

Approche exploratoire multirésolution basée sur le contenu d'une base d'images paléontologique

Content-based multiresolution exploratory approach of a paleontology images database

Jérôme Landré

Le2i - FRE 2309 CNRS

I.U.T. - 12, rue de la Fonderie

F - 71 200 Le Creusot

j.Landre@iutlecreusot.u-bourgogne.fr

Frédéric Truchetet

Le2i - FRE 2309 CNRS

I.U.T. - 12, rue de la Fonderie

F - 71 200 Le Creusot

f.Truchetet@iutlecreusot.u-bourgogne.fr

Abstract

This article deals with a work on content-based multiresolution retrieval of images in a paleontology images database. Our technique consists in extracting several multiresolution parameters from the wavelets coefficients of each image. Those parameters are automatically classified using k-means algorithm to build similar images families. For each family, a model image is built and is chosen to represent the family in a research tree. Our approach offers users a visual research tree to browse the database allowing multiresolution progressive retrieval techniques.

Keywords : Multiresolution analysis, content-based image retrieval, classification, images databases.

Cet article présente un travail sur la recherche et l'indexation d'images basée sur le contenu. La méthode consiste à transformer les images à l'aide des ondelettes pour en extraire des paramètres caractéristiques classés automatiquement à l'aide des nuées dynamiques. Les images sont regroupées en familles, et pour chaque famille, une image modèle est construite et est insérée dans un arbre proposé à l'utilisateur pour la recherche. L'approche multirésolution permet de commencer la recherche sur des images grossières

et d'améliorer la résolution au fur et à mesure de la progression dans l'arbre.

Mots-clés : Analyse multirésolution, recherche d'images par le contenu, classification, Base de données images.

1 Introduction

La recherche d'images est un domaine très vaste qui recouvre de nombreuses thématiques : l'analyse d'images, les bases de données, la reconnaissance de formes, l'intelligence artificielle... Les données manipulées sont des images qui ont un contenu syntaxique et sémantique très complexe pour lequel nous ne disposons pas de descripteur complet. Une représentation réduite de ces images comporte une grande part d'incertitude et d'éléments non mesurables. Nous verrons comment l'analyse multirésolution associée à une classification par les nuées dynamiques permet de gérer efficacement cette part d'incertain.

Le problème de la recherche d'images est apparu en même temps que celui de leur stockage. Chaque image, acquise par un moyen ou un autre, devait être stockée sous forme numérique. La première idée pour accéder à ces données images est inspirée des recherches textuelles. À chaque image sont associés plusieurs mots-clés. La recherche revient alors à une recherche textuelle sur ces mots-clés. Cette approche pose deux problèmes : l'annotation est subjective et dépend fortement de la personne qui en est chargée, l'annotation

de grandes bases d'images devient impossible manuellement.

La recherche d'images par le contenu propose d'utiliser pour la recherche les images elle-mêmes sans aucune autre information. Depuis quelques années elle suscite de la part de la communauté scientifique du traitement d'images un intérêt très particulier en raison des applications de recherche d'images sur Internet.

L'indexation a pour but de substituer à une image qui occupe une place non négligeable un représentant moins encombrant qui la caractérise le mieux possible et de ne travailler que sur ce modèle lors de la recherche. La difficulté provient de la définition même du représentant : quelles caractéristiques choisir pour quel résultat ? En général, les représentants sont d'ordre globaux (forme, couleurs, histogramme...), locaux (texture, extrema...) ou mixtes.

La section 2 présente les travaux relatifs à ce domaine. Elle est séparée en trois sous-sections : la recherche d'images par le contenu 2.1, l'analyse multirésolution 2.2 et les méthodes de classification automatiques 2.3. La section 3 décrit notre méthode, la partie 3.1 s'intéresse à l'analyse multirésolution, la partie 3.2 explique l'algorithme de classification des nuées dynamiques et le paragraphe 3.3 indique la procédure de construction de l'arbre. Les résultats expérimentaux sont donnés en section 4. La section 5 présente la conclusion et les perspectives.

2 Travaux antérieurs

2.1 recherche d'images par le contenu

La recherche d'images par le contenu est un domaine vaste qui recouvre de nombreuses thématiques. Le nombre de publications dans le domaine est en forte croissance chaque année, il est très difficile de donner une bibliographie exhaustive du domaine. Toutefois, Rui [15], Huang et Chang donnent un état de l'art de la discipline relativement complet. Vissac [16] propose un panorama général sur l'indexation d'images fixes.

