

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

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
29 avril 2010 (29.04.2010)

PCT

(10) Numéro de publication internationale
WO 2010/046599 A1

(51) Classification internationale des brevets :
G06F 17/30 (2006.01) **H04N 7/26** (2006.01)
G06K 9/00 (2006.01)

Plechatel (FR). **BERRANI, Sid Ahmed** [DZ/FR]; 13
boulevard de Metz, F-35700 Rennes (FR).

(21) Numéro de la demande internationale :
PCT/FR2009/052019

(74) Mandataire : **FROGER Marie-Hélène**; France Telecom
R&D/PIV/Brevets, 38-40 rue du Général Leclerc,
F-92794 Issy Moulineaux Cedex 9 (FR).

(22) Date de dépôt international :
22 octobre 2009 (22.10.2009)

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

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
0857258 24 octobre 2008 (24.10.2008) FR

(71) Déposant (pour tous les États désignés sauf US) :
FRANCE TELECOM [FR/FR]; 6 place d'Alleray,
F-75015 Paris (FR).

(72) Inventeurs; et

(84) États désignés (sauf indication contraire, pour tout titre
de protection régionale disponible) : ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ,

(75) Inventeurs/Déposants (pour US seulement) :
GENGEMBRE, Nicolas [FR/FR]; Brehil, F-35470

[Suite sur la page suivante]

(54) Title : DETERMINATION OF KEY ELEMENTARY ENTITY IN A MULTIMEDIA CONTENT

(54) Titre : DETERMINATION D'ENTITE ELEMENTAIRE CLE DANS UN CONTENU MULTIMEDIA

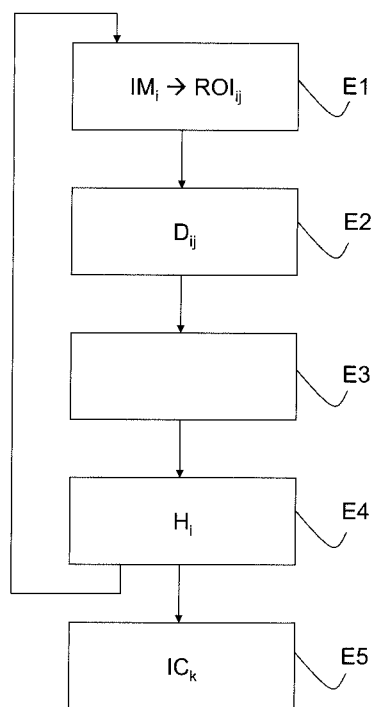


Fig. 1

(57) Abstract : The invention relates to a method of determining a key elementary entity in a multimedia content comprising a succession of elementary entities, characterized in that it comprises the following steps, for at least part of the succession of elementary entities (IM_i) of the multimedia content: determination (E1) of at least one region of interest (ROI_{ij}) per elementary entity of said part, determination (E2) of at least one local descriptor for said at least one region of interest, formation (E4) of a plurality of histograms (H_i) of the local descriptors for, respectively, a plurality of elementary entities of said part, selection (E5) of at least one key elementary entity (IC_k) as a function of a calculated distance between the histograms of said plurality of histograms.

(57) Abrégé : L'invention concerne un procédé de détermination d'entité élémentaire clé dans un contenu multimédia comportant une succession d'entités élémentaires, caractérisé en ce qu'il comporte les étapes suivantes, pour au moins une partie de la succession des entités élémentaires (IM_i) du contenu multimédia : détermination (E1) d'au moins une région d'intérêt (ROI_{ij}) par entité élémentaire de ladite partie, détermination (E2) d'au moins un descripteur local pour ladite au moins une région d'intérêt, formation (E4) d'une pluralité d'histogrammes (H_i) des descripteurs locaux pour, respectivement, une pluralité d'entités élémentaires de ladite partie, sélection (E5) d'au moins une entité élémentaire clé (IC_k) en fonction d'une distance calculée entre les histogrammes de ladite pluralité d'histogrammes.

WO 2010/046599 A1



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

Publiée :

- avec rapport de recherche internationale (Art. 21(3))
- avant l'expiration du délai prévu pour la modification des revendications, sera republiée si des modifications sont reçues (règle 48.2.h))

Déclarations en vertu de la règle 4.17 :

- relative à la qualité d'inventeur (règle 4.17.iv))

Détermination d'entité élémentaire clé dans un contenu multimédia

5

Le domaine de l'invention est celui de la transmission ou de l'échange de contenus multimédia. Plus précisément, l'invention concerne la détermination d'entité élémentaire clé dans un contenu multimédia, en vue notamment de l'analyse et le contrôle de tels contenus, comme par exemple la détection de copies d'un contenu référencé.

La généralisation des communications à haut débit, par exemple offert par l'ADSL, a induit l'émergence de nouveaux services permettant une consommation facilitée du contenu multimédia, comme les services de vidéo à la demande.

Des fournisseurs de contenus proposent ainsi en ligne leur catalogue vidéo. Les contenus multimédia proposés par ces services sont protégés, et soumis par exemple à un paiement pour pouvoir les télécharger.

Par ailleurs, sont apparus des sites d'échange de contenus multimédia. Ces contenus multimédia proviennent des utilisateurs eux-mêmes. Malheureusement, bien qu'une partie des contenus observés sur ces sites d'échanges provienne de contenus réellement créés par les utilisateurs, une autre partie est constituée de contenus illégalement proposés au téléchargement.

Il est donc souhaitable de pouvoir détecter les copies illicites d'un contenu protégé.

Plus précisément, la détection de copies vidéo permet :

- d'identifier les contenus référencés en catalogue, c'est-à-dire référencés dans une base de référence, afin de détecter les copies illicites des contenus de référence ;
- de relever les contenus fortement copiés (dé-doublonnage) afin de

déceler les contenus générateurs d'audience, ou de limiter les tailles de stockage.

Une telle détection doit pouvoir prendre en compte les altérations usuelles que peut subir un contenu dans ce contexte : forte compression, ré-
5 échantillonnage, recadrage, mais aussi incrustation de texte, logos, projections filmées (en anglais « camcording »), etc. En effet, un contenu copié subit généralement des transformations intentionnelles, afin de le rendre difficilement détectable, ainsi que des transformations non-intentionnelles, dues à l'enregistrement du contenu, à son transcodage, ou encore à des contraintes
10 éditoriales lors de sa republication.

