

**SYSTEME ET PROCEDE INTERACTIF POUR LA TRANSMISSION SUR
UN RESEAU BAS DEBIT D'IMAGES CLEFS SELECTIONNEES DANS UN
FLUX VIDEO**

- 5 L'invention concerne un système et un procédé interactif de transmission d'images et vidéos sur des réseaux à débit contraint. Elle adresse plus précisément les problématiques de transmission d'images pertinentes haute résolution au sein d'une séquence vidéo sur des réseaux bas débit.
- 10 L'invention peut, entre autre, être utilisée dans des applications mettant en œuvre la norme définie en commun par l'ISO MPEG et le groupe video coding de l'ITU-T dite H.264 ou MPEG-4 AVC (advanced video coding) qui est une norme vidéo fournissant une compression plus efficace que les normes vidéo précédentes tout en présentant une complexité de mise en
- 15 œuvre raisonnable et orientée vers les applications réseau. Elle s'applique également dans le cadre de systèmes de transmission d'images utilisant la norme JPEG 2000 et le protocole de communication JPIP (JPEG2000 Interactive Protocol).

Un des problèmes à résoudre réside dans le fait de pouvoir

20 transmettre des séquences vidéo haute résolution sur des réseaux à débit contraint. En effet, une séquence vidéo haute résolution, même après compression via un dispositif de codage source adapté présente un débit utile qui dépasse souvent la capacité du canal de transmission, notamment celle des réseaux sans fils. Une solution consiste à ne sélectionner et ne

25 transmettre au sein de la séquence que certaines images à une résolution inférieure mais se pose alors les problèmes de la sélection d'images pertinentes dans la séquence afin de transmettre la quasi-totalité de l'information utile contenue dans la vidéo. Un autre problème à résoudre concerne la méthode de transmission à mettre en œuvre pour transmettre et

30 recouvrer, côté récepteur, les images haute résolution. De plus, la mise en œuvre d'une forme d'interactivité entre l'opérateur distant et l'émetteur afin

de ne sélectionner qu'une partie du flux vidéo pour la transmission présente des avantages notables afin, par exemple, d'adapter la transmission aux besoins de l'opérateur, et de ne transmettre ainsi que l'information jugée pertinente. Enfin, le souci de complexité de mise en œuvre est un point
5 important à prendre en compte dans la réalisation d'une solution globale et qui satisfait, notamment, les contraintes de temps-réel inhérentes aux applications multimédia interactives.

Dans la suite du texte, l'expression « images pertinentes » ou « images clefs » fera référence à un sous ensemble d'images sélectionnées
10 au sein d'une séquence vidéo et présentant une priorité plus importante du point de vue de l'utilisateur final. Dans le contexte d'une transmission de ladite séquence vidéo sur un réseau bas débit, les images pertinentes sont, par exemple, celles qui présentent une différence de contenu significative entre elles. Dans le contexte d'une compression de ladite séquence vidéo
15 par un codeur vidéo adapté, les images clefs sont également celles qui seront compressées de la façon la plus efficace afin de garantir leur qualité finale en terme de rendu visuel, une fois décompressées. A cet effet, un résumé d'une séquence vidéo correspond à l'ensemble des « images pertinentes » ou « images clefs » de ladite séquence.

20

Sélection d'images pertinentes dans un flux vidéo

La problématique de la sélection d'images pertinentes au sein d'une séquence vidéo est souvent traitée, dans l'art antérieur, par le biais de solutions qui consistent à créer un résumé de ladite séquence en prenant en
25 compte le contenu global de la vidéo. Par exemple, la demande de brevet US2008/0232687 décrit une méthode permettant de sélectionner des images clefs au sein d'une séquence vidéo. Cette méthode permet également la segmentation temporelle de la séquence afin de produire plusieurs scènes. Ce type de procédé n'est pas adapté à la diffusion en temps réel d'un flux
30 vidéo car il nécessite le traitement de l'intégralité de la séquence pour produire l'ensemble des images clefs associées. A contrario, les contraintes

liées à la transmission vidéo nécessitent de traiter les images au fil de l'eau, la méthode de sélection d'images clefs ne bénéficie alors comme point d'entrée que de l'image courante et éventuellement de ses voisines proches temporellement, en particulier des images précédentes si l'on souhaite
5 minimiser le retard de transmission.

Techniques de compression de flux vidéo

Une séquence vidéo comporte de par sa nature même une importante redondance statistique tant dans le domaine temporel que spatial. La volonté
10 d'utiliser toujours plus efficacement la bande passante des média de transmissions sur lesquels transitent ces séquences et les objectifs de réduction du coût de leur stockage ont posé très tôt la question de la compression vidéo. Les techniques classiques de compression vidéo peuvent généralement se diviser en deux étapes. La première vise à réduire
15 la redondance spatiale et pour cela à compresser une image fixe. L'image est tout d'abord divisée en blocs de pixels (de 4x4 ou 8x8 selon les standards MPEG-1/2/4), un passage dans le domaine fréquentiel puis une quantification permet d'approximer ou de supprimer les hautes fréquences auxquelles l'œil est moins sensible, et enfin les données quantifiées sont
20 codées entropiquement. La seconde a pour but de réduire la redondance temporelle. Cette technique permet de prédire une image à partir d'une ou plusieurs autre(s) image(s) de référence(s) précédemment traitée(s) au sein de la même séquence (prédiction de mouvement). Cette technique consiste à chercher dans ces images de références le bloc auquel on souhaite
25 appliquer une prédiction, et seul est conservé un vecteur estimation de mouvement correspondant au déplacement du bloc entre les deux images ainsi qu'une erreur résiduelle permettant de raffiner le rendu visuel.

Granularité temporelle dans un standard vidéo

30 Un flux de données compressées selon une méthode permettant la granularité temporelle, ou en anglo-saxon « a temporally

scalable compressed bit-stream », suit un schéma de codage de type hiérarchique. Cette hiérarchie dans le schéma de codage permet la définition d'ensembles d'images qui sont accessibles par échelon ou résolution temporelle. Le premier échelon, appelé "résolution de base", est la séquence minimale permettant une reconstruction sans dégradation des trames qui la compose. Les autres échelons correspondent à des affinements de cette séquence de base. Généralement les échelons de raffinement ont des fréquences de trames multiples de celle de base, le rapport entre ces fréquences de trames est alors appelé facteur d'échelle. Dans un exemple d'une séquence à 30 trames par secondes suivant un schéma de codage avec une granularité temporelle de facteur d'échelle égale à deux et échelonné en trois niveaux, on obtiendrait un premier niveau de résolution (résolution de base) correspondant à un contenu vidéo à 7,5 trames par secondes. Dans cet exemple, si l'on a accès aux sous ensembles de base ainsi qu'à celui du premier niveau de raffinement, un contenu vidéo à 15 trames par seconde est alors accessible. Si l'on rajoute le dernier niveau de raffinement, un contenu vidéo à la résolution temporelle originale (30 trames par secondes) est accessible. Chacun de ces sous ensembles est supposé correspondre à une compression efficace de l'information qu'elle contient. La figure 1 schématise un exemple de granularité temporelle. Le niveau de base (0) correspond à la résolution temporelle minimale qui est transmise. Dans le cadre d'une transmission vidéo, le flux de code correspondant au niveau de base représente l'information minimum que doit recevoir le destinataire et doit donc être compressée de façon à assurer le moins de pertes possibles lors de la transmission. Typiquement, les images contenues dans cette résolution temporelle de base sont encodées de manière indépendante. Les résolutions temporelles de niveaux supérieurs (niveau de raffinement 1 et 2) peuvent être encodées en mettant en œuvre une prédiction par rapport aux images de la résolution de base (0). Une prédiction est possible entre images appartenant à la résolution de base, par contre les images contenues dans la

résolution temporelle de base ne peuvent être prédites à partir d'une image contenue dans un autre niveau de raffinement.

