

19) RÉPUBLIQUE FRANÇAISE  
INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE  
PARIS

11) N° de publication :  
(à n'utiliser que pour les  
commandes de reproduction)

2 863 080

21) N° d'enregistrement national : 03 13907

51) Int Cl<sup>7</sup> : G 06 F 17/30, G 06 T 7/00

12)

## DEMANDE DE BREVET D'INVENTION

A1

22) Date de dépôt : 27.11.03.

30) Priorité :

43) Date de mise à la disposition du public de la demande : 03.06.05 Bulletin 05/22.

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

60) Références à d'autres documents nationaux apparentés :

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

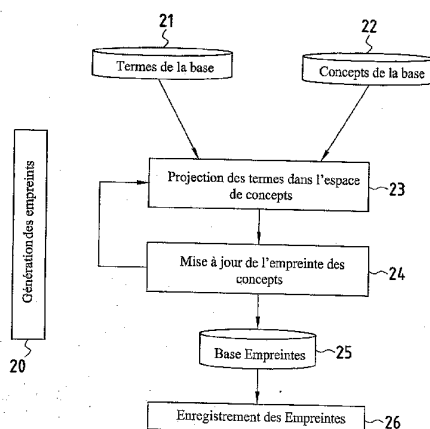
72) Inventeur(s) : ESSAFI HASSANE, GUEZOULI LARBI, SAYAH SALIMA, BEHLOUL ALI, MANDRI-DAKE CLARISSE et ESSAFI LOUAFI.

73) Titulaire(s) :

74) Mandataire(s) : CABINET BEAU DE LOMENIE.

54) PROCÉDE D'INDEXATION ET D'IDENTIFICATION DE DOCUMENTS MULTIMEDIAS.

57) Le procédé d'indexation de documents multimédias comprend les étapes suivantes: (a) identification et extraction pour chaque document de termes  $t_i$  constitués par des vecteurs caractérisant des propriétés du document; (b) stockage des termes  $t_i$  dans une base de termes comprenant  $P$  termes; (c) détermination d'un nombre maximum  $N$  de concepts souhaités regroupant les termes  $t_i$  les plus pertinents; (d) calcul de la matrice  $T$  de distances entre les termes  $t_i$  de la base de termes, (e) décomposition de l'ensemble  $P$  des termes  $t_i$  de la base de termes en  $N$  parties  $P_j$  ( $1 \leq j \leq N$ ) telles que  $P = P_1 \cup P_2 \dots \cup P_j \dots \cup P_N$ , chaque partie  $P_j$  comprenant un ensemble de termes  $t_{ij}$  et étant représentée par un concept  $c_i$ , les termes  $t_i$  étant répartis de telle façon que les termes les plus éloignés se trouvent dans des parties  $P_i, P_m$  distinctes et les termes proches se trouvent dans la même partie  $P_j$ ; (f) structuration du dictionnaire de concepts et (g) construction d'une base d'empreintes constituée de l'ensemble des concepts  $c_i$  représentant les termes  $t_i$  des documents, chaque document étant associé à une empreinte qui lui est propre.



FR 2 863 080 - A1



## Procédé d'indexation et d'identification de documents multimédias

5 La présente invention se rapporte aux procédés d'indexation et d'identification de documents multimédias.

Du point de vue général, l'identification d'un document multimédia comporte deux phases :

- 10 ▪ Une phase dite d'indexation, où l'on cherche à caractériser chaque document d'une base préalablement enregistrée par un nombre fini de paramètres pouvant facilement être stockés et manipulés ultérieurement.
- Une phase dite de recherche, où suite à une requête formulée par l'utilisateur, par exemple l'identification d'une image question, l'on cherche tous les documents multimédias similaires ou répondant à cette requête.

15 Il existe déjà de nombreux procédés d'indexation d'images qui mettent en oeuvre l'extraction des attributs de la forme des objets composants l'image s'ils existent, ainsi que ceux de la texture ou de la couleur de fond de l'image

20 Toutefois, les procédés connus s'appliquent dans des domaines très spécialisés ou impliquent le traitement d'un très grand nombre d'informations qui conduisent à une complexité et une lenteur dans le traitement de ces informations.

25 La présente invention vise à remédier aux inconvénients précités et à fournir un procédé d'indexation et d'identification de documents multimédias d'une application générale qui rationalise le processus de traitement et conduit à des temps de traitement plus courts tout en augmentant la qualité des résultats et leur fiabilité, ce qui permet notamment de procéder à des recherches efficaces par le contenu.

30 Ces buts sont atteints conformément à l'invention, grâce à un procédé d'indexation de documents multimédias, caractérisé en ce qu'il comprend les étapes suivantes :

- 35 (a) identification et extraction pour chaque document de termes  $t_i$  constitués par des vecteurs caractérisant des propriétés du document multimédia à indexer, telles que la forme, la texture, la couleur ou la

- structure d'une image, l'énergie, le taux d'oscillation ou des informations fréquentielles d'un signal audio, ou un groupe de caractères d'un texte ,
- 5 (b) stockage des termes  $t_i$  caractérisant des propriétés du document multimédia dans une base de termes comprenant  $P$  termes,
  - (c) détermination d'un nombre maximum  $N$  de concepts souhaités regroupant les termes  $t_i$  les plus pertinents,  $N$  étant un nombre entier inférieur à  $P$ , et chaque concept  $c_i$  étant prévu pour regrouper tous les termes voisins du point de vue de leurs caractéristiques,
  - 10 (d) calcul de la matrice  $T$  de distances entre les termes  $t_i$  de la base de termes,
  - (e) décomposition de l'ensemble  $P$  des termes  $t_i$  de la base de termes en  $N$  parties  $P_j$  ( $1 \leq j \leq N$ ) telles que  $P = P_1 \cup P_2 \dots \cup P_j \dots \cup P_N$ , chaque partie  $P_j$  contenant un ensemble de termes  $t_{ij}$  et étant représentée par un concept  $c_j$ , les termes  $t_i$  étant répartis de telle façon que les termes les plus éloignés se trouvent dans des parties  $P_l, P_m$  distinctes et les termes proches se trouvent dans la même partie  $P_l$ ,
  - 15 (f) structuration du dictionnaire de concepts de manière à constituer un arbre binaire où les feuilles contiennent les concepts  $c_i$  du dictionnaire et les nœuds de l'arbre contiennent les informations nécessaires à la scrutation de l'arbre lors d'une phase d'identification d'un document par comparaison avec les documents précédemment indexés, et
  - 20 (g) construction d'une base d'empreintes constituée de l'ensemble des concepts  $c_i$  représentant les termes  $t_i$  des documents à indexer, chaque document étant associé à une empreinte qui lui est propre.
  - 25

De façon plus particulière, on associe à chaque concept  $c_i$  de la base d'empreintes un ensemble d'informations comprenant le nombre  $NbT$  de termes dans les documents où le concept  $c_i$  est présent.

- 30

Selon un aspect particulier et l'invention, pour chaque document où un concept  $c_i$  est présent, on enregistre une empreinte du concept  $c_i$  dans le document, cette empreinte comprenant la fréquence d'occurrence du concept  $c_i$ , l'identification des concepts qui sont voisins du concept  $c_i$  dans le document et un score qui est une valeur moyenne des mesures de

- 35

similarités entre le concept  $c_i$  et les termes  $t_i$  du document qui sont les plus proches du concept  $c_i$ .

Avantageusement, le procédé selon l'invention comprend une étape d'optimisation de la partition de l'ensemble  $P$  des termes de la base de termes pour décomposer cet ensemble  $P$  en  $M$  classes  $C_i$  ( $1 \leq i \leq M$ , avec  $M \leq P$ ), de manière à réduire l'erreur de la répartition de l'ensemble  $P$  des termes de la base de termes en  $N$  parties ( $P_1, P_2, \dots, P_N$ ) où chaque partie  $P_i$  est représentée par le terme  $t_i$  qui sera pris comme concept  $c_i$ , l'erreur commise  $\varepsilon$  étant telle que  $\varepsilon = \sum_{i=1}^N \varepsilon_{t_i}$  où  $\varepsilon_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$  est l'erreur

commise lorsqu'on remplace les termes  $t_j$  d'une partie  $P_i$  par  $t_i$ .

Dans ce cas, le procédé peut comprendre les étapes suivantes :

- (i) on décompose l'ensemble  $P$  de termes en deux parties  $P_1$  et  $P_2$  ;
- (ii) on détermine les deux termes les plus éloignés  $t_i$  et  $t_j$  de l'ensemble  $P$  correspondant à la plus grande distance  $D_{ij}$  de la matrice  $T$  de distances ;
- (iii) pour chaque terme  $t_k$  de l'ensemble  $P$ , on examine si la distance  $D_{ki}$  entre le terme  $t_k$  et le terme  $t_i$  est plus petite que la distance  $D_{kj}$  entre le terme  $t_k$  et le terme  $t_j$ , si c'est le cas on affecte le terme  $t_k$  à la partie  $P_1$  et si ce n'est pas le cas on affecte le terme  $t_k$  à la partie  $P_2$  ;
- (iv) on itère l'étape (i) jusqu'à l'obtention du nombre  $N$  de parties souhaité et à chaque itération on applique les étapes (ii) et (iii) sur les termes des parties  $P_1$  et  $P_2$ .

Le procédé selon l'invention peut être plus particulièrement caractérisé en ce qu'il comprend une optimisation à partir des  $N$  parties disjointes  $\{P_1, P_2, \dots, P_N\}$  de l'ensemble  $P$  ainsi que des  $N$  termes  $\{t_1, t_2, \dots, t_N\}$  qui les représentent pour réduire l'erreur de décomposition de l'ensemble  $P$  en  $N$  parties, et en ce qu'il comprend les étapes suivantes :

- (i) calcul des centres de gravité  $C_i$  des parties  $P_i$
- (ii) calcul des erreurs  $\varepsilon_{C_i} = \sum_{t_j \in P_i} d^2(C_i, t_j)$  et  $\varepsilon_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$  lorsqu'on remplace les termes  $t_j$  de la partie  $P_i$  respectivement par  $C_i$  et par  $t_i$ ,
- (iii) comparaison de  $\varepsilon_{t_i}$  et  $\varepsilon_{C_i}$  et remplacement de  $t_i$  par  $C_i$  si  $\varepsilon_{C_i} \leq \varepsilon_{t_i}$ ,

- (iv) calcul de la nouvelle matrice T de distances entre les termes  $t_i$  de la base de termes et processus de décomposition de l'ensemble P des termes de la base de termes en N parties, sauf si une condition d'arrêt est remplie avec  $\frac{\varepsilon_{c_t} - \varepsilon_{c_{t+1}}}{\varepsilon_{c_t}} < \text{seuil}$ , où  $\varepsilon_{c_t}$  représente l'erreur commise à l'instant t.

Afin de faciliter la recherche et l'identification de documents, pour effectuer une structuration du dictionnaire de concepts, on produit de façon itérative à chaque itération une carte de navigation en commençant par scinder l'ensemble des concepts en deux sous-ensembles, puis en sélectionnant un sous-ensemble à chaque itération jusqu'à l'obtention du nombre de groupes souhaité ou jusqu'à ce qu'un critère d'arrêt soit satisfait.

Le critère d'arrêt peut être caractérisé par le fait que les sous-ensembles obtenus sont tous homogènes avec un écart-type faible.

De façon plus particulière, lors de la structuration du dictionnaire de concepts, on détermine des indicateurs de navigation à partir d'une matrice  $M = [c_1, c_2, \dots, c_N] \in \mathfrak{R}^{p \times N}$  de l'ensemble C des concepts  $c_i \in \mathfrak{R}^p$  où  $c_i$  représente un concept de p valeurs, selon les étapes suivantes :

- (i) on calcule un représentant  $w$  de la matrice  $M$ ,
- (ii) on calcule la matrice de covariance  $\tilde{M}$  entre les éléments de la matrice  $M$  et le représentant  $w$  de la matrice  $M$ ,
- (iii) on calcule un axe de projection  $u$  des éléments de la matrice  $M$ ,
- (iv) on calcule la valeur  $p_i = d(u, c_i) - d(u, w)$  et on décompose l'ensemble de concepts C en deux sous-ensembles C1 et C2 de la manière suivante :

$$\begin{cases} c_i \in C1 & \text{si } p_i \leq 0 \\ c_i \in C2 & \text{si } p_i > 0 \end{cases}$$

- (v) on stocke dans le nœud associé à C les informations  $\{u, w, |p_1|, p_2\}$  où  $p_1$  est le maximum de tous les  $p_i \leq 0$  et  $p_2$  est le minimum de tous les  $p_i > 0$ , l'ensemble des informations  $\{u, w, |p_1|, p_2\}$

constituant les indicateurs de navigation dans le dictionnaire de concept.

5 Selon un mode particulier de réalisation, on analyse à la fois les composantes structurelles et le complément de ces composantes structurelles constitué par les composantes texturales d'une image du document, et :

- (a) lors de l'analyse des composantes structurelles de l'image
    - 10 (a1) on procède à une répartition des zones frontières des structures de l'image en différentes classes selon l'orientation de la variation locale d'intensité de manière à définir des éléments de support structurel (ESS) de l'image, et
    - 15 (a2) on procède par analyse statistique à la construction de termes constitués par des vecteurs décrivant les propriétés locales et globales des éléments de support structurels,
  
  - (b) lors de l'analyse des composantes texturales de l'image
    - 20 (b1) on procède à une détection et une caractérisation paramétrique d'une composante purement aléatoire de l'image,
    - (b2) on procède à une détection et une caractérisation paramétrique d'une composante périodique de l'image,
    - (b3) on procède à une détection et une caractérisation paramétrique d'une composante directionnelle de l'image,
    - 25
  - (c) on regroupe dans un nombre limité de concepts l'ensemble des éléments descriptifs de l'image constitués par d'une part les termes décrivant les propriétés locales et globales des éléments de support structurels et d'autre part les paramètres des caractérisations paramétriques des composantes aléatoire, périodique et directionnelle définissant les composantes texturales de l'image, et
    - 30
  - (d) on définit pour chaque document une empreinte à partir des occurrences, des positions et des fréquences desdits concepts.
    - 35
- Avantageusement, les propriétés locales des éléments de support structurels prises en compte pour la construction de termes comprennent

au moins le type de support choisi parmi une bande linéaire ou un arc de courbe, les dimensions en longueur et largeur du support, la direction principale du support et la forme et les propriétés statistiques des pixels constituant le support.

- 5 Les propriétés globales des éléments de support structurels prises en compte pour la construction de termes comprennent au moins le nombre de chaque type de supports et leur disposition spatiale.

De préférence, lors de l'analyse des composantes structurelles de l'image on procède à un test préalable de détection de la présence d'au  
10 moins une structure dans l'image et, en cas d'absence de structure, on passe directement à l'étape de l'analyse des composantes texturales de l'image.

Avantageusement, pour procéder à une répartition des zones frontières des structures de l'image en différentes classes, à partir de  
15 l'image numérisée définie par l'ensemble des pixels  $\gamma(i,j)$  où  $(i,j) \in I \times J$ , avec  $I$  et  $J$  désignant respectivement le nombre de lignes et le nombre de colonnes de l'image, on calcule l'image gradient vertical  $g_v(i,j)$  avec  $(i,j) \in I \times J$  et l'image gradient horizontal  $g_h(i,j)$  avec  $(i,j) \in I \times J$  et on procède au  
20 partitionnement de l'image selon l'orientation locale de son gradient en un nombre fini de classes équidistantes, l'image contenant l'orientation du gradient étant définie par la formule

$$O(i,j) = \arctan \left[ \frac{g_h(i,j)}{g_v(i,j)} \right], \quad (1)$$

- 25 on identifie les classes constituant des régions de support susceptibles de contenir des éléments de support significatifs, et à partir des régions de support, on détermine les éléments de support significatifs et on les répertorie selon des critères prédéterminés.

Selon un aspect particulier de l'invention, on analyse les formes  
30 d'une image d'un document selon les étapes suivantes :

- (a) on procède à une multirésolution suivie d'une décimation de l'image,
- (b) on définit l'image dans l'espace logarithmique polaire.
- (c) on représente l'image ou la portion de l'image concernée par sa transformée de Fourier  $H$ ,

(d) on procède à une caractérisation de la transformée de Fourier H de la façon suivante :

(d1) on projette H dans plusieurs directions pour obtenir un ensemble de vecteurs dont la dimension est égale à la dimension du mouvement de projection,

5

(d2) on calcule les propriétés statistiques de chaque vecteur de projection, et

(e) on représente la forme de l'image par un terme  $t_i$  constitué des valeurs des propriétés statistiques de chaque vecteur de projection.

10

Selon un aspect particulier de l'invention, lors de l'indexation d'un document multimédia comportant des signaux vidéo, on choisit des termes  $t_i$  constitués par des images-clés représentant des groupes d'images homogènes consécutives, et on détermine des concepts  $c_i$  par regroupement de termes  $t_i$ .

15

Pour déterminer des images-clés constituant des termes  $t_i$ , on élabore d'abord un vecteur score VS comprenant un ensemble d'éléments VS(i) matérialisant la différence ou la similarité entre le contenu d'une image d'indice i et celui d'une image d'indice i-1, et on analyse le vecteur score VS afin de déterminer les images-clés qui correspondent aux maximums des valeurs des éléments VS(i) du vecteur score VS.

20

De façon plus particulière, une image d'indice j est considérée comme une image-clé si la valeur VS(j) de l'élément correspondant du vecteur score VS est un maximum et que la valeur VS(j) est située entre deux minimums min G et min D et que le minimum M1 tel que  $M1 = (|VS(j) - \min G|, |VS(j) - \min D|)$  est supérieur à un seuil donné.

25

On considèrera à nouveau l'indexation d'un document multimédia, comportant des composantes audio, on échantillonne et décompose le document en trames, qui sont ensuite regroupées en clips dont chacun est caractérisé par un terme  $t_i$  constitué par un vecteur de paramètre.

30

Une trame peut comprendre par exemple entre environ 512 et 2048 échantillons du document audio échantillonné.

Avantageusement, les paramètres pris en compte pour la définition des termes  $t_i$  comprennent des informations temporelles correspondant à au moins l'un des paramètres suivants : l'énergie des trames du signal audio, l'écart-type des énergies des trames dans les clips, le rapport des variations sonores, le rapport de basse énergie, le taux d'oscillation autour

35

d'une valeur prédéterminée, le haut taux d'oscillation autour d'une valeur prédéterminée,, la différence entre le nombre de taux d'oscillation au-dessus et au-dessous du taux d'oscillation moyen des trames de clips, la variance du taux d'oscillation, le rapport des trames silencieuses.

5           Toutefois, de façon alternative ou complémentaires, de façon avantageuse, les paramètres pris en compte pour la définition des termes  $t_i$  comprennent des informations fréquentielles correspondant à au moins l'un des paramètres suivants : le centre de gravité du spectre de fréquence de la transformée de Fourier courte du signal audio, la largeur  
10 de bande du signal audio, le rapport entre l'énergie dans une bande de fréquence et l'énergie totale dans toute la bande de fréquence du signal audio échantillonné, la valeur moyenne de la variation du spectre de deux trames adjacentes dans un clip, la fréquence de coupure d'un clip.

De façon plus particulière, les paramètres pris en compte pour la  
15 définition des termes  $t_i$  peuvent comprendre au moins la modulation d'énergie à 4 Hz.

D'autres caractéristiques et avantages de l'invention ressortiront de la description suivantes de modes particuliers de réalisation, donnés à titre d'exemples, en référence aux dessins annexés, sur lesquels :

- 20           - la Figure 1 est un schéma-bloc montrant le processus de production d'un dictionnaire de concepts à partir d'une base de documents, conformément à l'invention,  
              - la Figure 2 montre le principe de construction d'une base de concepts à partir de termes,  
25           - la Figure 3 est un schéma-bloc montrant le processus de structuration d'un dictionnaire de concepts, conformément à l'invention,  
              - la Figure 4 montre la structuration d'une base d'empreintes mise en œuvre dans le cadre du procédé selon l'invention,  
              - la Figure 5 est un organigramme montrant les différentes étapes  
30 de construction d'une base d'empreintes,  
              - la Figure 6 est un organigramme montrant les différentes étapes d'identification de documents,  
              - la Figure 7 est un organigramme montrant la sélection d'une première liste de réponses,

- la Figure 8 est un organigramme montrant les différentes étapes d'une phase d'indexation de documents conformément au procédé selon l'invention,
- la Figure 9 est un organigramme montrant les différentes étapes d'extraction de termes dans le cas du traitement d'images,
- la Figure 10 est un schéma résumant le processus de décomposition d'une image régulière et homogène,
- les Figures 11 à 13 montrent trois exemples d'images contenant différents types d'éléments,
- les Figures 14a à 14f montrent respectivement un exemple d'image originale, un exemple d'image après traitement en prenant en compte le module du gradient, et quatre exemples d'images traitées avec démantèlement des zones frontières de l'image,
- la Figure 15a représente un premier exemple d'image contenant un élément directionnel,
- la Figure 15a1 est une vue 3D du spectre de l'image de la Figure 15a,
- la Figure 15b représente un deuxième exemple d'image contenant un élément directionnel,
- la Figure 15b1 est une image module de Fourier de l'image de la Figure 15b,
- la Figure 15c représente un troisième exemple d'image contenant deux éléments directionnels,
- la Figure 15c1 est une image module de Fourier de l'image de la Figure 15c,
- la Figure 16 illustre des directions de projection pour des couples d'entiers  $(\alpha, \beta)$  dans le cadre du calcul de la transformée de Fourier Discrète d'une image,
- la Figure 17 illustre un exemple de mécanisme de projection avec l'exemple d'un couple d'entrées  $(\alpha_k, \beta_k) = (2, -1)$ ,
- la Figure 18a1 représente un exemple d'image contenant des composantes périodiques,
- la Figure 18a2 représente l'image module de la transformée de Fourier Discrète de l'image de la Figure 18a1,
- la Figure 18b1 représente un exemple d'image synthétique contenant une composante périodique,

- la Figure 18b2 représente une vue en 3D de la transformée de Fourier Discrète de l'image de la Figure 18b1, faisant apparaître une paire de pics symétriques,
- la Figure 19 est un organigramme montrant les différentes étapes de traitement d'une image avec établissement d'un vecteur caractérisant la distribution spatiale des propriétés iconiques de l'image,
- 5 - la Figure 20 montre un exemple de partitionnement d'une image et de création d'un vecteur caractéristique de cette image,
- la Figure 21 montre une rotation de  $90^\circ$  de l'image partitionnée de la Figure 20 et la création d'un vecteur caractéristique de cette image,
- 10 - la Figure 22 montre la décomposition d'un signal sonore en trames en clips,
- la Figure 23a montre la variation de l'énergie d'un signal de parole,
- 15 - la Figure 23b montre la variation de l'énergie d'un signal de musique,
- la Figure 24a montre le taux de passage par zéro d'un signal de parole,
- la Figure 24b montre le taux de passage par zéro d'un signal de musique,
- 20 - la Figure 25a montre le centre de gravité du spectre de fréquence de la transformée de Fourier courte d'un signal de parole,
- la Figure 25b montre le centre de gravité du spectre de fréquence de la transformée de Fourier courte d'un signal de musique,
- 25 - la Figure 26a montre la largeur de bande d'un signal de parole,
- la Figure 26b montre la largeur de bande d'un signal de musique,
- la Figure 27a montre pour trois sous-bandes de fréquence 1, 2, 3 le rapport d'énergie dans chaque sous-bande de fréquence sur l'énergie totale de toute la bande de fréquence, pour un signal de parole,
- 30 - la Figure 27b montre pour trois sous-bandes de fréquence 1, 2, 3 le rapport d'énergie dans chaque sous-bande de fréquence sur l'énergie totale de toute la bande de fréquence, pour un signal de musique,
- la Figure 28a montre le flux spectral d'un signal de parole,
- la Figure 28b montre le flux spectral d'un signal de musique,
- 35 - la Figure 29 est un graphique illustrant la définition de la fréquence de coupure d'un clip, et

- la Figure 30 illustre, pour un signal audio, la modulation de l'énergie autour de 4 Hz.

