

①9 RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
COURBEVOIE

①1 N° de publication : **3 148 485**

(à n'utiliser que pour les
commandes de reproduction)

②1 N° d'enregistrement national : **23 04517**

⑤1 Int Cl⁸ : **G 06 T 7/00 (2023.01), H 04 N 13/00, B 60 R 1/00,
B 60 W 40/00, 50/00, G 06 N 3/00, 20/00**

⑫

DEMANDE DE BREVET D'INVENTION

A1

②2 Date de dépôt : 05.05.23.

③0 Priorité :

④3 Date de mise à la disposition du public de la
demande : 08.11.24 Bulletin 24/45.

⑤6 Liste des documents cités dans le rapport de
recherche préliminaire : *Se reporter à la fin du
présent fascicule*

⑥0 Références à d'autres documents nationaux
apparentés :

Demande(s) d'extension :

⑦1 Demandeur(s) : *PSA AUTOMOBILES SA Société par
actions simplifiée (SAS) — FR.*

⑦2 Inventeur(s) : LI HAI.

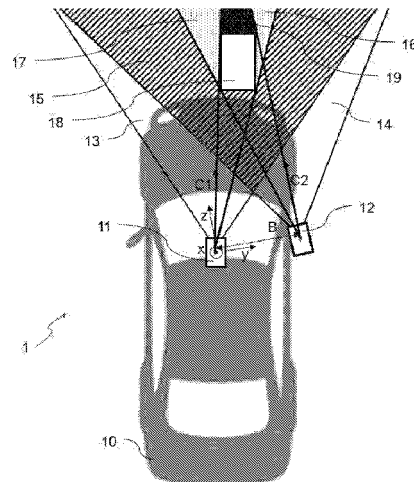
⑦3 Titulaire(s) : STELLANTIS AUTO SAS Société par
actions simplifiée.

⑦4 **Procédé(s) : dispositif de détermination d'un
masque de visibilité pour un système de vision
embarqué dans un véhicule.**

⑤7 L'invention concerne un procédé et un dispositif de détermination d'un masque de visibilité pour système de vision stéréoscopique embarqué dans un véhicule (10). Le système de vision comprend au moins deux caméras (11, 12) pour l'acquisition d'images d'une même scène tridimensionnelle selon des points de vue déterminés. A cet effet, des première

et deuxième images sont reçues, des profondeurs associées à un ensemble de pixels de la première image sont prédites, cet ensemble de pixels est reprojété dans la scène tridimensionnelle sous forme d'un ensemble de points en fonction des profondeurs prédites et une troisième image est générée par projection de l'ensemble de points. Un masque de visibilité est alors déterminé à partir des coordonnées dans l'espace des points de l'ensemble de points et à partir des coordonnées des pixels projetés dans la troisième image.

Figure pour l'abrégé : Figure 1



FR 3 148 485 - A1



Description

Titre de l'invention : Procédé et dispositif de détermination d'un masque de visibilité pour un système de vision embarqué dans un véhicule.

Domaine technique

[0001] La présente invention concerne les procédés et dispositifs de détermination d'un masque de visibilité pour système de vision embarqué dans un véhicule, par exemple dans un véhicule automobile. La présente invention concerne également un procédé et un dispositif de contrôle d'un ou plusieurs systèmes ADAS embarqués dans un véhicule à partir d'un masque de visibilité déterminé.

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] De la qualité des données émises par un système de vision dépend donc le bon fonctionnement des périphériques d'aides à la conduite utilisant ces données.

[0006] De nombreux systèmes de vision perçoivent un environnement autour d'un véhicule à partir de plusieurs images acquises par une ou plusieurs caméras. Lors de l'exploitation des images, des zones occluses des images qui correspondent à des zones de l'environnement qui ne sont pas présentes sur l'ensemble des images acquises sont

définies. Un masque de visibilité associé à une image définit alors par exemple un filtre permettant de déterminer des pixels associés à des zones ne se retrouvant pas dans les autres images.

- [0007] Des solutions pour détecter une occlusion, c'est-à-dire pour déterminer un masque de visibilité existent.
- [0008] Une première solution présentée par « Occlusion Aware Unsupervised Learning of Optical Flow » de Yang Wang, Yi Yang, Zhenheng Yang, Liang Zhao, Peng Wang et Wei Xu parue le 4 avril 2018 se base sur le flux optique inverse. Pour chaque pixel d'une première image représenté par ses coordonnées, l'algorithme vérifie si un pixel d'une deuxième image arrive à ce pixel de la première image avec le flux optique inverse en balayant tous les pixels de la deuxième image. Cette méthode peut être utilisée pour les deux sens du flux d'optique pour identifier les zones occluses des deux images.
- [0009] Une deuxième solution est décrite dans le document « Geometry-based Occlusion-Aware Unsupervised Stereo Matching for Autonomous Driving algorithm ». La détection des zones occluses se base sur une contrainte géométrique : le pixel occlus et un autre pixel qui le cache se retrouvent projetés dans un même pixel d'une image reconstruite.

Résumé de la présente invention

- [0010] Un objet de la présente invention est de résoudre au moins l'un des problèmes de l'arrière-plan technologique décrit précédemment.
- [0011] Un autre objet de la présente invention est de proposer une solution alternative pour déterminer un masque de visibilité pour tout système de vision afin d'améliorer la qualité des données issues de la ou les caméras du système de vision.
- [0012] Un autre objet de la présente invention est de réduire les ressources nécessaires pour la détermination d'un masque de visibilité.
- [0013] Selon un premier aspect, la présente invention concerne un procédé de détermination de masques de visibilité par un système de vision stéréoscopique embarqué dans un véhicule, le système de vision stéréoscopique comprenant un ensemble de caméras d'au moins deux caméras disposées de manière à acquérir chacune une image d'une scène tridimensionnelle selon un point de vue différent, ladite deuxième caméra se situant à droite de la première caméra d'un point de vue de la première caméra, le procédé étant caractérisé en ce qu'il comprend les étapes suivantes :
- réception de premières et deuxièmes données respectivement représentatives d'une première et deuxième images acquises par respectivement une première et deuxième caméras de l'ensemble de caméras à un même instant temporel d'acquisition ;
 - prédiction de profondeurs associées à un ensemble de pixels de la première image par le système de vision stéréoscopique à partir d'un modèle de prédiction appris,

chaque pixel de la première image ayant des coordonnées principales dans la première image ;

- reprojection dans la scène tridimensionnelle de l'ensemble de pixels sous forme d'un ensemble de points en fonction des profondeurs, d'une matrice intrinsèque de la première caméra et de paramètres extrinsèques du système de vision stéréoscopique ;

- détermination d'un premier masque de visibilité associé aux pixels de l'ensemble de pixels en fonction des coordonnées de points de l'ensemble de points, un pixel de l'ensemble de pixels étant non visible dans la deuxième image si les coordonnées d'un point dudit ensemble de points associé audit pixel le situent en dehors d'un champ de vision de la deuxième caméra déterminé en fonction d'une largeur de la deuxième image et d'une distance focale de la deuxième caméra ;

