

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

(19) Organisation Mondiale de la Propriété Intellectuelle
Bureau international



(10) Numéro de publication internationale
WO 2024/260628 A1

(43) Date de la publication internationale
26 décembre 2024 (26.12.2024)

(51) Classification internationale des brevets :
G06T 9/00 (2006.01) H04N 19/50 (2014.01)
G06N 3/02 (2006.01) H04N 19/82 (2014.01)

(21) Numéro de la demande internationale :
PCT/EP2024/062652

(22) Date de dépôt international :
07 mai 2024 (07.05.2024)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
FR2306455 22 juin 2023 (22.06.2023) FR

(71) Déposant : ORANGE [FR/FR] ; 111, quai du Président Roosevelt, 92130 Issy-les-Moulineaux (FR).

(72) Inventeurs : LADUNE, Theo ; ORANGE - Intellectual Property and Licensing, Orange Gardens - 44 Avenue de la République - CS 50010, 92326 Chatillon (FR). PHILIPPE, Pierrick ; ORANGE - Intellectual Property and Licensing, Orange Gardens - 44 Avenue de la République - CS 50010, 92326 Chatillon (FR). CLARE, Gordon ; ORANGE - Intellectual Property and Licensing, Orange Gardens - 44 Avenue de la République - CS 50010, 92326 Chatillon (FR). HENRY, Félix ; ORANGE - Intellectual Property and Licensing, Orange Gardens - 44 Avenue de la République - CS 50010, 92326 Chatillon (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

(54) Title: METHOD AND DEVICE FOR ENCODING AND DECODING IMAGE SEQUENCES

(54) Titre : PROCÉDÉ ET DISPOSITIF DE CODAGE ET DÉCODAGE DE SÉQUENCES D'IMAGES

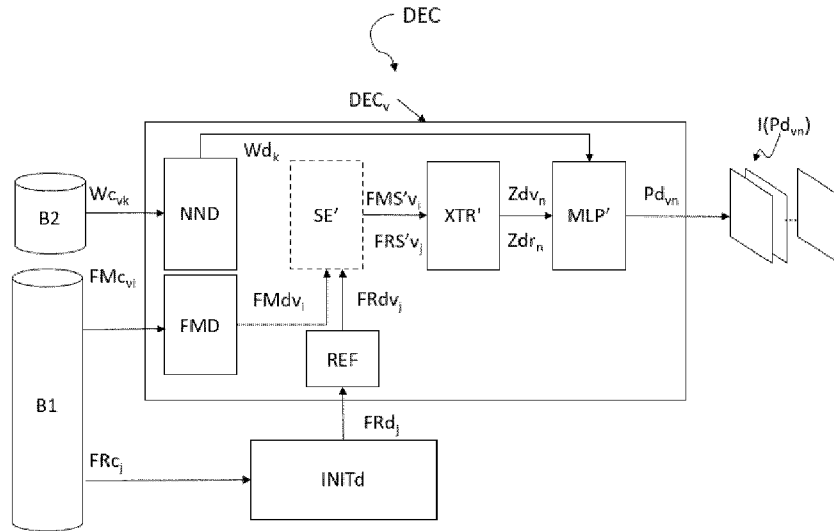


Figure 2

(57) Abstract: The invention relates to a method and device for encoding and decoding an image sequence. The decoding method comprises the following steps: - decoding (E35) a set of parameters representing an artificial neural network; - decoding (E31) a set of current feature maps including at least one current feature map representing a current image; - obtaining (E32) a set of reference feature maps, including at least one decoded map; and - for at least one current sample of the at least one current image, associated with a position in the image: - constructing (E34) a feature vector from the reference feature map and from the current feature map, according to said position of the current sample; and - processing (E37) of the vector by the artificial neural network in order to provide a vector representing a decoded value of the current sample.



WO 2024/260628 A1

HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) États désignés (*sauf indication contraire, pour tout titre de protection régionale disponible*) : ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, RU, TJ, TM), européen (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Publiée:

- avec rapport de recherche internationale (Art. 21(3))
 - en noir et blanc ; la demande internationale telle que déposée était en couleur ou en échelle de gris et est disponible sur PATENTSCOPE pour téléchargement.
-

(57) Abrégé : L'invention concerne un procédé et dispositif de codage et décodage d'une séquence d'images. Le procédé de décodage comporte les étapes suivantes : - décodage (E35) d'un ensemble de paramètres représentatifs d'un réseau de neurones de synthèse; - décodage (E31) d'un ensemble de cartes de caractéristiques courantes comprenant au moins une carte de caractéristiques courante représentative d'une image courante; - obtention (E32) d'un ensemble de cartes de caractéristiques de référence, comprenant au moins une carte décodée; - pour au moins un échantillon courant de ladite au moins une image courante, associé à une position dans l'image : - construction (E34) d'un vecteur caractéristique à partir de ladite carte de caractéristiques de référence et de ladite carte de caractéristiques courante, en fonction de ladite position dudit échantillon courant : - traitement (E37) dudit vecteur par ledit réseau de neurones de synthèse pour fournir un vecteur représentatif d'une valeur décodée de l'échantillon courant.

Description

Titre de l'invention : Procédé et dispositif de codage et décodage de séquences d'images.

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 en divisant l'image en plusieurs blocs rectangulaires, et en encodant ces blocs de pixels selon une séquence de traitement donnée. Dans les techniques existantes de compression vidéo, le traitement d'un bloc comprend typiquement une prédiction des pixels du bloc effectuée 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 ». Cette exploitation des redondances spatiales et/ou temporelles permet d'éviter de transmettre ou de stocker la valeur des pixels de chaque bloc de pixels, en représentant certains au moins des blocs par un résiduel représentant une différence entre les valeurs de prédiction des pixels du bloc et les valeurs réelles des pixels du bloc prédit.
- [0004] Les formats vidéo ne cessant d'évoluer pour compresser toujours davantage et s'adapter à la variété de formats attendus et des réseaux de communication, les possibilités de prédiction deviennent de plus en plus grandes et les algorithmes de codage et décodage classiques très complexes.
- [0005] En plus de ces approches classiques proposées par les normes de compression (MPEG, ITU), les approches basées sur l'intelligence artificielle, et notamment neuronales, tendent à se développer.
- [0006] Certaines de ces approches neuronales peuvent être vues comme une simple extension de la notion de compétition des techniques de compression susmentionnées, telles que la compétition de mode de prédiction et transformation en codage vidéo.
- [0007] D'autres approches utilisent le concept « d'auto-encodeur ». Les auto-encodeurs sont des algorithmes d'apprentissage à base de réseaux de neurones artificiels, qui permettent de construire une nouvelle représentation d'un jeu de données.

L'architecture d'un auto-encodeur est constituée de deux parties : l'encodeur et le décodeur. L'encodeur est constitué par un ensemble de couches de neurones, qui traitent les données afin de construire de nouvelles représentations dites « encodées », appelées aussi « représentations latentes ». À leur tour, les couches de neurones du décodeur reçoivent ces représentations et les filtrent afin d'essayer de reconstruire les données de départ. Les différences entre les données reconstruites et les données initiales permettent de mesurer l'erreur commise par l'auto-encodeur. L'entraînement consiste à modifier les paramètres de l'auto-encodeur afin de réduire l'erreur de reconstruction mesurée sur les différents échantillons du jeu de données. Les performances de tels systèmes à base d'auto-encodeur se font au prix d'une augmentation considérable de l'empreinte mémoire et de la complexité comparativement aux approches conventionnelles telles que proposées par les normes de compression. Ils peuvent avoir des millions de paramètres et peuvent nécessiter jusqu'à un million de MAC (multiplication-accumulation) pour décoder un seul pixel. Cela rend de tels décodeurs bien plus complexes que les décodeurs conventionnels, ce qui pourrait entraver l'adoption de la compression basée sur l'apprentissage.

- [0008] Plus récemment, une technique simple d'encodage d'image fondé sur un réseau de neurones a été décrite dans l'article « Compression with Implicit Neural representations » d'Émilien Dupont et al. (arXiv:2103.03123). La technique de codage proposée consiste à ajuster un réseau de neurones à une image, quantifier les poids du réseau et les transmettre. Au moment du décodage, le réseau de neurones est évalué à chaque position de pixel pour reconstruire l'image. Une telle technique reste cependant inefficace en termes de compression et suppose de coder indépendamment les images de la vidéo.
- [0009] Dans le domaine de la vidéo, une technique équivalente a été présentée dans l'article « Scalable Neural Video Representations with Learnable Positional Features » (NeurIPS 2022) de Subin Kim et al. La dimension temporelle est prise en compte. L'algorithme de codage génère un ensemble de trois images clés latentes bidimensionnelles, sur chaque axe spatio-temporel, pour apprendre les images communes de la vidéo. Il génère aussi une grille de représentation 3D latente pour capturer les détails locaux de la vidéo. Cette représentation est utilisée pour moduler un réseau de neurones entraîné pour traiter les trois images clés. Cependant, une telle approche manque de flexibilité dans le choix de représentation des images clés (dont le nombre est fixé à 3) et de la grille 3D (figée pour l'ensemble de la séquence). Par ailleurs elle nécessite le traitement d'une grille 3D transmise, qui est à la fois lourde en quantité de données transmises et en mémoire requise pour son traitement.
- [0010] Il existe donc un besoin pour une solution permettant de coder/compresser une séquence d'images de manière simple et efficace.

Exposé de l'invention

- [0011] L'invention vise un procédé de codage d'une pluralité d'images d'une séquence d'images comportant les étapes suivantes pour au moins une image courante à coder :
- une étape de construction, comportant les sous-étapes suivantes :
 - construction d'un ensemble de cartes de caractéristiques courantes, représentatives de ladite au moins une image courante, comprenant au moins une carte de caractéristiques courante ;
 - construction d'un ensemble de cartes de caractéristiques de référence, comprenant au moins une carte de caractéristiques de référence codée puis décodée ;
 - pour au moins un échantillon, dit échantillon courant, de ladite au moins une image courante, associé à une position dans l'image courante :
 - construction d'un vecteur caractéristique à partir de ladite au moins une carte de caractéristiques de référence et de ladite au moins une carte de caractéristiques courante, en fonction de ladite position dudit échantillon courant ;
 - traitement dudit vecteur caractéristique par un réseau de neurones artificiels, dit réseau de neurones de synthèse, défini par un ensemble de paramètres, pour fournir un vecteur représentatif d'une valeur décodée de l'échantillon courant ;
 - mise à jour d'au moins une valeur de ladite au moins une carte de caractéristiques courante et/ou d'au moins un paramètre dudit réseau, en fonction d'une mesure de performance de codage ;
 - une étape de codage dudit ensemble de cartes de caractéristiques courantes ;
 - une étape de codage dudit ensemble de paramètres dudit réseau de neurones de synthèse.
- [0012] L'invention vise aussi un procédé de **décodage** d'une séquence d'images d'un flux binaire, comportant les étapes suivantes pour au moins une image courante à décoder :
- décodage d'un ensemble de paramètres représentatifs d'un réseau de neurones de synthèse ;
 - décodage d'un ensemble de cartes de caractéristiques courantes comprenant au moins une carte de caractéristiques courante représentative de ladite au moins une image courante ;
 - obtention d'un ensemble de cartes de caractéristiques de référence, comprenant au moins une carte de caractéristiques de référence décodée ;
 - pour au moins un échantillon, dit échantillon courant, de ladite au moins une image courante à décoder, associé à une position dans l'image :
 - construction d'un vecteur caractéristique à partir de ladite au moins une carte de caractéristiques de référence et de ladite au moins une carte de caractéristiques courante, en fonction de ladite position dudit échantillon courant, et :

- traitement dudit vecteur par ledit réseau de neurones de synthèse pour fournir un vecteur représentatif d'une valeur décodée dudit au moins un échantillon courant.

- [0013] 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.
- [0014] 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).
- [0015] Par « au moins une image courante » on entend une sous-séquence d'au moins une image de la séquence.
- [0016] 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.
- [0017] Par « carte de caractéristiques », on entend une partie d'une représentation abstraite d'une image, comportant une pluralité de données scalaires variables, que l'on appelle aussi valeurs caractéristiques, par exemple des nombres réels, ou des nombres entiers, signés ou non. La représentation abstraite d'une image peut donc comporter plusieurs cartes de caractéristiques. La représentation abstraite d'une image peut aussi correspondre à une seule carte de caractéristiques comprenant un ensemble de valeurs associées à une position.
- [0018] Par carte de caractéristiques de référence, on entend une telle carte de caractéristiques obtenue pour l'ensemble des images de la séquence. Au contraire, une carte de caractéristiques courante ne concerne que la sous-séquence courante.
- [0019] Par « vecteur caractéristique construit à partir des cartes de caractéristiques en fonction d'une position » on entend un vecteur constitué d'un ou plusieurs éléments, ou données, de préférence discrètes, les données étant construites à partir des cartes de caractéristiques à une position déterminée par celle des échantillons en cours de traitement dans les images. Ce vecteur caractéristique est celui qui est appliqué à

l'entrée du réseau de neurones de synthèse. Dans le cas d'une image, un tel vecteur peut être constitué par exemple à partir d'une pluralité de valeurs prélevées dans chacune des cartes de caractéristiques aux mêmes coordonnées en abscisse et en ordonnée que la sous-séquence des échantillons à coder (resp. à décoder).

- [0020] Par « réseau de neurones de synthèse », on entend un réseau neuronal tels 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).
- [0021] 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, pondération, biais, valeur affectant le fonctionnement de la non-linéarité, etc.)
- [0022] Par « traitement par un réseau de neurones de synthèse », on entend l'application d'une fonction exprimée par un réseau de neurones au vecteur caractéristique d'entrée pour produire un vecteur de sortie représentatif de la sous-séquence des échantillons à coder (resp. décoder). Ce vecteur de sortie peut comporter une ou plusieurs données représentatives de la sous-séquence des échantillons.
- [0023] 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 les images courantes, 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.
- [0024] Par « étape de construction », on entend une étape qui vise à construire les paramètres représentatifs de l'image, avant leur codage effectif. Les sous-étapes de construction peuvent être répétées autant que nécessaire pour obtenir une mesure de performance acceptable.
- [0025] 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.

