

19 RÉPUBLIQUE FRANÇAISE
**INSTITUT NATIONAL
 DE LA PROPRIÉTÉ INDUSTRIELLE**
 COURBEVOIE

11 N° de publication : **3 153 176**

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

21 N° d'enregistrement national : **23 09711**

51 Int Cl⁸ : **G 06 T 9/00 (2023.01), H 04 N 19/90, G 06 N 3/02**

12

DEMANDE DE BREVET D'INVENTION

A1

22 Date de dépôt : 14.09.23.

30 Priorité :

43 Date de mise à la disposition du public de la
 demande : 21.03.25 Bulletin 25/12.

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

60 Références à d'autres documents nationaux
 apparentés :

○ Demande(s) d'extension :

71 Demandeur(s) : **ORANGE Société anonyme** — FR.

72 Inventeur(s) : **LADUNE Theo, PHILIPPE Pierrick et
 LEGUAY Thomas.**

73 Titulaire(s) : **ORANGE Société anonyme.**

54 **Procédé et dispositif de codage et décodage
 d'images avec mécanisme d'attention.**

57 Procédé et dispositif de codage et décodage d'images

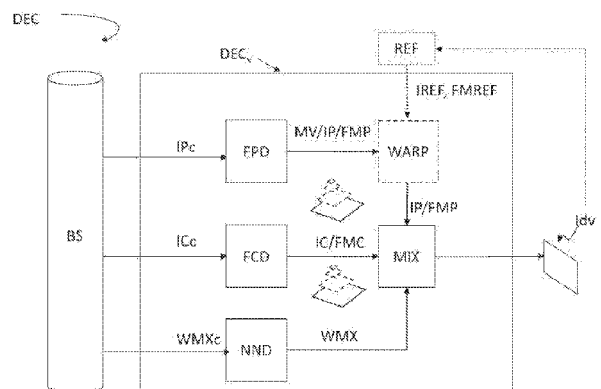
avec mécanisme d'attention

L'invention concerne un procédé et un dispositif de décodage d'une image courante (Iv) d'une séquence d'images (Si).

Le procédé comporte les étapes suivantes:

- obtention (E31) d'un ensemble d'informations de correction (IC, FMC) de ladite au moins une image courante;
- obtention (E32) d'un ensemble d'informations de prédiction (IP, FMP) de ladite au moins une image courante;
- décodage (E34) d'un ensemble de paramètres (WMX) représentatifs d'au moins un réseau de neurones de reconstruction (MIX);
- traitement (E35) desdites informations de correction et desdites informations de prédiction par ledit au moins un réseau de neurones de reconstruction pour produire une représentation de l'image courante décodée (FMREC, Ivd).

Figure pour l'abrégié : Fig. 1.



FR 3 153 176 - A1



Description

Titre de l'invention : Procédé et dispositif de codage et décodage d'images avec mécanisme d'attention.

Technique antérieure

- [0001] L'invention se rapporte au domaine général du codage de séquences d'images numériques. Elle concerne plus particulièrement la compression des vidéos numériques.
- [0002] Les vidéos numériques font en général l'objet d'un codage source visant à les compresser afin de limiter les ressources nécessaires à leur transmission et/ou à leur stockage. Il existe de nombreux standards de codage, tels que les standards des organismes ITU/MPEG (H.264/AVC, H.265/HEVC, H.266/VVC, etc.) ainsi que leurs extensions (MVC, SVC, 3D-HEVC, etc.).
- [0003] L'encodage d'une image est généralement effectué par une prédiction des pixels à l'aide de pixels précédemment codés puis décodés présents dans l'image en cours d'encodage, auquel cas on parle de « prédiction Intra », ou d'images précédemment codées, auquel cas on parle de « prédiction Inter ».
- [0004] Dans le cas d'une prédiction inter, une approche classique est souvent utilisée : après estimation et compensation du mouvement au codeur, une trame résiduelle est calculée en soustrayant la trame de prédiction de l'image originale. Au décodeur, la trame de prédiction et la trame résiduelle sont ajoutées pour obtenir la trame reconstruite. Cette simple addition n'étant pas optimale, de nouvelles approches, notamment neuronales, se sont développées récemment.
- [0005] Dans ce domaine, une approche neuronale de type conditionnelle est proposée par exemple dans le document "Conditional Residual Coding: A Remedy for Bottleneck Problems in Conditional Inter Frame Coding", de F. Brand et al (arXiv:2307.12864). Le codage conditionnel proposé modélise la distribution conditionnelle d'une image en fonction de son prédicteur. Toutefois ([Fig.3] de ce document), les données résiduelles et les données prédites sont considérées comme étant d'importance égale. Ceci limite la performance de compression.
- [0006] Il existe donc un besoin pour une solution permettant de combiner un signal résiduel à un signal prédit dans une séquence d'images de manière plus efficace.

Exposé de l'invention

- [0007] L'invention vise un procédé de codage d'une image courante au moins d'une séquence d'images, comportant les étapes suivantes :
- génération d'un ensemble d'informations de prédiction de ladite au moins une image courante ;
 - génération d'un ensemble d'informations de correction de ladite au moins une

image courante ;

- génération d'au moins un réseau de neurones de reconstruction, défini par un ensemble de paramètres de reconstruction ;
- traitement desdites informations de correction et desdites informations de prédiction par ledit au moins un réseau de neurones de reconstruction pour produire une représentation de l'image courante ;
- mise à jour d'une partie au moins desdites informations de prédiction et/ou desdites informations de correction et/ou desdits paramètres de reconstruction, en fonction d'une mesure de performance de codage ;
- codage d'une partie au moins desdites informations de correction et/ou de prédiction et desdits paramètres de reconstruction.

[0008] L'invention vise aussi un procédé de décodage d'une image courante d'une séquence d'images, comportant les étapes suivantes :

- obtention d'un ensemble d'informations de correction de ladite image courante ;
- obtention d'un ensemble d'informations de prédiction de ladite image courante ;
- décodage d'un ensemble de paramètres de reconstruction représentatifs d'au moins un réseau de neurones de reconstruction;
- traitement desdites informations de correction et desdites informations de prédiction par ledit au moins un réseau de neurones de reconstruction pour produire une représentation de l'image courante décodée.

[0009] Au sens de l'invention, on entend par encodage, ou « codage », l'opération qui consiste à représenter un ensemble d'échantillons, ou pixels, sous une forme compacte portée par exemple par un train binaire numérique. On entend par décodage l'opération qui consiste à traiter un train binaire numérique pour restituer des échantillons décodés.

[0010] Par « séquence d'images » on entend une pluralité d'images bidimensionnelles ordonnées, par exemple temporellement dans le cas d'une vidéo. Selon un exemple, la séquence correspond à une scène. Selon un exemple la séquence correspond à un ensemble d'images prédéfinies, par exemple un nombre fixe, ou, au sens des normes MPEG, un GOP (Group Of Pictures) comprenant les images situées entre deux images de type Intra-images, aussi appelé « période intra ». Selon un autre exemple, les images peuvent être des vues d'une même scène représentée en multivues. Selon un autre exemple, les images peuvent être une pluralité d'images temporelles et multivues (vidéo immersive).

[0011] Par « image courante » on entend une image de la séquence.

[0012] Par "informations de prédiction" on entend toute information utile à la génération d'une prédiction de l'image courante. Classiquement, ces informations peuvent être des informations de prédiction inter (mouvement, images de référence, etc.), de prédiction intra (direction de prédiction), de sélection de mode (mode intra ou inter ou bien

combiné intra/inter ou de combinaison), etc.

