

19 RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
COURBEVOIE

11 N° de publication :
(à n'utiliser que pour les
commandes de reproduction)

3 153 677

21 N° d'enregistrement national : 23 10306

51 Int Cl⁸ : G 06 T 7/593 (2023.01), G 06 V 10/82, 20/56

12 DEMANDE DE BREVET D'INVENTION

A1

22 Date de dépôt : 28.09.23.

30 Priorité :

43 Date de mise à la disposition du public de la
demande : 04.04.25 Bulletin 25/14.

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.

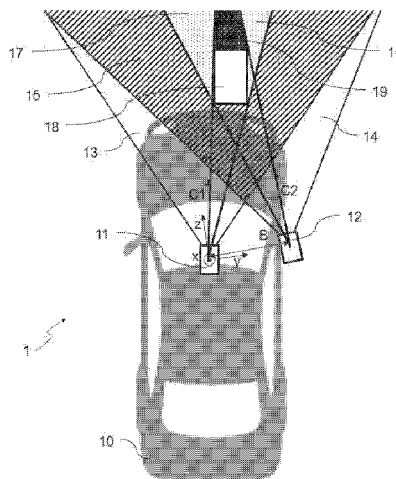
73 Titulaire(s) : STELLANTIS AUTO SAS Société par
actions simplifiée.

54 Mandataire(s) : dispositif de détermination d'une
profondeur par un système de vision
stéréoscopique auto-supervisé.

57 Procédé ou dispositif mettant en œuvre un procédé de détermination d'une profondeur par 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 selon un point de vue

différent, les axes optiques représentatifs d'une orientation du champ de vision de chaque caméra étant orientés de façon non parallèle, le procédé étant caractérisé en ce qu'il comprend les étapes de recadrage d'images acquises par le système de vision stéréoscopique, des prédictions de profondeurs par un réseau de neurones convolutif pour des pixels de chaque image, la génération d'images en fonction des profondeurs prédites et un calcul d'erreur par comparaison des images générées et des images acquises, le réseau de neurones étant appris par minimisation d'une erreur déterminée à partir d'une moyenne des erreurs précédemment déterminées.

Figure pour l'abrégé : Figure 1



Description

Titre de l'invention : Procédé et dispositif de détermination d'une profondeur par un système de vision stéréoscopique auto-supervisé.

Domaine technique

[0001] La présente invention concerne les procédés et dispositifs de détermination d'une profondeur par système de vision stéréoscopique embarqué dans un véhicule, par exemple dans un véhicule automobile. La présente invention concerne également un procédé et un dispositif de mesurage d'une telle profondeur. La présente invention concerne également un procédé et un dispositif de contrôle d'un ou plusieurs systèmes ADAS embarqués dans un véhicule à partir de la profondeur déterminée.

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.

Résumé de la présente invention

[0006] Un objet de la présente invention est de résoudre au moins l'un des problèmes de l'arrière-plan technologique décrit précédemment.

[0007] Un autre objet de la présente invention est d'améliorer la qualité des données issues

de ces caméras.

[0008] 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'au moins une caméra.

[0009] Selon un premier aspect, la présente invention concerne un procédé de détermination d'une profondeur 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,

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 ;

- obtention d'une première image recadrée par recadrage de la première image et obtention d'une deuxième image recadrée par recadrage de la deuxième image, un recadrage des première et deuxième images étant fonction de champs de vision des première et deuxième caméras ;

- prédiction de premières profondeurs associées à un premier ensemble de pixels de la première image recadrée à partir du premier ensemble de pixels et d'un deuxième ensemble de pixels de la deuxième image recadrée correspondant au premier ensemble de pixels et prédiction de deuxièmes profondeurs associées à un troisième ensemble de pixels de la deuxième image recadrée à partir du troisième ensemble de pixels et d'un quatrième ensemble de pixels de la première image recadrée correspondant au troisième ensemble de pixels,

les premières et deuxièmes profondeurs étant prédites par une méthode de flux optique mise en œuvre par un réseau de neurones convolutif ;

- génération d'une troisième image à partir de la première image recadrée et des premières profondeurs et génération d'une quatrième image à partir de la deuxième image recadrée et des deuxièmes profondeurs, les générations étant obtenues à partir de la formule suivante :

$$p_s = \pi \left(K \left[F \phi \left(p_t | K^{-1}, D_{st}(p_t) \right) \right] \right)$$

Avec :

- p_s un pixel de la troisième image, respectivement quatrième image,
- π une fonction pour passer de coordonnées homogènes à des coordonnées pixels en supprimant une dimension d'un vecteur,
- K une matrice intrinsèque de la deuxième caméra, respectivement de la première caméra, associée à la projection d'un point de la scène tridimensionnelle aux co-

ordonnées en 3 dimensions dans une image aux coordonnées en 2 dimensions,

- F une matrice extrinsèque du système de vision stéréoscopique,
- $\emptyset$ une fonction de projection dans la scène tridimensionnelle d'un pixel en fonction de sa profondeur,
- K' une matrice intrinsèque de la première caméra, respectivement de la deuxième caméra, associée à la projection d'un point de la scène aux coordonnées en 3 dimensions dans l'image aux coordonnées en 2 dimensions,
- $D_{st}(p_i)$ est une première profondeur du pixel p_i de la première image recadrée, respectivement une deuxième profondeur du pixel p_i de la deuxième image recadrée ;
- détermination d'une première erreur par comparaison des troisième image et deuxième image recadrée et détermination d'une deuxième erreur par comparaison des quatrième image et première image recadrée ;
- ajustement de paramètres d'entrée dudit réseau de neurones convolutif par minimisation d'une troisième erreur, ladite troisième erreur étant déterminée à partir d'une moyenne desdites première et deuxième erreurs.

[0010] Un tel procédé permet un entraînement du réseau de neurones convolutif à partir des images acquises par les caméras du système de vision stéréoscopique sans avoir à utiliser des données annotées. Un tel entraînement permet de qualifier le réseau de neurones convolutif comme « auto-supervisé », c'est-à-dire qu'il est entraîné à partir de données générées par le système de vision stéréoscopique. Les profondeurs prédites par ce système de vision stéréoscopique sont alors fiabilisées.

[0011] Selon une variante du procédé, la troisième erreur est déterminée par la fonction de perte suivante :

$$L'_s = \sum_p \text{avg}(L_{st}(p), L_{ss}(p)) \text{ Avec :}$$

- L'_s la troisième erreur,
- avg une moyenne d'arguments,
- $L_{st}(p)$ la première erreur pour un pixel p de la deuxième image recadrée,
- $L_{ss}(p)$ la deuxième erreur pour un pixel p de la première image recadrée.