- [0026] Le procédé de codage selon l'invention réalise une construction des paramètres de codage, à partir d'une séquence d'images d'entrée, en entraînant un réseau de neurones sur des vecteurs caractéristiques associés à une position d'un échantillon à coder dans l'une des images d'entrée d'une sous-séquence d'images en cours de traitement. La sous-séquence comporte au moins une image. Ces vecteurs caractéristiques sont construits à partir de cartes de caractéristiques de l'image elle-même, et de cartes de caractéristiques de référence de la séquence, représentatives des images de la séquence. Les cartes de caractéristiques peuvent être à la résolution des images d'entrée, ou à une résolution inférieure. Au cours de l'entraînement, ou construction, les paramètres du réseau de neurones et les valeurs des cartes de caractéristiques 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 et des valeurs des cartes de caractéristiques courantes peut être effectué et mémorisé ou transmis à destination du décodeur. Selon un mode de réalisation, lorsque toute la séquence a été traitée, les cartes de référence sont stabilisées et peuvent être encodées et mémorisées ou transmises à destination du décodeur.
- [0027] Avantageusement, le processus d'entraînement permet de raffiner les paramètres du réseau de neurones de synthèse et/ou les valeurs des cartes de caractéristiques 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 ou la sous-séquence d'images d'entrée en cours de codage. L'entraînement des valeurs des cartes de caractéristiques et des paramètres du réseau de neurones de synthèse peut être conjoint. Avantageusement, le procédé de codage selon l'invention permet de compresser efficacement le signal.
- [0028] Avantageusement, le procédé de décodage est simple puisqu'il suffit de décoder les cartes de caractéristiques de la sous-séquence d'images courante, les cartes de caractéristiques de référence et le réseau de neurones de synthèse pour reconstituer une version décodée de la sous-séquence courante.
- [0029] Avantageusement, les cartes courantes capturent de manière dynamique les caractéristiques locales de la sous-séquence courante, puisqu'elles sont codées (resp. décodées) au fur et à mesure du traitement de la séquence. Les cartes de référence, codées (resp. décodées) une seule fois pour la séquence, capturent les caractéristiques globales de la séquence. On peut ainsi bénéficier à la fois des deux types de caractéristiques. Par ailleurs, une fois les cartes de référence transmises et donc disponibles côté décodeur, les cartes courantes permettent de générer l'image courante, et on peut donc concevoir un système de transmission qui travaille image par image avec une faible latence, chaque image courante étant décodée dès réception des cartes

courantes associées.

- [0030] Un tel réseau de neurones de synthèse peut avantageusement être de structure très simple avec peu de paramètres.
- [0031] De surcroît, le décodage peut être effectué par séquence d'échantillons, de manière progressive.
- [0032] Selon des modes de réalisation du procédé de codage ou de décodage :
- [0033] - Le procédé de codage comprend une étape de construction et codage d'un groupe de cartes de caractéristiques de référence représentatives de la séquence comprenant ledit ensemble de cartes de caractéristiques de référence .
- [0034] - Le procédé de décodage comporte en outre :
- une étape de décodage d'un groupe de cartes de caractéristiques de référence représentatives de la séquence et
 - obtention dudit ensemble de cartes de caractéristiques de référence à partir dudit groupe.
- [0035] Avantageusement, le groupe de carte de références est construit et codé par le codeur, et décodé par le décodeur pour l'ensemble de la séquence. Puis pour chaque sous-séquence en cours de traitement, une partie du groupe de cartes de référence est considéré pour constituer l'ensemble de référence utilisé pour la sous-séquence. On peut ainsi simplifier le décodage en conservant un nombre restreint de cartes de référence. Selon un mode de réalisation, l'ensemble est plus petit que le groupe. Selon un mode de réalisation, l'ensemble est égal au groupe.
- [0036] - Le procédé comporte une étape de transformation dudit ensemble de cartes de caractéristiques courantes et/ou de référence pour obtenir un ensemble de cartes de caractéristiques transformées à la résolution de l'image d'entrée, le procédé étant caractérisé en ce que ledit vecteur caractéristique est construit à partir desdites cartes de caractéristiques transformées. Avantageusement selon ce mode, les cartes de caractéristiques sont divisées en deux ensembles dont l'un est réservé à l'extraction des vecteurs caractéristiques et l'autre au codage. On peut ainsi dissocier les deux procédés dont la finalité est différente : les cartes du premier groupe, à coder (resp. décoder), doivent être compressées le plus efficacement possibles, alors que les cartes du second groupe, ou cartes transformées, doivent pouvoir faciliter le processus d'extraction et de construction du vecteur caractéristique.
- [0037] – Selon une variante, une au moins des cartes de caractéristiques courantes et/ou de référence est de résolution inférieure à celle de la sous-séquence à coder (resp. décoder) et l'opération de transformation comporte un suréchantillonnage. Avantageusement selon ce mode, la compression des cartes de caractéristiques est plus efficace puisque l'une au moins des cartes de caractéristiques courantes et/ou de référence, à coder (resp. décoder), comporte moins de valeurs que si elle était à la

résolution du signal. Par exemple, dans le cas d'une image numérique, l'une des cartes de caractéristiques courantes et/ou de référence peut être à la résolution 1/2, c'est-à-dire qu'elle comporte deux fois moins de valeurs en abscisse et en ordonnée que l'image d'entrée comporte d'échantillons, soit au total 4 fois moins de valeurs qu'une carte de caractéristiques à la résolution de l'image. En revanche, la carte de caractéristiques transformée, qui correspond à une transformation de cette carte courante et/ou de référence, est de résolution identique à celle de l'image. La transformation comporte donc dans ce cas au moins une opération de suréchantillonnage pour obtenir le même nombre de valeurs dans cette carte transformée que l'image d'entrée (resp. à décoder) comporte d'échantillons.

[0038] – Une au moins desdites cartes de caractéristiques courantes et/ou de référence est de résolution identique à celle de la sous-séquence à coder (resp. à décoder).

Avantageusement selon ce mode, l'une au moins des cartes de caractéristiques, à la même résolution que le signal d'entrée à coder (resp. à décoder) permet une fidélité importante et le respect des détails de la résolution initiale du signal d'image. La transformation selon un mode de réalisation conserve dans ce cas le nombre de valeurs de la carte de caractéristiques transformée ; elle peut être réduite à l'identité (aucun traitement n'est effectué sur les valeurs de la carte courante) ou comporter une opération de filtrage, de quantification, une transformation de Fourier, etc. Lors de l'étape de codage, une quantification est indispensable au bon fonctionnement du système si les cartes de caractéristiques comportent par exemple des valeurs flottantes, ou réelles. Il est nécessaire de les quantifier avant de les coder et/ou de les fournir en entrée du réseau de neurones de synthèse et/ou de prédiction. Au décodage en revanche, la quantification inverse n'est pas nécessaire, selon les modes de réalisation.

[0039] - La construction du vecteur caractéristique comporte une sous-étape d'extraction d'une valeur d'une carte de caractéristiques courante et/ou de référence à une position identique à celle de l'échantillon courant dans l'image courante à coder ou à décoder. Avantageusement, il est possible d'extraire une valeur d'une carte de caractéristiques à la même position que l'échantillon dans l'image, pour constituer un élément du vecteur caractéristique. Ce procédé est simple à mettre en œuvre. Par exemple, si l'on dispose de NF cartes de caractéristiques en entrée ($NF = NFM + NFR$, où NFM représente le nombre de cartes de caractéristiques courantes, et NFR le nombre de cartes de caractéristiques de référence), de même résolution que l'image, une simple extraction des valeurs des cartes aux coordonnées de l'échantillon courant (à la même abscisse et la même coordonnée dans la carte de caractéristiques) permet de construire directement le vecteur caractéristique de J éléments.

[0040] - La construction dudit vecteur caractéristique comporte les sous-étapes suivantes :
- extraction d'une pluralité de valeurs desdites cartes de caractéristiques courantes

et/ou de référence en fonction de ladite position dudit échantillon courant ;
- traitement desdites valeurs extraites pour obtenir le vecteur caractéristique.

- [0041] Avantageusement selon ce mode, le vecteur caractéristique est extrait des cartes de caractéristiques, qui peuvent être quelconques, et notamment à une résolution inférieure à celle de l'image à coder (resp. décoder), avant de subir un traitement. Un tel traitement peut par exemple correspondre à une quantification des données extraites, ou à une mise à l'échelle, ou à un filtrage, etc. Au codage, une quantification est indispensable au bon fonctionnement du système si les cartes de caractéristiques comportent par exemple des valeurs flottantes, ou réelles. Il est nécessaire de les quantifier avant de les coder et/ou de les fournir en entrée du réseau de neurones de synthèse et/ou de prédiction. Au décodage en revanche, la quantification inverse n'est pas nécessaire, selon les modes de réalisation.
- [0042] - Le procédé comporte une étape de construction d'un groupe de cartes de caractéristiques supplémentaires et le vecteur caractéristique est en outre construit à partir desdites cartes de caractéristiques. Avantageusement, ces cartes supplémentaires, construites de manière identique au codeur et au décodeur, ne sont ni stockées, ni transmises au niveau du codeur, ni décodées au niveau du décodeur. Elles permettent ainsi de bénéficier de données supplémentaires pour améliorer la compression sans dégrader le débit. Elles peuvent par exemple comprendre des coordonnées, des données causales disponibles dans d'autres cartes, des données concernant d'autres images déjà traitées par le codeur ou décodeur, etc.
- [0043] - Lesdits ensembles de cartes de caractéristiques de référence et courantes contiennent le même nombre de cartes de caractéristiques ; avantageusement selon ce mode, la séquence et les différentes sous séquences d'images de la séquence sont associées au même nombre de cartes de caractéristiques, qui peuvent de surcroît être de même structure. Par exemple, la séquence et les sous-séquences sont associées à cinq cartes de caractéristiques présentant une hiérarchie de résolutions décroissantes. Ceci facilite la génération et l'extraction des valeurs des cartes de caractéristiques pour produire le vecteur caractéristique.
- [0044] - Ladite au moins une carte de référence est à la même résolution que ladite au moins une carte courante. Avantageusement selon ce mode, la construction du vecteur caractéristique est facilitée puisque la carte de référence est à la même résolution que la carte courante, il est très simple de réaliser l'extraction d'une valeur à la même position ou à une position proche de celle de la valeur de la carte courante en cours de codage.
- [0045] - Ladite au moins une carte de référence est indiquée dans le flux binaire. Avantageusement selon ce mode, un simple indicateur codé dans le flux permet au décodeur d'accéder directement à la ou les cartes de référence à utiliser pour le décodage de l'image ou la sous-séquence courante, par exemple via un indicateur du

numéro de la carte de référence dans le groupe de cartes de référence.

[0046] - Ladite au moins au moins une carte de référence est prédéterminée.

Avantageusement selon ce mode, la ou des cartes de caractéristiques à utiliser pour le décodage de l'image ou la sous-séquence courante sont prédéterminées. Par exemple la carte de caractéristiques de référence est celle qui comporte le même numéro d'ordre que la carte courante dans le groupe de cartes de caractéristiques auxquelles elles appartiennent respectivement. Selon un autre exemple, toutes les cartes de caractéristiques du groupe de référence servent de cartes de référence pour la carte courante.

[0047] - Ladite au moins une carte courante est décodée en fonction de ladite au moins une carte de référence. Avantageusement selon ce mode, la (au moins une) carte de référence est utilisée non seulement pour la construction du vecteur caractéristique, mais aussi pour le décodage (par exemple entropique) d'une carte courante.

[0048] - Ladite au moins une carte courante est décodée en fonction d'au moins une carte courante. Avantageusement selon ce mode, une carte courante déjà au moins partiellement décodée peut être utilisée pour coder plus efficacement la carte courante en cours de traitement.

[0049] Corrélativement, l'invention vise aussi un dispositif de codage et un dispositif de décodage.

[0050] 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.

[0051] 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.

[0052] 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.

[0053] Le programme selon l'invention peut être en particulier téléchargé sur un réseau de type Internet.

[0054] Alternativement, chaque support d'informations 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

[0055] 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.

[0056] [Fig.1] la [Fig.1] représente schématiquement un dispositif de codage utilisé dans le cadre de l'invention ;

[0057] [Fig.2] la [Fig.2] représente schématiquement un dispositif de décodage utilisé dans le cadre de l'invention ;

[0058] [Fig.3] la [Fig.3] illustre un exemple de réseau de neurones artificiels de synthèse utilisé dans un mode de réalisation de l'invention ;

[0059] [Fig.4] la [Fig.4] illustre un exemple de réseau de neurones artificiels de synthèse utilisé dans un autre mode de réalisation de l'invention ;

[0060] [Fig.5] La [Fig.5] illustre un exemple sélection de cartes de référence utilisé dans le cadre de l'invention ;

[0061] [Fig.6] la [Fig.6] est un logigramme représentant un exemple de procédé de codage qui peut être mis en œuvre par le dispositif de codage de la [Fig.1] ;

[0062] [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.2] ;

[0063] [Fig.8] la [Fig.8] représente de manière illustrée un procédé de décodage utilisé dans un mode de réalisation de l'invention.

Description des modes de réalisation

[0064] La [Fig.1] représente schématiquement un dispositif de codage de séquences d'images numériques ENC.

[0065] Le dispositif de codage ENC comprend un module INITc d'initialisation, chargé notamment de construire et coder le groupe de cartes de références de la séquence.

[0066] Une image peut représenter une image bidimensionnelle et la séquence une pluralité d'images bidimensionnelles (séquence vidéo, composantes stéréoscopiques, multiscopiques, série d'images médicales, etc.). La séquence est une succession quelconques d'images, par exemple un GOP.

[0067] Pour une sous-séquence courante à coder v, comprenant au moins une image

courante, un module de codage noté ENC_v comprend un module GEN de génération de cartes de caractéristiques courantes, un module REF d'obtention de cartes de caractéristiques de référence, un module SE de transformation, un module XTR d'extraction de données, un module MLP de synthèse correspondant à un réseau de neurones artificiels de synthèse, un module NNC de codage de réseau de neurones apte à coder le réseau de neurones de synthèse, un module FMC de codage des cartes de caractéristiques (qui peut être utilisé également par le module INIT_c pour coder les cartes de référence), un module EVAL d'évaluation d'une performance de codage, un module MAJ de mise à jour.

[0068] 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 susmentionnée et conçues pour effectuer les fonctionnalités du module concerné, notamment comme décrit ci-dessous, lorsque ces instructions sont exécutées par le processeur.

[0069] Le dispositif de codage ENC de la [Fig.1] reçoit en entrée une sous-séquence d'au moins une image à coder, notée $I(P_{v_n})$, comportant une pluralité d'échantillons P_{v_n} . L'indice v indique le ou les numéro(s) d'image(s) de la sous-séquence dans la séquence de V images (par exemple, $v=3$, ou $v= \{2,3,4\}$). L'indice n indique le numéro de pixel dans une image de N pixels. Dans un mode de réalisation, une image de la séquence est représentée au moyen d'au moins une représentation bidimensionnelle, telle qu'une matrice de pixels, chaque pixel comportant un vecteur de composantes rouge (R), verte (G), bleue (B), ou, en variante, une composante de luminance et au moins une composante de chrominance. L'emplacement de chaque pixel est défini par ses coordonnées en abscisse et ordonnée (x_n et y_n) dans l'image. Dans un mode de réalisation, la séquence est une séquence d'images en niveau de gris représentées au moyen d'une représentation bidimensionnelle, telle qu'une matrice de pixels, chaque pixel comportant une composante de niveaux de gris, ou luminance. Dans ce cas le vecteur représentatif du pixel est réduit à une seule composante. Dans un mode de réalisation, une sous-séquence est réduite à une seule image. Dans ce cas le codeur ENC traite successivement les échantillons courants P_{v_n} à coder de l'image. Dans un mode de réalisation, une sous-séquence comporte au moins deux images successives. Dans ce cas le codeur ENC traite simultanément les échantillons courants P_{v_n} à coder des images successives, considérés comme une sous-séquence d'échantillons courants. Dans un mode de réalisation, une sous-séquence comporte un nombre prédéfini d'images successives, par exemple 4.