- [0013] Par "information de correction" on entend toute information utile à la génération d'une correction de la prédiction de l'image courante. Classiquement, ces informations peuvent être des informations résiduelles, par exemple une image résiduelle.
- [0014] Les "informations de prédiction" et "informations de correction" sont les entrées du réseau de neurones de reconstruction. Elles peuvent contenir des données dans le domaine image ou dans le domaine latent. Par exemple, une image de prédiction (resp. correction) peut se composer de pixels ou de cartes de caractéristiques représentatives de l'image de prédiction (resp. correction) dans le domaine latent.
- [0015] On appelle "domaine latent" une représentation composée d'un ensemble de cartes de caractéristiques sous forme bidimensionnelle, dont la taille est proportionnelle à l'image d'entrée. Il peut aussi comprendre un ensemble de vecteurs unidimensionnels, ou un ensemble de valeurs scalaires.
- [0016] Par « réseau de neurones de reconstruction », on entend un réseau neuronal tel un réseau neuronal convolutif, un perceptron multicouche, un LSTM (pour « Long Short Term Memory » en anglais), etc. Le réseau de neurones est défini par exemple par une pluralité de couches de neurones artificiels et par un ensemble de fonctions d'activation, de pondération et d'addition (par exemple, une couche peut calculer $y = f(Ax+b)$, où y et b sont des vecteurs de dimension N , x un vecteur de dimension M , A est une matrice de dimension $M \times N$, et f est la fonction d'activation). De préférence le réseau de neurones de reconstruction est convolutif, c'est-à-dire qu'il comporte au moins une couche de convolution. Une pluralité de tels réseaux peut être mise en cascade.
- [0017] Par « paramètre du réseau de neurones », on entend une des valeurs qui caractérise le réseau de neurones, par un exemple un poids associé à l'un des neurones (coefficient de filtre ou de convolution, pondération, biais, valeur affectant le fonctionnement de la non-linéarité, etc.)
- [0018] Par « traitement (de valeurs d'entrée) par un réseau de neurones de reconstruction », on entend l'application d'une fonction exprimée par un réseau de neurones aux valeurs d'entrée pour produire des valeurs de sortie représentatives des échantillons de l'image courante à décoder (reconstruire) (resp. coder).
- [0019] Par "représentation de l'image courante" on entend sa représentation dans le domaine image (représentation sous forme de pixels) ou latent (représentation sous forme de cartes de caractéristiques).
- [0020] Par « mesure de performance », on entend une mesure entre au moins une valeur d'un échantillon à coder et une valeur décodée dudit échantillon. La mesure peut évaluer par exemple une distorsion, ou une erreur perceptive. Elle peut être effectuée sur un échantillon ou une pluralité d'échantillons (par exemple, les échantillons

courants, ou l'image courante, etc.). La mesure peut comporter aussi une mesure du débit, notamment associé au codage du réseau de neurones et/ou au codage des cartes de caractéristiques. La mesure peut être une mesure conjointe entre le débit et la distorsion au travers de leur pondération. Comme il est bien connu de l'état de l'art, il est généralement procédé à une minimisation de la valeur de cette mesure jusqu'à atteindre une valeur cible ou minimale, ou un temps prédéfini.

- [0021] Par « échantillon », on entend une valeur prélevée dans une image de la séquence. L'échantillonnage d'un signal produit une suite de valeurs discrètes nommées échantillons. Dans le cas d'un signal d'image, l'échantillon prend le nom de pixel qui peut être par exemple un pixel couleur représenté traditionnellement par un triplet de valeurs, par exemple (R,G,B) ou (Y,U,V). La position de l'échantillon est repérée par ses coordonnées en abscisse (x) et ordonnée (y) dans l'image.
- [0022] De manière générale, on considère que les étapes d'un procédé de codage ou décodage ne doivent pas être interprétées comme étant liées à une notion de succession temporelle. Autrement dit, les étapes peuvent être effectuées dans un ordre différent de celui indiqué dans la revendication indépendante de codage ou de décodage, voire en parallèle.
- [0023] Le procédé de codage selon l'invention réalise au cours de son apprentissage une construction des paramètres de codage, à partir d'une séquence d'images d'entrée, entraînant un réseau de neurones de reconstruction sur des ensembles de données d'informations de prédiction et de correction, pour obtenir une reconstruction fidèle de l'image. Avantageusement, le procédé de codage encode le réseau de neurones de reconstruction. Celui-ci peut être optimisé, et adapté en complexité ou en qualité, selon la complexité visée et le rapport visé entre débit et distorsion. De même, les informations de prédiction et de correction encodées et transmises peuvent ainsi être optimisées et adaptées.
- [0024] Au cours de l'entraînement, ou construction, ou apprentissage, les paramètres du réseau de neurones et les informations à coder sont mis à jour en fonction d'une mesure de performance, par exemple de type débit-distorsion. Lorsque l'entraînement est terminé, c'est-à-dire que la mesure de performance obtenue est satisfaisante, le codage effectif des paramètres du réseau de neurones de reconstruction ainsi que celui des informations de prédiction et de reconstruction peut être effectué et le résultat du codage mémorisé ou transmis à destination du décodeur. Avantageusement, le processus d'entraînement permet donc de raffiner les paramètres du réseau de neurones de reconstruction, ainsi que ses paramètres d'entrée, jusqu'à obtenir une représentation adéquate en termes de performance, par exemple un équilibre souhaité entre le débit généré et la distorsion subie par l'image d'entrée en cours de codage. Avantageusement, le procédé de codage selon l'invention permet de compresser efficacement

le signal.

- [0025] Avantageusement, le procédé de décodage est simple puisqu'il suffit d'obtenir les informations de correction et de prédiction et le réseau de neurones de reconstruction pour reconstituer une version décodée de l'image courante.
- [0026] Avantageusement, on peut concevoir un système de transmission qui travaille image par image avec une faible latence, chaque image courante étant décodée dès réception des paramètres du réseau de neurones et des informations de prédiction et de correction associées.
- [0027] Avantageusement, le décodage permet de travailler avec des données de prédiction et des données résiduelles ayant le même format et la même résolution, ce qui facilite l'implémentation matérielle.
- [0028] Selon des modes de réalisation du procédé de codage et/ou de décodage :
- [0029] - Ledit au moins un réseau de neurones de reconstruction comporte au moins un module d'attention. Avantageusement, le réseau de neurones utilise des mécanismes d'attention qui permettent de générer des masques d'attention. Le codage est ainsi amélioré.
- [0030] Par "module d'attention" on entend une couche d'un réseau neuronal comprenant un élément d'attention. Un élément d'attention est un élément qui permet d'appliquer un mécanisme d'attention pour générer un ou des masques qui sont utilisés pour pondérer (par multiplication ou utilisation de fonctions plus complexes) les caractéristiques de l'image afin par exemple d'adapter l'allocation du débit aux différentes régions et/ou caractéristiques de l'image en fonction de leur importance. Les masques d'attention sont appliqués aux entrées du réseau de neurones (informations de correction et de prédiction). Ils sont codés et décodés en tant que paramètres du réseau de neurones. Avantageusement, il n'est pas nécessaire d'utiliser des bits supplémentaires pour coder les masques.
- [0031] - Ledit au moins un réseau de neurones de reconstruction comporte au moins une couche de convolution ; ainsi, un traitement non localisé de l'image peut être effectué, ce qui permet d'améliorer la cohérence de l'image générée en limitant le bruit et en améliorant les contours présents dans l'image.
- [0032] - Ledit au moins un réseau de neurones de reconstruction applique le module d'attention sur lesdites informations de prédiction. Avantageusement selon ce mode, le module d'attention est utilisé pour corriger le canal de prédiction. Ainsi, une meilleure prise en compte de certaines zones de l'image peut être effectuée. On peut effectuer une pondération des informations de prédiction. Selon un exemple, si le masque généré "efface" l'information de prédiction, on aboutit à un codage de l'image en mode intra-image (seules les informations de correction sont utilisées).
- [0033] - Ledit au moins un réseau de neurones de reconstruction applique le module

d'attention sur lesdites informations de correction. Avantageusement selon ce mode, le module d'attention est utilisé pour corriger le canal de correction. Ainsi, une meilleure prise en compte de certaines zones de l'image à corriger peut être effectuée. On peut effectuer une pondération des informations de correction. Selon un exemple, si le masque généré "efface" l'information de correction, on aboutit à un codage de l'image en mode inter-image (seules les informations de prédiction sont utilisées).

- [0034] - Ledit au moins un réseau de neurones de reconstruction applique le module d'attention sur une combinaison desdites informations de correction et de prédiction. Avantageusement selon ce mode, le module d'attention est utilisé pour corriger conjointement les deux canaux d'entrée. Ainsi, un compromis peut être trouvé concernant l'importance, ou pondération, donnée à chacun des deux canaux. Le module d'attention peut également corriger des artefacts liés aux effets de la combinaison des deux entrées, ou canaux, de prédiction et de correction.
- [0035] - Une partie au moins de l'ensemble de paramètres de reconstruction est obtenue à partir de données dudit flux binaire ; avantageusement selon ce mode, les paramètres du réseau de neurones sont transmis dans le flux. Le module de reconstruction peut être transmis tout ou partie sous une forme quantifiée et codée sous une forme compacte à l'aide de tout quantificateur et codeur entropique accessible à l'homme du métier. Un format analogue à celui de la norme MPEG-7 NNR peut être employé.
- [0036] - Une partie au moins de l'ensemble de paramètres de reconstruction est décodée en fonction de paramètres prédéterminés ; avantageusement selon ce mode, une partie des paramètres de reconstruction du réseau de neurones est accessible du décodeur, par exemple ils peuvent être mémorisés dans une mémoire de stockage. Notamment, une partie du module de reconstruction peut être stocké sous une forme quantifiée et codée sous une forme compacte à l'aide de tout quantificateur et codeur entropique accessible à l'homme du métier. Un format analogue à celui de la norme MPEG-7 NNR peut être employé.
- [0037] - Une partie au moins de l'ensemble de paramètres de reconstruction est décodée en fonction de paramètres de référence ; avantageusement selon ce mode, certains paramètres du réseau de neurones sont décodés en prenant en compte des paramètres de référence qui peuvent être stockés dans une mémoire (mémoire de stockage, réseau, etc.) accessible du décodeur. Ainsi l'espace de stockage ou le débit de transmission sur le réseau peuvent être efficacement réduits : certains paramètres peuvent être codés/décodés de manière complémentaire (par exemple par mise à jour d'un noyau, d'une couche, d'un biais, etc.), d'autre ne pas être codés/décodés du tout, car disponibles dans les paramètres de référence. Les paramètres de référence peuvent constituer un réseau de référence. Selon des variantes de réalisation, qui peuvent être combinées entre elles :