- génération d'une troisième image par projection de l'ensemble de points en fonction d'une matrice intrinsèque de la deuxième caméra, des coordonnées secondaires dans la troisième image étant associées à chaque pixel de l'ensemble de pixels ;

- pour chaque ligne de la première image rectifiée, balayage de gauche à droite, selon un point de vue de la première caméra, de valeurs d'indices de colonne d'arrivée et détection d'un ensemble de pixels irréguliers dont l'indice de colonne d'arrivée ne suit pas une fonction monotone représentative d'une évolution d'indices de colonne d'arrivée en fonction d'un indice de colonne dans la première image rectifiée, et pour chaque pixel irrégulier de l'ensemble, identification d'un ensemble de pixels occlus dans la ligne à gauche de chaque pixel irrégulier dont un indice de colonne d'arrivée est supérieur ou égal à un indice de colonne d'arrivée de chaque pixel irrégulier,

un deuxième masque de visibilité étant déterminé comme une union des ensembles des pixels occlus ; et

- détermination d'un troisième masque de visibilité par association du premier masque de visibilité et du deuxième masque de visibilité.

[0014] Selon une variante de procédé, la reprojection d'un pixel dans la scène tridimensionnelle est réalisée à l'aide de la formule suivante :

$$p_{3D}(p_t) = T \cdot \phi(p_t | K^{-1}, D(p_t)) \text{ avec :}$$

- $p_{3D}(p_t)$ les coordonnées dans la scène tridimensionnelle du point issu de la reprojection du pixel p_t de la première image,

- T une matrice de déplacement entre une position de la première caméra et une position de la deuxième caméra,

- K la matrice intrinsèque de la première caméra associée à la projection d'un point de la scène tridimensionnelle dans une image acquise par la première caméra,

- ϕ une fonction de reprojection dans la scène tridimensionnelle d'un pixel en

fonction de sa profondeur,

- $D(p_i)$ est une profondeur du pixel p_i prédite par le système de vision stéréoscopique.

[0015] Selon une autre variante de procédé, la projection d'un point de la scène tridimensionnelle est réalisée à l'aide de la formule suivante :

$$p_s = \pi \left(K' \left[p_{3D}(p_i) \right] \right) \text{ avec } : - \pi \text{ une fonction pour passer de coordonnées}$$

homogènes dans l'espace à trois dimensions à des coordonnées pixels en deux dimensions en supprimant une dimension d'un vecteur,

- K' la matrice intrinsèque de la deuxième caméra associée à la projection d'un point de la scène tridimensionnelle dans une image acquise par la deuxième caméra,

- $p_{3D}(p_i)$ les coordonnées dans la scène tridimensionnelle du point issu de la re-projection du pixel p_i de la première image.

[0016] Selon encore une variante de procédé, le premier masque de visibilité est obtenu à l'aide de la formule suivante :

$$\bigcup_p p_i \text{ tel que } \frac{x_{p_{3D}(p_i)}}{D(p_i)} > \frac{W/2}{f} \text{ avec } :$$

- $x_{p_{3D}(p_i)}$ l'abscisse d'un point issu de la re-projection du pixel p_i de la première image, un axe des abscisses étant défini parallèlement à un axe suivant lequel se situent la première et la deuxième caméras,

- $D(p_i)$ la profondeur du pixel p_i de la première image prédite par le système de vision stéréoscopique,

- W la largeur de la deuxième image, et

- f la distance focale de la deuxième caméra.

[0017] Selon encore une autre variante de procédé, les profondeurs sont prédites par un réseau de neurones convolutif.

[0018] Selon une variante supplémentaire de procédé, le réseau de neurones convolutif est entraîné pour minimiser une erreur photométrique définie par la fonction de perte suivante :

$$L(p) = \sum_p \left[(1-\alpha) \cdot |I(p) - \hat{I}(p)| + \alpha \cdot \left(1 - \frac{1}{2} SSIM(I(p), \hat{I}(p)) \right) \right]$$

Avec :

- $I(p)$ une valeur du pixel p dans la deuxième image ;

- $\hat{I}(p)$ une valeur du pixel p dans la troisième image ;

- $SSIM$ une fonction qui prend en compte une structure locale ; et

- α un facteur de pondération dépendant d'un type d'environnement routier.

[0019] Selon un deuxième aspect, la présente invention concerne un dispositif de détermination d'un masque de visibilité 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.

- [0020] 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.
- [0021] 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.
- [0022] 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.
- [0023] 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.
- [0024] 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.
- [0025] 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.
- [0026] 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

- [0027] 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 5 annexées, sur lesquelles :
- [0028] [Fig.1] illustre schématiquement un système de vision stéréoscopique équipant un véhicule, selon un exemple de réalisation particulier et non limitatif de la présente

invention ;

[0029] [Fig.2] illustre schématiquement un système de vision stéréoscopique équipant un véhicule, selon un exemple de réalisation particulier et non limitatif de la présente invention ;

[0030] [Fig.3] illustre schématiquement un dispositif configuré pour la détermination d'un masque de visibilité par système de vision embarqué dans le véhicule de la [Fig.1], selon un exemple de réalisation particulier et non limitatif de la présente invention ;

[0031] [Fig.4] illustre un organigramme des différentes étapes d'un procédé de détermination d'un masque de visibilité par système 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.5] illustre une matrice présentant différents indices de colonne pour des pixels d'une ligne et un critère de visibilité pour un système de vision embarqué dans le véhicule de la [Fig.1], selon un exemple de réalisation particulier et non limitatif de la présente invention.

[0033] **Description des exemples de réalisation**

[0034] Un procédé et un dispositif de détermination d'un masque de visibilité pour un système de vision embarqué dans un véhicule vont maintenant être décrits dans ce qui va suivre en référence conjointement aux figures 1 à 5. Des mêmes éléments sont identifiés avec des mêmes signes de référence tout au long de la description qui va suivre.

[0035] Selon un exemple particulier et non limitatif de réalisation de la présente invention, un procédé de détermination d'un masque de visibilité pour un système de vision stéréoscopique embarqué dans un véhicule est par exemple mis en œuvre par un calculateur du système embarqué du véhicule contrôlant ce système de vision.

[0036] Le système de vision stéréoscopique comprend un ensemble de caméras d'au moins deux caméras disposées de manière à acquérir chacune une image d'une scène tridimensionnelle selon un point de vue différent, la deuxième caméra se situant à droite du point de vue de ladite première caméra.

[0037] A cet effet, le procédé de détermination d'un masque de visibilité par un système de vision stéréoscopique embarqué dans un véhicule comprend la réception de premières et deuxièmes données respectivement représentatives d'une première et deuxième images acquises selon un point de vue différent par les première et deuxième caméras à un même instant temporel d'acquisition.

