

Recherche par contenu visuel dans les grandes collections d’images

Valérie Gouet-Brunet

Mots-clés : descripteurs d’image, structures d’index multidimensionnels, malédiction de la dimension.

Résumé : à l’heure où la technologie numérique permet de produire, stocker et disséminer des volumes importants de données multimédia, automatiser la recherche d’images devient un enjeu crucial pour bon nombre d’applications. L’indexation d’une image à partir de l’analyse de son contenu visuel est une approche récente qui permet de réduire les ambiguïtés ou lacunes des techniques maintenant classiques d’indexation par le texte. Elle devient même la seule solution lorsque le volume d’images rend l’annotation textuelle irréalizable. La recherche par contenu visuel dans les grandes collections d’images est une problématique émergente née de l’intersection de deux domaines de recherche en informatique : l’analyse d’images et les bases de données. Dans ce chapitre, nous aborderons les solutions apportées par ces deux domaines pour cette problématique. Nous nous attacherons également à présenter les concepts et les nouvelles solutions qui en font désormais un domaine de recherche à part entière.

1 Introduction

La technologie numérique permet aujourd’hui de nouvelles applications multimédia pour de nombreuses communautés. Par exemple, les professionnels de l’audiovisuel tels que les grandes chaînes de télévision numérisent et archivent les journaux télévisés, de même que l’Institut national de l’audiovisuel avec le patrimoine audiovisuel français. À la Bibliothèque nationale de France, les chercheurs numérisent les fonds patrimoniaux, dont une grande partie est disponible gratuitement grâce au projet Gallica. Et tout simplement le grand public avec la démocratisation de l’imagerie numérique, au travers des magnétoscopes et autres appareils-photos numériques. Ces nouvelles fonctionnalités conduisent à la production et au stockage de grands volumes de données numériques. Celles-ci peuvent être de plus en plus facilement disséminées avec l’essor des nouvelles techniques de diffusion comme la télévision numérique ou des réseaux ouverts comme l’Internet (catalogues en ligne, connections pair-à-pair, etc). Rechercher une ou plusieurs images dans une telle masse de données numériques n’est plus trivial. Le volume produit rend désormais insuffisantes et inadaptées les approches classiques d’exploration par le texte telles que l’utilisation de mots-clés ou de méta-données associés aux images. D’autres modes doivent être envisagés. L’indexation d’une image à partir de l’analyse de son contenu visuel est une alternative récente qui permet de réduire les ambiguïtés ou lacunes des techniques d’indexation par le texte. Elle devient même la seule solu-

tion lorsque le volume d’images rend l’annotation textuelle irréalizable.

La problématique de la recherche par contenu visuel dans de grands volumes d’images se situe à l’intersection de deux domaines de l’informatique : l’analyse d’images et les bases de données. Les chercheurs en bases de données ont depuis longtemps proposé une large palette de méthodes d’optimisation permettant de faciliter la recherche dans les grands volumes de données. Jusqu’à récemment, les chercheurs en analyse d’images ont exclusivement concentré leurs efforts sur la mise au point d’approches de description du contenu visuel de l’image qui soient performantes en termes de reconnaissance. Les architectures développées étaient rarement organisées autour de bases de données et notamment travaillaient pour la plupart en mémoire principale, sans l’aide d’aucune structure d’index. Les principales raisons invoquées étaient que d’une part, les volumes de données manipulés permettaient de se passer d’une organisation physique particulière et que d’autre part, les structures d’index proposées en bases de données n’étaient pas bien adaptées aux informations complexes extraites des images. Mais aujourd’hui, l’augmentation du volume des bases d’images et le développement de techniques de reconnaissance de plus en plus sophistiquées font que la problématique du *passage à l’échelle* des structures traditionnelles d’index trouve un regain d’intérêt auprès de ces deux communautés.

Développer un système de recherche d’images par contenu visuel (SRI) qui soit efficace pour les

grandes collections d'images requiert ainsi une expertise à la fois en analyse d'images et en bases de données. En particulier, pour être performantes, les approches de description étudiées doivent tenir compte des problèmes inhérents aux espaces de grande dimension, de même que les structures d'index multidimensionnels doivent être adaptées à la distribution des descripteurs d'images.

Classiquement, un système de recherche d'images par contenu visuel comporte une phase hors ligne d'*indexation* de la base d'images et une phase en ligne de *recherche* à proprement dit, que nous décrivons dans les deux sections suivantes.

1.1 Indexation de la base d'images

La phase d'indexation suit deux étapes successives :

1. *Extraction des caractéristiques visuelles.* Cette étape consiste à extraire de l'image un résumé de son apparence visuelle par des méthodes d'analyse d'images. La méthode employée pour extraire ce résumé porte le nom de *descripteur* d'image. Après extraction des caractéristiques visuelles, le contenu visuel de l'image est décrit par un vecteur numérique (ou un ensemble de vecteurs), que l'on appelle généralement *signature* de l'image. La mise en place d'un descripteur consiste également à associer aux signatures une *mesure de similarité*. Au moment de la recherche, deux images seront jugées similaires si leurs signatures sont similaires au sens de cette mesure de similarité. Le descripteur n'est pas universel pour une image donnée, l'information extraite devant être pertinente pour l'usage visé. Il doit être fidèle au contenu *syntactique* et *sémantique* de l'image tout en étant *compact*. La section 2 passe en revue les principales catégories de descripteurs d'images existant.
2. *Structuration de l'espace des signatures.* Indexer l'espace des signatures consiste à organiser cet espace selon une structure d'index (généralement multidimensionnel) permettant d'accélérer la recherche. Si l'ensemble de l'espace et de la structure ne peuvent pas être chargés en mémoire principale, l'optimisation porte sur la réduction du nombre d'accès au disque. Les approches les plus récentes s'attachent également à réduire les coûts en temps CPU correspondant généralement au calcul de la mesure de similarité. La recherche via un index doit naturellement être plus rapide qu'un par-

cours exhaustif de l'espace. Pour être efficace, la structure développée doit prendre en compte les propriétés de l'espace des signatures. Cet aspect est traité à la section 3. Les principales techniques de structuration sont quant à elles décrites à la section 4.

Le rapprochement de deux communautés de recherche peut poser quelques problèmes de vocabulaire. Notamment, il est important de souligner que le terme "indexation" est employé dans la communauté de l'analyse d'images comme dans celle des bases de données, mais n'y revêt pas la même signification ! En analyse d'images, *indexer une image* correspond à l'étape (1) d'extraction des caractéristiques visuelles. Ici, l'index est la signature de l'image. La communauté des bases de données parle quant à elle d'*indexation de la base d'images*, ce qui correspond à l'étape (2) de structuration de l'espace des signatures. Dans ce cas, l'index est la structure issue de cette organisation.