Classiquement, la détection de copies de documents multimédia (images, sons, vidéos, etc) consiste à rechercher la présence ou non d'un document requête « suspect » dans une base de documents protégés. Une telle technique repose sur deux aspects essentiels :

- 15 - la description du contenu visuel du document vidéo, i.e. les descripteurs utilisés ;
- la technique d'indexation des descripteurs, i.e. la méthode utilisée pour structurer la base des descripteurs des documents protégés, qui permet d'exécuter efficacement des recherches.

20

Classiquement, les descripteurs des documents sont des vecteurs numériques de grande dimension auxquels est associée une mesure de similarité (souvent une distance). Cette mesure de similarité permet de quantifier la ressemblance entre deux documents en mesurant la proximité
25 entre leurs descripteurs. En particulier, ces descripteurs doivent être robustes aux altérations des documents.

Dans le domaine de l'analyse de contenus vidéo, on utilise principalement une description basée sur les images clés. Cette technique consiste à sélectionner à partir du contenu vidéo un sous-ensemble d'images, appelées images clés, et de décrire ces images clés. On distingue deux
30 approches :

- les approches locales : à partir de chaque image clé, un ensemble

de régions d'intérêt est sélectionné dans l'image. Un descripteur est ensuite calculé pour chaque région d'intérêt ;

- les approches globales : chaque image clé est décrite dans son ensemble en calculant un seul descripteur.

5 La détection de copies vidéo correspond donc à une détection de copies d'images provenant d'une même source vidéo dans un ordre temporel cohérent. Autrement dit, mettre en correspondance un document vidéo copié avec le document vidéo original, revient à mettre en correspondance les descripteurs des images clés des deux documents en vérifiant la cohérence
10 temporelle des images clés mises en correspondance. Cette approche est notamment décrite dans le document « *Statistical similarity search applied to content-based video copy detection* » de A. Joly et al (*icdew*, p. 1285, 21st International Conference on Data Engineering Workshops (ICDEW'05), 2005).

De nombreuses techniques d'analyse d'un contenu multimédia utilisent
15 une description locale, par exemple basée points d'intérêt pour la description d'images, considérant que les descripteurs locaux sont plus robustes que les descripteurs globaux. De plus, l'information décrivant le contenu multimédia est répartie en différentes régions de l'image. Par conséquent, l'altération de certaines de ces régions (par exemple lors de l'incrustation d'un logo dans une
20 image, ou encore lors du rognage de l'image, en anglais « cropping ») n'affecte pas les autres régions qui permettent d'identifier le document.

Ces descripteurs locaux sont très coûteux à la fois en stockage et en temps d'exécution des requêtes. En effet, la description d'une image par plusieurs descripteurs multiplie d'autant la taille de la base des descripteurs par
25 rapport à un schéma de descripteurs globaux d'une part, et augmente le coût d'une recherche d'autre part. En effet, pour retrouver l'image originale d'une image requête, il est nécessaire de rechercher les descripteurs similaires à chacun des descripteurs locaux de l'image requête puis d'intégrer l'ensemble des résultats.

30 Le problème du coût de la recherche affecte également les techniques de description globale, notamment quand la taille de la base des descripteurs

des contenus protégés est volumineuse.

L'analyse de contenu multimédia comporte deux phases. La première est une phase hors ligne dans laquelle des descripteurs associés à des contenus de référence sont déterminés et mémorisés dans une base de données. La

5 de référence sont déterminés et mémorisés dans une base de données. La seconde phase est une phase en ligne d'analyse proprement dite, pendant laquelle un contenu multimédia est comparé aux contenus de référence.

La présente invention concerne plus particulièrement la première phase.

L'article "A geometrical Key-frame selection method exploiting dominant

10 motion estimation in video" de Brigitte Fauvet, Patrici Bouthémy, Patrici Gros et Fabien Spindler, dans Proc of the Int. Conf. On Image and Video Retrieval, Dublin, juillet 2004, propose d'estimer le mouvement dominant dans la vidéo, pour évaluer les changements de cadrage. A chaque image, on estime la part de l'image correspondant à une zone nouvelle par rapport à l'image précédente.

15 Une image clé est définie quand la somme des parties nouvelles devient importante.

Selon cette technique, le changement de contenu visuel non lié à un changement de cadrage, par exemple le mouvement de personnes, le changement d'expression d'une personne, ne peut pas être correctement pris

20 en compte pour la détermination d'image clé.

En outre, cette technique est limitée par le fait qu'elle compare des images consécutives, sans prendre en compte les images antérieures. Ainsi, en cas de mouvements d'aller et retour de la caméra, des images clés redondantes entre elles seront extraites.

L'article "Adaptative key frame extraction using unsupervised clustering" par Yueting Zhuang, Yong Rui, T.S. Huang, S. Mehrotra, dans Proc. of the IEEE Int. Conf. On Image Processing, Chicago, Illinois, USA, octobre 1998, propose d'effectuer un regroupement ("clustering") d'images à partir d'une mesure de similarité définie par une distance entre histogrammes de couleurs.

25 Un groupe ("cluster") est défini par une image moyenne qui agrège le contenu des images du groupe. L'image la plus représentative d'un groupe est utilisée

30 comme image clé.

Les techniques de regroupement ("clustering") impliquent une détermination du nombre de groupes ("cluster") soit par un choix a priori, soit en fonction de paramètres de l'algorithme utilisé. Dans les deux cas, chaque groupe génère une image clé. Le nombre d'images clés dépend donc d'un paramétrage de l'algorithme utilisé, et non du contenu des images traitées.

En outre, cette technique est basée sur des descripteurs globaux, et est donc moins robuste à des modifications de l'image, comme exposé plus haut.