- les paramètres de référence peuvent constituer un ensemble de variantes d'un réseau de reconstruction : type de convolution, modules d'attention, etc. Ainsi, seuls ces paramètres ou leur identification nécessitent d'être codés, insérés dans le flux et décodés par le décodeur ;
- certains paramètres sont codés de manière complémentaire aux paramètres du réseau de référence. Ainsi il suffira au décodeur de décoder ces paramètres résiduels puis de les ajouter, les multiplier (ou les combiner de toute autre manière connue) à ceux du réseau de neurones de référence ;
- une partie des paramètres du réseau de référence est réutilisée pour le réseau de reconstruction. Par exemple, une couche complète du réseau peut être recopiée du réseau de référence, les paramètres de cette couche ne sont donc ni codés, ni transmis, ni décodés ;
- un indicateur de réseau de référence est transmis dans le flux. Il est ainsi possible d'indiquer un réseau à utiliser au décodage parmi une pluralité de réseaux possibles connus du décodeur.

[0038] - Ledit au moins un réseau de reconstruction est sélectionné parmi une pluralité de réseaux de reconstruction, et le flux binaire comporte un indicateur indiquant la sélection dudit réseau au moins de la pluralité.

[0039] - Ledit au moins un réseau de reconstruction correspond à une pluralité de réseaux de reconstruction en cascade. Avantagement, l'utilisation de plusieurs réseaux en cascade permet notamment de rehausser le filtrage et améliorer la cohérence de l'image générée en limitant son bruit et améliorant ses contours. En outre chaque réseau peut être appris individuellement ou figé indépendamment. Selon des variantes de réalisation, qui peuvent être combinées entre elles :

- le flux binaire comporte un indicateur indiquant l'utilisation d'un desdits réseaux au moins. Ainsi, en fonction de la qualité et de la complexité requises, un ou plusieurs réseaux peuvent être utilisés au décodage. L'indicateur indique par exemple que le réseau 1 est obligatoire, mais que les réseaux 2 et 3, qui se trouvent en cascade derrière le réseau 1, sont optionnels. Le décodeur peut alors faire le choix de les utiliser ou non : s'il les utilise, il obtiendra une meilleure qualité au détriment de la complexité, et vice-versa.
- Au moins deux des réseaux comportent des paramètres communs. Il est ainsi possible, avantagement, de transmettre des paramètres caractéristiques d'un réseau, qui peuvent être utilisés pour les autres, ce qui permet de réduire le coût de transmission. Le réseau est ainsi mutualisé au travers de paramètres communs.
- Au moins deux des réseaux ont une structure identique et seuls leurs poids diffèrent en partie ou totalement. Les mêmes éléments de calcul sont ainsi mu-

tualisés.

[0040] Corrélativement, l'invention vise aussi un dispositif de codage et un dispositif de décodage.

[0041] Les caractéristiques et avantages du procédé de codage ou décodage s'appliquent de la même façon au dispositif de codage ou décodage selon l'invention et vice versa.

[0042] L'invention vise également un programme d'ordinateur sur un support d'enregistrement, ce programme étant susceptible d'être mis en œuvre dans un ordinateur ou un dispositif de codage ou de décodage conforme à l'invention. Ce programme comporte des instructions adaptées à la mise en œuvre du procédé correspondant. Ce programme peut utiliser n'importe quel langage de programmation, et être sous la forme de code source, code objet, ou de code intermédiaire entre code source et code objet, tel que dans une forme partiellement compilée, ou dans n'importe quelle autre forme souhaitable.

[0043] L'invention vise aussi un support d'information ou un support d'enregistrement lisible par un ordinateur, et comportant des instructions de programme d'ordinateur mentionné ci-dessus. Les supports d'information ou d'enregistrement peuvent être n'importe quel entité ou dispositif capable de stocker les programmes. Par exemple, les supports peuvent comporter un moyen de stockage, tel qu'une ROM, par exemple un CD-ROM ou une ROM de circuit microélectronique, ou encore un moyen d'enregistrement magnétique, par exemple une disquette ou un disque dur, une séquence d'ADN, ou une mémoire flash. D'autre part, les supports d'information ou d'enregistrement peuvent être des supports transmissibles tels qu'un signal électrique ou optique, qui peut être acheminé via un câble électrique ou optique, par lien radio, par lien optique sans fil ou par d'autres moyens.

[0044] Le programme selon l'invention peut être en particulier téléchargé sur un réseau de type Internet.

[0045] Alternativement, chaque support d'information ou d'enregistrement peut être un circuit intégré dans lequel le programme est incorporé, le circuit étant adapté pour exécuter ou pour être utilisé dans l'exécution d'un procédé conforme à l'invention.

Brève description des dessins

[0046] D'autres caractéristiques et avantages de la présente invention ressortiront de la description faite ci-dessous, en référence aux dessins annexés qui en illustrent des exemples de réalisation dépourvus de tout caractère limitatif.

[0047] [Fig.1] la [Fig.1] représente schématiquement un dispositif de décodage utilisé dans le cadre de l'invention ;

[0048] [Fig.2] la [Fig.2] représente schématiquement un dispositif de codage utilisé dans le cadre de l'invention ;

- [0049] [Fig.3] La [Fig.3] illustre un exemple de module de mixage utilisé au codage et au décodage selon deux modes de mode de réalisation de l'invention
- [0050] [Fig.4] La [Fig.4] illustre un réseau de neurones de reconstruction utilisé au codage et au décodage selon un mode de réalisation de l'invention.
- [0051] [Fig.5] La [Fig.5] illustre une variante de réseau de neurones de reconstruction utilisé au codage et au décodage selon un mode de réalisation de l'invention.
- [0052] [Fig.6] La [Fig.6] illustre une autre variante de réseau de neurones de reconstruction utilisé au codage et au décodage selon un mode de réalisation de l'invention
- [0053] [Fig.7] la [Fig.7] est un logigramme représentant un exemple de procédé de décodage qui peut être mis en œuvre par le dispositif de décodage de la [Fig.1].

Description des modes de réalisation

- [0054] La [Fig.1] représente schématiquement un dispositif de décodage utilisé dans le cadre de l'invention.
- [0055] Le dispositif de décodage DEC_v de la [Fig.1] décode une image ou sous-séquence v de la séquence, en procédant image par image. A l'issue du décodage, les V images de la séquence sont décodées.
- [0056] Une sous-séquence courante à décoder notée Id_v comporte au moins une image courante à décoder, chaque image comportant respectivement une pluralité d'échantillons. Par la suite, on considère, sans perte de généralité, que la sous-séquence courante comporte une seule image courante I_v . Cependant une sous séquence courante à décoder peut comporter selon un mode de réalisation plusieurs images (qui peuvent être par exemple concaténées pour former une image de taille supérieure I_v).
- [0057] Le décodeur DEC_v dédié à l'image reçoit en entrée les données nécessaires pour la décoder :
- [0058] - des données encodées IP_c correspondant à des informations de prédiction. Ces informations de prédiction peuvent être par exemple :
- des informations de mouvement (champ de mouvement, numéros d'images de référence, etc.);
 - des informations de prédiction intra-image (mode de prédiction, directions de prédictions, etc.);
- [0059] - des informations de déformation du signal, comme des paramètres de transformations géométriques (translations, homothéties, etc.) ;
- [0060] - etc.
- [0061] - des données encodées IC_c correspondant à des informations de correction IC codées, par exemple des résidus de prédiction, ou d'une façon générale, des données visant à corriger la prédiction obtenue à partir des informations de prédiction.
- [0062] - les paramètres de reconstruction codés WMX_c d'au moins un réseau de neurones de

