

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
29 décembre 2004 (29.12.2004)

PCT

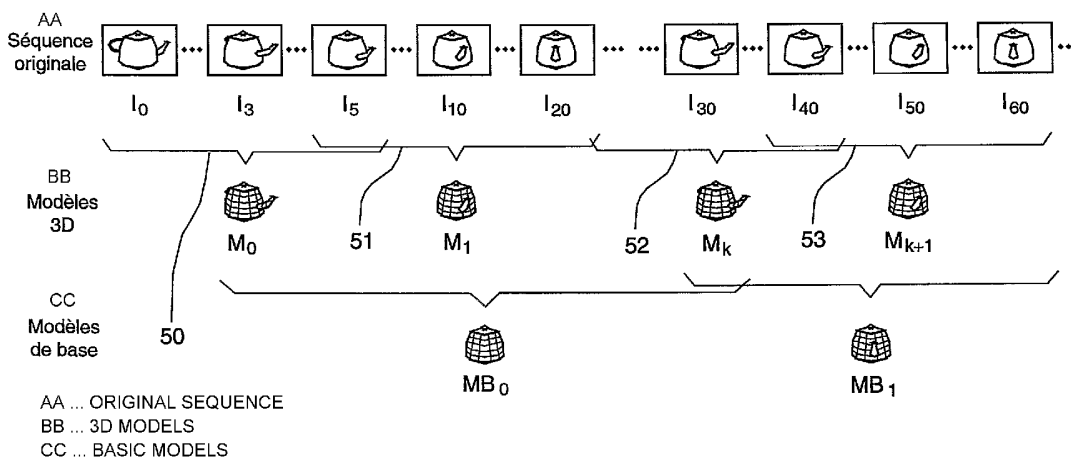
(10) Numéro de publication internationale
WO 2004/114669 A2

- (51) Classification internationale des brevets⁷ : H04N 7/26 F-35000 Rennes (FR). GIOIA, Patrick [FR/FR]; 31C rue Mirabeau, F-35700 Rennes (FR).
- (21) Numéro de la demande internationale : PCT/FR2004/001542 (74) Mandataire : VIDON, Patrice; Cabinet Vidon, 16B, rue de Jouanet, BP 90333, F-35703 Rennes Cedex 7 (FR).
- (22) Date de dépôt international : 18 juin 2004 (18.06.2004)
- (25) Langue de dépôt : français (81) États désignés (sauf indication contraire, pour tout titre de protection nationale disponible) : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (26) Langue de publication : français (84) États désignés (sauf indication contraire, pour tout titre de protection régionale disponible) : ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
- (30) Données relatives à la priorité : 03/07375 18 juin 2003 (18.06.2003) FR
- (71) Déposant (pour tous les États désignés sauf US) : FRANCE TELECOM [FR/FR]; 6, place d'Alleray, F-75015 Paris (FR).
- (72) Inventeurs; et
- (75) Inventeurs/Déposants (pour US seulement) : BAL-TER, Raphaèle [FR/FR]; 52, Boulevard de la Liberté,

[Suite sur la page suivante]

(54) Title: METHOD OF REPRESENTING A SEQUENCE OF PICTURES USING 3D MODELS, AND CORRESPONDING DEVICES AND SIGNAL

(54) Titre : PROCÉDE DE REPRESENTATION D'UNE SEQUENCE D'IMAGES PAR MODELES 3D, SIGNAL ET DISPOSITIFS CORRESPONDANTS



(57) Abstract: The invention relates to a method of representing a sequence of pictures which are grouped into sets comprising at least two successive pictures, known as groups of pictures (GOP), whereby a textured, meshed three-dimensional model is associated with each of said GOPs. According to the invention, the three-dimensional model associated with the n level GOP is represented with an irregular mesh taking account of at least one vertex of at least the irregular mesh representing the three-dimensional model that is associated with the n-1 level GOP, said vertex being known as the common vertex.

(57) Abrégé : L'invention concerne un procédé de représentation d'une séquence d'images regroupées en ensembles d'au moins deux images successives, appelés GOPs, un modèle tridimensionnel maillé texturé étant associé à chacun desdits GOPs. Selon l'invention, le modèle tridimensionnel associé au GOP de niveau n est représenté à l'aide d'un maillage irrégulier tenant compte d'au moins un sommet d'au moins le maillage irrégulier représentant le modèle tridimensionnel associé au GOP de niveau n-1, ledit sommet étant appelé sommet commun.

WO 2004/114669 A2



européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

Publiée :

— *sans rapport de recherche internationale, sera republiée dès réception de ce rapport*

Procédé de représentation d'une séquence d'images par modèles 3D, signal et dispositifs correspondants.

Le domaine de l'invention est celui du codage de séquences d'images. Plus précisément, l'invention concerne une technique de codage de séquences
5 d'images par flux de modèles tridimensionnels, ou 3D.

On rappelle que le codage vidéo par modèles 3D consiste à représenter une séquence vidéo par un ou plusieurs modèles 3D texturés. Les informations à transmettre à un codeur de la séquence d'images sont les modèles 3D, les images de textures qui leur sont associées, et les paramètres de la caméra ayant filmé la
10 séquence.

Ce type de codage permet donc d'atteindre des débits plus faibles que les techniques de codage classiques, selon lesquelles on représente généralement les vidéos par un ensemble de pixels, qui est bien plus coûteux à transmettre.

En outre, une telle technique de codage par modèles 3D permet, par rapport aux techniques de codage classiques, d'ajouter certaines fonctionnalités à la séquence reconstruite. Il est ainsi possible de changer l'illumination de la scène, de réaliser un affichage stéréoscopique, de stabiliser la séquence (lorsqu'il s'agit d'une séquence vidéo), d'ajouter des objets dans la scène ou enfin de changer le point de vue, de façon à simuler une navigation libre dans la scène (la navigation
15 libre peut en effet être définie comme un changement de trajectoire de la caméra par rapport au chemin original).
20

Il existe de ce fait, sur le marché du codage d'images, une demande importante de méthodes d'extraction de modèles 3D à partir de vidéos. En effet, en partant de scènes 3D réelles, on obtient par modélisation 3D un contenu
25 beaucoup plus photoréalistique que selon les méthodes de synthèse envisagées par le passé. En outre, grâce aux fonctionnalités citées ci-dessus, l'obtention de modèles virtuels des scènes réelles permet d'envisager un grand nombre d'applications telles que l'e-commerce, les jeux vidéo, la simulation, les effets spéciaux ou encore le repérage géographique.

On connaît à ce jour plusieurs techniques permettant la construction de modèles 3D à partir d'une vidéo.

Certaines techniques, dites actives, nécessitent de contrôler l'éclairage d'une scène réelle, et mettent généralement en œuvre une technologie laser, ou un grand nombre de caméras, afin d'acquérir plusieurs angles de vues et de nombreuses informations de profondeur.

D'autres techniques, dites passives, reposent quant à elles sur des algorithmes de calcul sophistiqués, et sont basées, soit sur les relations entre les images, soit sur les silhouettes. Elles diffèrent principalement les unes des autres par le niveau de calibration nécessaire et le degré d'interactivité autorisé. Elles consistent à reconstruire une information 3D à partir d'un ensemble de photographies ou d'images, et se heurtent principalement aux deux problèmes suivants :

- la mise en correspondance, qui consiste à trouver, pour une zone d'une image donnée, une zone correspondante dans les autres images (cette zone peut être réduite à un point de l'image) ;
- la calibration de la caméra, qui consiste à estimer les paramètres de formation de l'image (à savoir, les paramètres intrinsèques de la caméra (distance focale, etc.) et ses paramètres extrinsèques (positions de la caméra pour l'acquisition des différentes images de la séquence, etc.)).

La mise en correspondance est généralement gérée manuellement, comme décrit par V. M. Bove et al. dans "Semiautomatic 3D-model extraction from uncalibrated 2-D camera views," (en français, "extraction semi-automatique de modèles 3D à partir de vues bidimensionnelles non-calibrées de caméra") Proceedings Visual Data Exploration and Analysis, 1995.

La calibration est quant à elle un processus fastidieux, et les algorithmes de calcul qui lui sont associés sont souvent instables. De nombreuses méthodes reposent donc sur des séquences calibrées qui demandent, soit une intervention humaine (E. Boyer et al., "Calibrage et Reconstruction à l'aide de Parallélépipèdes

et de Parallélogrammes," Actes du treizième congrès francophone des reconnaissances de Formes et Intelligence Artificielle, 2002), soit un système d'acquisition compliqué, reposant sur une « turntable » (en français, « table tournante ») (W. Niem, "Robust and Fast Modeling of 3D Natural Objects from
5 Multiple Views," en français "modélisation rapide et robuste d'objets naturels 3D à partir de vues multiples", vci1994, 1994) ou sur l'utilisation d'un robot mobile (J. Wingbermuehle, "Automatic Reconstruction of 3D Object Using a Mobile Monoscopic Camera," en français "Reconstruction automatique d'objets 3D utilisant une caméra mobile monoscopique", Proceedings of the International
10 Conference on Recent Advances in 3D Imaging and Modelling, Ottawa, Canada, 1997).