La présente invention a pour but de résoudre les inconvénients de la technique antérieure en fournissant un procédé de détermination d'entité élémentaire clé dans un contenu multimédia comportant une succession d'entités élémentaires, dans lequel, pour au moins une partie de la succession des entités élémentaires du contenu multimédia, il comporte une étape de détermination d'au moins une région d'intérêt par entité élémentaire de ladite partie et une étape de détermination d'au moins un descripteur local pour ladite au moins une région d'intérêt. Le procédé est tel qu'il comporte en outre les étapes suivantes:

- quantification des descripteurs locaux ainsi déterminés;
- formation, par entité élémentaire de ladite partie, d'un histogramme représentatif des valeurs quantifiées des descripteurs locaux de l'entité élémentaire,
- sélection d'une pluralité d'histogrammes constituant des histogrammes retenus, en fonction d'une distance calculée entre les histogrammes des entités élémentaires;
- sélection d'au moins une entité élémentaire clé en fonction des histogrammes retenus.

Grâce à l'invention, les entités élémentaires clés déterminées représentent tout le contenu multimédia, avec le moins de redondance possible. En effet, les entités élémentaires clés sont choisies suffisamment distantes les unes des autres, grâce aux calculs de distances entre histogrammes représentatifs des valeurs quantifiées des descripteurs locaux des entités

élémentaires.

Ainsi, les entités élémentaires clés sont déterminées en comparant les distances des histogrammes entre elles. Les histogrammes sont des agrégations de descripteurs locaux et jouent le rôle de descripteurs globaux. Ils
5 sont plus faciles à traiter que les descripteurs locaux qu'ils représentent.

L'invention est donc particulièrement adaptée à l'indexation de contenus multimédias de référence. Ces contenus de référence sont ensuite par exemple utilisés pour de la détection de copie. La détermination des entités élémentaires
10 clés, utilisent des descripteurs de même type que ceux utilisés ultérieurement lors de la détection de copie. Les entités élémentaires clés sont donc sélectionnées en fonction de l'usage qui en sera fait ensuite.

Il est à noter que l'utilisation de descripteurs locaux dans l'invention rend la détection de copie robuste à des transformations de type découpe ou
15 rognage d'image, ou incrustation de logos, car une portion de l'entité élémentaire, par exemple de l'image, peut suffire à identifier une copie.

Selon une caractéristique préférée, la distance calculée entre les histogrammes appartient à la liste suivante :

- 20 – une intersection d'histogrammes,
- un coefficient de Batthacharyya,
- distances de Minkowski,
- distance euclidienne,
- distance de Manhattan,
- 25 – distance de Kullback-Leibler et ses dérivées,
- Entropie croisée,
- Test du chi-2.

Selon une caractéristique préférée, ladite étape de sélection comporte :
30 - la sélection d'un histogramme modèle,
- le calcul de la distance entre un histogramme candidat et l'histogramme modèle,

- l'ajout de l'histogramme candidat à une liste d'histogrammes retenus en fonction du résultat d'un test réalisé sur la distance calculée.

Selon une caractéristique préférée, la détermination d'au moins une entité élémentaire clé en fonction de la liste d'histogrammes retenus comporte

5 une sélection appartenant à la liste suivante :

- sélection d'au moins une entité élémentaire correspondant à au moins un histogramme retenu,

- sélection d'au moins une entité élémentaire entre deux entités élémentaires correspondant chacune à un histogramme retenu,

10 - sélection d'au moins une entité élémentaire recalculée en fonction d'entités élémentaires entre deux entités élémentaires correspondant chacune à un histogramme retenu.

La première possibilité implique moins de calculs et est donc plus simple à mettre en œuvre. La seconde et la troisième procurent une entité élémentaire

15 clé plus représentative des entités élémentaires entre deux entités élémentaires correspondant chacune à un histogramme retenu.

Selon une caractéristique préférée, la sélection d'au moins une entité élémentaire entre deux entités élémentaires correspondant chacune à un histogramme retenu appartient à la liste suivante :

20 - sélection d'une entité élémentaire sensiblement centrale par rapport à deux entités élémentaires correspondant chacune à un histogramme retenu,

- sélection d'une entité élémentaire qui contient le plus de descripteurs communs aux entités élémentaires entre deux entités élémentaires

25 correspondant chacune à un histogramme retenu.

La première possibilité est plus rapide à mettre en œuvre car elle nécessite moins de calcul, la seconde fournit une entité élémentaire clé plus représentative des entités élémentaires entre deux entités élémentaires correspondant chacune à un histogramme retenu.

30 Selon des caractéristiques alternatives, le test comporte la comparaison de la distance calculée à un seuil et l'ajout de l'histogramme candidat est

effectué si la distance calculée est supérieure au seuil, ou le test comporte une détection de saut dans une succession de distances calculées et l'ajout de l'histogramme candidat est effectué si un saut est détecté.

5 L'invention concerne aussi un procédé d'indexation de contenu multimédia, caractérisé en ce qu'il comporte le procédé de détermination d'entité élémentaire clé dans un contenu multimédia comportant une succession d'entités élémentaires, tel que précédemment décrit.

10 L'invention concerne encore un procédé de détection de contenu multimédia caractérisé en ce qu'il comporte le procédé de détermination d'entité élémentaire clé dans un contenu multimédia comportant une succession d'entités élémentaires, tel que précédemment décrit.

L'invention concerne un dispositif de détermination d'entité élémentaire clé dans un contenu multimédia comportant une succession d'entités
15 élémentaires, dans lequel sont compris des moyens de détermination d'au moins une région d'intérêt par entité élémentaire d'au moins une partie de la succession des entités élémentaires du contenu multimédia, des moyens de détermination d'au moins un descripteur local pour ladite au moins une région d'intérêt; Le dispositif est tel qu'il comporte en outre:
20 - des moyens de quantification des descripteurs locaux déterminés;
- des moyens de formation, par entité élémentaire de ladite partie, d'un histogramme représentatif des valeurs quantifiées des descripteurs locaux de l'entité élémentaire,
- des moyens de sélection d'une pluralité d'histogrammes constituant des
25 histogrammes retenus, en fonction d'une distance calculée entre les histogrammes des entités élémentaires;
- des moyens de sélection d'au moins une entité élémentaire clé en fonction des histogrammes retenus.