- reconstruction. D'autres réseaux de neurones peuvent être utilisés, comme des réseaux :
- [0063] - de synthèse des informations de prédiction ;
 - de synthèse des informations de correction ;
 - de décodage prédictif des cartes de caractéristiques latentes de prédiction ou de correction ;
 - [0064] - de suréchantillonnage et/ou de traitement de cartes de caractéristiques latentes de prédiction ou de correction ;
 - [0065] - de post-processing des données reconstruites après mixage ;
 - [0066] - etc.
 - [0067] Le module de décodage DEC_v comprend, pour une image courante, un module REF d'obtention d'images de référence, un module FPD de décodage d'information de prédiction, un module FCD de décodage d'information de correction, un module MIX de reconstruction par combinaison des images/cartes compensées IP/FMP et des images/cartes de correction IC/FMC, un module NND de décodage de réseau de neurones. Selon un mode de réalisation, DEC_v produit en sortie une image décodée courante, notée Idv , comportant une pluralité d'échantillons décodés.
 - [0068] Le module de décodage FCD décode des informations de correction ICc. Selon un mode de réalisation, il s'agit d'un décodeur classique, par exemple entropique, de type JPEG, ou MPEG, etc. qui produit en sortie une image de correction (de résidus) notée IC. Selon un autre mode de réalisation, il décode des données quantifiées, par exemple en utilisant un décodeur entropique, d'une ou plusieurs cartes de caractéristiques de correction dans le domaine latent. Selon un mode de réalisation, le module FCD peut comprendre à cette fin un réseau de neurones de synthèse d'informations de correction. Selon un mode de réalisation, les cartes décodées par le module FCD, au nombre de NFC, sont notées FMCet sont dites cartes de correction de l'image courante, ou cartes de correction courantes.
 - [0069] Le module de décodage FPD décode des informations de prédiction IPc. Selon un mode de réalisation, il s'agit d'un décodeur classique, par exemple entropique, de type JPEG, ou MPEG, etc. qui produit en sortie une image de prédiction notée IP. Selon un mode de réalisation, il s'agit d'un décodeur classique qui produit en sortie un champ de mouvement MV. Le module FPD peut comprendre à cette fin un réseau de neurones de synthèse de mouvement. Selon un autre mode de réalisation, il décode des données quantifiées, par exemple en utilisant un décodeur entropique, d'une ou plusieurs cartes de caractéristiques de prédiction. Les cartes décodées par le module FPD, au nombre de NFP, sont notées FMPet sont dites cartes de prédiction de l'image courante, ou cartes de prédiction courantes.
 - [0070] Le module REF facultatif constitue un ensemble de NFR images/cartes de référence, notées IREF/FMREF, correspondant à une ou plusieurs images/cartes précédemment

codées puis décodées de la séquence. Elles peuvent être prédéterminées ou leur référence peut être décodée dans le flux sous forme d'un indicateur. Dans un mode de réalisation, une seule image de référence est sélectionnée. Dans un mode de réalisation, plusieurs images de référence sont sélectionnées. Dans un mode de réalisation, un ensemble de cartes de référence est sélectionné dans le domaine latent. Dans un mode de réalisation, par exemple dans le cas où l'on décode une image intra, aucune carte de référence n'est sélectionnée et le module REF est absent.

- [0071] Le module WARP facultatif effectue une compensation de mouvement des images/cartes de référence à partir du champ de mouvement MV afin d'obtenir une ou plusieurs image(s) ou cartes de prédiction, IP/FMP. Le module WARP peut être absent si le décodeur FPD décode directement les image(s) ou cartes de prédiction. Dans un mode de réalisation, le module WARP produit les cartes FMP avec les données seulement issues du module FPD, c'est-à-dire que le module REF facultatif n'existe pas, ou que le module WARP ne prend pas en compte les images de référence qui lui sont soumises.
- [0072] Le module MIX utilise en entrée une ou plusieurs image(s) ou cartes de prédiction IP/FMP et une ou plusieurs image(s) ou cartes de correction IC/FMC et effectue une combinaison des deux informations pour obtenir l'image reconstruite. Le module MIX sera détaillé dans la suite de la description. Il comprend au moins un réseau de neurones CNN de type convolutif, c'est-à-dire qu'il comporte au moins une couche de convolution. Une couche de convolution peut inclure des éléments convolutifs, comportant une structure résiduelle ou non et un module d'attention ou non, comme il sera décrit par la suite.
- [0073] Les informations codées ICc, IPc et WMXc sont extraites à partir du flux BS qui peut être reçu sur un réseau de communication, ou stocké dans une mémoire accessible du décodeur.
- [0074] Le dispositif de décodage DEC peut être mis en œuvre au moyen d'un dispositif électronique comprenant un processeur et une mémoire, non représentés ; chacun des modules mentionnés ci-dessus peut alors être réalisé par la coopération du processeur et d'instructions de programme d'ordinateur mémorisées dans la mémoire susmentionnée et conçues pour effectuer les fonctionnalités du module concerné, notamment comme décrit ci-dessous, lorsque ces instructions sont exécutées par le processeur.
- [0075] La [Fig.2] représente schématiquement un dispositif de codage utilisé dans le cadre de l'invention.
- [0076] Le dispositif de codage ENCv de la [Fig.2] code une image ou sous-séquence courante Iv de la séquence SV, en procédant image par image. A l'issue du codage, les V images de la séquence sont codées.

- [0077] Le codeur ENC_v reçoit en entrée une séquence d'images comprenant l'image I_v et produit en sortie des paramètres codés. Ces paramètres comprennent :
- [0078] – des données encodées IP_c correspondant à des informations de prédiction.
- [0079] – des données encodées IC_c correspondant à des informations de correction, par exemple des résidus de prédiction.
- [0080] - les paramètres codés d'au moins un réseau de neurones de reconstruction. Comme mentionné lors de la description de la [Fig.1], d'autres réseaux de neurones peuvent être utilisés.
- [0081] Le module de codage ENC_v comprend, pour une image courante, un module NNC de codage de paramètres de réseau(x) de neurones, un module FCC de codage des informations de correction, un module FPC de codage des informations de prédiction, un module d'évaluation RD-OPT, un module d'initialisation et de mise à jour INIT/MAJ.
- [0082] Par ailleurs, de manière classique, le module de codage comprend un décodeur similaire au décodeur DEC_v qui a été précédemment décrit, afin de reconstruire l'image courante I'_v codée puis décodée.
- [0083] Durant le processus d'entraînement, ou de construction, du codage, c'est-à-dire tant que l'étape d'évaluation d'une performance n'est pas satisfaisante, les modules de codage réalisent une simulation de codage, suivie d'un décodage, à destination du module d'évaluation RD-OPT. Par la suite, ils effectuent le codage effectif des données. De manière connue, la simulation de codage peut être identique au codage effectif, ou en réaliser une approximation.
- [0084] Le module INIT/MAJ est chargé de l'initialisation et de la mise à jour des valeurs des images/cartes de caractéristiques et des paramètres du ou des réseaux de neurones. Il réalise une mise à jour des valeurs à encoder de l'image courante, en fonction des résultats de la fonction de performance. Une fois que les informations de prédiction et de correction et le(s) réseau(x) de neurones sont stabilisés, ils peuvent être codés.
- [0085] Le module RD-OPT réalise une évaluation et minimisation d'une performance de codage. La fonction d'évaluation est par exemple de type débit-distorsion. La distorsion peut être évaluée entre I_v et I'_v, le débit peut comprendre le cumul des débits lié aux contributions de IC_c, WMX_c et IP_c, ou seulement une sous-partie. La minimisation peut être effectuée par une descente de gradient, ou tout autre procédé d'optimisation à la portée de l'homme du métier.
- [0086] Le dispositif de codage ENC peut être mis en œuvre au moyen d'un dispositif électronique comprenant un processeur et une mémoire, non représentés ; chacun des modules mentionnés ci-dessus peut alors être réalisé par la coopération du processeur et d'instructions de programme d'ordinateur mémorisées dans la mémoire sus-mentionnée et conçues pour effectuer les fonctionnalités du module concerné, notamment comme décrit ci-dessous, lorsque ces instructions sont exécutées par le

processeur.