De nombreux systèmes existent d'ores et déjà sur le marché, dans le désordre on peut citer : QBIC d'IBM, Photobook du MIT Media lab, Virage de Virage software, Chabot, Visual-Seek, Retrievalware, Netra... Plus d'informations peuvent être trouvées dans les références bibliographiques ci-dessus.

2.2 analyse multirésolution

L'analyse multirésolution qui est employée pour notre méthode a été introduite par Mallat [11]. Elle permet d'obtenir une image transformée à plusieurs résolutions à partir d'une image originale. Chaque résolution donne une approximation de l'originale et une (ou plusieurs) image(s) de détails. Comme le décrit Mallat, l'analyse multirésolution est un outil qui permet de repérer les singularités d'un signal. La détection de contours et la discrimination de textures peuvent être directement obtenues à partir des coefficients de la transformée en ondelettes.

L'idée d'utiliser les ondelettes dans la recherche d'images par le contenu n'est pas nouvelle. Beaucoup d'articles traitent de leurs applications dans cette thématique. Dans leur article, Jacobs [8], Finkelstein et Salesin utilisent l'analyse multirésolution afin de créer un index à partir des indices et des signes des coefficients de plus grande amplitude. La comparaison est ensuite faite à plusieurs résolutions pendant la recherche entre l'image requête et les images de la base en utilisant une métrique construite pour leur besoin. Une autre méthode utilisée par Mandal [12], Aboulnasr et Panchanathan utilise les histogrammes des sous-bandes d'ondelettes pour l'indexation et la recherche à nouveau à partir d'une fonction de distance entre histogrammes. Guyader et Hérault [6] utilisent des filtres d'ondelettes de Gabor pour simuler le comportement du système rétinien humain. Les images sont regroupées grâce à une classification supervisée.

Dans leur article, Idris [7] et Panchanathan proposent une méthode basée sur la quantification vectorielle des paramètres extraits de l'analyse multirésolution dans laquelle la com-

paraison est effectuée sur les vecteurs obtenus pour chaque image. L'idée de recherche progressive d'images est donnée par Liang [10] et Kuo. Les coefficients d'ondelettes sont calculés à plusieurs niveaux qui sont utilisés au fur et à mesure de l'avancement de la recherche. Dans son rapport, Pečenović [14] donne un algorithme de décomposition en paquet d'ondelettes pour approximer la transformée de Karhunen-Loève afin d'indexer et de retrouver les images de la base. Do [5] préfère une méthode conçue à partir des moments des maxima des coefficients d'ondelettes par indexation locale. Chen [3], Li et Chien, quant à eux, travaillent sur la segmentation d'images couleur à plusieurs résolutions autorisant une recherche progressive.

2.3 classification automatique

De nombreuses méthodes de classification automatique existent. Le principe est de séparer un ensemble de paramètres en un nombre de classes donné. Plusieurs techniques tendent à résoudre ce problème. Dans leur papier, Ng [13] et Han illustrent une technique de regroupement d'informations basée sur un partitionnement statistique d'un espace de paramètres. Chang [1], Sheikholeslami, Wang et Zhang proposent une méthode multirésolution pour la recherche d'images qui donne des résultats encourageants. Chen [2], Bouman et Dalton utilisent une approche hiérarchique dans laquelle les paramètres extraits des images sont classés automatiquement en quatre familles à chaque niveau de l'arbre de recherche. L'image représentant chaque famille est l'image la plus proche du centre de gravité de la classe.

Notre précédente publication [9] décrit notre méthode. La figure 1 présente le schéma de notre approche. Le principe est assez simple : après l'extraction des paramètres multirésolution, une classification automatique en familles leur est appliquée. On obtient donc plusieurs images et plusieurs classes. La racine de l'arbre est composée des images représentatives obtenues. On réitère le processus sur les classes trouvées précédemment. On insère au niveau

inférieur de l'arbre les images représentatives construites à cette étape. Lorsque la classification des paramètres ne donne pas un bon résultat à un niveau : les classes ne sont plus séparables avec les paramètres à une certaine résolution, dans ce cas on relance la classification avec les paramètres plus fins (niveau suivant de la transformée en ondelettes). Ensuite, l'arbre ainsi construit est proposé à l'utilisateur pour sa recherche de façon exploratoire à l'aide d'un navigateur web. De cette manière, l'utilisateur peut naviguer directement dans la base de données et revenir en arrière s'il s'aperçoit que son choix n'était pas judicieux.