30 L'invention concerne aussi un produit programme d'ordinateur téléchargeable depuis un réseau de communication et/ou enregistré sur un support lisible par ordinateur et/ou exécutable par un processeur, caractérisé en

ce qu'il comprend des instructions de code de programme pour la mise en œuvre du procédé selon

Les procédés d'indexation, de détection, le dispositif, le produit programme d'ordinateur présentent des avantages similaires à ceux
5 précédemment exposés.

D'autres caractéristiques et avantages apparaîtront à la lecture de modes de réalisation préférés décrits en référence aux figures dans lesquelles :

- 10 - la figure 1 représente un mode de réalisation de procédé selon l'invention,
- la figure 2 représente des étapes du procédé selon l'invention,
- la figure 3 représente un mode de réalisation de dispositif selon l'invention.

15 Selon un mode de réalisation de l'invention représenté à la **figure 1**, le procédé de détermination d'entité élémentaire de référence dans un contenu multimédia est appliqué à une vidéo comportant une succession d'image. On suppose que la vidéo comporte N images, N étant un entier.

Le procédé comporte des étapes E1 à E5. Les étapes E1 à E4 sont
20 effectuées itérativement pour toutes les images IM_i de la vidéo, pour i entier variant de 1 à N. En variante, elles ne sont parcourues que pour une partie de la succession des images, par exemple pour une image toutes les X images, X étant un entier quelconque.

L'étape E1 est la détermination d'au moins une région d'intérêt ROI_{ij} dans
25 l'image courante IM_i , j étant un entier. La détermination de régions d'intérêt est classique. Une région d'intérêt est un ensemble de pixels saillants, c'est-à-dire présentant un fort contraste avec le voisinage de cette région. Il est possible de déterminer des régions d'intérêt ou des points d'intérêt, ou les deux. Dans le cas de la détermination d'un point d'intérêt, on définit en outre un voisinage de
30 points autour du point d'intérêt, de sorte que cela revient à une région d'intérêt.

L'étape suivante E2 est la détermination d'un descripteur local D_{ij} pour chaque région d'intérêt ROI_{ij} . Un descripteur est un vecteur dont les

composantes sont des réels. On utilise un descripteur de même nature pour toutes les régions d'intérêt de toutes les images considérées.

Par exemple, on utilise un descripteur de couleurs ou de formes, un descripteur différentiel, ou encore un descripteur de la répartition des gradients
5 de la région d'intérêt.

L'étape suivante E3 est une quantification des composantes du descripteur D_{ij} . Par exemple, si le descripteur est un vecteur à 3 composantes réelles comprises entre 0 et 10 et qu'après quantification, chaque composante peut prendre 5 valeurs discrètes distinctes (1..5), le descripteur après
10 quantification prendra $5^3 = 125$ valeurs discrètes différentes. Ainsi, le vecteur (7.21, 4.10, 1.85) sera représenté par le vecteur quantifié (4, 3, 1) et le vecteur (4.82, 9.10, 3.21) par le vecteur quantifié (3, 5, 2).

En variante, les composantes peuvent être en outre normalisées.

Dans les deux cas, le résultat de l'étape E3 est un descripteur quantifié.

15 L'étape suivante E4 est la formation d'un histogramme H_i , pour l'image IM_i . L'histogramme H_i représente les valeurs quantifiées prises par le ou les descripteur(s) de l'image. Plus précisément, l'histogramme représente la fréquence d'apparition dans l'image d'une valeur donnée de descripteur.

Cet histogramme H_i est donc un descripteur global à l'image qui agrège
20 un ensemble de descripteurs locaux décrivant des parties d'image. Cette façon de construire un descripteur d'image permet de rendre celui-ci indépendant du cadrage de l'image, ce qui n'est pas nécessairement le cas pour des descripteurs d'images globaux classiques, notamment pour les descripteurs construits à partir d'un découpage en blocs de l'image. En effet, dans le cas
25 présent, seule la valeur des descripteurs locaux apparaît et non pas la position du point d'intérêt correspondant. Ainsi, deux images partageant une partie de leur contenu auront des histogrammes de descripteurs proches, même si ce contenu commun apparaît à des positions différentes dans les deux images.

Lorsque l'ensemble des étapes E1 à E4 a été effectuées pour toutes les
30 images à prendre en compte, le résultat est une pluralité d'histogrammes correspondant respectivement aux images.

L'étape suivante E5 est effectuée après la détermination de cette pluralité d'histogramme. L'étape E5 est la sélection d'images clés IC_k , en fonction de distances calculées entre les histogrammes précédemment déterminés.

5

L'étape E5 est détaillée en référence à la **figure 2**. La sélection d'images clés comporte des sous-étapes E50 à E55.

L'étape E50 est la sélection d'un premier histogramme modèle HM. L'histogramme de la première image pour laquelle un histogramme a été déterminé est sélectionné.

10

L'étape suivante E51 est le calcul d'une distance $D(HM, H_i)$ entre l'histogramme H_i d'une image courante et l'histogramme modèle HM.

Plusieurs distances peuvent être utilisées, par exemple :

- Intersection d'histogrammes : $D(H_1, H_2) = \sum_{u=1}^N \min(H_1(u), H_2(u))$
- 15 - Coefficient de Batthacharyya : $D(H_1, H_2) = \sum_{u=1}^N \sqrt{H_1(u), H_2(u)}$
- distances de Minkowski : $D_p(H_1, H_2) = \left(\sum_{u=1}^N |H_2(u) - H_1(u)|^p \right)^{1/p}$
- distance euclidienne (qui est une distance de Minkowski pour une valeur de $p=2$),
- distance de Manhattan (distance de Minkowski pour une valeur de $p=1$),
- 20 - distance de Kullback-Leibler et ses dérivées,
- Entropie croisée,
- Test du chi-2.

Il est à noter que l'intersection d'histogramme est bien adaptée à la sélection d'image clé.

25

L'étape suivante E52 est un test sur la distance $D(HM, H_i)$ précédemment calculée.