- [0087] La [Fig.3] illustre un exemple de module de mixage utilisé au codage et au décodage selon deux modes de réalisation de l'invention (a) et (b).
- [0088] Le module MIX comporte un réseau de neurones de reconstruction CNN qui sera décrit ultérieurement à l'appui des figures 4, 5 et 6.
- [0089] Sur la gauche de la figure (a), selon un premier mode de réalisation, le module MIX reçoit en entrée une image IC de correction et une image IP de prédiction. Il génère en sortie une image reconstruite Ivd.
- [0090] Sur la droite de la figure (b), selon un second mode de réalisation, le module MIX reçoit en entrée un ensemble de cartes de caractéristiques de correction, FMC, et un ensemble de cartes de caractéristiques de prédiction FMP.
- [0091] Les cartes de caractéristiques FMC et FMD ne sont pas nécessairement en même nombre ni sous la même forme. Le module MIX est adapté pour traiter ces entrées sous leur format respectif.
- [0092] Selon un mode de réalisation, le module MIX génère en sortie une image reconstruite Ivd.
- [0093] Selon un mode de réalisation, le module MIX génère en sortie un ensemble de cartes de caractéristiques FMREC représentatives (dans le domaine latent) d'une image reconstruite Ivd. Dans ce cas, ces cartes peuvent être traitées par un réseau de neurones dit de combinaison, comprenant par exemple un réseau de type MLP (multi-layer perceptron) ou convolutif, pour produire l'image reconstruite Ivd.
- [0094] La [Fig.4] illustre un réseau de neurones de reconstruction (CNN) utilisé par le module de mixage MIX au codage et au décodage selon un mode de réalisation de l'invention.
- [0095] Le module MIX peut comporter un ou plusieurs réseaux de reconstruction, en cascade. Les réseaux peuvent être identiques dans leur structure et leurs paramètres, ou différents.
- [0096] Selon l'exemple de la [Fig.4], le réseau de reconstruction comporte une couche de mixage CC (Cross Channel) des deux ensembles d'entrée, images/cartes de correction IC/FMC sur la première entrée et images/cartes de prédiction IP/FMP sur la seconde entrée. La couche de mixage peut réaliser une simple concaténation des entrées ou une combinaison plus complexe incluant par exemple une convolution ou une combinaison par réseau de neurones sur les entrées concaténées. Le réseau de reconstruction comporte ensuite un ensemble de couches, ou sous-réseaux, N1, N2..NN.
- [0097] Les modules Ni (N1, N2..NN) comprennent des éléments de convolution, et des éléments de non-linéarité (RELU, etc.) sans aspect limitatif.
- [0098] Un élément de convolution peut être de type résiduel ou non, comporter des noyaux de n'importe quelle taille. Il est potentiellement suivi d'un biais comme cela est bien

connu de l'homme du métier. Toute fonction d'activation à la portée de l'homme du métier peut être utilisée (ReLU, sigmoïde, Softmax, Softplus, Softsign, etc.). Les éléments de convolution peuvent être composés de convolutions 2D, c'est-à-dire que le noyau de convolution s'applique conjointement sur les lignes et les colonnes de l'image ou de la carte de valeurs latentes. Les éléments de convolution peuvent être composés de convolutions 1D, sous une forme séparable. Dans ce cas, de manière connue, plutôt qu'effectuer la convolution sur l'image ou la carte de valeurs latentes bidimensionnelle, on opère un premier opérateur de convolution unidimensionnel (1D) sur les lignes, puis une seconde convolution sur les colonnes de la représentation obtenue après le premier opérateur. Le traitement ligne colonne peut être inversé pour effectuer un traitement par colonne puis par ligne.

- [0099] Selon un premier mode de réalisation, non représenté, le module N2 comporte, de même que les autres modules N_i , une ou plusieurs couches de convolution suivies de non-linéarités.
- [0100] La partie haute de la figure détaille un autre mode de réalisation, préféré, du module N2. Il reçoit en entrée un signal z intermédiaire représentatif d'une composition, ou mixage, par le module CC, des deux entrées, suivi éventuellement d'un premier réseau (N1), par exemple de convolution.
- [0101] Selon ce mode de réalisation, le module N2 comporte un élément d'attention CL. Le mécanisme d'attention permet d'appliquer un traitement différencié sur certaines zones de l'image à reconstruire, grâce à un masquage dans le domaine image ou latent. De telles opérations sont utilisées dans de nombreux systèmes de compression basés apprentissage. Par exemple, le document "Non-local Attention Optimized Deep Image Compression" de Liu et al. (disponible à l'adresse <https://arxiv.org/pdf/1904.09757.pdf>) décrit de tels mécanismes d'attention utiles dans le contexte du codage entropique. Notamment, l'introduction d'un masque d'attention permet de porter une importance particulière sur certaines zones, dans le contexte de cette publication il s'agit d'adapter l'allocation de débit aux différentes valeurs des cartes de caractéristiques (par exemple, en allouant plus de débit aux zones texturées de l'image).
- [0102] La branche supérieure de N2 reçoit en entrée le signal z issu du réseau N1 ou d'une combinaison des deux entrées. Le réseau (N2) apprend (au codeur) à construire un masque M qui est obtenu par traitement optionnel du signal d'entrée par un réseau neuronal (NN3) comprenant un ensemble de convolutions suivies de non-linéarités, puis traitement des valeurs par le module CL qui peut, de manière connue, être de type Clamp, sigmoïde, Arctg, etc.
- [0103] La branche centrale de N2 correspond au traitement principal du signal z par un réseau de neurones (NN4). Ce module est facultatif. La sortie du module NN4 est

multipliée par le masque M. Selon un mode de réalisation, seules certaines zones sont conservées à l'issue de la multiplication par les valeurs du masque M. Selon un mode de réalisation, les entrées sont effacées à l'issue de la multiplication par les valeurs du masque M

- [0104] Selon un mode de réalisation, M est un masque d'attention qui s'applique sur une image issue du réseau optionnel NN4 ou du signal z.
- [0105] Selon un mode de réalisation, M est un masque d'attention qui s'applique sur l'ensemble des latentes issues du réseau optionnel NN4 ou des latentes z.
- [0106] Selon un mode de réalisation, M est un masque d'attention qui comprend autant de latentes que la sortie du réseau optionnel NN4 ou le signal z.
- [0107] Selon un mode de réalisation, M est un masque d'attention dont les valeurs réelles se situent entre 0 et 1.
- [0108] Selon un autre mode réalisation, M contient uniquement les valeurs 0 ou 1.
- [0109] Une connexion résiduelle (optionnelle) est présente sur la branche basse, pour des raisons bien connues de stabilité des réseaux de neurones N3, N4.
- [0110] On réalise ainsi une combinaison des deux canaux (de prédiction et de correction) plus riche grâce au mécanisme de masquage. Cette combinaison est représentée par la sortie y du module N2.
- [0111] Le réseau NN est un réseau de neurones optionnel comprenant un réseau convolutif qui adapte la sortie du réseau précédent au format de l'image Ivd ou des cartes de caractéristiques latentes FMREC.
- [0112] Le réseau CNN de la [Fig.4] peut lui-même comporter une connexion résiduelle, présentée dans la partie basse de la figure, pour des raisons connues de stabilité.
- [0113] Le traitement des signaux IP/FMP et IC/FMC par le module CNN produit en sortie une image ou des cartes de caractéristique Ivd/FMREC.
- [0114] La [Fig.5] illustre une variante de réseau de neurones de reconstruction utilisé au codage et au décodage selon un mode de réalisation de l'invention.
- [0115] Le module N2 décrit précédemment est appliqué uniquement au signal de prédiction IP/FMP. En d'autres termes, seul le signal de prédiction est masqué par le module d'attention.
- [0116] Selon l'exemple de la [Fig.5], le canal d'entrée des images/cartes de prédiction IP/FMP alimente un module N2, qui a déjà été décrit précédemment à l'appui de la [Fig.4], d'entrée z et de sortie y, qui peuvent être des images, ou des cartes de caractéristiques de valeurs latentes. Le module N2 comporte au moins un élément d'attention CL, comme décrit précédemment. Dans le cas de la [Fig.5], la branche supérieure de N2 reçoit donc en entrée le signal z représentatif de l'entrée IP/FMP seule. Le réseau (N2) apprend (au codeur) à construire un masque M à partir de cette entrée. On réalise ainsi un traitement du canal IP/FMP plus riche grâce au mécanisme de masquage.

- [0117] Le réseau N2 produit une sortie y qui est le premier canal venant alimenter une couche de mixage CC (Cross Channel) possédant deux canaux d'entrée, le second canal étant alimenté par les images/cartes de correction IC/FMC. La couche de mixage peut réaliser une simple concaténation des entrées ou une combinaison plus complexe.
- [0118] Le réseau de reconstruction comporte ensuite un ensemble de couches, ou sous-réseaux, N1..NN. Les modules N1..NN comprennent des éléments de convolution, et des éléments de non-linéarité (RELU, etc.) sans aspect limitatif, déjà décrits précédemment.
- [0119] La [Fig.6] illustre une autre variante de réseau de neurones de reconstruction utilisé au codage et au décodage selon un mode de réalisation de l'invention.
- [0120] Le module N2 décrit précédemment est appliqué uniquement au signal de correction IC/FMC. En d'autres termes, seul le signal de correction est masqué par le module d'attention.
- [0121] Selon l'exemple de la [Fig.6], le canal d'entrée des images/cartes de prédiction IC/FMC alimente un module N2 qui a déjà été décrit précédemment à l'appui de la [Fig.4]. Il s'agit donc d'un mode d'implémentation similaire à celui de la [Fig.5], les rôles des signaux IP/FMP et IC/FMC étant inversés.
- [0122] On réalise ainsi un traitement du canal IC/MFC plus riche grâce au mécanisme de masquage.
- [0123] Le réseau N2 produit une sortie y qui est le premier canal venant alimenter une couche de mixage CC (Cross Channel) possédant deux canaux d'entrée, le second canal étant alimenté par les images/cartes de correction IP/FMP. La couche de mixage peut réaliser une simple concaténation des entrées ou une combinaison plus complexe.
- [0124] Le réseau de reconstruction comporte ensuite un ensemble de couches, ou sous-réseaux, N1..NN. Les modules N1..NN comprennent des éléments de convolution, et des éléments de non-linéarité (RELU, etc.) sans aspect limitatif, déjà décrits précédemment.
- [0125] La [Fig.7] est un logigramme représentant un exemple de procédé de décodage qui peut être mis en œuvre par le dispositif de décodage de la [Fig.1].
- [0126] Le décodage décrit concerne une sous-séquence d'images comportant au moins une image courante de la séquence à décoder.
- [0127] Lors d'une étape E30, le flux de données codées est obtenu. Il peut être reçu d'un réseau de communication, ou lu sur un support de stockage. En outre certaines informations (par exemple, un réseau de reconstruction de référence) peuvent être lues dans des mémoires accessibles du décodeur. Les données obtenues sont les informations de correction ICc, les données de prédiction IPc, et les paramètres WMXc d'un réseau de reconstruction.
- [0128] Lors d'une étape E31, les informations de correction ICc sont décodées.