On décrira d'abord en référence aux Figures 1 à 5 le principe général du procédé d'indexation de documents multimédias selon l'invention, qui conduit à la construction d'une base d'empreintes, chaque document indexé étant associé à une empreinte qui lui est propre.

A partir d'une base de documents multimédias 1, une première étape 2 consiste en l'identification et l'extraction, pour chaque document, de termes  $t_i$  constitués par des vecteurs caractérisant des propriétés du document à indexer.

A titre d'exemples, on décrira, en référence aux Figures 22 à 30, la façon dont il est possible d'identifier et d'extraire des termes  $t_i$  pour un document sonore.

Un document audio 140 est d'abord décomposé en trames 160 qui sont regroupées par la suite en clips 150 dont chacun va être caractérisé par un terme constitué par un vecteur de paramètres (Figure 22). Un document audio 140 sera donc caractérisé par un ensemble de termes  $t_i$  qui seront stockés dans une base de termes 3 (Figure 1).

Les documents audio dont on a extrait leur vecteur caractéristique peuvent être échantillonnés par exemple à 22 050 Hz afin d'éviter l'effet de crénelage. Le document est ensuite divisé en un ensemble de trames dont le nombre d'échantillons par trame est fixé en fonction du type de fichier à analyser.

Pour un document audio riche en fréquences et qui contient beaucoup de variations, comme les films par exemple, les émissions de variétés ou encore les émissions sportives, le nombre d'échantillons dans une trame doit être faible, de l'ordre de 512 échantillons par exemple. En revanche, pour un document audio homogène ne contenant que de la parole ou de la musique par exemple, ce nombre doit être important, par exemple de l'ordre de 2 048 échantillons.

Un clip de document audio peut être caractérisé par différents paramètres servant à constituer les termes et caractérisant des informations temporelles ou fréquentielles.

Il est possible d'utiliser tout ou partie des paramètres qui seront mentionnés ci-dessous pour former des vecteurs de paramètres

constituant les termes identifiant les clips successifs du document audio échantillonné.

L'énergie des trames du signal audio constitue un premier paramètre représentant une information temporelle.

5 L'énergie du signal audio varie beaucoup pour la parole alors qu'elle est plutôt stable pour la musique. Elle permet donc de discriminer la parole de la musique mais aussi de détecter les silences. L'énergie peut être couplée à un autre paramètre temporel tel que le taux d'oscillation (TO) autour d'une valeur, qui peut correspondre par exemple au taux de passage par zéro (TPZ). En effet un TO faible et une énergie forte sont synonymes d'un son voisé tandis qu'un TO élevé induit une zone non voisée.

10 La Figure 25a représente un signal 141 qui illustre la variation de l'énergie dans le cas d'un signal de parole.

15 La Figure 23b représente un signal 142 qui illustre la variation de l'énergie dans le cas d'un signal de musique.

Soit  $N$  le nombre d'échantillons dans une trame, le volume ou énergie  $E(n)$  est défini par :

$$E(n) = \frac{1}{N} \sum_{i=0}^{N-1} S_n^2(i). \quad (2)$$

20 où  $S_n(i)$  représente la valeur de l'échantillon  $i$  de la trame d'indice  $n$  d'un signal audio.

D'autres paramètres représentatifs d'informations temporelles peuvent être déduits de l'énergie, comme par exemple :

25 - l'écart type des énergies des trames dans les clips (encore appelé EEC ou VSTD) qui constitue un état défini comme la variance des volumes des trames dans un clip normalisé par le maximum du volume des trames du clip,

30 - le rapport des variations sonores (RVS) qui est constitué par la différence entre le maximum et le minimum des volumes des trames d'un clip divisé par le maximum des volumes de ces trames,

- le rapport de basse énergie (ou LER) qui est le pourcentage des trames dont le volume est inférieur à un seuil (qui est fixé par exemple à 95% du volume moyen d'un clip).

D'autres paramètres permettent de caractériser l'aspect temporel d'un clip, en particulier le taux d'oscillation autour d'une valeur prédéterminée, qui, lorsque cette valeur prédéterminée est zéro, définit un taux de passage par zéro (ou TPZ).

- 5 Le TPZ peut aussi être défini par le nombre de fois où l'onde passe par zéro .

$$Z(n) = \frac{1}{2} \left( \sum_{i=0}^{N-1} | \text{Sign}(S_n(i)) - (\text{Sign}(S_n(i-1))) | \right) \frac{f_s}{N} \quad (3)$$

$S_n(i)$  : Valeur de l'échantillon  $i$ , de la trame  $n$ .

$N$  : nombre d'échantillons dans une trame.

- 10  $f_s$  : fréquence d'échantillonnage.

Cette caractéristique est fréquemment utilisée pour la classification parole / musique. En effet, les brusques variations du TPZ sont significatives de l'alternance voisée / non voisée donc de la présence de parole. Pour la parole, le TPZ est faible pour les zones voisées et très élevé pour les zones non voisées alors que pour la musique, les variations du TPZ sont très faibles.

15 La Figure 24a montre une courbe 143 illustrant un exemple de TPZ pour un signal de parole.

20 La Figure 24b montre une courbe 144 illustrant un exemple de TPZ pour un signal de musique.

Un autre paramètre caractérisant l'aspect temporel d'un clip peut être constitué par le haut taux d'oscillation autour d'une valeur prédéterminée qui, lorsque cette valeur prédéterminée est zéro, définit un haut taux de passage par zéro (ou HTPZ).

- 25 Le HTPZ peut être défini comme étant le rapport du nombre de trames dont le TPZ est à une valeur  $\alpha$ , par exemple 1,5 au-dessus du TPZ moyen du clip (1s) :

$$HTPZ = \frac{1}{2N} \sum_{n=0}^{N-1} [ \text{sgn}(TPZ(n) - 1.5avTPZ) + 1 ] \quad (4)$$

$$\text{tel que : } avTPZ = \frac{1}{N} \sum_{n=0}^{N-1} TPZ(n) . \quad (5)$$

avec:

n : indice de la trame .

N : nombre de trames dans un clip.

5 Pour les segments de parole les clips sont de 0 à 200 s avec un HTPZ  
autour de 0,15.

En revanche, pour les segments de musique, les clips sont de 200 à 350 s  
et le HTPZ varie autour de 0,05 et est en général quasi nul.

Pour le son d'environnement les segments correspondant aux clips sont de  
10 351 à 450 s,

Le HTPZ est faible pour le bruit blanc et grand pour un son  
assourdissant (tambour par exemple).

15 On peut encore définir le paramètre DTPZ qui est constitué par la  
différence entre le nombre de TPZ au-dessus et en-dessous du TPZ moyen  
des trames d'un clip, ainsi que le paramètre VTPZ qui est constitué par la  
variance du TPZ.

Un autre paramètre caractérisant l'aspect temporel d'un clip est le  
rapport des trames silencieuses (RFS) qui est le pourcentage des trames  
non silencieuses dans un clip.

20 Une trame est non silencieuse si son volume dépasse un certain  
seuil (10) et si la valeur du TPZ est inférieure à un Tpz seuil.

Ainsi le rapport de trames non silencieuses dans un clip, permet de  
détecter le silence.

D'autres propriétés statistiques du TPZ peuvent être utilisées  
comme paramètres caractéristiques, telles que :

- 25 i) moment du troisième ordre de la moyenne,  
ii) le nombre de TPZ dépassant un certain seuil.

30 Les paramètres pris en compte pour la définition des termes  $t_i$   
peuvent comprendre également des informations fréquentielles qui  
prennent en compte le calcul de la transformée de Fourier rapide (FFT) du  
signal audio.

Ainsi, un paramètre appelé centroïde spectacle (CS) peut être défini  
comme étant le centre de gravité du spectre de fréquence de la  
transformée de Fourier courte (STFT) du signal audio :

$$CS(n) = \frac{\sum_{i=0}^{N-1} i S_n(i)}{\sum_{i=0}^{N-1} S_n(i)} \quad (6)$$

telle que  $S_n(i)$  : Puissance spectrale de la trame  $i$  du clip  $n$  .

Le paramètre CS est élevé pour la musique car les hauteurs sont  
 5 réparties sur une zone plus étendue que celle de la parole (en général 6  
 octaves pour la musique et 3 pour la parole ). Il a une relation avec la  
 sensation de la brillance du son qu'on entend. C'est un attribut perceptuel  
 important pour la caractérisation du timbre .

La Figure 25a montre une courbe 145 illustrant un exemple de CS  
 10 pour un signal de parole.

La Figure 25b montre une courbe 146 illustrant un exemple de CS  
 pour un signal de musique.

Un autre paramètre est constitué par la largeur de bande LB qui  
 peut être calculée à partir de la variance du paramètre précédent CS(n).

15

$$LB^2(n) = \frac{\sum_{i=0}^{N-1} (i - CS(n))^2 S_n(i)}{\sum_{i=0}^{N-1} S_n(i)} \quad (7)$$

La largeur de bande LB est importante tant en musique qu'en  
 parole.

20 La Figure 26a montre une courbe 147 illustrant un exemple de  
 largeur de bande d'un signal de parole.

La Figure 26b montre une courbe 148 illustrant un exemple de  
 largeur de bande d'un signal de musique.

Un autre paramètre utile est constitué par le rapport ERSB entre  
 25 l'énergie dans une sous-bande de fréquence  $i$  et l'énergie totale dans  
 toute la bande de fréquence du signal audio échantillonné.

En considérant les propriétés perceptuelles de l'oreille humaine, la  
 bande de fréquence a été divisée en quatre sous-bandes où ces dernières  
 correspondent aux filtres de Cochlear. Quand la fréquence d'échan-

tillonnage est de 22025 Hz, les bandes de fréquences sont : 0-630Hz , 630-1720Hz , 1720-4400Hz et 4400-11025Hz .Pour chacune de ces bandes on calcule son énergie ERSBi, qui correspond au rapport de l'énergie de cette dernière sur l'énergie dans toute la bande de fréquence .

5 La Figure 27a montre trois courbes 151, 152, 153 illustrant pour trois sous-bandes de fréquence 1, 2, 3 le rapport d'énergie dans chaque sous-bande de fréquence sur l'énergie totale de toute la bande de fréquence, pour un exemple de signal de parole.

10 La Figure 27b montre trois courbes 154, 155, 156 illustrant pour trois sous-bandes de fréquence 1, 2, 3 le rapport d'énergie dans chaque sous-bande de fréquence sur l'énergie totale de toute la bande de fréquence, pour un exemple de signal de musique.

15 Un autre paramètre est constitué par le flux spectral qui est défini comme la valeur moyenne de la variation du spectre de deux trames adjacentes dans un clip :

$$FS(n) = \frac{1}{N} \sum_{i=1}^N [\log(S_n(i) + \delta) - \log(S_n(i-1) + \delta)]^2 \quad (8)$$

où

$\delta$  : Une constante de faible valeur,

$S_n(i)$  : Puissance spectrale de la trame  $i$  du clip  $n$ .

20 Le flux spectral de la parole est en général plus important que celui de la musique, et celui du son d'environnement est le plus grand. Il varie considérablement en comparaison avec les deux autres signaux.

La Figure 28a montre une courbe 157 illustrant le flux spectral d'un exemple de signal de parole.

25 La Figure 28b montre une courbe 158 illustrant le flux spectral d'un exemple de signal de musique.

Un autre paramètre utile est constitué par la fréquence de coupure d'un clip (FCC).

30 La Figure 29 montre une courbe 149 illustrant le spectre d'amplitude en fonction de la fréquence  $f_e$ , et la fréquence de coupure  $f_c$  qui est la fréquence en dessous de laquelle 95% de l'énergie du spectre (la puissance spectrale) est concentrée.

Pour déterminer la fréquence de coupure du clip, on calcule la transformée de Fourier du clip  $DS(n)$

$$DS(n) = \sum_{i=0}^{N-1} S_n^2(i) \quad (9) . \text{ La fréquence } fc \text{ est déterminée par :}$$

$$5 \quad \sum_{i=0}^{fc} S_n^2(i) \geq 0.95 \times DS \quad (10) \text{ et } \sum_{i=0}^{fc-1} S_n^2(i) < 0.95 \times DS \quad (11)$$

La FCC est plus élevée pour un son non voisé (son riche en hautes fréquences ) que pour un son voisé (présence de parole où la puissance est concentrée dans les basses fréquences).

10 Cette mesure permet de caractériser les alternances voisées/non voisées de la parole car cette valeur est faible pour les clips contenant uniquement de la musique.

15 D'autres paramètres peuvent encore être pris en compte pour la définition des termes  $t_i$  d'un document audio, comme la modulation d'énergie autour de 4 Hz, qui constitue un paramètre issu à la fois d'une analyse fréquentielle et d'une analyse temporelle.

La modulation d'énergie à 4 Hz (4 ME) est calculée à partir du contour du volume, selon la formule suivante :

$$20 \quad 4ME = \frac{\sum_{i=0}^{N/T} (\sum_{j=0}^T W(j) S_n^2(j+i \times T)) / T}{\sum_{i=0}^N S_n^2(i)} \quad (12)$$

où  $S_n(i)$  : Puissance spectrale de la trame  $i$  du clip  $n$  .

$W(j)$  : Fenêtre triangulaire centrée à 4Hz.

T : Largeur d'un clip.

25 La parole a une 4ME plus importante que la musique car, pour la parole, les changements de syllabe se situent autour de 4Hz.

Une syllabe est en effet une combinaison d'une zone de faible énergie (consonne) et d'une zone de forte énergie (voyelle).

La Figure 30 montre une courbe 161 illustrant un exemple de signal audio et une courbe 162 montrant pour ce signal la modulation de l'énergie autour de 4 Hz.

On a décrit ci-avant le cas de documents multimédias comportant des composantes audio.

Dans le cas de l'indexation de documents multimédias comportant des signaux vidéo, on peut choisir des termes  $t_i$  constitués par des images-clés représentant des groupes d'images homogènes consécutives.

Les termes  $t_i$  peuvent à leur tour représenter par exemple les couleurs dominantes, les propriétés texturales, les structures de zones dominantes des images-clés du document vidéo.

D'une manière générale, dans le cas des images qui sera développé plus en détail plus loin, les termes peuvent représenter les couleurs dominantes, les propriétés texturales, les structures des zones dominantes de l'image. Plusieurs procédés peuvent être mis en œuvre de façon alternative ou cumulative, aussi bien sur la totalité de l'image que sur des portions de l'image, pour déterminer les termes  $t_i$  devant caractériser l'image.

Dans le cas d'un document contenant du texte, les termes  $t_i$  peuvent être constitués par des mots du langage parlé ou écrit, par des nombres et par d'autres identificateurs constitués de combinaisons de caractères (par exemple des combinaisons de lettres et de chiffres).

On considèrera à nouveau l'indexation d'un document multimédia comportant des signaux vidéo, pour lequel on choisit des termes  $t_i$  constitués par des images-clés représentant des groupes d'images homogènes consécutives, et on détermine des concepts  $c_i$  par regroupement des termes  $t_i$ .

La détection des images clés repose sur le regroupement des images d'un document vidéo en groupes contenant chacun uniquement des images homogènes. De chacun des groupes on extrait une ou plusieurs images (appelées images clés) représentant le document vidéo.

Le regroupement des images du document vidéo repose sur la production d'un vecteur score appelé VS représentant le contenu de la vidéo, il caractérise la variation des images consécutives de la vidéo

(l'élément  $VS_i$  matérialise la différence entre le contenu de l'image d'indice  $i$  et celui de l'image d'indice  $i-1$ ),  $VS$  est égal à zéro quand les contenus  $im_i$  et  $im_{i-1}$  sont identiques et il est important quand la différence entre les deux contenus est importante.

- 5 Pour calculer le signal  $VS$ , les trois bandes de chaque image  $im_i$  RGB d'indice  $i$  de vidéo sont additionnées pour ne constituer qu'une seule image qu'appelée  $TR_i$ . Ensuite, l'image  $TR_i$  est décomposée en plusieurs bandes de fréquence pour ne conserver que la composante basse fréquence  $TR_{Bi}$ . On utilise pour cela deux filtres à miroir (un filtre Passe bas PB et un filtre Passe Haut PH) qui sont appliqués successivement sur  
10 bas PB et un filtre Passe Haut PH) qui sont appliqués successivement sur les lignes et sur les colonnes de l'image. On considèrera deux types de filtre : l'ondelette de Haar et le filtre dont l'algorithme est le suivant :

Balayage ligne

- 15 A partir de  $TR_k$  on produit l'image Bas  
Pour chaque point  $a_{2x_i,j}$  de l'image TR faire  
Calculer le point  $b_{i,j}$  de l'image basse fréquence bas,  $b_{i,j}$  prend la valeur médiane de  $a_{2x_i,j-1}$ ,  $a_{2x_i,j}$  et  $a_{2x_i,j+1}$ .

20 Balayage Colonne

A partir des deux images Bas on produit l'image  $TR_{Bk}$   
Pour chaque point  $b_{i,2x_j}$  de l'image TR faire  
Calculer le point  $bb_{i,j}$  de l'image basse fréquence bas,  $bb_{i,j}$  prend la valeur médiane de  $b_{i,2x_j-1}$ ,  $b_{i,2x_j}$  et  $b_{i,2x_j+1}$

- 25 Les balayages ligne et colonne sont appliqués autant de fois qu'on le souhaite. Le nombre d'itérations dépend de la résolution des images de la vidéo. Pour des images de taille 512x512 ou peut fixer  $n$  à trois.

- L'image résultat  $TR_{Bi}$  est projetée dans plusieurs directions pour  
30 obtenir un ensemble de vecteurs  $V_k$ ,  $k$  est l'angle de projection (l'élément  $j$  de  $VO$ , vecteur obtenu suite à la projection horizontale de l'image, est égal à la somme de tous les points de la ligne  $j$  de l'image). Les vecteurs de directions de l'image  $TR_{Bi}$  sont comparés aux vecteurs de direction de  $TR_{Bi-1}$  pour obtenir un score  $i$  qui mesure la similarité entre  
35 ces deux images. Ce score est obtenu par la moyenne de toutes les distances des vecteurs de même direction : pour chaque  $k$  on calcule la

distance entre le vecteur  $V_k$  de l'image  $i$  et le vecteur  $V_k$  de l'image  $i-1$  puis on calcule toutes ces distances.

L'ensemble de tous les scores constitue le vecteur score  $VS$  : l'élément  $i$  de  $VS$  mesure la similarité entre l'image  $TRBi$  et l'image  $TRBi-1$ .

5 Le vecteur  $VS$  est lissé afin d'éliminer les irrégularités du au bruit engendré lors la manipulation de la vidéo.

On décrira ci-après un exemple de regroupement des images et d'extraction des images clés.

10 Le vecteur  $VS$  est analysé afin de déterminer les images clés qui correspondent aux maximums des valeurs de  $VS$ . Une image d'indice  $j$  est considérée comme une image-clé si la valeur  $VS(j)$  est un maximum et si  $VS(j)$  est situé entre deux minimums  $minG$  (minimum gauche) et  $minD$  (minimum droite) et si le minimum  $M1$  tel que

$M1 = \min ( |VS(j)-minG|, |VS(j)-minD| )$  est supérieur à un seuil donné.

15 Pour détecter les images-clés, on initialise  $minG$  avec  $VS(0)$  puis on parcourt le vecteur  $VS$  de la gauche vers la droite. A chaque étape, on détermine l'indice  $j$  correspondant à la valeur maximale située entre deux minimums ( $minG$  et  $minD$ ) puis en fonction du résultat de l'équation définissant  $M1$  on décide de considérer  $j$  comme un indice d'une image-clé ou non. Il est possible de prendre un groupe de plusieurs images-clés voisines, par exemple des images-clés d'indices  $j-1, j$  et  $j+1$ .

20 Trois cas se présentent si le minimum des deux pentes, définies par les deux minimums ( $minG$  et  $minD$ ) et la valeur maximale, n'est pas supérieur au seuil :

25 i) Si  $|VS(j) - minG|$  est inférieur au seuil et que  $minG$  ne correspond pas à  $VS(0)$ , le maximum  $VS(j)$  est ignoré et  $minD$  devient  $minG$ ,

30 ii) Si  $|VS(j) - minG|$  est supérieur au seuil et si  $|VS(j)-minD|$  est inférieure au seuil, le  $minD$  et le maximum  $VS(j)$  sont conservés et  $minD$  est ignoré sauf si le plus proche maximum à droite de  $minD$  est supérieur à un seuil. Dans ce cas, on conserve aussi  $minD$  et on déclare  $j$  comme un indice d'une image-clé. Dans le cas où  $minD$  est ignoré,  $minD$  prendra la valeur la plus proche du minimum situé à droite de  $minD$ .

35 iii) Si les deux pentes sont inférieures au seuil,  $minG$  est conservé et  $minD$  et  $j$  sont ignorés.

Après sélection d'une image-clé, on itère le processus. A chaque itération minD devient minG.

Si l'on se reporte à nouveau à la Figure 1; à partir d'une base de termes 3 comprenant P termes, on procède dans une étape 4 à un traitement des termes  $t_j$  et à leur regroupement en concepts  $c_i$  (Figure 2) destinés à être stockés dans un dictionnaire de concepts 5. Il s'agit ici d'élaborer un ensemble de signatures caractérisant une classe de documents. Les signatures sont des descripteurs qui, par exemple dans le cas de l'image, représentent la couleur, la forme et la texture. Un document peut alors être caractérisé et représenté par les concepts du dictionnaire.

Une empreinte d'un document peut alors être formée par les vecteurs signatures de chaque concept du dictionnaire 5. Le vecteur signature est constitué par les documents où le concept  $c_i$  est présent ainsi que par les positions et le poids de ce concept dans le document.

Les termes  $t_j$  extraits d'une base de documents 1 sont stockés dans une base de termes 3 et traités dans un module 4 d'extraction de concepts  $c_i$  qui sont eux-mêmes regroupés dans un dictionnaire de concepts 5. La Figure 2 illustre le processus de construction d'une base de concepts  $c_i$  ( $1 \leq i \leq m$ ) à partir de termes  $t_j$  ( $1 \leq j \leq n$ ) présentant des scores de similarité  $w_{ij}$ .

Le module de la production du dictionnaire de concepts reçoit en entrée l'ensemble P des termes de la base 3 et le nombre maximum N de concepts souhaité est fixé par l'utilisateur. Chaque concept  $c_i$  est prévu pour regrouper tous les termes voisins du point de vue de leurs caractéristiques.

Pour produire le dictionnaire de concepts, on commence par calculer la matrice de distance  $T$  entre les termes de la base 3, cette matrice est utilisée pour créer une partition dont le cardinal est égal au nombre N de concepts souhaité.

La création du dictionnaire de concepts s'effectue en deux phases :

Décomposition de P en N parties  $P = P_1 \cup P_2 \dots \cup P_N$

Processus d'optimisation de la partition qui décompose P en M classes

$P = C_1 \cup C_2 \dots \cup C_M$  avec M est inférieur ou égal à P.