Selon certaines autres méthodes automatiques, ou semi-automatiques, la mise en correspondance n'est pas gérée manuellement. On se réfèrera par exemple aux techniques de A. Fitzgibbon et al., ("Automatic Line Matching and 3D
15 Reconstruction of Buildings from Multiple Views," (en français, "Mise en correspondance automatique de lignes, et reconstruction 3D d'immeubles à partir de vues multiples") IAPRS, Munich, Allemagne, 1999) ou de C. Zeller et al., ("3-D Reconstruction of Urban Scene from Sequence of Images," (en français, "Reconstruction 3D de scènes urbaines à partir de séquences d'images) INRIA,
20 Information Technology 2572, 1995).

Cependant, ces méthodes semi-automatiques, ou automatiques, nécessitent de faire de nombreuses hypothèses sur les scènes à reconstruire, et ne s'appliquent par exemple qu'aux scènes architecturales.

Ces méthodes de reconstruction 3D automatiques mettent classiquement
25 en œuvre les étapes suivantes :

- détection de points ou de lignes particuliers ;
- mise en correspondance entre les images : cette étape consiste à suivre le long de la séquence vidéo les points ou lignes particuliers extraits lors de l'étape précédente ;
- 30 - mise en relation des différentes images ;

- reconstruction projective des points 3D ;
- autocalibration, ou raffinement de la calibration, si nécessaire, pour passer à un modèle 3D métrique (en effet, les manipulations interactives du modèle se font dans l'espace euclidien) ;
- 5 - estimation du modèle 3D texturé.

Certaines approches, basées sur l'algorithme ci-dessus, permettent de reconstruire un modèle 3D à partir de données fournies par une caméra monoculaire en mouvement (c'est à dire qu'on n'a aucune connaissance a priori, ni sur les paramètres intrinsèques ou extrinsèques de la caméra, ni sur la scène à
10 reconstruire). On peut se référer par exemple aux techniques de P. Debevec et al., "Panel Session on Visual Scene Representation," Smile2000, 2000, ou de G. Cross et al., "VHS to VRML: 3D Graphical Models from Video Sequences," en français "de la VHS au VRML : modèles graphiques 3D à partir de séquences vidéo", IEEE International Conference on Multimedia Computing and System,
15 Florence, 1999.

J. Rönning et al. dans "Modeling Structured Environments by a Single Moving Camera," (en français "Modélisation d'environnements structurés par une simple caméra mobile") Second International Conference on 3-D Imaging and Modelling, 1999 ont proposé une méthode qui estime un premier modèle à partir
20 de contours détectés et de filtres étendus de Kalman. Cependant, cette méthode présente l'inconvénient de reposer beaucoup sur les contours, et donc d'être mal adaptée aux scènes compliquées.

Dans "VHS to VRML: 3D Graphical Models from Video Sequences," IEEE International Conference on Multimedia Computing and System, Florence,
25 1999, G. Cross et al. ont présenté une méthode consistant à détecter des points par la méthode de Harris, et à les mettre en correspondance entre les différentes vues, simultanément à l'estimation de la géométrie. La mise en correspondance se fait par corrélation en croix, couplée à la géométrie épipolaire pour deux vues, ou à la géométrie trifocale pour trois vues, qui permettent de guider les appariements. Les
30 correspondances sont ensuite étendues à la séquence et optimisées par un

ajustement par faisceaux. On obtient donc des matrices de projection 3*4 et une structure 3D euclidienne (par autocalibration), sur laquelle on plaque la texture des images originales. Ceci permet de masquer les imperfections de la géométrie.

Cependant, un inconvénient de cette méthode est que le mouvement entre
5 deux images successives doit être relativement petit et que la séquence d'images doit être de taille raisonnable. Cette méthode n'est donc pas adaptée à une séquence d'images quelconque.

Deux approches ont également été proposées par l'université de Louvain.

Selon la première approche (M. Pollefeys, "Tutorial on 3D Modeling from
10 Images," eccv2000, 2000), les points ou lignes particuliers des images détectés sont extraits et mis en correspondance à l'aide de l'algorithme de Torr (décrit dans l'ouvrage cité ci-dessus). En parallèle, une calibration restreinte est évaluée, afin de pouvoir éliminer les correspondances incompatibles avec la calibration. La méthode de Beardsley (M. Pollefeys, "Tutorial on 3D Modeling from Images,"
15 eccv2000, 26 juin 2000, Dublin, Irlande) permet d'obtenir une ébauche de reconstruction projective pour les deux premières images, et les matrices de projection des autres vues. Une autocalibration, en fixant certaines inconnues à leurs valeurs par défaut et en appliquant le concept de la conique absolue, permet de retrouver les paramètres internes de la caméra, afin de passer à une
20 représentation métrique. Les données sont ensuite fusionnées en un modèle 3D commun, à l'aide d'une méthode qui concatène les points qui se correspondent sur plusieurs images, pour former deux chaînes (une chaîne descendante et une chaîne montante), à partir des cartes de disparités et des rotations calculées lors de la calibration. Pour les grands objets, une approche multi résolution est proposée.

25 Cependant, un inconvénient de cette technique est que l'approche multi résolution proposée pour les grands objets nécessite de disposer de plusieurs vidéos de la même scène, afin d'avoir accès non seulement à une vue d'ensemble mais également aux détails. En outre, cette méthode est de type semi-automatique.

Selon une seconde technique (Gool et al., "From image sequences to 3D
30 models," en français "des séquences d'images aux modèles 3D", Third

International Workshop on Automatic Extraction of Man-made Objects from Aerial and Space Images, 2001), les points ou lignes particuliers des images sont détectés par la méthode de Harris ou par la méthode de Shi et Tomasi (décrite par M. Pollefeys, dans "Tutorial on 3D Modeling from Images," eccv2000, 26 juin 5 2000, Dublin, Irlande). Ces caractéristiques sont ensuite mises en correspondance, ou suivies entre les différentes vues, suivant qu'il s'agit d'images ou de vidéo. A partir de ces correspondances, les relations entre les vues sont calculées par une méthode robuste comme celle de Torr ou de Fisher et Bolles. Pour la reconstruction projective, deux images sont sélectionnées, afin d'obtenir une 10 reconstruction initiale, en déterminant les matrices de projection pour des paramètres intrinsèques et une matrice de rotation approchée, et en triangulant. La position des caméras correspondant aux autres vues est ensuite déterminée à l'aide de la géométrie épipolaire. La structure est ensuite raffinée en utilisant un filtre de Kalman (décrit par M. Pollefeys, dans "Tutorial on 3D Modeling from Images," 15 eccv2000, 26 juin 2000, Dublin, Irlande) étendu pour chaque point. Lorsque la structure et le mouvement ont été obtenus pour toute la séquence, un ajustement de faisceaux est réalisé. On passe de la reconstruction projective à la reconstruction euclidienne grâce à l'autocalibration. Le modèle 3D virtuel est ensuite obtenu en élevant le maillage triangulaire sur l'une des images de la 20 séquence, en éliminant les points pour lesquels la profondeur n'est pas disponible.

Un inconvénient de cette méthode est qu'elle ne donne de bons résultats que sur les scènes simples, et n'est pas adaptée aux scènes complexes.

Plus généralement, toutes les techniques de l'art antérieur décrites ci-dessus présentent comme inconvénient de nécessiter de faire des hypothèses 25 simplificatrices sur l'acquisition de la séquence d'images (en termes par exemple de paramètres de la caméra), et/ou sur le contenu de la scène, ou encore sur la longueur de la séquence. En d'autres termes, ces différentes méthodes ne sont pas adaptées à une scène et une séquence d'images quelconques, éventuellement complexes.

Une dernière méthode, orientée codage, a été proposée par Franck Galpin dans "Représentation 3D de séquences vidéo: Schéma d'extraction automatique d'un flux de modèles 3D, applications à la compression et à la réalité virtuelle," Université de Rennes 1, 2002. Contrairement aux autres méthodes de l'art
5 antérieur, selon lesquelles on cherche à reconstruire un modèle 3D unique pour l'ensemble de la séquence d'images, l'idée principale de la méthode de Franck Galpin est de traiter la séquence vidéo par morceaux, afin d'obtenir plusieurs modèles, qui seront chacun valides pour une section de la séquence, appelée GOP (« Group Of Pictures », en français « groupe d'images »).

10 On suppose que la scène est statique (ou segmentée au sens du mouvement), filmée par une caméra monoculaire en mouvement, que les paramètres d'acquisition (paramètres intrinsèques et extrinsèques de la caméra) sont inconnus, que la focale de la caméra est constante et que la scène contient peu ou pas de surfaces spéculaires. Le contenu de la scène et les mouvements de
15 la caméra sont supposés quelconques.