[0012] Selon une autre variante du procédé, les première et respectivement deuxième erreurs sont déterminées par la fonction de perte suivante :

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

Avec :

- $L_*(p)$ la première, respectivement deuxième erreur,
- $I(p)$ une valeur du pixel p dans la deuxième image recadrée, respectivement la première image recadrée,
- $\hat{I}(p)$ une valeur du pixel p dans la troisième image, respectivement la quatrième

image'

- SSIM une fonction qui prend en compte une structure locale, et
- α un facteur de pondération dépendant notamment du type d'environnement.

[0013] Selon encore une variante du procédé, les première et respectivement deuxième erreurs sont déterminées par la fonction de perte suivante :

$$L_{smooth}(D_{st}(p_t), W, o) = \sum_{p_t} \sum_{d \in x, y} (W(p_t) |\nabla_d^o D_{st}(p_t)| e^{\beta |\nabla_d^2 I_t(p_t)|})$$

Avec :

- $L_{smooth}(O, W, o)$ la première, respectivement deuxième erreur,
- $D_{st}(p_t)$ est la première profondeur, respectivement deuxième profondeur D_{tm}

d'un pixel p_t prédite par le système de vision stéréoscopique ;

- W est une matrice de paramètres ;
- o est l'ordre d'un gradient de lissage ;
- une norme L1 des gradients de profondeur du second ordre est calculée avec $W=1$, et $o=2$;
- x et y sont les dimensions de la deuxième image, respectivement première image ;
- β est un hyperparamètre dépendant de l'environnement ; et
- $I_t(p_t)$ est la valeur du pixel p_t dans la troisième image, respectivement quatrième image.

[0014] Selon une variante du procédé supplémentaire, le recadrage est réalisé de manière à obtenir au moins 50% des pixels de chacune desdites première et deuxième images recadrées associés à des objets de la scène tridimensionnelle présents dans des champs de vision des première et deuxième caméras.

[0015] Selon une autre variante de procédé, les premières et deuxièmes profondeurs sont déterminées par une méthode dite de flux optique.

[0016] Selon encore une variante, le procédé comprend une étape de détermination d'une disparité induite par une position relative des première (11) et deuxième (12) caméras, les prédictions de premières et deuxièmes profondeurs étant, en outre, fonction de la disparité induite.

[0017] Selon un deuxième aspect, la présente invention concerne un dispositif de détermination d'une profondeur par un système de vision stéréoscopique 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.

[0018] 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.

[0019] 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.

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

- [0025] 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 :
- [0026] [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 ;
- [0027] [Fig.2] 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 ;
- [0028] [Fig.3] illustre schématiquement des images obtenues à partir d'un système de vision stéréoscopique embarqué dans le véhicule de la [Fig.1], selon un exemple de réa-

lisation particulier et non limitatif de la présente invention ;

- [0029] [Fig.4] 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.

Description des exemples de réalisation

- [0030] Un procédé et un dispositif de détermination d'une profondeur par système de vision stéréoscopique 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.
- [0031] 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.
- [0032] Selon un exemple particulier et non limitatif de réalisation de la présente invention, un procédé de détermination d'une profondeur par 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 stéréo.
- [0033] Le système stéréo 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, les axes optiques représentatifs d'une orientation du champ de vision de chaque caméra étant orientés de façon non parallèle. Le système de vision stéréoscopique est ainsi dit non parallèle.
- [0034] A cet effet, le procédé de détermination d'une profondeur 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 par respectivement une première et deuxième caméras de l'ensemble de caméras à un même instant temporel d'acquisition, le recadrage des première et deuxième images en fonction de champs de vision des première et deuxième caméras, la prédiction de premières profondeurs associées à un premier ensemble de pixels de la première image et la prédiction de deuxièmes profondeurs associées à un troisième ensemble de pixels de la deuxième image, les premières et deuxièmes profondeurs étant prédites par un réseau de neurones convolutif.
- [0035] Le procédé comprend également la génération d'une troisième image à partir de la première image et des premières profondeurs et la génération d'une quatrième image à partir de ladite deuxième image et des deuxièmes profondeurs, la détermination d'une première erreur par comparaison des troisième et deuxième images et la détermination

d'une deuxième erreur par comparaison des quatrième et première images.

- [0036] Des paramètres d'entrée du réseau de neurones convolutif sont alors ajustés par minimisation d'une troisième erreur, la troisième erreur étant déterminée à partir d'une moyenne des première et deuxième erreurs.
- [0037] 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.
- [0038] Un tel environnement 1 correspond, par exemple, à un environnement routier formé d'un réseau de routes accessibles pour le véhicule 10.
- [0039] 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.
- [0040] Le véhicule 10 comprend avantageusement plusieurs caméras 11, 12 embarquées, chacune configurée pour acquérir des images d'une scène dans l'environnement du véhicule 10. Cet ensemble de caméras 11, 12 forme le système stéréo. Deux caméras 11 et 12 sont illustrées sur la [Fig.1]. La présente invention ne se limite cependant pas à un système stéréo comprenant deux caméras mais s'étend à tout système stéréo comprenant 2 ou plus caméras, par exemple 2, 3, 4 ou 5 caméras.
- [0041] Les deux caméras 11, 12 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.
- [0042] 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.
- [0043] 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 dis-

torsions 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.

[0044] Ces deux caméras 11, 12 sont disposées de manière à acquérir chacune une image d'une scène 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.

[0045] 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.

[0046] 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.

[0047] Les paramètres extrinsèques sont déterminés, en outre, lors d'une phase de calibration du système de vision stéréoscopique.

[0048] 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, la base de référence doit atteindre 60cm pour les caméras couramment utilisées dans ce domaine.

[0049] Les deux caméras 11, 12 font l'acquisition d'images d'une scène 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 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 deux caméras 11, 12.

- [0050] 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 stéréo composé ici des trois champs 16, 17 et 19.
- [0051] 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.
- [0052] 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.
- [0053] 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.
- [0054] 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.
- [0055] 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 orientés horizontalement, c'est-à-dire que chacune des première (11) et deuxième (12) caméras dispose d'un axe optique horizontal. En d'autres termes, les deux axes optiques sont chacun disposés dans un plan horizontal, les deux plans supportant chacun un axe optique étant parallèles.
- [0056] Il est évident qu'il est possible d'utiliser un tel système de vision stéréoscopique pour prendre des images de scènes situées sur les côtés ou derrière le véhicule 10 en l'équipant de caméras placées et orientées différemment.
- [0057] 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 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).
- [0058] Les images acquises par les caméras 11, 12 représentent des vues d'une même scène prises à des points de vue différents, les positions des caméras étant distinctes. Sur cette scène 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.