- [0129] Selon un mode de réalisation, elles sont décodées par une technique classique à la portée de l'homme du métier, par exemple une prédiction suivie d'un décodage par un codeur entropique contextuel puis une déquantification, ou encore elles sont issues d'un décodeur standard (JPEG, MPEG, etc.)
- [0130] Selon un mode de réalisation, une image IC est décodée.
- [0131] Selon un mode de réalisation, une ou plusieurs cartes de caractéristiques de correction (latentes) FMC sont décodées.
- [0132] Lors d'une étape E32, les informations de prédiction IPc sont décodées.
- [0133] Selon un mode de réalisation, elles sont décodées par une technique classique à la portée de l'homme du métier, par exemple une prédiction suivie d'un décodage par un codeur entropique contextuel puis une déquantification, ou encore elles sont issues d'un décodeur standard (JPEG, MPEG, etc.)
- [0134] Selon un mode de réalisation, une image IP est décodée.
- [0135] Selon un mode de réalisation, une ou plusieurs cartes de caractéristiques de prédiction (latentes) FMP sont décodées.
- [0136] Selon un mode de réalisation, un champ de mouvement MV est décodé. Ce champ de mouvement peut être par exemple dense (une information de mouvement par pixel). Il peut être décodé par un réseau de neurones, à partir de cartes de caractéristiques décodées représentatives du mouvement de l'image courante ou estimé sur la base des plusieurs images/cartes de référence IREF/FMREF. De manière classique, ce champ de mouvement est utilisé pour compenser (warper) une ou plusieurs images/cartes de référence IREF/FMREF lors d'une étape E33 afin d'obtenir les images/cartes de prédiction IP/FMP compensées en mouvement.
- [0137] Lors d'une étape E34, les paramètres de reconstruction WMX d'un réseau de neurones de reconstruction du module MIX sont générés par décodage des valeurs WMXc. On peut utiliser à cette fin toute technique connue de décodage de réseau de neurones correspondant à celle de codage qui a été utilisée par le codeur, par exemple la norme de codage de réseaux de neurones proposée par le standard MPEG-7 partie 17, NNR.
- [0138] Selon un mode de réalisation, les paramètres de reconstruction sont reçus dans le flux binaire par le décodeur. Selon un mode de réalisation, certains paramètres du réseau de neurones de reconstruction sont accédés dans une mémoire accessible du décodeur. Selon un mode de réalisation, certains paramètres de référence du réseau de neurones de reconstruction sont accédés dans une mémoire accessible du décodeur.
- [0139] Selon un mode de réalisation, un indicateur reçu dans le flux permet de sélectionner un ou plusieurs réseaux de reconstruction parmi une pluralité de réseaux accessibles du décodeur.
- [0140] Lors d'une étape E35, une combinaison des images/cartes de correction IC/FMC et

des images/cartes de caractéristiques de prédiction IP/FMP est effectuée par le module MIX.

- [0141] Selon un mode de réalisation, le module MIX utilise un réseau de neurones de reconstruction CNN qui prend en entrée les images/cartes IC/FMC et IP/FMP et produit, par exemple à l'aide d'une série de couches convolutives et de modules d'attention, comme décrit auparavant, l'image I_{dv}.
- [0142] Selon un autre mode de réalisation, le module MIX utilise plusieurs réseaux de reconstruction en cascade. Ils peuvent être identiques ou différents.
- [0143] Selon un mode de réalisation, le module MIX produit en sortie une image reconstruite I_{dv}.
- [0144] Selon un autre mode de réalisation, le module MIX produit en sortie une ou plusieurs cartes de caractéristiques (latentes) FMREC. Les cartes de caractéristiques sont alors fournies en entrée d'un autre réseau de neurones, dit de combinaison, par exemple un réseau de type MLP, pour produire l'image finale I_{dv}.

Revendications

- [Revendication 1] Procédé de codage d'une image courante (Iv) d'une séquence d'images, comportant les étapes suivantes :
- génération d'un ensemble d'informations de prédiction (IP, FMP) de ladite au moins une image courante ;
 - génération d'un ensemble d'informations de correction (IC, FMC) de ladite au moins une image courante ;
 - génération d'au moins un réseau de neurones de reconstruction (MIX), défini par un ensemble de paramètres de reconstruction (WMX) ;
 - traitement desdites informations de correction et desdites informations de prédiction par ledit au moins un réseau de neurones de reconstruction pour produire une représentation de l'image courante (I'v);
 - mise à jour d'une partie au moins desdites informations de prédiction et/ou desdites informations de correction et/ou desdits paramètres de reconstruction, en fonction d'une mesure de performance de codage ;
 - codage d'une partie au moins desdites informations de correction et/ou de prédiction et desdits paramètres de reconstruction.
- [Revendication 2] Procédé de décodage d'une image courante (I_{dv}) d'une séquence d'images (SV), comportant les étapes suivantes :
- obtention (E31) d'un ensemble d'informations de correction (IC, FMC) de ladite image courante ;
 - obtention (E32) d'un ensemble d'informations de prédiction (IP, FMP) de ladite image courante ;
 - décodage (E34) d'un ensemble de paramètres de reconstruction (WMX) représentatifs d'au moins un réseau de neurones de reconstruction (MIX , MLP);
 - traitement (E35) desdites informations de correction et desdites informations de prédiction par ledit au moins un réseau de neurones de reconstruction pour produire une représentation décodée de ladite image courante (FMREC, I_{vd}).
- [Revendication 3] Procédé de décodage ou de codage selon l'une des revendications précédentes, caractérisé en ce que ledit au moins un réseau de neurones de reconstruction comporte au moins un module d'attention (N2, CL, M).
- [Revendication 4] Procédé de décodage ou de codage selon l'une des revendications précédentes, caractérisé en ce que ledit au moins un réseau de neurones de reconstruction comporte au moins une couche de convolution (N1...NN).

- [Revendication 5] Procédé de décodage ou de codage selon l'une des revendications précédentes, caractérisé en ce que ledit au moins un réseau de neurones de reconstruction applique le module d'attention (CL) sur lesdites informations de prédiction (IP, FMP).
- [Revendication 6] Procédé de décodage ou de codage selon l'une des revendications précédentes, caractérisé en ce que ledit au moins un réseau de neurones de reconstruction applique le module d'attention (CL) sur lesdites informations de correction (IC, FMC).
- [Revendication 7] Procédé de décodage ou de codage selon l'une des revendications précédentes, caractérisé en ce que ledit au moins un réseau de neurones de reconstruction applique le module d'attention (CL) sur une combinaison desdites informations de correction et de prédiction (IP, FMP).
- [Revendication 8] Procédé de décodage selon l'une des revendications 2 à 7, caractérisé en ce que une partie au moins de l'ensemble de paramètres de reconstruction (WMX) est obtenue à partir de données dudit flux binaire.
- [Revendication 9] Procédé de décodage selon l'une des revendications 2 à 8, caractérisé en ce que une partie au moins de l'ensemble de paramètres de reconstruction (WMX) est décodée en fonction de paramètres prédéterminés.
- [Revendication 10] Procédé de décodage selon l'une des revendications 2 à 9, caractérisé en ce que une partie au moins de l'ensemble de paramètres de reconstruction (WMX) est décodée en fonction de paramètres de référence.
- [Revendication 11] Procédé de décodage selon l'une des revendications 2 à 10, caractérisé en ce que ledit au moins un réseau de reconstruction est sélectionné parmi une pluralité de réseaux de reconstruction, et le flux binaire comporte un indicateur indiquant la sélection dudit réseau au moins de la pluralité.
- [Revendication 12] Procédé de décodage selon l'une des revendications 2 à 11, caractérisé en ce que ledit au moins un réseau de reconstruction correspond à une pluralité de réseaux de reconstruction en cascade.
- [Revendication 13] Dispositif de codage d'une image courante (I_v) d'une séquence d'images (S_v), ledit dispositif étant configuré pour mettre en œuvre :
- génération (INIT/MAJ) d'un ensemble d'informations de prédiction (IP, FMP) de ladite image courante ;
 - génération (INIT/MAJ) d'un ensemble d'informations de correction (IC, FMC) de ladite image courante ;
 - génération (INIT/MAJ) d'au moins un réseau de neurones de reconstruction (MIX), défini par un ensemble de paramètres (WMX) ;
 - traitement (MIX, CNN) desdites informations de correction et desdites

informations de prédiction par ledit au moins un réseau de neurones de reconstruction pour produire une représentation de l'image courante (I_v);

- mise à jour (INIT/MAJ) d'une partie au moins desdites informations de prédiction et/ou desdites informations de correction et/ou desdits paramètres de reconstruction, en fonction d'une mesure de performance de codage ;

- codage (FPC, FCP, NNC) d'une partie au moins desdites informations de correction et/ou de prédiction et desdits paramètres de reconstruction.