Le processus d'optimisation a pour but de réduire l'erreur de la répartition de P en N parties  $\{P_1, P_2, \dots, P_N\}$  où chaque partie  $P_i$  est

représentée par le terme  $t_i$  qui sera pris comme concept, l'erreur commise est alors égale à l'expression suivante :

$$\varepsilon = \sum_{i=1}^N \varepsilon_{t_i}, \quad \varepsilon_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$$

est l'erreur commise lorsqu'on

- 5 remplace les termes  $t_j$  de  $P_i$  par  $t_i$ .

On peut décomposer P en N parties de manière à répartir les termes de telle façon que les termes les plus éloignés se trouvent dans des parties distinctes et les termes proches se trouvent dans la même partie.

- 10 On décrira d'abord l'étape 1 de décomposition de l'ensemble de termes P en deux parties  $P_1$  et  $P_2$  :

(a) On détermine les deux termes les plus éloignés  $t_i$  et  $t_j$  de P correspondant à la plus grande distance  $D_{ij}$  de la matrice T.

- (b) Pour chaque  $t_k$  de P,  $t_k$  est affecté à  $P_1$  si la distance  $D_{ki}$  est plus petite que la distance  $D_{kj}$  et à  $P_2$  sinon.
- 15

On itère l'étape 1 jusqu'à l'obtention du nombre de parties souhaité et à chaque itération on applique les étapes (a) et (b) sur les termes de l'ensemble P1 et de l'ensemble P2.

On décrira maintenant une phase d'optimisation.

- 20 Le processus d'optimisation a pour point de départ les N parties disjointes de P  $\{P_1, P_2, \dots, P_N\}$  ainsi que les N termes  $\{t_1, t_2, \dots, t_N\}$  qui les représentent et il est utilisé afin de réduire l'erreur de décomposition de P en  $\{P_1, P_2, \dots, P_N\}$  parties.

- 25 On commence par calculer les centres de gravités  $C_i$  des  $P_i$ . Ensuite on calcule l'erreur  $\varepsilon_{C_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$  qu'on compare à  $\varepsilon_{t_i}$  et on remplace  $t_i$

- par  $C_i$  si  $\varepsilon_{C_i}$  est inférieur à  $\varepsilon_{t_i}$ . Puis, après avoir calculé la nouvelle matrice T et si la convergence n'est pas atteinte, on procède à une décomposition. La condition d'arrêt est définie par  $\frac{(\varepsilon_{C_t} - \varepsilon_{C_{t+1}})}{\varepsilon_{C_t}} < \text{seuil}$  qui est de l'ordre de  $10^{-3}$ .  $\varepsilon_{C_t}$  étant l'erreur commise à l'instant t qui
- 30 représente l'itération.

On présente ci-dessous une matrice T de distances entre les termes, où  $D_{ij}$  désigne la distance entre le terme  $t_i$  et le terme  $t_j$ .

	$t_0$		$t_i$		$t_k$		$t_j$		$t_n$
$t_0$	$D_{00}$		$D_{0i}$		$D_{0k}$		$D_{0j}$		$D_{0n}$
$t_i$	$D_{i0}$		$D_{ii}$		$D_{ik}$		$D_{ij}$		$D_{in}$
$t_k$	$D_{k0}$		$D_{ki}$		$D_{kk}$		$D_{kj}$		$D_{kn}$
$t_j$	$D_{j0}$		$D_{ji}$		$D_{jk}$		$D_{jj}$		$D_{jn}$
$t_n$	$D_{n0}$		$D_{ni}$		$D_{nk}$		$D_{nj}$		$D_{nn}$

La Figure 3 illustre, dans le cas de documents multimédias de contenus divers, un exemple de structuration du dictionnaire de concept 5.

Afin de faciliter la navigation à l'intérieur du dictionnaire 5 et de déterminer rapidement lors d'une phase d'identification le concept le plus proche d'un terme donné, le dictionnaire 5 est analysé et une carte de navigation 9 à l'intérieur du dictionnaire est établie.

La production de la carte de navigation 9 s'effectue de façon itérative. A chaque itération, on commence par scinder l'ensemble de concepts en deux sous-ensembles, puis à chaque itération on sélectionne un sous-ensemble jusqu'à l'obtention du nombre de groupes souhaité ou bien jusqu'à ce que le critère d'arrêt soit satisfait. Ce critère d'arrêt peut être par exemple que les sous-ensembles obtenus sont tous homogènes avec un écart-type faible par exemple. Le résultat final est un arbre binaire où les feuilles contiennent les concepts du dictionnaire et les nœuds de l'arbre contiennent les informations nécessaires à la scrutation de l'arbre lors de la phase d'identification d'un document.

On décrira ci-dessous un exemple de module 6 de répartition d'un ensemble de concepts.

L'ensemble de concepts  $C$  est représenté sous la forme d'une matrice  $M = [c_1, c_2, \dots, c_N] \in \mathfrak{R}^{p \times N}$ , avec  $c_i \in \mathfrak{R}^p$ , où  $c_i$  représente un concept de  $p$  valeurs. Différentes méthodes sont possibles pour assurer une répartition axiale. Dans ce cas, on commence par calculer le centre de

gravité C ainsi que l'axe utilisé pour décomposer l'ensemble en deux sous-ensembles.

Les étapes de traitement sont les suivantes :

- 5 Étape 1 : on calcule un représentant de la matrice M tel que le centroïde  $w$  de la matrice  $M$  :  $w = \frac{1}{N} \sum_{i=1}^N c_i$  (13)

- 10 Étape 2 : on calcule la matrice de covariance  $\tilde{M}$  entre les éléments de la matrice M et le représentant de la matrice M avec, dans le cas particulier ci-dessus  $\tilde{M} = M - we$ , où  $e = [1,1,1,\dots,1]$  (14)

Étape 3 : on calcule un axe de projection des éléments de la matrice M, par exemple le vecteur propre U associé à la plus grande valeur propre de la matrice de covariance.

- 15 Étape 4: on calcule la valeur  $p_i = u^T(c_i - w)$  et on décompose l'ensemble de concepts C en deux sous-ensembles C1 et C2 de la manière suivante :

$$\begin{cases} c_i \in C1 & \text{si } p_i \leq 0 \\ c_i \in C2 & \text{si } p_i > 0 \end{cases} \quad (15)$$

20

Les informations stockées dans le nœud associé à C sont  $\{u, w, |p1|, p2\}$  avec  $p1$  est le maximum de tous les  $p_i \leq 0$  et  $p2$  est le minimum de tous les  $p_i > 0$ .

- 25 L'ensemble  $\{u, w, |p1|, p2\}$  constitue les indicateurs de navigation dans le dictionnaire de concept. En effet pour déterminer, lors de la phase d'identification par exemple, le concept le plus proche d'un terme  $t_i$ , on calcule la valeur  $p_{ti} = u^T(t_i - w)$  puis on sélectionne le nœud associé à C1 si  $\|p_{ti}\| - |p1| < \|p_{ti}\| - p2$  et on sélectionne le nœud C2 si non. On itère le processus jusqu'à ce que l'on ait atteint une des feuilles de l'arbre.

30

Un module détecteur de singularité 8 peut être associé au module 6 de répartition des concepts.

Ce détecteur de singularité permet de sélectionner l'ensemble  $C_i$  à décomposer. Une des méthodes possibles consiste à sélectionner l'ensemble le moins compact.

Les Figures 4 et 5 illustrent l'indexation d'un document ou d'une  
5 base de documents et la construction d'une base d'empreintes 10.

La base d'empreintes 10 est constituée de l'ensemble des concepts représentant les termes des documents à protéger. A chaque concept  $C_i$  de la base d'empreintes 10 est associée une empreinte 11, 12, 13 constituée par un ensemble d'informations telles que le nombre de termes  
10 dans les documents où le concept est présent, et pour chacun de ces documents on enregistre une empreinte 11a, 11b, 11c comprenant l'adresse du document, le nombre de termes, le nombre d'occurrences du concept (fréquence), le score, ainsi que les concepts qui lui sont voisins dans le document. Le score est une valeur moyenne des mesures de  
15 similarité entre le concept et les termes du document qui sont les plus proches du concept. L'adresse d'un document donné est stockée dans une base 14 des adresses des documents protégés.

Le processus 20 de génération des empreintes ou signatures de documents à indexer est illustré sur la Figure 5.

20 Lors de l'enregistrement d'un document, on extrait les termes pertinents du document (étape 21) et on prend en compte le dictionnaire des concepts (étape 22). Chacun des termes  $t_i$  du document est projeté dans l'espace du dictionnaire de concepts afin de déterminer le concept  $c_i$  représentant le terme  $t_i$  (étape 23).

25 On met ensuite à jour l'empreinte du concept  $c_i$  (étape 24). Cette mise à jour s'effectue selon que le concept a déjà été rencontré, c'est-à-dire est présent dans les documents qui sont déjà enregistrés ou non.

Si le concept  $c_i$  n'est pas encore présent dans la base, on crée une nouvelle entrée dans la base (une entrée dans la base correspond à un  
30 objet dont les éléments sont des objets contenant la signature du concept dans les documents où ce concept est présent). On initialise l'entrée créée avec la signature du concept. La signature d'un concept dans un document est matérialisée principalement par les informations suivantes : NbTermes, Fréquence, Concepts Voisins et score.

Si le concept  $c_i$  existe dans la base, on ajoute à l'entrée associée au concept sa signature dans le document qui est composée de (Adresse du document, NbTermes, Fréquence, Concepts Voisins et score).

Lorsque la base d'empreintes est construite (étape 25), on procède à l'enregistrement de la base d'empreintes (étape 26).

La Figure 6 illustre un processus d'identification d'un document qui est implémenté sur une plate-forme 30 de recherche en ligne.

L'identification d'un document a pour objectif de déterminer si un document posé comme question est la réutilisation d'un document de la base. Elle est basée sur la mesure de similarité entre documents. Le but est d'identifier les documents contenant des éléments protégés. La reprise peut être totale ou partielle. Dans ce dernier cas, l'élément copie a subi des modifications telles que : suppression de phrases dans un texte, suppression de motif dans une image, suppression de plan ou de séquence dans un document vidéo,... changement d'ordre des termes ou substitution de termes par d'autres termes dans un texte.

Après présentation d'un document à identifier (étape 31), on procède à l'extraction des termes de ce document (étape 32).

En liaison avec une base d'empreintes (étape 25), on met en correspondance les concepts calculés à partir des termes extraits de la question, avec les concepts de base (étape 33), afin d'établir une liste de documents ayant des contenus similaires aux contenus du document question.

Le processus d'établissement de la liste est le suivant :

On note  $p_{dj}$  : le degré de ressemblance du document  $d_j$  au document question, avec  $1 \leq j \leq N$ ,  $N$  est le nombre de documents de la base de référence

On initialise à zéro tous les  $p_{dj}$

Pour chaque terme  $t_i$  de la question fourni à l'étape 331 (Figure 7) on détermine le concept  $C_i$  qui le représente (étape 332).

Pour chaque document  $d_j$  où le concept est présent on met à jour son  $p_{dj}$  de la manière suivante :

$p_{dj} = p_{dj} + f(\text{fréquence}, \text{score})$ , plusieurs fonctions  $f$  peuvent être utilisées par exemple  $f(\text{fréquence}, \text{score}) = \text{fréquence} \times \text{score}$ , fréquence désigne le

nombre d'occurrences du concept  $C_i$  dans le document  $d_j$  et score désigne

la moyenne des scores de ressemblance des termes du document  $d_j$  avec le concept  $C_j$ .

On ordonne les  $p_{dj}$  et on conserve ceux qui sont supérieurs à un seuil donné (étape 33). On procède ensuite à une confirmation et une validation des réponses (étape 34).

Confirmation des réponses : la liste des réponses est filtrée afin de n'en garder que les réponses les plus pertinentes. Le filtrage utilisé est basé sur la corrélation entre les termes de la question et de chacune des réponses.

Validation: elle permet de ne conserver que les réponses où il y a une grande certitude de reprise de contenu. Dans cette étape les réponses sont filtrées en tenant compte des propriétés algébriques et topologiques des concepts à l'intérieur d'un document : on exige que le voisinage dans le document question soit respecté dans les documents réponses, c'est à dire que deux concepts voisins dans le document question doivent être voisins dans le document réponse.

On fournit alors la liste des documents réponses (étape 35).

On considèrera maintenant plus particulièrement le cas de documents multimédias contenant des images.

On décrira en particulier pour la construction de la base d'empreintes qui servira d'outil pour l'identification d'un document, des procédés rapides et efficaces d'identification d'images qui tiennent compte de toutes les informations pertinentes contenues dans les images allant de la caractérisation des structures ou objets qui la composent, à celle des zones texturées et à la couleur de fond. Les objets de l'image sont identifiés par la production d'une table résumant différentes statistiques faites sur des informations des zones frontières des objets ainsi que des informations sur les voisinages de ces zones frontières. La caractérisation des zones texturées peut être effectuée à l'aide d'une description très fine à la fois spatiale et spectrale de la texture suivant trois caractéristiques fondamentales qui sont sa périodicité, son orientation globale et l'aspect aléatoire de son motif. La texture est ici assimilée à une réalisation de processus aléatoire bidimensionnel. La caractérisation de la couleur est un volet important de la méthode. Elle peut être utilisée comme un premier tri des réponses similaires basées sur la couleur, ou alors une dernière décision faite pour affiner la recherche.

Dans le premier volet de la phase de construction d'empreintes, on prend en compte des informations classifiées sous forme de composants appartenant à deux grandes catégories :

- 5 - les composants dits structurels qui décrivent la perception par l'œil d'un objet pouvant être isolé ou d'un ensemble d'objets disposé selon un arrangement spatial (images 81 et 82 des Figures 11 et 12),
- les composants dits texturaux qui sont le complément des composants structurels et qui traduisent la régularité ou l'homogénéité des motifs de texture (images 82 et 83 des Figures 12 et 13).

10 La Figure 11 montre ainsi une image 81 contenant des éléments structurels et ne présentant pas de motifs de texture.

La Figure 12 montre une image 81 contenant des éléments structurels et un fond texturé.

15 La Figure 13 montre une image 83 sans éléments structurels mais entièrement texturée.

Comme indiqué plus haut, lors de la phase de construction d'empreintes, chaque document de la base de documents est analysé afin d'en extraire les informations pertinentes. Ces informations seront ensuite répertoriées et analysées. Cette analyse se fait suivant un enchaînement de procédures qui se résume en trois étapes :

- 20 - Extraction pour chaque document de caractéristiques prédéfinies et stockage de ces informations dans un vecteur appelé terme.
- Regroupement dans un concept de tous les termes "voisins" du point de vue de leurs caractéristiques, ce qui permet de rendre la recherche plus concise.
- 25 - Construction d'une empreinte qui caractérise ce document par un nombre réduit d'entités. Chaque document est ainsi associé à une empreinte qui lui est propre.

30 La Figure 8 illustre le cas de l'indexation d'un document image 52 contenu dans une base d'images 51 préalablement enregistrée, pour caractériser cette image 52 par un nombre fini de paramètres pouvant être facilement stockés et manipulés ultérieurement. On procède à l'étape 53 à l'extraction de termes du document à chercher qui sont stockés dans une mémoire tampon (étape 54).

35 On opère à l'étape 55 une projection dans l'espace des termes de la base de références.

A l'étape 56, on procède à une description vectorielle donnant les valeurs de pertinence des termes dans le document à chercher.

L'étape 57 consiste en une répartition des termes dans N groupes 58 de concepts.

5 L'étape 59 consiste en une projection dans l'espace des concepts de chaque groupe 58 pour obtenir N partitions 61.

Enfin, une projection orthogonale 62 conduit à N ensembles 63 de descriptions vectorielles réduites.

10 Lors d'une phase ultérieure de recherche, suite à une requête formulée par un utilisateur, par exemple l'identification d'une image question, on recherche tous les documents multimédias similaires ou répondant à cette requête. Pour ce faire, comme indiqué plus haut, on calcule les termes du document question et on les compare aux concepts de la base afin de déduire le ou les documents de la base qui sont  
15 similaires au document question.

On décrira ci-dessous de façon plus détaillée la phase de construction des termes d'une image.

La phase de construction des termes d'une image met en œuvre utilement la caractérisation des supports structurels de l'image. Les  
20 supports structurels sont les éléments qui composent la scène de l'image. Les plus significatifs sont ceux qui délimitent les objets de la scène car ce sont eux qui caractérisent les différentes formes qui sont perçues lorsqu'on observe une quelconque image.

Cette étape concerne l'extraction de ces supports structurels. Elle  
25 consiste en un démantèlement des zones frontières des objets de l'image, qui sont caractérisées par des endroits entre deux zones où de fortes variations d'intensité sont observées. Ce démantèlement s'opère par un procédé qui consiste à répartir ces zones frontières parmi différentes  
30 « classes » selon l'orientation locale du gradient de l'image (orientation de la variation locale d'intensité). On obtient ainsi une multitude de petits éléments dénommés les « *Eléments de Support structurels* » (*ESS*). Chaque *ESS* appartenant effectivement à un contour d'une scène est caractérisé par une similarité au niveau de l'orientation locale de son gradient. Ceci est une première étape qui vise à répertorier tous les  
35 *éléments de Support structurels* de l'image.

La démarche suivante s'opère désormais à partir de ces *ESS*, à savoir la construction de *termes* décrivant les propriétés locales et globales des *ESS*.

5 Sont considérées comme propriétés locales les informations extraites de chaque support. Deux types de supports peuvent être distingués : les éléments de droites rectilignes (EDR) et les éléments d'arcs de courbes (EAC).

Les éléments de droites rectilignes EDR sont caractérisés par les propriétés locales qui sont :

- 10
- La dimension (longueur, largeur)
  - Direction principale (pente)
  - Propriétés statistiques des pixels constituant le support (valeur moyenne d'énergie, les moments)
  - 15 ▪ Informations du voisinage ( Transformée de Fourier locale)

Les éléments d'arcs de courbes EAC quant à eux sont caractérisés de la même façon que précédemment, en plus de la courbure des arcs.

20 Les propriétés globales englobent les statistiques telles que le nombre de chaque type de supports et leurs dispositions spatiales (associations géométriques entre les supports : connexités, gauche, droite, milieux...).

En résumé, pour une image donnée, les informations pertinentes extraites des objets la constituant sont regroupées sur le tableau 1.

Supports structuraux des objets d'une image		Type		
		ESS	EDR	EAC
<b>Propriétés globales</b>	<i>Nb total</i>	$n$	$n_1$	$n_2$
	<i>Nb longs (&gt; seuil)</i>	$nl$	$n_1l$	$n_2l$
	<i>Nb courts (&lt; seuil)</i>	$nc$	$n_1c$	$n_2c$
	<i>Nb de supports longs à une connexion gauche ou droite</i>	–	$n_1lgdx$	$n_2lgdx$
	<i>Nb d'une connexion milieu</i>	–	$n_1lgdx$	$n_2lgdx$
	<i>Nb de supports longs parallèles</i>	–	$n_1pll$	$n_2pll$
<b>Propriétés locales</b>	<i>Luminance (&gt;seuil)</i>	–		
	<i>Luminance (&lt;seuil)</i>	–		
	<i>Pente</i>	–		
	<i>Courbure</i>	–		
	<i>Caractérisation du voisinage des supports</i>	–		

Tableau 1

La phase de construction des termes d'une image met en œuvre également la caractérisation des informations texturales pertinentes de

l'image. Les informations venant de la texture de l'image sont divisées selon trois aspects visuels de l'image :

- 5           ▪ l'aspect aléatoire (comme une image de sable fin, ou d'herbe) où aucun arrangement particulier ne peut être décelé,
  - l'aspect périodique (comme un pull en jacquard) où une répétition de motifs (pixel ou groupement de pixels) dominants est observée,
  - et enfin l'aspect directionnel où les motifs tendent globalement à s'orienter vers une ou des directions privilégiées.
- 10           Ces informations sont obtenues en approchant l'image par des modèles ou représentations paramétriques. Chaque aspect est pris en compte par ses représentations spatiale et spectrale qui constituent les informations pertinentes de cette partie de l'image. La périodicité et l'orientation sont caractérisées par les supports spectraux tandis que
- 15 l'aspect aléatoire se traduit par l'estimation des paramètres d'un modèle autorégressif bidimensionnel.

Une fois toutes les informations pertinentes extraites, on peut procéder à la structuration des termes des textures.

<b>Supports spectraux et paramètres autorégressifs de la texture d'une image</b>		
<b>Composante périodique</b>	<i>Nb total d'éléments périodiques</i>	$np$
	<i>Fréquences</i>	<i>Couple <math>(\omega_p, \nu_p)</math>, <math>0 &lt; p \leq np</math></i>
	<i>Amplitudes</i>	<i>Couple <math>(C_p, D_p)</math>, <math>0 &lt; p \leq np</math></i>
<b>Composante directionnelle</b>	<i>Nb total d'éléments directionnels</i>	$nd$
	<i>Orientations</i>	<i>Couple <math>(\alpha_i, \beta_i)</math>, <math>0 &lt; i \leq nd</math></i>
	<i>Fréquences</i>	$\nu_i, 0 < i \leq nd$
<b>Composantes aléatoires</b>	<i>Ecart-type du bruit</i>	$\sigma$
	<i>Paramètres autorégressifs</i>	$\{a_{i,j}\} (i,j) \in S_{N,M}$

Tableau 2

La phase de construction des termes d'une image peut enfin mettre également en œuvre la caractérisation de la couleur de l'image.

- 5 La couleur est souvent représentée par les histogrammes de couleur, ces derniers sont invariants à la rotation et robustes contre l'occlusion et les changements de points de vue de la caméra.

- La quantification des couleurs peut se faire dans l'espace RVB (Rouge, Vert, Bleu), TSV (Teinte Saturation Valeur), ou l'espace LUV mais  
10 la méthode d'indexation par les histogrammes de couleurs a prouvé ses

limites car elle donne une information globale de l'image, et lors de l'indexation on peut trouver des images ayant le même histogramme de couleur, mais qui sont complètement différentes.

5 Beaucoup d'auteurs proposent des histogrammes de couleurs en intégrant l'information spatiale. Ceci consiste par exemple à distinguer les pixels cohérents des pixels incohérents, un pixel est cohérent s'il appartient à une région assez large regroupant des pixels identiques, il est classé incohérent s'il fait partie d'une région de taille réduite.

10 On décrira plus loin un procédé de caractérisation de la distribution spatiale des constituants de l'image (par exemple la couleur) qui est moins coûteux en temps de calcul que les méthodes citées ci dessus, et est robuste aux rotations et à la translation.

15 Les différentes caractéristiques extraites des éléments de support structurels ainsi que les paramètres des composants périodique, directionnel et aléatoire du champ de texture ainsi que les paramètres de la distribution spatiale des constituants de l'image constituent les *termes* pouvant servir à la description du contenu d'un document. Ces termes sont regroupés dans des *concepts* afin de réduire les *informations utiles* d'un document.

20 Les occurrences de ces concepts ainsi que leurs positions et leurs fréquences constituent ce qu'on appelle l'*empreinte* d'un document. Ces empreintes vont ensuite servir de trait d'union entre un document question et les documents d'une base, lors d'une phase de recherche de document.

25 Une image ne contient pas forcément tous les éléments et les caractéristiques décrits plus haut. Par conséquent, identifier une image commence par la détection de la présence de ses éléments constituants.

30 La Figure 9 montre un exemple d'organigramme d'un processus d'extraction des termes d'une image avec une première étape 71 de caractérisation des objets de l'image en supports structurels, qui peut le cas échéant être précédé d'un test de détection d'éléments structurels permettant d'omettre cette étape 71 dans les cas où les éléments structurels sont absents.