- [0059] 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.
- [0060] Un processus de détermination d'une profondeur par un système de vision stéréoscopique 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 stéréoscopique du véhicule 10.
- [0061] Dans une première opération, le calculateur reçoit des premières données représentatives d'une première image 31a acquise par la première caméra 11 à un instant temporel d'acquisition.
- [0062] Dans une deuxième opération, le calculateur reçoit des deuxièmes données représentatives d'une deuxième image 32a acquise par la deuxième caméra 12 au même instant temporel d'acquisition.
- [0063] Les champs de vision de la première et de la deuxième caméra étant concourants, une partie des pixels de la première image et une partie des pixels de la deuxième image sont associés à des mêmes objets présents dans la scène tridimensionnelle.
- [0064] La prédiction d'une profondeur d'un pixel d'une image acquise par une caméra du système de vision stéréoscopique via un algorithme associé à une vision stéréoscopique se fait pour un pixel de la première image ayant un pixel associé dans la deuxième image, c'est-à-dire que la prédiction de la distance d'un objet de la scène tridimensionnelle par un algorithme associé à une vision stéréoscopique est fiable si l'objet est présent dans le champ de vision commun des première 11 et deuxième 12 caméras, le troisième champ d'acquisition 15.
- [0065] Afin de garder des parties des première 31a et deuxième 32a images associées au troisième champ d'acquisition 15, une première image recadrée 31b est obtenue dans une troisième opération par recadrage de la première image 31a et une deuxième image recadrée 32b est obtenue dans une quatrième opération par recadrage de la deuxième image 32a. Le recadrage des première et deuxième images est donc fonction des champs de vision des première 11 et deuxième 12 caméras.
- [0066] Selon un exemple particulier de réalisation illustré par la [Fig.3], la première image recadrée 31b est obtenue par rognage d'une partie gauche 311 de la première image 31a et d'une partie inférieure 312 de cette première image 31a.

- [0067] Selon cet exemple particulier de réalisation, la deuxième image recadrée 32b est obtenue par rognage d'une partie droite 321 de la deuxième image 32a et d'une partie supérieure 322 de cette deuxième image 32a.
- [0068] Bien entendu, les parties des première 31a et deuxième 32a images rognées dépendent de la position relative et de l'orientation relative des première 11 et deuxième 12 caméras.
- [0069] La première image 31a est par exemple rognée dans sa partie droite et la deuxième image 32a rognée dans sa partie gauche si les positions des première 11 et deuxième 12 caméras sont inversées.
- [0070] On définit le centre optique d'une image par ses coordonnées dans l'image, l'origine de l'image étant par exemple située en haut à gauche de l'image.
- [0071] Le centre optique de la première image 31a a donc pour coordonnées :
- [0072] (
- [Math.1]
- $$c_x^{31a}, c_y^{31a} = \left(\frac{w}{2}, \frac{h}{2} \right)$$
- [0073] Avec w la largeur de la première image 31a et h la hauteur de la première image 31a.
- [0074] Si les première 31a et deuxième 32a images ont la même dimension, soit une largeur w et une hauteur h, les centres optiques des première 31a et deuxième 31b images ont tous deux pour coordonnées (w/2, h/2).
- [0075] Les différentes parties des première 31a et deuxième 32a images ont pour centres optiques respectifs :
- [0076] [Math.2]
- $$\begin{cases} c_x^{31b} = w^{31b} - \frac{w}{2} \\ c_y^{31b} = \frac{h}{2} \end{cases},$$
- [0077]
$$\begin{cases} c_x^{32b} = \frac{w}{2} \\ c_y^{32b} = h^{32b} - \frac{h}{2} \end{cases}$$
- [0078]
$$\begin{cases} c_x^{311} = \frac{w}{2} \\ c_y^{311} = \frac{h}{2} \end{cases}$$
- [0079]
$$\begin{cases} c_x^{312} = w^{312} - \frac{w}{2} \\ c_y^{312} = h^{312} - \frac{h}{2} \end{cases}$$
- [0080]
$$\begin{cases} c_x^{322} = \frac{w}{2} \\ c_y^{322} = \frac{h}{2} \end{cases}$$

$$[0081] \quad \begin{cases} c_x^{321} = w^{321} - \frac{w}{2} \\ c_y^{321} = \frac{h}{2} \end{cases}$$

[0082] L'indice x représente les abscisses suivant le sens horizontal, l'indice y les ordonnées suivant le sens vertical.

[0083] Les exposants représentent la partie d'une image tel que représenté par l'exemple illustré en [Fig.3].

[0084] Les centres optiques permettent alors de déterminer les matrices intrinsèques K liées aux différentes parties des images, la matrice intrinsèque K étant définie par :

[0085] [Math.3]

$$K = \begin{bmatrix} f_x & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix}$$

[0086] Avec :

- K une matrice intrinsèque d'une image ou partie d'image,
- f_x la distance focale en unité pixel dans le sens horizontal de l'image ou partie d'image,
- f_y la distance focale en unité pixel dans le sens vertical de l'image ou partie d'image,
- c_x l'abscisse du centre optique de l'image ou partie d'image,
- c_y l'ordonnée du centre optique de l'image ou partie d'image.

[0087] Selon un exemple de réalisation particulier, le recadrage est réalisé de manière à obtenir au moins 50%, 60% ou 80% des pixels de chacune des première 31b et deuxième 32b images recadrées associés à des objets de la scène tridimensionnelle présents dans des champs de vision des première 11 et deuxième 12 caméras.

[0088] Dans une cinquième opération, des premières profondeurs associées à un premier ensemble de pixels de la première image recadrée 31b sont prédites à partir du premier ensemble de pixels et d'un deuxième ensemble de pixels de la deuxième image recadrée 32b correspondant au premier ensemble de pixels.

[0089] Dans une sixième opération, deuxièmes profondeurs associées à un troisième ensemble de pixels de la deuxième image recadrée 32b sont prédites à partir du troisième ensemble de pixels et d'un quatrième ensemble de pixels de la première image recadrée 31b correspondant au troisième ensemble de pixels.

[0090] Les premières et deuxièmes profondeurs sont prédites par une méthode de flux optique par un réseau de neurones convolutif.

[0091] Ainsi, les cinquième et sixième opérations se font par mise en œuvre d'une méthode dite de calcul par flux optique (en anglais « optical flow »). Une telle méthode est notamment décrite dans « PWC-Net: CNNs for Optical Flow Using Pyramid, Warping,

and Cost Volume » de Deqing Sun, Xiaodong Yang, Ming-Yu Liu et Jan Kautz du 25 juin 2018 et dans « RAFT: Recurrent All-Pairs Field Transforms for Optical Flow » de Zachary Teed et Jia Deng du 25 août 2020.