Standard JPIP

5 Le standard JPIP (JPEG 2000 Interactive Protocol) définit un protocole dédié à la transmission progressive d'images conformes au standard JPEG 2000. Il permet de tirer profit des différents niveaux de granularité proposés par JPEG 2000 (granularité spatiale, granularité en résolution, ou en anglo-saxon « resolution », et en qualité). En effet, suite à
10 une requête formulée par l'opérateur, seules les informations nécessaires pour satisfaire cette requête sont transmises, et cela de façon progressive en terme de qualité. L'utilisation du protocole JPIP combiné au standard JPEG 2000 permet de ne pas retransmettre les informations déjà transmises. Ainsi les ressources en termes de débit transmis et de complexité de traitement
15 des deux cotés de la chaine de transmission sont allégées. De plus, l'envoi des informations étant hiérarchisé, il est possible de visualiser rapidement une partie de l'image à une faible qualité, qualité qui s'accroît de façon progressive au fur et à mesure que de nouvelles informations sont reçues. Ce standard peut être utilisé pour effectuer une transmission interactive avec
20 optimisation du débit dans le cas d'une transmission d'images JPEG2000 mais ne permet pas de mettre en œuvre le même type de procédé pour une transmission vidéo qui serait basée sur un standard différent. En particulier, la sélection d'une image par l'opérateur au sein d'un flux vidéo transmis n'est pas prise en compte par ce protocole.

25

L'art antérieur, tel que décrit précédemment, ne permet pas de résoudre un certain nombre de problèmes, notamment la transmission d'informations haute résolution dans un réseau présentant un débit contraint. L'accès interactif à une image, ou à une zone d'une image, au sein d'un flux
30 vidéo transmis en temps réel n'est pas non plus couvert par l'état de l'art.

Afin de répondre aux limitations de l'art antérieur, l'invention propose une nouvelle approche qui consiste à ne travailler que sur des images sous-résolues et sous-échantillonnées temporellement de façon intelligente afin de réduire au maximum la redondance et de s'adapter à la bande passante requise. La solution proposée permet également l'analyse de cette séquence sous-résolue de façon interactive via des requêtes effectuées par un opérateur distant. La présente invention est compatible des standards suivants. Le standard H.264 est défini par la norme ISO/IEC 14496-10. Le standard JPEG 2000 est défini par la norme ISO/IEC 15444-1. Le standard JPIP est défini par la norme ISO/IEC 15444-9.

A cet effet l'invention a pour objet un procédé de transmission en ligne d'une séquence vidéo haute résolution composée d'une succession de T images, caractérisé en ce qu'il comporte au moins une étape de sélection d'images pertinentes comprenant au moins les étapes suivantes :

- Etape 1 : Découper chaque image à transmettre à l'instant t en un nombre N de zones,
- Etape 2 : Pour chaque zone n déterminée à l'étape 1, calculer une valeur $v_n(t)$ représentative du contenu de ladite zone,
- Etape 3 : Pour chaque image $I(t)$ à transmettre, générer un vecteur $\mathbf{V}_N(t)$ de taille N représentatif du contenu de ladite image et contenant les valeurs obtenues à l'étape 2 pour chacune des zones déterminées à l'étape 1, $\mathbf{V}_N(t) = [v_0(t), \dots, v_{N-1}(t)]^T$, $[\cdot]^T$ étant l'opérateur de transposition d'un vecteur.
- Etape 4 : Calculer un coefficient de corrélation normalisé α entre le vecteur de référence $\mathbf{R}_N(k) = \mathbf{V}_N(t_1) = [r_0(k), \dots, r_{N-1}(k)]^T$ déterminé pour une image précédemment sélectionnée à l'instant t_1 $I^{sel}(k-1) = I(t_1)$ et celui $\mathbf{V}_N(t)$ calculé pour l'image courante $I(t)$ via le calcul suivant :

$$\alpha = \frac{\mathbf{R}_N^T(k-1)\mathbf{V}_N(t)}{\sqrt{\mathbf{R}_N^T(k-1)\mathbf{R}_N(k-1)}\sqrt{\mathbf{V}_N^T(t)\mathbf{V}_N(t)}}, \text{ où } k \text{ est l'indice temporel}$$

d'une image sélectionnée.

- Etape 5 : Prendre une décision sur la sélection (ou pas) pour transmission de l'image courante $I(t)$ en comparant la valeur du coefficient de corrélation normalisé α à un ou plusieurs seuils de valeurs fixées à priori.

Dans une variante de réalisation de l'invention, la valeur $v_n(t)$ calculée à l'étape 2 est obtenue en calculant un des critères dérivés d'un histogramme contenant les valeurs de luminance ou de chrominance des pixels de la zone n de l'image $I(t)$ parmi lesquels, l'énergie, la variance, le coefficient d'amortissement (ou Kurtosis), le coefficient de dissymétrie (ou Skewness) ou le centre de gravité dudit histogramme, ou en effectuant une combinaison linéaire d'au moins deux de ces critères.

Dans une variante de réalisation de l'invention, l'étape de sélection d'images pertinentes comporte en plus les étapes suivantes :

- A partir de chaque couple de vecteurs $(R_N(k-1), V_N(t))$, obtenus pour le couple d'images $(I^{set}(k-1), I(t))$, calculer un second vecteur $\Delta V_N(t) = (\Delta v_0(t), \dots, \Delta v_N(t))$ tel que
- $$\Delta v_n(t) = |v_n(t) - r_n(k-1)|^\beta, \beta \text{ étant un entier positif ou nul donné.}$$
- Calculer la norme $\|\Delta V_N(t)\|$ du vecteur $\Delta V_N(t)$ et diviser ledit vecteur par ladite norme pour obtenir un troisième vecteur
- $$\Delta \dot{V}_N(t) = \frac{\Delta V_N(t)}{\|\Delta V_N(t)\|} = (\Delta \dot{v}_0(t), \dots, \Delta \dot{v}_N(t))$$
- Calculer le critère additionnel suivant afin de déterminer s'il y a répartition uniforme ou localisée des différences entre les deux images considérées :

8

$$\circ \quad DC = (\Delta V_N^{MAX} - \frac{1}{N}) / (1 - \frac{1}{N}), \quad \text{avec}$$

$$\Delta V_N^{MAX} = MAX(\Delta v'_n(t)), 0 \leq n < N,$$

- 5 $\circ$ appliquer le test suivant, si $DC < S_{DC}$, les différences entre les deux images considérées sont réparties uniformément, sinon elles sont localisées sur une zone donnée desdites images.

Dans une variante de réalisation de l'invention, l'étape 5 est mise en œuvre en effectuant un test de comparaison entre le coefficient de corrélation α et un seuil prédéfini S , entier strictement positif, de la façon

10 suivante :

- si $\alpha < S$, l'image courante $I(t)$ est sélectionnée, le vecteur de référence est mis à jour $\mathbf{R}_N(k) = \mathbf{V}_N(t)$ et l'indice d'image sélectionnée est incrémenté $k = k + 1$
- dans le cas contraire, l'image n'est pas retenue comme image
- 15 pertinente.

Dans une variante de réalisation de l'invention, l'étape de sélection d'images pertinentes comporte les étapes supplémentaires suivantes :

- Définir un compteur d'images cpt_I , initialisé à 1 lorsqu'une image pertinente est sélectionnée à l'étape 5 et incrémenté de 1 à
- 20 chaque nouvelle image courante $I(t)$ reçue ainsi qu'une valeur maximum dudit compteur cpt_I_max ,
- Pour chaque image, effectuer le test de comparaison suivant : si $\alpha \geq S$ et $cpt_I \geq cpt_I_max$, sélectionner la dite image comme pertinente.

