permettant de disposer davantage de données d'apprentissage représentatives d'un ensemble de scènes tridimensionnelles.
- [0045] La [Fig.1] illustre schématiquement un système de vision équipant un véhicule, selon un exemple de réalisation particulier et non limitatif de la présente invention.
- [0046] Un environnement 1 correspond, par exemple, à un environnement routier formé d'un réseau de routes accessibles pour le véhicule 10.
- [0047] Dans cet exemple, le véhicule 10 correspond à un véhicule à moteur thermique, à moteur(s) électrique(s) ou encore un véhicule hybride avec un moteur thermique et un ou plusieurs moteurs électriques. Le véhicule 10 correspond ainsi, par exemple, à un véhicule terrestre tel qu'une automobile, un camion, un car, une moto. Enfin, le véhicule 10 correspond à un véhicule autonome ou non, c'est-à-dire un véhicule circulant selon un niveau d'autonomie déterminé ou sous la supervision totale du conducteur.
- [0048] Le véhicule 10 comprend avantageusement un ensemble de caméras comprenant une première caméra 11 et une deuxième caméra 12 embarquées, chacune configurée pour acquérir des images d'une scène tridimensionnelle dans l'environnement 1 du véhicule 10. Cet ensemble de caméras forme le système de vision. Deux caméras sont illustrées sur la [Fig.1]. La présente invention ne se limite cependant pas à un système de vision comprenant deux caméras mais s'étend à tout système de vision comprenant 2 ou plus caméras, par exemple 2, 3, 4 ou 5 caméras.
- [0049] Les première 11 et deuxième 12 caméras disposent de paramètres intrinsèques connus. Ces paramètres se composent notamment de :
- la distance focale f_1 de la première caméra 11 ;
 - la distance focale f_2 de la deuxième caméra 12 ;
 - les distorsions qui sont dues aux imperfections du système optique de chaque caméra ;
 - la direction C1 de l'axe optique de la première caméra 11 ;
 - la direction C2 de l'axe optique de la deuxième caméra 12 ; et
 - les résolutions respectives des caméras 11, 12.
- [0050] Les paramètres intrinsèques caractérisent la transformation qui associe, pour un point image, les coordonnées caméra aux coordonnées pixel, dans chaque caméra. Ces paramètres ne changent pas si l'on déplace la caméra.

- [0051] Les distorsions, qui sont dues aux imperfections du système optique telles que des défauts de forme et de positionnement des lentilles des caméras, vont dévier les faisceaux lumineux et donc induire un écart de positionnement pour le point projeté par rapport à un modèle idéal. Il est possible alors de compléter le modèle de caméra en y introduisant les trois distorsions qui génèrent le plus d'effets, à savoir les distorsions radiales, de décentrage et prismatiques, induites par des défauts de courbure, de parallélisme des lentilles et de coaxialité des axes optiques. Dans cet exemple, les caméras sont supposées parfaites, c'est-à-dire que les distorsions ne sont pas prises en compte ou que leur correction est traitée au moment de l'acquisition d'une image.
- [0052] Ces première 11 et deuxième 12 caméras sont disposées de manière à acquérir chacune une image d'une scène tridimensionnelle selon un point de vue différent, le premier point de vue est par exemple localisé sur ou dans le rétroviseur gauche du véhicule 10 ou en haut du pare-brise du véhicule 10, le deuxième point de vue est par exemple localisé sur ou dans le rétroviseur droit du véhicule 10 ou en haut du pare-brise du véhicule 10. Dans le cas où les deux caméras sont situées en haut du pare-brise du véhicule, celles-ci sont alors placées à une certaine distance. Dans cet exemple, la première caméra 11 est située en haut du pare-brise du véhicule 10, la deuxième caméra 12 est située dans le rétroviseur droit du véhicule 10.
- [0053] Un premier repère est associé à la première caméra 11 :
- la direction de l'axe y est définie par la position de la deuxième caméra 12, de manière à placer la deuxième caméra 12 sur l'axe y de la première caméra 11. La distance B séparant les deux caméras 11, 12 est appelée base de référence (en anglais « baseline ») et la direction séparant les deux caméras 11, 12 est celle de l'axe y ;
 - la direction de l'axe x est définie orthogonale à celle de l'axe y et orthogonale à celle de l'axe optique C1 de la première caméra 11 ;
 - la direction de l'axe z est définie orthogonale aux directions des axes x et y.
- Les trois axes x, y et z forment ainsi un repère orthonormé.
- [0054] Les paramètres extrinsèques liés à la position des caméras 11, 12 sont les paramètres suivants :
- 3 translations dans les directions x, y et z : T_x , T_y et T_z constituant le vecteur translation T ; et
 - 3 rotations autour des axes x, y et z : R_x , R_y et R_z , constituant la matrice de rotation R.
- [0055] La détermination des paramètres extrinsèques est effectuée par exemple lors de la calibration du système de vision.
- [0056] On appelle système de vision stéréoscopique un système de vision comprenant une pluralité de caméras, par exemple la première caméra 11 et la deuxième caméra 12. Une contrainte principale d'un système de vision stéréoscopique utilisé dans

l'automobile est, par exemple, la grande distance entre les deux caméras. En effet, pour pouvoir couvrir une plage de mesure de 200 mètres, la base de référence doit atteindre 60cm pour les caméras couramment utilisées dans ce domaine.