- [0092] La méthode de calcul par flux optique est réalisée par un réseau de neurones convolutif (en anglais « Convolutional Neural Networks », CNN). Ce type d'outil est couramment utilisé dans le cadre du traitement d'images.
- [0093] Une donnée intermédiaire de la cinquième opération est un flux optique représentatif d'un déplacement entre chaque pixel du premier ensemble de pixels et un deuxième pixel correspondant à chaque premier pixel dans la deuxième image recadrée 32b, l'ensemble des deuxièmes pixels correspondant au deuxième ensemble de pixels.
- [0094] De même, une donnée intermédiaire de la sixième opération est un flux optique représentatif d'un déplacement entre chaque pixel du troisième ensemble de pixels et un quatrième pixel correspondant à chaque troisième pixel dans la première image recadrée 31b, l'ensemble des quatrièmes pixels correspondant au quatrième ensemble de pixels.
- [0095] En effet, le calcul par flux optique permet de traiter les première 31b et deuxième 32b images recadrées dans les deux sens sans opération additionnelle.
- [0096] Les premières et deuxièmes profondeurs sont alors déterminées à partir de ces flux optiques, une profondeur étant proportionnelle à une norme d'un flux optique.
- [0097] La rotation et translation pour passer de l'axe optique de la première caméra 11 à la deuxième caméra 12 créent notamment une disparité dite « induite ». Cette disparité est, en outre, incluse dans le flux optique déterminé à partir des première 31b et deuxième 32b images recadrées.
- [0098] La rotation est par exemple due au fait que les axes optiques des première 11 et deuxième 12 caméras ne sont pas parallèles. La translation est par exemple due au fait que les première 11 et deuxième 12 caméras ne sont pas placées à une même hauteur sur le véhicule 10.
- [0099] Selon un exemple de réalisation particulier, afin de ne pas prendre en compte cette disparité induite susceptible de fausser la prédiction de premières et deuxièmes profondeurs, un réseau de neurones convolutif de type FCN (de l'anglais « Fully Convolutional network »), par la suite appelé un FCN, est mis en œuvre pour déterminer la disparité induite et filtrer la valeur de flux optique avant de prédire la profondeur d'un pixel.
- [0100] Ainsi, selon cet exemple de réalisation particulier, dans une septième opération, une première disparité induite est déterminée par le FCN à partir des première 31b et deuxième 32b images recadrées permettant de soustraire la première disparité induite du flux optique pour obtenir un flux optique corrigé, les premières profondeurs sont alors prédites en fonction de ce flux optique corrigé.

- [0101] De manière similaire, dans une huitième opération, une deuxième disparité induite est déterminée par le FCN à partir des deuxième 32b et première 31b images recadrées permettant de soustraire la deuxième disparité induite du flux optique pour obtenir un flux optique corrigé, les deuxième profondeurs sont alors prédites en fonction de ce flux optique corrigé.
- [0102] La donnée d'entrée d'un tel FCN est une matrice 'une dimension HxWxN avec H la hauteur d'une image, W la largeur d'une image et N un nombre de canaux incluant les éléments non nuls de la matrice de rotation (au maximum 9), les éléments non nuls du vecteur de translation (au maximum 3) ainsi que les deux flux optiques (horizontaux et verticaux). Le nombre maximal de canaux est donc de 14. Le nombre de couches intermédiaire et le nombre des canaux de la matrice à chaque couche dépend de la déviation du système parallèle.
- [0103] La structure de ce FCN est construite par expérimentation de manière à minimiser la taille du FCN tout en offrant un résultat acceptable.
- [0104] Dans une neuvième opération, une troisième image est générée à partir de la première image recadrée 31b et des premières profondeurs déterminées.
- [0105] De manière similaire, dans une dixième opération, une quatrième image est générée à partir de la deuxième image recadrée 32b et des deuxième profondeurs déterminées.
- [0106] Les générations des troisième et quatrième images sont obtenues à partir de la formule suivante :

[Math.4]

$$p_s = \pi \left(K \left[F \phi \left(p_t K^{-1}, D_{st}(p_t) \right) \right] \right)$$

- [0107] Avec :
- p_s un pixel de la troisième image, respectivement quatrième image,
 - π une fonction pour passer de coordonnées homogènes à des coordonnées pixels en supprimant une dimension d'un vecteur,
 - K une matrice intrinsèque de la deuxième caméra 12, respectivement de la première caméra 11, associée à la projection d'un point de la scène tridimensionnelle aux coordonnées en 3 dimensions dans une image aux coordonnées en 2 dimensions,
 - F une matrice extrinsèque du système de vision stéréoscopique,
 - ϕ une fonction de projection dans la scène tridimensionnelle d'un pixel en fonction de sa profondeur,
 - K' une matrice intrinsèque de la première caméra 11, respectivement de la deuxième caméra 12, associée à la projection d'un point de la scène aux coordonnées en 3 dimensions dans l'image aux coordonnées en 2 dimensions,
 - $D_{st}(p_t)$ est une première profondeur du pixel p_t de la première image recadrée 31b, respectivement une deuxième profondeur du pixel p_t de la deuxième image recadrée

32b.

[0108] Le principe de la génération d'image consiste à projeter un pixel d'une image acquise par la première caméra 11 vers un point dans la scène tridimensionnelle en fonction de sa première profondeur, puis à reprojeter ce point dans une image telle qu'elle serait acquise du point de vue de la deuxième caméra 12 et, respectivement, projeter un pixel d'une image acquise par la deuxième caméra 12 vers un point dans la scène tridimensionnelle en fonction de sa deuxième profondeur, puis à reprojeter ce point dans une image telle qu'elle serait acquise du point de vue de la première caméra 11.

[0109] Dans une onzième opération, une première erreur est déterminée par comparaison des troisième image et deuxième image recadrée 32b.

[0110] De manière similaire, dans une douzième opération, une deuxième erreur est déterminée par comparaison des quatrième image et première image recadrée 31b.

[0111] Selon un premier exemple de réalisation particulier, les première et respectivement deuxième erreurs sont déterminées par la fonction de perte suivante :

[Math.5]

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

[0112] Avec :

- $L_*(p)$ la première, respectivement deuxième erreur,
- $I(p)$ une valeur du pixel p dans la deuxième image recadrée (32b), respectivement la première image recadrée (31b),
- $\hat{I}(p)$ une valeur du pixel p dans la troisième image, respectivement la quatrième image,
- SSIM une fonction qui prend en compte une structure locale, et
- α un facteur de pondération dépendant notamment du type d'environnement.

[0113] Cette fonction de perte est basée sur l'erreur photométrique. Une fois une troisième image, ou respectivement quatrième image, générée, celle-ci est comparée à la véritable image acquise par le système de vision stéréoscopique, la deuxième image recadrée 32b ou respectivement la première image recadrée 31b.

[0114] Selon un deuxième exemple de réalisation particulier, les première et respectivement deuxième erreurs sont déterminées par la fonction de perte suivante :

[Math.6]

$$L_{smooth}(D_{st}(p_t), W, o) = \sum_p \sum_{d \in x,y} \left(W(p_t) |\nabla_d^o D_{st}(p_t)| e^{\beta |\nabla_d^2 I_t(p_t)|} \right)$$

[0115] Avec :

- $L_{smooth}(O, W, o)$ la première, respectivement deuxième erreur,
- $D_{st}(p_t)$ est la première profondeur, respectivement deuxième profondeur D_{tm} d'un pixel p_t prédite par le système de vision stéréoscopique ;

- W est une matrice de paramètres ;
- o est l'ordre d'un gradient de lissage ;
- une norme L1 des gradients de profondeur du second ordre est calculée avec $W=1$, et $o=2$;
- x et y sont les dimensions de la deuxième image, respectivement première image ;
- β est un hyperparamètre dépendant de l'environnement ; et
- $I_t(p_t)$ est la valeur du pixel p_t dans la troisième image, respectivement quatrième image.

