

(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
10 décembre 2009 (10.12.2009)

PCT

(10) Numéro de publication internationale
WO 2009/147181 A1

(51) Classification internationale des brevets :
H04N 7/26 (2006.01)

(21) Numéro de la demande internationale :
PCT/EP2009/056828

(22) Date de dépôt international :
3 juin 2009 (03.06.2009)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
0803065 3 juin 2008 (03.06.2008) FR

(71) Déposant (pour tous les États désignés sauf US) :
THALES [FR/FR]; 45, Rue de Villiers, F-92200 Neuilly
sur Seine (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) : **LE BARZ, Cédric** [FR/FR]; 6 rue Chaumisson, F-914710 Limours en Hurepoix (FR). **LENY, Marc** [FR/FR]; 90 rue de garches, F-92000 Nanterre (FR). **RENAN, Erwann** [FR/FR]; 67 rue de L'Aqueduc, F-75010 Paris (FR). **PRETEUX, Françoise** [FR/FR]; Laboratoire Artenis, 9, rue Charles Fourier, F-91011 Evry (FR).

(74) Mandataires : **DUDOUIT, Isabelle** et al.; Immeuble Visium, 22 avenue Aristide Briand, F-94117 Arcueil (FR).

(81) États désignés (sauf indication contraire, pour tout titre de protection nationale disponible) : AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(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), eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Suite sur la page suivante]

(54) Title : CONSTRUCTION OF A NOVEL STRUCTURE FOR REPRESENTING IMAGES BASED ON SEMANTIC OBJECTS WITHIN A COMPRESSED VIDEO STREAM, ASSOCIATED METHOD AND DEVICE

(54) Titre : CONSTRUCTION D'UNE NOUVELLE STRUCTURE DE REPRESENTATION DE L'IMAGE S'APPUYANT SUR DES OBJETS SEMANTIQUES AU SEIN D'UN FLUX VIDEO COMPRESSE, PROCEDE ET DISPOSITIF ASSOCIES

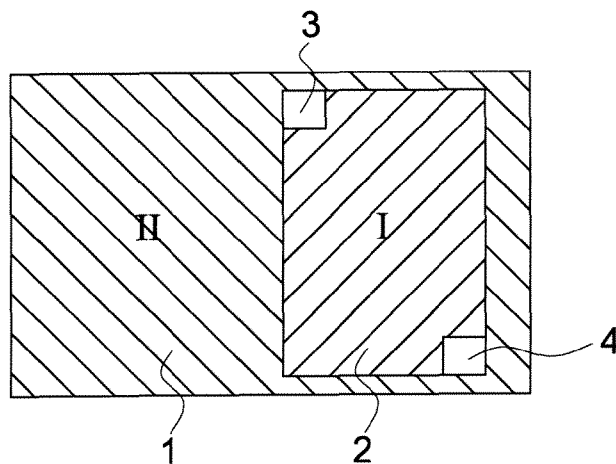


FIG.2

(57) Abstract : Method making it possible to define a structure of an image in a compressed video stream characterized in that it comprises at least the following steps: defining within an image i at least one object O_{i-i} or a group of objects go_{i-i} , a group of objects being defined as a set of objects related by one or more common descriptor(s) chosen when said stream is compressed, said object comprising one or more portions or "slices" S_{i-i} corresponding to a sub-part of the image consisting of macroblocks which belong to one and the same set, and inserting at the level of an object one or more descriptor(s) or to a group of objects one or more common descriptor(s).

(57) Abrégé :

[Suite sur la page suivante]



WO 2009/147181 A1



Publiée :

— avec rapport de recherche internationale (Art. 21(3))

Procédé permettant de définir une structure d'une image dans un flux vidéo compressé caractérisé en ce qu'il comporte au moins les étapes suivantes : définir au sein d'une image i au moins un objet O_{i-i} ou un groupe d'objets go_{i-i} , un groupe d'objets étant défini comme un ensemble d'objets liés par un ou plusieurs descripteur(s) commun(s) choisi au moment de la compression dudit flux, ledit objet comprenant une ou plusieurs portions ou « slices » S_{i-i} correspondant à une sous-partie de l'image constituée de macroblocs qui appartiennent à un même ensemble, et insérer au niveau d'un objet un ou plusieurs descripteur(s) ou à un groupe d'objets un ou plusieurs descripteur(s) commun(s).

**CONSTRUCTION D'UNE NOUVELLE STRUCTURE DE
REPRESENTATION DE L'IMAGE S'APPUYANT SUR DES OBJETS
SEMANTIQUES AU SEIN D'UN FLUX VIDEO COMPRESSE, PROCEDE
ET DISPOSITIF ASSOCIES**

5

L'invention concerne une nouvelle structure de représentation de l'image s'appuyant sur des objets sémantiques au sein d'un flux vidéo compressé, ainsi que le procédé permettant d'obtenir une telle structure. Elle sera, par exemple, utilisée dans des applications de vidéo surveillance.

- 10 Dans la suite de la description, le Demandeur définit sous le mot « objet », un ensemble connexe de pixels, blocs ou macroblocs d'une image qui présente un déplacement propre au sein de la séquence de l'image par rapport aux autres objets. La segmentation de l'objet est ainsi généralement effectuée au sens du mouvement. A partir de cette segmentation, un
- 15 ensemble de descripteurs peuvent être extraits et permettent de caractériser ce ou ces objet(s). Ces descripteurs, déterminés et/ou choisis par l'implantation réalisée du procédé, peuvent être bas, voire très bas niveau (couleur, histogramme de luminance, taille, etc.) ou de plus haut niveau (type d'objet, trajectoire, etc.). Ils enrichissent la segmentation d'un caractère
- 20 sémantique, permettant d'ajouter ces éléments à la description basée sur le mouvement. Il est ainsi possible de définir une personne, un véhicule ou encore le fond d'une image comme objet.

- Un objet ou un groupe d'objets pourra alors être définissable comme un ensemble d'objets liés par un descripteur commun choisi par l'utilisateur au
- 25 moment de la compression du flux vidéo. Un groupe d'objets contiendra ainsi par exemple l'ensemble des objets rouges, ceux se déplaçant vers la droite ou encore des personnes au sein d'une image.

- L'expression « portion » plus connue dans le domaine sous l'expression anglo-saxonne « slice » correspond à une sous-partie d'un objet constituée
- 30 de macroblocs qui appartiennent à un même ensemble défini par l'utilisateur.

Actuellement, plusieurs techniques sont utilisées pour la compression vidéo. Il est ainsi possible de citer les standards de compression vidéo MPEG, comme par exemple les standards MPEG-1, MPEG-2 ainsi que le standard MPEG-4. Le principe du standard de compression MPEG-4 est la
5 décomposition des scènes audiovisuelles en plusieurs objets médias hiérarchisés. Ces objets sont appelés objets audio visuels ou selon la dénomination anglo-saxonne Audio Visual Object (AVO). Un AVO peut être un flux vidéo, un flux audio, un texte, un graphique, etc. MPEG-4 fournit des méthodes de codage pour les objets individuels, par exemple MPEG-4 Part2
10 pour les objets vidéo et MPEG-4 Part3 pour les objets audio. Ainsi, une scène visuelle va être vue comme la composition d'AVO selon un script décrivant leurs relations spatiales et temporelles. L'accès individuel aux objets impose à la scène d'être représentée comme la composition d'objets divers qui seront ensuite assemblés pour recréer la scène. Cette approche
15 objet permet alors de très nombreuses opérations sur une séquence MPEG-4 : ajout, suppression ou déplacement d'un objet, transformation géométrique, changement de point de vue, etc.

