



Faculté des Sciences de L'Ingéniorat Année 2016-2017

Département d'Informatique

THESE

Présentée en vue de l'obtention
du diplôme de Doctorat

Fusion d'approches pour l'indexation d'images

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : TIVA (Traitement d'Images et Vision Artificielle)

par

Mr Ahmed BOULEMDEN

Directeur de Thèse

Mme Yamina TLILI

Professeur Université Badji Mokhtar-Annaba

DEVANT LE JURY

Président :

Mme Labiba SOUICI-MESLATI

Professeur Université Badji Mokhtar-Annaba

Examineurs :

Mr Rafik DJEMILI

Professeur Université 20 Août 1955 - Skikda

Mr Mohamed Taher KIMOUR

Professeur Université Badji Mokhtar-Annaba

Mr Imed BOUCHRIKA

MCA Université Mohamed-Chérif Messaadia –
Souk Ahras

إلى جدتي رحمها الله

إلى أبي و أمي حفظهما الله

إلى الأستاذة الكريمة يمينة تليلي

إلى من في قلبي

شكر

أحمد الله سبحانه و تعالى على أن وفقني لإجراز هذا العمل، كما و أدعوه أن يوفقني لأواصل في طريق العلم و جميع زملائي.

ثم أثنى بالشكر للأستاذة الدكتورة تليلى يمينة التي أولتني كامل رعايتها و توجيهها خلال مسيرة إشرافها على أبحاث تخرجي في الليسانس، الماستر و ختاماً برسالة الدكتوراه هذه.

أقدم خالص امتناني و شكري لجدتي رحمها الله و لأبي وأمي حفظهما الله على جهدهم و تعبهم خلال كامل حياتي و مسيرتي الدراسية كما ربوني صغيراً. كما و أقدم جزيل شكري لزوجتي.

لا أنسى أن أتقدم أيضاً بخالص الشكر لكافة أفراد أسرتي و عائلتي ممن ساعدوني و أعانوني خاصة بالذكر جميع أخوائي.

أتوجه كذلك بجزيل الشكر لأصدقائي و زملائي كافة و أخص منهم أخي و صديقي عمر.

أشكر أيضاً أعضاء شركة سيتروس إنوفيشن و جميع عمال قسم الإعلام الآلي بجامعة باجي مختار عناية.

أختم بالشكر و الدعاء لكامل أساتذتي خلال جميع المراحل الدراسية و الذين مهما دعوت لهم فلن أوفيهم حقهم فيما أفادوني به من علم و تكوين و أرجو من الله سبحانه و تعالى أن يوفقني إلى العمل بالعلم الذي لقتوني إياه و أن يجزل لهم الأجر.

Remerciements

Je remercie en premier lieu ALLAH qui m'a donné à la fois le courage, la volonté, et la patience afin d'élaborer cette thèse de recherche scientifique.

Mes premiers remerciements vont à Pr TLILI Yamina, la directrice de ce travail. Je suis heureux d'avoir eu la chance durant toutes ces années de partager ses réflexions, de disposer de son expertise scientifique et de son expérience pour mener à bien mes travaux.

Je souhaite associer à ces remerciements profonds ma famille surtout ma grande mère, mes chers parents, ma femme et mes sœurs.

Je remercie mes enseignants, mes amis et surtout Amor, mes collègues et tous ceux que je connais de près ou de loin.

Je tiens aussi à remercier mes amis les membres de la société Citrus Innovation.

Je tiens aussi à remercier le personnel du département de l'informatique de l'université Badji Mokhtar Annaba.

ملخص

نماذج المواضيع الإحصائية هي مجموعة من الخوارزميات التي تهتم باستخراج المفاهيم المخفية و التي تكون مسؤولة عن إنتاج الكلمات ضمن مستندات قواعد البيانات الكبرى. امتد استخدام هاته النماذج إلى معالجة الصور الرقمية ليغطي التعرف على الأشياء، التعليق النصي على الصور و فهرسة الصور و استرجاعها.

نقدم من خلال هذا العمل نظاما لاسترجاع الصور استنادا إلى المحتوى. يركز هذا النظام في عمله على نماذج المواضيع الإحصائية و يطبق هذا النموذج على تركيب من خصائص الألوان و النسخ إضافة إلى خصائص محلية و نصية.