1.2 Recherche dans la base d'images

Une fois la base d'images indexée, la phase de recherche peut revêtir plusieurs formes. La plus ancienne et la plus classique est la *recherche par similarité visuelle à partir d'un exemple*. Dans sa forme la plus simple, le système reçoit en entrée une image exemple et retourne l'ensemble des images les plus similaires à cette image, au sens de la mesure de similarité associée à la signature. La recherche implique de trouver *les plus proches voisins* de la signature associée à l'image requête (voir section 4). Ce mode est illustré par l'exemple (a) de la figure 1.1 obtenu à partir du moteur IKONA¹. La requête peut aussi porter sur une partie ou un objet de l'image requête, voir l'exemple (b) de la figure 1.1 ; dans ce cas, le système retourne les images impliquant les parties/objets les plus similaires. La recherche par l'exemple convient bien à une grande partie des besoins, comme nous le verrons à la section 5 à travers deux applications.

2 Aperçu des méthodes d'extraction de caractéristiques visuelles

Les méthodes de description de l'image dépendent des applications considérées. Pour des bases d'images au contenu hétérogène (on parle de bases d'images *génériques*), il existe deux classes de descripteurs d'images : les descripteurs *globaux* qui sont présentés à la section 2.1 et les descripteurs *locaux* présentés à la section 2.2. Lorsque les

¹IKONA est un moteur de recherche d'images par contenu visuel développé par l'INRIA : <http://www-rocq.inria.fr/imedia/ikona.html>.



(a) Recherche globale approximative, à partir de descripteurs de couleur, texture et forme

(b) Recherche locale précise, à partir de descripteur de points d'intérêt

FIG. 1.1 – Recherche d’images par contenu visuel : Deux scénarios de recherche par l’exemple dans une base d’images généraliste (sources : Images Du Sud et INA) à partir du moteur de recherche IKONA. Chaque requête porte sur le rectangle blanc de la photo présentée en haut à gauche dans l’interface de visualisation. La requête (a) est réalisée sur l’image entière et la requête (b) sur une partie de l’image (un tournesol) détournée manuellement par l’utilisateur. Les photos retournées sont classées par ordre décroissant de similarité visuelle, seules les neuf premières réponses sont présentées ici. On observe que la requête (a) retourne des images traduisant une ambiance globale de « jardins en fleur » alors que la requête (b) permet de retrouver des images contenant précisément l’objet d’intérêt.

bases d’images sont relatives à un domaine *spécifique*, des descripteurs spécifiques sont alors préconisés (cf. section 2.3).

2.1 Description globale de l’image

En indexation d’images, l’apparence visuelle globale d’une image est résumée par trois attributs de bas niveau : la *couleur*, la *texture* et la *forme*. Ces attributs ont été intégrés dans la norme MPEG-7². Nous les présentons brièvement ci-dessous et renvoyons le lecteur au chapitre 4.1 Indexation des données et recherche dans les bases multimédia de la section 4.1.2 Systèmes multimédia pour plus de détails.

La *couleur* est l’un des éléments les plus représentatifs du contenu d’une image. Les travaux de Swain et Ballard [Swain and Ballard, 1991] re-

latifs à l’indexation de la couleur constituent la référence et marquent même les débuts de l’indexation d’images. De nombreux espaces colorimétriques existent : RVB est le plus classique mais des espaces comme CIELab ou CIELuv s’avèrent mieux adaptés à la perception humaine de la couleur et donc à la recherche d’images par similarité visuelle. La distribution des couleurs dans l’image est le plus souvent décrite par un histogramme couleur qui représente la fréquence d’apparition des différentes couleurs. Un tel descripteur est robuste aux principales transformations géométriques que peut subir l’image, ce qui le rend bien adapté à une recherche approximative globale.

Caractériser la *texture* de l’image fournit une bonne information sur l’arrangement structurel des surfaces dans l’image. Deux grandes classes

²MPEG-7 : standard ISO/IEC pour la description de contenus multimédia. Parmi un large spectre d’applications, cette norme traite de la problématique de l’indexation d’images (fixes ou animées) par contenu visuel. Pour une introduction à MPEG-7, le lecteur peut consulter le chapitre 4.1 Indexation des données et recherche dans les bases multimédia de la section 4.1.2 Systèmes multimédia.

d'approches existent, l'une est basée sur des mesures statistiques alors que l'autre implique des mesures fréquentielles. Une approche statistique répandue est basée sur une analyse psychovisuelle de la texture particulièrement bien adaptée pour l'indexation d'images. Quant aux approches fréquentielles, elles proviennent largement du domaine du traitement du signal. Citons la transformée de Fourier et ses variantes, les filtres de Gabor et la transformée en ondelettes.

La description de la *forme* ou de la structure de l'image requiert en général deux niveaux de traitement. Elle nécessite en premier lieu une segmentation de l'image en régions ou au moins une extraction des contours. Ensuite, outre l'utilisation de caractéristiques simples comme la longueur ou la courbure des formes extraites, une technique simple consiste à caractériser la distribution de l'orientation des gradients sur les contours, sous la forme d'un histogramme par exemple. D'autres approches plus sophistiquées existent, citons le codage par chaînes (codage de Freeman, codage différentiel), les descripteurs de Fourier, les moments de Hu ou de Zernike.

Les trois classes de descripteurs qui viennent d'être présentées étant complémentaires, elles peuvent être combinées selon les besoins. La figure 1.1(a) illustre un scénario de recherche à partir d'une combinaison linéaire de ces descripteurs. Dans la mesure où elles caractérisent l'image dans sa globalité, ces approches permettent une recherche que l'on qualifie d'*approximative*. Elles rendent compte d'une *ambiance visuelle* mais ne permettent pas la recherche de parties d'images ou d'objets. Les sections 2.2 et 2.3 ci-après concernent ce type de requêtes.

2.2 Description locale de l'image

A l'opposé des approches globales, les approches de description locale visent à décrire le contenu de l'image *localement*. Elles offrent ainsi la possibilité d'effectuer une recherche sur une partie de l'image ou encore sur un objet présent dans l'image. La sélection étant définie explicitement sur une partie de l'image, on a l'habitude de leur associer le paradigme de *requêtes partielles*. Pour les bases d'images génériques, deux techniques de description entrent dans ce cadre. Elles mettent en jeu des primitives de l'image de nature différente :

Des régions d'intérêt : La phase d'indexation de l'image consiste à découper celle-ci en régions selon une approche de segmentation. Lorsque l'on n'a aucune connaissance a priori sur le contenu de la base d'images, il est difficile de réaliser une segmentation qui corresponde précisément aux objets que l'on souhaitera reconnaître par la suite. En général, il est préférable que la segmentation réalisée reste assez grossière,

de façon à produire un petit nombre de régions qui constituent des zones visuellement cohérentes et pertinentes comme requête. Pour être discriminantes, ces régions doivent présenter une certaine variabilité photométrique, qu'il faut décrire finement. Nous renvoyons le lecteur à l'ouvrage [Boujemaa et al., 2004] ainsi qu'au chapitre 4.4 Indexation des données et recherche dans les bases multimédia de la section 4.4 Systèmes multimédia pour un éventail des techniques de segmentation en régions et les méthodes de caractérisation associées.