- [0057] Les première 11 et deuxième 12 caméras font l'acquisition d'images d'une scène tridimensionnelle se situant devant le véhicule 10, la première caméra 11 couvrant seule un premier champ d'acquisition 13, la deuxième caméra 12 couvrant seule un deuxième champ d'acquisition 14 et les première 11 et deuxième 12 caméras couvrant toutes deux un troisième champ d'acquisition 15. Les premier et troisième champs d'acquisition 13, 15 permettent ainsi une vision monoscopique de la scène tridimensionnelle par la première caméra 11, les deuxième et troisième champs d'acquisition 14, 15 permettent une vision monoscopique de la scène par la deuxième caméra 12 et le troisième champ d'acquisition 15 permet une vision stéréoscopique de la scène par le système de vision stéréoscopique composé des première 11 et deuxième 12 caméras.
- [0058] Un obstacle 18 est placé dans le champ d'acquisition des caméras, par exemple dans le troisième champ d'acquisition 15. La présence de l'obstacle 18 définit un champ d'occlusion pour le système de vision stéréoscopique composé ici des trois champs 16, 17 et 19.
- [0059] Parmi ces trois champs, le champ 16 est visible depuis la deuxième caméra 12. La partie de la scène présente dans ce champ 16 est donc observable à l'aide du système de vision monoscopique composé de la deuxième caméra 12.
- [0060] Le champ 17 est quant à lui visible depuis la première caméra 11. La partie de la scène présente dans ce champ 17 est donc observable à l'aide du système de vision monoscopique composé de la première caméra 11.
- [0061] Enfin, le champ 19 n'est visible depuis aucune des caméras. La partie de la scène présente dans ce champ 19 n'est donc pas observable.
- [0062] Lorsque le véhicule 10 est en mouvement, chacune des première 11 et deuxième 12 caméras forme un système de vision monoscopique.
- [0063] Selon un exemple de réalisation, les directions C1, C2 des axes optiques représentatifs d'une orientation du champ de vision de chaque caméra sont orientées non parallèlement de manière à obtenir le troisième champ d'acquisition 15 de l'environnement 1 le plus large possible.
- [0064] Selon un autre exemple de réalisation, les directions C1, C2 des axes optiques représentatifs d'une orientation du champ de vision de chaque caméra sont orientées parallèlement.
- [0065] Il est évident qu'il est possible d'utiliser un tel système de vision, stéréoscopique ou monoscopique, pour prendre des images de scènes situées sur les côtés ou derrière le véhicule 10 en l'équipant de caméras placées et orientées différemment.
- [0066] Les images acquises par les première 11 et deuxième 12 caméras à un instant

temporel d'acquisition donné se présentent sous la forme de données représentant des pixels caractérisés par :

- des coordonnées dans chaque image ; et
- des données relatives aux couleurs et luminosité des objets de la scène observée sous forme par exemple de coordonnées colorimétriques RGB (de l'anglais « Red Green Blue », en français « Rouge Vert Bleu ») ou TSL (Ton, Saturation, Luminosité).

[0067] Les images acquises par les première 11 et deuxième 12 caméras à un même instant temporel représentent des vues d'une même scène prises à des points de vue différents, les position des caméras étant distinctes. Sur cette scène se trouvent par exemple des objets, chaque objet appartenant à un type d'objet tel que :

- des bâtiments ;
- des infrastructures routières ;
- d'autres usagers immobiles, par exemple un véhicule stationné ; et/ou
- d'autres usagers mobiles, par exemple un autre véhicule, un cycliste ou un piéton en mouvement.

[0068] Ces images sont envoyées à un calculateur d'un dispositif équipant le véhicule 10 ou stockées dans une mémoire d'un dispositif accessible à un calculateur d'un dispositif équipant le véhicule 10.

[0069] Un processus d'acquisition d'une base de données d'apprentissage d'un modèle de prédiction de profondeur pour le système de vision embarqué dans le véhicule 10 est avantageusement mis en œuvre par le véhicule 10, c'est-à-dire par un calculateur ou une combinaison de calculateurs du système embarqué du véhicule 10, par exemple par le ou les calculateurs en charge du système de vision du véhicule 10.

[0070] La base de données d'apprentissage comprend un premier ensemble d'éléments, chaque élément du premier ensemble d'éléments comprenant des premières données représentatives de :

- un type d'objet,
- une première profondeur, et
- un premier ensemble de quatre images.

[0071] Il est à noter que le procédé fonctionne y compris si la base de données d'apprentissage ne comprend pas encore d'élément, en effet le but de ce procédé est de compléter une base de données d'apprentissage existante ou de constituer une nouvelle base de données d'apprentissage.

[0072] Une première profondeur est par exemple une distance, exprimée en mètres, séparant le système de vision ayant fait l'acquisition d'un premier ensemble d'images d'un objet.

[0073] Un premier ensemble de quatre images comprend par exemple deux couples d'images acquises par deux caméras distinctes à deux instants temporels distincts. Par

exemple par la première caméra 11 et par la deuxième caméra 12 du système de vision embarqué dans le véhicule 10, ou par un autre couple de caméras appartenant à un système de vision similaire. En effet, selon un premier mode de réalisation particulier, la base de données est intégralement constituée d'éléments comprenant des données acquises par le système de vision du véhicule 10, tandis que selon un deuxième exemple de réalisation particulier, la base de données est communautaire ou provient d'un serveur distant et est constituée d'éléments comprenant des données acquises par différents systèmes de vision similaires embarqués sur plusieurs véhicules. Les systèmes de vision similaires embarqués dans une pluralité de véhicules sont par exemple des systèmes de vision comprenant des caméras de même spécifications techniques et disposées à des mêmes endroits dans la pluralité de véhicules.

[0074] Selon un exemple de réalisation particulier, chaque élément de la base de données d'apprentissage comprend en outre des données représentatives d'un type d'environnement routier. Un type d'environnement routier appartient par exemple à un ensemble de types d'environnement routiers comprenant :

- un parc de stationnement,
- un centre-ville en agglomération,
- une agglomération (hors centre-ville),
- une départementale ou une nationale en milieu rural,
- une route à accès réglementé, et
- une autoroute.

[0075] Dans une première opération, une première distribution de profondeurs associée à un type d'objet cible est déterminée à partir des premières profondeurs d'un deuxième ensemble d'éléments appartenant au premier ensemble d'éléments de la base de données d'apprentissage, chaque type d'objet d'un élément du deuxième ensemble d'éléments correspondant au type d'objet cible.

[0076] En d'autres termes, la première opération consiste à sélectionner dans la base de données d'apprentissage les éléments du premier ensemble d'éléments pour lesquels le type d'objet correspond à un type d'objet cible. Un tel type d'objet cible est par exemple déterminé par un utilisateur ou par un programme déterminant un type d'objet cible en fonction d'un besoin, par exemple afin d'acquérir des éléments supplémentaires correspondants à ce type d'objet cible. Une fois ces éléments sélectionnés, ceux-ci forment le deuxième ensemble d'éléments. De ce deuxième ensemble d'éléments sont extraites les premières profondeurs associées à chacun de ces éléments, et une première distribution de profondeurs est déterminée.