35 L'étape 72 consiste en un test pour déterminer s'il existe un fond texturé. Si c'est le cas, on passe à une étape 73 de caractérisation du fond texturé en supports spectraux et paramètres autorégressifs AR, puis à une étape 74 de caractérisation de la couleur de fond.

S'il n'existe pas de fond structuré, on passe directement de l'étape 72 à l'étape 74.

Enfin, une étape 75 réside dans le stockage des termes et la construction d'empreintes.

- 5 On reviendra maintenant de façon plus détaillée sur la caractérisation des éléments de support structurels d'une image.

Le principe de base de cette caractérisation consiste en un démantèlement des zones frontières des objets de l'image en multitudes de petits éléments de base appelés *éléments de supports significatifs* (ESS) qui véhiculent les *informations utiles* des zones frontières qui sont composées de bandes linéaires de taille variable, ou des coudes de différentes courbures. Des statistiques faites sur ces objets seront alors analysées et utilisées pour construire *les termes de ces supports structurels*.

10

- 15 Afin de décrire plus rigoureusement les principaux procédés composant cette approche, on notera une image numérisée par l'ensemble  $\{y(i, j), (i, j) \in I \times J\}$ , où  $I$  et  $J$  sont respectivement le nombre de lignes et de colonnes de l'image.

A partir des images *gradient vertical*  $\{g_v(i, j), (i, j) \in I \times J\}$  et *horizontal*  $\{g_h(i, j), (i, j) \in I \times J\}$  préalablement calculées, cette approche consiste à partitionner l'image selon l'orientation locale de son gradient en un nombre fini de classes équidistantes. L'image contenant l'orientation du gradient est définie par la formule :

20

$$O(i, j) = \arctan\left(\frac{g_h(i, j)}{g_v(i, j)}\right) \quad (1)$$

- 25 La partition n'est autre qu'une subdivision angulaire du plan 2D (de  $0^\circ$  à  $360^\circ$ ) par un pas de discrétisation bien défini. Le fait d'utiliser l'orientation locale du gradient comme critère de décomposition des zones frontières permet un meilleur groupement des pixels faisant partie d'une même zone frontière. Afin de résoudre le problème des points frontières qui peuvent être partagés entre deux *classes* juxtaposées, une deuxième partition avec le même nombre de *classes* que précédemment, mais décalées de  $\frac{1}{2}$  *classe* est utilisée. A partir des classes issues des deux
- 30

partitions, une simple procédure consiste à choisir celles qui totalisent le plus grand nombre de pixels. En effet, chaque pixel appartient à deux classes chacune issue des deux partitions. Sachant que chaque pixel est un élément potentiel d'un éventuel ESS, il vote alors pour la classe qui contient le plus de pixels parmi les deux. Il s'agit d'une région où la probabilité de trouver un ESS de taille plus élevée est la plus forte possible. A la suite des votes, on retient uniquement les classes qui totalisent plus de 50% des suffrages. Ce sont les *régions de support* susceptibles de contenir les ESS.

10 A partir de ces *régions de support*, on détermine les ESS, on les répertorie selon certains critères qui peuvent être :

- La longueur (on détermine pour cela un seuil  $l_0$  et on comptabilise les ESS inférieurs et supérieurs à ce seuil)
- L'intensité définie par la moyenne du module du gradient des pixels composant chaque ESS. ( un seuil noté  $I_0$  est alors défini, on répertorie ceux qui sont inférieurs et supérieurs à ce seuil).
- Le contraste défini par la différence entre le maximum et le minimum des pixels.

20 A cette étape du procédé, tous les éléments dits structurels sont connus et répertoriés conformément aux types de supports structurels pré-identifiés. Ils peuvent être extraits de l'image d'origine pour laisser place à la caractérisation du champ de textures.

25 A titre d'exemple, considérons l'image 81 de la Figure 11, reprise comme image 101 de la Figure 14a, les zones frontières sont illustrées sur l'image 102 de la Figure 14b. Les éléments de ces zones frontières sont ensuite démantelés et répartis selon l'orientation de leur gradient parmi différentes classes représentées par les images 103 à 106 des Figures 14c à 14f.. Ces différents éléments constituent les éléments de supports significatifs, et leurs analyses statistiques permettent de  
30 construire les termes du composant structurel.

Dans le cas des Figures 14c à 14f, à titre d'exemple, l'image 103 correspond à une classe 0 ( $0^\circ - 45^\circ$ ), l'image 104 correspond à une classe 1 ( $45^\circ - 90^\circ$ ), l'image 105 correspond à une classe 2 ( $90^\circ - 135^\circ$ ) et l'image 106 correspond à une classe 3 ( $135^\circ - 180^\circ$ ).

En l'absence d'éléments structurels, on suppose que l'image est texturée avec des motifs plus ou moins réguliers et on procède à une caractérisation du champ de la texture. Pour cela, on peut procéder à une décomposition de l'image en trois composantes qui sont :

- 5           ▪ Une composante texturale contenant les informations anarchiques ou aléatoires (comme une image de sable fin, ou d'herbe) où aucun arrangement particulier ne peut être décelé,
- Une composante périodique (comme un pull en jacquard) où une répétition de motifs dominants est observée,
- 10           ▪ et enfin une composante directionnelle où les motifs tendent globalement vers une ou des directions privilégiées.

L'objectif étant de caractériser parfaitement la texture de l'image à partir d'un ensemble de paramètres, ces trois composantes sont représentées par des modèles paramétriques.

15           Ainsi, la texture de l'image 15 régulière et homogène notée  $\{\tilde{y}(i, j), (i, j) \in I \times J\}$  est décomposée en trois composantes 16, 17, 18 comme illustré sur la Figure 10, conformément à la relation suivante :

$$\{\tilde{y}(i, j)\} = \{w(i, j)\} + \{h(i, j)\} + \{e(i, j)\} \quad (16)$$

20

Où  $\{w(i, j)\}$  est la composante purement aléatoire 16,  $\{h(i, j)\}$  est la composante harmonique 17 et  $\{e(i, j)\}$  la composante directionnelle 18. L'estimation des paramètres de ces trois composantes 16, 17, 18 termine cette étape d'extraction d'informations d'un document. Des méthodes

25 d'estimation sont décrites dans les paragraphes suivants.

On décrira d'abord un exemple de procédé de détection et caractérisation de la composante directionnelle de l'image.

Il s'agit dans un premier temps d'appliquer un modèle paramétrique à la composante directionnelle  $\{e(i, j)\}$ . Elle est constituée d'une somme

30 dénombrable d'éléments directionnels où chacun est associé à un couple d'entiers  $(\alpha, \beta)$  définissant une orientation d'angle  $\theta$  tel que  $\theta = \tan^{-1} \beta/\alpha$ . Autrement dit,  $e(i, j)$  est défini par  $e(i, j) = \sum_{(\alpha, \beta) \neq 0} e_{(\alpha, \beta)}(i, j)$  où chaque  $e_{(\alpha, \beta)}(i, j)$

est défini par :

$$e_{(\alpha,\beta)}(i,j) = \sum_{k=1}^{N_e} \left[ s_k^{\alpha,\beta}(i\alpha - j\beta) \times \cos\left(2\pi \frac{v_k}{\alpha^2 + \beta^2}(i\beta + j\alpha)\right) + t_k^{\alpha,\beta}(i\alpha - j\beta) \times \sin\left(2\pi \frac{v_k}{\alpha^2 + \beta^2}(i\beta + j\alpha)\right) \right] \quad (17)$$

où

- $N_e$  est le nombre d'éléments directionnels associés à  $(\alpha, \beta)$ ,
- $v_k$  est la fréquence du  $k^{\text{ième}}$  élément,
- $\{s_k(i\alpha - j\beta)\}$  et  $\{t_k(i\alpha - j\beta)\}$  sont les amplitudes.

La composante directionnelle  $\{e(i,j)\}$  est ainsi parfaitement définie par la connaissance des paramètres contenus dans le vecteur  $E$  suivant :

$$E = \left\{ \alpha_l, \beta_l, \left\{ v_{lk}, s_{lk}(c), t_{lk}(c) \right\}_{k=1}^{N_e} \right\}_{l=0} \quad (18)$$

Pour estimer ces paramètres, on utilise le fait que la composante directionnelle d'une image est représentée dans le domaine spectral par un ensemble de droites de pentes orthogonales à celles définies par les couples d'entiers  $(\alpha_l, \beta_l)$  du modèle qui seront notés  $(\alpha_l, \beta_l)^\perp$ . Ces droites peuvent être décomposées en un sous ensemble de droites de même pente associé chacun à un élément directionnel.

A titre d'illustration, les Figures 15a et 15b montrent des images 84, 86 contenant un élément directionnel et la Figure 15c montre une image 88 contenant deux éléments directionnels.

La figure 15a1 montre une vue 85 en trois dimensions du spectre de l'image 84 de la Figure 15a.

Les Figures 15b1 et 15c1 montrent des images module de Fourier 87, 89 respectivement des images 86 et 85 des Figures 15b et 15c.

Pour calculer les éléments du vecteur  $E$ , on peut adopter une approche fondée sur la projection de l'image suivant différentes directions. Le procédé consiste dans un premier temps à s'assurer de la présence de la composante directionnelle avant d'estimer ses paramètres.

La détection de la composante directionnelle de l'image est basée sur la connaissance des propriétés spectrales de celle-ci. Si on assimile le spectre de l'image à une image 3D  $(X, Y, Z)$ , où  $(X, Y)$  représentent les coordonnées des pixels et  $Z$  l'amplitude, les droites qu'on cherche à détecter sont représentées par un ensemble de pics concentrés le long de

droites dont les pentes sont définies par les couples  $(\alpha_l, \beta_l)$  recherchés (cf. figure 15a1). Pour déterminer la présence de ces droites, il suffit de comptabiliser les pics prédominants. Le nombre de ces pics renseigne sur la présence ou non de supports directionnels ou harmoniques.

5 On décrira maintenant un exemple de procédé de caractérisation de la composante directionnelle. Pour cela, on procède au calcul des couples de direction  $(\alpha_l, \beta_l)$  et à la détermination du nombre d'éléments directionnels.

10 On effectue d'abord le calcul de la transformée de Fourier Discrète (TFD) de l'image suivi d'une estimation des droites de pente rationnelle observées dans l'image transformée  $\Psi(i, j)$ .

Pour cela, on définit un ensemble de projections qui discrétise le domaine fréquentiel en différents angles de projection  $\theta_k$ ,  $k$  fini. Cet ensemble de projection peut être obtenu de différentes manières. On peut  
15 par exemple chercher tous couples d'entiers premiers entre eux  $(\alpha_k, \beta_k)$  définissant un angle  $\theta_k$ , tel que  $\theta_k = \tan^{-1} \frac{\alpha_k}{\beta_k}$  où  $0 \leq \theta_k \leq \frac{\pi}{2}$ . Un ordre  $r$  tel que  $0 \leq \alpha_k, \beta_k \leq r$  permet de contrôler le nombre de projections. Les propriétés de symétrie peuvent ensuite être utilisées pour obtenir tous les couples jusqu'à  $2\pi$ . Ces couples sont illustrés sur la Figure 16 pour  
20  $0 \leq \alpha_k, \beta_k \leq 3$ .

Des projections du module de la TFD de l'image sont effectuées suivant les  $\theta_k$ . Chaque projection engendre un vecteur de dimension 1,  $V_{(\alpha_k, \beta_k)}$ , noté  $V_k$  pour simplifier la notation, qui contient les informations directionnelles recherchées.

25 Chaque projection  $V_k$  est donnée par la formule :

$$V_k(n) = \sum_{\tau} \Psi(i + \tau\beta_k, j + \tau\alpha_k), \quad 0 < i + \tau\beta_k < I - 1, 0 < j + \tau\alpha_k < J - 1 \quad (19)$$

avec  $n = -i * \beta_k + j * \alpha_k$  et  $0 \leq |n| < N_k$  et  $N_k = |\alpha_k|(T-1) + |\beta_k|(L-1) + 1$ , où  $T * L$  est la taille de l'image.  $\Psi(i, j)$  est le module de transformée de Fourier de l'image à caractériser.

On sélectionne pour chaque  $V_k$  les éléments de fortes énergies ainsi que leurs positions spatiales. Ces éléments de forte énergie sont ceux qui présentent une valeur maximale par rapport à un *seuil* calculé selon la taille de l'image.

A cette étape de calcul, le nombre de droites est connu. On en déduit le nombre de composantes directionnelles  $N_e$  en utilisant les propriétés spectrales simples de la composante directionnelle d'une image texturée. Ces propriétés sont :

1. Les droites observées dans le domaine spectral d'une composante directionnelle sont symétriques par rapport à l'origine. On peut par conséquent réduire le domaine d'investigation sur une moitié seulement du domaine considéré.
2. Les maxima retenus dans le vecteur sont candidats à représenter des droites appartenant à des éléments directionnels. A partir de la connaissance des positions respectives des droites sur le module de transformée de Fourier discrète TFD, on en déduit le nombre exact d'éléments directionnels. La position du maximum droite correspond à l'argument du maximum du vecteur  $V_k$ , les autres droites du même élément sont situées tous les  $\min\{L, T\}$ .

Le mécanisme de projection est illustré sur la Figure 17 pour  $(\alpha_k, \beta_k) = (2, -1)$ .

Après traitement des vecteurs  $V_k$  et production des couples de direction  $(\hat{\alpha}_k, \hat{\beta}_k)$  on obtient les nombres de droites associés à chaque couple.

Ainsi on peut compter le nombre total d'éléments directionnels en utilisant les deux propriétés sus-mentionnées et on identifie les couples d'entiers  $(\hat{\alpha}_k, \hat{\beta}_k)$  associés à ces composantes qui sont les directions orthogonales à celles qui ont été retenues.

Pour tous ces couples  $(\hat{\alpha}_k, \hat{\beta}_k)$  l'estimation des fréquences de chaque élément détecté est immédiate. En effet, si l'on considère uniquement les points de l'image d'origine le long de la droite d'équation  $i\hat{\alpha}_k - j\hat{\beta}_k = c$ ,  $c$  est la position du maximum dans  $V_k$ , et ces points

5 constituent un signal monodimensionnel (1-D) harmonique d'amplitude constante et qui a pour fréquence  $\hat{\nu}_i^{(\alpha, \beta)}$ . Il suffit alors d'estimer la fréquence de ce signal 1-D par un procédé classique (localisation de la valeur maximale sur la TFD 1-D de ce nouveau signal).

En résumé, on peut mettre en œuvre le procédé comportant les

10 étapes suivantes :

On détermine le maximum de chaque projection.

Les maximums sont filtrés afin de ne garder que ceux supérieurs à un seuil.

- 15
  - Pour chaque maximum  $m_i$ , correspondant à un couple  $(\hat{\alpha}_k, \hat{\beta}_k)$
  - on détermine le nombre de droites associées à ce couple d'après les propriétés décrites plus haut.
  - on calcule la fréquence associée à  $(\hat{\alpha}_k, \hat{\beta}_k)$  qui correspond à l'intersection de la droite maximale (correspondant au
- 20 maximum de la projection retenue) avec l'axe horizontal.

On décrira maintenant le calcul des amplitudes  $\{\hat{s}_k^{(\alpha, \beta)}(t)\}$  et  $\{\hat{t}_k^{(\alpha, \beta)}(t)\}$  qui sont les autres paramètres contenus dans le vecteur E mentionné plus haut.

25 Connaissant la direction  $(\hat{\alpha}_k, \hat{\beta}_k)$  et la fréquence  $V_k$ , on peut déterminer les amplitudes  $\hat{s}_k^{(\alpha, \beta)}(c)$  et  $\hat{t}_k^{(\alpha, \beta)}(c)$ , pour  $c$  vérifiant la formule  $i\hat{\alpha}_k - j\hat{\beta}_k = c$ , en utilisant un procédé de démodulation. En effet,  $\hat{s}_k^{(\alpha, \beta)}(c)$  est égal à la moyenne des pixels le long de la droite d'équation

$i\hat{\alpha}_k - j\hat{\beta}_k = c$  de la nouvelle image obtenue en multipliant  $\tilde{y}(i, j)$  par  $\cos\left(\frac{\hat{v}_k^{(\alpha, \beta)}}{\hat{\alpha}_k^2 + \hat{\beta}_k^2}(i\hat{\beta}_k + j\hat{\alpha}_k)\right)$ . Ceci est traduit par l'équation

$$\hat{s}_k^{(\alpha, \beta)}(c) \equiv \frac{1}{N_s} \sum_{i\hat{\alpha} - j\hat{\beta} = c} \tilde{y}(i, j) \cos\left(\frac{\hat{v}_k^{(\alpha, \beta)}}{\hat{\alpha}_k^2 + \hat{\beta}_k^2}(i\hat{\beta}_k + j\hat{\alpha}_k)\right) \quad (20)$$

où  $N_s$  n'est autre que le nombre d'éléments de ce nouveau signal.

5 De la même manière, on obtient  $\hat{t}_k^{(\alpha, \beta)}(c)$  en appliquant l'équation :

$$\hat{t}_k^{(\alpha, \beta)}(c) \equiv \frac{1}{N_s} \sum_{i\hat{\alpha} - j\hat{\beta} = c} \tilde{y}(i, j) \sin\left(\frac{\hat{v}_k^{(\alpha, \beta)}}{\hat{\alpha}_k^2 + \hat{\beta}_k^2}(i\hat{\beta}_k + j\hat{\alpha}_k)\right) \quad (21)$$

Le procédé décrit ci-dessus peut être résumé par les étapes suivantes :

Pour tout élément directionnel  $(\hat{\alpha}_k, \hat{\beta}_k)$  faire

Pour toute droite (d) calculer

1. la moyenne des points (i,j) pondérée par le

$\cos\left(\frac{\hat{v}_k^{(\alpha, \beta)}}{\hat{\alpha}_k^2 + \hat{\beta}_k^2}(i\hat{\beta}_k + j\hat{\alpha}_k)\right)$ . Cette moyenne correspond à l'estimation de l'amplitude  $\hat{s}_k^{(\alpha, \beta)}(d)$

2. la moyenne des points (i,j) pondérée par le

$\sin\left(\frac{\hat{v}_k^{(\alpha, \beta)}}{\hat{\alpha}_k^2 + \hat{\beta}_k^2}(i\hat{\beta}_k + j\hat{\alpha}_k)\right)$ . Cette moyenne correspond à l'estimation de l'amplitude  $\hat{t}_k^{(\alpha, \beta)}(d)$

10

Le tableau 3 ci-dessous récapitule les étapes principales du procédé de projection.

<p>Etape 1. Calculer l'ensemble de couples de projection <math>(\alpha_k, \beta_k) \in P_r</math></p> <p>Etape 2. Calculer le module de la TFD de l'image <math>\tilde{y}(i, j)</math> :</p> $\Psi(\omega, \nu) =  TFD(y(i, j)) $
<p>Etape 3- Pour tout <math>(\alpha_k, \beta_k) \in P_r</math>,</p> <p>calculer le vecteur <math>V_k</math> : la projection de <math>\Psi(\omega, \nu)</math> suivant <math>(\alpha_k, \beta_k)</math> d'après la formule ( 19)</p>
<p>Etape 4- Détection de droites :</p> <p>Pour tout <math>(\alpha_k, \beta_k) \in P_r</math>,</p> <ul style="list-style-type: none"> <li>• déterminer : <math>M_k = \max_j \{V_k(j)\}_r</math>,</li> <li>• calculer <math>n_k</math> : le nombre de pixels de valeurs significatives rencontrés le long de la projection</li> <li>• sauvegarder <math>n_k</math> et <math>j_{\max}</math> l'indice du maximum dans <math>V_k</math>.</li> <li>• sélectionner les directions qui justifient le critère</li> </ul> $\frac{M_k}{n_k} > s_e$ <p>où <math>s_e</math> est un seuil à définir, dépendant de la taille de l'image,</p> <p>Les directions retenues sont considérées comme celles de droites recherchées.</p>
<p>Etape 5- Sauvegarder les couples <math>(\hat{\alpha}_k, \hat{\beta}_k)</math> recherchés qui sont les orthogonaux des couples <math>(\alpha_k, \beta_k)</math> retenus dans l'étape 4.</p>

**Tableau 3**

5 On décrira ci-dessous la détection et la caractérisation des informations texturales périodiques d'une image, qui sont contenues dans la composante harmonique  $\{h(i, j)\}$ . Cette composante peut être représentée par une somme finie de sinusoïdes 2-D :

$$h(i, j) = \sum_{p=1}^P C_p \cos 2\pi(i\omega_p + j\nu_p) + D_p \sin 2\pi(i\omega_p + j\nu_p), \quad (22)$$

où

- $C_p$  et  $D_p$  sont les amplitudes.
- $(\omega_p, \nu_p)$  est la  $p^{\text{ème}}$  fréquence spatiale.

On a représenté sur la Figure 18a1 une image 91 contenant des  
5 composantes périodiques et, sur la Figure 18b1, une image synthétique  
contenant une composante périodique.

La Figure 18a2 représente une image 92 qui est une image module  
de la TFD présentant un ensemble de pics.

La Figure 18b2 représente une vue 3D, 94, de la TFD qui montre la  
10 présence d'une paire de pics symétriques 95, 96.

Sur le domaine spectral, la composante harmonique apparaît ainsi  
comme des paires de pics isolés symétriques par rapport à l'origine (cf.  
Figure 18 (a2)-(b2)). Cette composante est le reflet de l'existence de  
périodicités dans l'image.

15 Les informations que l'on cherche à déterminer sont les éléments  
du vecteur :

$$H = \left\{ P, \left\{ C_p, D_p, \omega_p, \nu_p \right\}_{p=1}^P \right\} \quad (23)$$

Pour cela on commence par détecter la présence de cette  
composante périodique dans l'image module de Fourier et on estime  
20 ensuite ses paramètres.

La détection de la composante périodique consiste à déterminer la  
présence de pics isolés sur l'image module de la TFD. On opère de la  
même manière que dans le cas de la détermination de la composante  
directionnelle. D'après le procédé décrit dans le tableau 1, si la valeur  $n_k$   
25 *obtenue dans la phase 4 du procédé* décrit dans le tableau 1 est  
inférieure à un seuil, alors on est en présence de pics isolés qui  
caractérisent la présence de composante harmonique plutôt que de pics  
formant une droite continue.

La caractérisation de la composante périodique revient à localiser  
30 les pics isolés sur l'image module de la TFD.

Ces fréquences spatiales  $(\hat{\omega}_p, \hat{\nu}_p)$  correspondent à la position de ces pics :

$$(\hat{\omega}_p, \hat{\nu}_p) = \arg \max_{(\omega, \nu)} \Psi(\omega, \nu) \quad (24)$$

Pour le calcul des amplitudes  $(\hat{C}_p, \hat{D}_p)$ , on utilise un procédé de démodulation comme pour estimer les amplitudes de la composante directionnelle.