[0038] Le procédé comprend également la prédiction de profondeurs associées à un ensemble de pixels de la première image par le système de vision stéréoscopique à partir d'un modèle de prédiction appris, chaque pixel de la première image ayant des coordonnées principales dans la première image, la reprojexion dans la scène tridi-

mensionnelle des pixels de la première image sous forme d'un ensemble de points en fonction des profondeurs, d'une matrice intrinsèque de la première caméra et de paramètres extrinsèques du système de vision stéréoscopique.

- [0039] Le procédé détermine alors un premier masque de visibilité associé aux pixels de l'ensemble de pixels en fonction des coordonnées de points de l'ensemble de points, un pixel de l'ensemble de pixels étant non visible dans la deuxième image si les coordonnées d'un point de l'ensemble de points associé au pixel le situent en dehors d'un champ de vision de la deuxième caméra déterminé en fonction d'une largeur de la deuxième image et d'une distance focale de la deuxième caméra.
- [0040] Le procédé comprend ensuite la génération d'une troisième image par projection de l'ensemble de points en fonction d'une matrice intrinsèque de la deuxième caméra, des coordonnées secondaires dans la troisième image étant associées à chaque pixel de l'ensemble de pixels. La génération de cette troisième image permet alors la détermination d'un deuxième masque de visibilité associé aux pixels de l'ensemble de pixels. Pour chaque ligne de la première image rectifiée, des valeurs d'indices de colonne d'arrivée sont balayées et un ensemble de pixels irréguliers est détecté. Pour chaque pixel irrégulier de l'ensemble, un ensemble de pixels est occlus est identifié dans la ligne à gauche de chaque pixel irrégulier détecté. Un deuxième masque de visibilité est alors déterminé comme une union des ensembles des pixels occlus.
- [0041] Un troisième masque de visibilité est alors déterminé par association du premier masque de visibilité et du deuxième masque de visibilité.
- [0042] La [Fig.1] illustre schématiquement un système de vision stéréoscopique équipant un véhicule, selon un exemple de réalisation particulier et non limitatif de la présente invention.
- [0043] 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 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.
- [0044] Le véhicule 10 comprend avantageusement plusieurs caméras 11, 12 embarquées, chacune configurée pour acquérir des images d'une scène tridimensionnelle dans l'environnement du véhicule 10. Cet ensemble de caméras 11, 12 forme le système de vision stéréoscopique. Deux caméras 11, 12 sont illustrées sur la [Fig.1]. La présente invention ne se limite cependant pas à un système de vision stéréoscopique comprenant deux caméras mais s'étend à tout système de vision comprenant deux caméras ou plus, par exemple 3, 4 ou 5 caméras.
- [0045] Les caméras 11, 12 disposent de paramètres intrinsèques connus. Ces paramètres se

composent notamment de :

- distance focale f_1 de la première caméra 11 ;
- distance focale f_2 de la deuxième caméra 12 ;
- distorsions qui sont dues aux imperfections du système optique de chaque caméra ;
- direction C_1 de l'axe optique de la première caméra 11 ;
- direction C_2 de l'axe optique de la deuxième caméra 12 ;
- résolutions respectives des caméras 11, 12.

[0046] 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.

[0047] 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.

[0048] Ces caméras 11, 12 sont disposées de manière à acquérir chacune une image d'une scène tridimensionnelle selon un point de vue différent, la deuxième caméra 12 se situant à droite du point de vue de ladite première caméra 11, 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 deux caméras sont situées en haut du pare-brise du véhicule, celles-ci sont alors placées à une certaine distance.

[0049] 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 C_1 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.

[0050] 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 .

- [0051] Une contrainte principale du 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, le « baseline » doit atteindre 60cm pour les caméras couramment utilisées dans ce domaine.
- [0052] Les deux caméras 11, 12 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 deux caméras 11, 12 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 tridimensionnelle par la deuxième caméra 12 et le troisième champ d'acquisition 15 permet une vision stéréoscopique de la scène tridimensionnelle par le système de vision stéréoscopique composé des deux caméras 11, 12.
- [0053] 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.
- [0054] Parmi ces trois champs, le champ 16 est visible depuis la deuxième caméra 12. La partie de la scène tridimensionnelle présente dans ce champ 16 est donc observable à l'aide du système de vision monoscopique composé de la deuxième caméra 12.
- [0055] Le champ 17 est quant à lui visible depuis la première caméra 11. La partie de la scène tridimensionnelle présente dans ce champ 17 est donc observable à l'aide du système de vision monoscopique composé de la deuxième caméra 12.
- [0056] Enfin, le champ 19 n'est visible depuis aucune des caméras. La partie de la scène tridimensionnelle présente dans ce champ 19 n'est donc pas observable.
- [0057] Les directions C_1 , C_2 des axes optiques sont représentatifs d'une orientation du champ de vision de chaque caméra 11, 12.
- [0058] Il est évident qu'il est possible d'utiliser un tel système de vision stéréoscopique pour prendre des images de scènes tridimensionnelles situées sur les côtés ou derrière le véhicule 10 en l'équipant de caméras placées et orientées différemment.
- [0059] Les images acquises par les caméras 11, 12 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 tridimensionnelle 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).

- [0060] Les images acquises par les caméras 11, 12 représentent des vues d'une même scène tridimensionnelle prises à des points de vue différents, les position des caméras étant distinctes. Sur cette scène tridimensionnelle se trouvent par exemple :
- 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.
- [0061] 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.
- [0062] La [Fig.2] illustre schématiquement un système de vision stéréoscopique équipant un véhicule, selon un exemple de réalisation particulier et non limitatif de la présente invention.
- [0063] Des points 20, 21, 22 de la scène tridimensionnelle sont visibles du point de vue de la première caméra 11.
- [0064] Des points 21 sont également visibles du point de vue de la deuxième caméra 12.
- [0065] Des points 20 sont quant à eux occlus du point de vue de la deuxième caméra 12 car ils sont masqués par des points 21 se situant sur des mêmes axes dans le champ de vision de la deuxième caméra 12.
- [0066] Des points 22 sont situés en dehors du champ de vision de la deuxième caméra 12.
- [0067] Ainsi, lors de l'acquisition des première et deuxième images par respectivement la première caméra 11 et la deuxième caméra 12, des pixels associés aux points visibles 20, 21, 22 depuis le point de vue de la première caméra 11 seront présents dans la première image, alors que seuls des pixels associés aux points 21 visibles depuis le point de vue de la deuxième caméra seront présents dans la deuxième image.
- [0068] Les pixels associés aux points 20 visibles depuis le point de vue de la première caméra 11 et non visibles depuis le point de vue de la deuxième caméra 12 sont appelés par la suite pixels occlus.
- [0069] La [Fig.3] illustre schématiquement un dispositif 4 configuré pour la détermination d'un masque de visibilité pour un système de vision embarqué dans un véhicule 10, 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 calculateur.