Une première approche consiste à fixer un seuil puis à comparer la distance $D(HM, H_i)$ avec le seuil. Une approche alternative consiste à utiliser

une méthode de détection de saut (par exemple la méthode de Page-Hinckley) qui permet de détecter un saut dans un signal bruité.

Si la distance entre ces deux histogrammes est inférieure au seuil S , ou si aucun saut n'est détecté dans le signal constitué de la succession des distances, alors l'image courante possède un contenu similaire à celle qui a servi à construire le modèle. L'étape E52 est alors suivie de l'étape E53 à laquelle l'image courante est associée au modèle et on considère un histogramme suivant, tant que tous les histogrammes n'ont pas été traités. Les histogrammes sont considérés les uns après les autres, dans l'ordre temporel des images de la séquences auxquelles ils sont respectivement associés.

L'étape E53 est suivie de l'étape E51 précédemment décrite.

Si à l'étape E52 la distance est supérieure au seuil S , ou si un saut est détecté dans le signal constitué de la succession des distances, alors l'image courante ne peut pas être associée au modèle courant car son contenu est trop différent. L'instant correspondant à cette image (l'instant courant) est alors considéré comme le point de départ d'une nouvelle séquence d'images corrélées, c'est-à-dire partageant une certaine similarité. L'image précédente est quant à elle la dernière image correspondant à la séquence précédente.

L'étape E52 est alors suivie de l'étape E54 à laquelle l'histogramme de l'image courante devient histogramme modèle et on considère un histogramme suivant, tant que tous les histogrammes n'ont pas été traités. Comme à l'étape E53, les histogrammes sont considérés les uns après les autres, dans l'ordre temporel des images de la séquences auxquelles ils sont respectivement associés. L'étape E54 est suivie de l'étape E51 précédemment décrite.

Lorsque tous les histogrammes calculés pour la séquence vidéo ont été traités, les étapes E53 et E54 sont suivies de l'étape E55 à laquelle au moins une image clé IC_k est déterminée.

Pour cela, on considère chaque séquence délimitée par deux images dont l'histogramme a servi d'histogramme modèle. Ces séquences sont des groupes d'images consécutives qui partagent un contenu visuel similaire.

Pour chaque séquence ainsi délimitée, on extrait une image représentative de l'ensemble, dite image clé. Il peut s'agir de l'image centrale,

c'est-à-dire correspondant au temps moyen (éventuellement approché) des images constituant la séquence.

D'autres approches sont également envisageables, notamment celle consistant à analyser le contenu des images de la séquence et à choisir celle dont le contenu est le plus représentatif. Par exemple, on peut pour chaque image calculer la somme des distances entre son histogramme et tous les autres histogrammes de la séquence et choisir comme image clé celle pour laquelle cette grandeur est la plus faible.

Il est aussi possible de choisir comme images clé les images formant les bornes des séquences, c'est-à-dire les images associées aux histogrammes ayant servi de modèles.

Il est encore possible de calculer pour chaque séquence une image en fonction des images de la séquence. Cette image est par exemple une mosaïque d'images construite à partir de l'ensemble des images de la séquence considérée.

Le résultat du traitement de la vidéo selon l'invention est au moins une image clé, de manière générale une pluralité d'images clé. Ces images clés sont mémorisées pour servir ultérieurement à l'analyse de vidéos qui seront comparées à la vidéo traitée selon l'invention.

20

La **figure 3** représente un dispositif de détermination d'entité élémentaire clé dans un contenu multimédia, selon l'invention. Ce dispositif met en œuvre une technique de détermination d'entité élémentaire clé dans un contenu multimédia selon le mode de réalisation particulier décrit ci-dessus.

Un tel dispositif comprend notamment une mémoire 61 constituée d'une mémoire tampon, une unité de traitement 62, équipée par exemple d'un microprocesseur, et pilotée par le programme d'ordinateur 63, mettant en œuvre le procédé selon l'invention.

A l'initialisation, les instructions de code du programme d'ordinateur 63 sont par exemple chargées dans une mémoire RAM avant d'être exécutées par le processeur de l'unité de traitement 62. L'unité de traitement 62 reçoit en entrée un contenu multimédia 20 à traiter. Le microprocesseur de l'unité de

30

traitement 62 met en œuvre les étapes du procédé décrit précédemment, selon les instructions du programme d'ordinateur 63.

Pour cela, le dispositif de détermination d'entité élémentaire clé, dans un contenu multimédia comportant une succession d'entités élémentaires, 5 comporte :

- des moyens de détermination d'au moins une région d'intérêt par entité élémentaire d'au moins une partie de la succession des entités élémentaires du contenu multimédia,
- des moyens de détermination d'au moins un descripteur local pour ladite 10 au moins une région d'intérêt,
- des moyens de formation d'une pluralité d'histogrammes des descripteurs locaux pour, respectivement, une pluralité d'entités élémentaires de ladite partie,
- des moyens de sélection d'au moins une entité élémentaire clé en 15 fonction d'une distance calculée entre les histogrammes de ladite pluralité d'histogrammes.

Ces moyens sont pilotés par le microprocesseur de l'unité de traitement 62.

L'unité de traitement 62 délivre un résultat sous la forme d'au moins une 20 image clé, de manière générale d'une pluralité d'images clés. Ces images clés sont mémorisées pour servir ultérieurement à l'analyse de vidéos qui seront comparées à la vidéo traitée selon l'invention.

REVENDEICATIONS

- 5 **1.** Procédé de détermination d'entité élémentaire clé dans un contenu multimédia comportant une succession d'entités élémentaires, dans lequel, pour au moins une partie de la succession des entités élémentaires du contenu multimédia il comporte une étape de détermination d'au moins une région d'intérêt par entité élémentaire de ladite partie et une étape de détermination
- 10 d'au moins un descripteur local pour ladite au moins une région d'intérêt, caractérisé en ce qu'il comporte en outre les étapes suivantes:
- quantification des descripteurs locaux ainsi déterminés;
 - formation, par entité élémentaire de ladite partie, d'un histogramme représentatif des valeurs quantifiées des descripteurs locaux de l'entité
 - 15 élémentaire,
 - sélection d'une pluralité d'histogrammes constituant des histogrammes retenus, en fonction d'une distance calculée entre les histogrammes des entités élémentaires;
 - sélection d'au moins une entité élémentaire clé en fonction des
 - 20 histogrammes retenus.