Pour chaque élément périodique de fréquence  $(\hat{\omega}_p, \hat{\nu}_p)$ , l'amplitude correspondante est identique à la moyenne des pixels de la nouvelle image obtenue en multipliant l'image  $\{\tilde{y}(i, j)\}$  par  $\cos(i\hat{\omega}_p + j\hat{\nu}_p)$ . Ceci est traduit par les formules suivantes :

$$\hat{C}_p = \frac{1}{L \times T} \sum_{n=0}^{L-1} \sum_{m=0}^{T-1} y(n, m) \cos(n\hat{\omega}_p + m\hat{\nu}_p). \quad (25)$$

$$\hat{D}_p = \frac{1}{L \times T} \sum_{n=0}^{L-1} \sum_{m=0}^{T-1} y(n, m) \cos(n\hat{\omega}_p + m\hat{\nu}_p) \quad (26)$$

En résumé, un procédé d'estimation de la composante périodique comprend les étapes suivantes :

**Etape 1.** Localiser les pics isolés dans la deuxième moitié de l'image module de Fourier et compter leur nombre

**Etape 2.** Pour tout pic détecté :

- Calculer sa fréquence à l'aide de la formule (24)
- Calculer son amplitude à l'aide de la formule (25 - 26)

Les dernières informations à extraire sont contenues dans la composante purement aléatoire  $\{w(i, j)\}$ . Cette composante peut être

représentée par un modèle autorégressif 2-D de support demi-plan non symétrique (DPNS) fini défini par l'équation aux différences suivante:

$$w(i, j) = - \sum_{(k,l) \in S_{N,M}} a_{k,l} w(i-k, j-l) + u(i, j) \quad (27)$$

où  $\{a_{(k,l)}\}_{(k,l) \in S_{N,M}}$  sont les paramètres à déterminer pour tout  $(k,l)$  appartenant à  $S_{N,M} = \{(k,l)/k=0, 1 \leq l \leq M\} \cup \{(k,l)/1 \leq k \leq N, -M \leq l \leq M\}$ . Le couple  $(N,M)$  s'appelle l'ordre du modèle.

- $\{u(i, j)\}$  est un bruit blanc gaussien de variance finie  $\sigma_u^2$ .

Les paramètres du modèle sont donnés par :

$$W = \{(N, M), \sigma_u^2, \{a_{k,l}\}_{(k,l) \in S_{N,M}}\} \quad (28)$$

Les méthodes d'estimation des éléments de  $W$  sont nombreuses comme par exemple l'algorithme de Levinson 2D ou alors les méthodes adaptatives du type moindres carrés (MCR).

On décrira maintenant un procédé de caractérisation de la couleur d'une image dont on veut extraire des termes  $t_i$  représentant des caractéristiques iconiques de cette image, la couleur étant un exemple particulier de ces caractéristiques qui peuvent comprendre d'autres caractéristiques telles que les moments algébriques ou géométriques, les propriétés statistiques, les propriétés spectrales des moments de pseudo-Zernicke.

La méthode est fondée sur la caractérisation perceptuelle de la couleur. En premier lieu on effectue une transformation des composantes de la couleur de l'image de l'espace RVB (Rouge, Vert, Bleu) encore dénommé *RGB*, vers l'espace *TSV* (Teinte Saturation Valeur) encore dénommé *HSV*. On obtient ainsi trois composantes : *Teinte*, *Saturation*, *Valeur*. A partir de ces trois composantes on détermine  $N$  couleurs ou composantes iconiques de l'image. Chaque composante iconique  $C_i$  est représentée par un vecteur de  $M$  valeurs. Ces valeurs représentent la distribution angulaire et annulaire des points représentant chaque composante ainsi que le nombre de points de la composante en question.

La méthode développée est illustrée par la Figure 19 avec à titre d'exemple  $N = 16$  et  $M = 17$ .

Dans une première étape principale 110, à partir d'une image 11 de l'espace RGB, on procède à une transformation de l'image 111 de l'espace (R,G,B) à l'espace HSV (étape 112) pour obtenir une image dans l'espace HSV.

Le modèle HSV peut être défini de la façon suivante.

**Teinte (H) :** varie de [0 360] , et chaque angle représente une teinte.

**Saturation (S) :** varie de [0 1], elle mesure la pureté des couleurs, et permet de distinguer les couleurs "vives", "pastels", ou "délavées".

**Valeur (V) :** Elle prend des valeurs de [0 1], elle indique si une couleur est claire ou sombre et dans quelle mesure elle se rapproche du blanc ou du noir.

Le modèle *HSV* est une transformation non linéaire du modèle de l'espace (R,G,B). L'œil humain peut distinguer 128 teintes, 130 saturations, et 23 ombres.

Pour le *Blanc*  $V=1$  et  $S=0$ , le *noir* a une valeur  $V=0$  tandis que la teinte  $H$  et la saturation  $S$  sont indéterminées.

Lorsque  $V=1$  et  $S=1$  on a une couleur pure.

Chaque couleur est obtenue en rajoutant du blanc ou du noir à la couleur pure.

Pour avoir des couleurs plus claires on réduit  $S$  et on garde  $H$  et  $V$ , par contre, pour des couleurs foncées on rajoute du noir en réduisant  $V$  et on garde  $H$  et  $S$ .

Le passage de l'image couleur exprimée dans les coordonnées (R,G,B) en une image exprimée dans l'espace (H,S,V) (Teinte, Saturation, Valeur) s'effectue de la façon suivante :

Pour tout point de coordonnées  $(i,j)$  et de valeur  $(R_k, B_k, G_k)$  on produit un point de coordonnées  $(i,j)$  et valeur  $(H_k, S_k, V_k)$  avec :

$$\begin{aligned}
 5 \quad & V_k = \max(R_k, B_k, G_k) \\
 & S_k = \frac{V_k - \min(R_k, G_k, B_k)}{V_k} \\
 10 \quad & H_k = \begin{cases} \frac{G_k - B_k}{V_k - \min(R_k, G_k, B_k)} & \text{si } V_k \text{ est égal à } R_k \\ 2 + \frac{B_k - R_k}{V_k - \min(R_k, G_k, B_k)} & \text{si } V_k \text{ est égal à } G_k \\ 4 + \frac{R_k - G_k}{V_k - \min(R_k, G_k, B_k)} & \text{si } V_k \text{ est égal à } B_k \end{cases} \\
 15 \quad &
 \end{aligned}$$

On procède ensuite à une partition de l'espace HSV (étape 113).

A partir des valeurs de Teinte, Saturation, Valeur , on a défini N couleurs.  
 20 Dans le cas où N est égal à 16, on a : Noir, Blanc, Gris clair, Gris foncé, Gris moyen, Rouge, Rose, Orange, Marron, Olive, Jaune, Vert, Bleu Ciel , Bleu-vert, Bleu , Pourpre, Magenta.

Pour chaque pixel on évalue à quelle couleur il appartient. Ensuite on calcule le nombre de points de chaque couleur.

25 Dans une deuxième étape principale 120, on procède à une caractérisation des partitions obtenues lors de la première étape principale 110.

Dans cette étape 120, on cherche à caractériser chaque partition Ci obtenue précédemment. Une partition est définie par sa composante iconique et par les coordonnées des pixels qui la constituent. La description d'une partition est fondée sur la caractérisation de la répartition spatiale de ces pixels (nuage de points). La méthode commence par le calcul du centre de gravité, l'axe principal du nuage de points et l'axe perpendiculaire à cet axe. Le nouveau repère est utilisé  
 30 comme référence dans la décomposition de la partition Ci en plusieurs sous-partitions qui sont représentées par le pourcentage des points  
 35

constituant chacune des sous-partitions. Le processus de caractérisation d'une partition  $C_i$  est le suivant :

- calculer le centre de gravité et l'angle d'orientation des composantes  $C_i$  définissant le repère de partitionnement,
- 5        - calculer la distribution angulaire des points de la partition  $C_i$  dans les  $N$  directions dans le sens contraire des aiguilles d'une montre, en  $N$  sous-partitions définies par :

$$\left(0^\circ, \frac{360}{N}, \frac{2 \times 360}{N}, \dots, \frac{i \times 360}{N}, \dots, \frac{(N-1) \times 360}{N}\right)$$

- 10        - partitionner l'espace de l'image en carrés de rayons concentriques, avec calcul dans chaque rayon du nombre de points correspondant à chaque composante iconique.

Le vecteur caractéristique est obtenu à partir du nombre de points de chaque répartition de couleur  $C_i$ , du nombre de points dans les 08  
15 sous-répartitions angulaires ainsi que du nombre de points de l'image.

Ainsi le vecteur caractéristique est représenté par 17 valeurs dans l'exemple considéré.

Sur la Figure 19, on a illustré la deuxième étape 120 de traitement à partir des composantes iconiques  $C_0$  à  $C_{15}$  en montrant pour les  
20 composantes  $C_0$  (module 121) et  $C_{15}$  (module 131) les différentes étapes effectuées, à savoir le partitionnement angulaire 122, 132 conduisant à un nombre de points dans les 8 orientations considérées (étape 123, 133) et le partitionnement annulaire 124, 134 conduisant à un nombre de points dans les 8 rayons considérés (étape 125, 135), ainsi que la prise en  
25 compte du nombre de pixels de composante  $C_0$  respectivement  $C_{15}$  dans l'image (étape 126, respectivement 136).

Les étapes 123, 125, 126 conduisent à la production de 17 valeurs pour la composante  $C_0$  (étape 127) tandis que les étapes 133, 135, 136 conduisent à la production de 17 valeurs pour la composante  $C_{15}$  (étape  
30 137).

Naturellement, le processus est analogue pour les autres composantes  $C_1$  à  $C_{14}$ .

Les Figures 20 et 21 illustrent le fait que le procédé décrit ci-dessus est invariant à la rotation.

Ainsi, dans l'exemple de la Figure 20, l'image est partitionnée en deux sous-ensembles, l'un contenant les croix x, l'autre les ronds 0. Après  
5 calcul du centre de gravité ainsi que de l'angle d'orientation  $\theta$ , on obtient le repère d'orientation qui permettra d'obtenir les 04 sous-répartitions angulaires ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ ).

Par la suite, on effectue une répartition annulaire, on calcule le nombre de points dans un rayon égal à 1 puis 2. On obtient le vecteur V0  
10 caractéristique de l'image de la Figure 20 : 19 ; 6 ; 5 ; 4 ; 4 ; 8 ; 11.

L'image de la Figure 21 est obtenue en appliquant une rotation de  $90^\circ$  à l'image de la Figure 20. En appliquant le procédé ci-dessus à l'image de la Figure 21, on obtient un vecteur V1 caractérisant cette dernière qui montre que la rotation n'influence par le vecteur caractéristique. Cela  
15 permet de conclure que la méthode est invariante à la rotation.

Comme indiqué plus haut, les méthodes permettant d'obtenir pour une image les termes représentant les couleurs dominantes, les propriétés texturales ou les structures des zones dominantes de l'image, peuvent être appliquées aussi bien sur la totalité de l'image que sur des portions de l'image.  
20

On décrira brièvement ci-dessous des processus de segmentation d'un document qui permettent de produire les portions de l'image à caractériser.

Selon une première technique possible, on procède à une décomposition statique. L'image est décomposée en blocs avec recouvrement ou sans recouvrement.  
25

Selon une deuxième technique possible, on procède à une décomposition dynamique. Dans ce cas, la décomposition de l'image en portions est fonction du contenu de l'image.

30 Selon un premier exemple de technique de décomposition dynamique, les portions sont produites à partir des germes qui sont les points de singularité de l'image (les points d'inflexions). On commence par calculer les germes, qui sont ensuite fusionnés pour qu'il n'en reste qu'un nombre réduit et enfin les points de l'images sont fusionnés aux germes

ayant les mêmes propriétés visuelles (statistiques) pour produire les portions ou les segments de l'image à caractériser.

5 Selon une autre technique faisant appel à une segmentation hiérarchique, les points de l'image sont fusionnés pour former les n premières classes. Ensuite les points de chacune des classes sont décomposés en m classes et ainsi de suite jusqu'à atteindre le nombre de classes désiré. Lors de la fusion, les points sont affectés à la classe la plus proche. Une classe est représentée par le centre de gravité et/ou un délimiteur (boîte englobante, segment, courbe, ...).

10 On décrira maintenant les étapes principales d'un procédé de caractérisation des formes d'une image.

La caractérisation de la forme s'effectue en plusieurs étapes :

15 Pour une suppression d'effet de zoom ou variation dû aux mouvement des éléments non rigides de l'image (mouvement des lèvres, des feuilles d'arbre, ...), on procède par une multirésolution suivie d'une décimation de l'image.

Pour une réduction de l'effet de translation, l'image ou la portion de l'image est représentée par sa Transformée de Fourier.

20 Pour une réduction de l'effet de zoom, l'image est définie dans l'espace logarithmique polaire.

On peut mettre en œuvre les étapes suivantes :

a/ multirésolution  $f = \text{wavelet}(I, n)$  ; où I est l'image de départ et n est le nombre de décompositions

b/ projection de l'image dans l'espace logPolaire:

25  $g(l, m) = f(i, j)$  avec  $i = l \cdot \cos(m)$  et  $j = l \cdot \sin(m)$

c/ calcul de la transformée de Fourier de g :  $H = \text{FFT}(g)$  ;

d/ caractérisation de H :

d1/ projection de H dans plusieurs directions (0, 45, 90 , ...) : le résultat est un ensemble de vecteurs dont la dimension est égale à la dimension du segment de projection

30 d2/ calcul des propriétés statistiques de chaque vecteur de projection (moyenne, variance, les moments).

Le terme représentant la forme est constitué des valeurs des propriétés statistiques de chaque vecteur de projection.

## REVENDEICATIONS

1. Procédé d'indexation de documents multimédias, caractérisé en  
5 ce qu'il comprend les étapes suivantes :
- (a) identification et extraction pour chaque document de termes  $t_i$   
constitués par des vecteurs caractérisant des propriétés du document  
multimédia à indexer, telles que la forme, la texture, la couleur ou la  
10 structure d'une image, l'énergie, le taux d'oscillation ou des  
informations fréquentielles d'un signal audio, ou un groupe de  
caractères d'un texte ,
  - (b) stockage des termes  $t_i$  caractérisant des propriétés du document  
multimédia dans une base de termes comprenant  $P$  termes,
  - 15 (c) détermination d'un nombre maximum  $N$  de concepts souhaités  
regroupant les termes  $t_i$  les plus pertinents,  $N$  étant un nombre  
entier inférieur à  $P$ , et chaque concept  $c_i$  étant prévu pour regrouper  
tous les termes voisins du point de vue de leurs caractéristiques,
  - (d) calcul de la matrice  $T$  de distances entre les termes  $t_i$  de la base de  
20 termes,
  - (e) décomposition de l'ensemble  $P$  des termes  $t_i$  de la base de termes en  
 $N$  parties  $P_j$  ( $1 \leq j \leq N$ ) telles que  $P = P_1 \cup P_2 \dots \cup P_j \dots \cup P_N$ , chaque  
partie  $P_j$  comprenant un ensemble de termes  $t_{ij}$  et étant représentée  
par un concept  $c_j$ , les termes  $t_i$  étant répartis de telle façon que les  
25 termes les plus éloignés se trouvent dans des parties  $P_l, P_m$  distinctes  
et les termes proches se trouvent dans la même partie  $P_l$ ,
  - (f) structuration du dictionnaire de concepts de manière à constituer un  
arbre binaire où les feuilles contiennent les concepts  $c_i$  du  
dictionnaire et les nœuds de l'arbres contiennent les informations  
30 nécessaires à la scrutation de l'arbre lors d'une phase d'identification  
d'un document par comparaison avec les documents précédemment  
indexés, et
  - (g) construction d'une base d'empreintes constituée de l'ensemble des  
concepts  $c_i$  représentant les termes  $t_i$  des documents à indexer,  
35 chaque document étant associé à une empreinte qui lui est propre.

2. Procédé d'indexation selon la revendication 1, caractérisé en ce que l'on associe à chaque concept  $c_i$  de la base d'empreintes un ensemble d'informations comprenant le nombre NbT de termes dans les documents où le concept  $c_i$  est présent.

5

3. Procédé d'indexation selon la revendication 1 ou la revendication 2, caractérisé en ce que pour chaque document où un concept  $c_i$  est présent, on enregistre une empreinte du concept  $c_i$  dans le document, cette empreinte comprenant la fréquence d'occurrence du concept  $c_i$ , l'identification des concepts qui sont voisins du concept  $c_i$  dans le document et un score qui est une valeur moyenne des mesures de similarités entre le concept  $c_i$  et les termes  $t_i$  du document qui sont les plus proches du concept  $c_i$ .

15

4. Procédé d'indexation selon l'une quelconque des revendications 1 à 3, caractérisé en ce qu'il comprend une étape d'optimisation de la partition de l'ensemble P des termes de la base de termes pour décomposer cet ensemble P en M classes  $C_i$  ( $1 \leq i \leq M$ , avec  $M \leq P$ ), de manière à réduire l'erreur de la répartition de l'ensemble P des termes de la base de termes en N parties ( $P_1, P_2, \dots, P_N$ ) où chaque partie  $P_i$  est représentée par le terme  $t_i$  qui sera pris comme concept  $c_i$ , l'erreur commise  $\varepsilon$  étant telle que  $\varepsilon = \sum_{i=1}^N \varepsilon_{t_i}$  où  $\varepsilon_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$  est l'erreur commise lorsqu'on remplace les termes  $t_j$  d'une partie  $P_i$  par  $t_i$ .

25

5. Procédé d'indexation selon la revendication 4, caractérisé en ce qu'il comprend les étapes suivantes :

- (i) on décompose l'ensemble P de termes en deux parties  $P_1$  et  $P_2$  ;
- (ii) on détermine les deux termes les plus éloignés  $t_i$  et  $t_j$  de l'ensemble P correspondant à la plus grande distance  $D_{ij}$  de la matrice T de distances ;
- (iii) pour chaque terme  $t_k$  de l'ensemble P, on examine si la distance  $D_{ki}$  entre le terme  $t_k$  et le terme  $t_i$  est plus petite que la distance  $D_{kj}$  entre le terme  $t_k$  et le terme  $t_j$ , si c'est le cas on affecte le terme  $t_k$  à la partie  $P_1$  et si ce n'est pas le cas on affecte le terme  $t_k$  à la partie  $P_2$ ;

30

(iv) on itère l'étape (i) jusqu'à l'obtention du nombre  $N$  de points  $P_i$  souhaité et à chaque itération on applique les étapes (ii) et (iii) sur les termes des parties  $P_1$  et  $P_2$ .

5           6. Procédé d'indexation selon la revendication 4 ou la revendication 5, caractérisé en ce qu'il comprend une optimisation à partir des  $N$  parties disjointes  $\{P_1, P_2, \dots, P_N\}$  de l'ensemble  $P$  ainsi que des  $N$  termes  $\{t_1, t_2, t_N\}$  qui les représentent pour réduire l'erreur de décomposition de l'ensemble  $P$  en  $N$  parties, et en ce qu'il comprend les étapes suivantes :

10

(i) calcul des centres de gravité  $C_i$  des parties  $P_i$

(ii) calcul des erreurs  $\varepsilon_{C_i} = \sum_{t_j \in P_i} d^2(C_i, t_j)$  et  $\varepsilon_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$  lorsqu'on

remplace les termes  $t_j$  de la partie  $P_i$  respectivement par  $C_i$  et par  $t_i$ ,

15           (iii) comparaison de  $\varepsilon_{t_i}$  et  $\varepsilon_{C_i}$  et remplacement de  $t_i$  par  $C_i$  si  $\varepsilon_{C_i} \leq \varepsilon_{t_i}$ ,

(iv) calcul de la nouvelle matrice  $T$  de distances entre les termes  $t_i$  de la base de termes et processus de décomposition de l'ensemble  $P$  des termes de la base de termes en  $N$  parties, sauf si une condition d'arrêt est remplie avec  $\frac{\varepsilon_{C_t} - \varepsilon_{C_{t+1}}}{\varepsilon_{C_t}} < \text{seuil}$ , où  $\varepsilon_{C_t}$  représente l'erreur

20

commise à l'instant  $t$ .

25           7. Procédé d'indexation selon l'une quelconque des revendications 1 à 6, caractérisé en ce que pour effectuer une structuration du dictionnaire de concepts, on produit de façon itérative à chaque itération une carte de navigation en commençant par scinder l'ensemble des concepts en deux sous-ensembles, puis en sélectionnant un sous-ensemble à chaque itération jusqu'à l'obtention du nombre de groupes souhaité ou jusqu'à ce qu'un critère d'arrêt soit satisfait.

30           8. Procédé d'indexation selon la revendication 7, caractérisé en ce que le critère d'arrêt est constitué par le fait que les sous-ensembles obtenus sont tous homogènes avec un écart-type faible.

35           9. Procédé d'indexation selon la revendication 7 ou la revendication 8, caractérisé en ce que lors de la structuration du dictionnaire de

concepts, on détermine des indicateurs de navigation à partir d'une matrice  $M = [c_1, c_2, \dots, c_N] \in \mathfrak{R}^{p \times N}$  de l'ensemble  $C$  des concepts  $c_i \in \mathfrak{R}^p$  où  $c_i$  représente un concept de  $p$  valeurs, selon les étapes suivantes :

- (i) on calcule un représentant  $w$  de la matrice  $M$
- 5 (ii) on calcule la matrice de covariance  $\tilde{M}$  entre les éléments de la matrice  $M$  et le représentant  $w$  de la matrice  $M$ ,
- (iii) on calcule un axe de projection  $u$  des éléments de la matrice  $M$ ,
- (iv) on calcule la valeur  $p_i = d(u, c_i) - d(u, w)$  et on décompose l'ensemble de concepts  $C$  en deux sous-ensembles  $C_1$  et  $C_2$  de la
- 10 manière suivante :

$$\begin{cases} c_i \in C_1 & \text{si } p_i \leq 0 \\ c_i \in C_2 & \text{si } p_i > 0 \end{cases}$$

- 15 (v) on stocke dans le nœud associé à  $C$  les informations  $\{u, w, |p_1|, p_2\}$  où  $p_1$  est le maximum de tous les  $p_i \leq 0$  et  $p_2$  est le minimum de tous les  $p_i > 0$ , l'ensemble des informations  $\{u, w, |p_1|, p_2\}$  constituant les indicateurs de navigation dans le dictionnaire de concept.

20

10. Procédé d'indexation selon l'une quelconque des revendications 1 à 9, caractérisé en ce que l'on analyse à la fois les composantes structurelles et les compléments de ces composantes structurelles constitué par les composantes texturales d'une image du document, et en

25 ce que :

- (a) lors de l'analyse des composantes structurelles de l'image
- (a1) on procède à une répartition des zones frontières des structures de l'image en différentes classes selon l'orientation de la variation locale d'intensité de manière à définir des éléments de support structurel
- 30 (ESS) de l'image, et
- (a2) on procède par analyse statistique à la construction de termes constitués par des vecteurs décrivant les propriétés locales et globales des éléments de support structurels,

- (b) lors de l'analyse des composantes texturales de l'image
- (b1) on procède à une détection et une caractérisation paramétrique d'une composante purement aléatoire de l'image,
- 5 (b2) on procède à une détection et une caractérisation paramétrique d'une composante périodique de l'image,
- (b3) on procède à une détection et une caractérisation paramétrique d'une composante directionnelle de l'image,
- 10 (c) on regroupe dans un nombre limité de concepts l'ensemble des éléments descriptifs de l'image constitués par d'une part les termes décrivant les propriétés locales et globales des éléments de support structurels et d'autre part les paramètres des caractérisations paramétriques des composantes aléatoire, périodique et
- 15 directionnelle définissant les composantes texturales de l'image, et
- (d) on définit pour chaque document une empreinte à partir des occurrences, des positions et des fréquences desdits concepts.
- 20 11. Procédé d'indexation selon la revendication 10, caractérisé en ce que les propriétés locales des éléments de support structurels prises en compte pour la construction de termes comprennent au moins le type de support choisi parmi une bande linéaire ou un arc de courbe, les dimensions en longueur et largeur du support, la direction principale du
- 25 support et la forme et les propriétés statistiques des pixels constituant le support.
12. Procédé d'indexation selon la revendication 10 ou la revendication 11, caractérisé en ce que les propriétés globales des
- 30 éléments de support structurels prises en compte pour la construction de termes comprennent au moins le nombre de chaque type de supports et leur disposition spatiale.
13. Procédé d'indexation selon l'une quelconque des revendications
- 35 10 à 12, caractérisé en ce que lors de l'analyse des composantes structurelles de l'image on procède à un test préalable de détection de la

présence d'au moins une structure dans l'image et, en cas d'absence de structure, on passe directement à l'étape de l'analyse des composantes texturales de l'image.