- [0070] Le dispositif 4 est par exemple configuré pour la mise en œuvre des opérations et/ou étapes décrites en regard des figures 1, 2 et 4. 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 calculateur 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.
- [0071] 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.
- [0072] 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.
- [0073] 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.
- [0074] 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 Multimédia Haute Définition » en français) ;

- interface LIN (de l'anglais « Local Interconnect Network », ou en français « Réseau interconnecté local »).

- [0075] 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).
- [0076] 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.
- [0077] La [Fig.4] illustre un organigramme des différentes étapes d'un procédé 2 de détermination d'un masque de visibilité pour un système de vision embarqué dans le véhicule de la [Fig.1], le système de vision comprenant au moins une caméra 11 disposée de manière à acquérir une image d'une scène tridimensionnelle selon un point de vue déterminé, selon un exemple de réalisation particulier et non limitatif de la présente invention.
- [0078] Le procédé est par exemple mis en œuvre par un ou plusieurs processeurs d'un ou plusieurs calculateurs embarqués dans le véhicule 10, par exemple par un calculateur contrôlant le système de vision.
- [0079] Dans une première étape 31, le calculateur reçoit des premières données représentatives d'une première image acquise par une caméra première 11 à un instant temporel donné.
- [0080] Dans une deuxième étape 32, le calculateur reçoit des deuxièmes données représentatives d'une deuxième image acquise par une deuxième caméra 12 au même instant temporel donné.

- [0081] Les deux images reçues correspondent à deux vues d'une même scène tridimensionnelle se déroulant aux alentours du véhicule 10 prises depuis deux points de vue différents à un même instant temporel donné.
- [0082] Afin de faciliter l'analyse des deux images reçues, les première et deuxième images sont rectifiées suivant une méthode connue de l'homme du métier. Une telle méthode est décrite, par exemple, dans « Rectification Projective d'Images Stéréo non Calibrées Infrarouges avec prise en compte globale de la minimisation des distorsions » de Benoit Ducarouge, Thierry Sentenac, Florian Bugarin et Michel Devy du 16 juillet 2009.
- [0083] La méthode de rectification consiste à réorienter les lignes épipolaires pour qu'elles soient parallèles avec l'axe horizontal de l'image. Cette méthode est décrite par une transformation qui projette les épipoles à l'infini et dont les points correspondants sont nécessairement sur une même ordonnée.
- [0084] Un algorithme de rectification consiste, par exemple, en 4 étapes :
- Faire pivoter (virtuellement) la première caméra 11 pour que l'épipoles aille à l'infini le long de l'axe horizontal du repère qui lui est associé;
 - Appliquer la même rotation à la deuxième caméra 12 pour se retrouver dans la configuration géométrique initiale;
 - Faire pivoter la deuxième caméra de la rotation associée à la matrice de rotation 'R', correspondant au paramètre extrinsèque du système de vision stéréoscopique de départ;
 - Ajuster l'échelle dans les deux repères caméras.
- [0085] Il est à noter que la rectification simplifie la mise en correspondance des pixels des images stéréo, c'est-à-dire obtenues par un système de vision stéréoscopique. Le pixel correspondant dans la deuxième image à un pixel de la première image (et réciproquement) est positionné sur la même ligne. A partir de la connaissance de la géométrie épipolaire et donc d'une matrice fondamentale du système stéréo, l'objectif est alors de déterminer une paire de transformations projectives, appelée homographies, qui réorientent les projections épipolaires parallèlement aux lignes des images, donc à l'axe horizontal des caméras rectifiées.
- [0086] Dans une étape 33, des profondeurs associées à un ensemble de pixels de la première image sont prédites par le système de vision stéréoscopique à partir d'un modèle de prédiction appris.
- [0087] Un tel apprentissage auto-supervisé, c'est-à-dire ne nécessitant pas d'intervention extérieure ou une utilisation de données annotées, est par exemple réalisé en minimisant l'erreur photométrique calculée lors de reconstructions d'images.
- [0088] Des disparités sont déterminées à partir de l'obtention de première et deuxième images de la première caméra 11 et de la deuxième caméra 12 à un même instant

temporel, les disparités étant définies par la fonction suivante :

[0089] [Math.1]

$$p_{ss}^x = p_t^x - d(p_t)$$

[0090] avec :

- p_{ss}^x l'abscisse d'un pixel dans la deuxième image,
- p_t^x l'abscisse d'un pixel dans la première image, et
- $d(p_t)$ une disparité déterminée pour un pixel p_t de la première image.

[0091] A l'aide des disparités précédemment déterminées, des profondeurs sont calculées pour les pixels de la première image :

[0092] [Math.2]

$$D_{ts}(p_t) = f_1 \times \frac{B}{d(p_t)}$$

[0093] avec :

- $D_{ts}(p_t)$ la profondeur du pixel p_t de la première image prédite par le système de vision stéréoscopique ,
- $d(p_t)$ une disparité déterminée pour un pixel p_t de la première image, et
- f_1 la distance focale de la première caméra 11.

[0094] Une troisième image est reconstruite à partir de la première image et des profondeurs précédemment calculées via la formule suivante :

[0095] [Math.3]

$$p_s = \pi \left(K' \left[T \cdot \phi \left(p_t | K^{-1}, D(p_t) \right) \right] \right)$$

avec :

[Math.3]

-

[Math.3]

π

une fonction pour passer de coordonnées homogènes dans l'espace à trois dimensions à des coordonnées pixels en deux dimensions en supprimant une dimension d'un vecteur,

-

[Math.3]

K

la matrice intrinsèque de la première caméra 11 associée à la projection d'un point de la scène tridimensionnelle dans une image acquise par la première caméra 11,

-

[Math.3]

K'

la matrice intrinsèque de la deuxième caméra 12 associée à la projection d'un point de la scène tridimensionnelle dans une image obtenue par la deuxième caméra 12,

-

[Math.3]

T

une matrice de déplacement entre une position de la première caméra et une position de la deuxième caméra,

-

[Math.3]

\emptyset

une fonction de reprojection dans la scène tridimensionnelle d'un pixel en fonction de sa profondeur, et

-

[Math.3]

$D(p_t)$

est une profondeur du pixel

[Math.3]

p_t

de la première image prédite par le système de vision stéréoscopique .

[0096] L'image reconstruite est ensuite comparée à la deuxième image afin de déterminer une erreur photométrique :

[0097] [Math.4]

$$L(p) = \sum_p \left[(1-\alpha) \cdot |I(p) - \hat{I}(p)| + \alpha \cdot \left(1 - \frac{1}{2} SSIM(I(p), \hat{I}(p)) \right) \right]$$

[0098] avec :

- $I(p)$ une valeur du pixel p dans la deuxième image,

- $\hat{I}(p)$ une valeur du pixel p dans la troisième image,

- SSIM (de l'anglais « structural similarity index measure », en français « mesure de l'indice de similarité structurelle ») une fonction qui prend en compte une structure locale, et

- α un facteur de pondération dépendant d'un type d'environnement routier.