تعرض المذكرة لتقييم نموذجين من نماذج المواضيع في المجال المذكور و ذلك عن طريق تطوير نظام لفهرسة الصور مبني على أساس استخدام هاته النماذج مع مجموعة من الخصائص المرئية العامة و المحلية المستخرجة من الصور، إلى جانب خصائص نصية مصدرها النصوص المرفقة. تمت دراسة استخدام هاته الخصائص مع النماذج بشكل مستقل ثم إدماجها معا.

تم تطبيق التجارب على صور مستخرجة من قاعدة الصور IAPR TC 12 بالإضافة لقاعدة الصور الطبية MIAS.

الكلمات الدلالية

نموذج توزيع باشينكو، التحليل المعنوي الخفي، فهرسة الصور و استرجاعها، عزوم الألوان، خصائص الأنسجة، الخصائص المحلية و العامة، دمج الخصائص، دمج المقاربات.

Résumé

Les modèles de sujets probabilistes constituent un ensemble d'algorithmes qui permettent d'extraire les concepts cachés responsable de la génération des mots des documents de grands corpus. Ces modèles ont été utilisés dans le domaine de traitement d'images en les appliquant sur des problèmes tels que la reconnaissance d'objets, l'annotation automatique des images et l'indexation et la recherche d'images par le contenu.

Nous présentons dans le cadre de cette thèse un système de recherche d'images par le contenu (CBIRs). Le système est basé sur l'utilisation d'un modèle de sujet probabiliste qui est le Pachinko Allocation Model (PAM) avec une combinaison de caractéristiques visuelles et textuelles.

PAM permet de capturer la corrélation entre les mots et aussi entre les sujets eux même, ce qui présente plus de niveaux d'abstraction pour la modélisation. L'étude du modèle PAM pour l'indexation et la recherche d'images par le contenu est absente de la communauté de recherche en traitement d'images.

Nous visons via ce travail à évaluer l'utilité du modèle PAM pour les systèmes CBIRs. Nous avons réalisé cette évaluation en utilisant plusieurs modalités de caractéristiques seules et combinées. Dans ce contexte, PAM a été appliqué avec des caractéristiques visuelles globales (couleur et texture), locales (SIFT) et textuelles (à partir du texte associé à l'image) séparées et combinées.

Une deuxième contribution de la thèse vise à étudier le modèle LSA (Latent Semantic Analysis) pour l'indexation et la recherche des tumeurs mammaires.

Mots clés

Allocation de Pachinko, analyse de la sémantique latente, Indexation et recherche d'images, moments de couleur, caractéristiques de texture, modalité textuelles, caractéristiques locales, combinaison des caractéristiques.

Abstract:

Probabilistic topic models are a set of algorithms which aim to learn and discover hidden concepts responsible of generating words of documents in large archives. These models have been also used for image processing tasks such as object recognition, image annotation and image retrieval.

We present in this work a content based image retrieval system (CBIR) based on pachinko allocation model (PAM) and employing a combination of color, texture, local features and textual features.

PAM has presented more efficiency compared with other topic models by the way in which it captures correlation not only between words in documents but also between different topics (concepts) responsible of their generation. Although this advantage of PAM, there is no works which explore its utility for content based image retrieval tasks and using single and multimodal image features.

We aim to evaluate the use of PAM for CBIRs by implementing a system based on it. We are interested also in evaluating PAM with single and multimodal image features (i.e. the use of single and combined features). In this context, PAM was applied with different modalities of features, image global features (color and texture), local features extracting with SIFT algorithm and textual indexes (from associate texts with images). These features have been used separately and combined.

Mean Average Precision is evaluated. The use of PAM with combination of features has significantly improved results of using it with just one modality; this opens more perspective in order to enhance results. Images from the ImageCLEF IAPR 2012 dataset have been used for experiments.

As a secondary contribution, we have studied the use of the Latent Semantic Analysis model for medical image indexing and retrieval. The contribution uses especially mammographic images.

Keywords

Keywords: Pachinko allocation, image retrieval, color moments, texture features, global feature extraction, textual modality, features combination, local features.