- 2.** Procédé selon la revendication 1, caractérisé en ce que la distance calculée entre les histogrammes appartient à la liste suivante :
- une intersection d'histogrammes,
 - 25 - un coefficient de Batthacharyya,
 - distances de Minkowski,
 - distance euclidienne,
 - distance de Manhattan,
 - distance de Kullback-Leibler et ses dérivées,
 - 30 - Entropie croisée,
 - Test du chi-2.

3. Procédé selon la revendication 1 ou 2, caractérisé en ce que ladite étape de sélection des histogrammes retenus comporte :

- la sélection d'un histogramme modèle,
- le calcul de la distance entre un histogramme candidat et
5 l'histogramme modèle,
- l'ajout de l'histogramme candidat à une liste d'histogrammes retenus en fonction du résultat d'un test réalisé sur la distance calculée.

10 4. Procédé selon la revendication 1, caractérisé en ce que la détermination d'au moins une entité élémentaire clé en fonction de la liste d'histogrammes retenus comporte une sélection appartenant à la liste suivante :

- sélection d'au moins une entité élémentaire correspondant à au moins un histogramme retenu,
- 15 - sélection d'au moins une entité élémentaire entre deux entités élémentaires correspondant chacune à un histogramme retenu,
- sélection d'au moins une entité élémentaire recalculée en fonction d'entités élémentaires entre deux entités élémentaires correspondant chacune à un histogramme retenu.

20

5. Procédé selon la revendication 4, caractérisé en ce que la sélection d'au moins une entité élémentaire entre deux entités élémentaires correspondant chacune à un histogramme retenu appartient à la liste suivante :

- sélection d'une entité élémentaire sensiblement centrale par rapport à
25 deux entités élémentaires correspondant chacune à un histogramme retenu,
- sélection d'une entité élémentaire qui contient le plus de descripteurs communs aux entités élémentaires entre deux entités élémentaires correspondant chacune à un histogramme retenu.

30

6. Procédé selon la revendication 3, caractérisé en ce que le test

comporte la comparaison de la distance calculée à un seuil et en ce que l'ajout de l'histogramme candidat est effectué si la distance calculée est supérieure au seuil.

5 **7.** Procédé selon la revendication 3, caractérisé en ce que le test comporte une détection de saut dans une succession de distances calculées et en ce que l'ajout de l'histogramme candidat est effectué si un saut est détecté.

8. Procédé d'indexation de contenu multimédia, caractérisé en ce
10 qu'il comporte le procédé de détermination d'entité élémentaire clé dans un contenu multimédia comportant une succession d'entités élémentaires selon l'une quelconque des revendications 1 à 7.

9. Procédé de détection de contenu multimédia caractérisé en ce
15 qu'il comporte le procédé de détermination d'entité élémentaire clé dans un contenu multimédia comportant une succession d'entités élémentaires selon l'une quelconque des revendications 1 à 7.

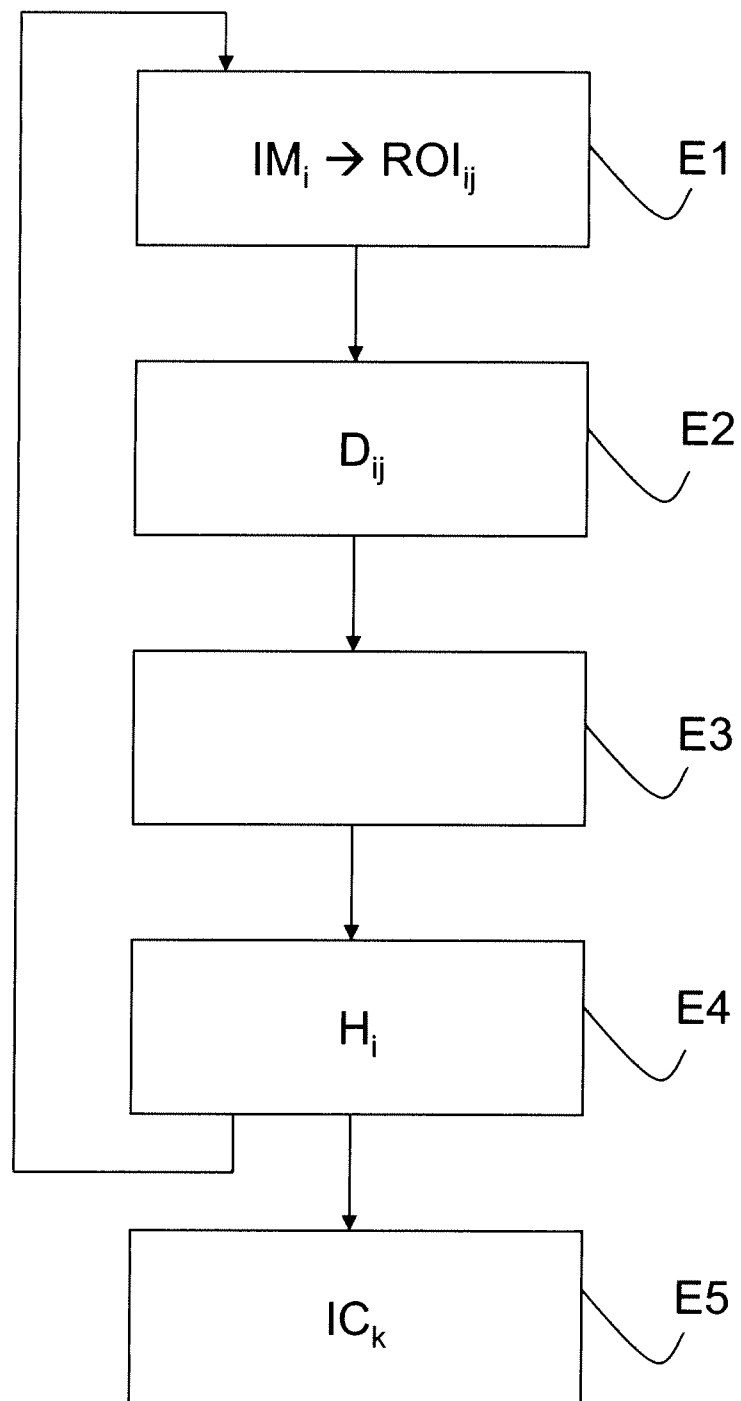
10. Dispositif de détermination d'entité élémentaire clé dans un contenu
20 multimédia comportant une succession d'entités élémentaires, dans lequel sont compris des moyens de détermination d'au moins une région d'intérêt par entité élémentaire d'au moins une partie de la succession des entités élémentaires du contenu multimédia, des moyens de détermination d'au moins un descripteur local pour ladite au moins une
25 région d'intérêt, caractérisé en ce qu'il comporte en outre:
- des moyens de quantification des descripteurs locaux déterminés;
- des moyens de formation, par entité élémentaire de ladite partie, d'un histogramme représentatif des valeurs quantifiées des descripteurs locaux de l'entité élémentaire,
30 - des moyens de sélection d'une pluralité d'histogrammes constituant des histogrammes retenus, en fonction d'une distance calculée entre les