[0099] Le réseau de neurones convolutif est alors appris pour minimiser l'erreur photométrique précédemment définie.

[0100] Ainsi, en sortie de l'étape 33, chaque pixel est défini en fonction de coordonnées principales (x,y) dans la première image et une profondeur prédite pour ce pixel.

[0101] Dans une étape 34, les pixels de l'ensemble de pixels sont reprojétés dans la scène tridimensionnelle sous forme d'un ensemble de points en fonction des profondeurs prédites lors de l'étape 33, de la matrice intrinsèque K de la première caméra 11 et de paramètres extrinsèques T du système de vision stéréoscopique.

[0102] Une telle reprojektion se fait par exemple à partir de la formule suivante :

[0103] [Math.5]

$$p_{3D}(p_t) = T \cdot \phi(p_t | K^{-1}, D(p_t))$$

[0104] avec :

- $p_{3D}(p_t)$ les coordonnées dans la scène tridimensionnelle du point issu de la reprojektion du pixel p_t de la première image,
- T une matrice de déplacement entre une position de la première caméra 11 et une position de la deuxième caméra 12,
- K la matrice intrinsèque de la première caméra 11 associée à la projection d'un point de la scène tridimensionnelle dans une image acquise par la première caméra 11,
- ϕ une fonction de reprojektion dans la scène tridimensionnelle d'un pixel en fonction de sa profondeur,
- $D(p_t)$ est une profondeur du pixel p_t de la première image prédite par le système de vision stéréoscopique .

[0105] Ainsi, l'ensemble de points est situé dans la scène tridimensionnelle à des positions telles que pourrait les voir la deuxième caméra 12.

[0106] Il est cependant possible que certains des points 22 projetés dans la scène tridimensionnelle se situent en dehors du champ de vision de la deuxième caméra 12. En effet, certaines parties de la scène tridimensionnelle ne sont pas visibles par les deux caméras 11, 12 à la fois.

[0107] Dans une étape 35, un premier masque de visibilité associé aux pixels de l'ensemble de pixels est déterminé en fonction des coordonnées de points de l'ensemble de points.

[0108] Un pixel de l'ensemble de pixels est défini comme non visible dans la deuxième image si les coordonnées du point 22 de l'ensemble de points associé au pixel se situent en dehors d'un champ de vision de la deuxième caméra 12 déterminé en fonction d'une largeur de la deuxième image et d'une distance focale de la deuxième caméra 12.

[0109] Par exemple, il est possible qu'un pixel soit la projection d'un point 22 de la scène tridimensionnelle qui se situe dans le premier champ d'acquisition 13 que seule la première caméra 11 perçoit. Dans ce cas, le point 22 de la scène tridimensionnelle est en dehors du champ de vision de la deuxième caméra 12. Il est alors nécessaire de détecter ce point 22 non visible pour la deuxième caméra 12 car celui-ci ne peut avoir

de pixel associé dans la deuxième image.

[0110] Le repère de coordonnées des points dans la scène tridimensionnelle est défini en fonction de l'orientation des caméras 11, 12. Ainsi, l'axe x du repère associé à la scène tridimensionnelle est parallèle à un axe défini par les positions des caméras 11, 12, les caméras étant placées sur cet axe, l'axe z du repère est l'axe focal de la deuxième caméra 12.

[0111] Le principe est de comparer le rapport entre une abscisse $x_{p_{3D}}$ d'un point de la scène tridimensionnelle et la profondeur du point de la scène tridimensionnelle au rapport entre la demi-largeur de la deuxième image et la distance focale de la deuxième caméra.

[0112] Le premier masque de visibilité est obtenu par exemple à l'aide de la formule suivante :

[0113] [Math.6]

$$\bigcup_p p_i \text{ tel que } \frac{x_{p_{3D}(p_i)}}{D(p_i)} > \frac{W/2}{f}$$

[0114] avec :

- $x_{p_{3D}(p_i)}$ l'abscisse du point dans la scène tridimensionnelle issu de la reprojction du pixel p_i de la première image, un axe des abscisses étant défini parallèlement à un axe suivant lequel se situent la première et la deuxième caméras 11, 12,
- $D(p_i)$ ladite profondeur du pixel p_i de la première image prédite par le système de vision stéréoscopique,
- W la largeur de la deuxième image, et
- f la distance focale de la deuxième caméra 12.

[0115] Ainsi, le premier masque de visibilité permet d'identifier les pixels de la première image pour lesquels les points 22 reprojétés dans la scène tridimensionnelle se situent en dehors du champ de vision de la deuxième caméra 12.

[0116] Dans une étape 36, une troisième image est générée par projection de l'ensemble de points en fonction d'une matrice intrinsèque K' de la deuxième caméra 12.

[0117] Des coordonnées secondaires (i,j) dans la troisième image sont associées à chaque pixel de l'ensemble de pixels.

[0118] La projection d'un point de la scène tridimensionnelle est réalisée par exemple à l'aide de la formule suivante :

$$p_s = \pi \left(K' \left[p_{3D}(p_i) \right] \right) \text{ avec } :- \pi \text{ une fonction pour passer de coordonnées}$$

homogènes dans l'espace à trois dimensions à des coordonnées pixels en deux dimensions en supprimant une dimension d'un vecteur,

- K' la matrice intrinsèque de la deuxième caméra 12 associée à la projection d'un point de la scène tridimensionnelle dans une image acquise par la deuxième caméra 12,

- $P_{3D}(P_i)$ les coordonnées dans la scène tridimensionnelle du point issu de la re-projection du pixel P_i de la première image.