Table des Matières

Remerciements	iv
ملخص.....	v
Résumé	v
Abstract:.....	vi
Table des Matières	7
Liste des Tableaux.....	10
Table des Illustrations.....	11
Liste des Figures.....	11
Liste des Acronymes	14
CHAPITRE – I.....	15
INTRODUCTION GENERALE	15
1.1 Introduction	16
1.2 Systèmes de recherche d'images par le contenu CBIRs	18
1.3 Objectifs et approches.....	20
1.4 Organisation du document	21
CHAPITRE – II.....	22
INDEXATION ET RECHERCHE D'IMAGES	22
CONCEPTS DE BASE.....	22
2.1 Introduction	23
2.2 Bases d'images	25
2.2.1 Bases d'images standards (Benchmark):	27
2.3 Extraction des caractéristiques des images.....	29
2.3.1 Caractéristiques textuelles	29
2.3.2 Caractéristiques visuelles	31
2.4 Localisation de régions d'intérêt.....	32
2.5 Segmentation des images	32
2.6 Extraction de caractéristiques locales	34
2.7 Extraction de caractéristiques globales.....	37
2.8 Conclusion.....	50
CHAPITRE – III.....	51
INDEXATION ET RECHERCHE D'IMAGES	51
CONCEPTS AVANCES.....	51
3.1 Introduction	52
3.2 Construction de la signature	52

3.3	Mesure de similarité.....	54
3.4	Indexation multimodale.....	55
3.5	Recherche d'images.....	56
3.6	Visualisation des résultats	58
3.7	Bouclage de pertinence	60
3.7.1	Contributions basée sur l'apprentissage automatique	60
3.7.2	Contributions basée sur les spécifications du BP	61
3.7.3	Contribution basées sur l'utilisateur.....	61
3.7.4	Contributions probabilistes.....	61
3.7.5	Contributions basées régions	61
3.8	Réduction du fossé sémantique	62
3.8.1	Les ontologies d'objets.....	62
3.8.2	L'apprentissage automatique.....	64
3.8.3	Le bouclage de pertinence	64
3.8.4	Les modèles sémantiques.....	64
3.8.5	Recherche d'image dans le web	64
3.9	Critères d'évaluation	65
3.10	Conclusion.....	66
CHAPITRE – IV		67
MODELES DE SUJETS		67
5.1.	Introduction	68
5.2.	Modèles à Concepts déterministes	69
5.2.1	Latent Semantic Analysis (LSA).....	69
5.3	Les modèles à concepts probabilistes	71
5.3.1	Probabilistic Latent Semantic Analysis.....	72
5.3.2	Latent Dirichlet Allocation.....	73
5.3.3	Correlated Topics Model (CTM).....	74
5.3.4	Pachinko Allocation Model	75
5.4	Extraction des concepts	76
5.4.1	SVD.....	76
5.4.2	Expectation Maximization	76
5.4.3	Gibbs Sampling	78
5.5	Application des modèles à concepts en traitement d'images	80
5.6	Conclusion.....	81
CHAPITRE – V.....		82
CONTRIBUTIONS.....		82
6.1	Introduction	83
6.2	Bases d'images utilisées dans les expérimentations.....	85
6.3	Modalités des caractéristiques	86
6.4	Préparation des données.....	87
6.4.1	L'extraction des caractéristiques textuelles.....	88
6.4.2	Caractéristiques visuelles locales.....	88
6.4.3	Caractéristiques visuelles globales	89
6.4.4	Construction des matrices de cooccurrences.....	90
6.5	Application du modèle PAM pour l'indexation	92
6.5.1	Résultats et perspectives.....	94

6.5.2	Les expérimentations de la thèse.	94
6.5.3	Les bases d'image utilisée.....	95
6.5.4	Extraction des caractéristiques.....	95
6.5.5	Application du modèle PAM.....	96
6.5.6	Evaluation et Résultat.....	96
6.5.6.1	Discussion détaillée des résultats.....	98
6.6	Indexation et reconnaissance des tumeurs mammaires par le modèle LSA	104
6.7	Conclusion.....	110
CONCLUSION GENERALE		112
ET PERSPECTIVES		112
7.1	Conclusion.....	113
7.2	Perspectives.....	114
Références		116

Liste des Tableaux

Tableau 1. Exemple de valeurs de RGB avec les couleurs correspondantes	39
Tableau 2 poids d'importance associés aux différentes balises du fichier textuel	88
Tableau 3 Les classes des images utilisées dans les expérimentations	95
Tableau 4 Résultats de PAM avec les caractéristiques globales, textes et fusion	96
Tableau 5 Résultats de PAM avec les caractéristiques locales, textuelles et fusion	96
Tableau 6 Résultats de PAM avec Combinaison des caractéristiques locales et globales	97
Tableau 7 Evaluation visuelles de trois critères visuels dans l'image	100
Tableau 8 Comparaison des résultats avec des travaux similaires	101
Tableau 9 Résultats de l'évaluation du MAP de LSA pour l'indexation	108