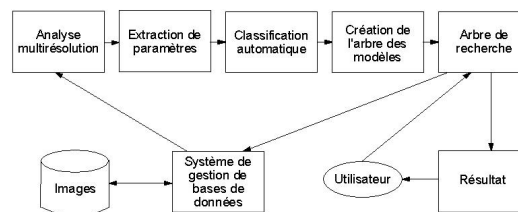


FIG. 1 – Principe général de la méthode.

Notre méthode est volontairement exploratoire plutôt que basée sur une comparaison d'images. L'utilisateur doit avoir une idée de ce qu'il cherche et navigue visuellement dans le système de recherche jusqu'à obtenir ce qu'il cherche. L'approche comparative est aussi possible, dans ce cas, il faut extraire les paramètres de l'image requête et la classer avec les autres puis commencer la recherche dans le sous-arbre dans lequel elle a été classée.

3 Méthode de recherche

La technique proposée emploie l'analyse multirésolution pour l'extraction des paramètres et l'algorithme des nuées dynamiques pour leur classification. Puis une image représentative est construite et un arbre est généré. Cet arbre sera proposé à l'utilisateur pour sa recherche exploratoire. Les paragraphes suivants décrivent chacune de ces étapes.

3.1 Analyse multirésolution

L'analyse multirésolution qui est employée par notre méthode a été définie par Mallat. De plus, elle possède des propriétés très intéressantes du point de vue de la recherche d'images :

- L'analyse multirésolution approche le comportement du système psychovisuel humain, c'est une propriété très intéressante pour un système de recherche d'images,
- Décomposition multirésolution : chaque image est décomposée en une image d'approximation et plusieurs images de détails à plusieurs niveaux de résolution. L'image d'approximation de basse résolution peut être utilisée comme icône pour représenter l'image d'origine,
- Des techniques de compression d'images sont basées sur la transformée en ondelettes et permettent de compresser les images de la base,
- Le tatouage d'images est possible à l'aide des coefficients d'ondelettes de basse résolution, il est donc possible de marquer les images pour des raisons de copyright,
- La transformée en ondelettes est réversible, les images peuvent être stockées sous leur forme transformée afin d'accélérer les traitements.

Le principe de l'analyse multirésolution est de projeter un signal $f(x) \in L^2(\mathbb{R})$ appartenant à un espace V_j sur un sous-espace V_{j+1} et un sous-espace W_{j+1} dans le but de réduire la résolution de moitié. Le schéma est donné en figure 2. Il existe donc un opérateur de projection A_j et un opérateur de projection D_j qui projettent respectivement le signal $f(x)$ sur V_{j+1} et W_{j+1} . V_{j+1} est le sous-espace d'approximation et W_{j+1} le sous-espace de détails. On peut démontrer qu'il existe une fonction d'échelle $\phi(x) \in L^2(\mathbb{R})$ qui engendre par dilatation et translation une base orthonormée de V_{j+1} et une fonction d'ondelettes $\psi(x) \in L^2(\mathbb{R})$ qui engendre par dilatation et translation une base orthonormée de W_{j+1} . Les espaces obtenus ne sont pas quelconques, ils possèdent des propriétés intéressantes. Par construction, les espaces d'approximation V_{j+1} et de détails W_{j+1} sont

supplémentaires : $V_j = V_{j+1} \oplus W_{j+1}$.

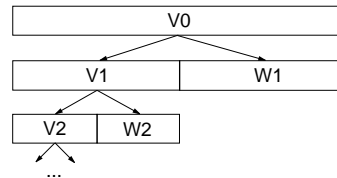


FIG. 2 – Principe de l'analyse multirésolution.