Un objet vidéo ou en anglo-saxon Video Object (VO) est un objet auquel l'utilisateur peut accéder et qu'il peut manipuler aisément. Un objet vidéo VO
20 à un instant donné est un plan objet vidéo ou en anglo-saxon Video Object Plane (VOP) correspondant donc à un objet vidéo de forme arbitraire à un instant fixé. Afin d'augmenter les possibilités de manipulation, le standard MPEG-4 a aussi prévu des hiérarchies de plans vidéo objet VOPs associées à différents degrés d'accessibilité, un VOP peut être sous-divisé en plusieurs
25 sous-VOPs. En général, trois types de VOP sont considérés. Un unique VOP, la représentation se ramène alors au cas bien connu d'une unique scène. Deux VOPs, cas qui correspond au partitionnement de la scène en divers VOPs, en utilisant les séquences vidéo 2D traditionnelles en entrée. Deux VOPs ou plus permettent par combinaison la reconstruction d'une
30 scène à partir de plusieurs sources. Cela suppose alors que les données sont exploitables au début du processus de production. Une scène est une

composition d'objets séparés. Les informations relatives à la forme, au mouvement et à la texture des VOPs sont codées dans des couches d'objets vidéo VOL (pour Video Object Layer selon la dénomination anglo-saxonne) séparées afin de permettre le décodage séparé des VOPs. Le modèle de
5 vérification vidéo MPEG-4 VM, selon la dénomination anglo-saxonne « *Verification Model* », utilise un algorithme identique pour coder des informations relatives à la forme, le mouvement et la texture dans chaque couche. Cependant, l'information concernant la forme n'est pas transmise si la séquence qui doit être codée ne contient que des images de taille
10 rectangulaire. Chaque VOP représente une image dont le type peut être Intra (I), Prédicatif (P) ou Bidirectionnel (B). Un VOP de type I est codé indépendamment des autres, il fournit les points d'accès dans le train binaire. Il sert de référence au codage des VOP de type "P" et "B". Un VOP de type P peut être prédit par compensation de mouvement à partir d'un VOP "I" ou "P"
15 le précédant dans le flux. Il sert aussi au codage des images de type "B". Enfin pour un VOP de type B, les macroblobs qui le composent peuvent être prédits soit par simple compensation de mouvement avant ou arrière à partir de VOP "P" ou "I", soit par double compensation de mouvement avant et arrière.

20 En général, les images en entrée qui doivent être codées pour chaque VOP sont de forme arbitraire et la position et la forme des images varient dans le temps en respect d'une fenêtre de référence. Toutes les VOL qui doivent être codées pour une séquence en entrée sont définies en référence à la fenêtre de référence dont la taille est constante. En fait, lors du codage, les objets
25 présents dans un flux sont codés les uns par rapport aux autres d'une manière dépendante.

Avec cette technique, pour masquer certaines informations contenues dans une image, des calques ou encore des masques sont utilisés. Un des inconvénients de cette technique est qu'elle nécessite un flux de données
30 compressées par objet, et par conséquent un processus de décodage par objet à visualiser.

L'utilisation des portions jusqu'à MPEG-4 Part10 / H.264 s'avère délicat : si l'on considère qu'il est possible de définir pour chaque image la forme d'un objet, permettant une déclaration image par image de la carte d'allocation des macroblocs (ou selon l'expression anglo-saxonne MBAmapping pour
5 MacroBlocs Allocation map), la gestion dynamique de cette dernière n'est couverte par aucune procédure décrite dans la norme. Dès lors, les décodeurs actuels attendront une image Intra après toute mise à jour de la MBAmapping, ce qui s'avère très restrictif. De plus, les profils définis dans H.264 limitent le nombre de groupes de portions ou « slices-group » : 1 pour le
10 profil principal, « *main* » selon la dénomination anglo-saxonne (ce qui rend impossible tout découpage par objet sémantique), 8 pour les profils de base et étendu, « *baseline* » et « *extended* » selon la dénomination anglo-saxonne, ce qui limite à 7 le nombre d'objets mobiles dans une scène (le 8^{ème} slices group étant alors attribué au fond de l'image). Les deux cas
15 restent trop restrictifs dans le cas d'une surveillance de site.

Au sein d'un « slices group », chaque slice peut être défini comme une sous-portion du flux de macroblocs. On peut ainsi choisir un nombre de lignes de macrobloc au sein d'un objet, mais pas précisément une zone telle qu'un rectangle derrière un objet mobile.

20

L'objet de l'invention concerne une nouvelle structure hiérarchique de représentation d'une image s'appuyant sur une approche sémantique. Elle présente entre autres avantages celui de coder les objets présents dans un flux vidéo de manière indépendante tout en ne produisant qu'un unique flux
25 sans surcoût pour la compression.

L'invention concerne un procédé permettant de définir une structure d'une image dans un flux vidéo compressé composé d'une structure présentant l'une des caractéristiques au moins énoncée précédemment caractérisé en
30 ce qu'il comporte au moins les étapes suivantes :

- 5
- définir au sein d'une image au moins un objet ou un groupe d'objets, un groupe d'objets étant défini comme un ensemble d'objets liés par un ou plusieurs descripteur(s) commun(s) choisi au moment de la compression dudit flux, ledit objet comprenant une ou plusieurs portions ou « slices » correspondant à une sous-partie de l'image constituée de macroblochs qui appartiennent à un même ensemble, et
 - insérer au niveau d'un objet un ou plusieurs descripteur(s) ou à un groupe d'objets un ou plusieurs descripteur(s) commun(s).

10 Le procédé peut comporter au moins les étapes suivantes :

- insérer un en-tête image correspondant à l'identité d'une image présente dans ledit flux,
- insérer au moins un en-tête objet associé à un ou plusieurs objets présents dans l'image,
- 15 ➤ insérer au moins un en-tête slice associé à un desdits objets,
- positionner les données représentatives des objets définis.

Il insère par exemple un en-tête groupe d'objets en amont d'un en-tête objet pour désigner un ensemble d'objets.

Le flux vidéo compressé étant composé d'une ou plusieurs images ayant une structure présentant l'une des caractéristiques précitées, le procédé est

20 caractérisé en ce qu'il comporte au moins une étape de décodage comprenant la sélection de descripteurs associés à un ou plusieurs objets ou un ou plusieurs groupes d'objets et une étape de décodage desdits objets ou groupes d'objets sélectionnés et une étape de décompression desdits objets

25 ou groupes d'objets.

L'invention concerne aussi un dispositif permettant de générer une structure de trame d'une image dans un flux vidéo caractérisé en ce qu'il comporte au moins un moyen adapté à exécuter les étapes du procédé selon l'une des revendications précédentes.