25 Dans une variante de réalisation de l'invention, l'étape de sélection d'images pertinentes comporte les étapes supplémentaires suivantes :

- Définir un compteur d'images cpt_I , initialisé à 1 lorsqu'une image pertinente est sélectionnée à l'étape 5 et incrémenté de 1 à

chaque nouvelle image courante $I(t)$ reçue et une valeur minimum dudit compteur cpt_I_min ,

- Pour chaque image, effectuer le test de comparaison suivant :
si $\alpha < S$ et $cpt_I < cpt_I_min$, ne pas retenir ladite image comme
pertinente.

5

Dans une variante de réalisation de l'invention, l'étape de sélection d'images pertinentes comporte les étapes supplémentaires suivantes :

- Définir un nombre N_s souhaité d'images sélectionnées par seconde,
- Définir un intervalle de temps de durée égale à T_h pour la mise à jour dudit seuil S .
- A chaque intervalle de temps de durée égale à T_h , calculer le nombre d'images N sélectionnées par le procédé pendant ledit intervalle de temps de durée égale à T_h précédent,
- A chaque intervalle de temps de durée égale à T_h , modifier la valeur du seuil S de la façon suivante :

10

15

Si $N < N_s$, $S = S - \delta S$

Si $N > N_s$, $S = S + \delta S$, où δS est une valeur entière positive incrémentale fixée initialement et/ou ajustée dynamiquement

20

Dans une variante de réalisation de l'invention, l'étape 5 consiste à effectuer un test de comparaison entre le coefficient de corrélation α et un nombre J strictement positif de seuils prédéfinis S_j , $1 \leq j < J$, de la façon suivante :

- Si $0 \leq \alpha < S_1$, l'image courante est sélectionnée avec un degré de priorité maximum,
- Si $S_j \leq \alpha < S_{j+1}$, pour $1 \leq j < J-1$ l'image courante est sélectionnée avec un degré de priorité moyen j ,
- Si $\alpha \geq S_J$, l'image n'est pas retenue comme pertinente, S_j est un entier strictement positif.

25

30

Dans une variante de réalisation de l'invention, l'étape 5 est effectuée de la façon suivante :

- Trier les T images de ladite séquence vidéo par ordre croissant de la valeur de leur coefficient de corrélation α ,
- 5 ➤ Grouper les T images triées en J+1 groupes hiérarchisés en fonction de la valeur de α , chacun desdits groupes d'images comportant le même nombre d'images et correspondant à un degré de priorité moyen j'

Dans une variante de réalisation de l'invention, la séquence vidéo haute résolution est compressée avant transmission via les étapes suivantes :

- Effectuer un sous échantillonnage spatial dudit flux vidéo haute résolution afin d'obtenir une séquence vidéo de résolution inférieure,
- 15 ➤ Appliquer l'étape de sélection d'images pertinentes à la séquence vidéo de résolution inférieure obtenue,
- Effectuer une compression de ladite séquence par un codeur vidéo adapté à générer une structure particulière de groupe d'images dans laquelle les images pertinentes sélectionnées définissent :
 - 20 ○ une résolution temporelle, contenant l'ensemble des images pertinentes sélectionnées dans le cas où l'étape de sélection d'images pertinentes ne délivre qu'un groupe d'images pertinentes ayant toutes le même degré de priorité,
 - 25 ○ plusieurs résolutions temporelles, contenant chacune les images pertinentes sélectionnées avec un degré de priorité différent, dans le cas où l'étape de sélection d'images pertinentes délivre plusieurs groupes d'images pertinentes avec des degrés de priorité différents.

30 Dans une variante de réalisation de l'invention, le codeur vidéo respecte le standard H.264.

Dans une variante de réalisation de l'invention, la séquence vidéo haute résolution à transmettre est sauvegardée dans une unité de stockage et qu'une image ou zone d'image de ladite séquence vidéo est sélectionnée afin d'être transmise ou retransmise selon les étapes suivantes :

- 5 ➤ Effectuer un sur échantillonnage spatial de ladite image afin d'obtenir une image haute résolution,
- Calculer la différence entre l'image sur échantillonnée et celle correspondante préalablement stockée afin d'obtenir un résiduel,
- Compresser ledit résiduel via un codeur image afin de produire un
10 flux compressé comportant au moins des paquets qui contiennent une information sur les hautes fréquences de l'image et d'autres paquets qui contiennent une information sur les basses fréquences de la dite image,
- Transmettre ledit flux compressé à un opérateur distant.

- 15 Dans une variante de réalisation de l'invention, l'image à retransmettre ou l'image associée à ladite zone d'image à retransmettre est comparée à ses images voisines temporellement au sein de ladite séquence vidéo haute résolution afin de déterminer laquelle est la moins floue selon les étapes suivantes :

- 20 ➤ Calculer sur l'image ou la zone d'image sélectionnée un critère représentatif de la netteté de ladite image ou zone d'image, ledit critère de netteté présentant un premier extremum lorsque ladite image ou zone d'image est parfaitement nette et un second extremum opposé au premier lorsque ladite image ou zone
25 d'image est totalement floue,
- Effectuer le même calcul sur les M images voisines temporellement, M étant une valeur prédéfinie,
- Sélectionner l'image la plus nette selon le critère précédent et retransmettre cette image ou la zone d'image associée, à la place
30 de celle initialement sélectionnée.

Dans une variante de réalisation de l'invention, le codeur image est adapté à définir l'ordre d'envoi des paquets composant le flux compressé en fonction, au moins, d'une information de débit disponible pour la transmission et fixée à priori.

5 Dans une variante de réalisation de l'invention, le codeur image respecte le standard JPEG2000.

Dans une variante de réalisation de l'invention, la retransmission d'une image ou zone d'image se fait via une requête de l'opérateur distant.

Dans une variante de réalisation de l'invention, les requêtes et les
10 réponses auxdites requêtes sont mises en œuvre via le standard JPIP.

L'invention a également pour objet un système interactif de transmission de séquences vidéo haute résolution caractérisé en ce qu'il comporte un émetteur et un récepteur comportant des moyens pour exécuter les étapes du procédé décrit précédemment.

15

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

- 20 ➤ La figure 1 une illustration de la granularité temporelle dans un flux vidéo,
- La figure 2 un synoptique du système de transmission interactive d'images et vidéos sur un réseau bas débit selon l'invention,
- 25 ➤ La figure 3 une illustration du procédé de sélection d'images pertinentes selon l'invention,
- La figure 4 une illustration d'une variante de réalisation du procédé de sélection d'images pertinentes selon l'invention,
- La figure 5 une illustration d'un exemple de mise en forme d'un
30 schéma de granularité temporelle avec une contrainte sur les images clefs,

- La figure 6 une illustration d'allocation de débit et de distorsion pour chaque résolution temporelle,
- La figure 7 une illustration d'une région d'intérêt dans une image.

5

La figure 2 représente un schéma du système global selon l'invention. Dans la partie gauche de la figure est illustré le système d'acquisition, compression, stockage et serveur de données. La partie droite de la figure correspond à l'opérateur ou expert distant qui réalise les requêtes vers l'émetteur. Les flèches en pointillés signifient qu'il y a transmission (soit de données vidéos ou images, soit de requêtes utilisateur) à travers un réseau à débit contraint connu.