Les fonctions de bases dilatées sont données par les relations :

$$\phi_{j,n}(x) = 2^{-j/2} \phi(2^{-j}x - n) \text{ avec } n \in \mathbb{Z} \quad \text{et} \\ \psi_{j,n}(x) = 2^{-j/2} \psi(2^{-j}x - n) \text{ avec } n \in \mathbb{Z}$$

On a donc $A_j f = \sum_n \langle f, \phi_{j,n} \rangle \phi_{j,n}$ et $D_j f = \sum_n \langle f, \psi_{j,n} \rangle \psi_{j,n}$ où $\langle f(t), g(t) \rangle$ désigne le produit scalaire de $f(t)$ par $g(t)$: $\langle f(t), g(t) \rangle = \int_{-\infty}^{+\infty} f(t) \overline{g(t)} dt$

Puisque les fonctions utilisées appartiennent toutes à $L^2(\mathbb{R})$, on a $\overline{\overline{g(t)}} = g(t)$. On pose $a_{j,n} = \langle f, \phi_{j,n} \rangle$ et $d_{j,n} = \langle f, \psi_{j,n} \rangle$. $a_{j,n}$ et $d_{j,n}$ sont respectivement les coefficients d'approximation et de détails de la transformée en ondelettes de la fonction f . C'est à partir de ces coefficients que les paramètres représentant les images seront extraits.

Le résultat de l'analyse multirésolution d'une image est donné en figure 3. On voit la diminution de la résolution, l'image d'approximation et les images de détails horizontaux, verticaux et diagonaux.

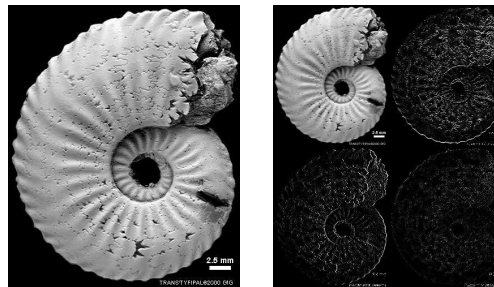


FIG. 3 – Un exemple de décomposition en ondelettes d'une image au premier niveau de résolution.

Une fois que les images sont transformées, il faut extraire des paramètres. Comme on l'a déjà signalé, les paramètres peuvent être de

trois types : locaux, globaux et mixtes. On peut imaginer des centaines de paramètres sur la transformée en ondelette, par exemple : l'énergie moyenne, les extrema d'énergie, les extrema de coefficients, dans l'approximation, dans chacune des images de détails. . . Les paramètres sont choisis en fonction de leur discrimination pour les images. Ils doivent avoir des valeurs proches pour des images visuellement proches et des valeurs éloignées pour les images éloignées. De leur choix dépend le temps de réponse et la performance globale du système de recherche d'images.

Les paramètres qui ont été utilisés sont les énergies normalisées des imagerie d'approximation et de détails. La normalisation d'un paramètre p indique simplement que l'on vérifie $p \in [0, 1]$ afin d'être homogène dans les calculs de distance de l'algorithme des nuées dynamiques.

Quand les paramètres sont extraits et normalisés, chaque image i est représentée par plusieurs vecteurs $v_{i,r} \in [0, 1]^p$ où p est le nombre de paramètres retenus et r est la résolution des ondelettes utilisées (quand r augmente, la résolution diminue). On utilise alors l'algorithme des nuées dynamiques dans l'hyper-espace $[0, 1]^p$ pour réaliser la classification des images.

3.2 Algorithme des nuées dynamiques

L'algorithme des nuées dynamiques est présenté en figure 4. Il permet de séparer un ensemble de points en un nombre donné de classes automatiquement. Plus on a de points à classer (un par image de la base, c'est-à-dire i), plus on a de paramètres (p) et plus on augmente la dimension du problème.

Les vecteurs $v_{i,r}$ sont donc classés et regroupés en familles grâce aux nuées dynamiques. Lorsque les classes ne sont plus suffisamment séparées (distance inter-centre très faible), on change de résolution en passant aux vecteurs $v_{i,r-1}$ et on relance l'algorithme avec une résolution plus fine. Après cette étape de classification, la construction de l'arbre proprement dite peut débuter.

-
- 1- Choisir au hasard C centres de classe (noyaux)
 - 2- Tant que non-stationnarité répéter {
 - 2.1- Affecter chaque point à la classe du centre le plus proche
 - 2.2- Calculer le centre de gravité des classes obtenues }
-

FIG. 4 – Algorithme des nuées dynamiques.

3.3 Construction de l'arbre de recherche

Dans notre méthode, les nuées dynamiques permettent une classification dans l'hyper-espace des coefficients d'ondelettes. Le nombre de familles C est paramétrable. Dans la suite, C sera égal à deux (arbre binaire de recherche). La condition de stationnarité est que les centres de gravité des classes soient identiques après deux itérations successives (c'est-à-dire que les familles obtenues soient stables).