.

[Revendication 14]

Dispositif de décodage d'une image courante (I_v) d'une séquence d'images (S_V), ledit dispositif étant configuré pour mettre en œuvre :

- obtention (FCD) d'un ensemble d'informations de correction (IC, FMC) de ladite image courante ;

- obtention (FPD) d'un ensemble d'informations de prédiction (IP, FMP) de ladite image courante ;

- décodage (NND) d'un ensemble de paramètres de reconstruction (WMX) représentatifs d'au moins un réseau de neurones de reconstruction (MIX);

- traitement (MIX, CNN) desdites informations de correction et desdites informations de prédiction par ledit au moins un réseau de neurones de reconstruction pour produire une représentation décodée de ladite image courante (FMREC, I_{vd}).

[Revendication 15]

Programme d'ordinateur comportant des instructions pour l'exécution des étapes d'un procédé de codage ou de décodage selon la revendication 1 ou 2 lorsque ledit programme est exécuté par un ordinateur.

[Fig. 1]

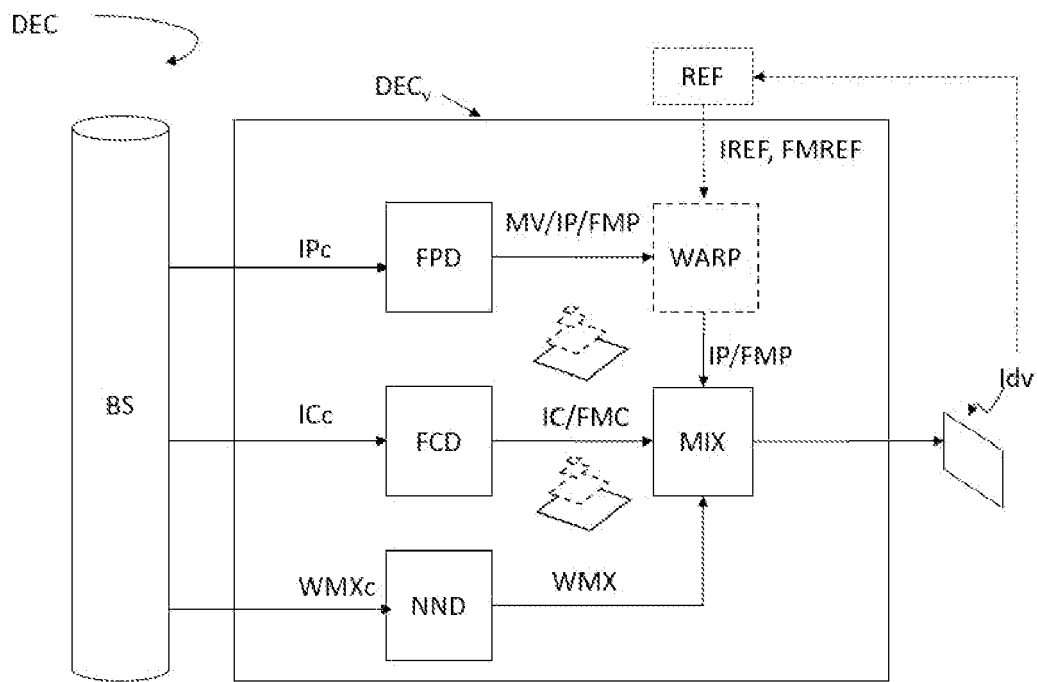


Figure 1

[Fig. 2]

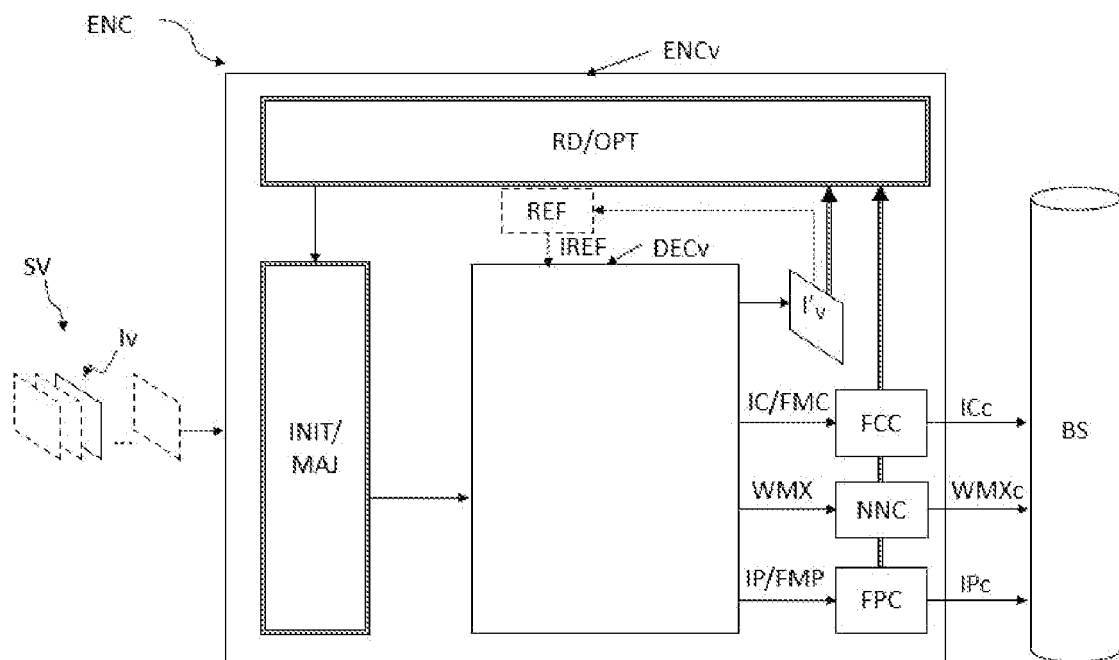


Figure 2

[Fig. 3]

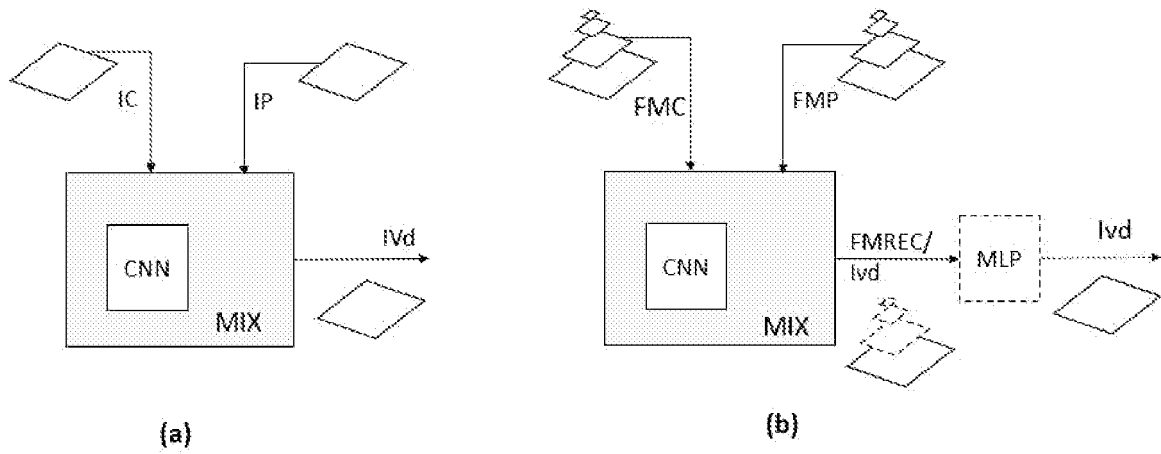


Figure 3

[Fig. 4]