Table des Illustrations

Liste des Figures

Figure 1. Une étude concernant les publications dans le domaine de la recherche d'images par le contenu pour la période 1995-2005. A gauche, les publications contenant les phrases 'image retrieval' et 'support vector'. A droite, les publications contenant la phrase 'image retrieval' pour trois grands publisher (Datta et al.,2008).....	16
Figure 2. Graphe qui représente le nombre de publications contenant les mots "Image retrieval" pour la période 2005-2016. Les données proviennent de Google Scholar.....	17
Figure 3. Architecture générale d'un système de recherche d'images par le contenu	18
Figure 4. Les différents axes de recherche dans le domaine de l'indexation et de la recherche d'images (Datta R. et al., 2008)	24
Figure 5. Le processus d'indexation et de recherche d'images par le contenu visuel	25
Figure 6. Des Images issues de la base Corel 5K (Web6, 2016)	27
Figure 7. Des images issues de la base IAPR TC 12	28
Figure 8. Des images issues de la base UW data-set	28
Figure 9. Exemple d'une image de web avec le texte associé (Web 7, 2009)	30
Figure 10. Etapes de la construction de la signature visuelle.....	31
Figure 11. Résultats de la segmentation avec l'algorithme de segmentation JSEG pour deux images (Liu et al., 2007)	33
Figure 12. Exemple de régions d'intérêts choisies pour l'extraction Des caractéristiques locales (Tollari S., 2006).....	35
Figure 13. Les phases d'extraction des caractéristiques locales	35
Figure 14. Détection des extrêmes par comparaison du pixel d'intérêt avec ses voisins du niveau courant ainsi que les niveaux adjacents (Lowe, 2004).....	36
Figure 15. 2x2 vecteurs descripteurs calculés à partir d'un échantillon de 8x8 (Lowe, 2004)	36
Figure 16. Une image et les points de caractéristiques locales extraites (Lowe D.G., 2004)	37

Figure 17. La perception de l'œil humain (Web 4, 2016).....	38
Figure 18. Cube des couleurs RGB (Manzanera A., 2008).....	40
Figure 19. Dimensions de l'espace RGB (Web5, 2016).....	40
Figure 20. Cercle de la teinte (Web5)	41
Figure 21. La composante teinte d'une image couleur (Manzanera, 2008).....	41
Figure 22. Variation de la saturation (Web5).....	42
Figure 23. La composante de saturation d'une image couleur (Manzanera, 2008) ..	42
Figure 24. Variation de la luminance (Web5).....	42
Figure 25. La composante de luminance d'une image couleur. (Manzanera, 2008)	43
Figure 26. Représentation de l'espace HSV (Web5)	43
Figure 27. Histogrammes de couleur RGB (Meskaldji et al., 2009).....	44
Figure 28. Exemple de textures ()	45
Figure 29. Matrice de cooccurrence calculée pour 4*4 pixels pour la direction 0°..	46
Figure 30. Classification des techniques d'extraction des caractéristiques de la forme (Yang et al., 2008).....	49
Figure 31. Les différentes formulations de la signature visuelle (Datta et al., 2008)	53
Figure 32. Classification des modalités de caractéristiques des images. (a) représente un exemple de modalité textuelles et (b) deux exemples de modalités visuelles.....	56
Figure 33. Exemples de formes de requêtes.....	57
Figure 34. Exemple de visualisation des résultats de la recherche du mot 'Tiger' par le moteur de recherche Google (google image, Juin 2014)	59
Figure 35. Résumé des approches de la réduction du gap sémantique (Liu et al., 2007)	62
Figure 36. Ontologie d'objets (Ying et al., 2007)	63
Figure 37. Matrice termes-documents, chaque élément de la matrice spécifie la fréquence d'apparition du mot de la ligne dans le document de la colonne (Steyvers and Griffiths, 2007).....	68
Figure 38. Exemple de quatre concepts latent extraits à partir du corpus TASA (Steyvers and Griffiths, 2007).....	70
Figure 39. Factorisation de la matrice pour le modèle LSA	71
Figure 40. La représentation graphique du modèle pLSA (Hörster, 2009).....	73
Figure 41. Modèle graphique du LDA (Hörster, 2009)	73
Figure 42. La représentation graphique du modèle CTM (Hörster, 2009)	74