La distance utilisée pour les nuées dynamiques est la distance euclidienne, ce n'est pas la plus rapide à calculer, mais elle est isotrope, c'est-à-dire qu'elle ne privilégie pas une direction particulière lors du calcul de distance. On a $d(u, v) = \sqrt{\sum_{j=1}^p (u_j - v_j)^2}$.

Pour chaque famille obtenue, une image représentative est construite. Dans notre démarche, l'image représentative de la base est la moyenne des images les plus proches (au sens de la distance euclidienne) du centre de la classe (dans la limite de 10% du nombre d'images total de la classe). Ces images sont ensuite organisées en arbre au fur et à mesure de la classification.

La partie suivante décrit les résultats expérimentaux mis en œuvre pour tester la méthode.

4 Résultats expérimentaux

La base d'images utilisée pour tester notre méthode est la base du laboratoire de Biogéosciences de l'université de Bourgogne.

Trans'tyfipal – pour transformée de types et figurés en paléontologie – contient des spécimens d'animaux et de plantes de l'apparition de la vie sur notre planète à nos jours. C'est une collection très importante (environ 60 000 images) dont les images sont assez hétérogènes. La figure 5 donne un exemple du contenu de Trans'tyfipal¹.

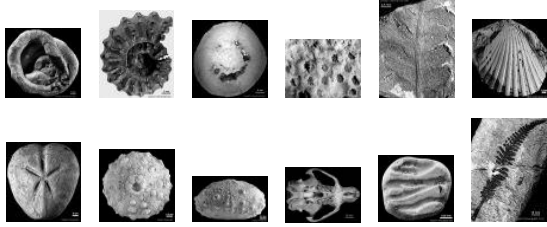


FIG. 5 – Quelques images issues de la base Trans'tyfipal.

La construction de l'arbre a été réalisée avec les paramètres suivants : deux classes à chaque niveau de l'arbre (deux représentants, arbre binaire de recherche), une transformée en ondelettes des images de la base à trois niveaux différents de résolution à l'aide des ondelettes de Haar. Les images de la base sont initialement codées en niveaux de gris.

L'arbre binaire de recherche obtenu et proposé à l'utilisateur est donné en figure 6. La recherche prend alors la forme d'une navigation à travers la base d'images.

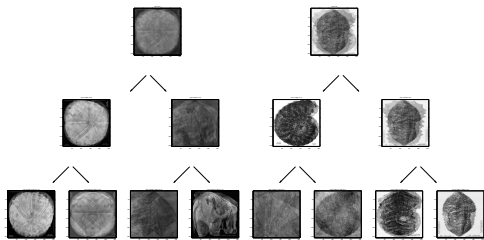


FIG. 6 – Arbre binaire de recherche où apparaissent les images modèle de chaque classe.

Le temps de calcul est une donnée primordiale dans les systèmes de recherche d'images, car il est difficile d'imaginer un utilisateur qui attend une heure devant sa machine pour obte-

¹Toutes les images de la base "Trans'tyfipal" sont la propriété du laboratoire de Biogéosciences de l'université de Bourgogne, Dijon, France.

nir un résultat de recherche. Les calculs ont été effectués sous Matlab 5 sur un Pentium II cadencé à 266 MHz équipé de 128 Mo de mémoire.

La figure 7 donne le temps de calcul pour différents niveaux de l'arbre et un nombre croissant de paramètres avec une base de 123 images. Il permet de mettre en avant le fait que le temps de calcul est pratiquement linéaire avec l'augmentation du nombre de paramètres. Le temps d'attente de l'utilisateur est réduit dans le cas d'un seul niveau pour l'arbre alors qu'il devient assez rapidement trop long lorsque l'arbre possède de nombreux niveaux.

Le temps de calcul peut bien sûr être amélioré en utilisant une machine plus puissante et un langage compilé pour l'exécution.

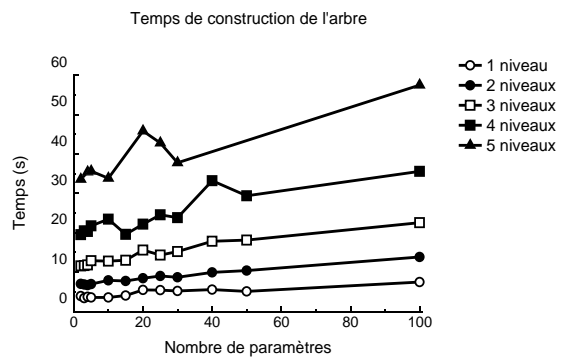


FIG. 7 – Temps de calcul en fonction du nombre de paramètres pour différents niveaux de l'arbre.