On réalise une estimation du mouvement dense, basée sur l'équation du flot optique ou sur un maillage 2D déformable, afin de permettre une estimation entre des images éloignées de la séquence (à savoir les images clefs qui délimitent les GOPs). Ces images clefs sont sélectionnées parallèlement et servent de support
20 à l'estimation du modèle 3D. Le calcul robuste des paramètres intrinsèques et extrinsèques des caméras est également réalisé sur les images clefs, et affiné simultanément avec la géométrie 3D, par une méthode d'ajustement de faisceaux par fenêtre glissante. Les positions des images intermédiaires sont estimées par localisation par Dementhon (voir notamment « Représentation de séquence
25 vidéo : schéma d'extraction automatique d'un flux de modèles 3D, applications à la compression et à la réalité virtuelle », Université de Rennes 1, janvier 2002, par Franck Galpin) afin de pouvoir reconstruire la séquence originale, comme illustré sur la figure 1.

La séquence initiale comprend une pluralité d'images I_k successives,
30 regroupées en groupes d'images appelées GOPs. Ainsi, les images I_0 à I_5 sont

regroupées au sein d'un premier GOP référencé 1, auquel est associé un modèle 3D M_0 . Les images I_5 à I_{13} sont rassemblées au sein d'un deuxième GOP référencé 2, auquel est associé un deuxième modèle M_1 .

Cette dernière méthode de l'art antérieur permet d'obtenir, en termes de
5 codage, des résultats très supérieurs aux autres méthodes décrites précédemment dans ce document. Les figures 2a à 2e illustrent les résultats obtenus, en bas débit, selon cette technique d'une part, et selon la technique H26L d'autre part. Plus précisément, la figure 2a présente l'évolution du PSNR, les figures 2b et 2c
10 présentent respectivement une image et une zone de détail de cette image obtenues selon la technique H26L (ou H264, voir notamment « Sliding adjustment for 3D video representation », Franck Galpin et Luce Morin, eurasip 2002, pages 1088 à 2001) pour un débit de 82kb/s, et les figures 2d et 2e présentent les mêmes images obtenues selon la méthode de flux de modèles 3D de Franck Galpin.

Sur la figure 2a, la première courbe (la plus haute sur la figure) est la
15 qualité objective de la séquence reconstruite, obtenue par reprojexion des modèles 3D suivant la méthode de Franck Galpin dans l'espace texture, i.e. sans prendre en compte les distorsions géométriques. Les deux autres courbes de la figure 2a indiquent la qualité objective pour les séquences reconstruites obtenues par la méthode de Franck Galpin et par le codeur H264 dans l'espace image.

20 Bien qu'en mesure objective (c'est-à-dire en termes de PSNR « Peak Signal to Noise Ratio », « rapport signal à bruit crête ») les performances obtenues soient similaires pour le codeur de Franck Galpin et le codeur H26L, on notera que, d'un point de vue visuel, la qualité obtenue est supérieure avec le codeur basé sur un flux de modèles 3D, notamment en termes de respect des
25 détails, d'absence d'effets blocs, etc.

En outre, cette technique de codage basée sur un flux de modèles 3D permet d'atteindre de très bas débits pour une qualité visuelle satisfaisante, ainsi qu'illustré par les figures 3a à 3c, qui présentent respectivement :

- l'évolution du PSNR ;
- 30 - une image obtenue selon cette technique ;

- une zone de détail de cette image,
pour un débit de 16kb/s.

Bien que la méthode de Franck Galpin, reposant sur l'extraction d'un flux de modèles 3D, ne présente pas certains inconvénients inhérents aux méthodes
5 d'extraction d'un modèle 3D unique décrites précédemment, elle se heurte cependant à certains problèmes.

Notamment, un inconvénient de cette technique de l'art antérieur est que l'ensemble des modèles 3D obtenus pour une séquence d'images ne sont que partiellement redondants, ce qui rend cette technique inadaptée aux applications
10 de navigation libre dans une scène.

En effet, les différents modèles 3D obtenus sont exprimés dans des repères différents, et présentent de nombreuses imperfections (en termes notamment de dérive, de points aberrants, etc.).

Un autre inconvénient de cette technique de l'art antérieur est que, bien
15 qu'elle soit orientée vers le codage (contrairement aux autres approches décrites précédemment), elle n'est échelonnable (en anglais « scalable ») que du point de vue de la texture des images, et non de la géométrie.

Cette méthode n'est donc pas, ou mal adaptée à une mise en œuvre sur des terminaux de visualisation de capacités de traitement très diverses, ou sur des
20 réseaux de transmission de débit variable.

L'invention a notamment pour objectif de pallier ces inconvénients de l'art antérieur.

Plus précisément, un objectif de l'invention est de fournir une technique de représentation d'une séquence d'images par modèle 3D qui soit adaptée à tout
25 type de séquence d'images fixes ou statiques, ou de scène, y compris complexe. Notamment, l'invention a pour objectif de mettre en œuvre une telle technique qui permette la reconstruction d'une scène, sur laquelle on ne formule aucune hypothèse, qui est acquise avec un appareil grand public, dont on ne connaît ni les caractéristiques, ni le déplacement.

Un autre objectif de l'invention est de mettre en œuvre une telle technique qui permette d'obtenir une séquence reproduite par reprojction de bonne qualité visuelle, même lorsqu'on s'éloigne de la trajectoire originale de la caméra ayant servi à l'acquisition de la séquence.

5 L'invention a encore pour objectif de fournir une telle technique qui soit adaptée aux bas et très bas débits.

L'invention a également pour objectif de mettre en œuvre une telle technique qui soit particulièrement bien adaptée aux scènes de grandes dimensions.

10 L'invention a encore pour objectif de fournir une telle technique qui convienne aux applications de codage et de navigation virtuelle.

L'invention a aussi pour objectif de mettre en œuvre une telle technique qui permette d'obtenir des représentations échelonnables (en anglais « scalable ») des séquences d'images, de façon à permettre une transmission sur des réseaux de débits divers, en vue notamment d'applications portables.

15 Encore un objectif de l'invention est de fournir une telle technique, qui permette, à même débit, la représentation de scènes de meilleure qualité visuelle que selon la technique de Franck Galpin décrite ci-dessus.

L'invention a aussi pour objectif de mettre en œuvre une telle technique
20 qui permette, pour la représentation d'une séquence d'images de même qualité visuelle, une réduction du débit par rapport à la technique de Franck Galpin décrite ci-dessus.

Ces objectifs, ainsi que d'autres qui apparaîtront par la suite, sont atteints à l'aide d'un procédé de représentation d'une séquence d'images regroupées en
25 ensembles d'au moins deux images successives, appelés GOPs, un modèle tridimensionnel maillé texturé étant associé à chacun desdits GOPs.

Selon l'invention, le modèle tridimensionnel associé au GOP de niveau n est représenté à l'aide d'un maillage irrégulier tenant compte d'au moins un sommet d'au moins le maillage irrégulier représentant le modèle tridimensionnel
30 associé au GOP de niveau n-1, ledit sommet étant appelé sommet commun.

Ainsi, l'invention repose sur une approche tout à fait nouvelle et inventive de la représentation d'une séquence d'images par modèles 3D. En effet, comme pour la méthode proposée par Franck Galpin, l'invention propose une approche reposant, non pas sur l'extraction d'un modèle 3D unique pour l'ensemble des
5 images de la séquence, mais sur l'extraction d'un flux de modèles 3D, associés chacun à un groupe d'images, appelé GOP.

En outre, l'invention propose une amélioration inventive de la technique de Franck Galpin, en établissant une correspondance entre les différents modèles 3D associés à chacun des GOPs, de façon, notamment, à accroître leur
10 redondance. L'invention permet donc avantageusement des applications de type navigation interactive.

Une telle correspondance entre modèles 3D successifs est rendue possible en utilisant un maillage irrégulier des images, qui s'adapte particulièrement bien aux singularités des images. Le maillage irrégulier d'un modèle 3D prend ainsi en
15 compte au moins un sommet singulier (et plus généralement les points ou lignes particuliers de l'image) du maillage irrégulier du sommet 3D précédent.

L'invention permet donc, à qualité visuelle égale, de réduire le débit de transmission de la séquence d'images, du fait de la redondance entre les différents modèles 3D. Elle permet également, pour un même débit, d'obtenir une meilleure
20 qualité visuelle de la représentation de la séquence d'images, grâce au suivi des singularités de l'image entre modèles 3D successifs.

Selon une caractéristique avantageuse de l'invention, on associe également à au moins deux modèles tridimensionnels consécutifs un modèle de base construit à partir desdits sommets communs auxdits au moins deux modèles
25 tridimensionnels.

Selon la nature de la séquence d'images, il est possible qu'à tous les modèles 3D associés à la séquence corresponde un même maillage de base. Ce maillage de base, ou maillage grossier dont les différents modèles 3D constituent des raffinements, correspond à la structure géométrique commune à tous les
30 modèles 3D qui lui sont associés.

Préférentiellement, on passe d'un desdits modèles tridimensionnels à un autre par transformation en ondelettes, à l'aide d'un premier ensemble de coefficients d'ondelettes.

Avantageusement, un desdits modèles tridimensionnels est obtenu à partir
5 dudit modèle de base associé par transformation en ondelettes, à l'aide d'un deuxième ensemble de coefficients d'ondelettes.