Figure 43. La représentation graphique du modèle PAM (Li and McCallum, 2006)	75
Figure 44. Algorithme EM (Urli, 2007).....	78
Figure 45. Architecture de l'étude de la première contribution de la thèse.....	85
Figure 46 Un exemple d'une image de la base IAPR TC 12 et le fichier contenant sa description textuelle	86
Figure 47. Exemple des images issues de la base IAPR TC 12	87
Figure 48. Une image de la base IAPR-TC 12 et les points de caractéristiques locales extraites en utilisant le package de (Web7)	89
Figure 49. Un exemple d'une image divisée en 16 régions. Chaque région est étiquetée par numéro et elle est considérée comme un mot visuel lors de la construction des mots visuels.....	89
Figure 50. Un fragment de texte généré à la base des caractéristiques visuelles locales. Les mots sont des mots visuels	91
Figure 51. Un fragment de texte généré à la base des caractéristiques visuelles globales. Les mots sont des mots visuels	91
Figure 52. Structure de PAM à 4 niveaux utilisé dans la thèse (Li W and McCallum, 2006)).....	93
Figure 53. Exemple des deux distributions inférées pour une image de la base, à gauche les sous-sujets et à droite les super-sujets (cas des mots visuels global)	94
Figure 54. 5 premiers résultats pour une image requête. Cas de fusion locales-texte	97
Figure 55 Exemple des images utilisées pour la contribution.....	106
Figure 56 Architecture de l'étude de LSA pour l'indexation des tumeurs mammaires	106
Figure 57 graphes des résultats de MAP par catégories de requêtes, à gauche les requêtes malignes et à droite les requêtes bénignes	109
Figure 58 résultats de l'évaluation de la reconnaissance de type de tumeur	110
Figure 59 Distribution gaussienne (normale) d'écart-type 1 et de moyenne 0.....	136
Figure 60. Courbes représentant l'une des dimensions d'une distribution de Dirichlet à deux dimensions ($m = 2$) pour différentes valeurs de u ($u = \alpha_1 = \alpha_2$).137	
Figure 61. Les attributs de la classe PAM4L	140
Figure 62. Les deux constructeurs de la classe PAM4L	140
Figure 63. Une partie du code de la méthode d'estimation dans la class PAM4L..	141
Figure 64. Exemple d'utilisation de la classe PAM4L.....	142

Liste des Acronymes

CBIR	Content Based Image Retrieval
CTM	Correlated Topic Model
EM	Expectation Maximization
LDA	Latent Dirichlet Allocation
LSA	Latent Semantic Analysis
pLSA	Probabilistic Latent Semantic Analysis
PAM	Pachinko Allocation Model

CHAPITRE – I

INTRODUCTION GENERALE

1.1. Introduction:

Le développement rapide dans le domaine des caméras numériques ainsi que les outils d'acquisition, de transfert et de stockage d'images pose un défi majeur pour les utilisateurs. Ce défi se résume dans la nécessité d'accéder et d'utiliser efficacement l'importante quantité d'images disponibles.

La réalisation d'outils permettant l'utilisation optimale de bases d'images et la manipulation satisfaisante par les utilisateurs est assurée à travers les différents travaux dans le domaine de l'indexation et de la recherche d'images. Ce domaine constitue l'une des applications de techniques du traitement d'images et de plusieurs domaines de la technologie de l'information.