Des points d'intérêt : L'image est décrite dans son intégralité par un ensemble de *points d'intérêt*, également appelés *points clés* ou *points saillants*, qui sont extraits automatiquement pendant la phase d'indexation. Ces points représentent les sites de l'image caractérisés par une forte variabilité locale, jugée discriminante et plutôt stable face au bruit et aux transformations photométriques ou géométriques que peut subir l'image. C'est cette variabilité locale modélisée pour chaque point extrait qui est retenue comme descripteur du point. L'ensemble de vecteurs obtenu constitue un *résumé compact* de l'image. Au moment de la recherche, l'utilisateur définit dynamiquement la partie de l'image qui l'intéresse et seul le sous-ensemble de points d'intérêt correspondant est recherché dans la base. En outre, la répartition spatiale des points dans l'image rend cette technique robuste aux occultations. La figure 1.1(b) illustre la notion de requêtes partielles par points d'intérêt. De nombreuses méthodes d'extraction et de caractérisation de points d'intérêt existent dans la littérature, un point de départ est [Boujemaa et al., 2004]. Le lecteur peut aussi se référer au chapitre 4.4 Le traitement des images numériques de la section 4.4 Calcul Scientifique qui traite de l'extraction de telles primitives dans les images.

Les deux approches venant d'être présentées ne décrivent pas la même information visuelle dans l'image, ce qui les destine à des usages différents : les points d'intérêt se prêtent mieux à la recherche fine d'objets ou de parties d'images au contenu hétérogène, alors que l'approche basée sur une segmentation en régions se focalise sur les zones dominantes de l'image. Le support région étant déterminé pendant la phase d'indexation, la sélection au moment de la recherche se limite à choisir parmi les régions proposées. Le support point permet quant à lui une sélection interactive de la zone de recherche. En contrepartie, le prix à payer pour cette flexibilité est l'extrême lourdeur du descripteur, qui impose l'utilisation de structures d'index multidimensionnelles pour être opérationnel. Le choix entre ces deux approches doit être motivé par l'application ou par les be-

soins de l'utilisateur en terme de précision de la recherche au prix du temps de réponse. En tout cas, leur complémentarité en fait un atout pour répondre aux différents scénarios de requêtes partielles [Boujemaa et al., 2004].

2.3 Descripteurs spécifiques

Les descripteurs sus-cités ne supposent aucune connaissance a priori sur la nature des images/objets à indexer. Par descripteurs spécifiques, on entend des descripteurs dédiés à des bases d'images ou des classes d'objets relatifs à un domaine particulier, la biométrie ou la numismatique par exemple. Ces descripteurs exploitent la spécificité de la classe des objets à reconnaître et s'attachent à caractériser la variabilité des objets dans cette classe. Idéalement, cette variabilité est estimée par apprentissage statistique à partir d'un ensemble de vues représentatives des objets de la classe. Pour la reconnaissance de visages, les approches *EigenFaces* et *FisherFaces*, étudiées et comparées par exemple dans [Belhumeur et al., 1997], illustrent bien ce type d'approche.

2.4 Composition de requêtes

Rechercher dans une base de photos aériennes une image représentant "une caserne à côté d'un aéroport dans le désert" implique l'utilisation conjointe de plusieurs des descripteurs que nous avons passés en revue. Dans ce cas, la requête est composée de plusieurs parties d'image. Mais surtout, elle met en jeu des relations ou contraintes spatiales ("à côté de", "dans") entre les différents éléments qui la composent. Ces relations sont souvent modélisées par des graphes d'adjacence entre régions. Le lecteur peut par exemple consulter [Rigaux et al., 2001] pour plus de détails sur les bases de données spatiales.

3 Structure des espaces de description

Les différents descripteurs d'images que nous avons passés en revue dans la section précédente font apparaître plusieurs catégories d'espaces de description pour lesquels nous proposons une typologie à la section 3.1. Dans la section 3.2, nous introduisons quelques-uns des phénomènes bien connus qui se produisent lorsque la dimension de l'espace de description est grande. Une alternative pour minimiser ces effets est de réduire cette dimension. Les approches de réduction de la dimension sont décrites à la section 3.3.

3.1 Une typologie des espaces de description

La distribution des signatures dans l'espace de description des images dépend naturellement des descripteurs employés et des images de la base. Néanmoins à l'heure actuelle, une signature

d'image se place généralement dans l'une des deux catégories suivantes :

- elle est représentée par un seul vecteur par image. C'est le cas avec les descripteurs globaux ou bien avec certains descripteurs spécifiques comme les descripteurs de visage. L'espace engendré est généralement de grande dimension et moyennement peuplé (par rapport à l'autre catégorie). Par exemple, une base de 100000 images décrites par un histogramme couleur échantillonné sur 216 couleurs conduit à un espace de 216 dimensions contenant 100000 points ;
- elle est représentée par plusieurs vecteurs de même dimension par image, avec potentiellement des relations ou contraintes entre vecteurs. C'est le cas avec l'approche par descripteurs locaux. L'espace engendré est de moindre dimension mais contient un grand nombre de vecteurs. Par exemple, une base de 100000 images décrites par 300 points d'intérêt caractérisés par 17 invariants conduit à un espace de 17 dimensions contenant 3000000 points reliés par des contraintes géométriques.

Une telle typologie fait apparaître des catégories d'espaces de description spécifiques aux bases d'images, qui diffèrent fondamentalement en termes de *dimension des données* et de *taille de la population des données*. Pour être efficaces, nous verrons que les techniques d'indexation multidimensionnelle ne doivent négliger aucun de ces deux paramètres. Des exemples de signatures de différente nature sont donnés dans le tableau 1.1.

3.2 La malédiction de la dimension

Des phénomènes mathématiques particulièrement non intuitifs peuvent être observés quand la dimension de l'espace des données augmente. Ces effets portent le nom de *malédiction de la dimension*, qui est une traduction de l'anglais *curse of dimensionality*. Ils sont notamment passés en revue dans les études [Weber et al., 1998, Böhm et al., 2001] menées pour l'indexation multidimensionnelle, mais seulement dans le cadre d'une *distribution uniforme* des données. De nombreuses techniques d'indexation sont basées sur le partitionnement de l'espace en volumes englobant les données. Or on peut observer que les volumes et les surfaces augmentent de manière exponentielle avec la dimension de l'espace, ce qui conduit à une forte augmentation du temps de réponse des algorithmes de recherche. Au-delà d'une certaine dimension, un parcours séquentiel de l'espace devient même plus performant qu'un parcours d'index.

Une autre observation révèle que le nombre

Descripteur	Classe	Taille/image	Dimension
Histogramme couleur	global	1	> 200
Filtres de Gabor	global	1	~48
Régions d'intérêt	local	2-20	~20
Points d'intérêt	local	100-500	8-30

TAB. 1.1 – Caractéristiques des espaces de description associés à différents descripteurs d'images. Les taille et dimension données sont des ordres de grandeur représentatifs des descripteurs existants.

de cellules obtenues par partitionnement de l'espace croît exponentiellement avec la dimension. Elle porte généralement le nom de *phénomène de l'espace vide* qui indique que plus la dimension est grande, plus le nombre de cellules vides est important pour un nombre de vecteurs fixe.

Un autre phénomène, dit de *concentration des mesures*, est lié aux distances entre vecteurs : en grande dimension, les distances deviennent statistiquement identiques. Dans ces conditions, le bruit provenant de l'acquisition des données perturbe trop les métriques habituelles.