L'invention permet donc une transmission scalable de la séquence d'images, adaptable en fonction des caractéristiques du réseau ou du terminal de visualisation. Les éléments à transmettre pour une reconstruction de la séquence
10 sont, outre les paramètres de la caméra, le maillage de base d'une part, et les coefficients d'ondelettes permettant de reconstruire les différents modèles 3d d'autre part. En transmettant un nombre plus ou moins grand de coefficients d'ondelettes, on obtient une qualité de reconstruction plus ou moins élevée, adaptée au débit du réseau de transmission ou à la capacité du terminal de
15 visualisation.

De manière préférentielle, ledit maillage irrégulier de niveau n est un maillage irrégulier bidimensionnel de l'une des images dudit GOP de niveau n.

De manière avantageuse, ladite image maillée est la première image dudit
GOP de niveau n.

20 Préférentiellement, chacun desdits modèles tridimensionnels est obtenu par élévation dudit maillage irrégulier le représentant.

On combine ainsi des informations de profondeur au maillage 2D, pour obtenir par élévation une carte de profondeur maillée.

Selon une première variante avantageuse de l'invention, ledit maillage
25 bidimensionnel irrégulier est obtenu par simplifications successives d'un maillage triangulaire régulier de ladite image.

Par exemple, on part de triangles de côté 1, pour couvrir tous les points de l'image.

Selon une deuxième variante avantageuse de l'invention, ledit maillage bidimensionnel irrégulier est obtenu à partir d'un maillage de Delaunay de points d'intérêt prédéterminés de ladite image.

Ces points d'intérêt sont détectés au préalable, par exemple par
5 l'algorithme de Harris et Stephen.

Préférentiellement, deux GOPs successifs ont au moins une image commune.

Ainsi, la dernière image d'un GOP est aussi la première image du GOP
suivant.

10 Selon une caractéristique avantageuse de l'invention, lesdits sommets communs auxdits niveaux $n-1$ et n sont détectés par estimation de mouvement entre la première image dudit GOP de niveau $n-1$ et la première image dudit GOP de niveau n .

Avantageusement, un tel procédé comprend une étape de stockage desdits
15 sommets communs détectés.

Ces sommets communs stockés peuvent ensuite être utilisés pour la construction du modèle associé au GOP suivant.

De manière préférentielle, ledit maillage irrégulier représentant ledit
modèle associé au GOP de niveau n tient également compte d'au moins un
20 sommet d'au moins le maillage irrégulier représentant le modèle associé au GOP de niveau $n+1$.

En procédant ainsi de manière bidirectionnelle, on accroît encore la qualité visuelle lors de la reconstruction.

Avantageusement, ledit deuxième ensemble de coefficients d'ondelettes
25 est généré par application d'au moins un filtre d'analyse sur un remaillage semi-régulier dudit modèle tridimensionnel associé.

On rappelle qu'un maillage semi-régulier est un maillage dont les sommets qui n'ont pas six voisins sont isolés sur le maillage (c'est-à-dire qu'ils ne sont pas voisins entre eux).

Préférentiellement, lesdites ondelettes sont des ondelettes de deuxième génération.

De manière préférentielle, lesdites ondelettes appartiennent au groupe comprenant :

- 5
- les ondelettes affines par morceaux ;
 - les ondelettes polynomiales ;
 - les ondelettes basées sur le schéma de subdivision de Butterfly.

L'invention concerne aussi un signal représentatif d'une séquence d'images regroupées en ensembles d'au moins deux images successives, appelés
10 GOPs, un modèle tridimensionnel maillé texturé étant associé à chacun desdits GOPs.

Selon l'invention, un tel signal comprend :

- au moins un champ contenant un modèle de base construit à partir de sommets communs à au moins deux maillages irréguliers, représentant chacun un
15 modèle tridimensionnel, lesdits au moins deux modèles tridimensionnels étant associés à au moins deux GOPs successifs ;
- au moins un champ contenant un ensemble de coefficients d'ondelettes permettant de construire, par transformation en ondelettes à partir dudit modèle de base, au moins un modèle tridimensionnel associé à l'un desdits
20 GOPs ;
- au moins un champ contenant au moins une texture associée à l'un desdits modèles tridimensionnels ;
- au moins un champ contenant au moins un paramètre de position de caméra.

25 L'invention concerne encore un dispositif de représentation d'une séquence d'images mettant en œuvre le procédé de représentation décrit précédemment.

L'invention concerne notamment un dispositif de représentation d'une séquence d'images regroupées en ensembles d'au moins deux images successives,

appelés GOPs, un modèle tridimensionnel maillé texturé étant associé à chacun desdits GOPs.

Selon l'invention, un tel dispositif comprend :

- 5 - des moyens de construction desdits modèles tridimensionnels, par transformation en ondelettes d'au moins un modèle de base, élaboré à partir de sommets communs à au moins deux maillages irréguliers représentant deux modèles tridimensionnels successifs ;
- 10 - des moyens de représentation desdites images de la séquence à partir desdits modèles tridimensionnels, d'au moins une image de texture et d'au moins un paramètre de position de caméra.

L'invention concerne aussi un dispositif de codage d'une séquence d'images regroupées en ensembles d'au moins deux images successives, appelés GOPs, un modèle tridimensionnel maillé texturé étant associé à chacun desdits GOPs.

- 15 Selon l'invention, un tel dispositif de codage comprend des moyens de codage d'un modèle tridimensionnel associé au GOP de niveau n, ledit modèle tridimensionnel étant représenté à l'aide d'un maillage irrégulier tenant compte d'au moins un sommet d'au moins le maillage irrégulier représentant le modèle tridimensionnel associé au GOP de niveau n-1.

- 20 D'autres caractéristiques et avantages de l'invention apparaîtront plus clairement à la lecture de la description suivante d'un mode de réalisation préférentiel, donné à titre de simple exemple illustratif et non limitatif, et des dessins annexés, parmi lesquels :

- 25 - la figure 1, déjà commentée en relation avec l'art antérieur, présente le principe de la reconstruction d'une séquence vidéo par un flux de modèles 3D ;
 - les figures 2a à 2e, déjà commentées en relation avec l'art antérieur, illustrent un comparatif des résultats visuels obtenus selon une technique de type H26L d'une part, et selon la technique de codage de la figure 1 d'autre part ;
- 30

- les figures 3a à 3c, déjà commentées en relation avec l'art antérieur, présentent les résultats obtenus selon la technique de la figure 1 pour un débit faible de 16kb/s ;
- la figure 4 illustre le principe général de la reconstruction d'une séquence vidéo à partir d'un modèle 3D ;
- la figure 5 illustre le principe général de la présente invention, reposant sur l'extraction d'un flux de modèles 3D, associés chacun à un modèle de base, commun à un ou plusieurs modèles 3D ;
- la figure 6 présente les différents coefficients d'ondelettes utilisés pour le codage des modèles 3D de la figure 4 ;
- la figure 7 présente un synoptique des différentes étapes mises en œuvre selon l'invention pour le codage des images de la séquence.

Le principe général de l'invention repose sur l'extraction d'un flux de modèles 3D auxquels on associe des maillages irréguliers, adaptés au contenu des images de la séquence, et qui prennent en compte les correspondants des sommets du maillage irrégulier du modèle 3D précédent.

On rappelle brièvement, en relation avec la figure 4, le principe général de la reconstruction d'une séquence vidéo par l'intermédiaire d'un modèle tridimensionnel.

On considère une scène réelle, en l'occurrence un objet 41 (ici, une théière), que l'on filme (42) au moyen d'une caméra 43. On ne fait aucune hypothèse, ni sur la nature de cette caméra, qui peut être un appareil grand public, ni sur les paramètres d'acquisition de la vidéo.

Après numérisation 44 de la vidéo, on obtient une séquence d'images 45, que l'on appelle séquence originale.

Par analyse 46 de cette séquence originale, on construit au moins un modèle 3D 47 (une pluralité de modèles 3D selon l'invention), à partir duquel on peut reconstruire (48) une séquence d'images 49, en vue de son affichage sur un terminal de visualisation.

On présente désormais, en relation avec la figure 5, le principe général de l'invention, qui est basé, d'une part, sur un flux de modèles 3D maillés texturés, et d'autre part, sur la mise en œuvre de transformations en ondelettes.

Chaque modèle 3D correspond à une partie de la séquence d'images d'origine, c'est-à-dire à un GOP (en anglais « Group of Pictures »). Les modèles 3D considérés sont des cartes d'élévations maillées irrégulièrement, sous la contrainte de la prise en compte des correspondants des sommets du modèle précédent. Cette contrainte permet de garantir des correspondances précises entre les sommets des modèles successifs.

Les transformations permettant de passer d'un modèle à un autre sont décomposées en ondelettes, ce qui permet d'adapter la précision de la transformation au débit, grâce à la scalabilité naturelle des ondelettes.

L'invention repose en outre sur la reconstruction de modèles de base, que l'on associe à un ou plusieurs GOPs successifs, ainsi qu'illustré par la figure 4.