30 Le dispositif peut comporter au moins les éléments suivants :
un codeur comportant au moins :

- un moyen d'analyse d'images d'un premier flux F1 afin de segmenter les images et d'en extraire les descripteurs souhaités pour un groupe d'objets ou un objet,
 - 5 ➤ un module permettant de définir une structure de représentation d'image précitée,
 - un codeur permettant de compresser les différents objets ou les groupes d'objets et les transmettre sous la forme d'un flux F2,
- un décodeur comportant au moins :
- 10 ➤ un module adapté à décoder les différentes en-têtes insérés dans ledit flux vidéo F2 et interpréter les informations qu'ils contiennent,
 - une mémoire tampon permettant d'exécuter un « pass-over » lors de la reconstitution d'un objet, permettant de laisser un bloc tel qu'il est dans la mémoire tampon du décodeur au moment de son décodage,
 - 15 ➤ un module adapté à identifier les différents objets ou groupes d'objets satisfaisants une requête de décodage,
 - un décodeur permettant de décompresser les données associées aux différents objets.

20 D'autres caractéristiques et avantages du dispositif selon l'invention apparaîtront mieux à la lecture de la description qui suit d'un exemple de réalisation donné à titre illustratif et nullement limitatif annexé des figures qui représentent :

- 25 ➤ les figures 1A et 1B, un exemple de structure image selon l'invention décrite dans la représentation de flux de données et aussi selon une distribution d'arbre,
- la figure 2, la manière dont est décrite une portion d'objet nécessitant une reconstruction différente du reste de l'objet,
- la figure 3, un exemple de mise en œuvre du décodage des objets dans une image ayant une structure selon l'invention,
- 30 ➤ la figure 4,

- la figure 5, la notion de passage multiréférence correspondant à la dénomination anglo-saxonne « multi-reference pass-over », et
- la figure 6, un exemple de dispositif associé.

5 L'objet de la présente invention concerne notamment une nouvelle structure hiérarchique de représentation de l'image qui met en œuvre une approche par objet visuel au sein d'un unique flux vidéo compressé. Cela permet entre autre de décoder les objets présents dans un unique flux de manière indépendante et sans coût additionnel. La nouvelle structure :

10

Image
-> Groupe d'objets
-> Objet
-> slice

15 repose sur la hiérarchie suivante au niveau de chaque image : une Image est composée de Groupe d'Objets, eux-mêmes comprenant des Objets, constitués de « Slices ». Une telle structure hiérarchique est, par exemple, décrite aux figures 1A et 1B. Contrairement au standard habituellement utilisé, l'objet de la présente invention concerne un unique flux vidéo ayant

20 une structure telle que celle décrite à la figure 1A. Dans certains cas d'application, les groupes d'objets peuvent comprendre plusieurs objets ayant un paramètre commun. Ainsi, un groupe d'objets est défini par exemple, par un descripteur commun choisi par l'utilisateur à la compression tel qu'il a été défini précédemment.

25 La figure 1A schématise un exemple de structure pour une image dans un unique flux vidéo selon l'invention. Le flux vidéo est découpé en plusieurs zones correspondant chacune à une image, désignée Z_1 pour l'image 1 et Z_i de manière plus générale.

Selon la structure hiérarchique proposée, le flux vidéo est composé de

30 plusieurs images, elles-mêmes composées d'un ou plusieurs groupe(s) d'objets ; chaque groupe d'objets est composé d'un ou plusieurs objet(s), eux-mêmes constitués de une ou plusieurs « slice(s) ». Un groupe d'objets peut se réduire dans certains cas un objet. Chacun de ces éléments de la

hiérarchie sera identifié au sein du flux par un en-tête propre, permettant une navigation et un accès optimisés en terme de recherche de sous portion du flux. En effet, les en-têtes pourront directement contenir les descripteurs qui ont permis de distinguer les slices / objets / groupes d'objets, et qui seront
5 donc exploitables au moment de la navigation. D'une part, la segmentation proposée repose sur le mouvement, puisque chaque objet est défini comme un ensemble connexe de pixels, blocs ou macroblocs présentant un déplacement propre au sein de la séquence d'images par rapport aux autres objets. D'autre part, c'est à partir de cette segmentation qu'un ensemble de
10 descripteurs sera extrait pour chacun des objets détectés, et qu'une classification (supervisée ou non) sera effectuée pour former les groupes d'objets. Le résultat obtenu permet donc, de par les différentes étapes qui y mènent, de parler d'indexation par le mouvement.

Le flux vidéo comprend, par exemple, un en-tête E_i correspondant à l'en-tête
15 de l'image i et contenant toutes les informations relatives à cette image, comme, par exemple, le nombre de groupes d'objets ou d'objets dans l'image, leur disposition, ainsi que les paramètres de compression de l'image. L'en-tête E_i est suivi d'un en-tête groupe d'objets $E_{go_{i-j}}$ qui contient notamment les descripteurs utilisés pour caractériser les groupes d'objets.
20 L'en-tête $E_{go_{i-j}}$ est suivi d'un en-tête objet $E_{o_{i-j-k}}$ qui contient notamment les descripteurs caractérisant cet objet. L'en-tête objet $E_{o_{i-j-k}}$ est suivi d'un en-tête slice $E_{s_{i-j-k-l}}$ qui sera utilisé pour la reconstruction de l'image au niveau d'un décodeur vidéo. Un en-tête slice est suivi des données associées $D_{i-j-k-l}$ contenues dans la portion ou « slice » de l'objet contenu dans l'image i .

25 Dans l'exemple des figures 1A et 1B, l'image 1 est décrite par ces en-têtes et données selon la description suivante. Au sein du flux peuvent être identifiés dans l'ordre de lecture : un en-tête image E_1 , un entête groupe d'objet $E_{go_{1-1}}$, un en-tête objet $E_{o_{1-1-1}}$, un en-tête slice $E_{s_{1-1-1-1}}$, des données $D_{1-1-1-1}$, un entête objet $E_{o_{1-1-2}}$, un en-tête slice $E_{s_{1-1-2-1}}$, des données $D_{1-1-2-1}$, un en-tête
30 slice $E_{s_{1-1-2-2}}$, des données $D_{1-1-2-2}$, un en-tête groupe d'objets $E_{go_{1-2}}$, un en-tête objet $E_{o_{1-2-1}}$, un en-tête slice $E_{s_{1-2-1-1}}$, des données $E_{s_{1-2-1-1}}$, un en-tête

slice $Es_{1-2-1-2}$, des données $D_{1-2-1-2}$ et un nouvel en-tête image E_2 . L'ordre de ces paquets est donné à titre illustratif et n'est aucunement restrictif. On pourra par exemple regrouper l'ensemble des en-têtes au commencement d'une nouvelle image dans le flux, qui seront suivis de l'ensemble des données (auxquelles on accédera via un pointeur présent dans l'en-tête correspondant ou un code de départ, ou *start code* selon l'expression anglo-saxonne, spécifique). Il est également possible d'ordonner les objets dans le flux selon leur taille : les petits objets seront privilégiés (a priori les objets mobiles) en début de structure image, ceux plus importants (dont fera vraisemblablement partie le fond) en fin de structure. Cela permettra d'accélérer encore le processus de décodage en cas de visualisation partielle.

La structure hiérarchique permet ainsi de reconstituer une scène complète décrite par l'arbre aux figures 1B et 3 : au sein de l'image 1 (en-tête E_1) se trouvent deux groupes d'objets (Ego_{1-1} et Ego_{1-2}). Le premier comprend deux objets (Eo_{1-1-1} et Eo_{1-1-2}). Le premier objet est constitué d'un unique slice ($Es_{1-1-1-1}$) alors que le second en comporte deux ($Es_{1-1-2-1}$ et $Es_{1-1-2-2}$). Le deuxième groupe d'objet est constitué d'un seul objet (E_{1-2-1}) comprenant deux slices déclarés par les en-têtes $Es_{1-2-1-1}$ et $Es_{1-2-1-2}$.