- 5           14. Procédé d'indexation selon l'une quelconque des revendications 10 à 13, caractérisé en ce que pour procéder à une répartition des zones frontières des structures de l'image en différentes classes, à partir de l'image numérisée définie par l'ensemble des pixels  $y(i,j)$  où  $(i,j) \in I \times J$ , avec I et J désignant respectivement le nombre de lignes et le nombre de  
10 colonnes de l'image, on calcule l'image gradient vertical  $g_v(i,j)$  avec  $(i,j) \in I \times J$  et l'image gradient horizontal  $g_h$  avec  $(i,j) \in I \times J$  et on procède au partitionnement de l'image selon l'orientation locale de son gradient en un nombre fini de classes équidistantes, l'image contenant l'orientation du gradient étant définie par la formule

15

$$O(i,j) = \arctan \left[ \frac{g_h(i,j)}{g_v(i,j)} \right],$$

- on identifie les classes constituant des régions de support susceptibles de contenir des éléments de support significatifs, et à partir des régions de  
20 support, on détermine les éléments de support significatifs et on les répertorie selon des critères prédéterminés.

15. Procédé d'indexation selon l'une quelconque des revendications 1 à 9, caractérisé en ce que lors de l'indexation d'un document multimédia comportant des signaux vidéo, on choisit des termes  $t_i$  constitués par des  
25 images-clés représentant des groupes d'images homogènes consécutives, et on détermine des concepts  $c_i$  par regroupement de termes  $t_i$ .

16. Procédé d'indexation selon la revendication 15, caractérisé en ce que pour déterminer des images-clés constituant des termes  $t_i$ , on élabore d'abord un vecteur score VS comprenant un ensemble d'éléments  
30 VS(i) matérialisant la différence ou la similarité entre le contenu d'une image d'indice i et celui d'une image d'indice i-1, et on analyse le vecteur score VS afin de déterminer les images-clés qui correspondent aux maximums des valeurs des éléments VS(i) du vecteur score VS.

17. Procédé d'indexation selon la revendication 16, caractérisé en ce qu'une image d'indice  $j$  est considérée comme une image-clé si la valeur  $VS(j)$  de l'élément correspondant du vecteur score  $VS$  est un maximum et que la valeur  $VS(j)$  est située entre deux minimums  $\min G$  et  $\min D$  et que le minimum  $M1$  tel que  $M1 = (|VS_{(j)} - \min G|, |VS_{(j)} - \min D|)$  est supérieur à un seuil donné.

18. Procédé d'indexation selon l'une quelconque des revendications 1 à 9, caractérisé en ce que lors de l'indexation d'un document multimédia comportant des composantes audio, on échantillonne et décompose le document en trames, qui sont ensuite regroupées en clips dont chacun est caractérisé par un terme  $t_i$  constitué par un vecteur de paramètre.

19. Procédé d'indexation selon la revendication 18, caractérisé en ce qu'une trame comprend entre environ 512 et 2048 échantillons du document audio échantillonné.

20. Procédé d'indexation selon la revendication 18 ou la revendication 19, caractérisé en ce que les paramètres pris en compte pour la définition des termes  $t_i$  comprennent des informations temporelles correspondant à au moins l'un des paramètres suivants : l'énergie des trames du signal audio, l'écart-type des énergies des trames dans les clips, le rapport des variations sonores, le rapport de basse énergie, le taux d'oscillation autour d'une valeur prédéterminée, le haut taux d'oscillation autour d'une valeur prédéterminée, la différence entre le nombre de taux d'oscillation au-dessus et au-dessous du taux d'oscillation moyen des trames de clips, la variance du taux d'oscillation, le rapport des trames silencieuses.

21. Procédé d'indexation selon l'une quelconque des revendications 18 à 20, caractérisé en ce que les paramètres pris en compte pour la définition des termes  $t_i$  comprennent des informations fréquentielles correspondant à au moins l'un des paramètres suivants : le centre de gravité du spectre de fréquence de la transformée de Fourier courte du signal audio, la largeur de bande du signal audio, le rapport entre l'énergie dans une bande de fréquence et l'énergie totale dans toute la bande de fréquence du signal audio échantillonné, la valeur moyenne de la variation

du spectre de deux trames adjacentes dans un clip, la fréquence de coupure d'un clip.

22. Procédé d'indexation selon l'une quelconque des revendications  
5 18 à 21, caractérisé en ce que les paramètres pris en compte pour la  
définition des termes  $t_i$  comprennent au moins la modulation d'énergie à  
4 Hz.

23. Procédé d'indexation selon l'une quelconque des revendications  
10 1 à 14, caractérisé en ce que l'on analyse les formes d'une image d'un  
document selon les étapes suivantes :

- (a) on procède à une multirésolution suivie d'une décimation de l'image,
- (b) on définit l'image dans l'espace logarithmique polaire.
- (c) on représente l'image ou la portion de l'image concernée par sa  
15 transformée de Fourier H,
- (d) on procède à une caractérisation de la transformée de Fourier H de la  
façon suivante :
  - (d1) on projette H dans plusieurs directions pour obtenir un ensemble  
de vecteurs dont la dimension est égale à la dimension du  
20 mouvement de projection,
  - (d2) on calcule les propriétés statistiques de chaque vecteur de  
projection, et
  - (e) on représente la forme de l'image par un terme  $t_i$  constitué des  
valeurs des propriétés statistiques de chaque vecteur de  
25 projection.

1/15

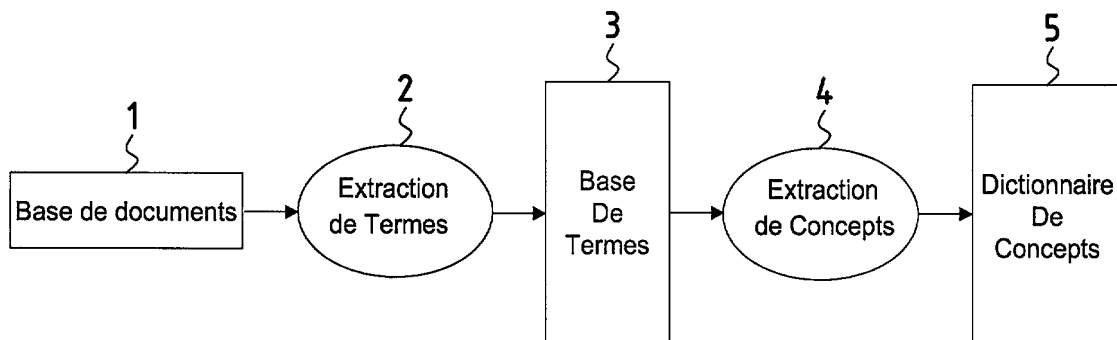


FIG.1

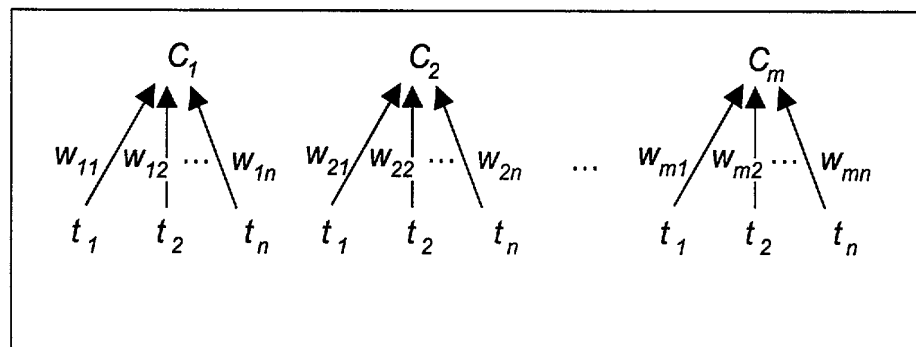


FIG.2

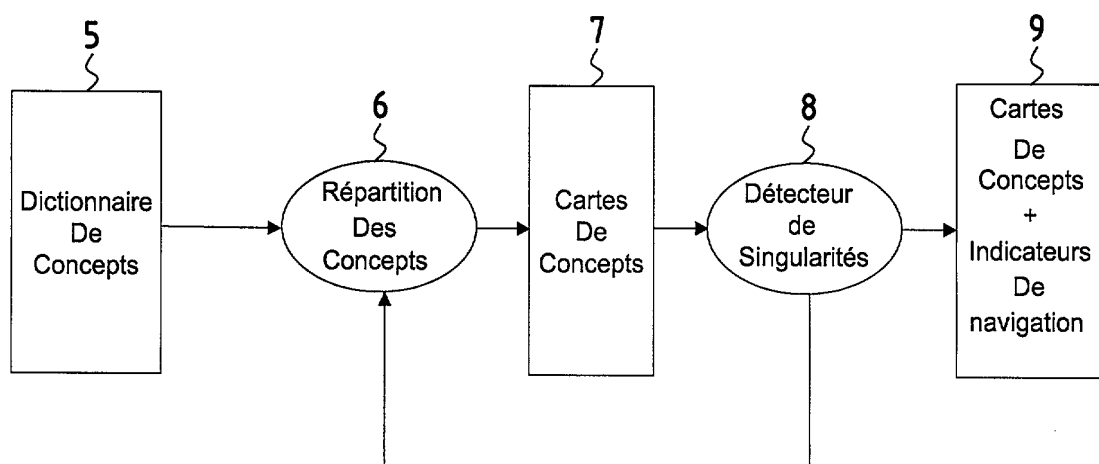


FIG.3

2/15

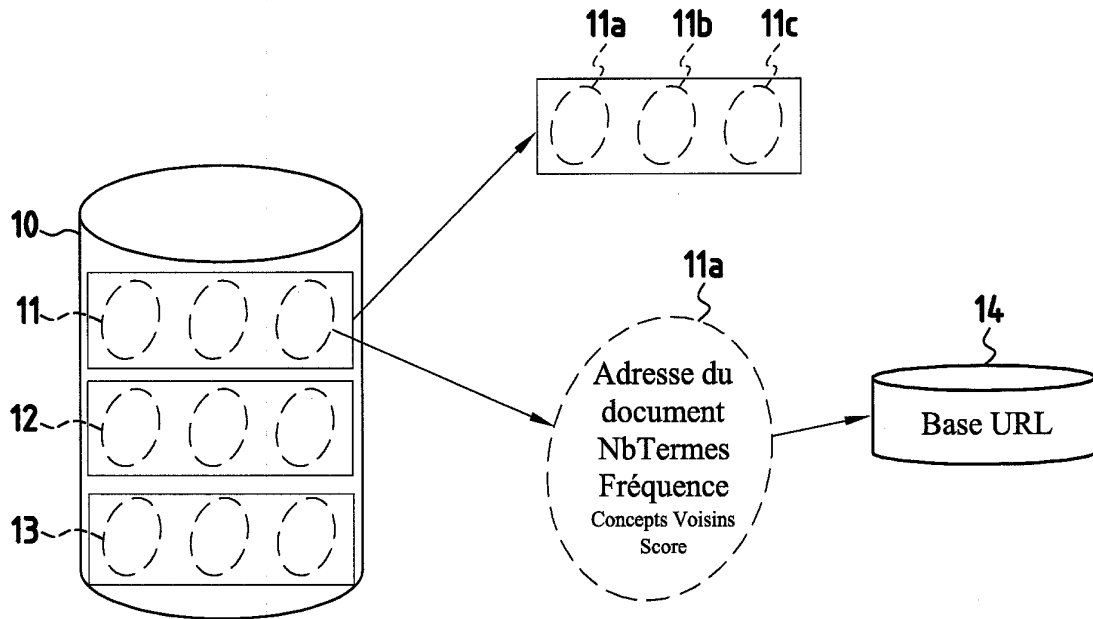


FIG.4

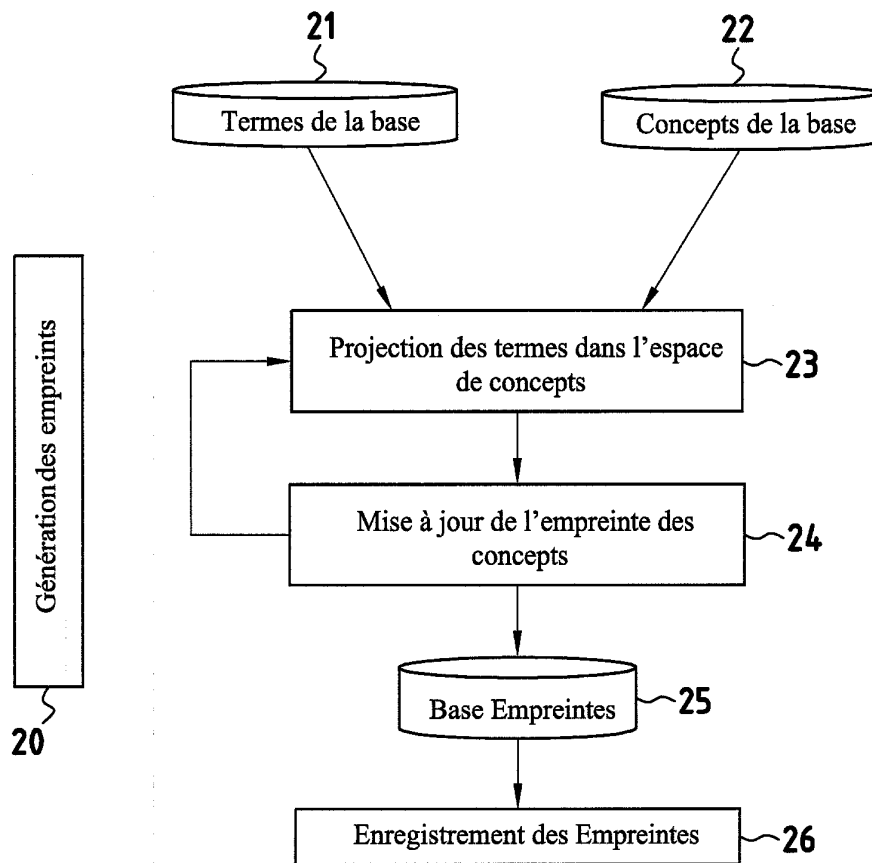


FIG.5

3/15

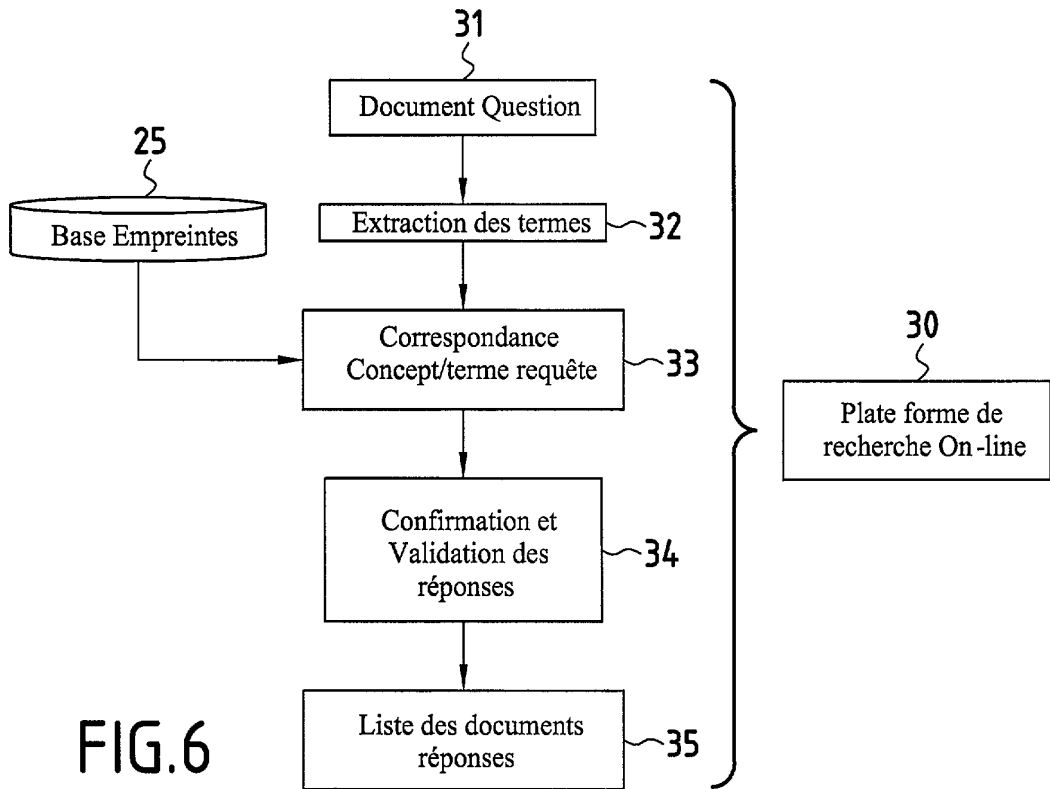


FIG. 6

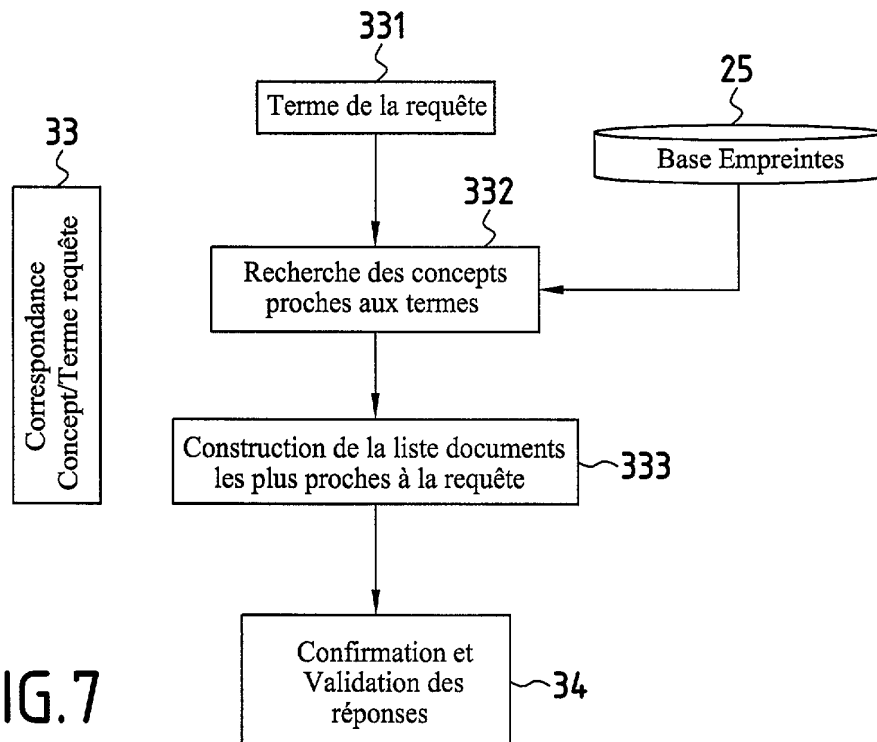


FIG. 7

4/15

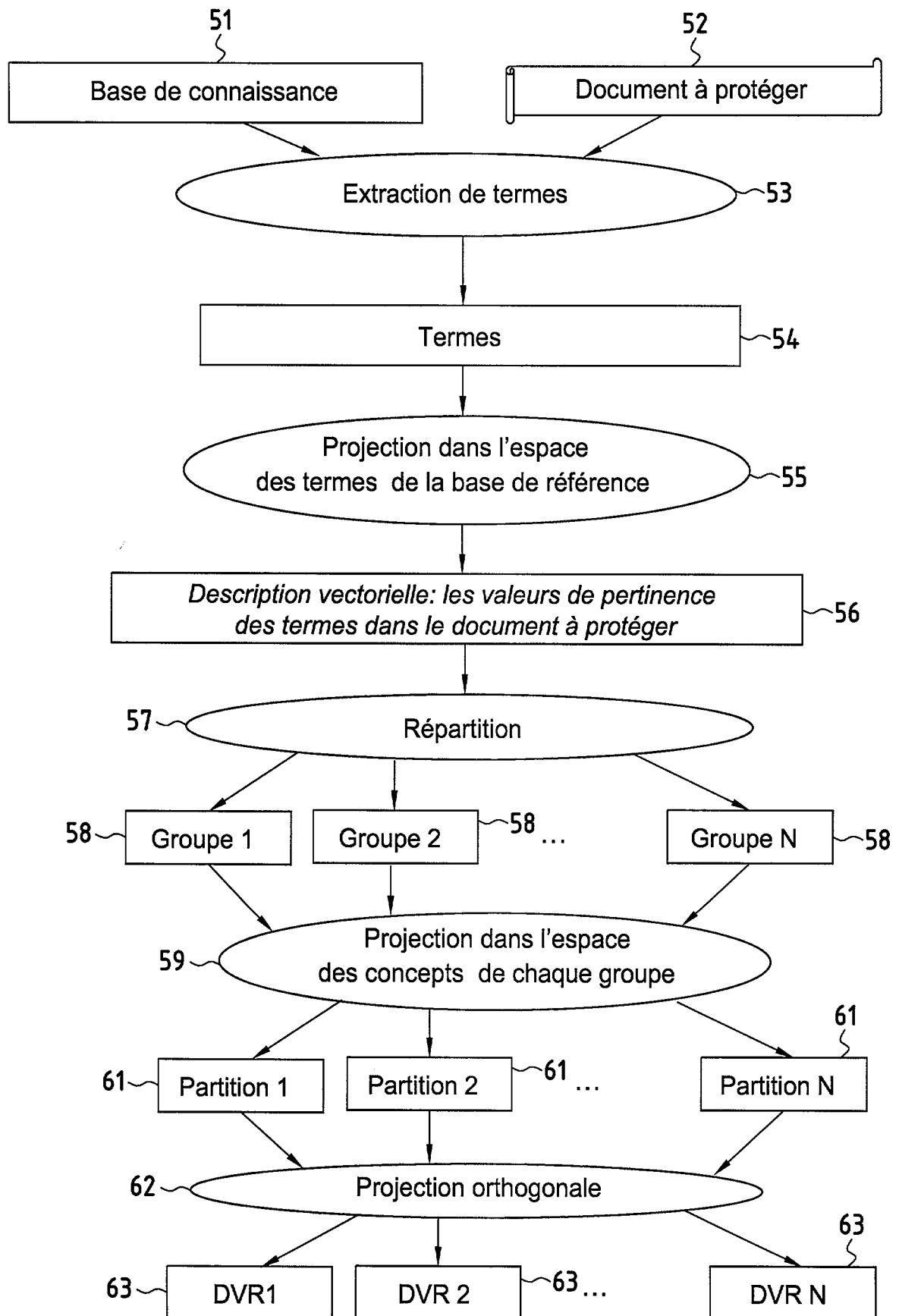
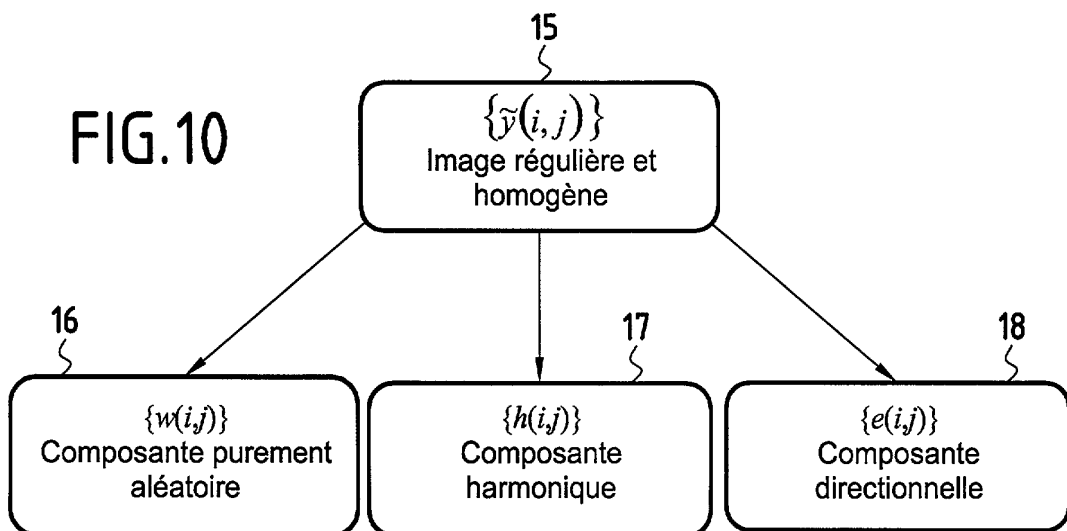
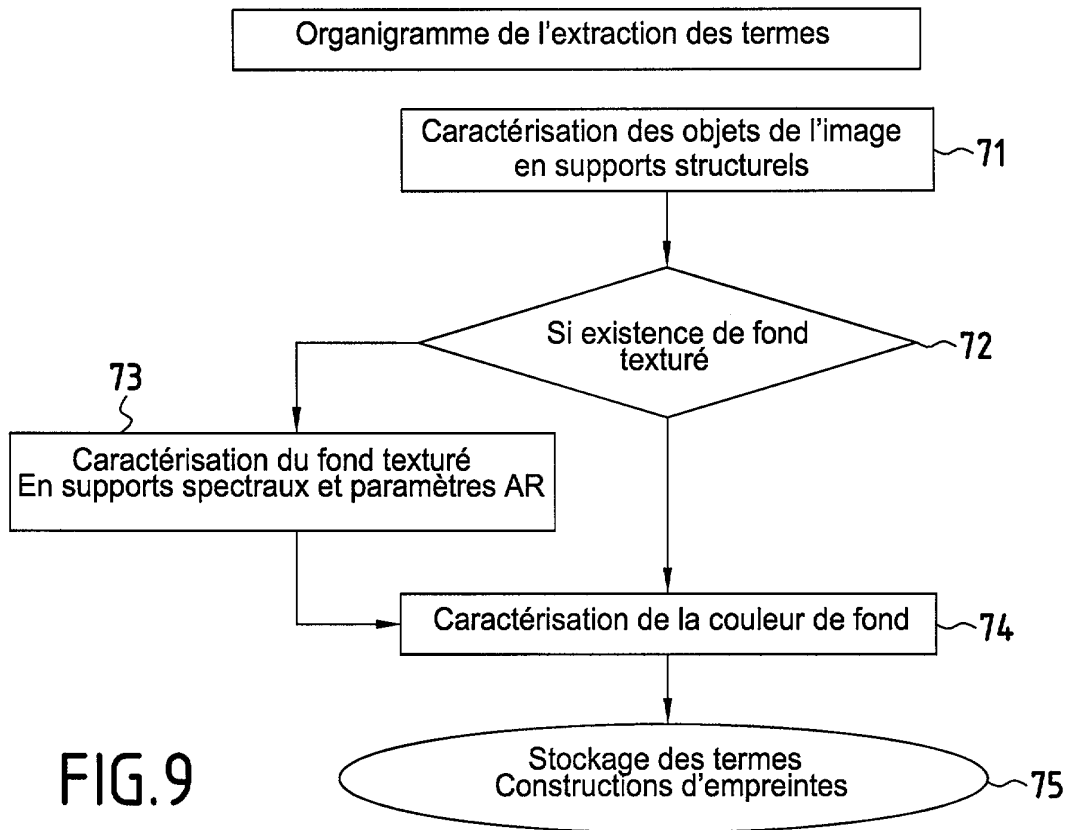