- [0116] Cette deuxième fonction de perte est généralement utilisée pour traiter la discontinuité à la bordure des objets (en anglais « edge aware smoothness »).
- [0117] Selon un troisième exemple de réalisation particulier, les première et deuxième erreurs sont déterminées par une fonction combinant la fonction présentée dans le premier exemple de réalisation particulier et par la fonction présentée dans le deuxième exemple de réalisation particulier, par exemple en réalisant la somme des deux fonctions ou en appliquant l'une ou l'autre des fonctions selon une localisation des pixels des première 31b et deuxième 32b images recadrées.
- [0118] Les opérations permettant de prédire les premières et deuxième profondeurs sont mises en œuvre, en outre, par un réseau de neurones convolutif auto-supervisé. Cet apprentissage à partir des données obtenues du système de vision stéréoscopique permet d'affiner les paramètres du réseau de neurones convolutif et de fiabiliser ainsi les données intermédiaires telles que les disparités ou flux optiques ou de sortie telles que les premières et deuxième profondeurs.
- [0119] Un tel réseau de neurones convolutif est connu de l'homme du métier, comme PWCNet (de l'anglais « Pyramidal processing, Warping, and the use of a Cost volume Network », en français « Réseau de traitement pyramidal, déformation et utilisation d'un volume de coûts »).
- [0120] Selon un exemple de réalisation particulier, pour trouver les bonnes valeurs des paramètres d'entrée (appelés « weights » en anglais) d'un réseau de neurones convolutif choisi, les première et deuxième erreurs sont minimisées dans une treizième opération.
- [0121] Dans une quatorzième opération, les paramètres d'entrée du réseau de neurones convolutif sont ajustés par minimisation d'une troisième erreur, la troisième erreur étant déterminée à partir d'une moyenne des première et deuxième erreurs.
- [0122] Selon un exemple de réalisation particulier, la troisième erreur est déterminée par la fonction de perte suivante :

[Math.7]

$$L'_s = \sum_p \text{avg}(L_{st}(p), L_{ss}(p))$$

- [0123] Avec :

- L_s' la troisième erreur,
- avg une moyenne d'arguments,
- $L_{st}(p)$ la première erreur pour un pixel p de la deuxième image recadrée (32b),
- $L_{ss}(p)$ la deuxième erreur pour un pixel p de la première image recadrée (31b).

- [0124] La moyennisation des première et deuxième erreurs permet, en outre, de prendre en considération chacune des première et deuxième erreurs de manière équivalente.
- [0125] La [Fig.2] illustre un organigramme des différentes étapes d'un procédé 2 de détermination d'une profondeur par système de vision stéréoscopique embarqué dans le véhicule de la [Fig.1], selon un exemple de réalisation particulier et non limitatif de la présente invention.
- [0126] Le procédé 2 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 stéréo.
- [0127] Dans une première étape 21, des premières données représentatives d'une première 31a image acquise par une première caméra 11 sont reçues.
- [0128] Dans une deuxième étape 22, des deuxième données représentatives d'une deuxième 32a image acquise par une deuxième caméra 12 sont reçues.
- [0129] Les deux images reçues correspondent à deux vues d'une même scène se déroulant aux alentours du véhicule 10 à même instant temporel d'acquisition.
- [0130] Dans une troisième étape 23a, une première image recadrée 31b est obtenue par recadrage de la première image 31a.
- [0131] Dans une quatrième étape 23b, une deuxième image recadrée 32b est obtenue par recadrage de la deuxième image 32a.
- [0132] Un recadrage des première et deuxième images est fonction de champs de vision des première 11 et deuxième 12 caméras.
- [0133] Dans une cinquième étape 24a, des premières profondeurs associées à un premier ensemble de pixels de la première image recadrée 31b sont prédites à partir du premier ensemble de pixels et d'un deuxième ensemble de pixels de la deuxième image recadrée 32b correspondant au premier ensemble de pixels.
- [0134] Dans une sixième étape 24b, des deuxième profondeurs associées à un troisième ensemble de pixels de la deuxième image recadrée 32b sont prédites à partir du troisième ensemble de pixels et d'un quatrième ensemble de pixels de la première image recadrée 31b correspondant au troisième ensemble de pixels.
- [0135] Les premières et deuxième profondeurs sont prédites par une méthode de flux optique mise en œuvre par un réseau de neurones convolutif.
- [0136] Dans une septième étape 25a, une troisième image est générée à partir de la première image recadrée 31b et des premières profondeurs prédites.
- [0137] Dans une huitième étape 25b, une quatrième image est générée à partir de la deuxième

image recadrée 32b et des deuxièmes profondeurs.

[0138] Les générations d'images sont obtenues à partir de la formule suivante :

[Math.8]

$$p_s = \pi \left(K \left[F \phi \left(p_r K^{-1}, D_{st}(p_t) \right) \right] \right)$$

[0139] Avec :

- P_s un pixel de la troisième image, respectivement quatrième image,
- π une fonction pour passer de coordonnées homogènes à des coordonnées pixels en supprimant une dimension d'un vecteur,
- K une matrice intrinsèque de la deuxième caméra 12, respectivement de la première caméra 11,
- F une matrice extrinsèque du système de vision stéréoscopique,
- ϕ une fonction de projection dans la scène tridimensionnelle d'un pixel en fonction de sa profondeur,
- K' une matrice intrinsèque de la première caméra 11, respectivement de la deuxième caméra 12,
- $D_{st}(p_t)$ est une première profondeur du pixel p_t de la première image recadrée 31b, respectivement une deuxième profondeur du pixel p_t de la deuxième image recadrée 32b.

[0140] Dans une neuvième étape 26a, une première erreur est déterminée par comparaison des troisième image et deuxième image recadrée 32b.

[0141] Dans une dixième étape 26b, une deuxième erreur est déterminée par comparaison des quatrième image et première image recadrée 31b.

[0142] Dans une onzième étape 27, des paramètres d'entrée du réseau de neurones convolutif sont ajustés par minimisation d'une troisième erreur, la troisième erreur étant déterminée à partir d'une moyenne des première et deuxième erreurs

[0143] Ainsi, les prédictions des premières et deuxièmes profondeurs sont affinées, améliorant la fiabilité des données de sortie du système de vision stéréoscopique.

[0144] Un ADAS utilisant les premières ou deuxièmes profondeurs 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.

[0145] La [Fig.4] illustre schématiquement un dispositif 4 configuré pour la détermination

d'un masque de visibilité par 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.