histogrammes des entités élémentaires;

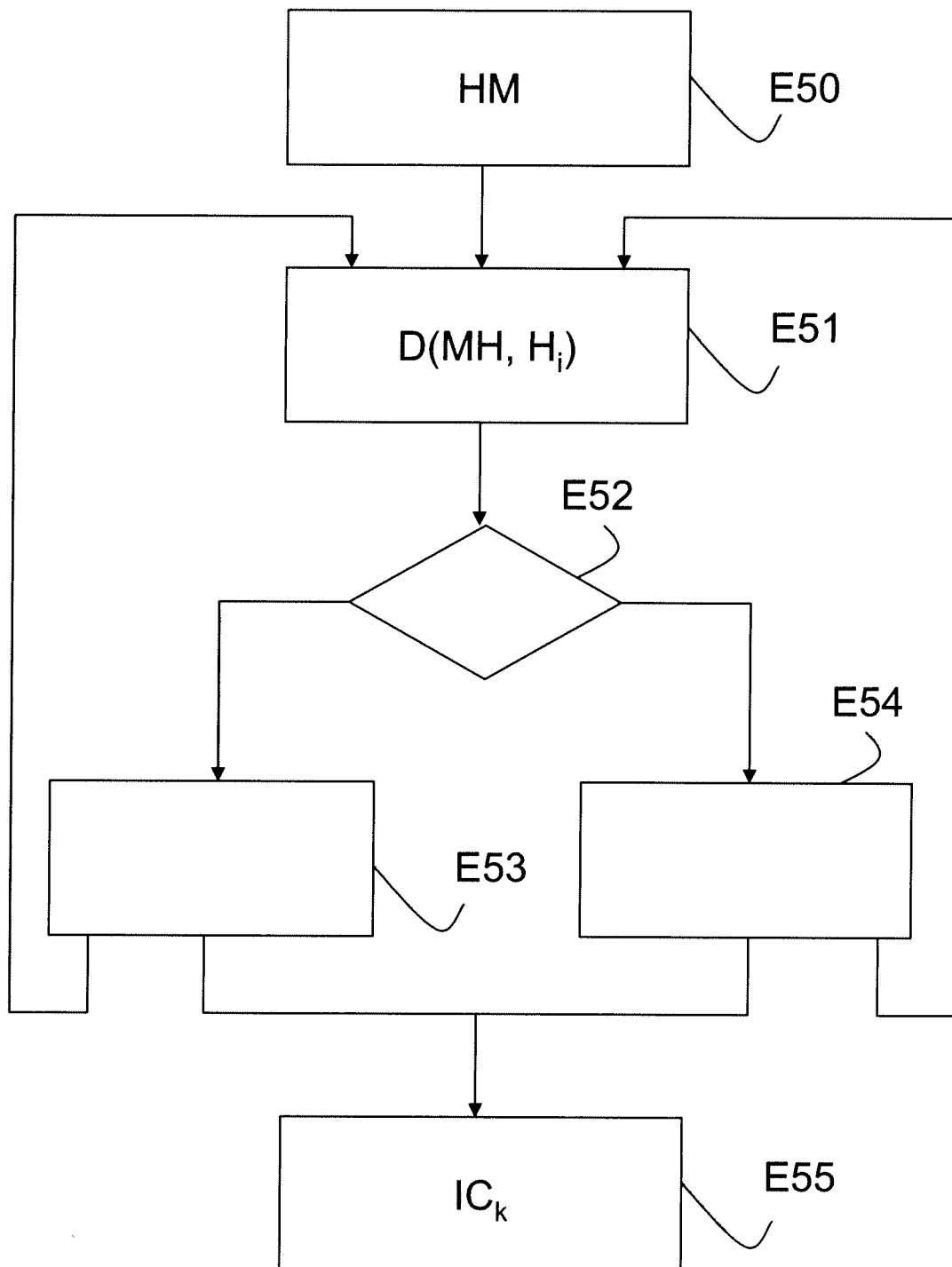
- des moyens de sélection d'au moins une entité élémentaire clé en fonction des histogrammes retenus.

- 5 **11.** Produit programme d'ordinateur téléchargeable depuis un réseau de communication et/ou enregistré sur un support lisible par ordinateur et/ou exécutable par un processeur, caractérisé en ce qu'il comprend des instructions de code de programme pour la mise en œuvre du procédé selon l'une au moins des revendications 1 à 9.