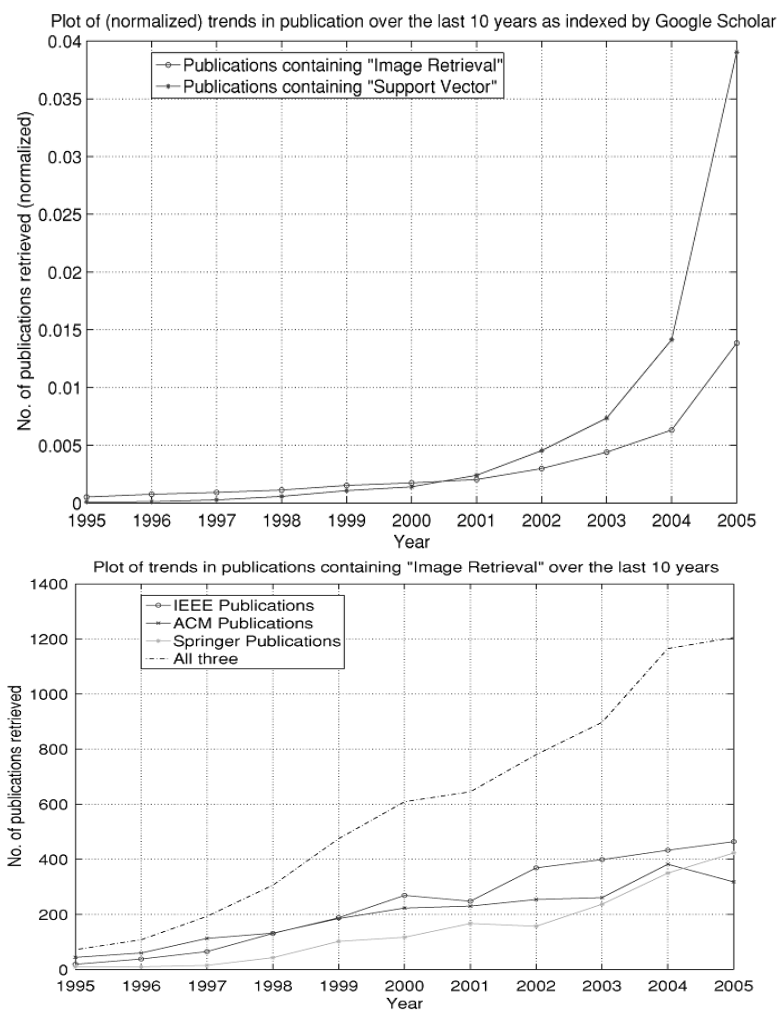


Figure 1. Une étude concernant les publications dans le domaine de la recherche d'images par le contenu pour la période 1995-2005. A gauche, les publications contenant les phrases 'image retrieval' et 'support vector'. A droite, les publications contenant la phrase 'image retrieval' pour trois grands publisher (Datta et al.,2008)

La recherche en indexation des images a connu une grande émergence grâce aux différentes techniques et approches traitées et proposées à partir des années quatre vingt. La figure 1 montre des courbes présentées par (Datta et al., 2008) qui illustrent la croissance des publications du sujet de la recherche d'images (Image Retrieval) entre 1995-2005. Tandis que la figure 2 présente la variation annuelle de nombre de publication contenant le terme "Image Retrieval" et cela entre 2005-2016.

Le domaine de recherche en indexation d'images reste toujours ouvert pour d'autres travaux et propositions. La satisfaction des besoins des utilisateurs, la diversité de leurs intentions ainsi que l'hétérogénéité et la dynamique des bases d'images constituent des défis majeurs pour l'amélioration de la performance de la recherche.

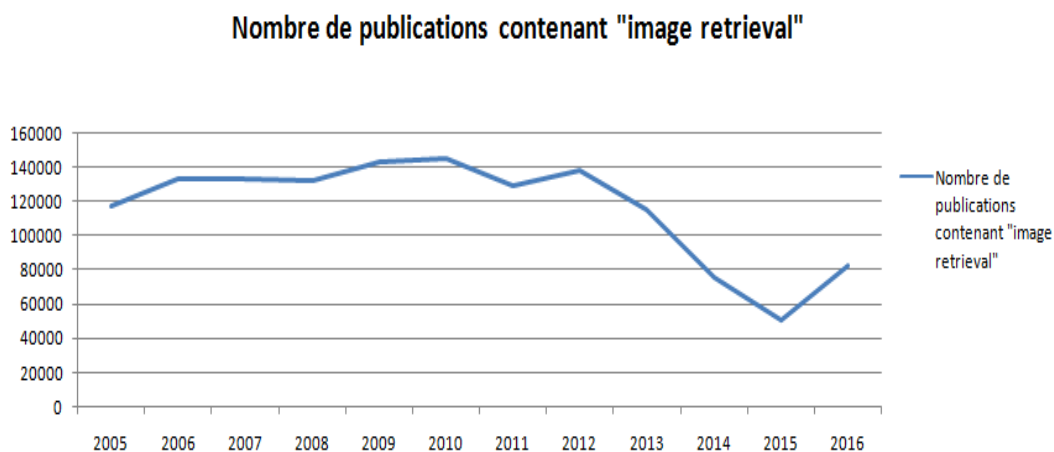


Figure 2. Graphe qui représente le nombre de publications contenant les mots "Image retrieval" pour la période 2005-2016. Les données proviennent de Google Scholar (Octobre 2016).

Les Systèmes de recherche d'images à base de contenu (CBIRS pour Content Based Image Retrieval Systems) sont au centre de ce domaine de recherche. Ces systèmes représentent l'outil efficace pour les utilisateurs utilisé pour effectuer leurs recherches. Le centre d'intérêt du domaine de la recherche de l'indexation se résume dans la performance de ces systèmes, donc l'objectif est de développer des systèmes performants pour permettre la satisfaction efficace et efficiente des besoins des différents utilisateurs.