[0070] Comme cela sera décrit plus en détail plus loin en référence aux figures 3 à 7 :

- [0071] Le module GEN de génération de cartes de caractéristiques est configuré pour générer une pluralité de NFM cartes de caractéristiques courantes, notées FMv_i pour la sous-séquence courante v . Ces cartes sont nommées cartes de caractéristiques courantes. Selon un mode de réalisation, un module facultatif non représenté peut générer une ou plusieurs cartes supplémentaires, qui ne seront ni codées ni transmises.
- [0072] Le module REF est configuré pour générer une pluralité de NFR cartes de caractéristiques de référence, notées FRv_j , pour la séquence v , à partir d'un groupe de cartes de caractéristiques de référence, correspondant à la séquence, au nombre de $NFRMAX$. On a donc $NFR \leq NFRMAX$. Par exemple, les cartes de caractéristiques de référence considérées pour la sous-séquence en cours sont toutes celles du groupe de référence, ou sont sélectionnées comme un sous-ensemble du groupe de référence. Dans ce dernier cas, elles peuvent être prédéterminées ou leur référence peut être codée dans le flux sous forme d'un indicateur.
- [0073] L'ensemble des NFM cartes courantes et des NFR cartes de référence est utilisé par le module de codage courant $ENCv$ pour coder la sous-séquence Iv . Le nombre de cartes de caractéristiques utilisées est donc de $NFM + NFR$.
- [0074] Dans un mode de réalisation, le module SE de transformation réalise une transformation de l'ensemble de cartes de caractéristiques courantes et/ou de référence (FMv_i, FRv_j) pour générer un ensemble de cartes de caractéristiques transformées ($FMSv_i, FRSv_j$) à la même résolution que le signal d'entrée.
- [0075] Le module SE, optionnel, peut réaliser une quantification des données extraites de cet ensemble de $NFM + NFR$ cartes de caractéristiques, ou du vecteur caractéristique constitué à partir de ces données. On rappelle que la quantification d'une valeur fait référence à la mise en correspondance de cette valeur avec un membre d'un ensemble discret de symboles de code possibles. Par exemple, l'ensemble de symboles de code possibles peut être constitué de valeurs entières, et le système de quantification réalise un simple arrondi d'une valeur réelle à une valeur entière. Selon un autre exemple, la quantification consiste en une multiplication par une valeur donnée puis un arrondi. Puis le module SE peut effectuer une transformation des valeurs d'au moins une des cartes de caractéristiques, par exemple un suréchantillonnage, une interpolation, un filtrage, etc. À l'issue de la transformation, une carte de caractéristiques transformée est de même résolution que les images de la sous-séquence d'entrée. Avantagusement selon ce mode, les cartes de caractéristiques qui sont codées peuvent être de résolution inférieure à celle des images à coder, alors que les cartes transformées, qui servent à la construction des vecteurs caractéristiques, sont à la même résolution que la séquence d'images, ce qui facilite l'extraction des valeurs.
- [0076] Dans un mode de réalisation, le module SE est absent, dans ce cas les valeurs qui serviront à construire le vecteur caractéristique sont extraites des cartes de

caractéristiques courantes et de référence non transformées.

[0077] Le module XTR réalise une extraction de valeurs dans des cartes de caractéristiques courantes et de référence, optionnellement transformées, pour un échantillon courant ou une sous-séquence d'échantillons courants Pv_n à coder, en fonction de ses (leurs) coordonnées dans le signal d'entrée. Par exemple, si l'on cherche à coder l'échantillon ou la sous-séquence d'échantillons Pv_n aux coordonnées (x_n, y_n) d'une image d'entrée, le module XTR réalise une extraction de valeurs dans des cartes à des positions imposées par les coordonnées (x_n, y_n) . Dans un mode de réalisation, les valeurs extraites constituent le vecteur caractéristique Z_n . Z_n est un J-uplet, c'est-à-dire qu'il comporte J éléments, ou données z_i . Z_n comporte deux ensembles de valeurs : l'un (Zv_n) est extrait des cartes de caractéristiques courantes de la sous-séquence v , et l'autre (Zr_n) est extrait des cartes de caractéristiques de référence. Le vecteur Z_n d'indice n réfère au vecteur caractéristique du pixel ou de la séquence de pixels $P'v_n$. Selon un mode de réalisation, $J = NFM + NFR$.

[0078] Dans un mode de réalisation, le module XTR réalise un traitement des valeurs extraites, pour générer le vecteur Z_n . Le module XTR peut réaliser une quantification des données extraites de l'ensemble de cartes de caractéristiques. Le traitement peut comporter d'autres opérations, comme un filtrage, une mise à l'échelle, etc. En particulier, si le module SE n'est pas utilisé et si les cartes de caractéristiques sont à des résolutions inférieures à celle des images de la séquence, le module XTR peut prendre en compte les coordonnées des valeurs dans les cartes de résolutions inférieures.

[0079] On notera que l'un au moins des modules SE ou XTR doit effectuer une quantification des valeurs extraites des cartes de caractéristiques courantes de la sous-séquence courante.

[0080] Le module MLP contient un réseau de neurones de synthèse défini par K paramètres W_k , apte à traiter le vecteur Z_n , ou J-Uplet, en entrée, pour générer en sortie un vecteur représentatif de l'échantillon ou de la sous-séquence d'échantillons Pv_n à coder de la sous-séquence courante. Le réseau de neurones de synthèse contient selon un mode de réalisation un MLP, ou Multi Layer Perceptron, constitué d'une couche d'entrée adaptée au format d'entrée (le J-uplet), optionnellement une ou plusieurs couche(s) cachée(s), et une couche de sortie adaptée au format de sortie du vecteur de sortie, de manière générale un vecteur comportant NC (nombre de composantes) éléments pour chaque échantillon. Selon un mode de réalisation, NC est égal à 3 et le vecteur de sortie est le triplet (R, G, B) du pixel ou des pixels $P'v_n$ codé(s) puis décodé(s) de la sous-séquence courante. Selon un autre mode de réalisation, le module MLP contient un réseau de neurones convolutif.

[0081] Le module NNC réalise le codage du réseau de neurones de synthèse de l'image ou

de la sous-séquence courante, notamment de ses paramètres W_{v_k} . 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, le module NNC réalise une simulation de codage, suivi d'un décodage, à destination du module d'évaluation. Par la suite, il effectue le codage effectif des paramètres du réseau de neurones de synthèse W_{v_k} . Les paramètres codés de l'image ou de la sous-séquence courante sont notés W_{vc_k} . De manière connue, la simulation de codage peut être identique au codage effectif, ou en réaliser une approximation.

[0082] Le module INITc réalise la génération et le codage des cartes FR_j , c'est-à-dire des valeurs des cartes de caractéristiques de la séquence. Les cartes codées de référence sont notées FRC_j . Par exemple, lors de l'initialisation du codeur, toutes les cartes (NFRMAX) de la séquence sont construites, puis codées et décodées afin de pouvoir être utilisées comme référence pour une sous-séquence à coder. Selon un autre mode de réalisation, les cartes de référence sont générées dynamiquement au fur et à mesure du codage des sous-séquences, de la même manière que les cartes courantes. Selon un mode de réalisation, le codage prend en compte le voisinage d'une valeur à coder de la carte de caractéristiques de référence, dans la carte elle-même et/ou dans une autre carte de référence.

[0083] Lors du codage des sous-séquences, le module FMC réalise le codage des cartes FMv_i , c'est-à-dire des valeurs des cartes de caractéristiques de la sous-séquence courante (à l'exclusion des cartes transformées, résultant optionnellement d'un suréchantillonnage ou autre transformation par le module SE). 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, le module FMC réalise une simulation de codage, suivi d'un décodage, à destination du module d'évaluation. Par la suite, il effectue le codage effectif des valeurs des cartes FMv_i et selon un mode de réalisation, comme expliqué plus haut, des cartes de référence. Les cartes codées de la sous-séquence courante sont notées $FMcv_i$. De manière connue, la simulation de codage peut être identique au codage effectif, ou en réaliser une approximation. Le module de codage quantifie si nécessaire la représentation latente des valeurs des cartes en utilisant un quantificateur pour générer une collection ordonnée de valeurs quantifiées. Puis le module de codage compresse les données quantifiées, par exemple en utilisant un codeur entropique.

[0084] Selon un mode de réalisation, il met en œuvre un codage prenant en compte le voisinage d'une valeur à coder de la carte de caractéristiques. Selon un mode de réalisation, la position voisine est sélectionnée dans la carte en cours de codage (par exemple, la valeur voisine en haut à gauche de celle en cours de codage). Selon un mode de réalisation, ce voisinage est sélectionné dans une autre carte courante, à une position voisine. Selon un autre mode de réalisation, le voisinage est sélectionné dans

une carte de référence, à une position voisine. Selon encore un autre mode de réalisation, le voisinage est sélectionné dans une carte de caractéristiques codée et décodée correspondant à une ou des sous-séquences v' de la séquence distinctes de la sous-séquence courante v , à une position voisine, par exemple la sous-séquence précédente dans la séquence, ou deux sous-séquences passées de la séquence, ou une sous-séquence passée et d'une future, dans la mesure où l'ordre de décodage de la séquence le permet (de manière connue, l'ordre de décodage peut être différent de l'ordre chronologique de la séquence). Tous ces modes de réalisation peuvent être combinés, c'est-à-dire que le voisinage peut être constitué de valeurs extraites de la carte courante, des cartes courantes, des cartes de référence, et/ou de cartes courantes d'une autre sous-séquence. Le module FMC peut comprendre à cette fin un réseau de neurones de prédiction de carte de caractéristiques dont les paramètres peuvent être codés et transmis au décodeur, qui prend en entrée un vecteur de voisinage constitué des valeurs de voisinage et produit en sortie une prédiction de la valeur courante de la carte de caractéristiques en cours de codage, qui peut se présenter, de manière connue, sous forme d'une valeur de prédiction ou de données de probabilité.

- [0085] Le module EVAL réalise une évaluation et minimisation d'une performance de codage. La fonction d'évaluation est par exemple de type débit-distorsion. La minimisation peut être effectuée par une descente de gradient, ou tout autre procédé à la portée de l'homme du métier.
- [0086] Le module MAJ réalise une mise à jour des valeurs des cartes FMv_i à encoder de la sous-séquence courante, en fonction des résultats de la fonction de performance. Selon un mode de réalisation, il réalise aussi une mise à jour des valeurs des cartes de référence à encoder de la séquence, en fonction des résultats de la fonction de performance.
- [0087] Une fois que les cartes de caractéristiques courantes sont stabilisées, elles peuvent être codées. Selon un mode de réalisation, elles sont codées en utilisant les cartes de référence.
- [0088] La [Fig.2] représente schématiquement un dispositif de décodage DEC.
- [0089] Le dispositif de codage DEC de la [Fig.2] décode les N séquences d'échantillons Pdv_n de la sous-séquence v , en procédant par sous-séquence. A l'issue du décodage, les V images de la séquence sont décodées.
- [0090] Le dispositif de codage DEC comprend un module INITd d'initialisation, chargé notamment de décoder le groupe de cartes de références de la séquence SV , noté FRd_j . Par exemple, lors de l'initialisation du décodeur, toutes les cartes ($NFRMAX$) de la séquence sont décodées afin de pouvoir être utilisées comme référence pour une sous-séquence à décoder par le module $DECv$.
- [0091] Une sous-séquence courante à décoder v comporte au moins une image courante à

décoder notée $I(Pd_{v_n})$, chaque image comportant respectivement une pluralité d'échantillons. Le décodeur DEC_v dédié à la sous-séquence reçoit en entrée les données nécessaires pour décoder la sous-séquence. Comme pour le codeur, l'indice v indique le ou les numéro(s) d'image(s) de la sous-séquence dans la séquence de V images (par exemple, $v=3$, ou $v= \{2,3,4\}$). L'indice n indique le numéro de pixel dans une image de N pixels ou une sous-séquence d'images de N pixels. Le sous-module de décodage reçoit en entrée au moins :

- des données encodées organisées en NFM cartes de caractéristiques FMc_{v_i} (les cartes de caractéristiques courantes encodées) ;
- les paramètres encodés Wvc_k du réseau de neurones de synthèse correspondant à la sous-séquence courante, optionnellement les paramètres d'autres réseaux de neurones comme un réseau de neurones de suréchantillonnage, de prédiction de carte de caractéristiques, etc.;

[0092] Par ailleurs, le module DEC_v obtient (par exemple, lit dans la mémoire du décodeur) un ensemble de cartes de caractéristiques de référence, notées FRd_{v_j} , issues du groupe de cartes de référence FRd_j .

[0093] Le module de décodage DEC_v comprend, pour une sous-séquence courante, un module NND de décodage de réseau(x) de neurones apte à décoder le réseau de neurones de synthèse MLP', un module FMD de décodage des cartes de caractéristiques courantes (qui peut être utilisé également par le module INITd pour décoder les cartes de référence), un module REF' d'obtention de cartes de caractéristiques de référence, un module XTR' d'extraction de données, un module SE' de transformation inverse, un module MLP' correspondant à un réseau de neurones de synthèse, optionnellement un module de génération de cartes de caractéristiques supplémentaires. Selon un mode de réalisation, il produit en sortie une sous-séquence décodée courante, comprenant au moins une image notée $I(Pd_{v_n})$, comportant une pluralité d'échantillons ou séquences d'échantillons décodés Pd_{v_n} .

[0094] Le module de décodage FMD décompresse les données quantifiées, par exemple en utilisant un codeur entropique. Selon un mode de réalisation, il met en œuvre un décodage prenant en compte le voisinage d'une valeur à décoder de la carte de caractéristiques. Les modes de réalisation proposés pour le codeur s'appliquent ici en remplaçant le codage par un décodage des valeurs. Notamment, le voisinage peut être sélectionné dans une carte courante (y compris celle en cours de décodage), une carte de référence, ou une carte courante d'une séquence distincte de la sous-séquence courante. Le module FMD peut comprendre à cette fin un réseau de neurones de prédiction de carte de caractéristiques, comme décrit pour le module FMC. Les cartes décodées par le module FMD, au nombre de NFM, sont notées FMd_{v_i} (cartes de caractéristiques de l'image courante).