La séquence d'images originale est constituée d'images I_k successives. On a plus particulièrement représenté sur la figure 4 les images $I_0, I_3, I_5, I_{10}, I_{20}, I_{30}, I_{40}, I_{50}$, et I_{60} . Cette séquence peut être de longueur quelconque, aucune hypothèse restrictive n'étant nécessaire selon la présente invention.

La séquence d'images I_k est divisée en groupes d'images successifs, appelés GOPs. Ainsi, le premier GOP 50 comprend les images référencées I_0 à I_5 , le deuxième GOP 51 comprend les images I_5 à I_{20} , un $(k+1)^{\text{ième}}$ GOP 52 comprend notamment les images I_{30} à I_{40} et un $(k+2)^{\text{ième}}$ GOP 53 comprend les images I_{40} à I_{60} . On notera que, dans le mode de réalisation préférentiel de la figure 4, la dernière image d'un GOP est également la première image du GOP suivant : ainsi, l'image I_5 par exemple appartient au premier GOP 50 et au deuxième GOP 51.

On construit, pour chacun de ces GOPs 50 à 53, un modèle 3D M_k . Le modèle 3D M_0 est associé au GOP 50, le modèle 3D M_1 est associé au GOP 51, etc.

On construit également un ensemble de modèles de base, notés MB_k , dont les modèles 3D M_k constituent des raffinements. Ainsi, sur la figure 4, le modèle

de base MB_0 est associé aux modèles 3D M_0 à M_k , et le modèle de base MB_1 est associé aux modèles 3D M_k , M_{k+1} et suivants.

On choisit d'associer un tel modèle grossier MB_k aux modèles 3D de l'ensemble des GOPs le long desquels on peut suivre un ensemble de points particuliers prédéterminés. Lorsque certains de ces points ne sont plus apparents dans le modèle 3D suivant, on choisit de passer à un nouveau modèle de base MB_{k+1} .

On peut ainsi décomposer en ondelettes les différents modèles 3D M_k , qui ont été obtenus séparément, mais qui s'appuient tous sur un même maillage de base, à savoir celui du modèle grossier commun associé.

Suivant la nature des images de la séquence originale, et l'existence de zones communes entre celles-ci en plus ou moins grand nombre, le maillage de base MB_k pourra être valide pour un nombre variable de GOPs, voire même éventuellement pour toute la séquence d'images.

Grâce à ces modèles de base MB_k , on peut donc exprimer chaque modèle 3D M_k estimé, par le maillage de base lui correspondant d'une part, et par un ensemble de coefficients d'ondelettes d'autre part.

Cette représentation est résumée dans le schéma de la figure 6, où les coefficients t_i^k représentent les coefficients d'ondelettes relatifs à une transformation de passage d'un modèle 3D M_k au suivant et où les coefficients r_i^k représentent les coefficients d'ondelettes relatifs à un raffinement entre un modèle de base MB_k et un modèle 3D M_k associé.

Ainsi, les coefficients d'ondelettes $t_0^{k,k+1}$ à $t_n^{k,k+1}$ sont utilisés pour passer d'un modèle 3D M_k au modèle 3D M_{k+1} . Les coefficients d'ondelettes r_0^k à r_n^k illustrent quant à eux le passage d'un modèle 3D M_k au modèle de base associé (en l'espèce, le modèle MB_1).

Le premier ensemble de coefficients d'ondelettes t_i^k définit donc les liens entre les différents modèles M_k , ce qui permet de passer de l'un à l'autre, et de générer des modèles intermédiaires, soit par une interpolation linéaire entre les correspondants, soit de manière implicite grâce aux ondelettes.

Le deuxième jeu d'ondelettes r_i^k assure une transmission progressive et efficace (en terme de débit) des différents modèles. Ainsi, la technique de l'invention peut être adaptée à tous types de terminaux, quelle que soit leur capacité de traitement, et à tous types de réseaux de transmission, quel que soit
5 leur débit.

On présente désormais, en relation avec la figure 7, les différentes étapes mises en œuvre selon l'invention, lors du codage des modèles et des textures associées, pour la représentation d'une séquence d'images originale.

En entrée de l'algorithme, on a un ensemble d'images naturelles I_n à I_m ,
10 correspondant à différentes prises de vue d'une scène ou d'un objet du monde réel, comme illustré précédemment en relation avec la figure 4. Dans un mode de réalisation préféré de l'invention, ces images sont au format ppm et au format pgm. L'invention s'applique bien sûr également à tout autre format d'image.

On procède tout d'abord à une estimation de mouvement 71 entre les
15 différentes images de la séquence originale, de façon à déterminer le champ de mouvement $C_{n,n+p}$ entre les images I_n et I_{n+p} , ainsi que l'ensemble des points supports de l'estimation de l'information 3D, à savoir l'ensemble $\varepsilon_{n,n+p}$ des sommets du maillage utilisé pour l'estimation de mouvement entre les images I_n et I_{n+p} , ayant les plus hauts scores avec le détecteur de Harris et Stephen et décimés
20 régulièrement.

On sélectionne ensuite (72) les images clefs K_k de la séquence originale, qui délimitent les différents GOPs de la séquence.

Si la séquence originale est une séquence vidéo, la sélection 72 des images clefs K_k délimitant les GOPs est réalisée selon l'algorithme développé par Franck
25 Galpin et al. dans "Sliding Adjustment for 3D Video Representation" EURASIP Journal on Applied Signal Processing 2002 :10 (voir notamment le paragraphe 5.1. Selection Criteria). Cette sélection 72 des images de début et de fin de GOP repose donc sur la validation de trois critères:

- un mouvement moyen suffisant pour la reconstruction de l'information
30 3D ;

- un pourcentage de points communs relativement élevé entre les deux images extrêmes du GOP ;
- la validité de la géométrie estimée (évaluée grâce au résidu épipolaire).

La première image clef sélectionnée est quant à elle la première image I_0
5 de la séquence originale.

Pour l'extraction des modèles 3D M_k , c'est-à-dire pour l'estimation de la matrice fondamentale et pour l'estimation des matrices de projection et des positions de caméra 73, on exploite également les techniques développées par Franck Galpin dans "Représentation 3D de séquences vidéo: Schéma d'extraction
10 automatique d'un flux de modèles 3D, applications à la compression et à la réalité virtuelle," Université de Rennes 1, 2002 et dans "Sliding Adjustment for 3D Video Representation" EURASIP Journal on Applied Signal Processing 2002 :10. Ces techniques reposent sur les algorithmes classiques de la modélisation 3D.

Dans le cas, non pas d'une séquence vidéo, mais d'un ensemble d'images,
15 le principe est le même pour l'extraction de l'information 3D. En revanche, le support de cette estimation est un ensemble de points particuliers de l'image courante, ayant un fort score pour le détecteur de Harris et Stephen (dans "A Combined Corner and Edge Detector," en français "un détecteur combiné de sommets et d'arêtes", Proc. 4th Alvey Vision Conf., 1988), dont on cherche les
20 correspondants dans l'image suivante par « block matching » (ou estimation de mouvement par blocs). On limite en outre le nombre de modèles à transmettre, en mettant en œuvre une sélection 72 des images à prendre en compte pour la reconstruction de la séquence originale. Cette sélection 72 est basée sur les mêmes critères que la sélection des images clefs dans le cas d'une vidéo.

25 Après sélection 72 des images clefs K_k du GOP k , on détermine donc le champ de mouvement C_k associé au GOP k comme étant le champ de mouvement entre les deux images de début et de fin de GOP k .

On procède également à une calibration 75, permettant de déterminer l'ensemble des paramètres intrinsèques et extrinsèques de la caméra ayant servi à

l'acquisition de la séquence d'images, et notamment la position P_k de la caméra associée à l'image I_k .

Connaissant cette position P_k d'une part, et le champ de mouvement C_k associé au GOP k d'autre part, on estime (74) la carte de profondeur Z_k associée au GOP k .

On sauvegarde également (76) l'ensemble des images clefs K_k de la séquence originale associées aux GOPs k .

On pourra se référer aux deux publications de Franck Galpin citées précédemment pour le mode de fonctionnement plus particulier des blocs référencés 71 à 76 sur la figure 7.

En vue de la reconstruction, on réalise un maillage irrégulier bidimensionnel 77 des cartes de profondeur Z_k , sous la contrainte de la prise en compte des correspondants des sommets du modèle associé au GOP précédent, contenus dans l'image K_k .

Ce maillage 2D peut être calculé de deux manières:

- par des simplifications successives à partir d'un maillage régulier de triangles de côté 1 (i.e. tous les points de l'image) ;
- par un maillage de Delaunay de points d'intérêt détectés au préalable.

Lorsque le maillage a été déterminé au niveau n , on estime (78), à l'aide du champ de mouvement C_n , les correspondants de ces points dans la dernière image du GOP n (qui est aussi, dans un mode de réalisation préféré de l'invention, la première image du GOP $n+1$). Cette liste de sommets correspondants est alors stockée (78) et utilisée lors du maillage 77 du modèle associé au GOP $n+1$.

Dans le cas du maillage 2D obtenu par simplification, on contraint que les points de cette liste 78 soient présents dans le maillage final.