- [0146] Le dispositif 4 est par exemple configuré pour la mise en œuvre des opérations décrites en regard des figures 1 et 3 et/ou des 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 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.
- [0147] 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.
- [0148] 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.
- [0149] 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.
- [0150] 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 »).

[0151] 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).

[0152] 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.

[0153] 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'une profondeur par système de vision stéréoscopique 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.

[0154] 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 de détermination d'une profondeur 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, ledit procédé étant caractérisé en ce qu'il comprend les étapes suivantes :

- réception (21, 22) de premières et deuxièmes données respectivement représentatives d'une première (31a) et deuxième (32a) images acquises par respectivement une première et deuxième caméras (11, 12) dudit ensemble de caméras à un même instant temporel d'acquisition ;
 - obtention (23a) d'une première image recadrée (31b) par recadrage de la première image (31a) et obtention (23b) d'une deuxième image recadrée (32b) par recadrage de la deuxième image (32a), un recadrage des première et deuxième images étant fonction de champs de vision desdites première (11) et deuxième (12) caméras ;
 - prédiction (24a) de premières profondeurs associées à un premier ensemble de pixels de la première image recadrée (31b) à partir dudit premier ensemble de pixels et d'un deuxième ensemble de pixels de la deuxième image recadrée (32b) correspondant audit premier ensemble de pixels et prédiction (24b) de deuxièmes profondeurs associées à un troisième ensemble de pixels de la deuxième image recadrée (32b) à partir dudit troisième ensemble de pixels et d'un quatrième ensemble de pixels de la première image recadrée (31b) correspondant audit troisième ensemble de pixels,
- lesdites premières et deuxièmes profondeurs étant prédites par une méthode de flux optique mise en œuvre par un réseau de neurones convolutif ;
- génération (25a) d'une troisième image à partir de ladite première image recadrée (31b) et des premières profondeurs et génération (25b) d'une quatrième image à partir de ladite deuxième image recadrée (32b) et des deuxièmes profondeurs, les générations étant obtenues à partir de la formule suivante :

$$p_s = \pi \left(K \left[F \phi \left(p_t K^{-1}, D_{st}(p_t) \right) \right] \right)$$

Avec :

- p_s un pixel de la troisième image, respectivement quatrième image,

- π une fonction pour passer de coordonnées homogènes à des coordonnées pixels en supprimant une dimension d'un vecteur,
- K une matrice intrinsèque de la deuxième caméra (12), respectivement de la première caméra (11),
- F une matrice extrinsèque du système de vision stéréoscopique,
- $\emptyset$ une fonction de projection dans la scène tridimensionnelle d'un pixel en fonction de sa profondeur,
- K' une matrice intrinsèque de la première caméra (11), respectivement de la deuxième caméra (12),
- $D_{st}(p_i)$ est une première profondeur du pixel p_i de la première image recadrée (31b), respectivement une deuxième profondeur du pixel p_i de la deuxième image recadrée (32b) ;
- détermination (26a) d'une première erreur par comparaison des troisième image et deuxième image recadrée (32b) et détermination (26b) d'une deuxième erreur par comparaison des quatrième image et première image recadrée (31b) ;
- ajustement (27) de paramètres d'entrée dudit réseau de neurones convolutif par minimisation d'une troisième erreur, ladite troisième erreur étant déterminée à partir d'une moyenne desdites première et deuxième erreurs.

[Revendication 2]

Procédé selon la revendication 1, pour lequel ladite troisième erreur est déterminée par la fonction de perte suivante :

$$L'_s = \sum_p \text{avg}(L_{*t}(p), L_{*s}(p)) \text{ Avec :}$$

- L'_s la troisième erreur,
- avg une moyenne d'arguments,
- $L_{*t}(p)$ la première erreur pour un pixel p de la deuxième image recadrée (32b),
- $L_{*s}(p)$ la deuxième erreur pour un pixel p de la première image recadrée (31b).

[Revendication 3]

Procédé selon l'une des revendications 1 à 2, pour lequel lesdites première et respectivement deuxième erreurs sont déterminées 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 :

- $L_*(p)$ la première, respectivement deuxième erreur,
- $I(p)$ une valeur du pixel p dans la deuxième image recadrée (32b), res-

pectivement la première image recadrée (31b),

- $\hat{I}(p)$ une valeur du pixel p dans la troisième image, respectivement la quatrième image
- SSIM une fonction qui prend en compte une structure locale, et
- α un facteur de pondération dépendant notamment du type d'environnement.

[Revendication 4]

Procédé selon l'une des revendications 1 à 3, pour lequel lesdites première et respectivement deuxième erreurs sont déterminées par la fonction de perte suivante :

$$L_{smooth}(D_{st}(p_t), W, o) = \sum_{p_t} \sum_{d \in x, y} (W(p_t) |\nabla_d^o D_{st}(p_t)| e^{\beta |\nabla_d^2 I_t(p_t)|})$$

Avec :

- $L_{smooth}(O, W, o)$ la première, respectivement deuxième erreur,
- $D_{st}(p_t)$ est la première profondeur, respectivement deuxième profondeur D_{tm} d'un pixel p_t prédite par le système de vision stéréoscopique ;
- W est une matrice de paramètres ;
- o est l'ordre d'un gradient de lissage ;
- une norme L1 des gradients de profondeur du second ordre est calculée avec $W=1$, et $o=2$;
- x et y sont les dimensions de la deuxième image, respectivement première image ;
- β est un hyperparamètre dépendant de l'environnement ; et
- $I_t(p_t)$ est la valeur du pixel p_t dans la troisième image, respectivement quatrième image.

[Revendication 5]

Procédé selon l'une des revendications 1 à 4, pour lequel ledit recadrage est réalisé de manière à obtenir au moins 50% des pixels de chacune desdites première et deuxième images recadrées (31b, 32b) associés à des objets de la scène tridimensionnelle présents dans des champs de vision des première (11) et deuxième (12) caméras.

[Revendication 6]