- [0095] Les paramètres du réseau de neurones de synthèse (MLP') décodés par le module NND sont notés Wdv_k .
- [0096] Le module REF' constitue un ensemble de NFR cartes de caractéristiques, dites cartes de caractéristiques de référence, notées $FRdv_j$, issues d'un groupe de cartes de caractéristiques de référence précédemment codées puis décodées, correspondant à la séquence, au nombre de NFRMAX. On a donc $NFR \leq NFRMAX$. Le module REF' est identique à celui du codeur et les mêmes exemples s'appliquent. Par exemple, les cartes de caractéristiques de référence considérées pour la sous-séquence en cours sont toutes celles du groupe de référence, ou sont sélectionnées comme un sous-ensemble du groupe de référence. Dans ce dernier cas, elles peuvent être prédéterminées ou leur référence peut être décodée dans le flux sous forme d'un indicateur.
- [0097] L'ensemble des NFM cartes courantes et des NFR cartes de référence est utilisé par le module de décodage courant pour décoder la sous-séquence courante v . Le nombre de cartes de caractéristiques utilisées est donc de $NFM + NFR$.
- [0098] Le décodeur peut aussi générer une ou plusieurs cartes supplémentaires, identiques aux cartes supplémentaires générées par le codeur pour la sous-séquence courante.
- [0099] Dans un mode de réalisation, le module SE' de transformation réalise une transformation de l'ensemble de cartes de caractéristiques courantes et/ou de référence ($FMdv_i, FRdv_j$) pour générer un ensemble de cartes de caractéristiques transformées ($FMS'v_i, FRS'v_j$) à la même résolution que le signal d'entrée.
- [0100] Le module SE' est identique au module SE de la [Fig.1]. Le module SE', optionnel, peut réaliser une quantification inverse des données extraites de cet ensemble de $NFM+NFR$ cartes de caractéristiques. Puis le module SE' peut effectuer une transformation des valeurs d'au moins une des cartes de caractéristiques, par exemple un suréchantillonnage, une interpolation, un filtrage, etc. À l'issue de la transformation, une carte de caractéristiques transformée est de même résolution que les images de la sous-séquence à décoder. Avantageusement selon ce mode, les cartes de caractéristiques qui sont décodées peuvent être de résolution inférieure à celle des images à décoder, alors que les cartes transformées, qui servent à la construction des vecteurs caractéristiques, sont à la même résolution que la séquence d'images, ce qui facilite l'extraction des valeurs.
- [0101] Dans un mode de réalisation, le module SE' est absent, dans ce cas les valeurs qui serviront à construire le vecteur caractéristique sont extraites des cartes de caractéristiques courantes et de référence non transformées.
- [0102] Le module XTR' est identique au module XTR de la [Fig.1]. Il réalise une extraction de valeurs dans des cartes de caractéristiques courantes et de référence, pour un échantillon courant ou une sous-séquence d'échantillons courants Pdv_n à décoder, en fonction de ses (leurs) coordonnées dans le signal d'entrée. Par exemple, si l'on

cherche à décoder l'échantillon ou la sous-séquence d'échantillons Pv_n aux coordonnées (x_n, y_n) d'une image d'entrée, le module XTR' réalise une extraction de valeurs dans des cartes à des positions imposées par les coordonnées (x_n, y_n) . Dans un mode de réalisation, les valeurs extraites constituent le vecteur caractéristique Zd_n . Zd_n comporte deux ensembles de valeurs : l'un (Zdv_n) est extrait des cartes de caractéristiques courantes de la sous-séquence v , et l'autre (Zdr_n) est extrait des cartes de caractéristiques de référence. Zd_n est un J-uplet, c'est-à-dire qu'il comporte J éléments, ou données zd_i . Le vecteur Zd_n d'indice n réfère au vecteur caractéristique du pixel ou de la sous-séquence de pixels Pdv_n . selon un mode de réalisation, $J = NFM + NFR$.

[0103] Dans un mode de réalisation, le module XTR' réalise un traitement des valeurs extraites, pour générer le vecteur Zd_n . Le module XTR' peut réaliser une quantification inverse des données extraites de l'ensemble de cartes de caractéristiques. Le traitement peut comporter d'autres opérations, comme un filtrage, une mise à l'échelle, etc. En particulier, si le module SE' n'est pas utilisé et si les cartes de caractéristiques sont à des résolutions inférieures à celle des images de la séquence, le module XTR' peut prendre en compte les coordonnées des valeurs dans les cartes de résolutions inférieures.

[0104] On notera que l'un au moins des modules SE' ou XTR' doit effectuer une quantification inverse des valeurs extraites des cartes de caractéristiques courantes de la sous-séquence courante, si une quantification a été effectuée au codeur.

[0105] Le module MLP' contient un réseau de neurones dit de synthèse, défini par K paramètres Wd_k , apte à traiter le vecteur Zd_n , ou J-Uplet, en entrée, pour générer en sortie un vecteur représentatif de l'échantillon ou de la sous-séquence d'échantillons Pdv_n à décoder. Le réseau de neurones de synthèse contient selon un mode de réalisation un MLP, ou Multi Layer Perceptron, constitué d'une couche d'entrée adaptée au format d'entrée (le J-uplet), optionnellement une ou plusieurs couche(s) cachée(s), et une couche de sortie adaptée au format de sortie du vecteur de sortie, de manière générale un vecteur comportant NC (nombre de composantes) éléments pour chaque échantillon. Il peut aussi comprendre un réseau de neurones de convolution. Selon un mode de réalisation, $K=3$ et le vecteur de sortie est le triplet (R, G, B) ou (Y, U, V) de chaque pixel Pdv_n décodé. Le module MLP' est de structure identique au module MLP, et ses paramètres sont soit identiques si le codage de ses paramètres W_k est sans perte, soit différents si le codage est réalisé avec pertes.

[0106] Lorsque tous les échantillons Pdv_n de la sous-séquence v ont été décodés, on dispose d'un signal reconstruit $I(Pdv_n)$, selon un mode de réalisation une image Iv comportant N échantillons décodés sous la forme de N vecteurs Pdv_n , selon un autre mode une sous-séquence d'images Iv comportant chacune N échantillons décodés sous la forme de N vecteurs Pdv_n .

- [0107] 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
- [0108] La [Fig.3] illustre un exemple de réseau de neurones artificiels de synthèse utilisé au codage et au décodage selon un mode de réalisation de l'invention.
- [0109] Le réseau de neurones artificiels de synthèse utilisé au codage, MLP, et le réseau de neurones artificiels de synthèse utilisé au décodage, MLP', sont définis par une structure identique, comprenant par exemple une pluralité de couches de neurones artificiels, et par un ensemble de poids et fonctions d'activation associés respectivement aux neurones artificiels du réseau concerné.
- [0110] Une représentation vectorielle d'un échantillon courant (un vecteur Z_n ou Zd_n issu des cartes de caractéristiques $FMv_j/FMSv_j$ et $FRv_j/FRSv_j$ au codeur ou $FMdv_j/FMS'v_j$ et $FRdv_j/FRS'v_j$ au décodeur) est appliquée en entrée (c'est-à-dire sur une couche d'entrée) du réseau de neurones artificiels de synthèse MLP ou MLP'. Le vecteur d'entrée comporte des valeurs issues des cartes courantes (Zv_n, Zdv_n) et des valeurs issues des cartes de référence (Zr_n, Zdr_n).
- [0111] De manière générale, le réseau de neurones artificiels MLP/MLP' produit en sortie une pluralité de vecteurs P'_{vn} ou Pd_{vn} représentatifs des échantillons reconstruits (au codage) ou décodés (au décodage), selon un mode de réalisation les composantes de couleur (R, G, B ou Y, U, V) constitutives des pixels couleur de la sous-séquence d'images.
- [0112] Sur la [Fig.3], la sous-séquence d'images correspond à une image, chaque sous-séquence d'échantillons comporte donc un seul échantillon et le vecteur correspondant est noté P'_{1n} (au codeur) ou Pd_{1n} (au décodeur). Selon l'exemple représenté en [Fig.3], 4 valeurs sont extraites de chaque ensemble en entrée. Le réseau de neurones artificiels de synthèse produit en sortie les composantes constitutives d'un pixel de l'image.
- [0113] La concaténation de tous ces pixels reconstruits dans l'image constitue l'image décodée, ou reconstruite.
- [0114] Au codeur, le réseau de neurones artificiels de synthèse MLP est entraîné sur l'image, de sorte à minimiser les différences entre la représentation en entrée de la sous-séquence courante $I(Pv_n)$ et sa représentation $I(P'v_n)$ en sortie, tout en minimisant également la quantité de données à encoder. Le module EVAL effectue une mesure de performance en ce sens. Une fois que l'entraînement du codeur est

terminé, les paramètres du réseau sont encodés, soit sans pertes, auquel cas le réseau de neurones MLP' est identique à MLP, soit avec pertes, auquel cas le réseau MLP' peut être légèrement différent de MLP.

- [0115] La [Fig.4] illustre un exemple de réseau de neurones artificiels de synthèse utilisé au codage et au décodage selon un autre mode de réalisation de l'invention.
- [0116] La [Fig.4] est identique à la [Fig.3] pour ce qui concerne l'entrée du réseau de neurones.
- [0117] De manière générale, le réseau de neurones artificiels MLP/MLP' produit en sortie une pluralité de vecteurs P'_{vn} ou Pd_{vn} représentatifs des échantillons reconstruits (au codage) ou décodés (au décodage), selon un mode de réalisation les composantes de couleur (R, G, B ou Y, U, V) constitutives des pixels couleur de la sous-séquence d'images.
- [0118] Sur la [Fig.4], la sous-séquence d'images correspond à deux images, une sous-séquence d'échantillons comporte donc deux échantillons et les deux vecteurs (R, G, B) ou (Y, U, V) correspondants sont notés P'_{1n}/Pd_{1n} et P'_{2n}/Pd_{2n} . La concaténation de tous ces pixels reconstruits dans les deux images constitue la sous-séquence décodée, ou reconstruite.
- [0119] La [Fig.5] illustre un exemple de sélection de cartes de référence dans une séquence vidéo.
- [0120] La séquence représentée sur la [Fig.5] comporte 4 images, numérotées de 1 à 4. Par exemple, il s'agit d'une période Intra d'un codeur vidéo standard, ou d'un GOP, etc. Elle est constituée de 3 sous-séquences $v=1$, $v=\{2,3\}$ et $v=4$. Les cartes de références FR_j (resp. FRd_j) de la séquence sont au nombre de 3 ($NFR = 3$) et de résolutions différentes.
- [0121] La sous-séquence $v=\{2,3\}$ est codée (resp. décodée) en utilisant 3 cartes courantes $FM_{\{2,3\}_{i=1,2,3}}$ (resp. $FMd_{\{2,3\}_{i=1,2,3}}$) et une carte de référence $FR_{\{2,3\}_{j=1}}$ (resp. $FRd_{\{2,3\}_{j=1}}$) de la séquence. Au total, 4 cartes caractéristiques sont présentées (après avoir été éventuellement transformées par le module SE/SE', puis leurs vecteurs caractéristiques extraits par le module XTR/XTR') à l'entrée du réseau de synthèse MLP (resp. MLP') qui synthétise les échantillons des images codées (resp. décodées) 2 et 3.
- [0122] La sous-séquence $v=\{4\}$ est codée (resp. décodée) en utilisant 3 cartes courantes $FM_{4_{i=1,2,3}}$ (resp. $FMd_{4_{i=1,2,3}}$) et trois cartes de référence $FR_{4_{j=1,2,3}}$ (resp. $FRd_{4_{j=1,2,3}}$) de la séquence. Au total, 6 cartes caractéristiques sont présentées (après avoir été éventuellement transformées par le module SE/SE', puis leurs vecteurs caractéristiques extraits par le module XTR/XTR') à l'entrée du réseau de synthèse MLP (resp. MLP') qui synthétise les échantillons de l'image codée (resp. décodée) numéro 4.
- [0123] Les exemples numériques choisis sont indicatifs.
- [0124] La [Fig.6] est un logigramme représentant un exemple de procédé de codage qui peut

être mis en œuvre par le dispositif de codage de la [Fig.1].

- [0125] Selon un mode de réalisation, la sous-séquence courante est une séquence d'images bidimensionnelles, chaque sous-séquence d'échantillon à coder est donc un ensemble de pixels Pv_n de coordonnées (x_n, y_n) dans les images respectives $I(Pv_n)$ de la séquence à coder. Selon un mode de réalisation, la sous-séquence courante comporte une seule image.
- [0126] L'encodage se déroule en deux phases principales :
- [0127] Dans une première phase, dite phase de construction, un premier apprentissage est réalisé à l'étape d'initialisation **E0** pour construire, coder et décoder un groupe d'images de référence FRc_j de la séquence SV . Selon un mode de réalisation, SV comporte une série d'images, par exemple des images temporelles d'une séquence vidéo. Selon un mode de réalisation, les images sont des composantes multivues ou 3D d'une image ou séquence d'images multivues ou 3D. Selon un mode de réalisation, les images correspondent à une période Intra ou un GOP d'un codeur vidéo classique.
- [0128] Ces cartes peuvent être insérées dans le flux B1 après codage. Selon un mode de réalisation, ces cartes sont générées avant le traitement des sous-séquences. Selon un autre mode de réalisation, elles sont générées au fil du traitement des sous-séquences, par optimisation, de la même manière que les cartes de caractéristiques courantes.
- [0129] Pour une sous-séquence d'entrée v comprenant au moins une image $I(Pv_n)$, un apprentissage est réalisé, afin de déterminer les valeurs des cartes FMv_i et des paramètres W_k pour optimiser une fonction de coût globale. L'apprentissage est par exemple réalisé par une descente de gradient, suivie d'une mise à jour des paramètres du réseau de neurones MLP et des valeurs des cartes de caractéristiques courantes et/ou de référence. Comme il est connu dans l'état de l'art, la fonction de coût peut être de type débit-distorsion, ou débit, ou distorsion, ou perceptuelle. Pour mesurer le débit R , il est nécessaire de simuler le codage des cartes FMv_i , optionnellement des cartes FRv_j , puis de mesurer le débit de codage associé (la taille du flux B1). Selon un mode de réalisation, on ne simule pas le codage des paramètres W_k car leur influence est moins importante que celle des cartes de caractéristiques. Selon un mode de réalisation, on simule aussi le codage des paramètres W_k et on mesure le débit associé (la taille du flux B2). Pour mesurer la distorsion D , il est nécessaire de simuler le codage puis le décodage d'une partie au moins de la séquence d'images, pour obtenir au moins une séquence de pixels $P'v_n$ résultant d'une simulation de codage puis de décodage des échantillons d'indice n , puis de mesurer l'écart entre cette sous-séquence d'images $I(Pv_n)$ en entrée et une partie correspondante de la séquence $I(P'v_n)$ codée puis décodée.
- [0130] Puis lors d'une deuxième phase, dite phase de codage, les cartes FMv_i et les paramètres W_k sont encodés pour produire les valeurs codées $FMcv_i$ et Wcv_k avant

transmission ou stockage. Ils constituent la représentation compressée de la séquence d'entrée $I(Pv_n)$.