D'autres effets encore sont observables [Weber et al., 1998, Böhm et al., 2001]. Tous mettent clairement en lumière que les structures d'index multidimensionnels, très souvent pensées pour des petites dimensions, peuvent devenir inefficaces en grande dimension. Ceci nous amène naturellement à poser la question suivante : *quelle est la limite entre les espaces de faible et de grande dimension ?* La littérature actuelle ne fournit malheureusement pas de réponse unique et définitive. Au premier abord, certains considèrent qu'un espace est de grande dimension lorsqu'il dépasse les trois dimensions du monde réel, devenant ainsi non visualisable.

Dans le cadre de l'indexation multidimensionnelle, une solution consiste à dire qu'un espace est de grande dimension dès lors qu'un parcours séquentiel est plus performant qu'un parcours d'index. L'analyse quantitative faite par Weber et al. [Weber et al., 1998] sur des données suivant une distribution uniforme, montre ainsi qu'à partir d'une dimension moyenne de 10, un parcours exhaustif est plus rapide qu'un parcours d'index tels que le X-tree ou le R*-tree.

L'année d'après, en apportant une preuve mathématique au phénomène de concentration des mesures, Beyer et al. [Beyer et al., 1999] ont validé expérimentalement que, sous certaines hypothèses d'indépendance sur la distribution des données, la recherche des plus proches voisins devient réellement instable lorsque la dimension dépasse 20.

Il n'y a cependant pas d'école unique : les chercheurs en apprentissage statistique sont aujourd'hui eux aussi confrontés à des données de

grande dimension, ce qui les obligent à repenser les modes de représentation, d'accès et de traitement des données issus des années 90. Dans ce contexte, une proposition intéressante a été faite plus récemment dans [Verleysen, 2003] : elle consiste à mettre en rapport la dimension avec la taille de la population de l'espace. Faisant référence au phénomène de l'espace vide sus-cité, l'auteur pose qu'un espace est *de grande dimension* (ou *de petite taille*) quand la taille de sa population n'augmente pas exponentiellement avec sa dimension. Bien que difficilement évaluable, cette analyse illustre bien l'idée que *taille* et *dimension* des données sont intimement liées en indexation multidimensionnelle. Appliquée aux bases d'images, une telle proposition étiquette la première catégorie d'espace introduite à la section 3.1 comme étant de grande dimension.

Face à la problématique de la malédiction de la dimension, le domaine de l'analyse des données peut fournir des alternatives : celles-ci consistent à réduire la dimension de l'espace en réorganisant l'information portée par les données. Nous en abordons les grands principes dans la section suivante.

3.3 Réduction de la dimension

Par manque de connaissance sur la distribution des données collectées, il peut arriver que des relations existent entre composantes de vecteurs, ou même que certaines composantes véhiculent une information non réellement pertinente. L'espace engendré exprime alors plusieurs fois le même type d'information ou est caractérisé par des dimensions superflues. Cette information inutile peut être éliminée ou au moins réduite en déterminant la *dimension intrinsèque* des données : c'est la dimension du plus petit sous-espace qui permet de représenter les données tout en conservant la même quantité d'information.

Les méthodes qui permettent de transformer les données dans l'espace correspondant à leur dimension intrinsèque portent le nom de *méthodes de réduction de la dimension*. Lorsqu'elles sont linéaires, ces méthodes consistent à estimer une transformation linéaire entre l'espace original et un nouvel espace dans lequel chaque axe repré-

sente la quantité d'information portée par la composante associée. L'approche linéaire la plus répandue est certainement l'analyse en composantes principales (ACP), pour laquelle il existe plusieurs variantes [Gerbrands, 1981]. Le principe général de l'ACP réside dans l'estimation d'un nouvel espace dont les axes sont orthogonaux et assurent une dispersion maximale de chaque composante. La réduction de la dimension s'effectue en éliminant les axes porteurs de peu d'information, autrement dit ceux pour lesquels la variance des données est faible.

Des approches non linéaires ont également été proposées. Elles permettent de tenir compte de dépendances non linéaires entre composantes et sont en général moins sensibles au bruit des données que les approches linéaires. En revanche, leur inconvénient majeur est que la dimension de l'espace réduit doit être fixée a priori. La plupart d'entre elles sont itératives et consistent à minimiser une fonction de coût relative aux distances entre vecteurs. Les plus connues relèvent principalement de la statistique, citons l'analyse en composantes curvilinéaires [Demartines and Héroult, 1997] et le cadrage multidimensionnel [Cox and Cox, 2001]. En parallèle, l'approche *FastMap* [Faloutsos and Lin, 1995] a été spécifiquement développée à des fins d'indexation de grands volumes de données multidimensionnelles. Elle est purement heuristique, mais contrairement aux autres approches, elle possède les avantages d'avoir une complexité linéaire et de permettre les insertions.

A l'heure actuelle, il n'existe malheureusement pas de solution idéale pour réduire efficacement les espaces de données de grande dimension. En effet, il est rare que dans la pratique les données aient des composantes en dépendance linéaire et soient exemptes de bruit, limitant ainsi les approches linéaires. Quant aux approches non linéaires, elles semblent difficilement exploitables puisqu'elles nécessitent généralement la connaissance de la dimension intrinsèque des données.

4 Indexation multidimensionnelle

Les requêtes associées aux bases de données image sont différentes de celles associées aux bases de données traditionnelles, où la recherche dite *exacte* est la plus courante ("Je recherche les employés âgés de 25 ans"). Comme cela a déjà été dit dans la section 1, c'est principalement la recherche d'images par similarité à partir d'un exemple qui fait référence en indexation d'images. Elle peut revêtir plusieurs formes, que nous explicitons à la section 4.1. La section 4.2 dresse ensuite un panorama des principales structures d'index multidimensionnelles existantes et est illustrée en donnant le principe général de la structure M-Tree. Enfin,

pour clôturer ce panorama, nous présentons à la section 4.3 le nouveau concept de recherche qu'est la recherche approximative.

4.1 Taxonomie des types de requêtes

Qu'il s'agisse de descripteurs globaux, locaux ou spécifiques, la recherche d'images par similarité visuelle à partir d'un exemple passe par la recherche des vecteurs voisins d'un vecteur requête q . Cette opération peut se décliner en trois stratégies :

1. *Recherche par intervalle* : chercher les vecteurs r dont les composantes r_i sont telles que : $|q_i - r_i| < s_i$, où les $\{s_i\}$ représentent les seuils associés à chaque dimension. En d'autres termes, il s'agit de déterminer les vecteurs présents dans un hyper-rectangle centré en q .
2. *Recherche des k plus proches voisins* : chercher les k vecteurs les plus proches de q , au sens de la mesure de similarité associée aux vecteurs.
3. *Recherche à ϵ près* : chercher les vecteurs se trouvant à une distance inférieure à un seuil ϵ , au sens de la même mesure de similarité. Si par exemple cette mesure est la distance euclidienne, cela revient à déterminer l'ensemble des vecteurs présents dans l'hyper-sphère (q, ϵ) .