Une séquence vidéo haute résolution, 1, issue d'un capteur, 10, est tout d'abord sous échantillonnée spatialement, 11, afin de réduire la résolution initiale et indirectement le débit utile de la séquence à transmettre. Parallèlement, la séquence, 1, est également stockée localement dans une unité de stockage 12. Un procédé, selon l'invention, de sélection d'images pertinentes, 13, récupère le flux sous échantillonné, 2, afin de produire un résumé de la séquence vidéo, ceci en temps réel. Ce résumé permet de déterminer quelles sont les images clefs, 3, au sein de la séquence. Ladite séquence d'images clefs, 3, ainsi obtenue est une suite d'images dont le contenu entre chaque image diffère de façon significative. Pour chaque image transmise, le procédé de sélection, 13, détermine si elle est une image clef et fournit cette indication à l'encodeur vidéo 14. Ledit encodeur a donc à son entrée un flux sous échantillonné, 2, et une indication d'image clef, 3, et utilise ces deux sources d'information pour produire deux flux vidéos compressés. Le premier flux, 4, correspond à la résolution temporelle de base ou séquence basse-résolution, qui doit contenir, obligatoirement, les images clefs précédemment déterminées. Le second, 5, correspond au flux vidéo intégral, sous échantillonné spatialement puis encodé. Au moins un de ces deux flux est transmis, à travers un réseau à débit contraint, au décodeur

vidéo, 18, distant qui effectue le décodage et génère une séquence vidéo décompressée qui peut être affichée. Le choix de transmettre l'un ou l'autre des flux disponibles dépend de l'opérateur et de la bande passante disponible. Le flux, 4, correspondant à la résolution temporelle de base sera
5 transmis en priorité. Les deux flux compressés, sont également stockés dans une unité de stockage, 15, avant d'être transmis. Dans le cas d'utilisation d'un codeur vidéo de type H.264, seul le flux prioritaire, 4, est stocké.

Du côté de la réception, un opérateur, 19, peut effectuer plusieurs types de requêtes vers le serveur vidéo. Une première requête, 8, permet à
10 l'opérateur de demander la retransmission d'une image spécifique ou l'ensemble des images clefs à la résolution d'origine (associée au capteur 10). Cette requête peut également s'effectuer sur une partie d'une image, afin d'effectuer un zoom sur un élément précis que l'opérateur souhaite afficher en haute résolution. La requête 8 est donc transmise à une étape 17
15 qui effectue la récupération de l'image, du groupe d'images ou de la partie d'image requise. L'image récupérée est donc décodée à partir du flux stocké dans l'unité de stockage 15 puis est sur-échantillonnée spatialement, 16, pour revenir à la résolution d'origine. La différence, 25, entre l'image (ou la partie d'image) haute résolution d'origine préalablement stockée dans l'unité,
20 12, et l'image (ou la partie d'image) sur-échantillonnée, 6, est calculée afin d'obtenir un résiduel, 9.

Ledit résiduel obtenu est ensuite compressé via un encodeur image, 21, respectant, par exemple, le standard JPEG2000. C'est ce résiduel qui est ensuite transmis à l'opérateur via, par exemple, l'utilisation du protocole
25 bidirectionnel JPIP, 7. En réception, le flux est décodé progressivement via un décodeur image, 22, le résiduel décodé ainsi obtenu est communiqué à un module de restauration d'image, 23, qui à partir de ce résiduel et de la première image basse qualité et basse résolution qu'il a reçu (celle qui lui a permis de faire sa requête pour obtenir des informations complémentaires)
30 fournit une image décompressée complète à l'opérateur. Ce dernier va ainsi

être en mesure de reconstruire de façon progressive la partie de l'image qu'il souhaite analyser et ce jusqu'à sa pleine résolution et qualité maximale.

L'opérateur 19 peut interagir avec le système selon l'invention selon plusieurs scénarios. Par exemple, il souhaite la retransmission en haute
5 résolution d'une image ou d'une zone d'image déjà transmise en basse résolution. L'opérateur 19 peut également requérir la transmission du flux vidéo ou des images haute résolution entre deux images déjà transmises en basse résolution.

Le procédé, selon l'invention, de sélection d'images pertinentes, 13,
10 permet de déterminer, en temps réel et en cours de transmission du flux vidéo, les images clefs de la séquence vidéo. L'ensemble de ces images représente un résumé de l'information de la séquence initiale haute résolution. Dans la suite de la description, les notations suivantes seront utilisées :

- 15 ➤ T est le nombre total d'images de la séquence vidéo traitée, t est l'indice temporel associé à une image de ladite séquence,
- N est le nombre de zones de l'image déterminé par le procédé, n est l'indice d'une zone,
- K est le nombre d'images sélectionnées via le procédé parmi
20 les T images disponibles, k est l'indice temporel associé à une image parmi l'ensemble des images sélectionnées.
- L'image courante de la séquence vidéo traitée est notée $I(t)$

Le procédé de sélection d'images pertinentes présente les étapes
25 suivantes :

Etape 1 : Chaque image $I(t)$ reçue à l'instant t est découpée en un nombre N de zones qui peut varier, typiquement N est par exemple un nombre entier supérieur à 9.

Etape 2 : A chaque zone n est associée une valeur $v_n(t)$ représentative du contenu de cette zone de l'image et obtenue selon un critère précis.

Etape 3 : Un vecteur $V_N(t)$ de taille N représentatif du contenu de l'image $I(t)$ est généré en concaténant les valeurs obtenues à l'étape 2 pour chaque zone de l'image :

$V_N(t) = [v_0(t), \dots, v_{N-1}(t)]^T$ où $[.]^T$ est l'opérateur de transposition. Lorsqu'une image courante d'indice t est sélectionnée par le procédé de sélection d'images pertinentes, on met à jour un vecteur de référence :

$$R_N(k) = V_N(t) = [r_0(k), \dots, r_{N-1}(k)]^T$$

L'image sélectionnée est notée $I^{sel}(k) = I(t)$ et pour chaque nouvelle image sélectionnée par le procédé selon l'invention, k est incrémenté de 1 ($k=k+1$).

Etape 4 : Un coefficient de corrélation normalisé α est calculé entre le vecteur de référence $R_N(k-1)$ déterminé pour l'image précédemment sélectionnée $I^{sel}(k-1)$ et le vecteur $V_N(t)$ représentatif du contenu de l'image courante $I(t)$

$$\alpha = \frac{\mathbf{R}_N^T(k-1) \cdot \mathbf{V}_N(t)}{\sqrt{\mathbf{R}_N^T(k-1) \cdot \mathbf{R}_N(k-1)} \sqrt{\mathbf{V}_N^T(t) \cdot \mathbf{V}_N(t)}}$$

20

ou encore

$$\alpha = \frac{\sum_{n=0}^{N-1} r_n(k-1) \cdot v_n(t)}{\sqrt{\sum_{n=0}^{N-1} r_n(k-1) \cdot r_n(k-1)} \sqrt{\sum_{n=0}^{N-1} v_n(t) \cdot v_n(t)}}$$

Etape 5 : Un test de comparaison est effectué entre le coefficient de corrélation α et un seuil prédéfini S . Si $\alpha < S$, l'image courante $I(t)$ est

sélectionnée, le vecteur de référence est mis à jour $R_N(k) = V_N(t)$ et l'indice k est incrémenté de 1. Dans le cas contraire, l'image n'est pas retenue comme pertinente.

Le procédé tel que mis en œuvre via les étapes 1 à 5 décrites précédemment est illustré à la figure 3. Le procédé est initialisé en sélectionnant la première image $I^{sel}(0) = I(t=0)$ de la séquence puis une nouvelle image $I^{sel}(k) = I(t_1)$ est sélectionnée à chaque fois que son critère α associé est inférieur au seuil S .