- [0131] On va maintenant décrire les étapes d'un procédé selon un mode de réalisation de l'invention.
- [0132] Lors d'une étape **E20**, une séquence d'entrée $I(Pv_n)$ à coder, comportant au moins une image comportant une pluralité de N échantillons Pv_n , est fournie en entrée du procédé.
- [0133] Lors d'une étape **E21**, NFM cartes de caractéristiques courantes et optionnellement de référence sont initialisées. Par la suite, les paramètres W_k du réseau de neurones MLP et les valeurs des cartes doivent être optimisés durant la phase de construction.
- [0134] Selon un mode de réalisation, les cartes de caractéristiques sont initialisées par des valeurs constantes prédéfinies. Selon un autre mode de réalisation, les cartes de caractéristiques sont initialisées par un ensemble de valeurs réelles aléatoires. Les cartes de caractéristiques sont par la suite mises à jour, ou raffinées, lors d'une étape **E22**, par le module de mise à jour MAJ du codeur au cours de son apprentissage.
- [0135] Lors d'une étape **E23**, NFR cartes de référence sont obtenues par le module REF, parmi le groupe des cartes de caractéristiques associées à la séquence. Le module REF constitue un ensemble de NFR cartes de caractéristiques de référence, notées FR_{v_j} , issues d'un groupe de cartes de caractéristiques de référence FR_j , précédemment décodées, au nombre de NFRMAX ($NFR \leq NFRMAX$). Par exemple, les cartes de caractéristiques de référence considérées pour la sous-séquence en cours de codage sont toutes celles du groupe de référence ($FR_{v_j} = FR_j$) ou sont sélectionnées comme un sous-ensemble du groupe de référence. Dans ce dernier cas, elles peuvent être prédéterminées. Par exemple, on peut utiliser les cartes de caractéristiques de référence de même taille que les cartes de caractéristiques courantes. Leur référence peut alternativement être codée dans le flux sous forme d'un indicateur. Par exemple, la carte de référence utilisée pour le codage/décodage des images $v=\{2,3\}$ de la [Fig.5] est l'image numéro 1 du groupe de cartes de référence. Cette indication peut être codée dans le flux.
- [0136] L'ensemble des NFM cartes courantes et des NFR cartes de référence constitue un ensemble de cartes de caractéristiques pour le réseau de synthèse MLP.
- [0137] Selon un mode de réalisation, les cartes de caractéristiques courantes et/ou de référence sont de même résolution que les images de la séquence d'entrée $I(Pv_n)$ et comportent donc chacune le même nombre de valeurs N qu'il y a d'échantillons Pv_n à coder dans chaque image. Selon un mode de réalisation, les cartes de caractéristiques courantes et/ou de référence sont de résolution inférieure ou égale à celle des images de la séquence d'entrée $I(Pv_n)$ et comportent donc, pour au moins l'une d'entre elles, un nombre N' de valeurs à coder inférieur à N . Selon une variante, la première carte

courante (resp. carte de référence) est à la résolution des images et chaque carte suivante est à une résolution moitié de la précédente. Selon un mode de réalisation, plusieurs cartes sont de même résolution, inférieure à celle des images de la séquence d'entrée $I(Pv_n)$.

- [0138] Lors d'une étape **E24**, selon un mode de réalisation, les NFM cartes de l'ensemble de cartes courantes FMv_i et/ou les NFR cartes de référence FRv_j sont transformées par le module SE pour générer des cartes transformées $FMSv_i$, $FRSv_j$ à la résolution des images d'entrée. Dans ce mode, les vecteurs caractéristiques sont extraits de préférence des cartes transformées, et non directement des cartes courantes à coder et/ou des cartes de référence. Les cartes transformées ne sont ni stockées ni transmises, elles servent uniquement à la construction des vecteurs caractéristiques.
- [0139] Selon un mode de réalisation, une ou plusieurs cartes supplémentaires, formant un groupe supplémentaire de cartes de caractéristiques supplémentaires, sont générées, et complètent l'ensemble de cartes de caractéristiques constitué. Elles servent à la construction du vecteur caractéristique mais ne sont ni stockées ni transmises.
- [0140] Selon un mode de réalisation, NFM + NFR cartes $FMSv_i$, $FRSv_j$ sont générées. Selon un mode de réalisation, chaque carte FMv_i est transformée en une carte $FMSv_i$. Selon un mode de réalisation, chaque carte FRv_j de référence est transformée en une carte $FRSv_j$.
- [0141] Selon un mode de réalisation, au moins une carte FMv_i , FRv_j est de résolution inférieure à celle des images de la séquence à coder et l'opération de transformation comporte un suréchantillonnage pour que la carte transformée comporte le même nombre d'échantillons que les images de la séquence. Le suréchantillonnage consiste à rajouter des valeurs dans les cartes $FMSv_i$ et/ou $FRSv_j$ pour atteindre la résolution des images de la séquence d'entrée. Il peut être simple (par réplication du plus proche voisin) ou comporter une interpolation (linéaire, polynomiale, par filtrage, etc.). La transformation peut comporter un réseau de neurones de suréchantillonnage.
- [0142] Lors d'une étape **E25**, des valeurs sont extraites par le module XTR :
- des cartes de caractéristiques courantes FMv_i (ou éventuellement $FMSv_i$ transformées, et optionnellement supplémentaires).
 - des cartes de caractéristiques de référence FRv_j , ou éventuellement $FRSv_j$ transformées.
- [0143] Cette extraction est réalisée en fonction des coordonnées (x_n, y_n) de la sous-séquence d'échantillons Pv_n courants de la sous-séquence courante. Elle peut aussi être réalisée en fonction de la résolution de la carte considérée. Les séquences d'échantillons à coder sont par exemple traitées par ordre séquentiel, de $n=1$ à $n=N$.
- [0144] Selon un mode de réalisation, le vecteur Z_n caractéristique résulte directement de cette extraction.

- [0145] Selon un mode de réalisation, le vecteur Z_n caractéristique est construit par traitement des valeurs extraites. Le traitement peut comporter une quantification des valeurs extraites ou du vecteur Z_n constitué, si nécessaire. Le traitement peut comporter d'autres opérations, comme un filtrage, une mise à l'échelle, l'application d'une fonction quelconque, de préférence monotone, etc. Dans un mode de réalisation, Z_n est un J-uplet construit à partir de valeurs prélevées dans les cartes de caractéristiques courantes et de référence (et optionnellement supplémentaires) en appliquant le traitement à une ou plusieurs valeurs des cartes, par exemple un filtrage des valeurs voisines de la valeur ciblée dans une carte. Par exemple, dans une carte qui se trouve à la même résolution que le signal d'entrée, on peut extraire les valeurs situées aux coordonnées (x_n, y_n) , (x_n-1, y_n) , (x_n, y_n-1) et (x_n-1, y_n-1) et appliquer un traitement à ces valeurs (filtrages, moyennage, interpolation, etc.) pour obtenir la valeur finale (z_i) de l'élément i du vecteur Z_n relative à cette carte. Selon un autre exemple, dans une carte qui se trouve à une résolution moitié du signal d'entrée, on peut considérer les valeurs situées aux coordonnées $(x_n/2, y_n/2)$, $(x_n/2-1, y_n/2)$, $(x_n/2, y_n/2-1)$ et $(x_n/2-1, y_n/2-1)$ et appliquer un traitement à ces valeurs (filtrages, moyennage, interpolation, etc.) pour obtenir la valeur finale (z_i) de l'élément i du vecteur Z_{V_n} ou Z_{R_n} relatif à cette carte de caractéristiques courante, de référence (ou supplémentaire).
- [0146] Dans un mode de réalisation, Z_n comporte autant de valeurs que de cartes en entrée.
- [0147] Dans un mode de réalisation, Z_n est un J-uplet comportant des valeurs situées aux coordonnées (x_n, y_n) d'un pixel courant P_{V_n} , comme il sera illustré à l'appui de la [Fig.8]. Le vecteur Z_n d'indice n réfère au vecteur caractéristique du pixel ou de la séquence de pixels P_{V_n} . Selon un mode de réalisation, $J = N_{FM} + N_{FR}$. Le vecteur Z_n ($z_{V_1} \dots z_{V_{N_{FM}}}, z_{R_1} \dots z_{R_{N_{FR}}}$), comporte deux ensembles de valeurs : l'une (Z_{V_n}) est extraite des cartes de caractéristiques courantes de la sous-séquence v , et l'autre (Z_{R_n}) est extraite des cartes de caractéristiques de référence. Optionnellement il peut comporter des valeurs issues des cartes supplémentaires.
- [0148] Dans un mode de réalisation, Z_n est un J-uplet construit à partir de valeurs prélevées dans les cartes de caractéristiques courantes et de référence (et optionnellement supplémentaires) à des coordonnées qui peuvent être différentes selon les cartes. Par exemple, si les cartes sont à des résolutions différentes, les coordonnées sont adaptées (par une mise à l'échelle) pour correspondre à la résolution de chaque carte.
- [0149] Lors d'une étape **E26**, le vecteur Z_n est traité par le réseau de neurones MLP pour générer en sortie la sous-séquence d'échantillons P_{V_n} à coder, selon un mode de réalisation les triplets (R, G, B) des échantillons P'_{V_n} (les échantillons P_{V_n} codés puis décodés) des images de la sous-séquence.
- [0150] Lors d'une étape **E27**, les cartes FM_{V_i} courantes sont codées par le module FMC du codeur. Durant la phase de construction, cette opération est une simulation de codage.

Durant la phase de codage, cette opération est un codage effectif et les valeurs codées constituent le flux B1. La simulation peut être identique au codage effectif mais elle peut aussi être différente (par exemple, simplifiée). Pour ce codage, on peut utiliser toute technique connue visant à compresser les valeurs des cartes. Dans un mode de réalisation, les cartes de référence sont aussi codées à cette étape.

- [0151] Dans un mode de réalisation, les cartes sont codées dans l'ordre (FM_{v1}, FM_{v2}, ..., FM_{v4}), et les variables de chaque carte dans un ordre prédéfini, par exemple lexicographique. Dans un mode de réalisation, les cartes de référence sont aussi codées dans l'ordre. Chaque carte subit un codage entropique. Le codage entropique produit un flux compressé B1 dont le débit est mesuré ultérieurement au cours d'une étape E29.
- [0152] Lors d'une étape **E28**, les paramètres W_k du réseau de neurones MLP sont quantifiés et codés. Durant la phase de construction, cette opération est une simulation de codage. Durant la phase de codage, cette opération est un codage effectif et les valeurs codées constituent le flux B2. La simulation peut être identique au codage effectif mais elle peut aussi être différente (par exemple, simplifiée). On peut utiliser à cette fin toute technique connue, par exemple la norme de codage de réseaux de neurones proposée par le standard MPEG-7 partie 17, aussi appelée NNR (Neural Network Representation). On notera que dans ce cas, il faut choisir la quantité de dégradation que le codage apporte aux paramètres W_k . La structure et les paramètres W_k du réseau de neurones sont initialisés par exemple lors de la première itération de cette étape. Ces paramètres sont par la suite mis à jour, ou raffinés, au cours de la phase de construction, lors des itérations ultérieures du procédé.
- [0153] Selon un mode de réalisation, les paramètres du réseau de neurones sont initialisés par des valeurs prédéfinies connues pour donner un résultat satisfaisant (par exemple, à la suite d'un entraînement sur un corpus de séquences d'images). Selon un autre mode de réalisation, les paramètres W_k du réseau de neurones sont initialisés par un ensemble de valeurs aléatoires.
- [0154] Lors d'une étape **E29**, une mesure de performance est évaluée.
- [0155] À cet effet, les débits de simulation de codage associés aux cartes de caractéristiques (simulation du flux B1 par codage des cartes FM_{v_i}, FR_{v_j}) et optionnellement aux paramètres du réseau de neurones (simulation du flux B2 par codage des paramètres W_k) sont mesurés.
- [0156] Selon un mode de réalisation, la fonction de coût est de type débit-distorsion, notée (D+L*R), où D par exemple l'erreur quadratique mesurée entre les images d'entrée et les images décodées (ou l'erreur mesurée sur un sous-ensemble d'échantillon des images). Selon un autre exemple, D est calculé à partir d'une fonction perceptuelle telle que le SSIM (pour *Structural SIMilarity*), ou MSSSIM (pour *Multi-scale Structural*

SIMilarity). Selon un mode de réalisation, R est le débit simulé du flux B1 ; selon un autre mode de réalisation, R est le débit total utilisé pour coder cette image, c'est-à-dire la somme des débits simulés de B1 et B2. L est un paramètre qui règle le compromis débit-distorsion. D'autres fonctions de coût sont possibles.

- [0157] Tant que la fonction de coût n'a pas atteint son minimum, la mesure de performance n'est pas satisfaisante, et le procédé est réitéré à partir de l'étape E22. Alternativement, le procédé peut être interrompu au bout d'un temps prédéfini ou d'un nombre d'itérations prédéfini, de façon à en contrôler la complexité ou la durée. Cette minimisation peut être effectuée par un mécanisme connu comme une descente de gradient avec mise à jour des paramètres au cours de l'étape E22 pour les valeurs des cartes de caractéristiques et E27 pour les paramètres du réseau.
- [0158] Lors d'une étape EF, si la fonction de coût a atteint son minimum, l'entraînement s'arrête. Si une version codée correspondant à la dernière simulation des paramètres du réseau de neurones et des cartes de caractéristiques est disponible, les flux B1 et B2 peuvent en être constitués. Selon un autre mode de réalisation, le codage effectif des paramètres mis à jour du réseau de neurones et des valeurs des cartes de caractéristiques (FMv_i) est effectué à cette étape pour produire les paramètres encodés Wc_k et FMc_k qui constituent des flux B1 et B2. Selon un mode de réalisation, les cartes FRv_j sont codées à la fin de l'apprentissage, lorsque toute la séquence a été traitée.
- [0159] Les flux B1 et B2 peuvent être concaténés pour produire un flux final. Selon un mode de réalisation, le flux B2 des paramètres codés du réseau de neurones est stocké ou transmis avant le flux B1, afin de pouvoir être décodé avant le flux B1.
- [0160] La [Fig.6] est un logigramme représentant un exemple de procédé de décodage qui peut être mis en œuvre par le dispositif de décodage de la [Fig.2].
- [0161] Le décodage décrit concerne une sous-séquence d'images comportant au moins une image de la séquence à décodé on suppose que le groupe de cartes de référence de la séquence FRC_j , a été décodé par le module d'initialisation. Selon un mode de réalisation, on utilise une technique de prédiction d'une valeur de carte de caractéristiques par son voisinage, comme il a été décrit à l'appui de la [Fig.1]. Dans un mode de réalisation, les cartes de référence sont décodées dans l'ordre, et les valeurs de chaque carte dans un ordre prédéfini, par exemple lexicographique.
- [0162] Lors d'une étape **E30**, les flux B1 et B2 sont extraits du flux encodé. Ils contiennent respectivement, pour une sous-séquence courante de la séquence, les représentations codées des cartes courantes $FMcv_i$, et des paramètres Wcv_k du réseau de neurones.
- [0163] Lors d'une étape **E31**, les NFM cartes $FMdv_i$ de la sous-séquence courante sont générées par décodage des valeurs $FMcv_i$. Pour ce décodage, on utilise par exemple un décodage entropique. Selon un mode de réalisation, on utilise une technique de prédiction d'une valeur de carte de caractéristiques par son voisinage, comme il a été

décrit à l'appui de la [Fig.1]. Dans un mode de réalisation, les cartes $FMdv_i$ sont décodées dans l'ordre ($FMdv_1, FMdv_2, \dots, FMdv_4$), et les valeurs de chaque carte dans un ordre prédéfini, par exemple lexicographique.