Très classique en bases de données spatio-temporelles où la dimension des données est faible (≤ 4), la recherche par intervalle n'a que peu d'applications en recherche d'images par contenu visuel. En effet, si la signature d'une image est représentée par un ou plusieurs vecteurs, rechercher des images similaires revient plutôt à déterminer les vecteurs les plus proches au sens de la mesure de similarité qui leur est associée.

Les stratégies (2) et (3) sont donc mieux adaptées à la recherche d'images. Elles diffèrent fondamentalement de part la nature de leurs paramètres : k est fonction du contenu de la base puisqu'il doit correspondre au nombre de vecteurs similaires au vecteur requête, alors que le paramètre ϵ est généralement lié à la méthode de description de l'image et plus précisément à la variabilité de la signature d'image pour un ensemble d'images similaires. En conséquence, ces stratégies ont certains avantages et inconvénients : k est en général facile à fixer et la recherche des k plus proches voisins garantit k vecteurs résultats. En revanche, les réponses peuvent impliquer des vecteurs trop éloignés correspondant à des images non similaires. Ce problème est contourné avec la recherche à ϵ près, mais celle-ci nécessite une bonne connaissance de la distribution des vecteurs pour déterminer ϵ . Une telle connaissance est souvent diffi-

cile à modéliser. Une solution consiste à estimer ϵ par apprentissage de la variabilité de la description sur un jeu de données, ce que l'application ne rend pas toujours possible.

La plupart des structures d'index qui vont être présentées ci-après prend en compte ces deux stratégies de recherche.

4.2 Structures d'index multidimensionnels

Une grande partie des structures multidimensionnelles a été développée pour l'indexation des bases de données en général ou des bases de données spatiales. Elles s'attachent à traiter l'accès à la mémoire secondaire. D'autres sont plus récentes et affrontent notamment les problèmes inhérents aux espaces de grande dimension. Nous ne rentrons pas dans les spécificités de ces approches qui sont trop nombreuses pour être détaillées ici.

Principe général

L'idée maîtresse des méthodes de structuration des espaces de données réside dans le regroupement des données en *paquets*. Chaque paquet regroupe un ensemble de données proches dans une zone de l'espace. Les paquets proches peuvent être assemblés pour constituer un plus gros paquet, et ainsi de suite, donnant naissance à une structure hiérarchique arborescente dans laquelle chaque nœud est associé à un paquet de données. Le principe de la recherche des plus proches voisins d'un vecteur requête consiste alors à mesurer la distance entre ce vecteur et un paquet de la hiérarchie. Si ce dernier est jugé suffisamment proche, il faut continuer la recherche dans les paquets de niveau inférieur, sinon cette partie de l'espace peut être d'emblée éliminée de la recherche. Une telle structuration permet de minimiser le nombre de vecteurs de l'espace à considérer pendant la recherche. Par conséquent, le temps CPU global, correspondant au calcul des mesures de similarité, s'en trouve notablement réduit. Si la structure et les données sont intégralement contenues en mémoire principale, alors le nombre d'entrées/sorties reste inchangé. Dans le cas contraire, les entrées/sorties correspondant à la lecture des données sur le disque (ou mémoire secondaire) peuvent être avantageusement réduites si le contenu d'un nœud à visiter est regroupé sur un minimum de pages du disque. Cette organisation physique favorise une gestion efficace des données sur disque. Elle est à l'origine des structures d'index en bases de données, la toute première étant le B-tree³.

Selon la structure d'index considérée, il existe de nombreuses stratégies pour regrouper les données en paquets. Ces derniers sont la plupart du temps répartis dans des volumes de forme géométrique élémentaire, comme des hypercubes, des hypersphères ou parfois leur intersection. S'assurer alors de la proximité du vecteur requête de telles formes n'est pas cher en temps de calcul. Les algorithmes associés diffèrent bien sûr si la requête est de type recherche des k plus proches voisins ou bien recherche à ϵ près [Böhm et al., 2001].

Les structures d'index multidimensionnels traditionnelles en bases de données se répartissent en deux catégories qui diffèrent de la façon dont les paquets de données sont construits : les techniques basées sur le *partitionnement des données* et celles basées sur le *partitionnement de l'espace*. Pour la plupart, ces structures favorisent une gestion efficace des données en mémoire secondaire et visent donc à réduire les entrées/sorties. Nous passons en revue ces deux catégories ci-dessous, avant de présenter d'autres types d'approches.

Partitionnement des données

Ces techniques consistent à regrouper les données en paquets en tenant compte de leur distribution dans l'espace. La technique la plus connue est le R-tree pour *Rectangle tree* [Guttman, 1984]. Toutes les autres techniques de cette catégorie en sont des variantes. À l'origine, le R-tree a été développé pour les bases de données spatiales. Dérivée de la classique structure arbre B, c'est une structure arborescente qui regroupe les données en rectangles. Chaque rectangle est le rectangle minimal englobant les données ; il est associé à un nœud de l'arbre et à une page disque qui en détermine la taille. Les feuilles de l'arbre contiennent les données alors que les nœuds intermédiaires stockent la hiérarchie de rectangles englobants. Le lecteur peut se reporter au chapitre *Bases de données spatio-temporelles* pour un exemple de R-tree.

Les R⁺-tree, R*-tree et X-tree sont des adaptations de cette technique aux espaces multidimensionnels [Böhm et al., 2001]. Elles s'attachent en particulier à éliminer ou à réduire le recouvrement entre les hyper-rectangles, de façon à minimiser les branches d'arbre à parcourir. Le R-tree et ses variantes ont été évalués pour l'indexation d'espaces multidimensionnels. Malgré toutes les précautions prises, elles ont montré une dégradation des performances dans les grandes dimensions. En effet, le volume du recouvrement entre hyper-rectangles augmente avec la dimension, ce

³Le B-tree, pour *Balanced tree* [Bayer and McCreight, 1972], permet de structurer les données selon l'une des dimensions. Consulter par exemple le chapitre *Architecture des SGBD* pour davantage de détails sur cette structure ainsi qu'un exemple.

qui a pour conséquence de faire tendre la recherche vers un parcours complet de l'arbre.

Les techniques SS-tree (*Similarity Search tree*) et SR-tree (*Sphere Rectangle tree*) [Böhm et al., 2001] sont assez proches de la famille R-tree. Ces techniques s'appuient sur certaines des propriétés des hypersphères pour minimiser les phénomènes liés à la malédiction de la dimension. Le SS-tree a de meilleures performances que le R*-tree mais n'atteint pas celles du X-tree. Le SR-tree peut être vu comme la combinaison du R*-tree et du SS-tree, les nœuds étant définis comme l'intersection de rectangles et de sphères. Des évaluations ont montré que le SR-tree est plus performant que le SS-tree et le R*-tree, mais nous n'avons répertorié aucune comparaison avec le X-tree.

D'autres techniques de partitionnement des données existent, nous invitons le lecteur intéressé à consulter par exemple [Böhm et al., 2001] pour un état de l'art avancé.

Partitionnement de l'espace

Les approches par partitionnement de l'espace consistent à découper *a priori* l'espace en régions, sans tenir compte de la distribution des données. Les structures d'index qui en découlent sont en général simples à mettre en place et à gérer. Les régions ainsi générées ont l'avantage de ne présenter aucun recouvrement. L'inconvénient est que l'espace entier est représenté dans ce type de structures, certaines parties vides seront donc inutilement visitées pendant la recherche.