[0077] Selon un exemple de réalisation, la première distribution de profondeurs est déterminée par classes de profondeurs. Les classes de profondeur sont par exemple de largeur égale ou de largeur différente, progressive en fonction de la profondeur

moyenne de chaque.

[0078] La [Fig.3] illustre un exemple de première distribution de profondeurs par classes sous la forme d'un histogramme 3. Chaque barre 31, 32, 33, 34, 35 de l'histogramme 3 représente le ratio entre le nombre d'éléments du deuxième ensemble d'éléments dont la première longueur est comprise dans une classe et le nombre d'éléments du deuxième ensemble d'éléments. Par exemple :

- la première classe représentée par une première barre 31 comprend l'ensemble des éléments du deuxième ensemble d'éléments dont la première profondeur est comprise entre 0 et D1,
- la deuxième classe représentée par une deuxième barre 32 comprend l'ensemble des éléments du deuxième ensemble d'éléments dont la première profondeur est comprise entre D1 et D2,
- la troisième classe représentée par une troisième barre 33 comprend l'ensemble des éléments du deuxième ensemble d'éléments dont la première profondeur est comprise entre D2 et D3,
- la quatrième classe représentée par une quatrième barre 34 comprend l'ensemble des éléments du deuxième ensemble d'éléments dont la première profondeur est comprise entre D3 et D4, et
- la cinquième classe représentée par une cinquième barre 35 comprend l'ensemble des éléments du deuxième ensemble d'éléments dont la première profondeur est supérieure à D4.

[0079] L'axe des abscisses représente la profondeur D (en anglais « Depth »), par exemple exprimée en mètres (m) tandis que l'axe des ordonnées représente le ratio ou la probabilité P que représente chaque classe. Le ratio est par exemple exprimé en pourcentage (%).

[0080] Selon un exemple de réalisation particulier, chaque classe a un ratio compris entre un seuil minimal $S_{\min}$ et un seuil maximal $S_{\max}$. De telles limites permettent ainsi de s'assurer que chaque classe est représentée de manière homogène. Ainsi, pour chaque plage de profondeur, un nombre similaire d'éléments correspondants au type d'objet cible est enregistré dans la base de données d'apprentissage.

[0081] Les seuils minimal et maximal sont par exemple définis en fonction d'un nombre total d'éléments du deuxième ensemble d'éléments et d'un nombre total de classes de profondeur, par exemple le seuil minimal est déterminé par la fonction suivante :

$$S_{\min} = 2 \times N_T / (3 \times I_T)$$

Avec :

- $S_{\min}$ le seuil minimal,
- N_T le nombre total d'éléments du deuxième ensemble d'éléments, et
- I_T le nombre d'intervalles,

et le seuil maximal $S_{\max}$ est déterminé par la fonction suivante :

$$S_{\max} = 2 \times S_{\min}.$$

- [0082] Dans une deuxième opération, des deuxièmes données représentatives d'un deuxième ensemble de quatre images sont reçues. Le deuxième ensemble de quatre images comprend :
- une première et une deuxième images acquises respectivement par la première caméra 11 et la deuxième caméra 12 à un premier instant temporel,
 - une troisième et une quatrième images acquises respectivement par la première caméra 11 et la deuxième caméra 12 à un deuxième instant temporel différent du premier instant temporel.
- [0083] Les images acquises respectivement par la première caméra 11 et la deuxième caméra 12 à un premier instant temporel permettent par exemple de déterminer une profondeur d'un objet présent dans ces images par un système de vision stéréoscopique comprenant les première 11 et deuxième 12 caméras.
- [0084] Les images acquises par la première caméra 11 aux deux instants temporels distincts permettent par exemple de déterminer une profondeur d'un objet présent dans ces images par un système de vision monoscopique comprenant la première caméra 11.
- [0085] Les images acquises par la deuxième caméra 12 aux deux instants temporels distincts permettent par exemple de déterminer une profondeur d'un objet présent dans ces images par un système de vision monoscopique comprenant la deuxième caméra 12.
- [0086] Dans une troisième opération, un objet courant est détecté dans chaque image du deuxième ensemble de quatre images, un type de l'objet courant correspondant au type d'objet cible et dans une quatrième opération, une boîte englobante comprenant des pixels associés à l'objet courant est déterminée dans chaque image du deuxième ensemble de quatre images.
- [0087] La détection d'un objet dans une image et la détermination d'une boîte englobante sont connues de l'homme du métier, ces opérations étant par exemple réalisées à l'aide d'algorithmes tels que Yolo7®, centernet® ou vfnet®.
- [0088] Selon un exemple de réalisation particulier, les opérations suivantes ne sont mises en œuvre que si l'objet courant est entièrement visible dans chacune des images du deuxième ensemble de quatre images. On entend ici par entièrement visible le fait qu'aucun autre objet ne vienne masquer partiellement l'objet courant, c'est-à-dire que l'objet courant ne soit pas occlus par un autre objet, ou par le fait qu'aucun partir de l'objet ne soit située en dehors du champ de vision de chacune des caméras aux deux instants temporels d'acquisition des images du deuxième ensemble de quatre images. Cela permet de ne sauvegarder par la suite que des éléments courants comprenant le deuxième ensemble de quatre images dans lequel l'objet courant est entièrement visible.