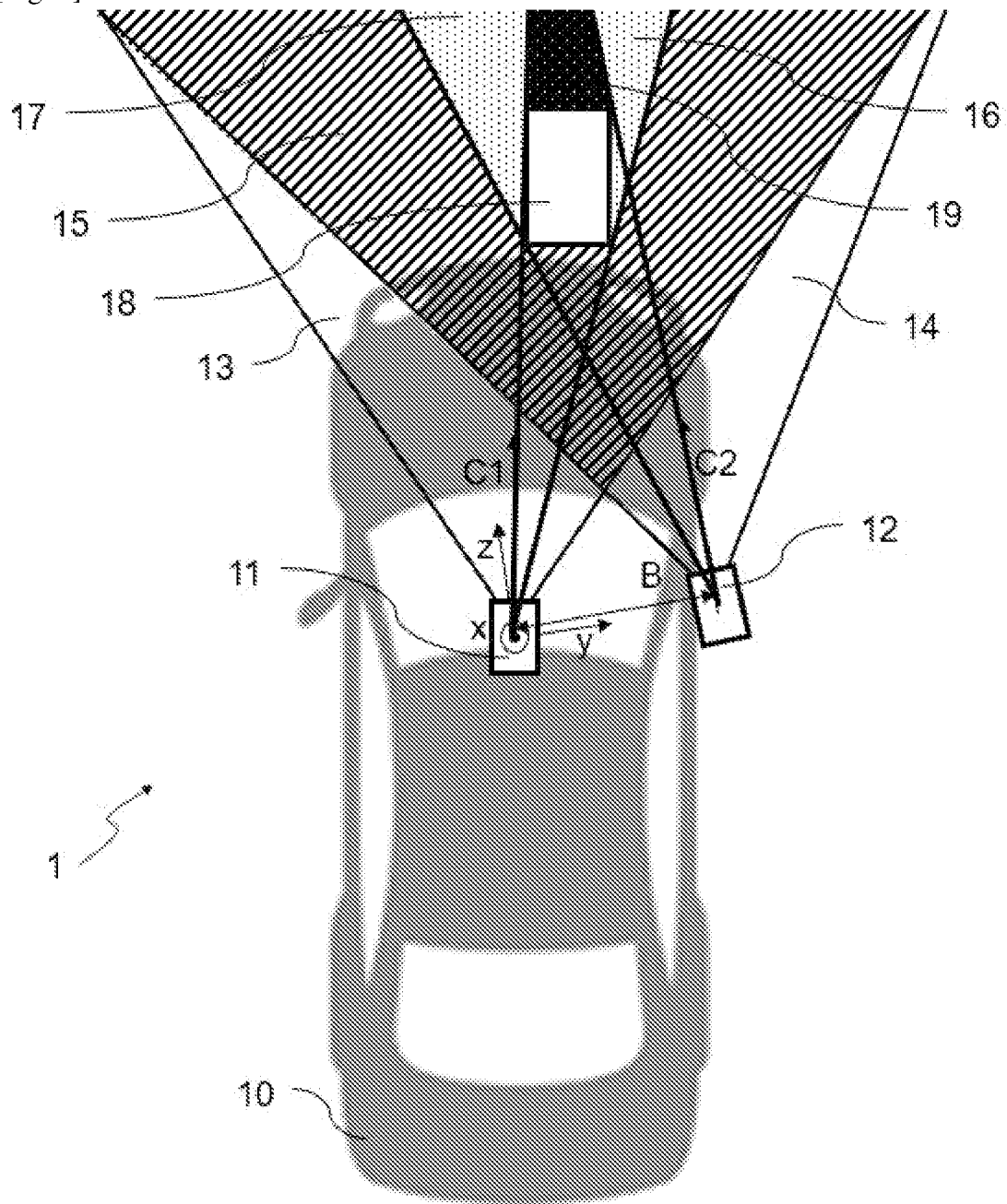
Procédé selon l'une des revendications 1 à 5, comprenant une étape de détermination d'une disparité induite par une position relative des première (11) et deuxième (12) caméras, lesdites prédictions de premières et deuxièmes profondeurs étant, en outre, fonction de la disparité induite.

[Revendication 7]

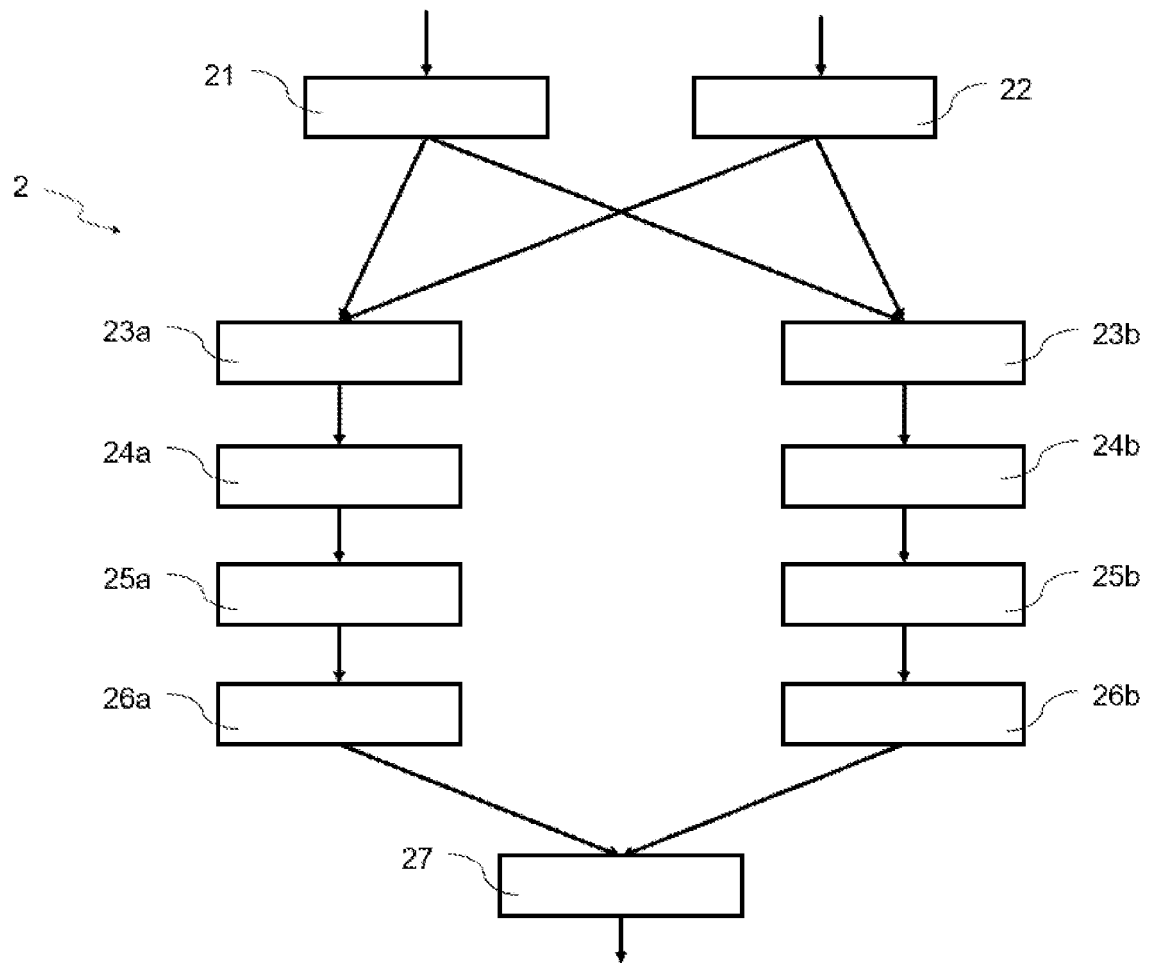
Programme d'ordinateur comportant des instructions pour la mise en œuvre du procédé selon l'une quelconque des revendications pré-

- cédentes, lorsque ces instructions sont exécutées par un processeur.
- [Revendication 8] 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 les revendications 1 à 6.
- [Revendication 9] Dispositif (4) de détermination d'une profondeur par un système de vision stéréoscopique 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 10] Véhicule (10) comprenant le dispositif (4) selon la revendication 9.