Ces techniques sont dérivées du k-d-tree [Bentley, 1979], qui est un arbre binaire multidimensionnel consistant pour chaque niveau de l'arbre à partitionner l'espace en deux sous-espaces successivement selon chaque dimension.

Le LSD-tree [Böhm et al., 2001] pour *Local Split Decision tree* est une structure d'index simple similaire au k-d-tree mais organisée en deux niveaux : le premier niveau est entièrement stocké en mémoire principale et caractérise le partitionnement réalisé (à la façon du k-d-tree). Le deuxième niveau représente les feuilles qui contiennent les pages de données. Lorsque la base de données est trop volumineuse, les nœuds sont répartis sur deux niveaux, dont l'un est stocké sur le disque. Le LSD^h-tree [Böhm et al., 2001] est une amélioration du LSD-tree pour les données de très grande dimension, qui s'attache à minimiser le nombre de sous-espaces vides, très nombreux en grande dimension.

Citons également les algorithmes de type grille ou *grid-file* introduits dans [Nievergelt et al., 1984] qui se basent sur le même algorithme de partitionnement que le k-d-tree. Le

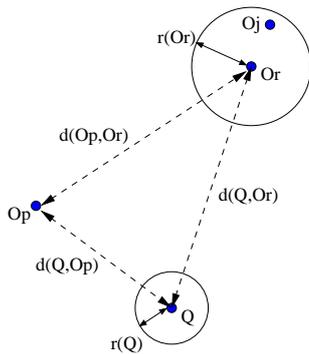
contenu de chaque partition obtenue est stocké sur une page disque et les adresses de ces partitions sont représentées en mémoire par une table de hachage dynamique.

Indexation dans les espaces métriques

Certaines données peuvent ne pas être traduites sous forme de vecteurs ou être géométriquement difficiles à interpréter. Cependant, si on peut leur associer une mesure de similarité exprimable sous forme de distance, alors il est possible d'appliquer des structures d'index dédiées aux espaces métriques. Plusieurs structures de ce type existent dans la littérature : elles définissent pour la plupart des points de l'espace spécifiques appelés *pivots* à partir desquels la propriété d'inégalité triangulaire entre distances est judicieusement exploitée, de façon à réduire le temps CPU lié au nombre de distances à calculer au moment de la recherche. Leur pertinence dépend bien sûr du rapport entre le coût d'un calcul de distance et d'une comparaison de distances. La toute première structure appliquant ce principe est le VP-tree (*Vantage Points tree*), mais on lui préférera une extension plus récente : le M-tree pour *Metric-tree* [Ciaccia et al., 1997], qui s'attache à minimiser le nombre de calculs de distances, tout en ayant une structure similaire à celle du R-tree, lui permettant de gérer les accès au disque à la façon des méthodes traditionnelles. Dans un M-tree, les données sont stockées au niveau des feuilles. Chaque nœud interne contient un ou plusieurs objets dits *routeurs*. Chaque objet routeur O_p a tous ses descendants O_r tels que $d(O_p, O_r) < r(O_p)$, d étant la distance associée aux objets. Le paramètre $r(O_p)$ représente le *rayon de couverture* de O_p , la forme de la région de l'espace effectivement couverte dépendant de la distance d utilisée. La recherche par le biais d'une telle structure est fondée sur deux lemmes exploitant l'inégalité triangulaire. Ils sont présentés à la figure 1.2 pour la recherche à ϵ près. Les auteurs du M-tree ont montré que l'utilisation de ces lemmes permet d'économiser jusqu'à 40% de calcul de distances.

Filtrage des données

Les approches procédant par filtrage des données consistent à accélérer le parcours séquentiel de l'espace en utilisant une version quantifiée des vecteurs pendant la recherche des plus proches voisins. Ces approches peuvent être qualifiées de *pessimistes* : quitte à parcourir tout l'espace, autant le faire le plus vite possible. La technique la plus connue est celle du VA-File, pour *Vector Approximation File* [Weber et al., 1998]. Ici l'espace est régulièrement subdivisé en cellules et chaque vecteur est approché par la cellule qui l'englobe.



Lemme 1 :

Si $d(Q, O_r) > r(Q) + r(O_r)$ alors on a : $d(Q, O_j) > r(Q)$.

Lemme 2 :

Si $|d(Q, O_p) - d(O_r, O_p)| > r(Q) + r(O_r)$ alors on a : $d(Q, O_r) > r(Q) + r(O_r)$.

FIG. 1.2 – Illustration de la stratégie de recherche à ϵ près dans un M-tree, par l’utilisation de l’inégalité triangulaire pour éviter certains calculs de distances. Q est l’objet requête (ici $\epsilon = r(Q)$), O_r un objet routeur de parent O_p et O_j un descendant de O_r . Le calcul de la distance $d(Q, O_j)$ peut être évité si le lemme 1 est vérifié. Celui-ci suppose la donnée de la distance $d(Q, O_r)$ dont le calcul peut également être évité si le lemme 2 est vérifié. La distance $d(O_r, O_p)$ entre un objet routeur et son parent a été calculée au moment de la construction de l’arbre.

Seules ces approximations sont chargées en mémoire principale et elles sont toutes passées en revue durant la recherche. La quantification utilisée permet de mettre en place une distance peu coûteuse entre le vecteur requête et les vecteurs quantifiés sous forme de cellules. Lors de la recherche, le parcours des vecteurs non candidats est ainsi plus rapide qu’un parcours séquentiel classique des vecteurs réels. Lorsqu’il y a ambiguïté, c’est-à-dire lorsque les approximations ne permettent pas de décider, on charge depuis le disque les données réelles afin de réaliser la comparaison. Cette technique de filtrage des données permet de réduire considérablement les entrées/sorties ainsi que le temps CPU relatif au calcul des distances qui ne sont plus réalisées que sur une très faible partie de l’espace.

Les auteurs de cette technique montrent qu’au delà de six dimensions pour une distribution uniforme des données, le VA-File atteint des performances supérieures à la plupart des structures d’index hiérarchiques telles que R*-tree et X-tree. Ceci est vrai à condition que le fichier des approximations tienne en mémoire principale. Dans le cas contraire, ses performances deviennent inférieures à un parcours séquentiel.

Plusieurs optimisations du VA-File ont été proposées depuis : le VA⁺-File, le IQ-tree et le LPC-File [Böhm et al., 2001].

Linéarisation de l’espace

Les courbes de remplissage, plus communément appelées *Space Filling Curves* sont des approches non hiérarchiques. Elles peuvent être apparentées aux techniques de réduction de la di-

mension puisqu’elles consistent à déterminer une transformation bijective entre l’espace multidimensionnel et un espace à une dimension, ou en d’autres termes à faire passer une courbe par tous les points de l’espace (d’où leur nom). La projection opérée ne préserve pas réellement les distances entre les données mais celles qui étaient proches dans l’espace original restent proches dans l’espace projeté. Ensuite, une structure d’index simple (un B-tree par exemple), peut être utilisée pour parcourir efficacement l’espace à une dimension. Plusieurs manières de positionner la courbe dans l’espace original existent, citons la technique de l’ordre Z, les courbes de Hilbert ou les *Gray codes*. Pour un panorama de ces techniques, le lecteur peut consulter [Sagan, 1994].