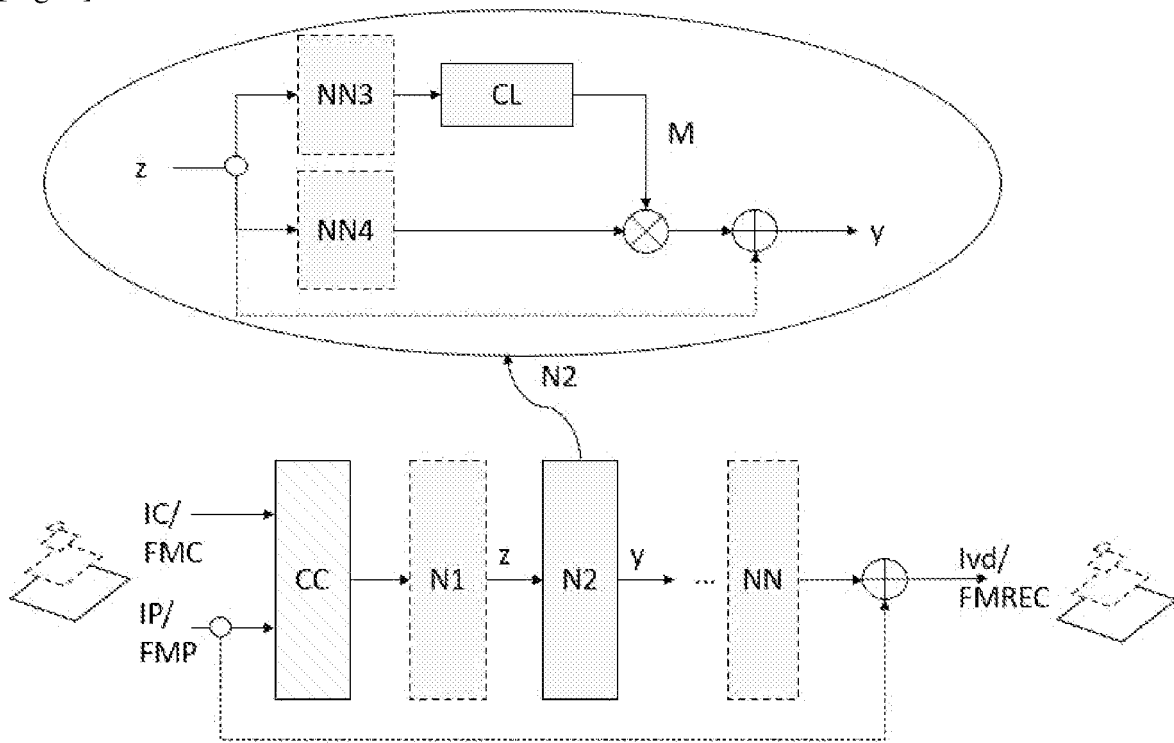


Figure 4

[Fig. 5]

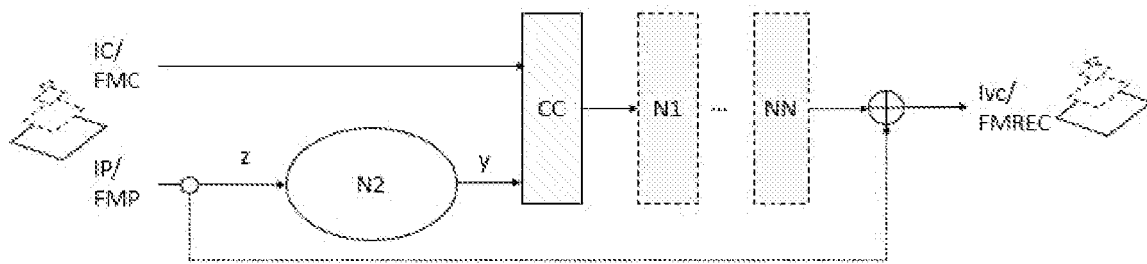


Figure 5

[Fig. 6]

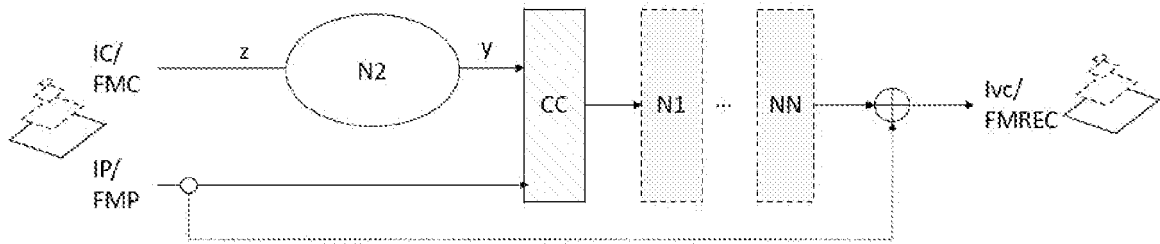


Figure 6

[Fig. 7]

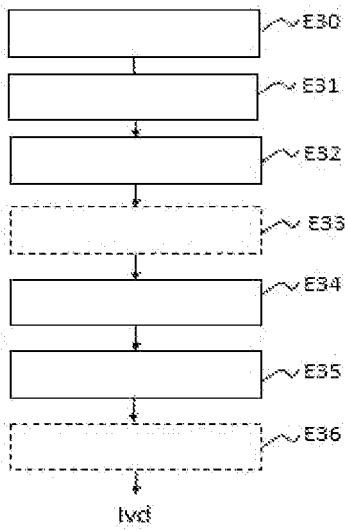


Figure 7

**RAPPORT DE RECHERCHE
PRÉLIMINAIRE**

N° d'enregistrement
national

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

FA 922444
FR 2309711

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X,D	<p>Brand Fabian ET AL: "Conditional Residual Coding: A Remedy for Bottleneck Problems in Conditional Inter Frame Coding", arXiv.org, 24 juillet 2023 (2023-07-24), pages 1-12, XP093139119, Ithaca DOI: 10.48550/arxiv.2307.12864 Extrait de l'Internet: URL:https://arxiv.org/pdf/2307.12864v1.pdf [extrait le 2024-03-07] * le document en entier *</p> <p style="text-align: center;">-----</p>	1-15	<p>G06N 3/02 G06T 9/00 H04N 19/90</p> <hr/> <p>DOMAINES TECHNIQUES RECHERCHÉS (IPC)</p> <p>H04N G06N</p>
A	<p>WO 2017/178782 A1 (MAGIC PONY TECH LTD [GB]) 19 octobre 2017 (2017-10-19) * le document en entier *</p> <p style="text-align: center;">-----</p>	1-15	
A	<p>Ladune Théo ET AL: "Conditional Coding for Flexible Learned Video Compression", arXiv (Cornell University), 3 mai 2021 (2021-05-03), pages 1-18, XP093139126, Ithaca DOI: 10.48550/arxiv.2104.07930 Extrait de l'Internet: URL:https://hal.science/hal-03192548/document [extrait le 2024-03-07] * le document en entier *</p> <p style="text-align: center;">-----</p> <p style="text-align: center;">-/--</p>	1-15	
Date d'achèvement de la recherche		Examineur	
8 mars 2024		Lombardi, Giancarlo	
<p>CATÉGORIE DES DOCUMENTS CITÉS</p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant</p>			

1
EPO FORM 1503 12.99 (P04C14)



**RAPPORT DE RECHERCHE
PRÉLIMINAIRE**

N° d'enregistrement
national

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

FA 922444
FR 2309711

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
A	<p>BAO WENBO ET AL: "MEMC-Net: Motion Estimation and Motion Compensation Driven Neural Network for Video Interpolation and Enhancement", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 43, no. 3, 1 mars 2021 (2021-03-01), pages 933-948, XP055838113, USA ISSN: 0162-8828, DOI: 10.1109/TPAMI.2019.2941941 Extrait de l'Internet: URL:https://arxiv.org/pdf/1810.08768.pdf> * le document en entier *</p> <p style="text-align: center;">-----</p>	1-15	
A,D	<p>Liu Haojie ET AL: "Non-local Attention Optimized Deep Image Compression", arXiv.org, 22 avril 2019 (2019-04-22), pages 1-10, XP093139111, Ithaca DOI: 10.48550/arxiv.1904.09757 Extrait de l'Internet: URL:https://arxiv.org/pdf/1904.09757.pdf [extrait le 2024-03-07] * le document en entier *</p> <p style="text-align: center;">-----</p>	1-15	<p>DOMAINES TECHNIQUES RECHERCHÉS (IPC)</p>
Date d'achèvement de la recherche		Examineur	
8 mars 2024		Lombardi, Giancarlo	
<p>CATÉGORIE DES DOCUMENTS CITÉS</p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant</p>			

1
EPO FORM 1503 12.99 (P04C14)

**ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 2309711 FA 922444**

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.
Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **08-03-2024**
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 2017178782 A1	19-10-2017	EP 3298783 A1	28-03-2018
		ES 2853700 T3	17-09-2021
		US 2018124422 A1	03-05-2018
		WO 2017178782 A1	19-10-2017