Dans ce cadre, la motivation principale pour ce travail est de contribuer à l'amélioration de la performance des systèmes (CBIRS). Cette amélioration nécessite

la concentration sur les problèmes majeurs de ces systèmes résumés dans les points suivants:

- La pertinence des caractéristiques extraites pour la représentation des images. Ce problème nous a motivé à étudier l'utilisation d'un ensemble de caractéristiques de différents types et modalités.
- La performance des techniques et modèles appliqués pour l'indexation des images. Dans ce contexte, nous avons choisis une famille de modèles statistiques connue par sa performance dans la représentation des données.

Nous allons présenter dans la section suivante les systèmes de recherche d'images à base de contenu (CBIRs).

1.2. Systèmes de recherche d'images par le contenu CBIRs:

Le but de la recherche dans le domaine de l'indexation des images est d'améliorer la performance (et plus particulièrement la précision) du système de recherche d'images par le contenu CBIRs.

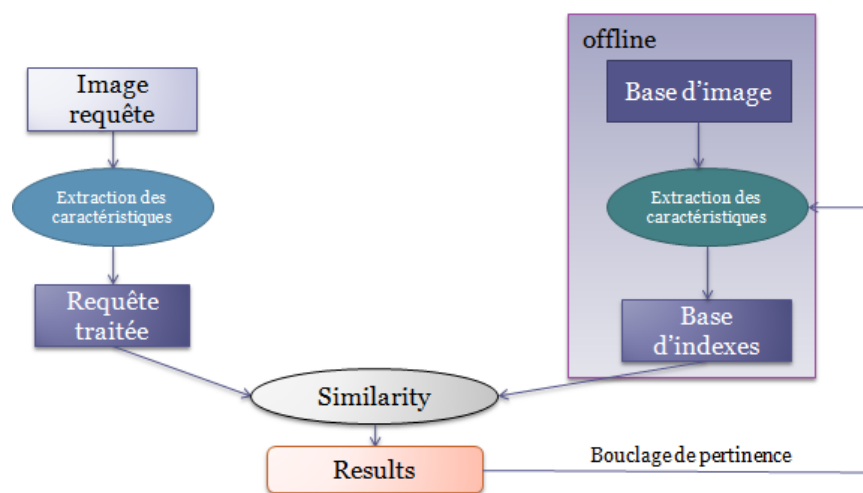


Figure 3. Architecture générale d'un système de recherche d'images par le contenu

Les systèmes de recherche d'images représentent un cas spécial des systèmes de recherche d'information. Les images sont l'information que l'utilisateur cherche à récupérer de la base en utilisant ces systèmes. L'architecture de base de CBIRs est illustrée par la figure 3.

La phase hors ligne est une phase d'indexation où les images trouvées dans la base sont indexées en utilisant un ensemble de caractéristiques généralement extraites à partir de ces images. Cette représentation du contenu de l'image en utilisant l'ensemble de caractéristiques extraites permet de réduire la dimension utilisée durant la recherche. La pertinence de ces caractéristiques est primordiale pour assurer une bonne représentation du contenu des images de la base.

La phase de recherche des images est une phase en ligne. La requête fournie par l'utilisateur doit être traitée afin et transformée en une représentation adéquate semblable à celle des caractéristiques des images dans la base.

Le système effectue un calcul de similarité entre l'index de la requête de l'utilisateur et les images indexées dans la base pour répondre à l'utilisateur avec la liste des images les plus similaires (pertinentes) à sa requête dans la base.

Eventuellement, une phase de bouclage de pertinence peut être utilisée dans le système. Cette phase est le résultat de l'interaction des utilisateurs à travers plusieurs sessions de recherche. L'utilisateur participe à l'évaluation des résultats retournés par le système en jugeant leur pertinence vis-à-vis de sa requête. Cette participation de la part d'utilisateurs aide le système à ajuster ces paramètres internes utilisés dans l'indexation.

La recherche dans le domaine de l'indexation a passé par trois périodes. Durant la première qui précède les années 90s, l'indexation a été effectuée manuellement par des opérateurs humains. Ces derniers indexent les images des bases avec des mots clés, et la recherche est effectuée par l'utilisateur dans le même mode textuelle. Durant la période des années 90s, le domaine de recherche s'est orienté vers l'utilisation des caractéristiques visuelles des images pour pouvoir automatiser le processus d'indexation et donc indexer les grandes bases d'image qui commencent à apparaître largement durant cette période. Il faut noter que l'utilisation de ces caractéristiques visuelles a été proposée pour la première fois durant les années 80.