- [0119] Lorsque cette opération est effectuée, chaque pixel de la première image est ainsi défini par ses coordonnées (x,y) dans la première image, par un indice de colonne d'arrivée i et par un indice de ligne d'arrivée j dans ladite troisième image.
- [0120] Dans une étape 37, un deuxième masque de visibilité associé aux pixels de l'ensemble de pixels est déterminé.
- [0121] Le principe est de déterminer les points de la scène tridimensionnelle qui sont masqués par d'autres points plus proches de la deuxième caméra 12, c'est-à-dire les pixels occlus associés à ces points.
- [0122] L'axe des abscisses de la première image est orienté positivement de gauche à droite.
- [0123] Pour chaque ligne de la première image rectifiée, un balayage des pixels est réalisé de gauche à droite selon un point de vue de la première caméra 11. Une matrice, telle que présentée dans la [Fig.5], est constituée, présentant des valeurs d'indices de colonne d'arrivée i pour chaque pixel de la ligne dont la colonne de départ est définie par 'x'.
- [0124] En l'absence de pixels occlus, c'est-à-dire si tous les pixels d'une ligne 'y' de la première image rectifiée ont un pixel d'arrivée différent sur la ligne 'j' de la troisième image rectifiée, alors les indices 'i' de colonne d'arrivée sont répartis dans la matrice suivant une fonction monotone croissante.
- [0125] On appelle pixel irrégulier un pixel de la première image dont l'indice i' de colonne d'arrivée ne suit pas la fonction monotone précédemment décrite. Un tel pixel irrégulier, placé à un rang n de la matrice, est alors détecté lorsque $i'_n < i'_{n-1}$.
- [0126] A la suite de la détection en position n d'un pixel irrégulier, l'ensemble des pixels de la ligne balayée masqués par ce pixel irrégulier est alors identifié, un pixel masqué ou occlus étant un pixel dont l'indice de colonne d'arrivée i_k est supérieur ou égale à l'indice de colonne d'arrivée i'_n du pixel irrégulier. Ainsi, un pixel à gauche du pixel irrégulier est occlus par le pixel irrégulier si $i_k \geq i'_n$.
- [0127] Le deuxième masque de visibilité est ainsi défini comme l'union des pixels occlus précédemment identifiés.
- [0128] Par exemple, sur la [Fig.5], le pixel en sixième position ($x(p) = 6$) est un pixel irrégulier. En effet, son indice de colonne d'arrivée est égal à 3 alors que l'indice de colonne d'arrivée du pixel qui le précède ($x(p) = 5$) est égal à 5.
- [0129] Les pixels masqués ou occlus situés à gauche du pixel irrégulier sont alors identifiés, il s'agit des pixels en troisième, quatrième et cinquième position. En effet, leurs indices de colonne d'arrivée dans la troisième image sont respectivement supérieurs ou égaux à l'indice de colonne d'arrivée du pixel en sixième position dans la matrice. Ainsi, si

$V(p)$ représente la visibilité d'un pixel p dans la deuxième image, alors $V(p)=0$ pour les pixels masqués précédemment identifiés.

[0130] A l'inverse $V(p)=1$ pour les pixels p de la ligne visibles dans la deuxième image.

[0131] Dans une étape 38, un troisième masque de visibilité est déterminé comme étant l'association du premier masque de visibilité et du deuxième masque de visibilité.

[0132] Ainsi, ce troisième masque de visibilité prend en considération l'ensemble des pixels de la première image associés à des points de la scène tridimensionnelle qui se situent en dehors du champ de vision de la deuxième caméra 12 et l'ensemble des pixels de la première image occlus.

[0133] L'avantage d'une telle définition d'un masque de visibilité est de le déterminer sans calcul additionnel, des reprojections et projections étant déjà utilisées pour l'apprentissage et des profondeurs étant déjà prédites par le système de vision stéréoscopique. Cette solution permet également de se passer de calcul de flux optique souvent utilisé pour ce type d'application et nécessitant beaucoup de ressources pour des calculs.

[0134] Cette définition d'un masque de visibilité permet d'identifier les pixels visibles dans la première image et non visibles dans la seconde image.

[0135] L'utilisation de ce troisième masque de visibilité lors de l'apprentissage du réseau de neurones convolutif permet d'améliorer la pertinence de la définition de paramètres d'entrée de ce réseau de neurones convolutif, l'apprentissage est ainsi plus efficace.

[0136] Si l'ADAS utilise des données d'entrée telles que les profondeurs déterminées par le système de vision stéréoscopique 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 si la profondeur prédite est fiable lorsque le pixel est bien visible dans les première et deuxième images.

[0137] Bien entendu, la présente invention ne se limite pas aux exemples de réalisation décrits ci-avant mais s'étend à un procédé de détermination d'un masque de visibilité pour un système de vision embarqué dans un véhicule, 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.

[0138] 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.3].

Revendications

[Revendication 1]

Procédé de détermination de masques de visibilité par un système de vision stéréoscopique embarqué dans un véhicule (10), le système de vision stéréoscopique comprenant un ensemble de caméras d'au moins deux caméras (11, 12) disposées de manière à acquérir chacune une image d'une scène tridimensionnelle selon un point de vue différent, ladite deuxième caméra (12) se situant à droite de ladite première caméra (11) d'un point de vue de ladite première caméra (11), ledit procédé étant caractérisé en ce qu'il comprend les étapes suivantes :

- réception (31, 32) de premières et deuxièmes données respectivement représentatives d'une première et deuxième images acquises par respectivement une première et deuxième caméras (11, 12) dudit ensemble de caméras à un même instant temporel d'acquisition ;
- prédiction (33) de profondeurs associées à un ensemble de pixels de la première image par ledit système de vision stéréoscopique à partir d'un modèle de prédiction appris, chaque pixel de la première image ayant des coordonnées principales (x,y) dans la première image ;
- reprojection (34) dans la scène tridimensionnelle dudit ensemble de pixels sous forme d'un ensemble de points en fonction desdites profondeurs, d'une matrice intrinsèque de la première caméra (11) et de paramètres extrinsèques dudit système de vision stéréoscopique ;
- détermination (35) d'un premier masque de visibilité associé aux pixels dudit ensemble de pixels en fonction des coordonnées de points dudit ensemble de points, un pixel dudit ensemble de pixels étant non visible dans ladite deuxième image si les coordonnées d'un point dudit ensemble de points associé audit pixel le situent en dehors d'un champ de vision de ladite deuxième caméra (12) déterminé en fonction d'une largeur de la deuxième image et d'une distance focale de la deuxième caméra (12) ;
- génération (36) d'une troisième image par projection dudit ensemble de points en fonction d'une matrice intrinsèque de ladite deuxième caméra (12), des indices de colonne d'arrivée (i) et des indices de ligne d'arrivée (j) dans ladite troisième image étant associées à chaque pixel dudit ensemble de pixels ;
- pour chaque ligne de ladite première image rectifiée, balayage de gauche à droite, selon un point de vue de la première caméra (11), de valeurs d'indices de colonne d'arrivée (i) et détection d'un ensemble de

pixels irréguliers dont l'indice de colonne d'arrivée (i') ne suit pas une fonction monotone représentative d'une évolution d'indices de colonne d'arrivée (i) en fonction d'un indice de colonne (x) dans la première image rectifiée, et

pour chaque pixel irrégulier dudit ensemble, identification d'un ensemble de pixels occlus dans ladite ligne à gauche dudit chaque pixel irrégulier dont un indice de colonne d'arrivée (i) est supérieur ou égal à un indice de colonne d'arrivée (i') dudit chaque pixel irrégulier,

un deuxième masque de visibilité étant déterminé (37) comme une union desdits ensembles des pixels occlus ; et

- détermination (38) d'un troisième masque de visibilité par association dudit premier masque de visibilité et dudit deuxième masque de visibilité.