- [0089] Dans une cinquième opération, une deuxième profondeur associée à l'objet courant est déterminée par le modèle de prédiction de profondeur à partir des boîtes englobantes. Le modèle de prédiction de profondeur est par exemple associé au système de vision stéréoscopique comprenant les première 11 et deuxième 12 caméras et connu de l'homme du métier, par exemple en utilisant le modèle CoEx®, ou au système de vision monoscopique comprenant la première caméra 11 en mouvement ou la deuxième caméra 12 en mouvement et connu de l'homme du métier, par exemple en utilisant le modèle HR-depth®. Cette opération consiste premièrement en la détermination de profondeurs des pixels inclus dans les boîtes englobantes.
- [0090] Selon un exemple de réalisation particulier, les profondeurs sont prédites pour des pixels inclus dans des boîtes englobantes réduites, les boîtes englobantes réduites étant comprises dans les boîtes englobantes précédemment déterminées auxquelles sont soustraites des pixels périphériques, les boîtes englobantes réduites étant par exemple d'une largeur et d'une hauteur réduite de 25% par rapport aux boîtes englobantes précédemment déterminées et centrées sur ces mêmes boîtes englobantes. Cette réduction des boîtes englobantes permet ainsi de ne pas prendre en compte des pixels d'arrière-plan lors de la détermination de profondeurs et qui fausseraient la détermination de la profondeur de l'objet courant.
- [0091] Deuxièmement, la deuxième profondeur est déterminée comme égale à une profondeur moyenne d'un ensemble de pixels de la boîte englobante de la première image.
- [0092] Dans une sixième opération, la deuxième profondeur est ajoutée à la première distribution de profondeurs pour générer une deuxième distribution de profondeurs.
- [0093] Selon l'exemple de réalisation particulier précédemment présenté dans lequel chaque élément de la base de données d'apprentissage comprend en outre des données représentatives d'un type d'environnement routier, un type d'environnement routier courant associé à l'élément courant est déterminé dans une septième opération à partir des deuxièmes données.
- [0094] Dans une huitième opération, un élément courant comprenant des troisièmes données représentatives de l'objet courant et de la deuxième profondeur et les deuxièmes données est enregistré dans la base de données d'apprentissage en fonction de la deuxième distribution de profondeurs.
- [0095] Selon l'exemple de réalisation particulier précédemment présenté dans lequel chaque élément de la base de données d'apprentissage comprend en outre des données représentatives d'un type d'environnement routier, des données représentatives du type d'environnement routier courant sont également enregistrées dans la base de données d'apprentissage.
- [0096] Selon un exemple de réalisation particulier, l'élément courant est enregistré si un

nombre total d'éléments du deuxième ensemble d'éléments est inférieur à un seuil total. Une telle condition permet ainsi d'éviter de surcharger ou saturer la base de données et évite d'avoir recours à une trop grande capacité mémoire pour stocker la base de données d'apprentissage.

- [0097] Selon encore un exemple de réalisation particulier, l'élément courant est enregistré si un nombre d'éléments d'une partie du deuxième ensemble d'éléments comprenant une première profondeur appartenant à une même classe de profondeur que la deuxième profondeur est comprise entre le seuil minimal et le seuil maximal précédemment définis. Cette condition permet de n'enregistrer dans la base de données que des éléments dont la deuxième profondeur est incluse dans une classe de profondeurs n'étant pas saturées en nombre d'éléments. Ainsi, l'élément courant n'est sauvegardé que s'il correspond à une classe de profondeurs ne comprenant pas le nombre d'éléments maximal, correspondant au seuil maximal $S_{\max}$.
- [0098] Il est à noter que chaque premier ensemble de quatre images et le deuxième ensemble de quatre images comprennent quatre images. Cependant l'invention ne se limite pas à un système de vision comprenant deux caméras et s'étend à tout système de vision comprenant une pluralité de caméras. Les premier et deuxième ensembles de quatre images peuvent alors être des ensembles de six, huit ou plus images, le nombre d'images étant par exemple le double du nombre de caméras que comprend le système de vision embarqué dans le véhicule. Chaque élément de la base de données comprend alors des premières données représentatives d'un ensemble d'une pluralité d'images et les différentes opérations du processus sont alors mises en œuvre sur cette pluralité d'images.
- [0099] Selon un exemple de réalisation particulier, le modèle de prédiction de profondeur est entraîné à partir des deuxièmes données lorsque le type d'objet courant correspond au type d'objet cible.
- [0100] Selon encore un exemple de réalisation particulier, le modèle de prédiction de profondeur est entraîné à partir de la base de données acquise et comprenant l'élément courant. L'entraînement d'un tel modèle de prédiction de profondeur est notamment connu de l'homme du métier. En effet, de nombreux modèles de prédiction de profondeur sont mis en œuvre par des réseaux de neurones convolutifs. Le réseau de neurones convolutif est alors entraîné, c'est-à-dire que ses paramètres d'entrée sont ajustés, par exemple par minimisation de fonctions de perte, par exemple déterminées à partir d'erreurs de reconstruction. L'entraînement d'un tel modèle de prédiction de profondeur à partir des de la base de données d'apprentissage est également appelé auto-supervision et est connu de l'homme du métier, par exemple à travers l'utilisation d'algorithmes tels que UnOs® pour un modèle de prédiction associé à un système de vision stéréoscopique ou monodepth2® pour un modèle de prédiction associé à un

système de vision monoscopique.

- [0101] Ces données d'apprentissage sont ainsi utilisées pour ajuster les paramètres du modèle de prédiction de profondeur associé au système de vision embarqué dans le véhicule, fiabilisant les profondeurs prédites et augmentant la robustesse du modèle de prédiction de profondeur.
- [0102] Si un ADAS utilise des profondeurs prédites par ce modèle de prédiction comme donnée d'entrée pour déterminer la distance entre une partie du véhicule 10, par exemple le pare-chocs avant, et un autre usager présent sur la route, l'ADAS est alors en mesure de déterminer précisément cette distance. Par exemple, si l'ADAS a pour fonction d'agir sur un système de freinage du véhicule 10 en cas de risque de collision avec un autre usager de la route et que la distance séparant le véhicule 10 de ce même usager de la route diminue fortement, alors l'ADAS est en mesure de détecter ce rapprochement soudain et d'agir sur le système de freinage du véhicule 10 pour éviter un éventuel accident, y compris si l'autre usager n'est vu que par une seule caméra. La sécurité des usagers est ainsi améliorée.
- [0103] Cette base de données peut en outre être partagée, par exemple en l'enregistrant sur un serveur distant accessible à la pluralité de véhicules embarquant systèmes de vision similaires. Ces derniers bénéficient alors d'une base de données d'apprentissage pertinente permettant d'entraîner le modèle de prédiction de profondeur qu'ils embarquent.
- [0104] La [Fig.2] illustre un organigramme des différentes étapes d'un procédé 2 d'acquisition d'une base de données d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule comprenant une première caméra 11 et une deuxième caméra 12, par exemple dans le véhicule 10 de la [Fig.1], selon un exemple de réalisation particulier et non limitatif de la présente invention. Le procédé 2 est par exemple mis en œuvre par un dispositif embarqué dans le véhicule 10 ou par le dispositif 4 de la [Fig.4].
- [0105] La base de données d'apprentissage comprend notamment un premier ensemble d'éléments, chaque élément du premier ensemble d'éléments comprenant des premières données représentatives de :
- un type d'objet,
 - une première profondeur, et
 - un premier ensemble de quatre images.
- [0106] Dans une étape 21, une première distribution de profondeurs associée à un type d'objet cible est déterminée à partir des premières profondeurs d'un deuxième ensemble d'éléments appartenant au premier ensemble d'éléments de la base de données d'apprentissage, chaque type d'objet d'un élément du deuxième ensemble d'éléments correspondant au type d'objet cible.