A la suite des dernières données représentatives de la première image, seront disposées de manière similaire les différents en-têtes et zones de données correspondant à une deuxième image, dont la répartition commence à partir de l'en-tête E_2 correspondant à la deuxième image. Cette structure se retrouve créée au niveau du codeur vidéo et permet au niveau du décodeur un accès direct à tout type d'objet ou groupe d'objet présent dans le flux. Un exemple de codeur et de décodeur sont décrits à la figure 5.

Les portions, ou slices, telles que définies au sein de cette structure hiérarchique, correspondent chacune à une région rectangulaire au sein d'un objet. Elles seront ainsi déclarées via deux macroblocs extrêmes sur le même principe qu'une boîte englobante, comme illustré dans la figure 2. L'un des principaux intérêts de cette définition est de permettre des codages

différents de sous-parties d'un même objet. En effet, la segmentation telle que proposée dans la structure permet de séparer les objets sémantiques différents. Mais même s'il est sémantiquement cohérent, un objet peut requérir des traitements différents selon les sous-portions qui le constituent.

5 Ainsi par exemple, dans la figure 4, la voiture à gauche V1 entre dans le champ de vision entre les deux images intégralement représentées. La voiture représente un objet sémantique cohérent et sera donc déclarée au niveau hiérarchique « objet ». Pour reconstruire la voiture V1 dans la seconde image I2, une estimation de mouvement peut être pratiquée sur
10 l'avant AV1 de la voiture qui était déjà présent dans l'image précédente I1. En revanche l'arrière ARV1 n'était pas dans le champ de vision de la caméra et ne peut donc être estimé par compensation de mouvement ; il sera donc codé Intra. La définition des « slices » telle que proposée ici permet donc au sein d'un même objet de fournir deux types de reconstruction différents, sans
15 avoir à déclarer celle-ci pour chaque bloc ou macrobloc, tout en préservant la structure hiérarchique. A l'exemple précédemment cité, on peut ajouter la possibilité de reconstruction après occlusion par mécanisme de *multi-reference pass-over*, tel que décrit ci-après, et rendu possible par cette définition des slices.

20 La figure 2 illustre un tel découpage en slice pour un objet, en représentant un objet 1 dans lequel une partie 2 nécessite une reconstruction différente : deux macroblocs extrêmes (3 et 4) définissent un slice correspondant à la partie hachurée I et nécessitant par exemple une reconstruction par blocs intra contre une prédiction normale pour la partie II qui restera par exemple
25 inchangée par rapport à l'image précédente via un vecteur estimation de mouvement nul ou un mécanisme de pass-over comme décrit ci-après. Cette nouvelle définition de slice détermine une portion rectangulaire au sein d'un objet et non une sous-portion dans le flux vidéo comme il est généralement trouvé.

30 La figure 3 représente une scène qui est décrite selon la hiérarchie présentée à la figure 1B : le flux y est représenté par deux images, l'image n

et l'image n+1. L'image n est sous-divisée en 2 groupes d'objets ; le premier correspond aux objets mobiles (2 véhicules), le second au fond. Le groupe « objets mobiles » comprend donc deux véhicules. La première voiture est reconstruite à partir d'un unique « slice » $Es_{1-1-1-1}$ ou portion, par estimation de mouvement par exemple. Le second véhicule comporte quant à lui « deux slices » $Es_{1-1-2-1}$ et $Es_{1-1-2-2}$: cela pourra s'expliquer par le fait qu'il 'rentre' dans l'image, permettant une estimation de mouvement sur l'avant du véhicule alors que l'arrière devra être codé intra. Un deuxième groupe d'objets (groupe n°2) de l'image n correspond au fond de l'image, immobile ou quasi-immobile. A ce groupe d'objets 2 est associé un unique objet 2-1 auquel sont associés deux « slices » $Es_{1-2-1-1}$ et $Es_{1-2-1-2}$.

Le découpage des objets s'accompagne d'une restriction de la compensation de mouvement d'une image sur l'autre à l'intérieur de l'objet. Ainsi, chaque macrobloc de l'objet dans l'image N ne sera prédit qu'à partir de ce même objet (dans la ou les image(s) référence selon le type d'image en cours de compression). Cette étape requiert la prédiction de la position de l'objet dans l'image en cours au moment de la compression. Cela peut être réalisé en utilisant une méthode de suivi, plus connue sous la dénomination anglo-saxonne « tracking ». Avant d'initialiser l'estimation et la compensation de mouvement, chaque objet est ainsi déclaré et positionné. Chaque objet est alors uniquement prédit à partir de ses précédentes occurrences, permettant un codage indépendant des objets dans le flux.

La déclaration des objets se fait par exemple en utilisant une MBAmapping semblable à celle présente dans le standard MPEG-4 part 10/H.264. Cette carte pourra elle-même être compressée en appliquant une compensation de mouvement similaire à celle utilisée sur le reste de l'image. Le procédé prendra en compte cette mise à jour image par image de la MBAmapping, modifiant constamment la disposition des slices/objets dans l'image. Le décodeur sera fortement impacté par ce renouvellement constant.

De fait, la nouvelle structure de données selon l'invention, permet de décoder et éventuellement d'afficher sur un moniteur uniquement le ou les objets

requis au cours de la séquence sans rencontrer de problèmes de reconstruction. Les objets non décodés laisseront quant à eux des espaces sans information. Cela peut être très pertinent selon les applications visées. Par exemple, seuls les véhicules sont nécessaires pour la reconnaissance
5 de plaque d'immatriculation ; seuls le fond et les acteurs présents dans une sous-fenêtre de l'image doivent être décodés si l'utilisateur a zoomé lors du visionnage.

Ce découpage qui s'accompagne donc de la possibilité de décoder indépendamment les objets, apporte un nouvel accès aux données. Il sera
10 alors possible d'accéder directement à un niveau de hiérarchie donné dans la structure ou trame correspondant à l'image, donc aux parties de l'image considérée comme significative, en naviguant rapidement au sein des entêtes image/groupe d'objets/objets ou « slice » dans le flux et de fait atteindre plus vite l'information pertinente. Cette navigation peut s'effectuer de
15 plusieurs façons. Soit l'utilisateur sait ce qu'il cherche (par exemple un objet rouge) auquel cas lorsqu'il demandera le décodage du flux vidéo il déclenchera une requête demandant de ne décoder que les groupes d'objets ou objets dont les descripteurs correspondent à ce qu'il souhaite. La navigation peut également s'effectuer de façon interactive, lors d'une
20 première étape de « parsing » du flux vidéo, les différents descripteurs présents dans le flux peuvent être récupérés. Connaissant ces descripteurs l'utilisateur peut alors faire une requête spécifique indiquant quelle(s) image(s), quel(s) groupe(s) d'objets ou quel(s) objet(s) il souhaite décoder. Le décodage « interactif » peut également s'effectuer en décodant une
25 image sur laquelle l'utilisateur va sélectionner (en cliquant dessus par exemple) les portions de l'image qu'il souhaite continuer à visualiser avant de lancer la décompression de la suite du flux. Le procédé selon l'invention met ainsi en œuvre une navigation dite « sémantique » basée sur une définition de hiérarchie ou de catégorie d'objets plutôt qu'une autre, contrairement à
30 l'art antérieur qui utilise une sous-portion temporelle ou spatiale d'une séquence.