Dans le cas du maillage de Delaunay, les sommets du maillage associé au GOP $n+1$ obtenu par une triangulation de Delaunay sont:

- les points particuliers détectés par l'algorithme de Harris et Stephen ("A Combined Corner and Edge Detector," Proc. 4th Alvey Vision

Conf., 1988), ou tout autre détecteur de points d'intérêt adéquat, sur l'image clef K_{n+1} du GOP $n+1$,

- les correspondants des sommets du maillage associé au GOP n .

La liste des correspondants $C(E_n)$ calculés au niveau n permet de prendre
5 en compte les sommets du modèle du GOP n qui ne feraient pas partie des sommets détectés par Harris sur l'image clé du GOP $n+1$.

Ainsi, on est assuré de la présence des correspondants des sommets d'un modèle dans le modèle suivant, ce qui facilite amplement le lien 79 entre ces deux modèles. En effet, les correspondances 79 entre les modèles pourront être
10 obtenues de manière précise grâce au champ de mouvement.

Dans une variante de réalisation de l'invention, pour obtenir une transformation 79 encore plus précise, on rend cette étude bidirectionnelle, en contraignant le maillage du modèle courant à la prise en compte des correspondants, non seulement des sommets du modèle précédent, mais
15 également des sommets du modèle suivant.

Les maillages 3D M_k , correspondant à la géométrie des modèles 3D représentant les GOPs, sont obtenus par élévation des maillages 2D estimés, ainsi qu'illustré par le bloc référencé 80.

Les correspondances 78 établies entre les sommets de deux modèles
20 successifs permettent d'exprimer la transformation 79, permettant de passer d'un modèle M_k à un modèle M_{k+1} , à l'aide de coefficients d'ondelettes.

L'intérêt d'exprimer cette transformation à l'aide d'ondelettes est que l'on peut adapter la précision de la transformation au débit grâce à la scalabilité naturelle des ondelettes.

25 Les ondelettes utilisées pour la décomposition sont des ondelettes de deuxième génération, c'est-à-dire qu'elles sont définissables sur des ensembles qui n'ont pas de structure d'espace vectoriel. En l'espèce, avec les notations de la Figure 6, les ondelettes sont définies sur les modèles de base MB_0, MB_1 , etc.

Disposant du maillage de base MB_i et de la correspondance géométrique
30 entre MB_i et le modèle 3D M_i , la génération des coefficients d'ondelettes se fait

par application de filtres d'analyse sur un remaillage semi-régulier de M_i . Les coefficients d'ondelettes d sont la solution du système linéaire suivant :

$$Td = c$$

où T est la matrice de synthèse globale et où c est l'ensemble des positions des sommets sur le remaillage semi-régulier de M_i .

T dépend du type d'ondelettes utilisées. Trois schémas sont privilégiés selon l'invention : les ondelettes affines par morceaux, les ondelettes polynomiales (notamment les ondelettes de Loop), et les ondelettes basées sur le schéma de subdivision de Butterfly (J. Warren et al., "Multiresolution Analysis for Surfaces of Arbitrary Topological Type," *ACM Transactions on Graphics*, vol. 16, pp. 34--73, 1997).

Ainsi, la matrice T est de la forme

$$T = (P \ Q)$$

où P est une sous-matrice qui représente uniquement le schéma de subdivision (Affine, Loop, Butterfly,...) et où la sous-matrice Q est l'interprétation géométrique des coefficients d'ondelettes.

Dans un mode de réalisation préféré de l'invention, Q est choisie de telle sorte que les coefficients d'ondelettes aient un moment nul. Dans le cas général, P et Q peuvent être arbitraires dans la mesure où T reste inversible.

La figure 7 résume l'approche qui vient d'être exposée pour le GOP k .

Les notations utilisées sur cette figure sont les suivantes:

- $I_n \dots I_m$ sont les images d'entrée ;
- $C_{n,n+p}$ est le champ de mouvement entre les images I_n et I_{n+p} ;
- C_k est le champ de mouvement associé au GOP k ;
- $C(V)$ est l'ensemble des correspondants des points de l'ensemble V trouvés par le champ de mouvement ;
- ε_m est l'ensemble des points supports de l'estimation de l'information 3D (sommets du maillage utilisé pour l'estimation de mouvement ayant les plus hauts scores avec le détecteur de Harris et Stephen et décimés régulièrement) ;

- E_k est l'ensemble des sommets du modèle 3D associé au GOP k ;
- Z_k est la carte de profondeur associée au GOP k ;
- K_k est l'image de la séquence originale correspondant à l'image clef associée au GOP k ;
- 5 - M_k est le modèle 3D associé au GOP k ;
- P_m est la position de caméra associée à l'image I_m ;
- θ_k est l'ensemble des coefficients d'ondelettes définissant la transformation de passage entre M_k et M_{k+1} ;
- V_k est l'ensemble des sommets du maillage correspondant au modèle
- 10 M_k .

Le codeur 81 reçoit en entrée les positions P_k de la caméra pour les différentes images I_k de la séquence originale, l'estimation M_k du modèle 3D texturé, et les coefficients d'ondelettes permettant de transformer le modèle M_{k-1} en modèle M_k .

15 Simultanément à l'estimation des modèles 3D M_k de chacun des GOPs k , illustrée en figure 7, on reconstruit des modèles de base MB_i valides pour plusieurs GOPs successifs.

On suit pour cela, grâce au champ de mouvement calculé C_k , l'ensemble de points particuliers détectés dans la première image du GOP k le long de plusieurs

20 images de la séquence. Plus précisément, on détecte la présence des correspondants de ces points le long de plusieurs GOPs successifs, jusqu'à ce que le nombre de correspondants inclus dans l'image analysée soit inférieur à un seuil prédéterminé. Ce seuil doit être choisi de façon à assurer la possibilité de la reconstruction (i.e. de l'estimation de la matrice fondamentale) ; on le choisit par

25 exemple égal à 7. Lorsque le nombre de points particuliers détectés dans un GOP est inférieur au seuil, on en déduit que ce GOP ne doit pas être associé au même modèle de base MB_i que les GOPs précédents.

A partir de ce sous-ensemble de points particuliers, que l'on a suivis de GOP en GOP, on reconstruit un modèle de base MB_i dont les sommets sont tous

présents dans les modèles M_k associés aux GOPs k le long desquels on a pu suivre ces points.

Ces modèles de base, ou modèles grossiers MB_i sont ensuite individuellement décomposés en ondelettes. Pour ce faire, on met en œuvre la
5 méthode décrite par P. Gioia, dans "Reducing the number of wavelet coefficients by geometric partitioning," *Computational geometry, Theory and applications*, vol. 14, 1999, en s'appuyant sur le même maillage de base. Chaque modèle 3D M_k est considéré comme un raffinement du modèle de base grossier MB_i .

Ainsi, les coefficients t_i^k de la figure 6 sont obtenus de la manière
10 suivante : les maillages de base issus d'un même GOP sont identiques, et génèrent après subdivision, le même maillage semi-régulier. Par conséquent, les coefficients r_i^k sont indexés par les mêmes sommets géométriques lorsque k varie dans un même GOP. Pour chaque k intermédiaire, on peut donc définir une fonction f^k qui fait correspondre à chacun de ces sommets la différence entre les
15 coefficients r_i^k et r_i^{k+1} . Cette fonction f^k est alors décomposée, comme précédemment, en coefficients d'ondelettes, qui sont les coefficients t_i^k .

L'invention permet donc de transmettre la géométrie des modèles associés à la séquence originale de manière peu coûteuse, puisqu'on transmet, d'une part, les maillages de base et d'autre part, les coefficients d'ondelettes associés aux
20 différents modèles.

Les applications envisageables dans le cadre de l'invention sont nombreuses. L'invention s'applique ainsi tout particulièrement au codage d'images représentant une même scène fixe (qui peuvent être un ensemble d'images indépendantes ou une vidéo). Les taux de compression atteints par ce
25 type de représentation se situent dans les bas et très bas débits (typiquement de l'ordre de 20 kbits/s) et on peut donc envisager des applications portables.

De plus, la séquence virtuelle obtenue par reprojexion (au décodage) possède toute les fonctionnalités permises par la 3D, telles que le changement d'illumination, la stabilisation de la séquence, la navigation libre, l'ajouts d'objet...