[0164] Lors d'une étape **E32**, NFR cartes de référence sont obtenues par le module REF', parmi le groupe des cartes de caractéristiques associées à la séquence. Le module REF' constitue un ensemble de NFR cartes de caractéristiques de référence, notées $FRdv_j$, issues du groupe de cartes de caractéristiques de référence FRd_j , précédemment décodées, au nombre de NFRMAX (NFR \leq NFRMAX). Par exemple, les cartes de caractéristiques de référence considérées pour la sous-séquence en cours sont toutes celles du groupe de référence ($FRv_j = FR_j$) ou sont sélectionnées comme un sous-ensemble du groupe de référence. Dans ce dernier cas, elles peuvent être prédéterminées. Par exemple, on peut utiliser les cartes de caractéristiques de référence de même taille que les cartes de caractéristiques courantes. Leur référence peut alternativement être codée dans le flux sous forme d'un indicateur. Par exemple, la carte de référence utilisée pour le décodage des images $v = \{2, 3\}$ de la [Fig.5] est l'image numéro 1 du groupe de cartes de référence. Cette indication peut être lue et décodée dans le flux. Au contraire, toutes les cartes de référence sont utilisées pour le décodage de l'image $v = 4$ de la [Fig.5]. Il n'est donc pas nécessaire de lire cette indication dans le flux.

[0165] L'ensemble des NFM cartes courantes et des NFR cartes de référence constitue un ensemble de cartes de caractéristiques pour le réseau de synthèse.

[0166] Selon des modes de réalisation tels que décrits pour le codeur :

[0167] - Les cartes de caractéristiques courantes et/ou de référence sont de même résolution que la séquence I (Pdv_n) à reconstituer, c'est-à-dire qu'elles comportent N valeurs.

[0168] - Les cartes de caractéristiques courantes et/ou de référence sont à une résolution inférieure ou égale à celle de la sous-séquence I (Pdv_n) à reconstituer. Selon une variante, la première carte courante (resp. carte de référence) est à la résolution des images et chaque carte suivante est à une résolution moitié de la précédente.

[0169] - Plusieurs cartes de caractéristique courantes et/ou de référence sont de même résolution, inférieure à la résolution de la sous-séquence.

[0170] Lors d'une étape **E33**, selon un mode de réalisation, les NFM cartes de l'ensemble de cartes courantes $FMdv_i$ et/ou les NFR cartes de référence $FRdv_j$ sont transformées par le module SE' pour générer des cartes transformées $FMS'v_i, FRS'v_j$ à la résolution des images à décoder. Cette étape est similaire à l'étape E24 qui a été décrite pour le codeur à l'appui de la [Fig.6], et les modes de réalisation s'appliquent. Notamment :

[0171] Selon un mode de réalisation, NFM + NFR cartes $FMS'v_i, FRS'v_j$ sont générées.

[0172] Selon un mode de réalisation, chaque carte $FMdv_i$ est transformée en une carte $FMS'v_i$. Selon un mode de réalisation, chaque carte $FRdv_j$ est transformée en une carte $FRS'v_j$.

- [0173] Selon un mode de réalisation, au moins une carte FMd_{v_i}/FRd_{v_j} est de résolution inférieure à celle des images à décoder et l'opération de transformation comporte un suréchantillonnage pour que la carte transformée comporte le même nombre d'échantillons que l'image d'entrée. Le suréchantillonnage consiste à rajouter des valeurs dans les cartes FMS'_{v_i} et/ou FRS'_{v_j} pour atteindre la résolution de l'image à décoder. Il peut être simple (par réplication du plus proche voisin) ou comporter une interpolation (linéaire, polynomiale, par filtrage, etc.) La transformation peut optionnellement comporter une quantification inverse des valeurs extraites, si nécessaire. Cependant la quantification inverse n'est pas obligatoire. La transformation peut comporter un réseau de neurones de suréchantillonnage.
- [0174] Lors d'une étape **E34**, des valeurs sont extraites par le module XTR' :
- des cartes de caractéristiques courantes FMd_{v_i} (ou éventuellement FMS'_{v_i} transformées, et optionnellement supplémentaires).
 - des cartes de caractéristiques de référence FRd_{v_j} ou éventuellement FRS'_{v_j} transformées.
- [0175] Cette extraction est réalisée en fonction des coordonnées (x_n, y_n) de la sous-séquence d'échantillons Pv_n courants de la sous-séquence courante. Elle peut aussi être réalisée en fonction de la résolution de la carte considérée. Cette étape est similaire à l'étape **E25** qui a été décrite pour le codeur à l'appui de la [Fig.6], et les modes de réalisation s'appliquent. Notamment :
- [0176] Dans un mode de réalisation, Zd_n comporte autant de valeurs que de cartes en entrée.
- [0177] Selon un mode de réalisation, le vecteur Zd_n caractéristique est construit par traitement des valeurs extraites. Le traitement peut comporter une quantification des valeurs extraites ou du vecteur Zd_n constitué, si nécessaire. Le traitement peut comporter d'autres opérations, comme un filtrage, une mise à l'échelle, l'application d'une fonction quelconque, de préférence monotone, etc. Les exemples cités pour le codeur s'appliquent.
- [0178] Dans un mode de réalisation, Zd_n est un J-uplet comportant des valeurs situées aux coordonnées (x_n, y_n) d'un pixel courant Pd_{v_n} , comme il sera illustré à l'appui de la [Fig.8]. Le vecteur Zd_n d'indice n réfère au vecteur caractéristique du pixel ou de la séquence de pixels Pd_{v_n} . Selon un mode de réalisation, $J = NFM + NFR$. Le vecteur Zd_n ($zd_{v_1} \dots zd_{v_{NFM}}, zdr_1 \dots zdr_{NFR}$), comporte deux ensembles de valeurs : l'une (Zd_{v_n}) est extraite des cartes de caractéristiques courantes de la sous-séquence v , et l'autre (Zdr_n) est extraite des cartes de caractéristiques de référence. Optionnellement il peut comporter des valeurs issues des cartes supplémentaires.
- [0179] Dans un mode de réalisation, Zd_n est un J-uplet construit à partir de valeurs prélevées dans les cartes de caractéristiques courantes et de référence (et optionnellement supplémentaires) à des coordonnées qui peuvent être différentes selon les cartes. Par

exemple, si les cartes sont à des résolutions différentes, les coordonnées sont adaptées (par une mise à l'échelle) pour correspondre à la résolution de chaque carte.

[0180] Les échantillons à décoder sont par exemple traités par ordre séquentiel, de $n=1$ à $n=N$.

[0181] Lors d'une étape **E35**, les paramètres Wd_{v_k} du réseau de neurones de synthèse MLP' du sont générés par décodage des valeurs Wc_{v_k} du flux B2. On peut utiliser à cette fin toute technique connue de décodage correspondant à celle de codage qui a été utilisée par le codeur. Le réseau de neurones de synthèse MLP' est semblable au réseau de synthèse MLP, c'est-à-dire qu'il est de même structure et comporte les mêmes paramètres, au codage près, qui peut être réalisé avec ou sans pertes.

[0182] Lors d'une étape **E36**, le vecteur Zd_n caractéristique est traité par le réseau de neurones de synthèse MLP' pour générer en sortie un vecteur représentatif de l'échantillon ou de la sous-séquence d'échantillons Pd_{v_n} à décoder, selon un mode de réalisation un ou plusieurs triplets qui sont injectés dans l'image décodée $I(Pd_{v_n})$ aux positions (x_n, y_n) des composantes couleur (Rd, Gd, Bd). Cette étape est similaire à l'étape **E26** qui a été décrite pour le codeur à l'appui de la [Fig.6].

[0183] Lorsque tous les échantillons du signal ont été traités, la sous-séquence décodée courante correspondant par exemple à l'image $I(Pd_{v_n})$ est disponible.

[0184] Selon un mode de réalisation, le flux B2 est décodé avant le flux B1 (à l'exception des cartes de référence), afin de pouvoir disposer du réseau de neurones de synthèse MLP' avant de commencer à décoder les échantillons.

[0185] La [Fig.8] représente de manière illustrée un procédé de décodage utilisé dans un mode de réalisation de l'invention.

[0186] Dans ce mode de réalisation, les cartes décodées courantes FMd_{v_i} sont au nombre de 3. Dans un mode préféré, elles sont au nombre de 7.

[0187] Dans ce mode de réalisation, la première carte FMd_{v_1} a la même résolution que l'image $I(Pd_{v_n})$ à décoder, et comporte donc $W \times H$ variables, où W représente la largeur de l'image en pixels, et H sa hauteur. La seconde carte FMd_{v_2} est de résolution moitié (dans chaque dimension) de la carte FMd_{v_1} . Chaque carte supplémentaire est de résolution moitié de la carte précédente. Cette structure permet de réduire le nombre de variables des cartes de caractéristiques ce qui facilite le décodage tout en minimisant le coût de codage.

[0188] La carte FMd_{v_2} est suréchantillonnée d'un facteur 2 par le module SE' dans chaque dimension, selon tout procédé de suréchantillonnage à la portée de l'homme du métier. La carte FMd_{v_3} est suréchantillonnée d'un facteur 4 dans chaque dimension.

[0189] Dans ce mode de réalisation, les cartes décodées de référence FRd_j sont au nombre de 3. Dans un mode préféré, elles sont au nombre de 7. Dans ce mode de réalisation, la première carte FRd_1 a la même résolution que l'image, et comporte donc $W \times H$

variables, où W représente la largeur de l'image en pixels, et H sa hauteur. La seconde carte FRd_2 est de résolution moitié (dans chaque dimension) de la carte FRd_1 . Chaque carte supplémentaire est de résolution moitié de la carte précédente. Cette structure permet de réduire le nombre de variables des cartes de caractéristiques ce qui facilite le décodage tout en minimisant le coût de codage.

- [0190] Les trois cartes du groupe de référence sont sélectionnées par le décodeur pour constituer l'ensemble de référence. ($FRdv_1=FRd_1, FRdv_2=FRd_2, FRdv_3=FRd_3$).
- [0191] La carte FRd_2 est suréchantillonnée d'un facteur 2 dans chaque dimension par le module SE', selon tout procédé de suréchantillonnage à la portée de l'homme du métier. La carte FRd_3 est suréchantillonnée d'un facteur 4 dans chaque dimension.
- [0192] Les cartes transformées FMS'_v et FRS'_v sont de même résolution que l'image à décoder, et comportent donc $W \times H$ valeurs, où W représente la largeur de l'image en pixels, et H sa hauteur.
- [0193] Dans ce mode de réalisation, le vecteur Zd_n est un 6-uplet ($zV_3, zV_2, zV_1, zR_3, zR_2, zR_1$) constitué des 3 valeurs des cartes FMS'_v situées aux coordonnées (x_n, y_n) du pixel courant Pdv_n , et des 3 valeurs des cartes FRS'_v situées aux coordonnées (x_n, y_n) du pixel courant Pdv_n . Le vecteur Zd_n est optionnellement déquantifié puis traité par le réseau de neurones de synthèse MLP' pour générer en sortie les triplets (R, G, B) représentatifs des échantillons Pdv_n à décoder. Le triplet (R, G, B) est inséré dans chaque image décodée $I(Pdv_n)$ de la sous-séquence v aux coordonnées (x_n, y_n) dans les composantes couleur (Rd, Gd, Bd).
- [0194] Selon une variante présentée en pointillés, les cartes sont au nombre de 7 : une carte supplémentaire notée FME'_0 a été introduite. Dans ce mode de réalisation, le vecteur Zd_n est un 7-uplet.
- [0195] Cette carte supplémentaire comporte typiquement des données qui peuvent assister le réseau MLP' dans la tâche de reconstruction des images. Ainsi, les cartes ajoutées peuvent être une ou plusieurs parmi la liste suivante, non limitative :
- [0196] – Une carte comportant en chaque point l'abscisse de ce point.
- [0197] – Une carte comportant en chaque point l'ordonnée de ce point.
- [0198] – Une carte comportant en chaque point un codage positionnel.
- [0199] – Une carte représentant une image distincte des images en cours de traitement, susceptible d'apporter des informations sur les images à coder, par exemple une image ou séquence d'images précédemment traitée.
- [0200] – Une carte contenant une donnée représentative de la différence temporelle entre les images de la vidéo en train d'être codée. Par exemple, si la première et la dernière image de la vidéo sont distantes de 8 images, tous les échantillons de la carte contiennent la valeur 8.
- [0201] – Une carte représentant une carte de caractéristiques d'une image distincte des

images en cours de traitement, susceptible d'apporter des informations sur les images à coder, par exemple une carte précédemment traitée.

- [0202] – Une carte comportant la valeur d'un échantillon déjà décodé de la même carte, par exemple l'échantillon précédent dans l'ordre de décodage.

Revendications

[Revendication 1]

Procédé de **codaged**'une séquence d'images (S_v) comportant les étapes suivantes pour au moins une image courante à coder :

- une étape de construction, comportant les sous-étapes suivantes :
- construction (E21, E22) d'un ensemble de cartes de caractéristiques courantes (FM_{v_i}) représentatives de ladite au moins une image courante ($v, I(P_{v_n})$), comprenant au moins une carte de caractéristiques courante ;
- construction (E23) d'un ensemble de cartes de caractéristiques (FR_{v_j}) de référence, comprenant au moins une carte de caractéristiques de référence codée puis décodée ;
- pour au moins un échantillon, dit échantillon courant (P_{v_n}), de ladite au moins une image courante, associé à une position (x_n, y_n) dans ladite au moins une image courante :
- construction (E25) d'un vecteur caractéristique (Z_n) à partir de ladite au moins une carte de caractéristiques de référence (Z_{r_n}) et de ladite au moins une carte de caractéristiques courante (Z_{v_n}), en fonction de ladite position (x_n, y_n) dudit au moins un échantillon courant (P_{v_n}) ;
- traitement (E26) dudit vecteur caractéristique (Z_n) par un réseau de neurones artificiels (MLP), dit réseau de neurones de synthèse, défini par un ensemble de paramètres (W_{v_k}), pour fournir un vecteur (P'_{v_n}) représentatif d'une valeur décodée dudit au moins un échantillon courant ;
- mise à jour (E22, E29) d'au moins une valeur de ladite au moins une carte de caractéristiques courante et/ou d'au moins un paramètre dudit réseau, en fonction d'une mesure de performance de codage ;
- une étape de codage (E24) dudit ensemble de cartes de caractéristiques courantes (FM_{v_i}) ;
- une étape de codage (E28) dudit ensemble de paramètres (W_{v_k}) dudit réseau de neurones de synthèse.

[Revendication 2]

Procédé de codage d'une séquence d'images selon la revendication 1, caractérisé en ce qu'il comporte en outre :

- une étape de construction (FR_j) et codage (FR_{c_j}) d'un groupe de cartes de caractéristiques de référence représentatives de la séquence comprenant ledit ensemble de cartes de caractéristiques de référence (FR_{v_j}).