La présence des descripteurs au cœur du flux permet également un procédé d'indexation et donc de requête à partir d'un unique flux. Les métadonnées sont noyées dans la vidéo compressée et ne requièrent pas de transmission et/ou codage séparé. Une implantation du procédé permettra donc d'utiliser ces données sans avoir à recourir à d'autres outils comme c'est le cas dans MPEG-4, où la partie 2 ou 10 de codage vidéo doit être accompagnée de la partie de gestion des métadonnées (MPEG-4 Part1).

Sur la figure 5 est représentée la notion de passage multiréférence correspondant à la dénomination anglo-saxonne « multi-reference pass-over », notion qui n'est pas connue actuellement à la connaissance du Demandeur. Le « pass-over » permet de laisser un bloc tel qu'il est dans la mémoire tampon du décodeur au moment de son décodage. Il correspond à une prédiction de mouvement avec un vecteur nul. L'apport de la multiréférence joue notamment lorsqu'un objet fixe, le fond d'une image par exemple, subit une occlusion pendant plusieurs images et est ensuite à nouveau présent. Avant l'occlusion, le procédé utilise des « pass-over » d'une image, c'est-à-dire que la mémoire tampon du dispositif permettant la reconstitution d'un objet reste inchangé dans la zone définie. Lors de l'occlusion il n'y a plus de reconstruction de l'objet considéré sur la zone masquée, derrière l'objet mobile. Enfin, lorsque la partie masquée est à nouveau visible il est alors possible de reconstruire par un pointeur sur une image précédant l'occlusion, cette image référence étant indiquée N-x sur la figure 4. Une unique déclaration au niveau slice permettra de reconstruire toute une zone par mécanisme de pass-over, sans nécessiter d'information supplémentaire pour chaque bloc ou macrobloc. Cette nouvelle reconstruction de l'image nécessite de conserver dans la mémoire tampon plusieurs images précédant l'image actuelle, vers lesquels pointeront les pass-over définis au niveau du codeur. Au niveau du codeur, il sera donc nécessaire de déterminer les occlusions d'une partie d'un objet pour chaque image, ce qui pourra se faire par exemple par une étude du mouvement global de chaque objet par rapport aux autres. Les parties précédemment

masquées seront alors préférentiellement reconstruites par pass-over et non plus par blocs Intra comme ce pouvait être le cas avec les standards actuels. Sur le plan de la compression, la transmission de la table d'allocation MBAmapping en utilisant la compensation de mouvement et du fait de sa faible
5 résolution n'aura que peu d'impact en terme de débit pour le flux vidéo.

En revanche, avec un découpage de l'image selon le principe énoncé précédemment et qui retranscrit au plus près la segmentation au sens du mouvement des différents objets indépendamment les uns des autres, il devient possible de faire de l'effacement de vecteur estimation de
10 mouvement, ceux-ci pouvant être reconstruits au niveau du décodeur par rapport au déplacement, la forme, la taille ou encore toute autre particularité de l'objet.

Pour mettre en œuvre l'invention le dispositif selon l'invention, représenté 10 sur la figure 5, comporte un codeur et un décodeur comprenant, par
15 exemple :

au niveau du codeur

- un moyen d'analyse d'images 11 du flux F1 afin de segmenter les images et d'en extraire les descripteurs souhaités pour un groupe d'objets ou un objet représenté par la même référence 12,
- 20 ➤ un module 13 permettant de définir une structure de représentation d'image selon l'invention. Ce module, à partir de l'analyse précédemment réalisée, va créer les en-têtes appropriés, par exemple No_i en-têtes objets pour une image i , éventuellement Ngo_i en-têtes groupe d'objets pour les groupes d'objets contenus dans l'image i ,
25 Nso_i pour les en-têtes des slices inclus dans les objets de l'image i ,
- un codeur 14 permettant de compresser les différents objets ou les groupes d'objets et les transmettre sous la forme d'un flux F2.

au niveau du décodeur 20

- un module 21 capable de décoder les différentes en-têtes insérés
30 dans le flux vidéo et d'interpréter les informations qu'ils contiennent,

- une mémoire tampon 22 permettant d'exécuter un mécanisme de « pass-over » lors de la reconstitution d'un objet, le module mémoire tampon 22 a pour entrées le module 21 et le module décodeur 24, et pour sortie le module décodeur 24 (24 est en E/S puisqu'un bouclage a lieu entre les anciennes images et celles reconstruites).
5
- un module 23 permettant d'identifier les différents objets ou groupes d'objets satisfaisants la requête de décodage effectuée par l'utilisateur,
- un décodeur 24 permettant de décompresser les données associées aux différents objets.
10

L'objet de la présente invention offre notamment comme avantages d'offrir une structure de trame dans laquelle les objets sont accessibles individuellement et décodables indépendamment. Cela permet de gagner en
15 temps de décodage de l'image car il sera alors possible d'accéder directement à un niveau de hiérarchie donné, donc aux parties de l'image considérées comme significatives, en naviguant rapidement au sein des entêtes image / groupes d'objets / objets ou slice dans le flux. La nouvelle structure selon l'invention permet notamment d'évoquer la notion de
20 navigation sémantique, puisqu'au lieu de choisir une sous-portion spatiale ou temporelle d'une séquence, elle permet d'opter pour une catégorie d'objets plutôt qu'une autre. Cela complète et étend la notion « d'accès universel au flux » apparue dans MPEG-4. Le nouveau concept de multi-référence pass-over permet une reconstruction avec un coût de codage minimal après
25 occlusion.

Contrairement au standard actuel tel que MPEG-4, la représentation d'une scène sémantiquement ne requiert pas de schéma de description de scène ni de placement de chaque objet dans la scène. Les Métadonnées sont présentes directement dans le flux. Il n'est donc pas nécessaire d'implanter
30 tout un standard, seule la sous-partie liée au codeur vidéo permet d'accéder à divers éléments du flux.

Le procédé et la structure nouvelle d'images selon l'invention offrent la possibilité de déclarer autant d'objets et groupes d'objets qu'il est nécessaire et décidé par un utilisateur. Ainsi la non limitation du nombre d'objets et groupe d'objets permet d'envisager le traitement de séquences haute
5 résolution dans l'avenir.

REVENDICATIONS

1 - Procédé permettant de définir une structure d'une image dans un flux vidéo compressé composé d'images (t) ayant une image (i) comprenant au moins un objet, ledit objet comprenant une ou plusieurs portions ou "slices" correspondant à une sous-partie de l'image constitué de macroblochs qui appartiennent à un même ensemble caractérisé en ce qu'il comporte au moins les étapes suivantes :

- définir au sein d'une image (i) au moins un objet o_{i-i-i} ou un groupe d'objets go_{i-i-i} , un groupe d'objets étant défini comme un ensemble d'objets liés par un ou plusieurs descripteur(s) commun(s) choisi au moment de la compression dudit flux, ledit objet comprenant une ou plusieurs portions ou « slices » $s_{i-i-i-i}$ correspondant à une sous-partie de l'image constituée de macroblochs qui appartiennent à un même ensemble, et
- insérer au niveau d'un objet un ou plusieurs descripteur(s) ou à un groupe d'objets un ou plusieurs descripteur(s) commun(s).