REVENDEICATIONS

1. Procédé de représentation d'une séquence d'images regroupées en ensembles d'au moins deux images successives, appelés GOPs, un modèle tridimensionnel maillé texturé étant associé à chacun desdits GOPs,
5 caractérisé en ce que le modèle tridimensionnel associé au GOP de niveau n est représenté à l'aide d'un maillage irrégulier tenant compte d'au moins un sommet d'au moins le maillage irrégulier représentant le modèle tridimensionnel associé au GOP de niveau n-1, ledit sommet étant appelé sommet commun.
- 10 2. Procédé de représentation selon la revendication 1, caractérisé en ce qu'on associe également à au moins deux modèles tridimensionnels consécutifs un modèle de base construit à partir desdits sommets communs auxdits au moins deux modèles tridimensionnels.
- 15 3. Procédé de représentation selon l'une quelconque des revendications 1 et 2, caractérisé en ce qu'on passe d'un desdits modèles tridimensionnels à un autre par transformation en ondelettes, à l'aide d'un premier ensemble de coefficients d'ondelettes.
- 20 4. Procédé de représentation selon l'une quelconque des revendications 1 à 3, caractérisé en ce qu'un desdits modèles tridimensionnels est obtenu à partir dudit modèle de base associé par transformation en ondelettes, à l'aide d'un deuxième ensemble de coefficients d'ondelettes.
- 25 5. Procédé de représentation selon l'une quelconque des revendications 1 à 4, caractérisé en ce que ledit maillage irrégulier de niveau n est un maillage irrégulier bidimensionnel de l'une des images dudit GOP de niveau n.
6. Procédé de représentation selon la revendication 5, caractérisé en ce que ladite image maillée est la première image dudit GOP de niveau n.
7. Procédé de représentation selon l'une quelconque des revendications 1 à 6, caractérisé en ce que chacun desdits modèles tridimensionnels est obtenu par élévation dudit maillage irrégulier le représentant.

8. Procédé de représentation selon l'une quelconque des revendications 5 à 7, caractérisé en ce que ledit maillage bidimensionnel irrégulier est obtenu par simplifications successives d'un maillage triangulaire régulier de ladite image.
- 5 9. Procédé de représentation selon l'une quelconque des revendications 5 à 7, caractérisé en ce que ledit maillage bidimensionnel irrégulier est obtenu à partir d'un maillage de Delaunay de points d'intérêt prédéterminés de ladite image.
- 10 10. Procédé de représentation selon l'une quelconque des revendications 1 à 9, caractérisé en ce que deux GOPs successifs ont au moins une image commune.
- 15 11. Procédé de représentation selon l'une quelconque des revendications 1 à 10, caractérisé en ce que lesdits sommets communs auxdits niveaux $n-1$ et n sont détectés par estimation de mouvement entre la première image dudit GOP de niveau $n-1$ et la première image dudit GOP de niveau n .
12. Procédé de représentation selon la revendication 11, caractérisé en ce qu'il comprend une étape de stockage desdits sommets communs détectés.
- 20 13. Procédé de représentation selon l'une quelconque des revendications 1 à 12, caractérisé en ce que ledit maillage irrégulier représentant ledit modèle associé au GOP de niveau n tient également compte d'au moins un sommet d'au moins le maillage irrégulier représentant le modèle associé au GOP de niveau $n+1$.
- 25 14. Procédé de représentation selon l'une quelconque des revendications 4 à 13, caractérisé en ce que ledit deuxième ensemble de coefficients d'ondelettes est généré par application d'au moins un filtre d'analyse sur un remaillage semi-régulier dudit modèle tridimensionnel associé.
15. Procédé de représentation selon l'une quelconque des revendications 3 à 14, caractérisé en ce que lesdites ondelettes sont des ondelettes de deuxième génération.

16. Procédé de représentation selon l'une quelconque des revendications 3 à 15, caractérisé en ce que lesdites ondelettes appartiennent au groupe comprenant :

- les ondelettes affines par morceaux ;
- 5 - les ondelettes polynomiales ;
- les ondelettes basées sur le schéma de subdivision de Butterfly.

17. Signal représentatif d'une séquence d'images regroupées en ensembles d'au moins deux images successives, appelés GOPs, un modèle tridimensionnel maillé texturé étant associé à chacun desdits GOPs,

10 caractérisé en ce qu'il comprend :

- au moins un champ contenant un modèle de base construit à partir de sommets communs à au moins deux maillages irréguliers, représentant chacun un modèle tridimensionnel, lesdits au moins deux modèles tridimensionnels étant associés à au moins deux GOPs successifs ;
- 15 - au moins un champ contenant un ensemble de coefficients d'ondelettes permettant de construire, par transformation en ondelettes à partir dudit modèle de base, au moins un modèle tridimensionnel associé à l'un desdits GOPs ;
- au moins un champ contenant au moins une texture associée à l'un desdits modèles tridimensionnels ;
- 20 - au moins un champ contenant au moins un paramètre de position de caméra.

18. Dispositif de représentation d'une séquence d'images mettant en œuvre le procédé de représentation de l'une quelconque des revendications 1 à 16.

25

19. Dispositif de représentation d'une séquence d'images regroupées en ensembles d'au moins deux images successives, appelés GOPs, un modèle tridimensionnel maillé texturé étant associé à chacun desdits GOPs, caractérisé en ce qu'il comprend :

- des moyens de construction desdits modèles tridimensionnels, par transformation en ondelettes d'au moins un modèle de base, élaboré à partir de sommets communs à au moins deux maillages irréguliers représentant deux modèles tridimensionnels successifs ;
- 5 - des moyens de représentation desdites images de la séquence à partir desdits modèles tridimensionnels, d'au moins une image de texture et d'au moins un paramètre de position de caméra.

20. Dispositif de codage d'une séquence d'images regroupées en ensembles d'au moins deux images successives, appelés GOPs, un modèle
10 tridimensionnel maillé texturé étant associé à chacun desdits GOPs, caractérisé en ce qu'il comprend des moyens de codage d'un modèle tridimensionnel associé au GOP de niveau n, ledit modèle tridimensionnel étant représenté à l'aide d'un maillage irrégulier tenant compte d'au moins un
15 sommet d'au moins le maillage irrégulier représentant le modèle tridimensionnel associé au GOP de niveau n-1.