- [0107] Dans une étape 22, des deuxièmes données représentatives d'un deuxième ensemble de quatre images sont reçues, le deuxième ensemble de quatre images comprenant :
- une première et une deuxième images acquises respectivement par la première caméra 11 et la deuxième caméra 12 à un premier instant temporel,
 - une troisième et une quatrième images acquises respectivement par la première caméra 11 et la deuxième caméra 12 à un deuxième instant temporel différent du premier instant temporel.
- [0108] Dans une étape 23, un objet courant est détecté dans chaque image du deuxième ensemble de quatre images, un type de l'objet courant correspondant au type d'objet cible.
- [0109] Dans une étape 24, une boîte englobante comprenant des pixels associés à l'objet courant est déterminée dans chaque image du deuxième ensemble de quatre images.
- [0110] Dans une étape 25, une deuxième profondeur associée à l'objet courant est déterminée par le modèle de prédiction de profondeur à partir des boîtes englobantes.
- [0111] Dans une étape 26, la deuxième profondeur est ajoutée à la première distribution de profondeurs pour générer une deuxième distribution de profondeurs.
- [0112] Dans une étape 27, un élément courant comprenant des troisièmes données représentatives de l'objet courant et de la deuxième profondeur et les deuxièmes données est enregistré dans la base de données d'apprentissage, l'enregistrement étant fonction de la deuxième distribution de profondeurs
- [0113] Selon une variante, les variantes et exemples des opérations décrits en relation avec les figures 1 et 3 s'appliquent aux étapes du procédé 2 de la [Fig.2].
- [0114] La [Fig.4] illustre schématiquement un dispositif 4 configuré pour l'acquisition d'une base de données d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule, par exemple dans le véhicule 10 de la première figure, selon un exemple de réalisation particulier et non limitatif de la présente invention. Le dispositif 4 correspond par exemple à un dispositif embarqué dans le premier véhicule 10, par exemple un ordinateur.
- [0115] Le dispositif 4 est par exemple configuré pour la mise en œuvre des opérations décrites en regard des figures 1 et 3 et/ou étapes décrites en regard de la [Fig.2]. Des exemples d'un tel dispositif 4 comprennent, sans y être limités, un équipement électronique embarqué tel qu'un ordinateur de bord d'un véhicule, un ordinateur électronique tel qu'une UCE (« Unité de Commande Electronique »), un téléphone intelligent, une tablette, un ordinateur portable. Les éléments du dispositif 4, individuellement ou en combinaison, peuvent être intégrés dans un unique circuit intégré, dans plusieurs circuits intégrés, et/ou dans des composants discrets. Le dispositif 4 peut être réalisé sous la forme de circuits électroniques ou de modules logiciels (ou informatiques) ou encore d'une combinaison de circuits électroniques et de modules

logiciels.

- [0116] Le dispositif 4 comprend un (ou plusieurs) processeur(s) 40 configurés pour exécuter des instructions pour la réalisation des étapes du procédé et/ou pour l'exécution des instructions du ou des logiciels embarqués dans le dispositif 4. Le processeur 40 peut inclure de la mémoire intégrée, une interface d'entrée/sortie, et différents circuits connus de l'homme du métier. Le dispositif 4 comprend en outre au moins une mémoire 41 correspondant par exemple à une mémoire volatile et/ou non volatile et/ou comprend un dispositif de stockage mémoire qui peut comprendre de la mémoire volatile et/ou non volatile, telle que EEPROM, ROM, PROM, RAM, DRAM, SRAM, flash, disque magnétique ou optique.
- [0117] Le code informatique du ou des logiciels embarqués comprenant les instructions à charger et exécuter par le processeur est par exemple stocké sur la mémoire 41.
- [0118] Selon différents exemples de réalisation particuliers et non limitatifs, le dispositif 4 est couplé en communication avec d'autres dispositifs ou systèmes similaires (par exemple d'autres calculateurs) et/ou avec des dispositifs de communication, par exemple une TCU (de l'anglais « Telematic Control Unit » ou en français « Unité de Contrôle Télématique »), par exemple par l'intermédiaire d'un bus de communication ou au travers de ports d'entrée / sortie dédiés.
- [0119] Selon un exemple de réalisation particulier et non limitatif, le dispositif 4 comprend un bloc 42 d'éléments d'interface pour communiquer avec des dispositifs externes. Les éléments d'interface du bloc 42 comprennent une ou plusieurs des interfaces suivantes :
- interface radiofréquence RF, par exemple de type Wi-Fi® (selon IEEE 802.11), par exemple dans les bandes de fréquence à 2,4 ou 5 GHz, ou de type Bluetooth® (selon IEEE 802.15.1), dans la bande de fréquence à 2,4 GHz, ou de type Sigfox utilisant une technologie radio UBN (de l'anglais Ultra Narrow Band, en français bande ultra étroite), ou LoRa dans la bande de fréquence 868 MHz, LTE (de l'anglais « Long-Term Evolution » ou en français « Evolution à long terme »), LTE-Advanced (ou en français LTE-avancé) ;
 - interface USB (de l'anglais « Universal Serial Bus » ou « Bus Universel en Série » en français) ;
 - interface HDMI (de l'anglais « High Definition Multimedia Interface », ou « Interface Multimedia Haute Definition » en français) ;
 - interface LIN (de l'anglais « Local Interconnect Network », ou en français « Réseau interconnecté local »).
- [0120] Selon un autre exemple de réalisation particulier et non limitatif, le dispositif 4 comprend une interface de communication 43 qui permet d'établir une communication avec d'autres dispositifs (tels que d'autres calculateurs du système embarqué) via un

canal de communication 430. L'interface de communication 43 correspond par exemple à un transmetteur configuré pour transmettre et recevoir des informations et/ou des données via le canal de communication 430. L'interface de communication 43 correspond par exemple à un réseau filaire de type CAN (de l'anglais « Controller Area Network » ou en français « Réseau de contrôleurs »), CAN FD (de l'anglais « Controller Area Network Flexible Data-Rate » ou en français « Réseau de contrôleurs à débit de données flexible »), FlexRay (standardisé par la norme ISO 17458) ou Ethernet (standardisé par la norme ISO/IEC 802-3).