[Revendication 2]

Procédé selon la revendication 1, pour lequel la reprojection d'un pixel dans la scène tridimensionnelle est réalisée à l'aide de la formule suivante :

$$p_{3D}(p_t) = T \cdot \phi(p_t | K^{-1}, D(p_t)) \text{ avec :}$$

- $p_{3D}(p_t)$ les coordonnées dans la scène tridimensionnelle du point

issu de la reprojection du pixel p_t de la première image,

- T une matrice de déplacement entre une position de la première caméra (11) et une position de la deuxième caméra (12),

- K la matrice intrinsèque de la première caméra (11) associée à une projection d'un point de la scène tridimensionnelle dans une image acquise par la première caméra (11),

- ϕ une fonction de reprojection dans la scène tridimensionnelle d'un pixel en fonction de sa profondeur,

- $D(p_t)$ est une profondeur du pixel p_t de la première image prédite par le système de vision stéréoscopique.

[Revendication 3]

Procédé selon l'une des revendications 1 à 2, pour lequel la projection d'un point de la scène tridimensionnelle est réalisée à l'aide de la formule suivante :

$$p_s = \pi(K' [p_{3D}(p_t)]) \text{ avec : - } \pi \text{ une fonction pour passer de co-}$$

ordonnées homogènes dans l'espace à trois dimensions à des coordonnées pixels en deux dimensions en supprimant une dimension d'un vecteur,

- K' la matrice intrinsèque de la deuxième caméra (12) associée à une

projection d'un point de la scène tridimensionnelle dans une image acquise par la deuxième caméra (12),

- $p_{3D}(p_t)$ les coordonnées dans la scène tridimensionnelle du point issu de la projection du pixel p_t de la première image.

[Revendication 4] Procédé selon l'une des revendications 1 à 3, pour lequel ledit premier masque de visibilité est obtenu à l'aide de la formule suivante :

$$\bigcup_p p_t \text{ tel que } \frac{x_{p_{3D}(p_t)}}{D(p_t)} > \frac{W/2}{f} \text{ avec :}$$

- $x_{p_{3D}(p_t)}$ l'abscisse d'un point issu de la projection du pixel p_t de la première image, un axe des abscisses étant défini parallèlement à un axe suivant lequel se situent la première et la deuxième caméras (11, 12),

- $D(p_t)$ ladite profondeur du pixel p_t de la première image prédite par le système de vision stéréoscopique,

- W la largeur de la deuxième image, et

- f la distance focale de la deuxième caméra (12).

[Revendication 5] Procédé selon l'une des revendications 1 à 4, pour lequel lesdites profondeurs sont prédites par un réseau de neurones convolutif.

[Revendication 6] Procédé selon la revendication 5, pour lequel le réseau de neurones convolutif est entraîné pour minimiser une erreur photométrique définie par la fonction de perte suivante :

$$L(p) = \sum_p \left[(1-\alpha) \cdot |I(p) - \hat{I}(p)| + \alpha \cdot \left(1 - \frac{1}{2} SSIM(I(p), \hat{I}(p)) \right) \right]$$

Avec :

- $I(p)$ une valeur du pixel p dans la deuxième image ;

- $\hat{I}(p)$ une valeur du pixel p dans la troisième image ;

- $SSIM$ une fonction qui prend en compte une structure locale ; et

- α un facteur de pondération dépendant d'un type d'environnement routier.

[Revendication 7] Programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé selon l'une quelconque des revendications précédentes, lorsque ces instructions sont exécutées par un processeur.

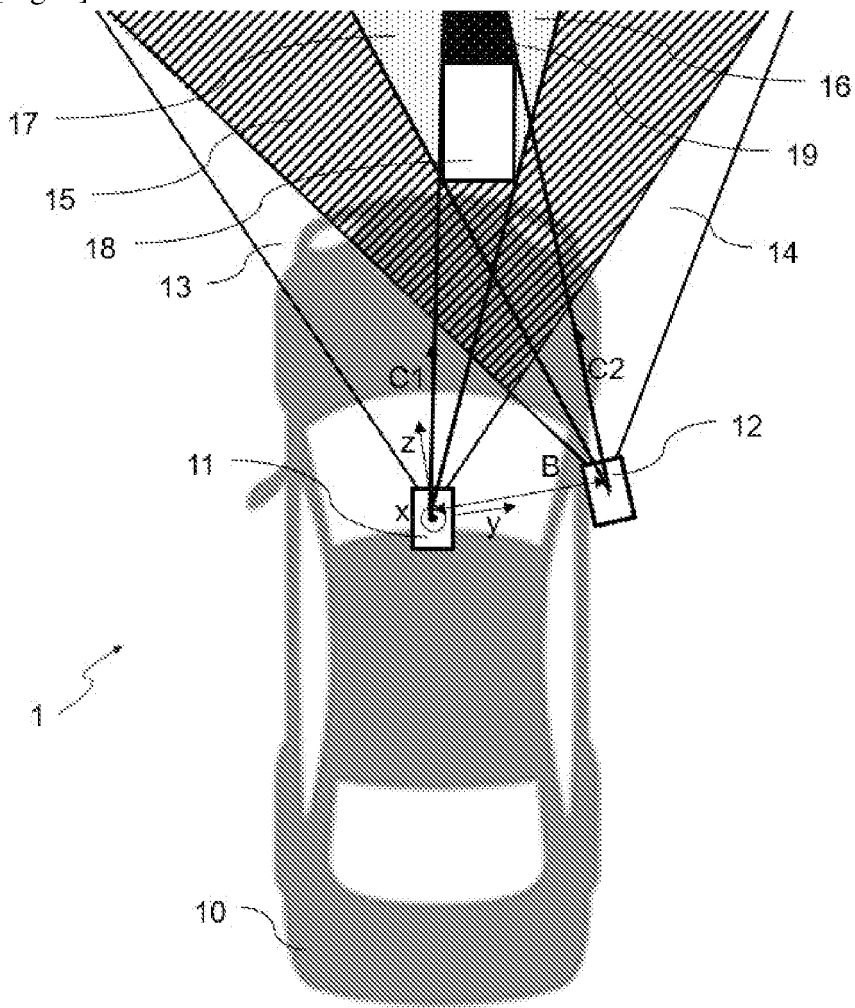
[Revendication 8] Dispositif (4) de détermination d'un masque de visibilité pour un système de vision embarqué dans un véhicule (10), ledit dispositif (4) comprenant une mémoire (41) associée à au moins un processeur (40) configuré pour la mise en œuvre des étapes du procédé selon l'une quelconque des revendications 1 à 6.