Dans une variante de réalisation, on peut définir J seuils S_j ($1 \leq j \leq J$) et $S_j < S_{j+1}$ afin de sélectionner $J+1$ groupes d'images. Lesdits groupes obtenus seront hiérarchisés en fonction de leur degré de priorité. Par exemple, pour $J = 3$, le groupe d'images tel que $0 \leq \alpha < S_1$ est le plus prioritaire, le second groupe d'images dans l'ordre de priorité est celui obtenu pour $S_1 \leq \alpha < S_2$ et ainsi de suite. Une autre possibilité pour obtenir $J+1$ groupes d'images consiste à diviser l'ensemble des images disponibles en $J+1$ groupes comportant chacun un nombre égal d'images, la distribution des images dans chaque groupe se fait par hiérarchisation des images en fonction de la valeur de leur critère α associé. La figure 4 illustre cette variante de réalisation dans le cas où 12 images sont initialement disponibles. Trois groupes d'images sont définis par le procédé de sélection selon l'invention. Dans un cas, les seuils S_1, S_2 et S_3 sont utilisés pour déterminer la répartition des images dans les trois groupes. Le groupe 1 contient les images (représentées par leur index temporel) $\{1,2\}$, le groupe 2 contient les images $\{3,4,5,6,7,8\}$ et le groupe 3 contient les images $\{9,10,11,12\}$. Dans le second cas, les images sont hiérarchisées en fonction de la valeur de leur critère α et sont regroupées en trois groupes contenant autant d'éléments chacun. Le groupe 1 contient alors les images $\{1,2,3,4\}$, le groupe 2 contient les images $\{5,6,7,8\}$ et le groupe 3 contient les images $\{9,10,11,12\}$.

La valeur $v_n(t)$ représentative du contenu d'une zone n d'une image reçue à l'instant t ainsi que le vecteur représentatif de ladite image $V_N(t)$ peuvent être calculés selon différents critères tels que décrits dans la suite de ce paragraphe.

- 5 Un critère possible est un critère basé sur la luminance moyenne de l'image. Dans ce cas, chaque valeur $v_n(t)$ est égale à la valeur moyenne de la luminance calculée sur la zone n de l'image $I(t)$. Ce critère est utilisable mais sa performance est fortement conditionnée au pas de quantification utilisé pour la fonction de distribution de la luminance. Le procédé selon
- 10 l'invention permet d'améliorer les performances de ce critère en modélisant plus finement la distribution de la luminance en utilisant, par exemple, une modélisation à partir de plusieurs fonctions Gaussiennes. Le modèle utilisé pour cela est un modèle de mélanges gaussiens, ou « Gaussian Mixture Model » (GMM) en anglais, connu de l'Homme du métier. L'estimation de ce
- 15 modèle peut se faire avec un algorithme espérance-maximisation connu sous le terme anglo saxon « Expectation-Maximisation » (EM). Les coefficients obtenus via cet algorithme peuvent soit être utilisés directement comme valeurs $v_n(t)$ sans pondération particulière, soit pondérés en utilisant les poids associés à chaque Gaussienne et déterminés par
- 20 l'algorithme EM. Enfin, le nombre de Gaussiennes utilisées peut être fixé a priori ou estimé de façon dynamique par un critère d'information connu. Cependant la méthode GMM étant complexe à mettre en œuvre, l'invention propose de modéliser la distribution de la luminance par un histogramme. Le dit histogramme est calculé de manière classique, en utilisant soit la
- 25 résolution initiale de la luminance, c'est-à-dire le nombre de valeurs que peut prendre la luminance sur une image soit une résolution inférieure en regroupant plusieurs valeurs voisines. Une fois cet histogramme construit, plusieurs critères peuvent en dériver tels que, par exemple, l'énergie, la variance, le centre de gravité, le coefficient d'aplatissement (plus connu sous

le terme Kurtosis) ou le coefficient de dissymétrie (connu sous le terme anglo saxon Skewness). Ces paramètres permettent de caractériser plus ou moins finement la distribution de la luminance.

La valeur $v_n(t)$ représentative d'une zone de l'image peut être diverse. Elle

- 5 dépend de l'application. Elle peut être basée sur la luminance, communément notée Y, sur les chrominances, communément notées Cr et Cb, ou sur une combinaison linéaire de la luminance et des chrominances. Les coefficients de la combinaison linéaire sont déterminés de façon à représenter au mieux le contenu de la zone n de l'image pour une application
- 10 visée. D'une façon similaire, la valeur $v_n(t)$, au lieu d'être basée sur les composantes YCrCb de l'image, peut être basée sur les composantes des différents espaces de couleur connu par l'Homme du Métier (RGB, LUV, Lab,...). D'une manière générale, cette valeur représentative peut également résulter d'une transformation appliquée sur les composantes de l'image.
- 15 Cette transformation a, dans ce cas, comme objectif de mettre en exergue une caractéristique de l'image intéressante pour l'application visée, comme par exemple la texture.

- 20 Enfin, un critère additionnel au vecteur représentatif $V_N(t)$ peut être mis en place afin de déterminer si la différence estimée (via le vecteur $V_N(t)$) entre l'image courante et l'image de référence est répartie uniformément sur l'ensemble de l'image ou localisée sur une zone de cette image uniquement. A partir d'un des critères précédemment proposés, un vecteur représentatif
- 25 $V_N(t)$ est disponible pour l'image courante, ainsi qu'un vecteur de référence $R_N(k-1)$ correspondant à la dernière image sélectionnée. Chaque composante d'un vecteur représente une zone de l'image. Un troisième vecteur $\Delta V_N(t) = (\Delta v_0(t), \dots, \Delta v_N(t))$ peut être calculé, avec

$\Delta v_n(t) = |v_n(t) - r_n(k-1)|^\beta, \beta \geq 0$ la valeur absolue, éventuellement élevée à une puissance β , des différences des composantes des vecteurs. La valeur de la puissance β peut être déterminée par simulation. Le vecteur $\Delta V_N(t)$ ainsi obtenu est ensuite normé. La répartition des différences entre les deux images peut alors s'interpréter en observant l'homogénéité des composantes du vecteur normé : si la répartition des différences est parfaitement uniforme, toutes les composantes du vecteur sont égales à $1/N$ où N est la taille du vecteur (égal au nombre de zones de l'image considérées), et si les différences sont parfaitement localisées sur une seule zone, toutes les composantes du vecteur sont nulles sauf une. Un critère normalisé est alors défini permettant de donner une information sur l'homogénéité ou la localisation des différences entre deux images consécutives. Ledit critère utilise deux valeurs :

$$\begin{aligned}
 & \triangleright \Delta V_N^{MAX} = MAX(\Delta v_n(t)), 0 \leq n < N \\
 & \triangleright DC = (\Delta V_N^{MAX} - \frac{1}{N}) / (1 - \frac{1}{N})
 \end{aligned}$$

S'il y a répartition uniforme des différences entre l'image courante et l'image de référence, $\Delta V_N^{MAX} = \frac{1}{N}$ et $DC = 0$

S'il y a localisation des différences sur une zone donnée,

$$\Delta V_N^{MAX} = 1 \text{ et } DC = 1.$$

Le critère additionnel consiste à comparer la valeur de la variable DC à un seuil prédéfini, S_{DC} compris entre 0 et 1 et dont la valeur est déterminée en fonction du scénario d'utilisation. Si $DC < S_{DC}$, alors on conclue à la répartition uniforme des différences entre l'image courante et l'image de référence, dans le cas contraire, on conclue à la localisation des différences sur une zone donnée.