- [Revendication 3] Procédé de **décodage** d'une séquence d'images (S_v) d'un flux binaire, comportant les étapes suivantes pour au moins une image courante (v) :
- décodage (E35) d'un ensemble de paramètres (W_{dv_k}) représentatifs d'un réseau de neurones (MLP') de synthèse ;
 - décodage (E31) d'un ensemble de cartes de caractéristiques courantes (FM_{dv_j}) comprenant au moins une carte de caractéristiques courante représentative de ladite au moins une image courante;
 - obtention (E32) d'un ensemble de cartes de caractéristiques (FR_{dv_j}) de référence, comprenant au moins une carte de caractéristiques de référence décodée ;
 - pour au moins un échantillon, dit échantillon courant (P_{dv_n}), de ladite au moins une image courante (v) à décoder, associé à une position (x_n, y_n) dans l'image :
 - construction (E34) d'un vecteur caractéristique (Z_{d_n}) à partir de ladite au moins une carte de caractéristiques de référence (Z_{dr_n}) et de ladite au moins une carte de caractéristiques courante (Z_{dv_n}), en fonction de ladite position (x_n, y_n) dudit au moins un échantillon courant, et :
 - traitement (E37) dudit vecteur (Z_{d_n}, ZR_{d_n}) par ledit réseau de neurones de synthèse (MLP') pour fournir un vecteur représentatif d'une valeur décodée dudit au moins un échantillon courant (P_{dv_n}).
- [Revendication 4] Procédé de décodage selon la revendication 3, caractérisé en ce qu'il comporte en outre :
- une étape de décodage (E31) d'un groupe de cartes de caractéristiques (FR_{d_j}) de référence représentatives de la séquence et
 - obtention (E32) dudit ensemble de cartes de caractéristiques de référence (FR_{dv_j}) à partir dudit groupe (FR_{d_j}).
- [Revendication 5] Procédé de codage ou de décodage selon l'une des revendications précédentes, caractérisé en ce que l'ensemble de cartes de référence et l'ensemble de cartes courantes (FR_{dv_j} , FM_{dv_j}) contiennent le même nombre (NFM, NFR) de cartes de caractéristiques.
- [Revendication 6] Procédé de codage ou de décodage selon l'une des revendications précédentes, caractérisé en ce que ladite au moins une carte de référence est à la même résolution que ladite au moins une carte courante.
- [Revendication 7] Procédé de codage ou de décodage selon l'une des revendications précédentes, caractérisé en ce que ladite au moins une carte de

référence est indiquée dans le flux binaire.

[Revendication 8]

Procédé de codage ou de décodage selon l'une des revendications précédentes, caractérisé en ce que ladite au moins une carte de référence est prédéterminée.

[Revendication 9]

Procédé de codage ou de décodage selon l'une des revendications précédentes, caractérisé en ce qu'il comporte en outre une étape de transformation dudit ensemble de cartes de caractéristiques courantes et/ou de référence pour obtenir un ensemble de cartes de caractéristiques transformées à la résolution de l'image courante, le procédé étant caractérisé en ce que ledit vecteur caractéristique est construit à partir desdites cartes de caractéristiques transformées.

[Revendication 10]

Procédé de décodage selon l'une des revendications 3 à 9, caractérisé en ce que ladite au moins une carte courante est décodée en fonction de ladite au moins une carte de référence et/ou d'une carte de caractéristique courante.

[Revendication 11]

Procédé de décodage selon la revendication 10 caractérisé en ce que le décodage est effectué par un réseau de neurones de prédiction.

[Revendication 12]

Dispositif de codage d'une séquence d'images (S_v), ledit dispositif étant configuré pour mettre en œuvre :

- une étape de construction, comportant les sous-étapes suivantes :

- construction (GEN, MAJ) d'un ensemble de cartes de caractéristiques courantes, (FMv_i) représentatives de ladite au moins une image courante (v , $I(Pv_n)$), comprenant au moins une carte de caractéristique courante ;

- construction (INITc, REF) d'un ensemble de cartes de caractéristiques (FRv_j) de référence, comprenant au moins une carte de caractéristiques de référence codée puis décodée ;

- pour au moins un échantillon, dit échantillon courant (Pv_n), de ladite au moins une image courante, associé à une position (x_n, y_n) dans ladite au moins une image courante :

- construction (XTR) d'un vecteur caractéristique (Z_n) à partir de ladite carte de caractéristiques au moins une carte de caractéristiques de référence (Zr_n) et de ladite au moins une carte de caractéristiques courante (ZV_n), en fonction de ladite position (x_n, y_n) dudit au moins un échantillon courant (Pv_n) ;

- traitement (MLP) dudit vecteur caractéristique (Z_n) par un réseau de neurones artificiels, dit réseau de neurones de synthèse, défini par un ensemble de paramètres (Wv_k), pour fournir un vecteur ($P'v_n$)

représentatif d'une valeur décodée dudit au moins un échantillon courant ;

- mise à jour (E22, E29) d'au moins une valeur de ladite au moins une carte de caractéristiques courantes et/ou d'au moins un paramètre dudit réseau, en fonction d'une mesure de performance de codage,
- une étape de codage (E24) dudit ensemble de cartes de caractéristiques courantes (FM_{v_i}) ;
- une étape de codage (E28) dudit ensemble de paramètres (W_{v_k}) dudit réseau de neurones de synthèse.

[Revendication 13]

Dispositif de décodage à partir d'un flux binaire d'une séquence d'images (S_v), ledit dispositif étant configuré pour mettre en œuvre, pour au moins une image courante (v, I(Pd_{v_n})) de la séquence :

- décodage (NND) d'un ensemble de paramètres (Wd_{v_k}) représentatifs d'un réseau de neurones (MLP') de synthèse ;
- décodage (FMD) d'un ensemble de cartes de caractéristiques courantes comprenant au moins une carte de caractéristiques (FMd_{v_i}) courante représentative de ladite au moins une image courante;
- obtention (INITd, REF') d'un ensemble de cartes de caractéristiques (FRd_{v_j}) de référence, comprenant au moins une carte de caractéristiques de référence décodée ;
- pour au moins un échantillon, dit échantillon courant (Pd_{v_n}), de ladite au moins une image courante (v) à décoder, associé à une position (x_n, y_n) dans l'image :
- construction (E34) d'un vecteur caractéristique (Zd_n) à partir de ladite au moins une carte de caractéristiques de référence (Zdr_n) et de ladite au moins une carte de caractéristiques courante (Zdv_n), en fonction de ladite position (x_n, y_n) dudit échantillon courant, et :
- traitement (E37) dudit vecteur (Zd_n) par ledit réseau de neurones de synthèse (MLP') pour fournir un vecteur représentatif d'une valeur décodée dudit au moins un échantillon courant (Pd_{v_n}).

[Revendication 14]

Programme d'ordinateur comportant des instructions pour l'exécution des étapes d'un procédé de codage ou de décodage selon la revendication 1 ou 3 lorsque ledit programme est exécuté par un ordinateur.

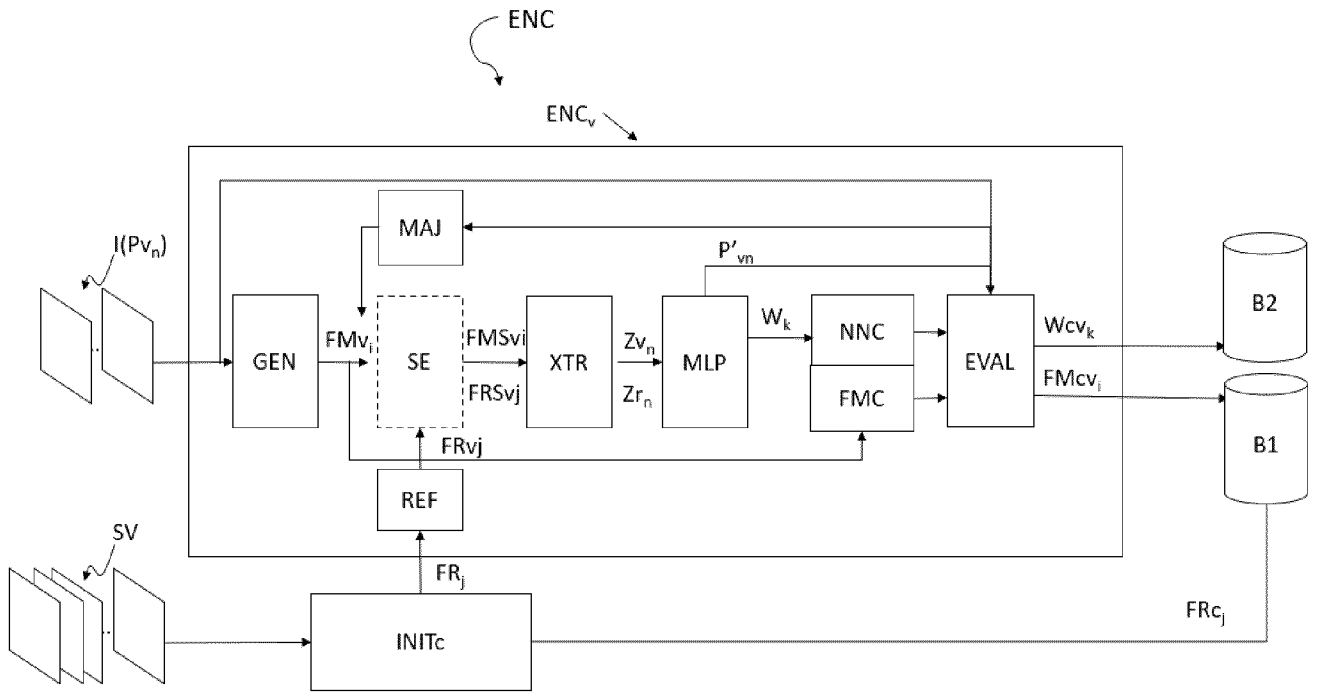


Figure 1

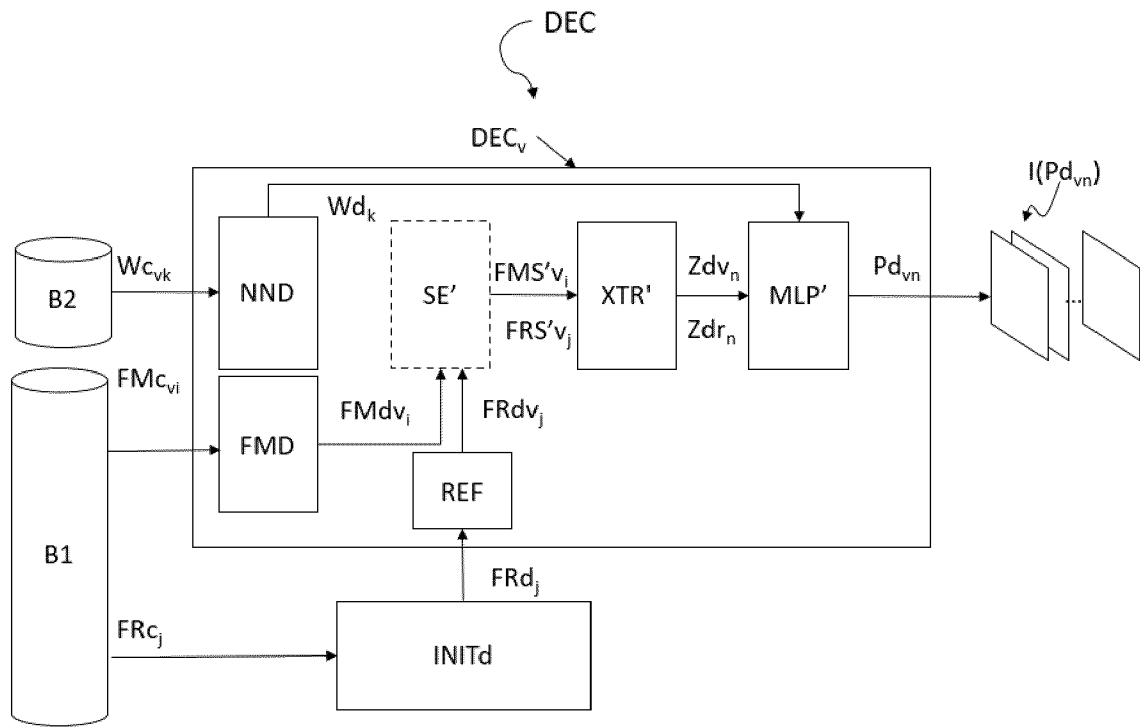


Figure 2

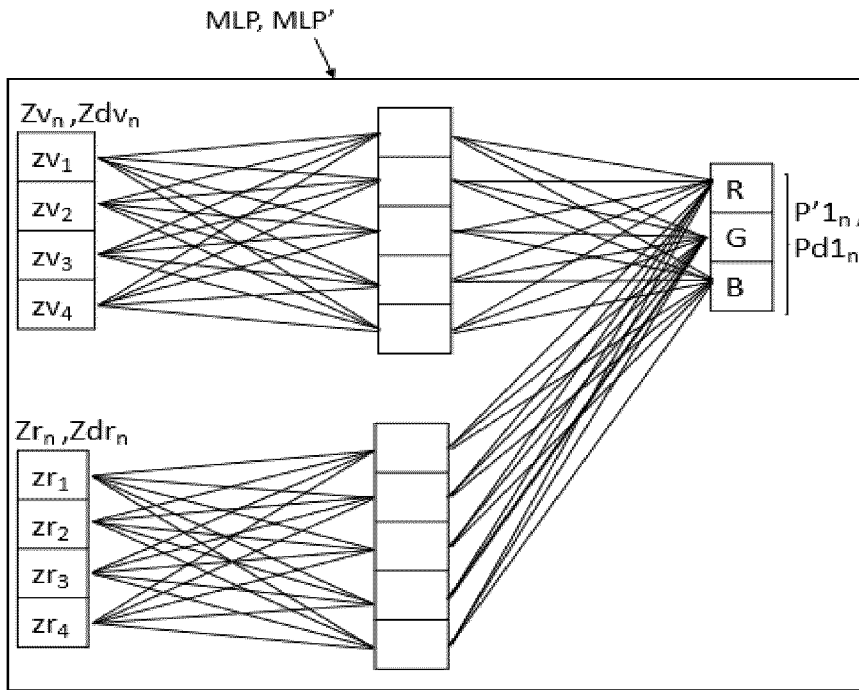


Figure 3

MLP, MLP'