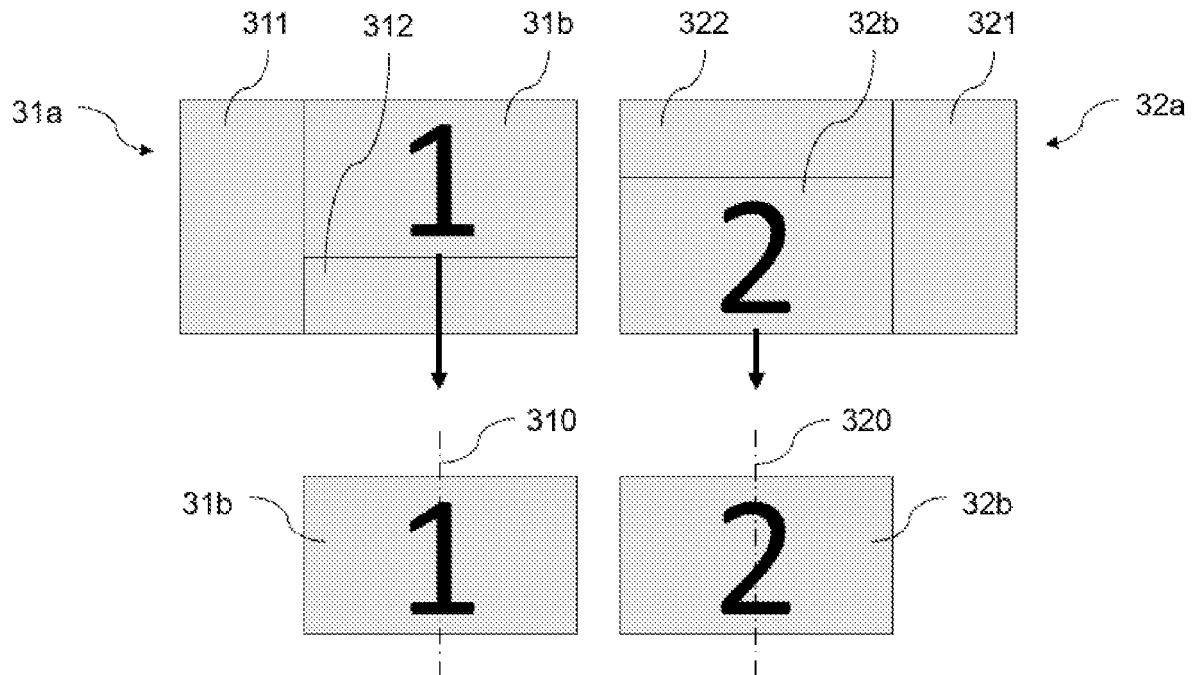
[Fig. 1]



[Fig. 2]

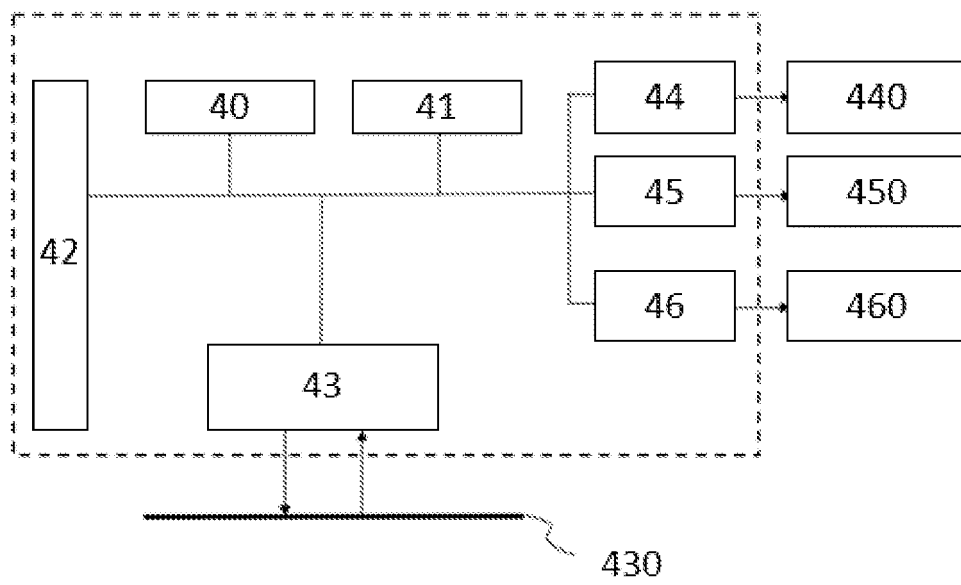


[Fig. 3]



[Fig. 4]

4



RAPPORT DE RECHERCHE PRÉLIMINAIRE

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

N° d'enregistrement
national

FA 922152
FR 2310306

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	WO 2021/096806 A1 (ZOOX INC [US]) 20 mai 2021 (2021-05-20)	1, 2, 5-10	G06T 7/593 G06V 10/82 G06V 20/56
Y	* abrégé; figures 1, 2, 3, 4, 6 * * alinéas [0011], [0012], [0019], [0055], [0096], [0105], [0057], [0122], [0113] * -----	3, 4	
A	US 2021/326694 A1 (WANG JIALIANG [US] ET AL) 21 octobre 2021 (2021-10-21) * le document en entier * -----	1-10	
Y	WANG YANG ET AL: "UnOS: Unified Unsupervised Optical-Flow and Stereo-Depth Estimation by Watching Videos", 2019 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), IEEE, 15 juin 2019 (2019-06-15), pages 8063-8073, XP033686469, DOI: 10.1109/CVPR.2019.00826 [extrait le 2020-01-08]	3, 4	
A	* abrégé * * section "Rigid and occlusion-aware structural matching"; page 8075 * * section "Regularization with edge-aware smoothness"; page 8074 * -----	1, 2, 5-10	DOMAINES TECHNIQUES RECHERCHÉS (IPC) G06T
Date d'achèvement de la recherche		Examineur	
24 avril 2024		Rodes Arnau, Isabel	
CATÉGORIE DES DOCUMENTS CITÉS 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		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	

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 2310306 FA 922152

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 **24-04-2024**
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
WO 2021096806 A1		20-05-2021	CN	114981834 A		30-08-2022
			EP	4058974 A1		21-09-2022
			JP	2023503827 A		01-02-2023
			WO	2021096806 A1		20-05-2021

US 2021326694 A1		21-10-2021	CN	113538575 A		22-10-2021
			DE	102021108927 A1		21-10-2021
			GB	2596392 A		29-12-2021
			US	2021326694 A1		21-10-2021