Ce critère additionnel peut être mis en œuvre notamment dans le cas d'utilisation d'un capteur en position fixe, pour lequel on cherche à

sélectionner une image sur une différence localisée, même si le critère initial n'active pas la sélection.

Dans une variante de réalisation, il est possible de contraindre le système à sélectionner une image même si le coefficient de corrélation α demeure au-dessus du seuil S en utilisant un compteur d'images cpt_I . Ce compteur est incrémenté à chaque image reçue, et est remis à un lorsqu'une image pertinente est sélectionnée. Le procédé consiste à fixer une valeur maximale cpt_I_max pour ledit compteur. Lorsque le compteur atteint cette valeur maximale, l'image courante $I(t)$ est sélectionnée même si le critère défini à l'étape 5 n'est pas vérifié.

D'une façon analogue, il est possible de fixer une valeur minimale cpt_I_min pour le compteur d'images. Si l'image courante remplit le critère de sélection (étape 5) mais que le compteur d'images cpt_I n'a pas dépassé la valeur minimale prédéfinie, l'image courante $I(t)$ n'est pas sélectionnée. Cette variante de réalisation présente notamment l'avantage de répondre à des contraintes de débit plus importantes sur le lien de transmission.

Dans une autre variante de réalisation, le seuil S utilisé pour mettre en œuvre le critère de sélection d'une image pertinente peut être rendu adaptatif. En particulier, le seuil S peut varier au cours du temps en fonction, par exemple, d'une contrainte de débit moyen. L'adaptation du seuil au cours du temps peut être réalisée en utilisant un algorithme de type gradient, connu de l'Homme du métier, qui vise à minimiser la différence entre le nombre d'images N sélectionnées, via l'étape 5, par seconde et le nombre souhaité N_s d'images sélectionnées par seconde.

On définit un horizon de temps T_h au bout duquel une mise à jour du seuil S est effectué. Par exemple, mais pas uniquement, cet horizon de temps T_h est pris égal à une seconde.

A chaque intervalle de temps T_h , un test de comparaison est effectué, et le seuil S est modifié comme suit :

Si $N < N_s$, $S = S - \delta S$

Si $N > N_s$, $S = S + \delta S$, où δS est un incrément fixé initialement qui correspond au pas d'adaptation.

Le pas d'adaptation δS peut également être ajusté dynamiquement au cours du temps.

- 5 De plus, une valeur minimum et une valeur maximum dudit seuil S sont fixées afin d'éviter les problèmes de divergence, si $S < S_{\min}$ alors $S = S_{\min}$, si $S > S_{\max}$ alors $S = S_{\max}$.

La valeur de l'incrément δS peut être adaptée en fonction de la valeur absolue de l'erreur $E = |N - N_s|$. Un exemple simple de fonction est une

- 10 fonction linéaire avec saturation définie par :

$\delta S = \delta S_{\max}$ si $|N - N_s| > E_{\max}$, $\delta S = \delta S_{\min}$ si $|N - N_s| < E_{\min}$,

δS est interpolé linéairement entre $\delta S_{\min}$ et $\delta S_{\max}$, si $E_{\min} \leq |N - N_s| \leq E_{\max}$, où $\delta S_{\max}$, $\delta S_{\min}$, $E_{\max}$ et $E_{\min}$ sont des données d'entrée.

- 15 Les images clefs sélectionnées par le procédé décrit précédemment doivent être transmises avec une priorité maximale. Le codeur vidéo, 14, adapté à mettre en œuvre le procédé selon l'invention, utilise lesdites images clefs définies par l'étape précédente pour générer une structure particulière de groupe d'images, ou en anglais, « Group of Pictures » (GOP). Le fait que
- 20 les images clefs soient définies à la volée au cours de la transmission du flux vidéo impose au codeur vidéo d'adapter les structures des GOP de façon dynamique. Le flux compressé obtenu en sortie du codeur vidéo sera composé de plusieurs résolutions temporelles, la résolution minimale devant contenir impérativement l'ensemble des images clefs et devant également
- 25 être compressée de façon à assurer la meilleure qualité en regard du débit disponible sur le canal de transmission. La résolution temporelle qui comporte les images clefs doit être reçue en priorité par l'opérateur. La figure 5 illustre un exemple de mise en forme d'un schéma de granularité temporelle sous contrainte de priorité sur certaines images préalablement
- 30 identifiées comme pertinentes. La contrainte imposée au codeur vidéo réside dans le fait que chaque image codée appartenant à une résolution

temporelle donnée ne peut pas être prédite à partir d'une image codée appartenant aux résolutions temporelles supérieures.

Dans le mode de réalisation où le module de sélection d'images pertinentes fournit un degré de priorité définissant plusieurs classes d'images clefs, et non plus une seule, le codeur vidéo générera alors plusieurs résolutions temporelles, hiérarchisées en fonction du degré de priorité des images clefs.

Une fois que la structure d'un GOP est définie, le codeur, pour chaque résolution temporelle définit une allocation en débit et en distorsion spécifique qui doit permettre de répondre aux contraintes imposées par la transmission, en termes de débit disponible notamment. Cette allocation permet, pour chaque type d'image reçue (image clef ou image moins prioritaire) de déterminer le type de codage à appliquer, en particulier codage spatial ou codage par prédiction d'une autre trame. La figure 6 illustre schématiquement ce principe.

15

Les requêtes effectuées par l'opérateur pour interagir avec le système selon l'invention peuvent respecter le protocole de communications JPIP. A cet effet, un rappel des fonctionnalités de ce protocole est effectué avant d'introduire l'adaptation qui en est faite dans le cadre de l'invention.

L'architecture typique d'un système JPIP se compose de plusieurs clients communiquant avec un serveur. L'application cliente dispose d'une interface graphique où s'affiche l'image décodée et l'application serveur renvoie des données lorsque le client modifie, par un zoom ou un déplacement, sa zone de visualisation. La communication entre le client et le serveur se compose d'une succession de requêtes et de réponses. La requête définit les paramètres de la zone de visualisation (tels que la résolution, la taille ou la position de la fenêtre de zoom) de façon complètement transparente par rapport à la syntaxe du standard JPEG2000. Une des propriétés du protocole JPIP est de fournir des portions d'images JPEG2000 connues par l'Homme du métier sous le terme régions d'intérêt, en réponse à une requête du client. Trois couples de paramètres permettent de définir une région d'intérêt

30

comme illustré par la figure 7. Le premier couple de paramètres (Fx et Fy) permet de définir la résolution à laquelle on souhaite l'image, ou la partie de l'image. Les deux autres couples de paramètres vont s'appliquer pour cette résolution. (X_offset et Y_offset) sont les paramètres de position de la partie
5 d'image souhaitée au sein de la résolution définie par le couple (Fx,Fy) et (Rx , Ry) définissent sa taille également au sein de la résolution définie par le couple (Fx,Fy).

De façon classique, en réponse à une requête JPIP, le serveur JPEG2000 transmet en priorité les basses fréquences contenues dans l'image puis les
10 hautes fréquences de façon incrémentale. Dans la mise en œuvre du système selon l'invention, le protocole JPIP est appliqué à une image résiduelle, 9, obtenue par les étapes illustrées sur la figure 2. Dans ce cas, les basses fréquences ont déjà été envoyées partiellement ou complètement dans le flux vidéo transmis, 5. Il est donc judicieux d'envoyer en priorité les
15 hautes fréquences de l'image correspondant à la résolution requise par l'opérateur d'autant plus qu'elles contiennent d'avantage d'informations. Cependant, la contrainte de débit imposée au codeur vidéo doit également être prise en compte. En effet plus le débit de la séquence contenant les images clefs est faible plus les hautes fréquences seront détériorées et dans
20 le pire cas, les basses fréquences peuvent également être endommagées. Dans ce cas, il est nécessaire d'envoyer d'abord les résidus des basses fréquences avant d'envoyer les hautes fréquences correspondant à la requête utilisateur. Le module JPIP utilisé dans le cadre de l'invention est donc adapté afin de déterminer l'ordre d'envoi des paquets JPEG2000 en
25 fonction du débit alloué par le codeur vidéo et des paramètres de résolution, qualité et zone spatiale demandés par l'opérateur. L'objectif est de réduire au maximum le temps de transmission, ce qui permettra à l'opérateur de prendre une décision rapide.