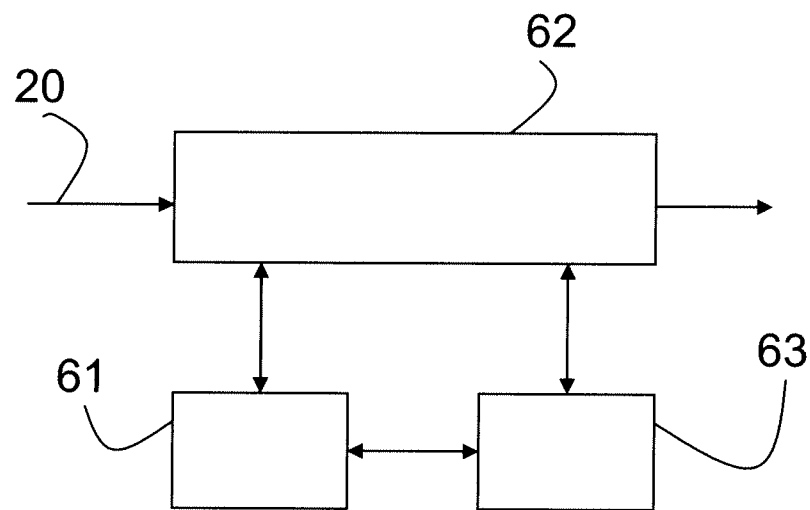
1/3

**Fig. 1**

2/3

**Fig. 2**

3/3

**Fig. 3**

INTERNATIONAL SEARCH REPORT

International application No
PCT/FR2009/052019

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06F17/30 G06K9/00 H04N7/26

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F H04H G06K H04N

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

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

EPO-Internal, INSPEC, WPI Data, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NAGASAKA A ET AL: "AUTOMATIC VIDEO INDEXING AND FULL-VIDEO SEARCH FOR OBJECT APPEARANCES" VISUAL DATABASE SYSTEMS, II. SECOND WORKING CONFERENCE, 30 SEP. - 3 OCT. 1991, BUDAPEST, HU, AMSTERDAM, NL, 1 January 1991 (1991-01-01), pages 113-127, XP000473004 page 114 page 115, paragraph 2.1 page 118 page 119, paragraph 2.4 ----- -/-	1-11

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

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

14 janvier 2010

Date of mailing of the international search report

16/02/2010

Name and mailing address of the ISA/

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

Authorized officer

Laurentowski, A

INTERNATIONAL SEARCH REPORT

International application No

PCT/FR2009/052019

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PARK M-H ET AL: "Shot boundary detection using scale Invariant feature matching" PROCEEDINGS OF SPIE - THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING, VISUAL COMMUNICATIONS AND IMAGE PROCESSING CONFERENCE 2006, vol. 6077, 17 January 2006 (2006-01-17), pages 1-9, XP002563321 ISSN: 0277-786X abstract; figures 2-6 page 2, paragraph 2 - page 6, paragraph 4 -----	1-11
X	KOPRINSKA I ET AL: "Temporal video segmentation: A survey" SIGNAL PROCESSING. IMAGE COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 16, no. 5, 1 January 2001 (2001-01-01), pages 477-500, XP004224651 ISSN: 0923-5965 page 477, right-hand column; figure 1 pages 481-482, paragraph 2.1.3 pages 483-484, paragraph 2.1.3.2 -----	1-11
X	COTSACES C ET AL: "Video shot detection and condensed representation. A review" IEEE SIGNAL PROCESSING MAGAZINE IEEE USA, vol. 23, no. 2, March 2006 (2006-03), pages 28-37, XP002524623 ISSN: 1053-5888 pages 28-29 page 30, right-hand column, line 21 - page 31, right-hand column, last line page 34, right-hand column, last paragraph page 35, left-hand column, last paragraph -----	1-11
A	ANJULAN ARASANATHAN ET AL: "Invariant region descriptors for robust shot segmentation" PROCEEDINGS OF THE SPIE, USA, vol. 6073, 1 January 2006 (2006-01-01), pages 1-12, XP008093691 ISSN: 0277-786X pages 1-7 -----	1-11

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°

PCT/FR2009/052019

A. CLASSEMENT DE L'OBJET DE LA DEMANDE

INV. G06F17/30 G06K9/00 H04N7/26

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

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

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

G06F H04H G06K 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, INSPEC, WPI Data, COMPENDEX

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	NAGASAKA A ET AL: "AUTOMATIC VIDEO INDEXING AND FULL-VIDEO SEARCH FOR OBJECT APPEARANCES" VISUAL DATABASE SYSTEMS, II. SECOND WORKING CONFERENCE, 30 SEP. - 3 OCT. 1991, BUDAPEST, HU, AMSTERDAM, NL, 1 janvier 1991 (1991-01-01), pages 113-127, XP000473004 page 114 page 115, alinéa 2.1 page 118 page 119, alinéa 2.4 ----- -/--	1-11

☒ 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

14 janvier 2010

Date d'expédition du présent rapport de recherche internationale

16/02/2010

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

Laurentowski, A

C(suite). DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	PARK M-H ET AL: "Shot boundary detection using scale Invariant feature matching" PROCEEDINGS OF SPIE - THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING, VISUAL COMMUNICATIONS AND IMAGE PROCESSING CONFERENCE 2006, vol. 6077, 17 janvier 2006 (2006-01-17), pages 1-9, XP002563321 ISSN: 0277-786X abrégé; figures 2-6 page 2, alinéa 2 - page 6, alinéa 4 -----	1-11
X	KOPRINSKA I ET AL: "Temporal video segmentation: A survey" SIGNAL PROCESSING. IMAGE COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 16, no. 5, 1 janvier 2001 (2001-01-01), pages 477-500, XP004224651 ISSN: 0923-5965 page 477, colonne de droite; figure 1 pages 481-482, alinéa 2.1.3 pages 483-484, alinéa 2.1.3.2 -----	1-11
X	COTSACES C ET AL: "Video shot detection and condensed representation. A review" IEEE SIGNAL PROCESSING MAGAZINE IEEE USA, vol. 23, no. 2, mars 2006 (2006-03), pages 28-37, XP002524623 ISSN: 1053-5888 pages 28-29 page 30, colonne de droite, ligne 21 - page 31, colonne de droite, dernière ligne page 34, colonne de droite, dernier alinéa page 35, colonne de gauche, dernier alinéa -----	1-11
A	ANJULAN ARASANATHAN ET AL: "Invariant region descriptors for robust shot segmentation" PROCEEDINGS OF THE SPIE, USA, vol. 6073, 1 janvier 2006 (2006-01-01), pages 1-12, XP008093691 ISSN: 0277-786X pages 1-7 -----	1-11