- [0121] Selon un exemple de réalisation particulier et non limitatif, le dispositif 4 peut fournir des signaux de sortie à un ou plusieurs dispositifs externes, tels qu'un écran d'affichage 440, tactile ou non, un ou des haut-parleurs 450 et/ou d'autres périphériques 460 (système de projection) via respectivement les interfaces de sortie 44, 45, 46. Selon une variante, l'un ou l'autre des dispositifs externes est intégré au dispositif 4.
- [0122] Bien entendu, la présente invention ne se limite pas aux exemples de réalisation décrits ci-avant mais s'étend à un procédé d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule à partir de données d'apprentissage acquises par le système de vision embarqué, qui inclurait des étapes secondaires sans pour cela sortir de la portée de la présente invention. Il en serait de même d'un dispositif configuré pour la mise en œuvre d'un tel procédé.
- [0123] La présente invention concerne également un véhicule, par exemple automobile ou plus généralement un véhicule autonome à moteur terrestre, comprenant le dispositif 4 de la [Fig.4].

Revendications

[Revendication 1]

Procédé d'acquisition d'une base de données d'apprentissage d'un modèle de prédiction de profondeur pour un système de vision embarqué dans un véhicule, le système de vision comprenant un ensemble de caméras comprenant une première caméra (11) et une deuxième caméra (12) disposées de manière à acquérir chacune une image d'une scène tridimensionnelle selon un point de vue différent, ladite base de données d'apprentissage comprenant un premier ensemble d'éléments, chaque élément dudit premier ensemble d'éléments comprenant des premières données représentatives de :

- un type d'objet,
- une première profondeur, et
- un premier ensemble de quatre images,

ledit procédé étant caractérisé en ce qu'il comprend les étapes suivantes :

- détermination (21) d'une première distribution de profondeurs associée à un type d'objet cible à partir des premières profondeurs d'un deuxième ensemble d'éléments appartenant audit premier ensemble d'éléments de la base de données d'apprentissage, chaque type d'objet d'un élément dudit deuxième ensemble d'éléments correspondant audit type d'objet cible ;

- réception (22) de deuxièmes données représentatives d'un deuxième ensemble de quatre images, ledit deuxième ensemble de quatre images comprenant :

- une première et une deuxième images acquises respectivement par la première caméra (11) et la deuxième caméra (12) à un premier instant temporel,
- une troisième et une quatrième images acquises respectivement par la première caméra (11) et la deuxième caméra (12) à un deuxième instant temporel différent du premier instant temporel ;

- détection (23) d'un objet courant, dans chaque image dudit deuxième ensemble de quatre images, un type de l'objet courant correspondant audit type d'objet cible ;

- détermination (24), dans chaque image dudit deuxième ensemble de quatre images, d'une boîte englobante comprenant des pixels associés audit objet courant ;

- détermination (25) d'une deuxième profondeur associée audit objet

courant par ledit modèle de prédiction de profondeur à partir des boîtes englobantes ;

- ajout (26) de ladite deuxième profondeur à ladite première distribution de profondeurs pour générer une deuxième distribution de profondeurs ;
et

- enregistrement (27) dans ladite base de données d'apprentissage d'un élément courant comprenant des troisièmes données représentatives de l'objet courant et de la deuxième profondeur et les deuxièmes données, ledit enregistrement étant fonction de ladite deuxième distribution de profondeurs.

- [Revendication 2] Procédé selon la revendication 1, chaque élément de ladite base de données d'apprentissage comprenant en outre des données représentatives d'un type d'environnement routier, comprenant en outre une étape de détermination d'un type d'environnement routier courant associé audit élément courant à partir desdites deuxièmes données, des données représentatives dudit type d'environnement courant étant enregistrées dans ladite base de données d'apprentissage.
- [Revendication 3] Procédé selon l'une des revendications 1 à 2, pour lequel ledit modèle de prédiction de profondeur est entraîné à partir des deuxièmes données lorsque ledit type d'objet courant correspond audit type d'objet cible.
- [Revendication 4] Procédé selon l'une des revendications 1 à 3, pour lequel ladite deuxième profondeur est égale à une profondeur moyenne d'un ensemble de pixels de la boîte englobante de la première image.
- [Revendication 5] Procédé selon l'une des revendications 1 à 4, pour lequel ledit élément courant est enregistré lorsque ledit objet courant est entièrement visible dans chacune des images du deuxième ensemble de quatre images.
- [Revendication 6] Procédé selon l'une des revendications 1 à 5, pour lequel ledit élément courant est enregistré si un nombre total d'éléments du deuxième ensemble d'éléments est inférieur à un seuil total.
- [Revendication 7] Procédé selon l'une des revendications 1 à 5, pour lequel ledit élément courant est enregistré si un nombre d'éléments d'une partie du deuxième ensemble d'éléments comprenant une première profondeur appartenant à une même classe de profondeur que la deuxième profondeur est compris entre un seuil minimal et un seuil maximal, lesdits seuils minimal et maximal étant définis en fonction d'un nombre total d'éléments du deuxième ensemble d'éléments et d'un nombre total de classes de profondeur.
- [Revendication 8] Procédé selon la revendication 7, pour lequel ledit seuil maximal est

déterminé par la fonction suivante :