30 Dans une variante de réalisation, l'opérateur peut préciser, lors de sa requête, qu'il souhaite effectuer une recherche au voisinage de l'image

sélectionnée pour être retransmise. Cette recherche a pour but de trouver, éventuellement, une image dont le contenu est proche mais qui sera plus pertinente selon un critère de netteté. Dans ce cas, les étapes suivantes sont mises en œuvre lors de la requête de l'opérateur :

- 5 ➤ Un calcul sur la zone sélectionnée (par l'opérateur) de l'image sélectionnée d'un critère représentatif de la netteté de ladite zone est effectué.
- Le même calcul est appliqué sur les mêmes zones des images situées dans un voisinage temporel proche de l'image
- 10 sélectionnée.
- L'image dont la zone sélectionnée est la plus nette selon le critère choisi est sélectionnée et récupérée via l'étape 17 (figure 2). Elle est ensuite traitée par la suite du procédé selon l'invention décrite précédemment et illustré par la figure 2.
- 15 Le critère de netteté utilisé peut être, par exemple, basé sur un principe similaire à celui pouvant être utilisé pour réaliser une fonction d'autofocus au niveau du zoom optique d'un appareil photo ou d'une caméra. Un critère, connu de l'Homme du métier, tel que celui décrit dans l'article suivant : « *Autofocusing Algorithm Selection in Computer Microscopy* », Yu Sun, S.
- 20 Duthaler and B. J. Nelson, *In Proceedings of IEEE IROS (International Conference on Intelligent Robots and Systems)*, pp 70-76, 2005 peut être utilisé. Un tel critère de netteté doit pour être efficace présenter une valeur maximum (respectivement minimum) lorsque l'image ou la zone d'image sur laquelle est appliqué ce critère est parfaitement nette, et à l'opposé il doit
- 25 présenter une valeur décroissante (respectivement croissante) au fur et à mesure que l'image ou la zone d'image devient moins nette.

Le système et le procédé selon l'invention présentent notamment les avantages suivants :

- La séquence haute résolution à transmettre est adaptée aux contraintes du medium de transmission en effectuant une sélection des informations les plus pertinentes pour l'utilisateur final.
- L'opérateur peut, grâce à des requêtes interactives, obtenir des informations temporelles ou spatiales supplémentaires à celles qu'il a reçu initialement.
- Tous les traitements effectués sont compatibles du temps-réel et nécessitent des ressources systèmes faibles, de plus le temps de latence introduit est minimal.
- L'ensemble du système proposé permet de s'adapter aux contraintes de débit, sans pour autant sacrifier la résolution et la qualité des images auxquelles l'opérateur peut avoir accès.

REVENDECATIONS

1. Procédé de transmission en ligne d'une séquence vidéo haute
 5 résolution composée d'une succession de T images, caractérisé en ce
 qu'il comporte au moins une étape de sélection d'images pertinentes
 (13) comprenant au moins les étapes suivantes :
- Etape 1 : Découper chaque image à transmettre à l'instant t en un
 nombre N de zones,
 - 10 ➤ Etape 2 : Pour chaque zone n déterminée à l'étape 1, calculer une
 valeur $v_n(t)$ représentative du contenu de ladite zone,
 - Etape 3 : Pour chaque image $I(t)$ à transmettre, générer un
 vecteur $\mathbf{V}_N(t)$ de taille N représentatif du contenu de ladite image
 et contenant les valeurs obtenues à l'étape 2 pour chacune des
 15 zones déterminées à l'étape 1, $\mathbf{V}_N(t) = [v_0(t), \dots, v_{N-1}(t)]^T$, $[\cdot]^T$ étant
 l'opérateur de transposition d'un vecteur.
 - Etape 4 : Calculer un coefficient de corrélation normalisé α entre le
 vecteur de référence $\mathbf{R}_N(k) = \mathbf{V}_N(t_1) = [r_0(k), \dots, r_{N-1}(k)]^T$ déterminé
 pour une image précédemment sélectionnée à l'instant t_1
 20 $I^{sel}(k-1) = I(t_1)$ et celui $\mathbf{V}_N(t)$ calculé pour l'image courante $I(t)$ via
 le calcul suivant :

$$\alpha = \frac{\mathbf{R}_N^T(k-1) \cdot \mathbf{V}_N(t)}{\sqrt{\mathbf{R}_N^T(k-1) \cdot \mathbf{R}_N(k-1)} \cdot \sqrt{\mathbf{V}_N^T(t) \cdot \mathbf{V}_N(t)}}, \text{ où } k \text{ est l'indice temporel}$$
 d'une image sélectionnée.
 - Etape 5 : Prendre une décision sur la sélection (ou pas) pour
 25 transmission de l'image courante $I(t)$ en comparant la valeur du
 coefficient de corrélation normalisé α à un ou plusieurs seuils de
 valeurs fixées a priori.

2. Procédé selon la revendication 1, caractérisé en ce que la valeur $v_n(t)$ calculée à l'étape 2 est obtenue en calculant un des critères dérivés d'un histogramme contenant les valeurs de luminance ou de chrominance des pixels de la zone n de l'image $I(t)$ parmi lesquels, l'énergie, la variance, le coefficient d'amortissement (ou Kurtosis), le coefficient de dissymétrie (ou Skewness) ou le centre de gravité dudit histogramme, ou en effectuant une combinaison linéaire d'au moins deux de ces critères.
3. Procédé selon l'une des revendications 1 à 2, caractérisé en ce que l'étape de sélection d'images pertinentes (13) comporte en plus les étapes suivantes :
- A partir de chaque couple de vecteurs $(R_N(k-1), V_N(t))$, obtenus pour le couple d'images $(I^{sel}(k-1), I(t))$, calculer un second vecteur $\Delta V_N(t) = (\Delta v_0(t), \dots, \Delta v_N(t))$ tel que $\Delta v_n(t) = |v_n(t) - r_n(k-1)|^\beta$, β étant un entier positif ou nul donné.
 - Calculer la norme $\|\Delta V_N(t)\|$ du vecteur $\Delta V_N(t)$ et diviser ledit vecteur par ladite norme pour obtenir un troisième vecteur $\Delta V'_N(t) = \frac{\Delta V_N(t)}{\|\Delta V_N(t)\|} = (\Delta v'_0(t), \dots, \Delta v'_N(t))$
 - Calculer le critère additionnel suivant afin de déterminer s'il y a répartition uniforme ou localisée des différences entre les deux images considérées :
 - $DC = (\Delta V_N^{MAX} - \frac{1}{N}) / (1 - \frac{1}{N})$, avec $\Delta V_N^{MAX} = MAX(\Delta v'_n(t)), 0 \leq n < N$,
 - appliquer le test suivant, si $DC < S_{DC}$, les différences entre les deux images considérées sont réparties uniformément,

sinon elles sont localisées sur une zone donnée desdites images.