Un autre concept, celui du *Pyramid-tree* [Berchtold et al., 1998] n’est pas très éloigné de celui qui vient d’être présenté. L’espace multidimensionnel est partitionné selon un ensemble de *pyramides* ayant comme sommet commun le centre de l’hyper-rectangle englobant les données et comme bases les côtés de cet hyper-rectangle. Cette construction est une sorte de repère dans lequel il est assez facile de représenter chaque donnée par une seule coordonnée qui est une combinaison astucieuse du numéro de la pyramide qui la contient et de sa position dans cette pyramide. Cette technique peut donc être vue comme une projection de l’espace multidimensionnel dans un espace à une dimension.

4.3 Compromis entre efficacité et précision : la recherche approximative

Les techniques de recherche dite *approximative* sont récentes, puisque les premiers travaux

datent de 1999. Elles représentent une alternative intéressante lorsque la recherche dans des espaces de données de grande dimension est confrontée à la malédiction de la dimension. Leur principe est judicieux : il s'agit d'échanger une certaine imprécision dans les réponses contre une forte réduction du temps de recherche. Ces approches sont développées pour la recherche des k plus proches voisins, qui revient alors à chercher k vecteurs parmi les voisins plus ou moins proches du vecteur requête. Selon le cas, le degré d'imprécision dépend de la proportion de réponses *exactes* retournées ou bien de la distance entre le vecteur requête et les voisins retournés. Les solutions les plus évoluées permettent à l'utilisateur de connaître mais surtout de fixer le niveau d'imprécision du résultat, ce qui les rend exploitables. Parmi ces approches, certaines sont l'adaptation à la recherche approximative de structures d'index jusqu'alors dédiées à la recherche classique, comme le V-File et le M-tree. Des approches développées spécifiquement pour la recherche approximative ont aussi vu le jour ; elles sont pour la plupart basées sur le regroupement des données en classes. Pour plus de détails, le lecteur est invité à se reporter par exemple à [Berrani et al., 2003] qui explore la recherche approximative avec contrôle probabiliste de la précision.

5 Deux cas concrets d'utilisation

Nous présentons maintenant deux utilisations réelles d'un SRI et qui sont d'utilité publique. Elles sont issues de domaines très différents pour lesquels la recherche par contenu visuel dans les grandes collections d'images constitue une avancée notable.

5.1 Traçabilité de corpus multimédia

L'essor des nouveaux médias numériques comme les chaînes de télévision satellitaires ou bien Internet contribue largement à la dissémination de contenu multimédia. Ceci n'est pas sans susciter de nouveaux dangers pour la propriété intellectuelle et les droits de diffusion de ces données⁴. Plus généralement, faire de la veille de documents multimédia au sein des réseaux de diffusion standard (ex : statistiques de diffusion d'une publicité à la télévision, d'une photo de presse sur l'Internet) n'est plus trivial. Dans ce contexte, la problématique de la traçabilité de corpus d'images peut être vue comme la *détection de copies par analyse du contenu visuel des images*. Il se distingue du tatouage d'images qui consiste à introduire dans le signal une information cachée et dis-

criminante identifiant son propriétaire. Si les techniques de tatouage permettent d'exhiber l'antériorité du dépôt de l'image tout en caractérisant de manière irréfutable son propriétaire, elles restent à l'heure actuelle fortement sensibles aux traitements que la copie a pu subir (recadrage, échantillonnage, ajustement colorimétrique, incrustation de logos, etc). Les techniques de détection de copie s'apparentent quant à elles à celles de l'indexation et de la recherche d'images par contenu visuel. Elles possèdent les avantages d'être plus robustes aux transformations de l'image et de pouvoir s'appliquer à des documents diffusés qui n'ont pas été déposés. En revanche, elles ne peuvent pas fournir de preuve légale de copie, ce qui implique que dans le cas particulier de la recherche de copies illicites, les images semblables retrouvées restent soumises à une expertise humaine.

De telles techniques de détection de copies sont depuis peu utilisées par l'Institut national de l'audiovisuel. L'INA, qui a pour mission la conservation et l'exploitation du patrimoine audiovisuel français, est actuellement détenteur d'un fonds télévisé particulièrement conséquent dont 200 000 heures sont déjà numérisées. À partir de descripteurs locaux d'images tels que ceux décrits à la section 2.2 couplés avec une structure d'index adéquate, l'INA a ainsi développé un dispositif permettant de détecter et de comptabiliser automatiquement les éventuelles retransmissions de ce fonds sur les grandes chaînes de télévision française, avec comme principal objectif d'en contrôler les droits de diffusion.

5.2 Lutte contre le crime

La Division nationale pour la répression des atteintes aux personnes et aux biens (DNRAPB) de la Direction centrale de la police judiciaire (DCPJ) a mis en place un service de lutte contre la pornographie infantile. Ce service collecte de nombreuses images par le biais d'Internet et de perquisitions effectuées chez des particuliers suspectés de pédophilie. À l'heure actuelle, le nombre d'images accumulées (environ 300000 clichés en constante augmentation) ne leur permet plus de croiser manuellement les informations visuelles pertinentes. Les techniques de recherche automatique par contenu ont été mises à profit et sont utilisées depuis 2000 comme outils d'aide à l'investigation : au fur et à mesure de leur saisie, les clichés sont indexés et classés en fonction de divers critères visuels tel que le vêtement de l'enfant impliqué ou le détail d'un décor. Cette structuration permet de regrouper automatiquement les

⁴L'adoption en 2004 du projet de loi pour la confiance dans l'économie numérique montre l'intérêt que porte le législateur aux dangers engendrés par la dissémination des contenus numériques.

images relatives à une scène donnée et ainsi de faciliter l'identification de l'auteur de ces clichés ou de l'adulte impliqué.

Les bases de données ainsi constituées sont désormais mises en commun entre la police et la gendarmerie française. Plus encore, la politique menée actuellement par la Commission européenne vise une coopération européenne et même internationale des polices. Un tel déploiement rend les techniques d'indexation par contenu visuel prépondérantes pour la lutte contre le crime.

6 Conclusions et perspectives

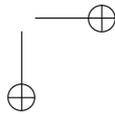
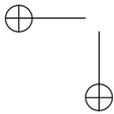
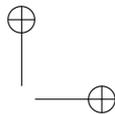
Dans ce chapitre, nous avons présenté la problématique de la recherche par contenu visuel dans les grands volumes d'images. Abordée du point de vue des bases de données, le principal obstacle scientifique à surmonter reste le *passage à l'échelle* des techniques d'optimisation physique. Plusieurs solutions ont été proposées dans ce sens récemment. Néanmoins, le volume d'information grandissant de manière spectaculaire, il conviendra certainement assez rapidement d'envisager d'autres solutions. Dans un certain nombre de cas, la variabilité des images contenues dans des bases volumineuses doit permettre de répartir ces images en sous-bases ou catégories pour lesquelles

les techniques d'indexation et de recherche actuelles restent applicables. Dans les autres cas, la seule solution passera certainement par la conception de systèmes de recherche exploitant des bases de données réparties sur plusieurs machines. C'est déjà ce que fait le moteur de recherche Google pour la recherche par le texte, avec plus de 100000 processeurs et 250000 disques (données 2004).