1/5

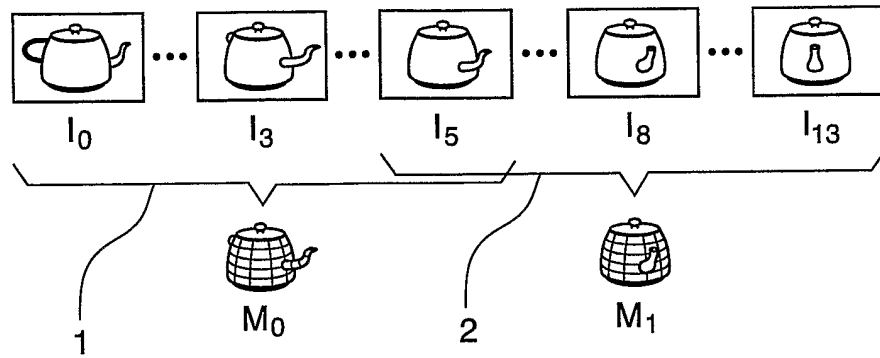


Fig. 1

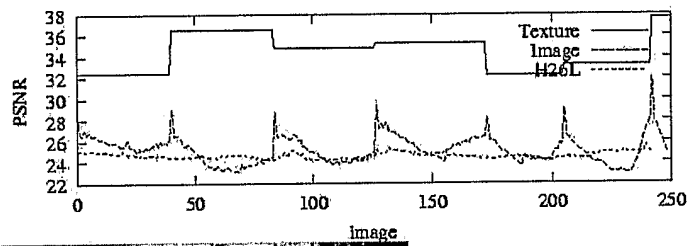


Fig. 2a

Fig. 2b



Fig. 2c



Fig. 2d



Fig. 2e



2/5

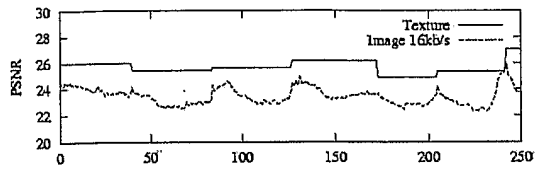


Fig. 3a



Fig. 3b

Fig. 3c

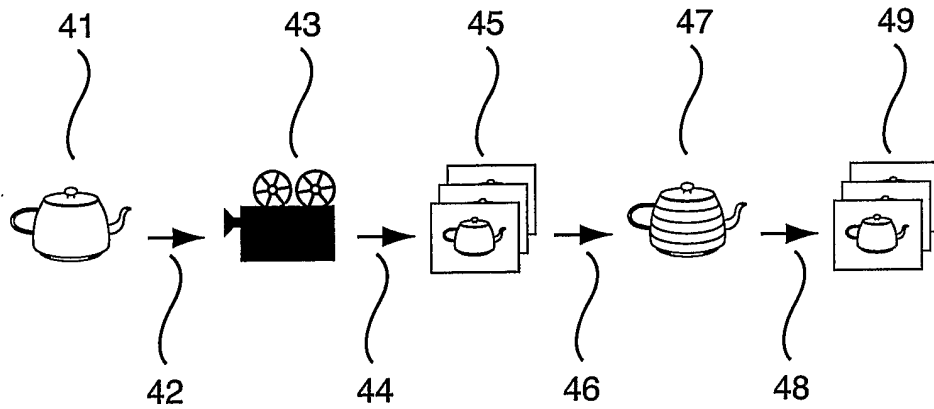


Fig. 4

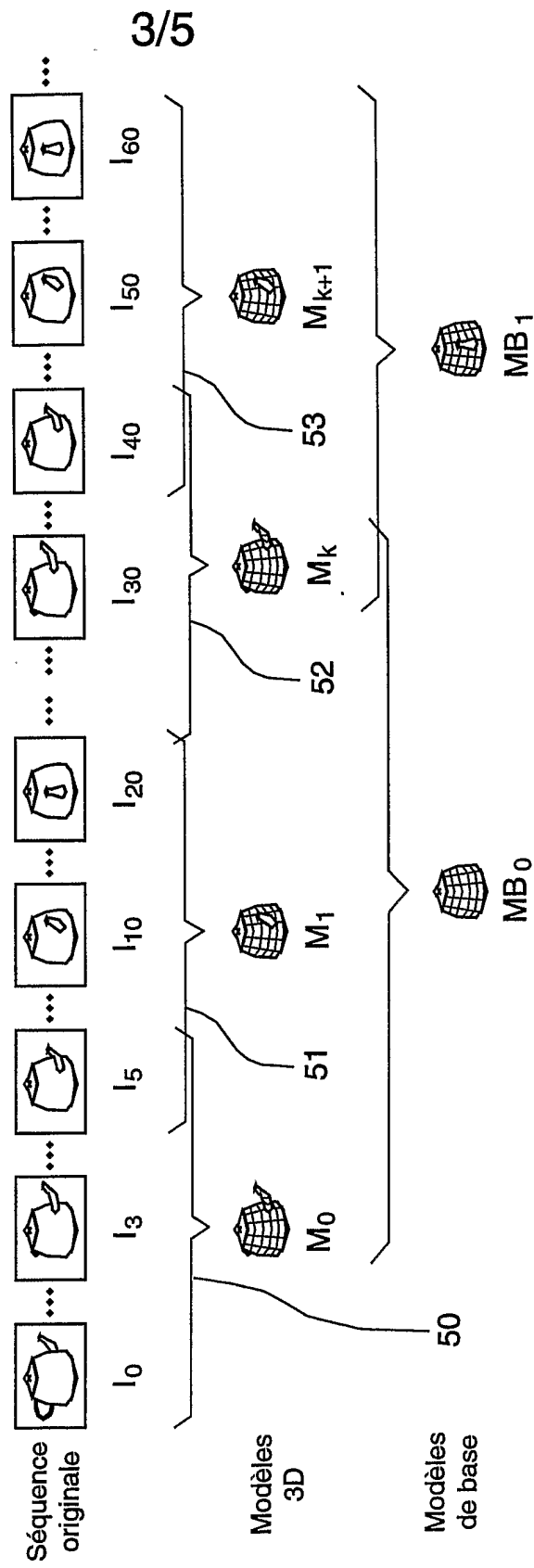


Fig. 5

4/5

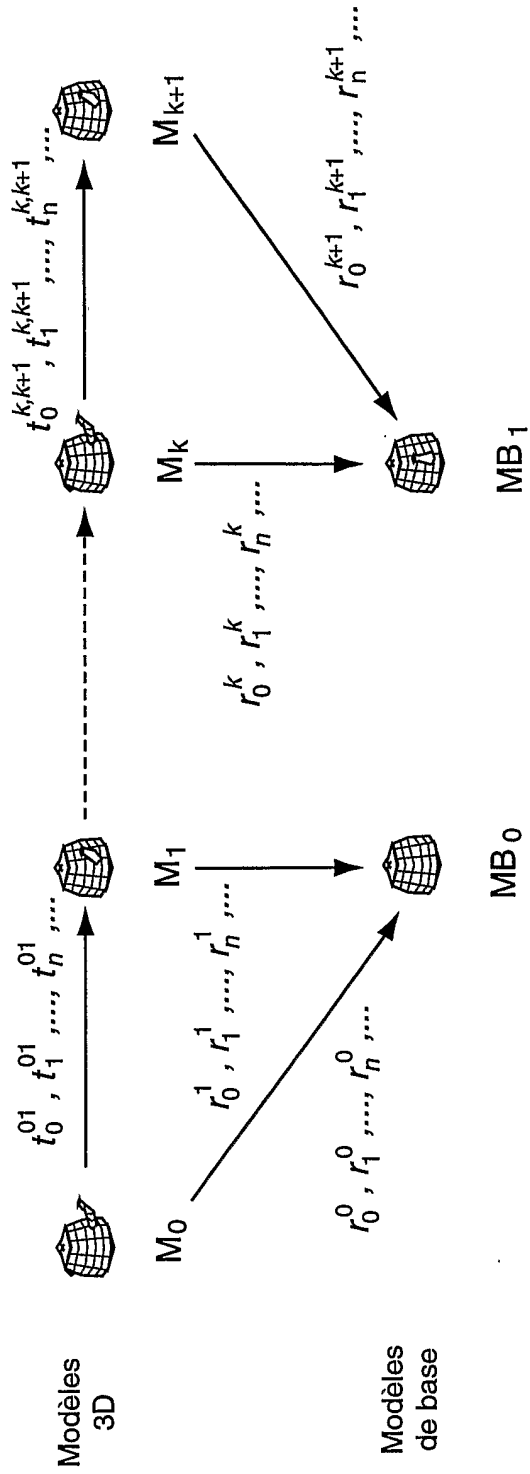


Fig. 6

5/5

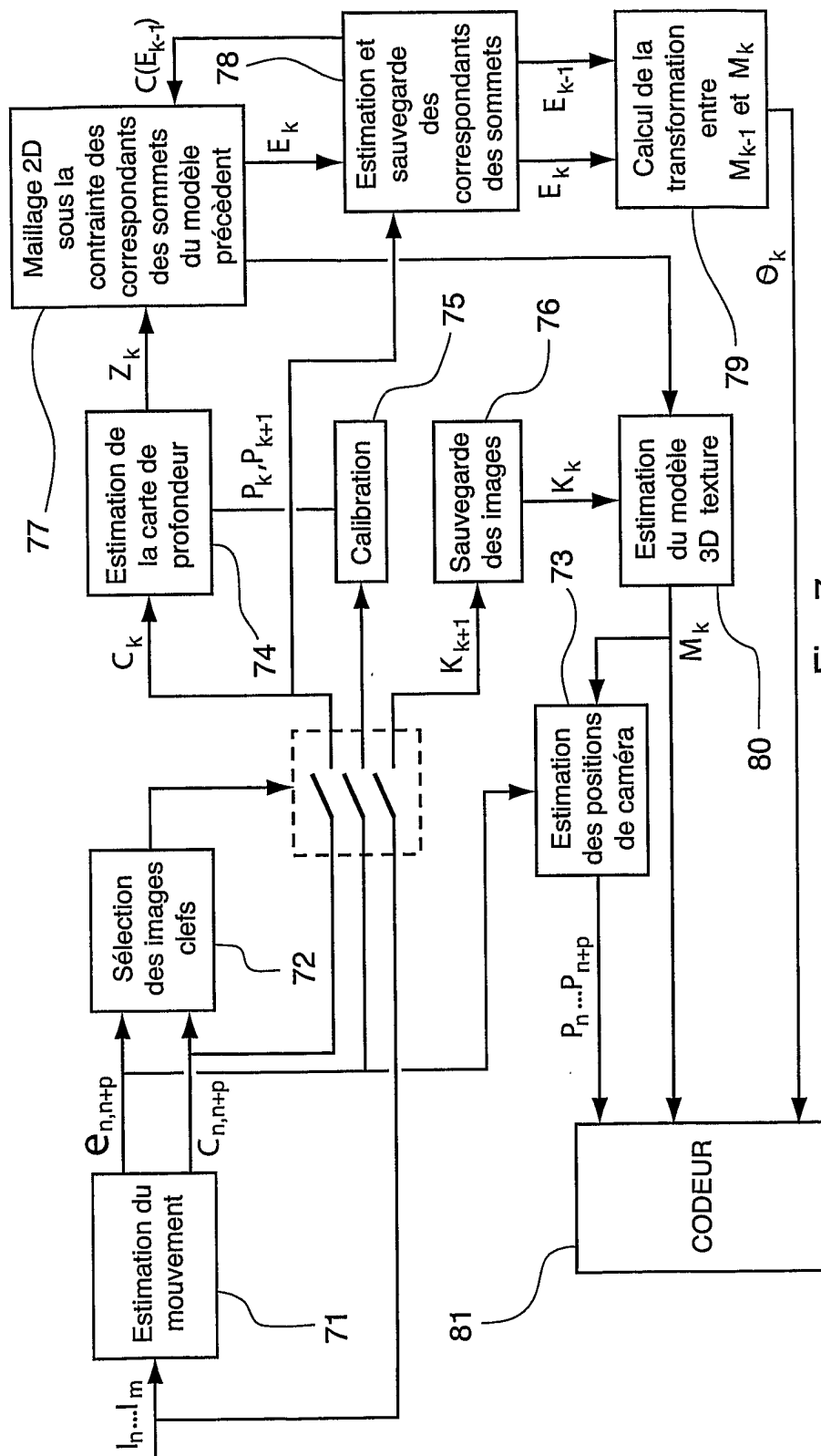


Fig. 7