4. Procédé selon l'une des revendications 1 à 3, caractérisé en ce que l'étape 5 est mise en œuvre en effectuant un test de comparaison entre le coefficient de corrélation α et un seuil prédéfini S, entier strictement positif, de la façon suivante :
 - si $\alpha < S$, l'image courante $I(t)$ est sélectionnée, le vecteur de référence est mis à jour $\mathbf{R}_N(k) = \mathbf{V}_N(t)$ et l'indice d'image sélectionnée est incrémenté $k = k + 1$
 - dans le cas contraire, l'image n'est pas retenue comme image pertinente.
5. Procédé selon les revendications 1 à 4, caractérisé en ce que l'étape de sélection d'images pertinentes (13) comporte les étapes supplémentaires suivantes :
 - Définir un compteur d'images cpt_I , initialisé à 1 lorsqu'une image pertinente est sélectionnée à l'étape 5 et incrémenté de 1 à chaque nouvelle image courante $I(t)$ reçue ainsi qu'une valeur maximum dudit compteur cpt_I_max ,
 - Pour chaque image, effectuer le test de comparaison suivant : si $\alpha \geq S$ et $cpt_I \geq cpt_I_max$, sélectionner la dite image comme pertinente.
6. Procédé selon les revendications 1 à 4, caractérisé en ce que l'étape de sélection d'images pertinentes (13) comporte les étapes supplémentaires suivantes :
 - Définir un compteur d'images cpt_I , initialisé à 1 lorsqu'une image pertinente est sélectionnée à l'étape 5 et incrémenté de 1 à

chaque nouvelle image courante $I(t)$ reçue et une valeur minimum dudit compteur cpt_I_min ,

- Pour chaque image, effectuer le test de comparaison suivant :
si $\alpha < S$ et $cpt_I < cpt_I_min$, ne pas retenir ladite image comme pertinente.

5

7. Procédé selon les revendications 1 à 6, caractérisé en ce que l'étape de sélection d'images pertinentes (13) comporte les étapes supplémentaires suivantes :

10

- Définir un nombre N_s souhaité d'images sélectionnées par seconde,
- Définir un intervalle de temps de durée égale à T_h pour la mise à jour dudit seuil S .

15

- A chaque intervalle de temps de durée égale à T_h , calculer le nombre d'images N sélectionnées par le procédé pendant ledit intervalle de temps de durée égale à T_h précédent,
- A chaque intervalle de temps de durée égale à T_h , modifier la valeur du seuil S de la façon suivante :

Si $N < N_s$, $S = S - \delta S$

20

Si $N > N_s$, $S = S + \delta S$, où δS est une valeur entière positive incrémentale fixée initialement et/ou ajustée dynamiquement

25

8. Procédé selon les revendications 1 à 3 caractérisé en ce que l'étape 5 consiste à effectuer un test de comparaison entre le coefficient de corrélation α et un nombre J strictement positif de seuils prédéfinis S_j , $1 \leq j < J$, de la façon suivante :

- Si $0 \leq \alpha < S_1$, l'image courante est sélectionnée avec un degré de priorité maximum,

- Si $S_j \leq \alpha < S_{j+1}$, pour $1 \leq j < J-1$ l'image courante est sélectionnée avec un degré de priorité moyen j ,
- Si $\alpha \geq S_J$, l'image n'est pas retenue comme pertinente, S_j est un entier strictement positif.

5

9. Procédé selon les revendications 1 à 3 caractérisé en ce que l'étape 5 est effectuée de la façon suivante :

- Trier les T images de ladite séquence vidéo par ordre croissant de la valeur de leur coefficient de corrélation α ,
- 10 ➤ Grouper les T images triées en J+1 groupes hiérarchisés en fonction de la valeur de α , chacun desdits groupes d'images comportant le même nombre d'images et correspondant à un degré de priorité moyen j'

15 10. Procédé selon l'une des revendications précédentes caractérisé en ce que la séquence vidéo haute résolution est compressée avant transmission via les étapes suivantes :

- Effectuer un sous échantillonnage spatial (11) dudit flux vidéo haute résolution afin d'obtenir une séquence vidéo de résolution inférieure (2),
- 20 ➤ Appliquer l'étape de sélection d'images pertinentes (13) à la séquence vidéo de résolution inférieure obtenue,
- Effectuer une compression de ladite séquence (2) par un codeur vidéo (14) adapté à générer une structure particulière de groupe d'images dans laquelle les images pertinentes sélectionnées définissent :
- 25 ○ une résolution temporelle, contenant l'ensemble des images pertinentes sélectionnées dans le cas où l'étape de sélection d'images pertinentes (13) ne délivre qu'un groupe d'images pertinentes ayant toutes le même degré de
- 30 priorité,

- plusieurs résolutions temporelles, contenant chacune les images pertinentes sélectionnées avec un degré de priorité différent, dans le cas où l'étape de sélection d'images pertinentes (13) délivre plusieurs groupes d'images pertinentes avec des degrés de priorité différents.

5

11. Procédé selon la revendication 10 caractérisé en ce que le codeur vidéo (14) respecte le standard H.264.

- 10 12. Procédé selon l'une des revendications 10 ou 11 caractérisé en ce que la séquence vidéo haute résolution à transmettre est sauvegardée dans une unité de stockage (12) et qu'une image ou zone d'image de ladite séquence vidéo est sélectionnée afin d'être transmise ou retransmise selon les étapes suivantes :

- 15 ➤ Effectuer un sur échantillonnage spatial (16) de ladite image afin d'obtenir une image haute résolution (6),
- Calculer la différence entre l'image sur échantillonnée (6) et celle correspondante préalablement stockée (12) afin d'obtenir un résiduel (9),
- 20 ➤ Compresser ledit résiduel (9) via un codeur image (21) afin de produire un flux compressé comportant au moins des paquets qui contiennent une information sur les hautes fréquences de l'image (12) et d'autres paquets qui contiennent une information sur les basses fréquences de la dite image,
- 25 ➤ Transmettre ledit flux compressé à un opérateur distant (19).

13. Procédé selon la revendication 12 caractérisé en ce que l'image à retransmettre ou l'image associée à ladite zone d'image à retransmettre est comparée à ses images voisines temporellement au sein de ladite séquence vidéo haute résolution afin de déterminer laquelle est la moins floue selon les étapes suivantes :

30

- Calculer sur l'image ou la zone d'image sélectionnée un critère représentatif de la netteté de ladite image ou zone d'image, ledit critère de netteté présentant un premier extremum lorsque ladite image ou zone d'image est parfaitement nette et un second extremum opposé au premier lorsque ladite image ou zone d'image est totalement floue,
5
 - Effectuer le même calcul sur les M images voisines temporellement, M étant une valeur prédéfinie,
 - Sélectionner l'image la plus nette selon le critère précédent et retransmettre cette image ou la zone d'image associée, à la place de celle initialement sélectionnée.
10
14. Procédé selon l'une des revendications 12 ou 13 caractérisé en ce que le codeur image (21) est adapté à définir l'ordre d'envoi des paquets composant le flux compressé en fonction, au moins, d'une information de débit disponible pour la transmission et fixée à priori (14).
15
15. Procédé selon l'une des revendications 12 à 14 caractérisé en ce que le codeur image (21) respecte le standard JPEG2000.
20
16. Procédé selon l'une des revendications 12 à 15 caractérisé en ce que la retransmission d'une image ou zone d'image se fait via une requête (8) de l'opérateur distant.
25
17. Procédé selon la revendication 16 caractérisé en ce que les requêtes (8) et les réponses auxdites requêtes sont mises en œuvre via le standard JPIP.
- 30 18. Système interactif de transmission de séquences vidéo haute résolution caractérisé en ce qu'il comporte un émetteur et un

récepteur comportant des moyens pour exécuter les étapes du procédé selon l'une des revendications 1 à 17.