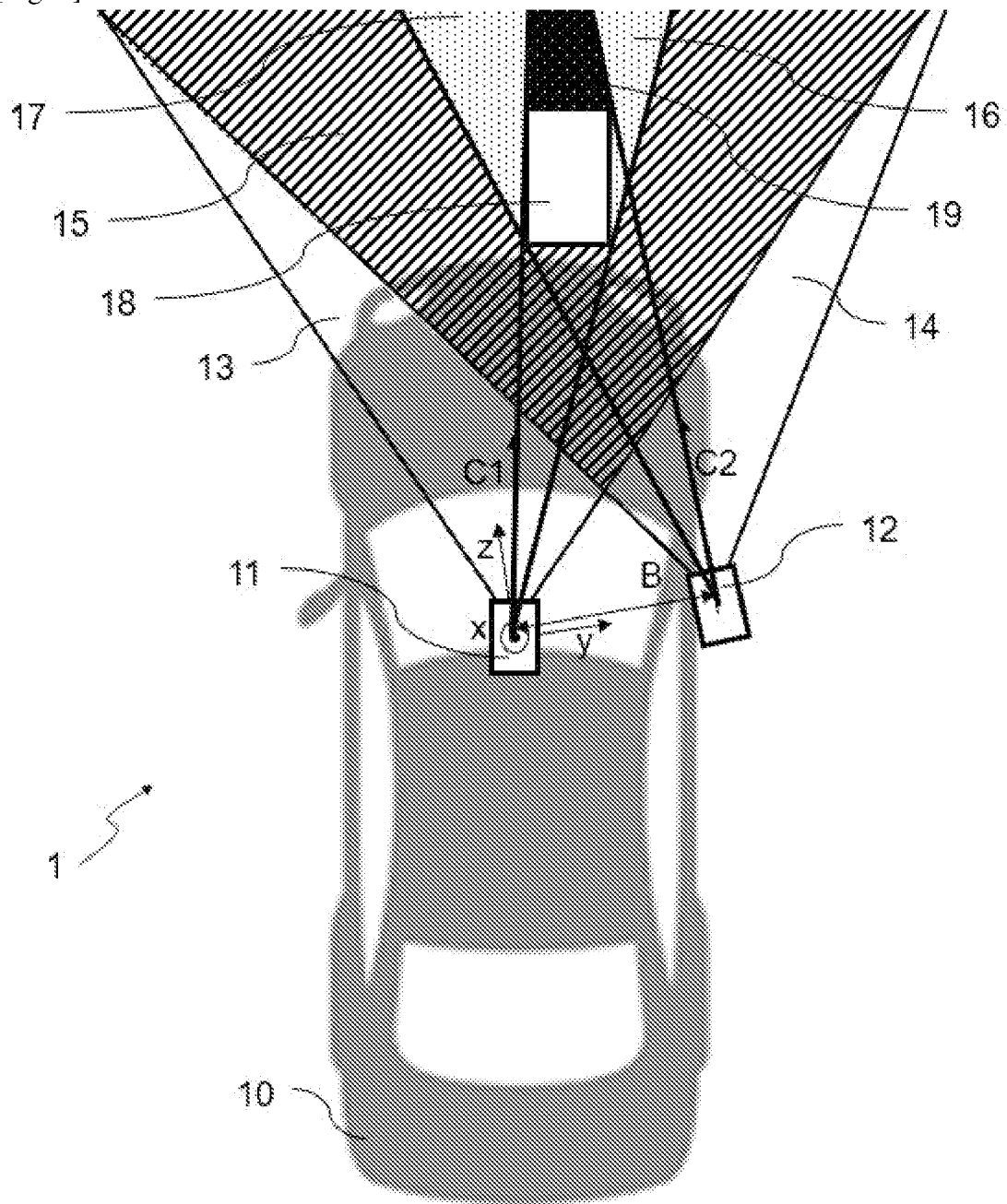
$$S_{\max} = 4 \times N_T / (3 \times I_T)$$

avec :

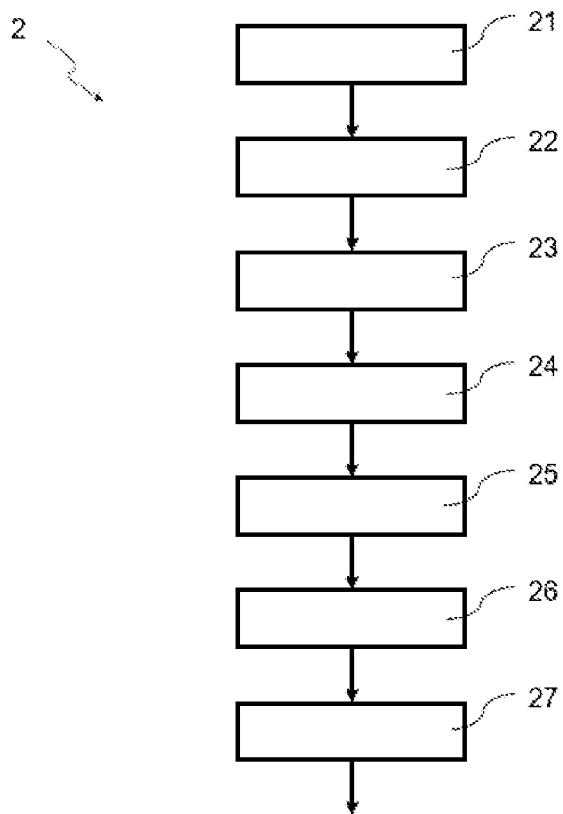
- $S_{\max}$ le seuil minimal,
- N_T le nombre total d'éléments du deuxième ensemble d'éléments, et
- I_T le nombre d'intervalles.

- [Revendication 9] Dispositif (4) d'acquisition d'une base de données d'apprentissage d'un système de vision embarqué dans un véhicule (10), ledit dispositif (4) comprenant une mémoire (41) associée à au moins un processeur (40) configuré pour la mise en œuvre des étapes du procédé selon l'une quelconque des revendications 1 à 8.
- [Revendication 10] Véhicule (10) comprenant le dispositif (4) selon la revendication 9.