Si l'on se place ensuite du point de vue de l'analyse d'images, le principal défi à résoudre est le problème du *fossé sémantique*. Résumer le contenu visuel d'une image par l'analyse automatique de ses pixels ne peut pas se mesurer à ce que l'humain voit, ou plutôt perçoit lorsqu'il regarde cette image. Pour s'en rapprocher, l'analyse des images doit être combinée avec d'autres modes, le texte étant sémantiquement le plus riche. Au moment de l'indexation de la base d'images, l'analyse du contenu visuel peut permettre de propager automatiquement des mots-clés d'images annotées vers des images visuellement similaires mais non annotées. Au moment de la recherche, l'interrogation pluri-modale doit pouvoir permettre de réduire les ambiguïtés ou lacunes du texte par l'apparence visuelle, comme de combler l'insuffisance sémantique des descripteurs image.

Index

- B-tree, 8
- bases
 - d’images
 - indexation, 1–12
 - génériques, 2
 - spécifiques, 3
- descripteur d’image, 2–7
 - descripteur global, 3
 - descripteur local, 4
 - descripteur spécifique, 5
 - espaces de description, 5
 - signature, 2
- dimension
 - malédiction de la, 6–7
 - réduction de la, 5–6
- grid-file, 9
- image
 - indexation, 1–12
 - recherche par contenu visuel, 1–12
- index
 - multidimensionnels, 7–11
 - principe général, 8
 - partitionnement des données, 8
 - partitionnement de l’espace, 9
 - espaces métriques, 9
 - filtrage des données, 9
 - linéarisation de l’espace, 10
- indexation
 - d’images, 1–12
 - multidimensionnelle, 7–11
- k-d-tree, 9
- LSD-tree, 9
- M-tree, 9
- malédiction de la dimension, 5–6
 - phénomène de concentration des mesures, 6
 - phénomène de l’espace vide, 5
- points d’intérêt, 4
- Pyramid-tree, 10
- R-tree, 8
- réduction de la dimension, 6–7
 - Analyse en Composantes Principales, 6
 - FastMap, 7
- recherche
 - d’images, 1
 - globale (approximative), 3
 - locale (précise), 3
 - par contenu visuel, 1
 - par l’exemple, 2
 - types de requêtes, 7
 - recherche à ϵ près, 7
 - recherche approximative, 10
 - recherche des k plus proches voisins, 7
 - recherche par intervalle, 7
- segmentation en régions, 4
- signature d’image, 2–7
- Space Filling Curves, 10
- SR-tree, 9
- SRI, voir système de recherche d’images
- SS-tree, 9
- structuration, 2
- structures d’index, 8–11
- système de recherche d’images, 2
- VA-File, 9
- voisin
 - recherche des plus proches, 7
- X-tree, 8



Bibliographie

- [Bayer and McCreight, 1972] Bayer, R. and McCreight, E. (1972). Organization and maintenance of large ordered indices. *Acta Informatica*, 1(3) :173–189.
- [Belhumeur et al., 1997] Belhumeur, P., Hespanha, J., and Kriegman, D. (1997). Eigenfaces vs. fisherfaces : Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :711–720.
- [Bentley, 1979] Bentley, J. (1979). Multidimensional binary search in database applications. *IEEE Transactions on Software Engineering*, 4(5) :333–340.
- [Berchtold et al., 1998] Berchtold, S., Böhm, C., and Kriegel, H. (1998). The pyramid-technique : towards indexing beyond the curse of dimensionality. In *ACM SIGMOD Int. Conf. on Management of Data*, pages 142–153, Seattle, WA.
- [Berrani et al., 2003] Berrani, S.-A., Amsaleg, L., and Gros, P. (2003). Approximate searches : k-neighbors + precision. In *Conference on Information and Knowledge Management*, pages 24–31.
- [Beyer et al., 1999] Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful? In *7th International Conference on Database Theory*, pages 217–235, Jerusalem, Israel.
- [Böhm et al., 2001] Böhm, C., Berchtold, S., and Keim, D. (2001). Searching in high-dimensional spaces - index structures for improving the performance of multimedia databases. *ACM Computing Survey*, 33(3) :322–373.
- [Boujemaa et al., 2004] Boujemaa, N., Fauqueur, J., and Gouet, V. (2004). What's beyond query by example? In L. Shapiro, H.P. Kriegel, R. V., editor, *Trends and Advances in Content-Based Image and Video Retrieval*. LNCS, Springer Verlag.
- [Ciaccia et al., 1997] Ciaccia, P., Patella, M., and Zezula, P. (1997). M-tree : An efficient access method for similarity search in metric spaces. In *23th Int. Conf. on Very Large Databases*, pages 426–435, Greece.
- [Cox and Cox, 2001] Cox, T. and Cox, M. (2001). *Multidimensional Scaling, Second edition*. Chapman & Hall, London.
- [Demartines and Héroult, 1997] Demartines, P. and Héroult, J. (1997). Curvilinear component analysis : A self-organizing neural network for non linear mapping of data sets. *IEEE Transactions on Neural Networks*, 8(1) :148–154.
- [Faloutsos and Lin, 1995] Faloutsos, C. and Lin, K. (1995). Fastmap : A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *ACM SIGMOD International Conference on Management of Data*, pages 163–174, San Jose, CA, USA.
- [Gerbrands, 1981] Gerbrands, J. (1981). On the relationship between SVD, KLT and PCA. *Pattern Recognition*, 14(1-6) :375–381.
- [Guttman, 1984] Guttman, A. (1984). R-trees : a dynamic index structure for spatial searching. In *ACM SIGMOD Int. Conf. on Management of Data*, pages 47–57, Boston, MA.
- [Nievergelt et al., 1984] Nievergelt, J., Hinterberger, H., and Sevcik, K. C. (1984). The Grid File : An adaptable, symmetric multikey file structure. *ACM Transactions on Database Systems*, 9(1) :38–71.
- [Rigaux et al., 2001] Rigaux, P., Scholl, M., and Voisard, A. (2001). *Spatial Databases*. Morgan Kaufmann.
- [Sagan, 1994] Sagan, H. (1994). *Space Filling Curves*. Springer, Berlin/Heidelberg/New york.
- [Swain and Ballard, 1991] Swain, M. and Ballard, D. (1991). Color indexing. *International Journal of Computer Vision*, 7(1) :11–32.
- [Verleysen, 2003] Verleysen, M. (2003). *Limitations and future trends in neural computation*, chapter Learning high-dimensional data, pages 141–162. IOS Press.
- [Weber et al., 1998] Weber, R., Schek, H., and Blott, S. (1998). A quantitative analysis and performance study for similarity search methods in high-dimensional spaces. In *24th International Conference on Very Large Databases*, New york, NY, USA.