La figure 8 fournit le temps de calcul pour $p = 100$ paramètres avec augmentation progressive de la taille de la base de données. On constate à nouveau que le temps d'attente de l'utilisateur est pratiquement linéaire avec l'augmentation de la taille de la base. La même amélioration que précédemment peut être envisagée, c'est-à-dire la compilation de l'algorithme et l'augmentation de la puissance de calcul.

En ce qui concerne les images retrouvées, on s'aperçoit que dans l'ensemble, la séparation visuelle est assez bonne (avec toujours quelques exceptions, notamment pour les images en bordure de regroupement). La fi-

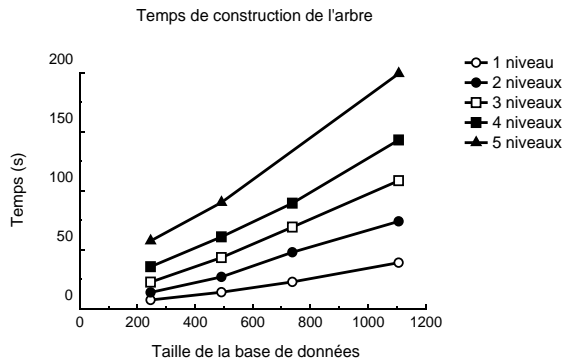


FIG. 8 – Temps de calcul en fonction de la taille de la base de données par différents niveaux de l'arbre.

gure 9 donne un exemple d'une famille de 29 images ainsi que le représentant construit à partir des trois images les plus proches du centre de la classe. La séparation visuelle des familles d'objets paléontologiques est assez bonne en dehors des images éloignées du centre de gravité de la classe.

Les graphiques classiques de recherche dans les bases d'images que sont précision et rappel ne sont pas portés dans ce document pour la simple raison qu'ils ne peuvent pas y être calculés. En effet, les noyaux des nuées dynamiques étant affectés au hasard, l'algorithme ne donne pas le même résultat s'il est lancé deux fois successivement. On ne peut donc pas évaluer le rappel et la précision dans notre cas.

5 Conclusion

Cet article a présenté un travail sur la recherche d'images par le contenu dans une base d'images de paléontologie. La difficulté principale du sujet est que l'information portée par une image n'est pas quantifiable. Il existe donc une part d'incertitude quant aux résultats obtenus. Pour assouplir cette contrainte, il est possible d'autoriser certaines images à faire partie de plusieurs classes lorsqu'elles sont très éloignées du centre de la classe.

L'utilisation de paramètres multirésolution permet une recherche à plusieurs niveaux de résolution (des résolutions grossières aux résolutions fines), ce qui permet une recherche

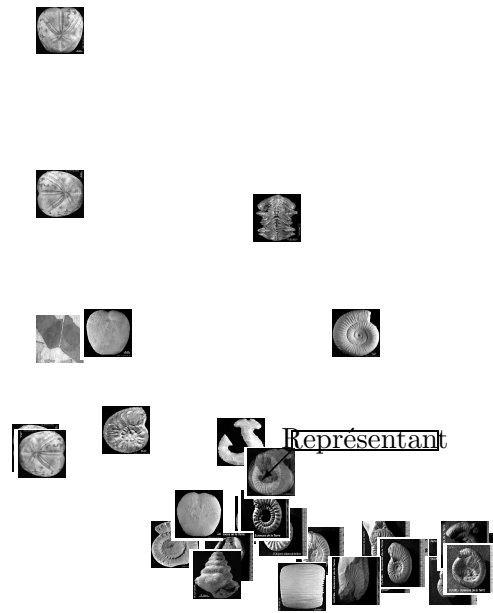


FIG. 9 – Un exemple de famille formé à partir de 29 images, les images sont représentées en fonction de leur distance réelle au centre de la classe (image représentative indiquée sur le schéma).

rapide au départ et plus poussée à la fin. La navigation par un client web simplifie l'accès aux bases d'images. Les temps de calculs restent faibles et ceci même avec une augmentation importante du nombre de paramètres et du nombre de niveaux dans l'arbre.