2 – Procédé selon la revendication 1, caractérisé en ce qu'il comporte au moins les étapes suivantes :

- insérer un en-tête image E_i correspondant à l'identité d'une image (i) présente dans ledit flux,
- insérer au moins un en-tête objet Eo_{i-j-k} associé à un ou plusieurs objets présents dans l'image,
- insérer au moins un en-tête slice $Es_{i-j-k-l}$ associé à un desdits objets,
- positionner les données $D_{i-j-k-l}$ représentatives des objets définis.

3 – Procédé selon la revendication 2, caractérisé en ce qu'il insère un en-tête groupe d'objets en amont d'un en-tête objet pour désigner un ensemble d'objets.

4 – Procédé selon la revendication 1, caractérisé en ce que l'on détermine une portion rectangulaire au sein d'un objet Eo_{i-i-i} au moyen d'une portion ou slice déclaré par un en-tête Es_{i-i-i} et en ce que ladite portion est délimitée par deux macroblocs extrêmes.

5

5 - Procédé selon l'une des revendications précédentes dans lequel le flux vidéo compressé est composé d'une ou plusieurs images ayant une structure selon la revendication 1, caractérisé en ce qu'il comporte au moins une étape de décodage comprenant la sélection de descripteurs associés à un ou
10 plusieurs objets ou un ou plusieurs groupes d'objets et une étape de décodage desdits objets ou groupes d'objets sélectionnés et une étape de décompression desdits objets ou groupes d'objets.

6 – Dispositif permettant de générer une structure de trame d'une image
15 dans un flux vidéo, caractérisé en ce qu'il comporte au moins les éléments suivants :

un codeur comportant au moins :

- un moyen d'analyse d'images (11) d'un premier flux F1 afin de segmenter les images (i) et d'en extraire les descripteurs souhaités
20 pour un groupe d'objets ou un objet,
- un module (13) permettant de définir une structure de représentation d'image selon l'une des revendications 1 à 10,
- un codeur (14) permettant de compresser les différents objets ou les groupes d'objets et les transmettre sous la forme d'un flux F2,

25 un décodeur comportant au moins :

- un module (21) adapté à décoder les différentes en-têtes insérés dans ledit flux vidéo F2 et interpréter les informations qu'ils contiennent,
- une mémoire tampon (22) permettant d'exécuter un « pass-over » lors
30 de la reconstitution d'un objet, permettant de laisser un bloc tel qu'il est dans la mémoire tampon du décodeur au moment de son décodage,

- un module (23) adapté à identifier les différents objets ou groupes d'objets satisfaisants une requête de décodage,
- un décodeur (24) permettant de décompresser les données associées aux différents objets.

FIG. 1A

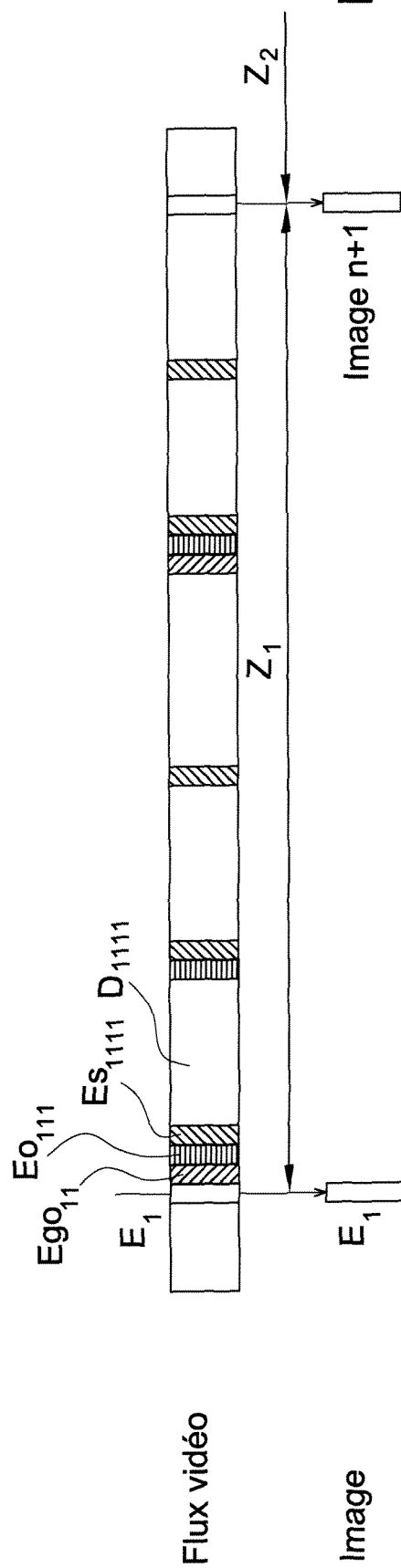
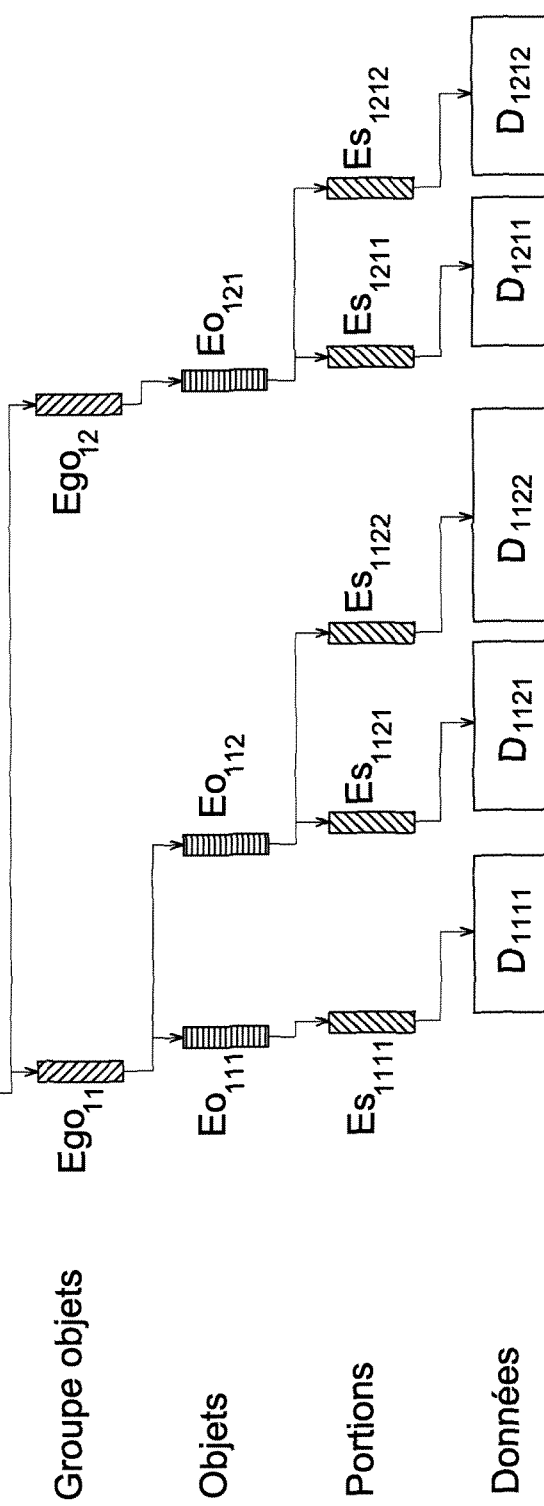