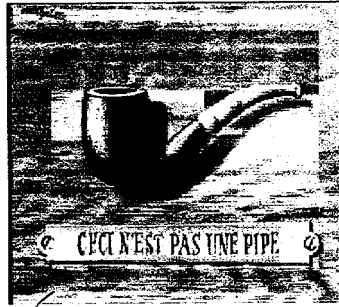
FIG.8

5/15

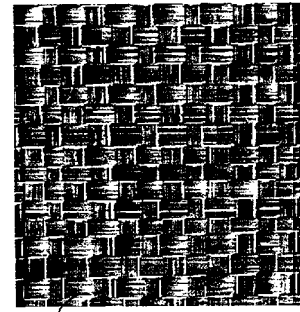




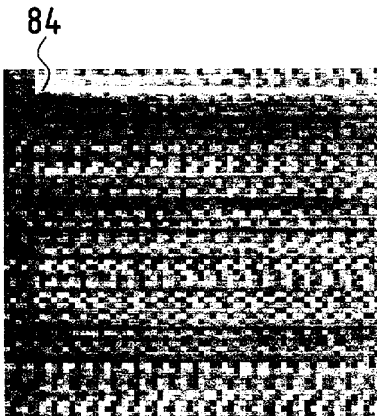
81 FIG. 11



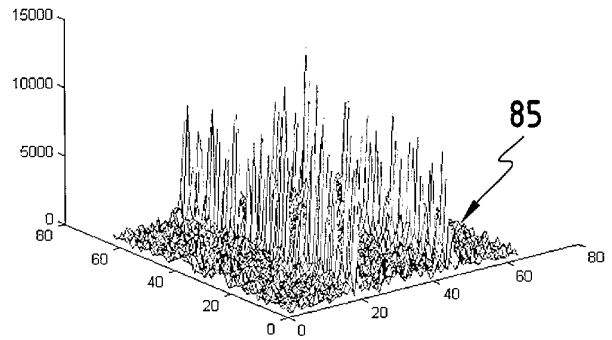
82 FIG. 12



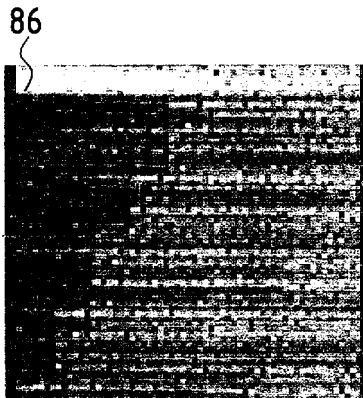
83 FIG. 13



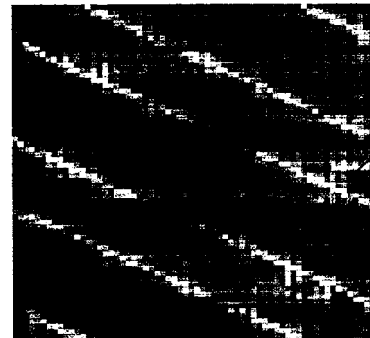
84 FIG. 15a



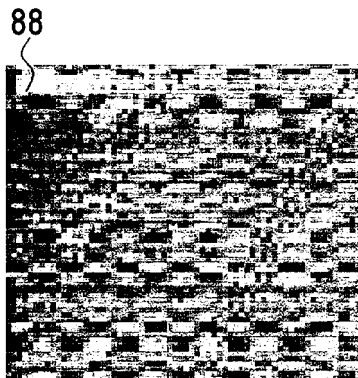
85 FIG. 15a1



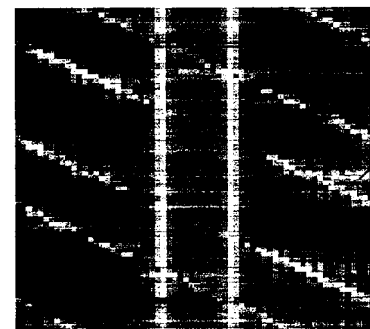
86 FIG. 15b



87 FIG. 15b1



88 FIG. 15c



89 FIG. 15c1



FIG. 14a

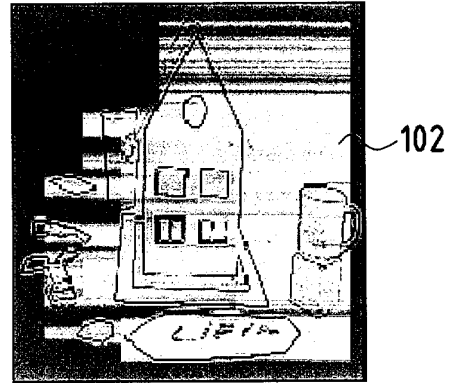


FIG. 14b

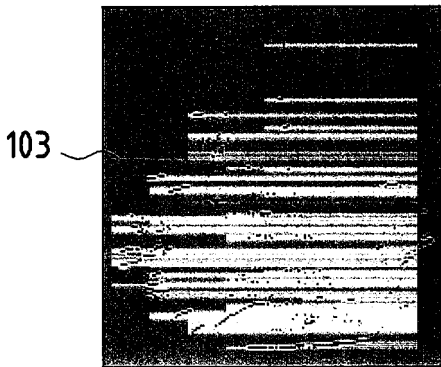


FIG. 14c

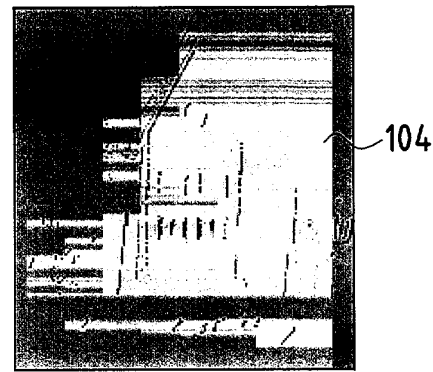


FIG. 14d

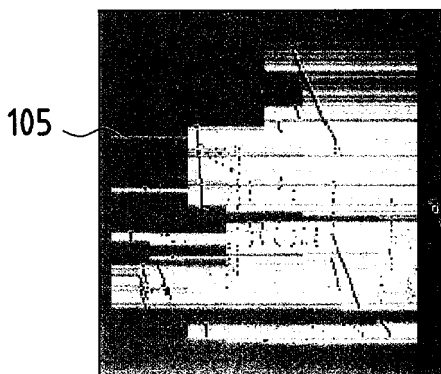


FIG. 14e

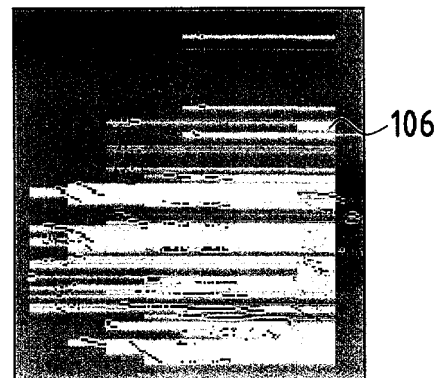


FIG. 14f

8/15

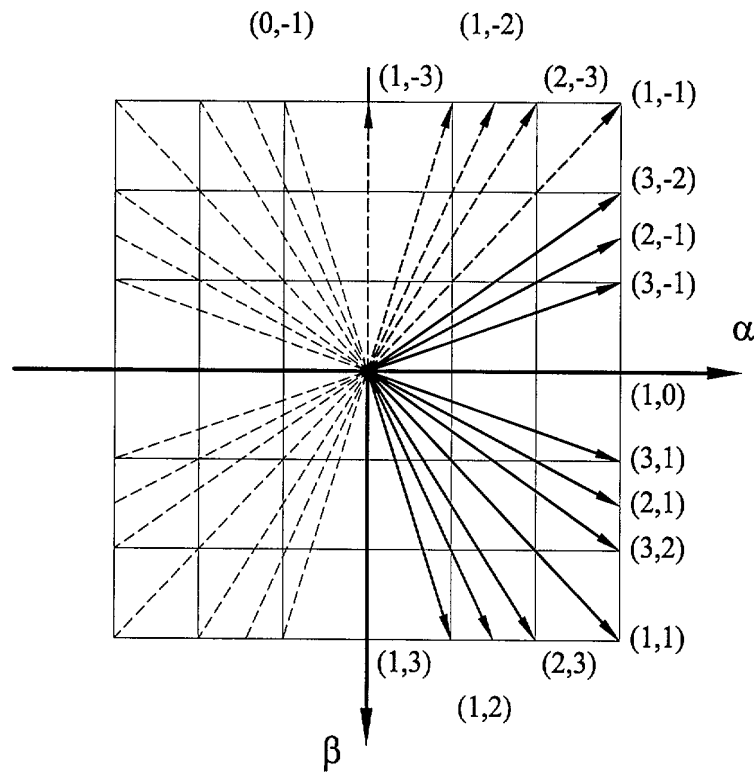


FIG.16

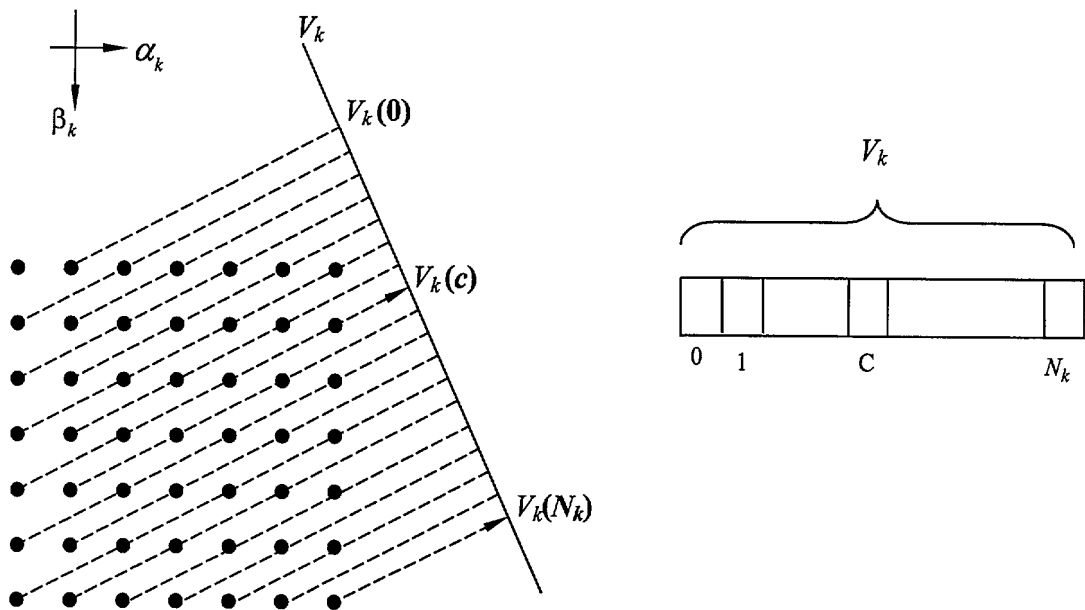


FIG.17

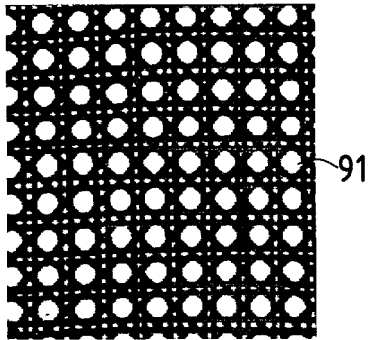


FIG.18a1

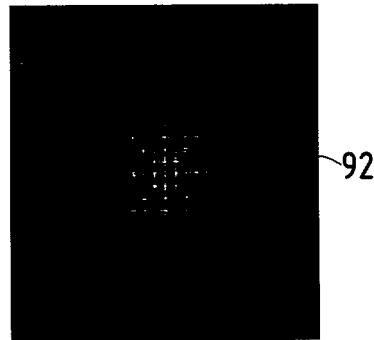


FIG.18a2

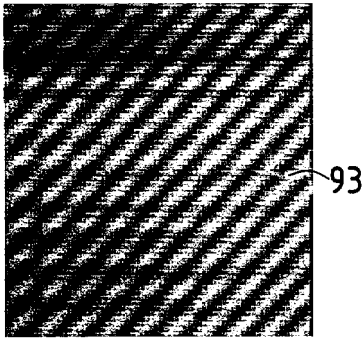


FIG.18b1

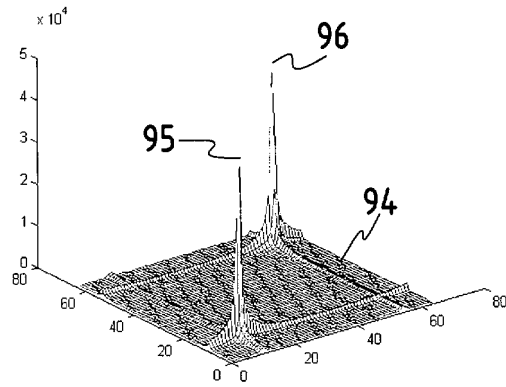
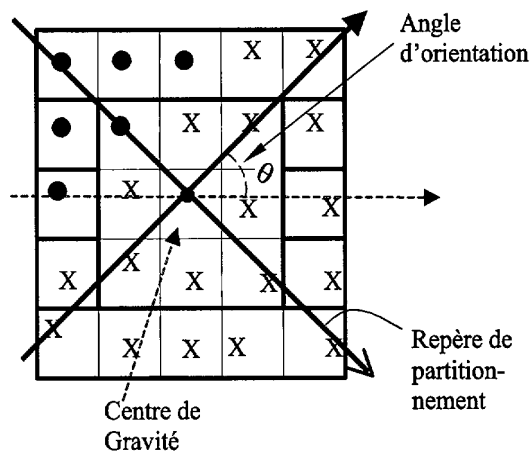