Il est possible d'imaginer une classification automatique en K classes (au lieu de deux pour notre publication) avec K variable d'un niveau de l'arbre à un autre (l'arbre obtenu peut être non-balancé avec des sous-arbre K -aire où K varie). Lors de la création de l'image modèle de la classe, une pondération des images en fonction de leur distance au centre de gravité de la classe est envisageable et permettrait de donner un poids plus important aux images proches du centre de la classe.

L'ajout des informations de couleur dans un espace RGB ou HSV par exemple peut permettre d'améliorer les informations dont on dispose afin de dégager des paramètres d'ondelettes discriminants. L'utilisation d'ondelettes plus élaborées que celles de Haar comme les symlettes ou les splines permettrait d'ob-

tenir une meilleure séparation des coefficients d'ondelettes à chaque niveau de résolution.

Les travaux de D.E.A. de A. Da Costa [4] encadré par M.-N. Terrasse et E. Leclercq proposent des améliorations pour la génération aléatoire des noyaux, l'ajout de données sémantiques...

De nombreux travaux sont encore nécessaires pour arriver aux systèmes de recherche d'images de la prochaine génération...

Références

- [1] Wendy Chang, Gholamhosein Sheikholeslami, Jia Wang, and Aidong Zhang. Data resource selection in distributed visual information systems. *TKDE*, 10(6) :926–946, 1998.
- [2] Jau-Yuen Chen, Charles A. Bouman, and John C. Dalton. Hierarchical browsing and search of large image databases. *IEEE Transactions on Image Processing*, 9(3) :442–455, March 2000.
- [3] Li Chen and Li Chien. Color image segmentation using progressive wavelet transform for image retrieval. *Proceedings of ISIVC'2000 - Rabat - Morocco*, pages 245–252, 2000.
- [4] Arnaud Da Costa, Eric Leclercq, and Marie-Noëlle Terrasse. Recherche d'images par le contenu : une approche exploratoire multirésolution paramétrable. Technical report, D.E.A. Instrumentation et Informatique de l'Image - université de Bourgogne - Dijon - France, 2001.
- [5] Minh N. Do. Invariant image retrieval using wavelet maxima moment. Technical report, École Polytechnique Fédérale de Lausanne, 1998.
- [6] Nathalie Guyader and Jeanny Hérault. Représentation espace-fréquence pour la catégorisation d'images. *Actes du colloque GRETSI (CD-ROM) article numéro 397*, septembre 2001.
- [7] F. Idris and S. Panchanathan. Image indexing using wavelet vector quantization. *SPIE*, 2606 :269–275, 1995.
- [8] Charles E. Jacobs, Adam Finkelstein, and David H. Salesin. Fast multiresolution image querying. *Proceedings of SIGGRAPH95, Los Angeles, California*, August 1995.
- [9] Jérôme Landré, Frédéric Truchetet, Sophie Montuire, and Bruno David. Automatic building of a visual interface for content-based multiresolution indexing and retrieval of paleontology images. *Journal of Electronic Imaging - special issue*, 10(4) :957–965, october 2001.
- [10] Kai-Chieng Liang and C. C. Jay Kuo. Progressive indexing, retrieval and transmission of wavelet compressed image database. *SPIE*, 3169 :190–199, 1997.
- [11] Stéphane G. Mallat. A theory for multiresolution signal decomposition : The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7) :674–693, July 1989.
- [12] M. K. Mandal, T. Aboulnasr, and S. Panchanathan. Illumination invariant image indexing using moments and wavelets. *Journal of electronic Imaging*, 7(2) :282–293, 1998.
- [13] Raymond T. Ng and Jiawei Han. Efficient and effective clustering methods for spatial data mining. *Proceedings of the 20th Very Large Database Conference - Santiago - Chile*, pages 144–155, 1994.
- [14] Zoran Pečenović. Finding rainbows on the internet. Technical report, École Polytechnique Fédérale de Lausanne, 1998.
- [15] Yong Rui, Thomas S. Huang, and Shih-Fu Chang. Image retrieval : Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, 10 :1–23, 1999.
- [16] Matthieu Vissac and Jean-Luc Dugelay. Un panorama sur l'indexation d'images fixes. *Proceedings of ISIVC'2000 - Rabat - Morocco*, pages 72–80, 2000.