FIG. 1B



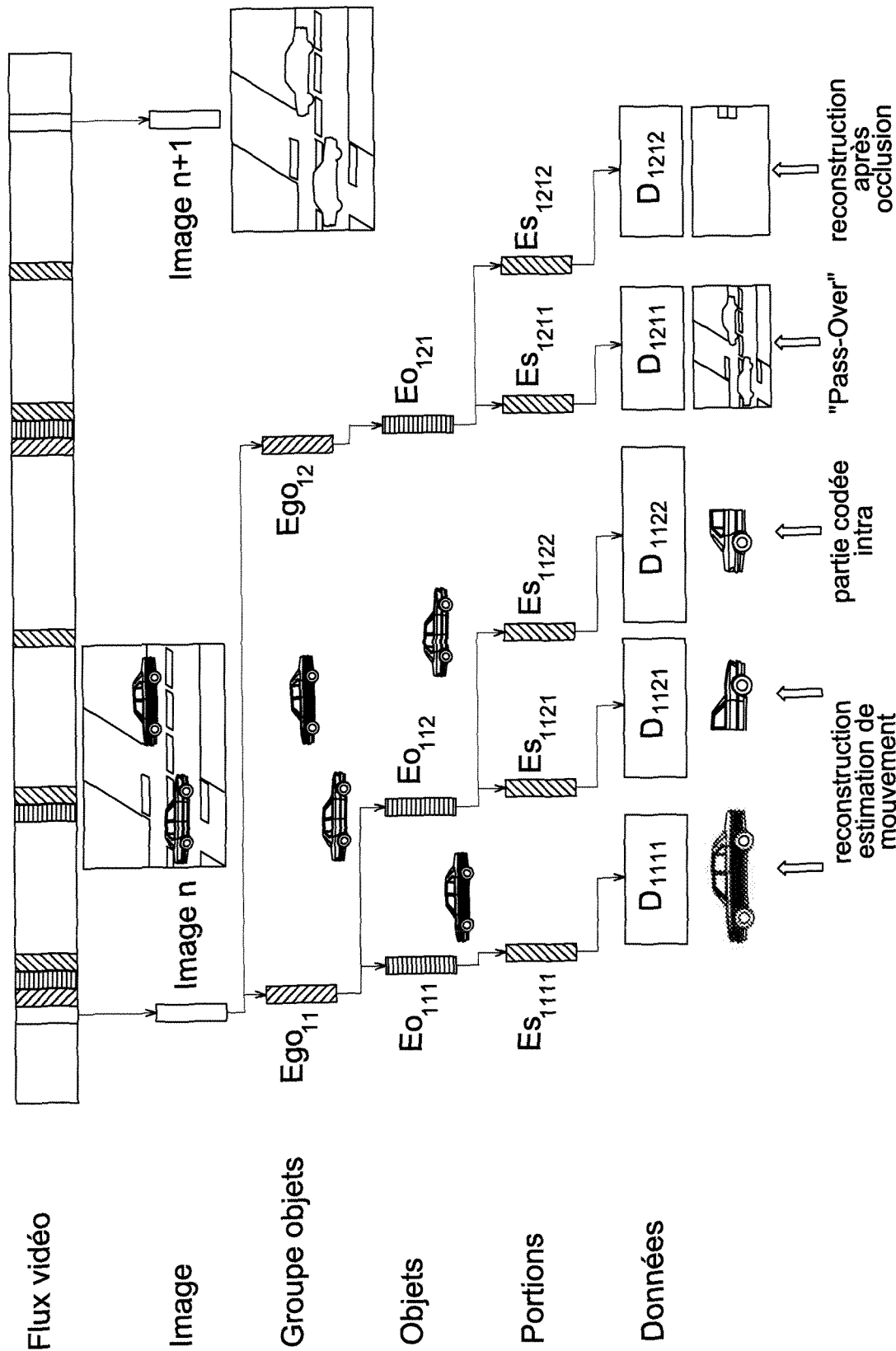


FIG.3

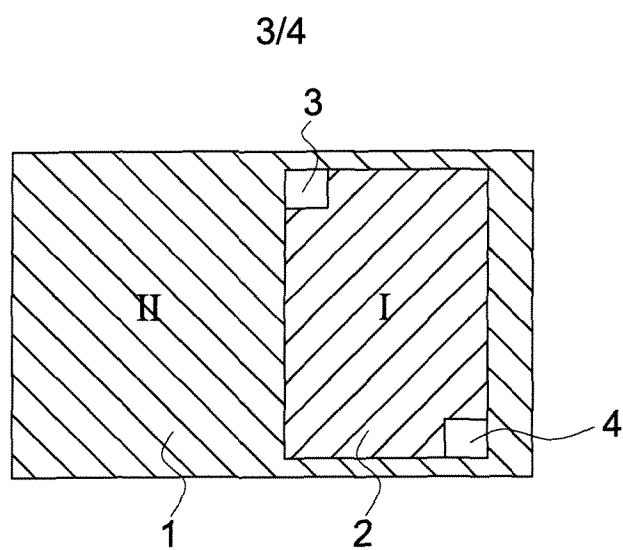


FIG. 2

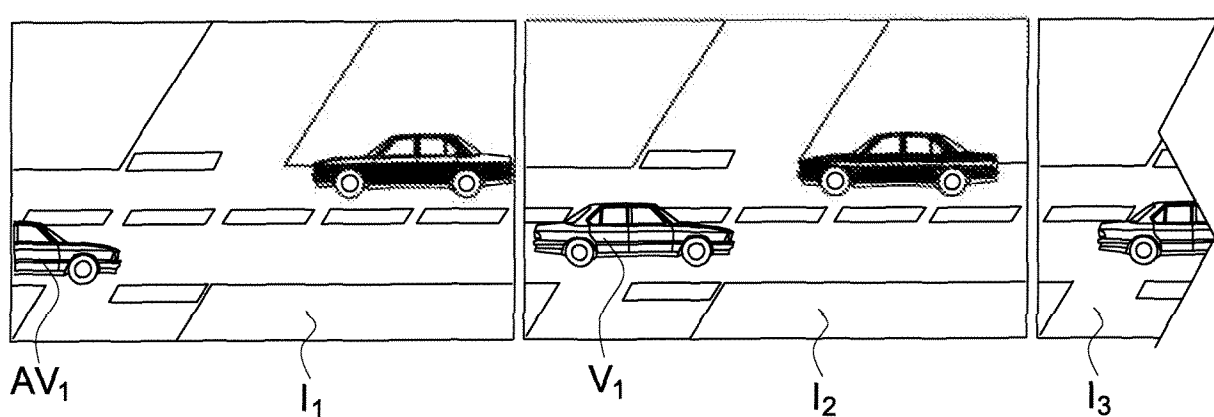


FIG. 4

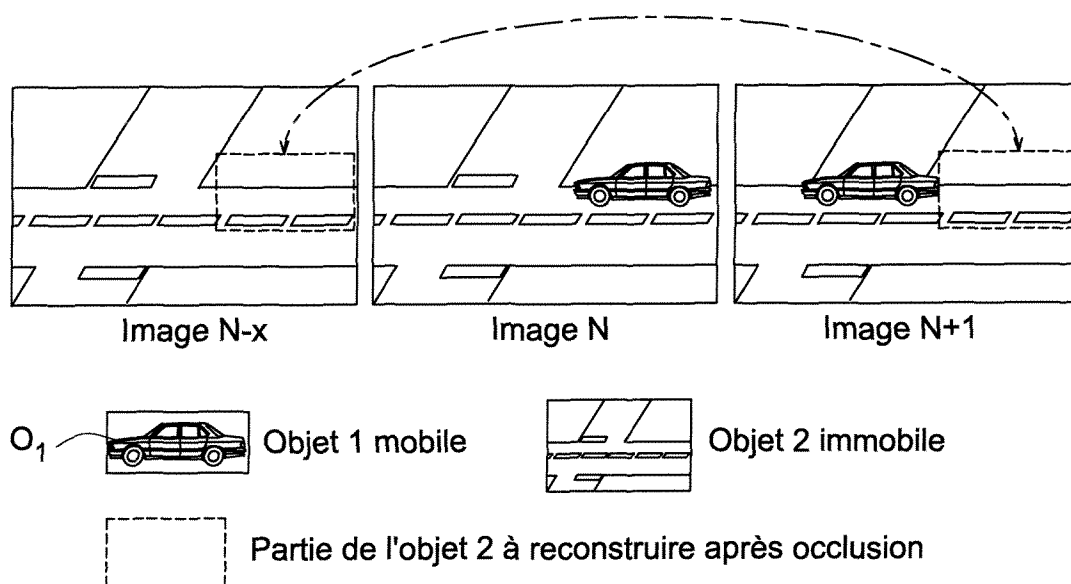


FIG. 5

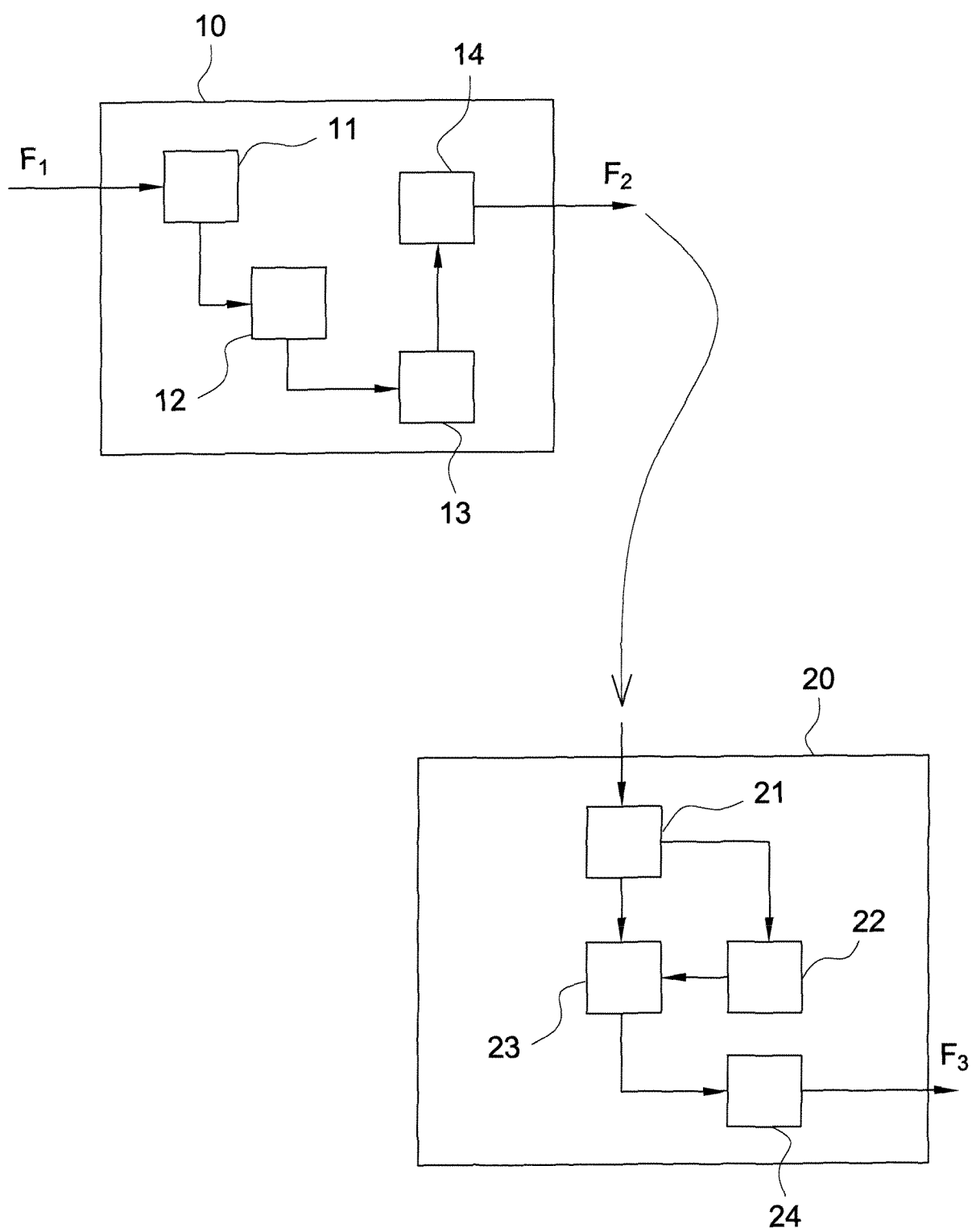


FIG.6

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2009/056828

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DHONDT Y ET AL: "Flexible macroblock ordering as a content adaptation tool in H.264/AVC" PROCEEDINGS OF THE SPIE, SPIE, BELLINGHAM, VA; US, vol. 6015, 24 October 2005 (2005-10-24), pages 601506/1-9, XP007901849 the whole document paragraph [0002] ----- -/--	1-6

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *G* document member of the same patent family

Date of the actual completion of the international search

5 August 2009

Date of mailing of the international search report

18/08/2009

Name and mailing address of the ISA/
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Schneiderlin, Jean

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/056828

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	RONG LUO ET AL: "A Hierarchical Scheme of Flexible Macroblock Ordering for ROI based H.264/AVC Video Coding" ADVANCED COMMUNICATION TECHNOLOGY, 2008. ICACT 2008. 10TH INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 17 February 2008 (2008-02-17), pages 1579-1582, XP031245410 ISBN: 978-89-5519-136-3 the whole document paragraph [0003] -----	1-3,5

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°

PCT/EP2009/056828

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
INV. H04N7/26

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

H04N

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

EPO-Internal

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	DHONDT Y ET AL: "Flexible macroblock ordering as a content adaptation tool in H.264/AVC" PROCEEDINGS OF THE SPIE, SPIE, BELLINGHAM, VA; US, vol. 6015, 24 octobre 2005 (2005-10-24), pages 601506/1-9, XP007901849 le document en entier alinéa [0002] ----- -/--	1-6



Voir la suite du cadre C pour la fin de la liste des documents



Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

- *A* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- *E* document antérieur, mais publié à la date de dépôt international ou après cette date
- *L* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- *O* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- *P* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

T document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

X document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

Y document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

Z document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

5 août 2009

Date d'expédition du présent rapport de recherche internationale

18/08/2009

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Schneiderlin, Jean

C(suite). DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	RONG LUO ET AL: "A Hierarchical Scheme of Flexible Macroblock Ordering for ROI based H.264/AVC Video Coding" ADVANCED COMMUNICATION TECHNOLOGY, 2008. ICACT 2008. 10TH INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 17 février 2008 (2008-02-17), pages 1579-1582, XP031245410 ISBN: 978-89-5519-136-3 le document en entier alinéa [0003] -----	1-3,5