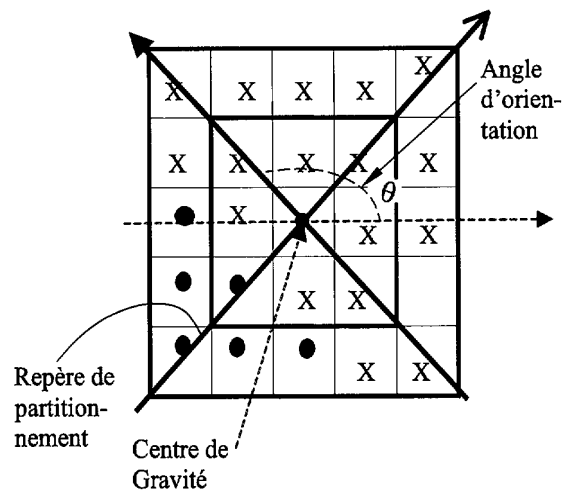
FIG.18b2



19	6	5	4	4	8	11
----	---	---	---	---	---	----

V0

FIG.20



19	6	5	4	4	8	11
----	---	---	---	---	---	----

V1

FIG.21

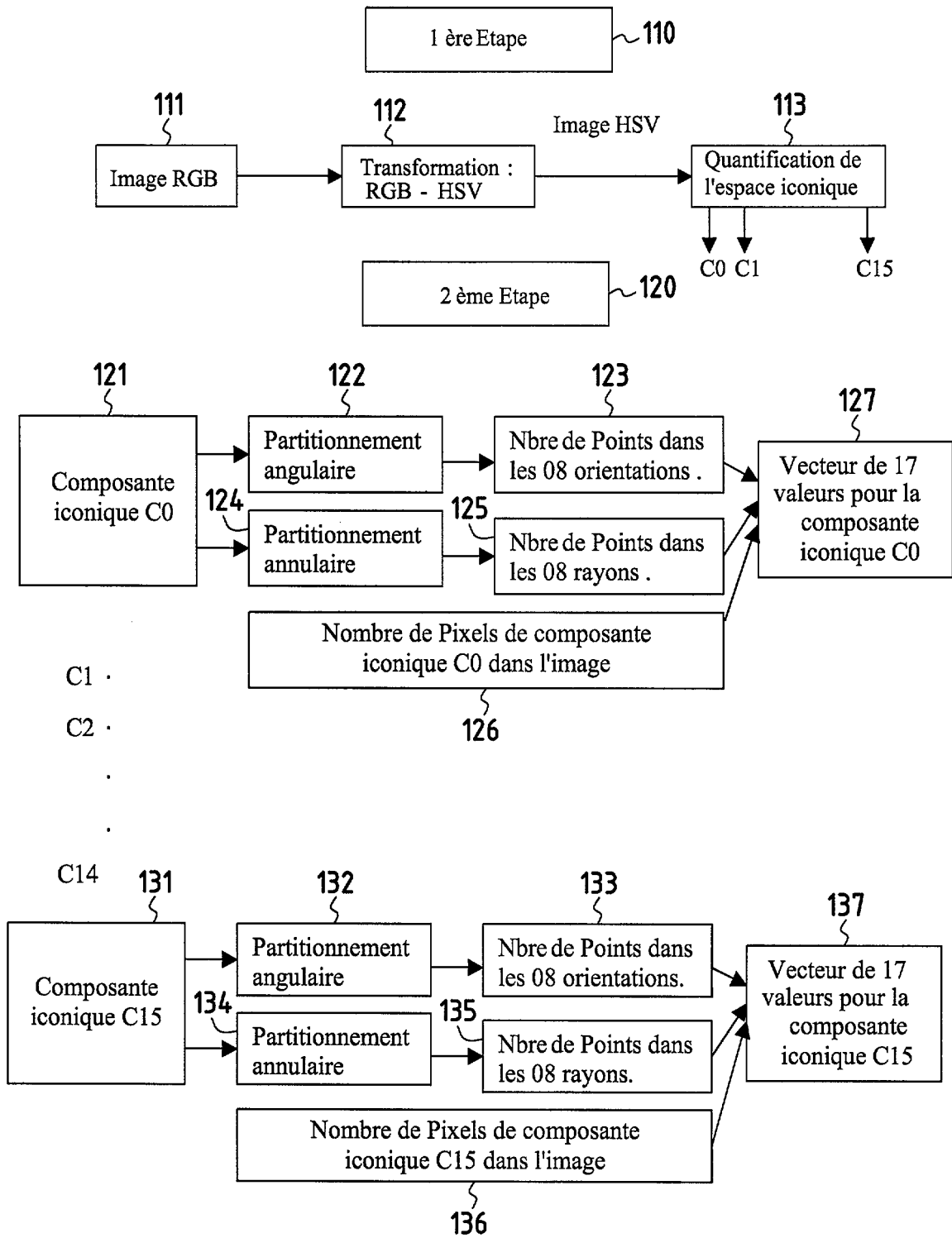


FIG.19

11/15

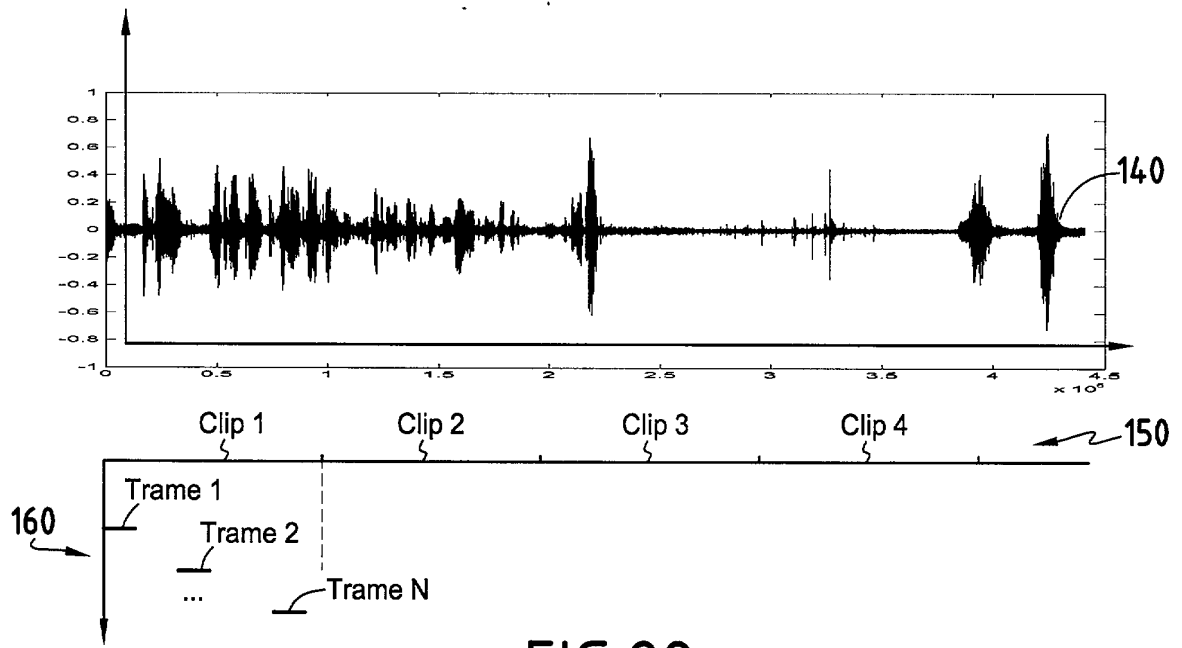


FIG.22

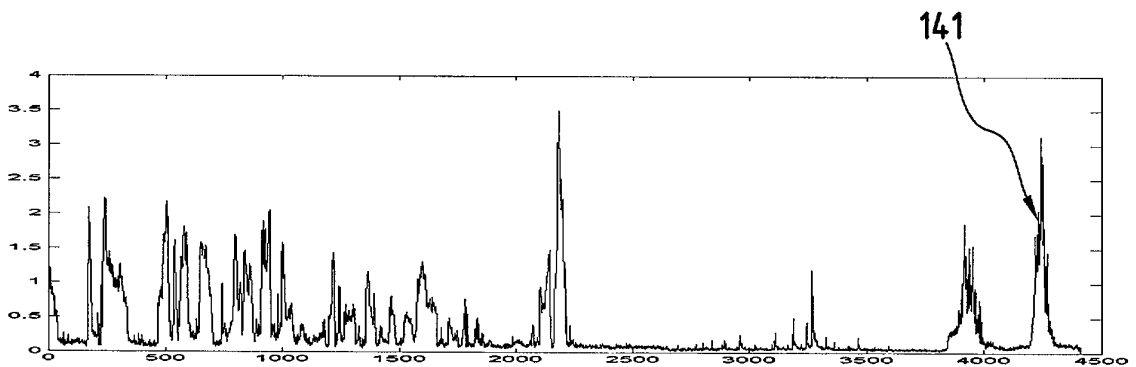


FIG.23a

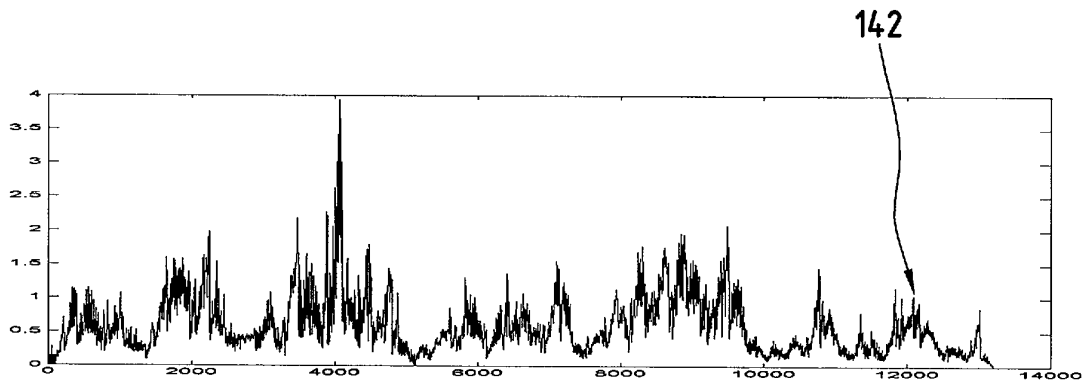


FIG.23b

12/15

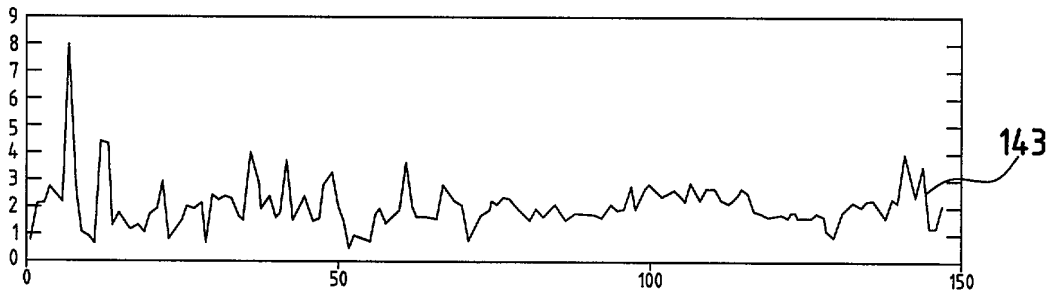


FIG.24a

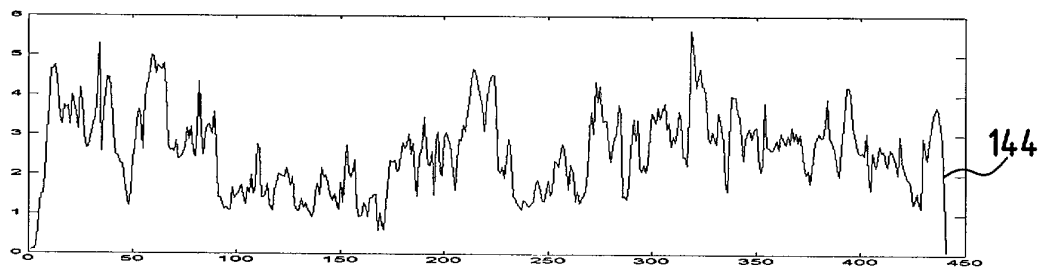


FIG.24b

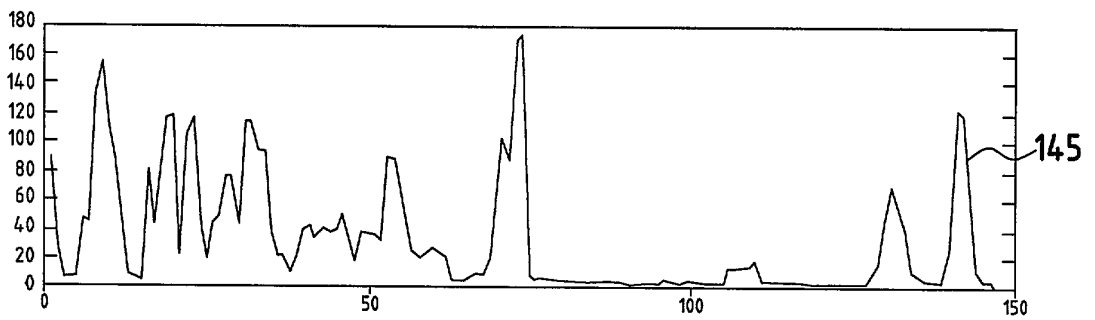


FIG.25a

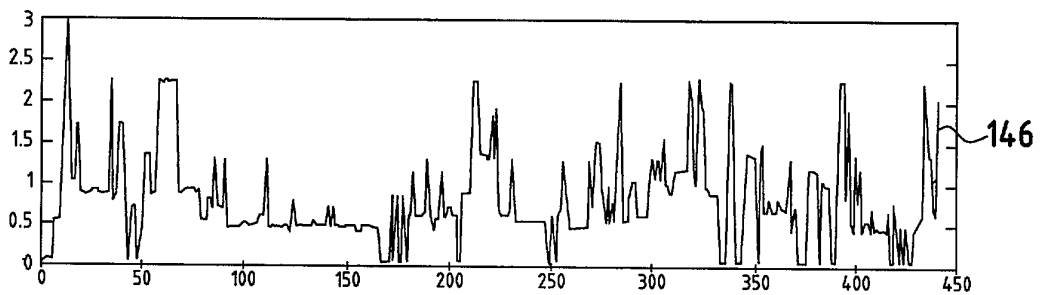


FIG.25b

13/15

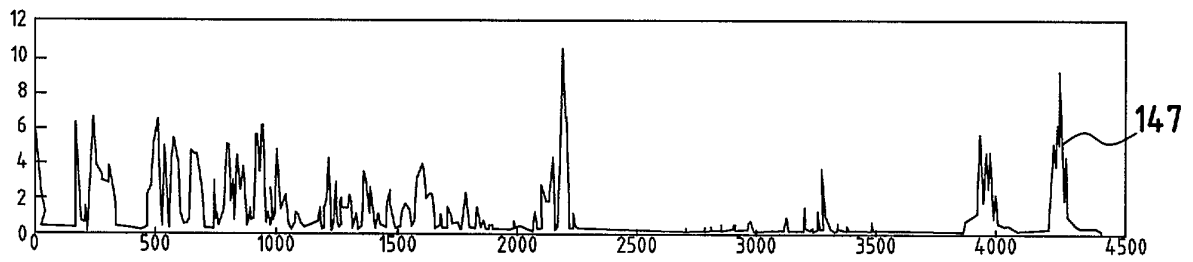


FIG. 26a

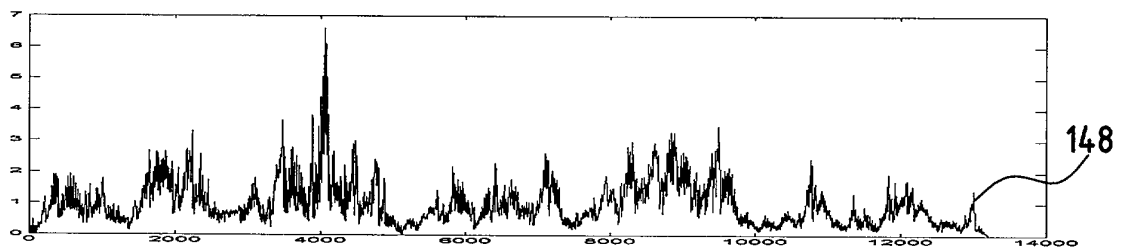


FIG. 26b

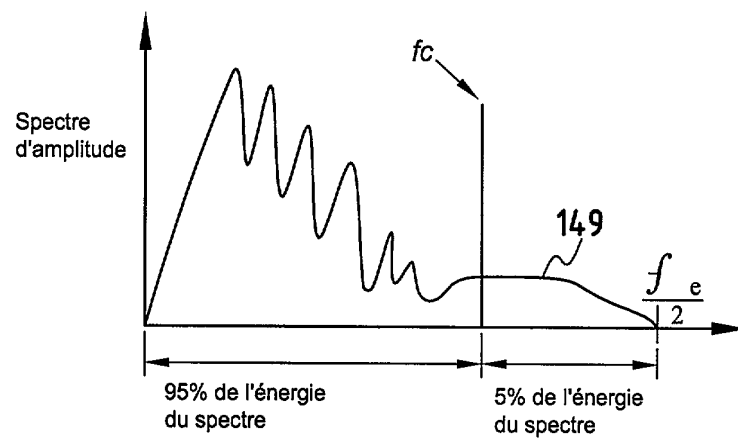


FIG. 29

14/15

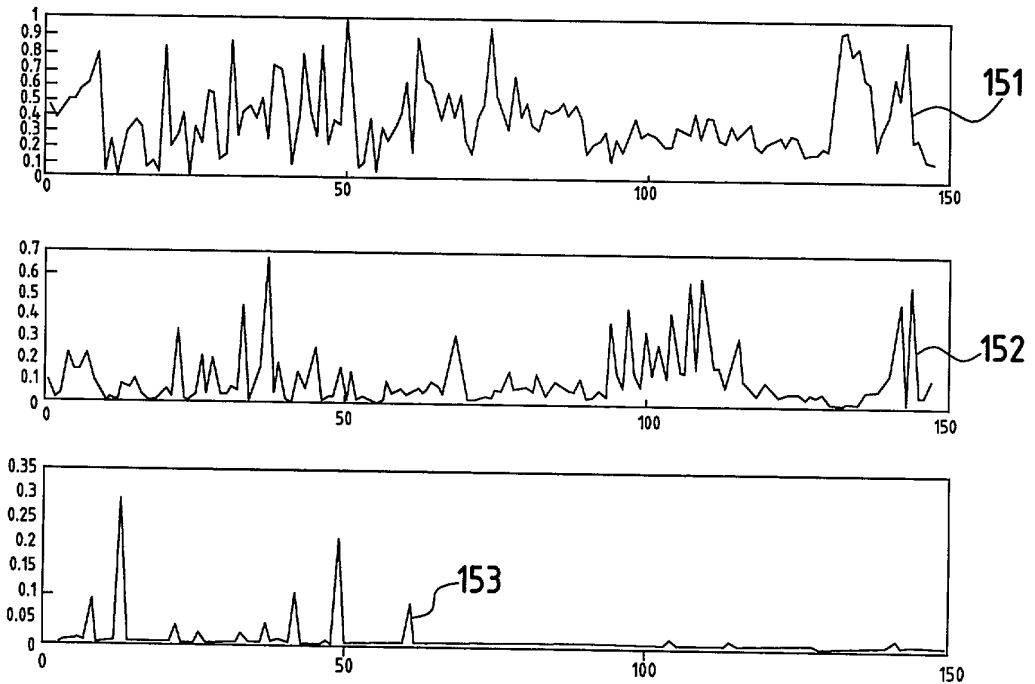


FIG.27a

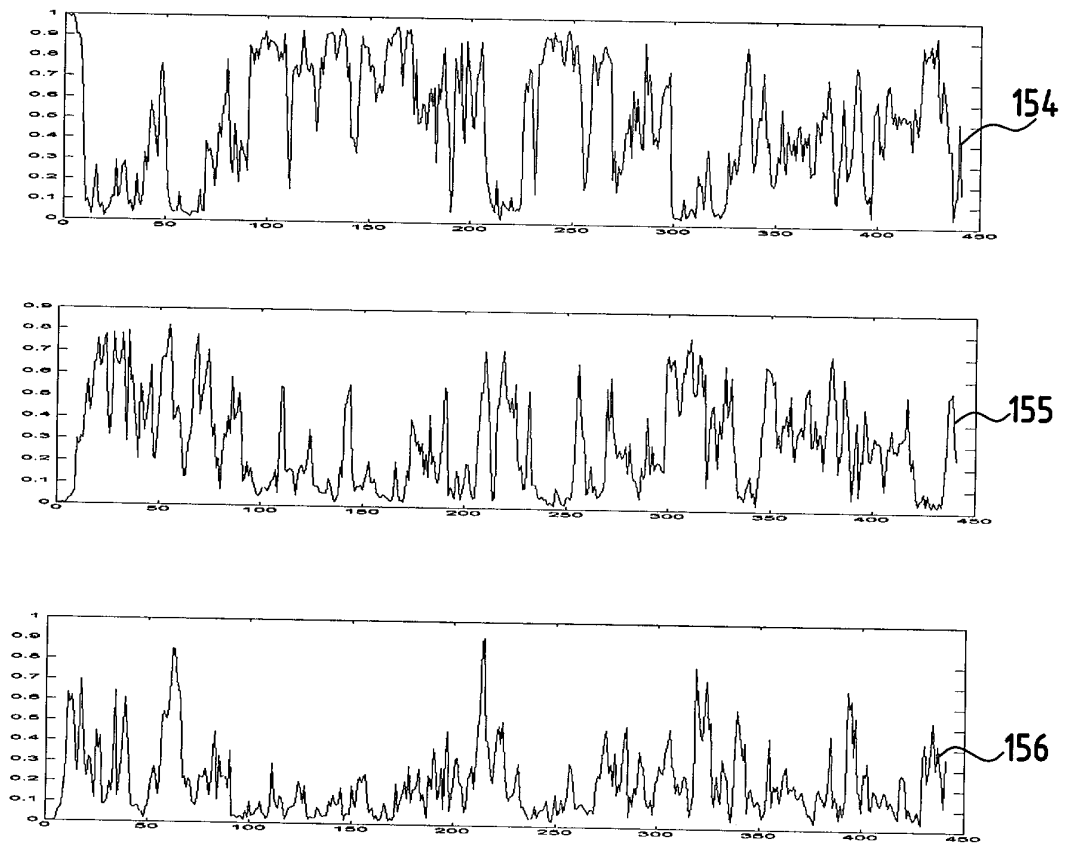


FIG.27b

15/15

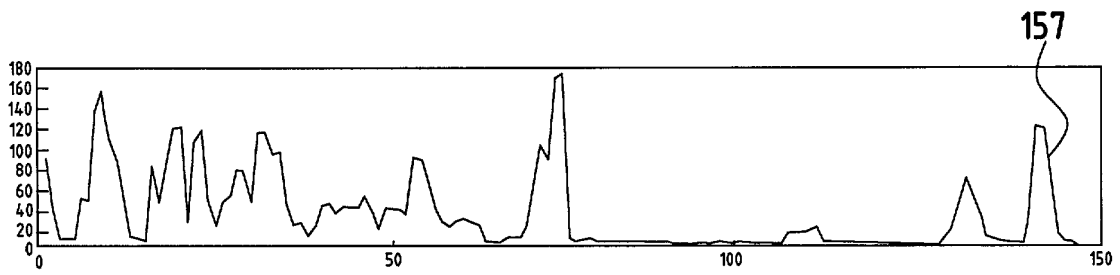


FIG.28a

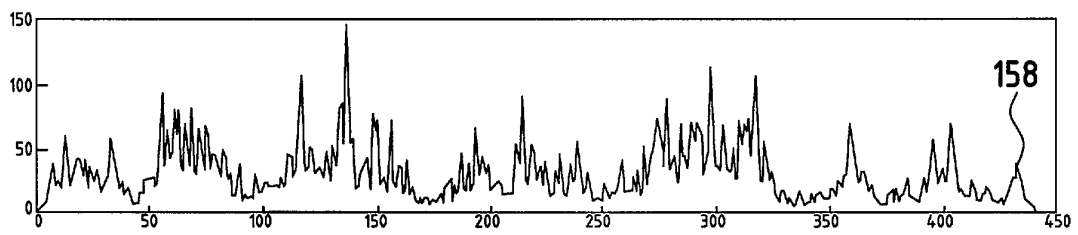


FIG.28b

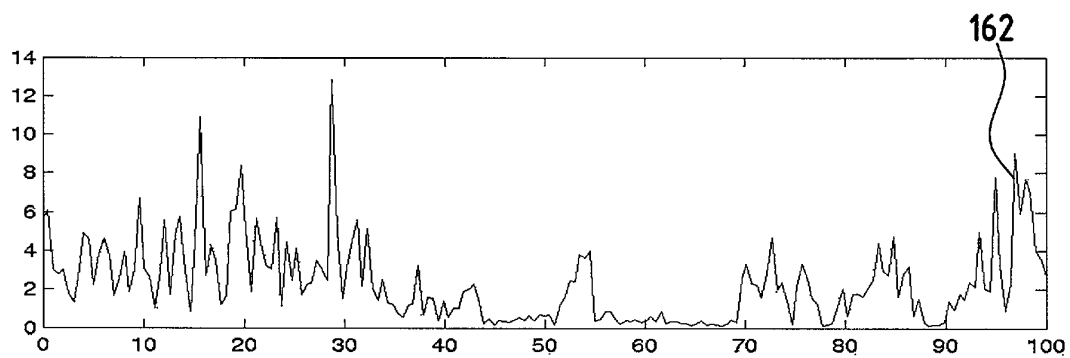
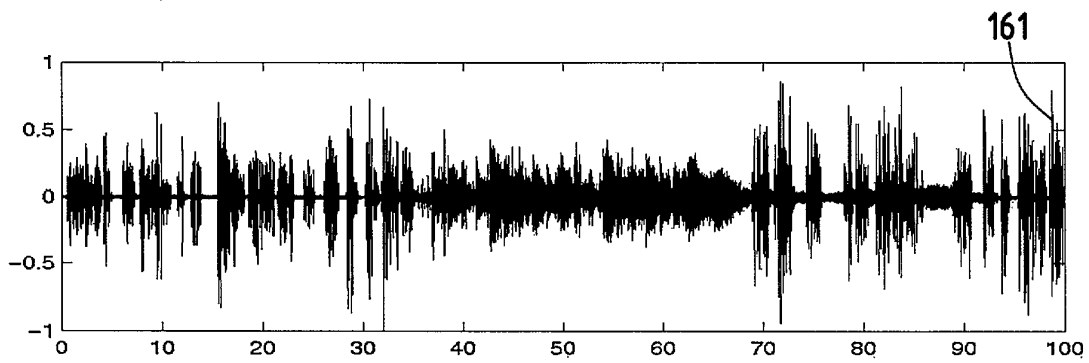


FIG.30



**RAPPORT DE RECHERCHE  
PRÉLIMINAIRE**  
établi sur la base des dernières revendications  
déposées avant le commencement de la recherche

N° d'enregistrement  
national

FA 644642  
FR 0313907

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
A	Y. CHANG, I. CHOI, J. CHOI, M. KIM, AND V. V. RAGHAVAN: "Conceptual retrieval based on feature clustering of documents" PROCEEDINGS OF ACM SIGIR WORKSHOP ON MATHEMATICAL-FORMAL METHODS IN INFORMATION RETRIEVAL, [Online] août 2002 (2002-08), pages 1-10, XP002297293 TAMPERE, FINLAND Extrait de l'Internet: URL:http://www.cacs.louisiana.edu/Publications/Raghavan/CCCK02.pdf> [extrait le 2004-09-15] * abrégé * * page 3, alinéa 3; figure 1 *	1,2	G06F17/30 G06T7/00
A	DIAMANTINI C ET AL: "A conceptual indexing method for content-based retrieval" DATABASE AND EXPERT SYSTEMS APPLICATIONS, 1999. PROCEEDINGS. TENTH INTERNATIONAL WORKSHOP ON FLORENCE, ITALY 1-3 SEPT. 1999, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 1 septembre 1999 (1999-09-01), pages 192-197, XP010352474 ISBN: 0-7695-0281-4 * abrégé *	1	DOMAINES TECHNIQUES RECHERCHÉS (Int.CL.7) G06F
A	CARRE M ET AL: "INDEXATION AUDIO: UN ETAT DE L'ART STATE OF THE ART IN AUDIO INDEXING" ANNALES DES TELECOMMUNICATIONS - ANNALS OF TELECOMMUNICATIONS, PRESSES POLYTECHNIQUES ET UNIVERSITAIRES ROMANDES, LAUSANNE, CH, vol. 55, no. 9/10, septembre 2000 (2000-09), pages 507-525, XP000994593 ISSN: 0003-4347 * page 508, colonne de droite, alinéa II.1 - page 512, colonne de gauche, ligne 6 *	1	
Date d'achèvement de la recherche		Examineur	
21 septembre 2004		Fournier, C	
<p>CATÉGORIE DES DOCUMENTS CITÉS</p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire</p>		<p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons ..... &amp; : membre de la même famille, document correspondant</p>	