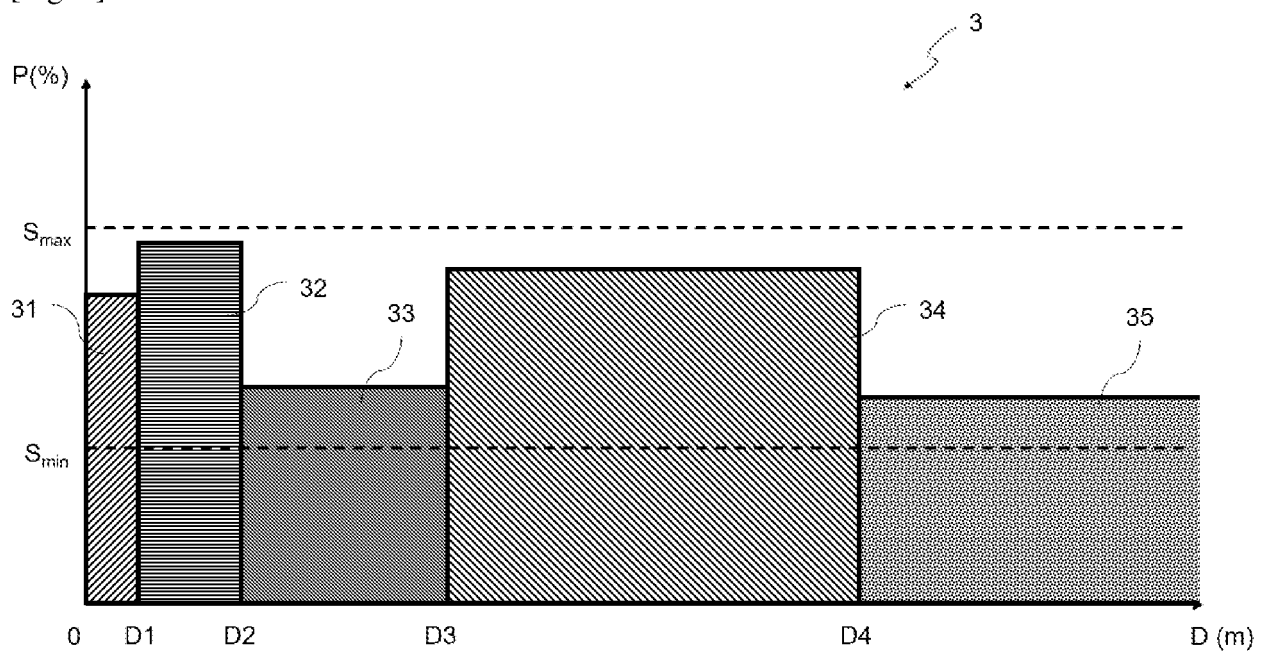
[Fig. 1]



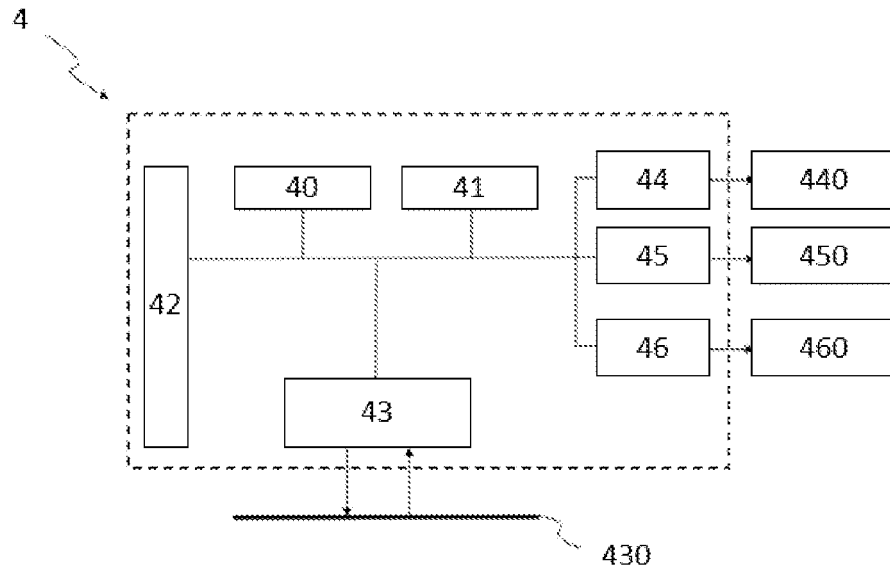
[Fig. 2]



[Fig. 3]



[Fig. 4]



**RAPPORT DE RECHERCHE
PRÉLIMINAIRE**

établi sur la base des dernières revendications
déposées avant le commencement de la recherche

N° d'enregistrement
national

FA 924717
FR 2312183

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X	US 2018/348780 A1 (ZHENG HAO [US] ET AL) 6 décembre 2018 (2018-12-06) * alinéas [0056], [0063], [0114], [0113] * * alinéa [0064] - alinéa [0065] * -----	1-10	G06F 16/50 G06N 20/00 G06T 7/00 G06T 7/593 G06V 10/10 G06V 20/58
A	US 2022/301202 A1 (PARK DENNIS [US] ET AL) 22 septembre 2022 (2022-09-22) * le document en entier * -----	1-10	
A	US 2023/037731 A1 (GUIZILINI VITOR [US] ET AL) 9 février 2023 (2023-02-09) * le document en entier * -----	1-10	
A	WO 2022/165722 A1 (HUAWEI TECH CO LTD [CN]) 11 août 2022 (2022-08-11) * le document en entier * -----	1-10	
A	US 11 805 236 B2 (NIANTIC INC [US]) 31 octobre 2023 (2023-10-31) * le document en entier * -----	1-10	
			DOMAINES TECHNIQUES RECHERCHÉS (IPC)
			G06V
Date d'achèvement de la recherche		Examineur	
2 mai 2024		Stan, Johann	
CATÉGORIE DES DOCUMENTS CITÉS		T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire			

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE **RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 2312183 FA 924717**

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.
 Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **02-05-2024**
 Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche		Date de publication		Membre(s) de la famille de brevet(s)	Date de publication
US 2018348780 A1	06-12-2018	CN	110753953 A		04-02-2020
		CN	110799982 A		14-02-2020
		EP	3635496 A1		15-04-2020
		US	2018348780 A1		06-12-2018
		US	2018348781 A1		06-12-2018
		US	2022350340 A1		03-11-2022
		WO	2018224875 A1		13-12-2018
		WO	2018224878 A1		13-12-2018

US 2022301202 A1	22-09-2022	JP	2022142784 A		30-09-2022
		JP	2022142787 A		30-09-2022
		JP	2022142789 A		30-09-2022
		JP	2022142790 A		30-09-2022
		US	2022300746 A1		22-09-2022
		US	2022300768 A1		22-09-2022
		US	2022301202 A1		22-09-2022
		US	2022301203 A1		22-09-2022

US 2023037731 A1	09-02-2023	US	2022084230 A1		17-03-2022
		US	2023037731 A1		09-02-2023

WO 2022165722 A1	11-08-2022	AUCUN			

US 11805236 B2	31-10-2023	TW	202201348 A		01-01-2022
		US	2021352261 A1		11-11-2021
		WO	2021229455 A1		18-11-2021