[Revendication 9] Système de détermination d'un masque de visibilité pour un système de

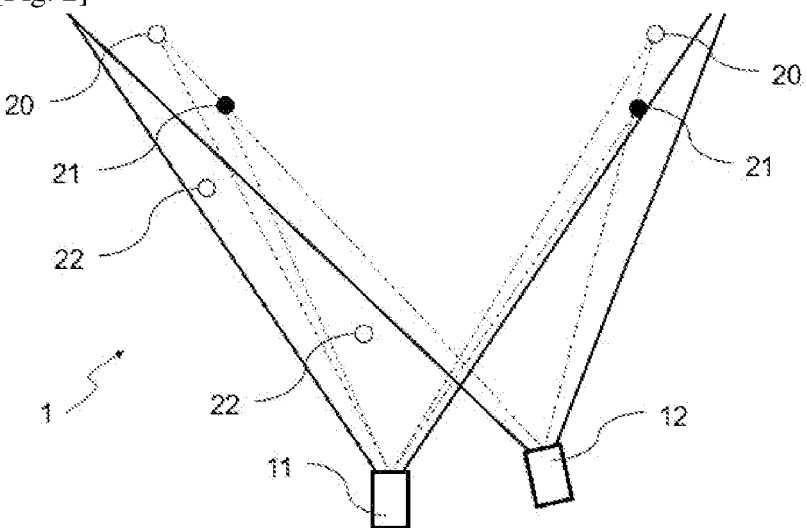
vision embarqué dans un véhicule (10) comprenant au moins deux caméras (11, 12) et un dispositif selon la revendication 7.

[Revendication 10] Véhicule (10) comprenant le dispositif (4) selon la revendication 8 ou le système selon la revendication 9.

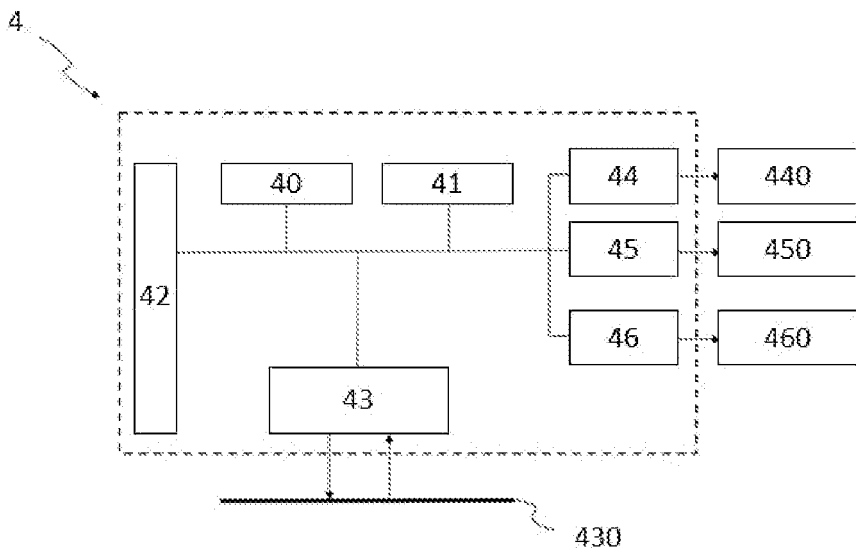
[Fig. 1]



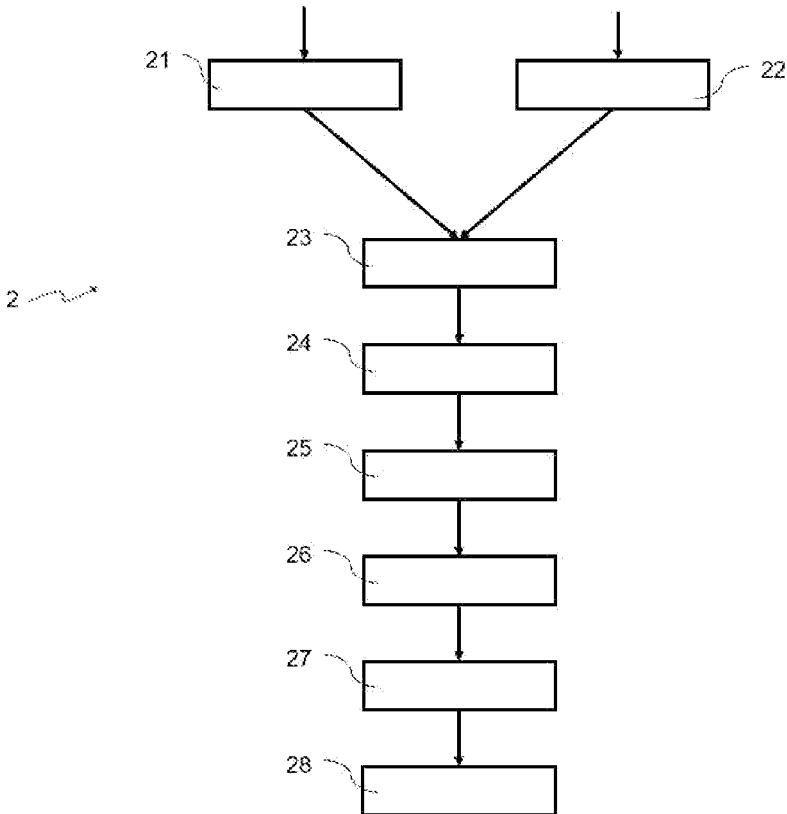
[Fig. 2]



[Fig. 3]



[Fig. 4]



[Fig. 5]

	i					\bar{i}		
$x(p)$	1	2	3	4	5	6	7	8
$i(p)$	1	2	3	4	5	3	4	5
$V(p)$	1	1	0	0	0	1	1	1

**RAPPORT DE RECHERCHE
PRÉLIMINAIRE**

N° d'enregistrement
national

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

FA 919817
FR 2304517

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		
A	<p>GUANGMING WANG ET AL: "Unsupervised Learning of Depth, Optical Flow and Pose with Occlusion from 3D Geometry", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 2 mars 2020 (2020-03-02), XP081744385, DOI: 10.1109/TITS.2020.3010418 * Abstract. Sec. III-A, III-B. Eq (1). Fig. 1. *</p> <p style="text-align: center;">-----</p>	1-10	<p>B60R 1/00 B60W 40/00 B60W 50/00 G06N 20/00 G06N 3/00 G06T 7/00 H04N 13/00</p>
A	<p>LI XIANGYU ET AL: "DVONet: Unsupervised Monocular Depth Estimation and Visual Odometry", 2019 IEEE VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP), IEEE, 1 décembre 2019 (2019-12-01), pages 1-4, XP033693910, DOI: 10.1109/VCIP47243.2019.8965952 [extrait le 2020-01-21] * le document en entier *</p> <p style="text-align: center;">-----</p>	1-10	<p>DOMAINES TECHNIQUES RECHERCHÉS (IPC)</p> <p>G06T</p>
A	<p>SAAD IMRAN ET AL: "Increased-Range Unsupervised Monocular Depth Estimation", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 23 juin 2020 (2020-06-23), XP081708363, * le document en entier *</p> <p style="text-align: center;">-----</p>	1-10	
Date d'achèvement de la recherche		Examineur	
18 décembre 2023		Nicolau, Stephane	
<p>CATÉGORIE DES DOCUMENTS CITÉS</p> <p>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</p> <p>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</p>			