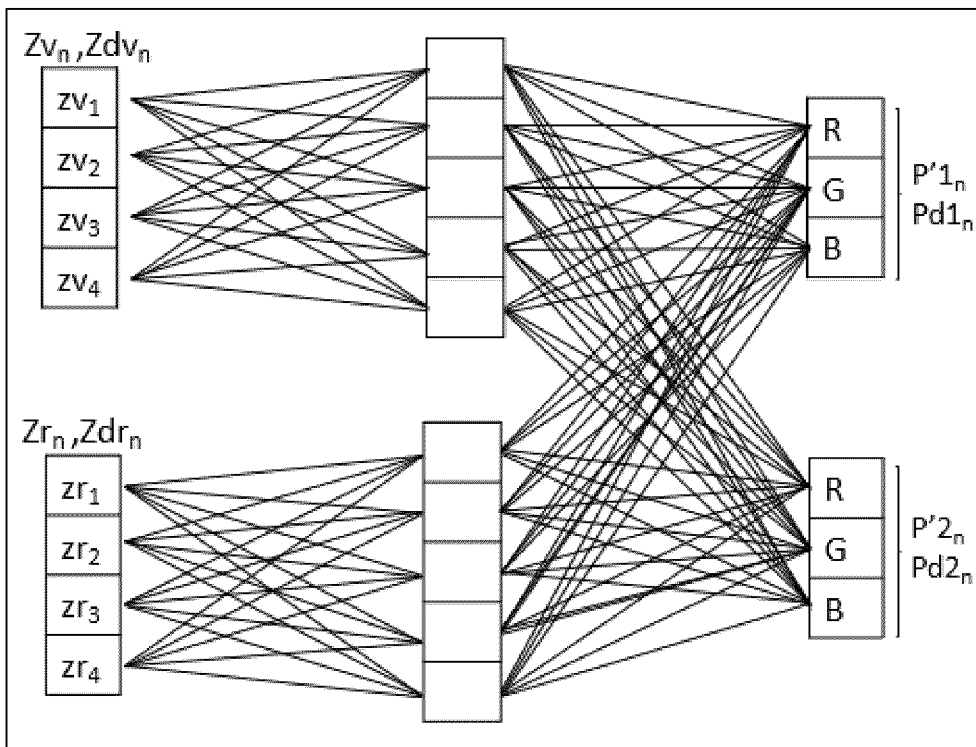


Figure 4

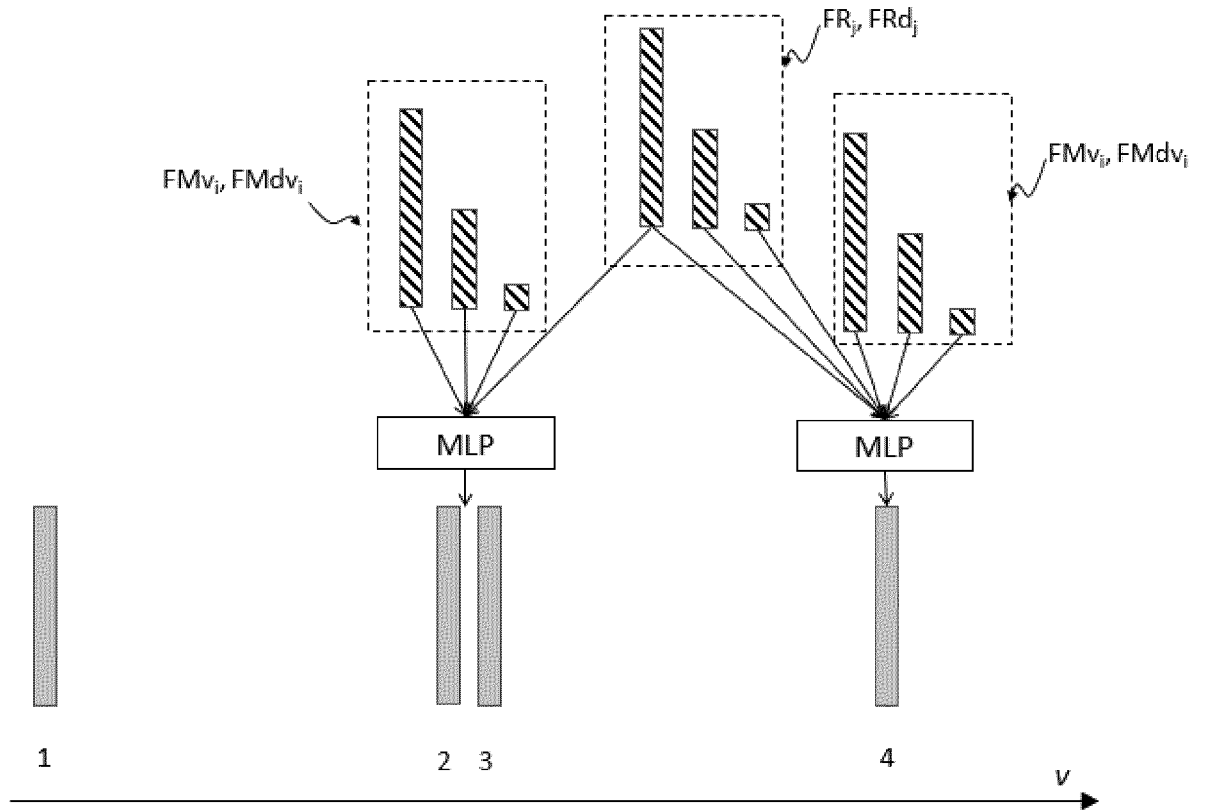


Figure 5

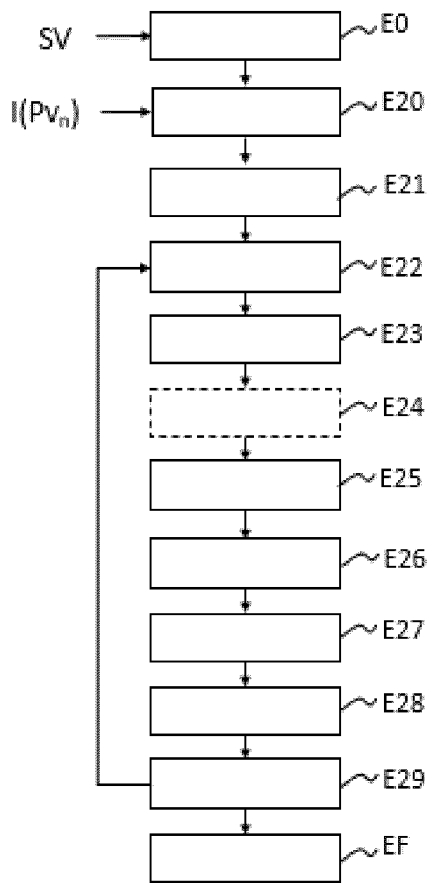


Figure 6

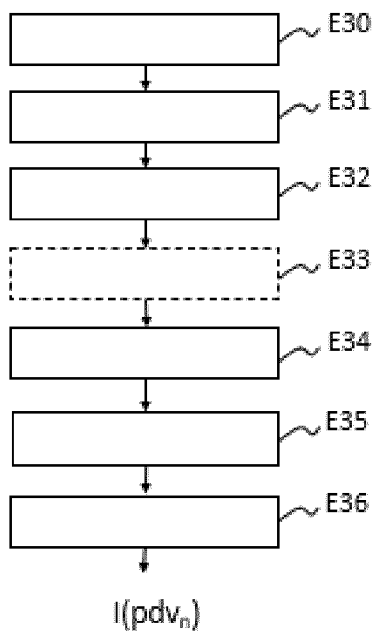


Figure 7

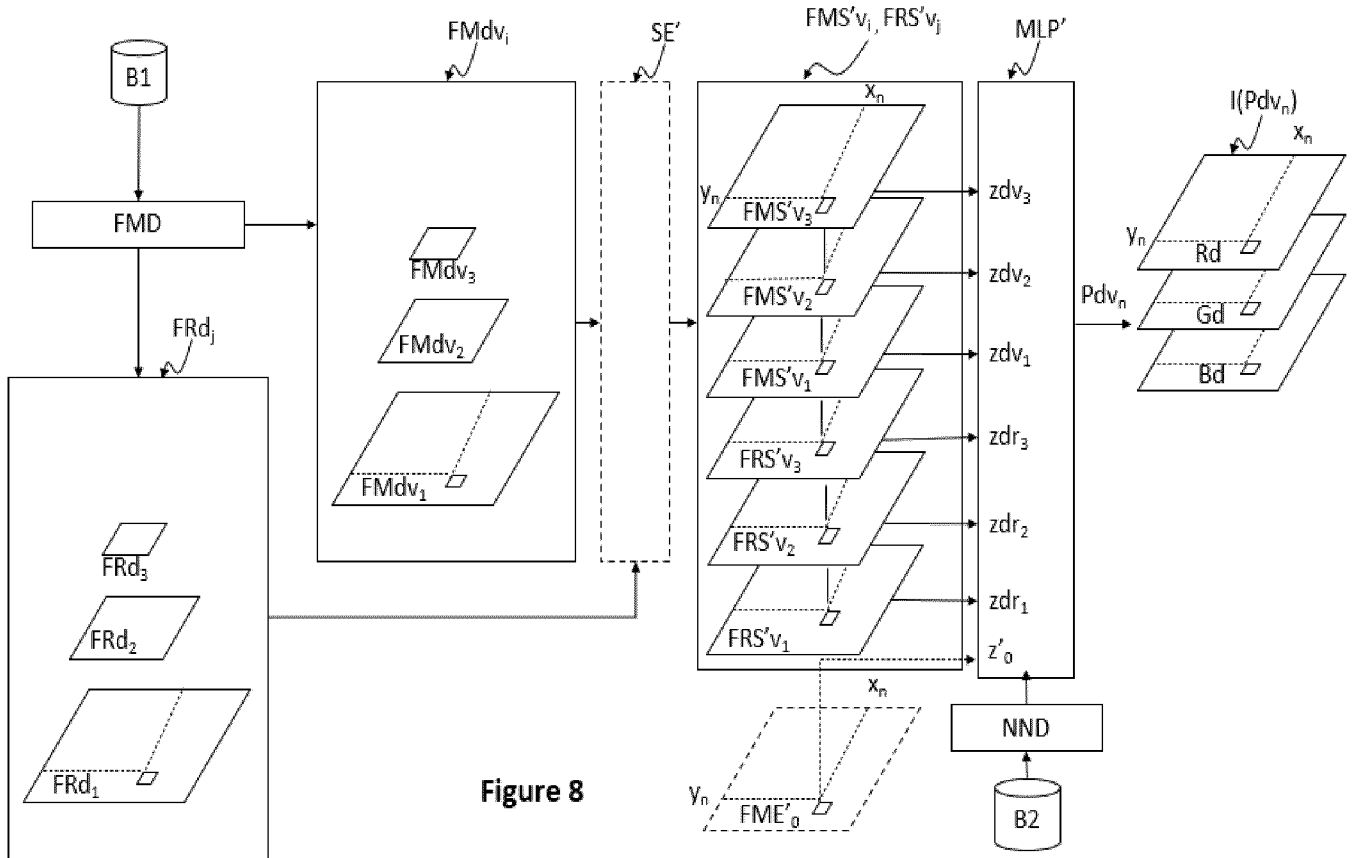


Figure 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2024/062652

A. CLASSIFICATION OF SUBJECT MATTER		
<i>G06T 9/00</i> (2006.01)i; <i>G06N 3/02</i> (2006.01)i; <i>H04N 19/50</i> (2014.01)i; <i>H04N 19/82</i> (2014.01)i		
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) G06T; G06N; 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 practicable, search terms used) EPO-Internal, WPI Data, INSPEC		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Siwei Ma ET AL. "Image and Video Compression With Neural Networks: A Review" 10 April 2019 (2019-04-10), pages 1-16, Retrieved from the Internet: https://arxiv.org/pdf/1904.03567.pdf [retrieved on 2021-01-15] XP055765818 abstract sections I., II., III., IV., V., VI.	1-14
A	CHEN TONG ET AL. "DeepCoder: A deep neural network based video compression" <i>2017 IEEE VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP), IEEE</i> , 10 December 2017 (2017-12-10), pages 1-4 DOI: 10.1109/VCIP.2017.8305033 XP033325713 abstract	1-14
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"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</p> <p>"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</p> <p>"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</p> <p>"&" document member of the same patent family</p>		
Date of the actual completion of the international search 17 June 2024		Date of mailing of the international search report 27 June 2024
Name and mailing address of the ISA/EP European Patent Office p.b. 5818, Patentlaan 2, 2280 HV Rijswijk Netherlands (Kingdom of the) Telephone No. (+31-70)340-2040 Facsimile No. (+31-70)340-3016		Authorized officer Fassnacht, Carola Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REZASOLTANI SHIMA ET AL. "Hyperspectral Image Compression Using Implicit Neural Representations" 2023 20TH CONFERENCE ON ROBOTS AND VISION (CRV), IEEE, 06 June 2023 (2023-06-06), pages 248-255 DOI: 10.1109/CRV60082.2023.00039 XP034412598 the whole document	1-14
A	WO 2023280558 A1 (NOKIA TECHNOLOGIES OY [FI]) 12 January 2023 (2023-01-12) abstract paragraphs [0087], [0089] - [0091], [0094], [0114] - [0117], [0208] - [0210], [0244] figures 5-14, 17	1-14

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/EP2024/062652

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO	2023280558	A1	12 January 2023	CN	117730537	A	19 March 2024
				EP	4367889	A1	15 May 2024
				WO	2023280558	A1	12 January 2023

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°

PCT/EP2024/062652

A. CLASSEMENT DE L'OBJET DE LA DEMANDE

INV. G06T9/00 G06N3/02 H04N19/50 H04N19/82
ADD.

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)

G06T G06N 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, WPI Data, INSPEC

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>Siwei Ma ET AL: "Image and Video Compression With Neural Networks: A Review", , 10 avril 2019 (2019-04-10), pages 1-16, XP055765818, Extrait de l'Internet: URL:https://arxiv.org/pdf/1904.03567.pdf [extrait le 2021-01-15] abrégé sections I., II., III., IV., V., VI. ----- -/-</p>	1-14



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
"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

17 juin 2024

Date d'expédition du présent rapport de recherche internationale

27/06/2024

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

Fassnacht, Carola

C(suite). DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	<p>CHEN TONG ET AL: "DeepCoder: A deep neural network based video compression", 2017 IEEE VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP), IEEE, 10 décembre 2017 (2017-12-10), pages 1-4, XP033325713, DOI: 10.1109/VCIP.2017.8305033 abrégé</p> <p style="text-align: center;">-----</p>	1 - 14
A	<p>REZASOLTANI SHIMA ET AL: "Hyperspectral Image Compression Using Implicit Neural Representations", 2023 20TH CONFERENCE ON ROBOTS AND VISION (CRV), IEEE, 6 juin 2023 (2023-06-06), pages 248-255, XP034412598, DOI: 10.1109/CRV60082.2023.00039 le document en entier</p> <p style="text-align: center;">-----</p>	1 - 14
A	<p>WO 2023/280558 A1 (NOKIA TECHNOLOGIES OY [FI]) 12 janvier 2023 (2023-01-12) abrégé alinéas [0087], [0089] - [0091], [0094], [0114] - [0117], [0208] - [0210], [0244] figures 5-14, 17</p> <p style="text-align: center;">-----</p>	1 - 14

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale n°

PCT/EP2024/062652

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 2023280558 A1	12-01-2023	CN 117730537 A	19-03-2024
		EP 4367889 A1	15-05-2024
		WO 2023280558 A1	12-01-2023