L'utilisation du mode visuel de caractéristiques a conduit à la définition du problème majeur du domaine depuis ce temps, qui est le problème du gap sémantique déclaré par (Smeulders et al., 2000). Le gap sémantique a été définie en tant que *"un manque de relation (coïncidence) entre l'information visuelle numérique qu'on peut extraire*

à partir des images, et l'interprétation sémantique de l'image par un utilisateur dans un contexte donné". Ce problème concerne le manque de relation entre ce qui est numérique comme caractéristiques de l'image et l'interprétation sémantique souvent cherchée par l'utilisateur dans les images.

A partir des années 2000s, les chercheurs ont commencé à s'intéresser à la proposition des modèles d'indexation pouvant réduire ce problème de gap sémantique.

1.3. Objectifs et approches

Les aspects traités durant la préparation de la thèse sont liées principalement à l'amélioration de la performance (plus précisément la précision) des systèmes de recherche d'image par le contenu visuelle (CBIRs).

Les études menées durant la préparation des différents travaux visent principalement deux aspects importants. Le premier est celui du modèle d'indexation et pour lequel nous avons étudiés l'utilité et la performance des modèles d'indexations proposés. Le deuxième aspect visé concerne l'étude de la pertinence des caractéristiques, ce contexte nous a amenés à étudier l'utilisation de différentes modalités de caractéristique indépendamment et en combinaison.

Les modèles d'indexation proposés sont basés sur l'utilisation de deux modèles de sujets. Le premier est le modèle probabiliste de l'allocation de Pachinko (PAM pour Pachinko Allocation Model) proposé par (Wei and McCallum, 2006). Tandis que Le deuxième est un modèle déterministe de l'analyse de la sémantique latente (LSA pour Latent Semantic Analysis) proposé par (Deerwester et al., 1990). La famille de modèles de sujets contient des modèles qui ont été proposés initialement pour la modélisation de grands corpus de textes. Ils ont constitués par la suite une source d'inspiration pour le domaine de la recherche en traitement d'images.

L'adaptation de la représentation des caractéristiques visuelles extraites est nécessaire pour produire une représentation similaire aux textes. Cette représentation est indispensable pour l'application des modèles de sujets sur les images. La représentation construite est appelée sac de mots visuelles (BoVW pour Bag of Visual Words) et elle est basée sur le même principe de la représentation fréquentielle des sacs de mots textuelles.

Les travaux de la thèse sont décomposés en deux parties selon la nature des images traitées. La première partie s'intéresse par le domaine des images généraliste. Tandis que la deuxième est orientée vers le domaine des images médicales et plus précisément les tumeurs mammaires.

1.4. Organisation du document

Ce manuscrit est organisé de la manière suivante :

- Le premier chapitre présente le cadre d'étude puis la problématique de la recherche en indexation et recherche d'images par le contenu. Il est clos par la présentation de l'organisation de la thèse.
- Le deuxième et le troisième chapitre introduisent ensuite l'état de l'art de l'indexation et la recherche d'images basée sur le contenu visuelle. Le chapitre II présente les différents concepts de base en relation avec le processus d'implémentation des systèmes d'indexation et de recherche d'images. Tandis que le chapitre III illustre les concepts avancés du domaine, une dernière partie présente quelques exemples des systèmes de recherche d'images existants.
- Le quatrième chapitre est consacré à la présentation des modèles de sujets. Ces modèles ont été initialement proposés pour la modélisation des corpus de textes et par la suite constituent une source d'inspiration pour le traitement d'image.
- Le cinquième chapitre présente les différentes propositions qui constituent les bases de nos contributions.
- Enfin le dernier chapitre conclut sur les perspectives d'amélioration et les autres pistes de recherche suggérées par ce travail.

CHAPITRE – II

INDEXATION ET RECHERCHE

D’IMAGES

CONCEPTS DE BASE

2.1. Introduction

L'indexation et la recherche d'image est un domaine de recherche vaste. L'indexation est le module central d'un système de recherche d'image par le contenu visuel. Le rôle de ce module est l'extraction d'un ensemble de caractéristiques à partir des images et l'organiser sous forme de signature (index) pour l'utiliser par la suite dans la phase de la recherche. Cette signature doit être de taille réduite pour permettre une recherche rapide dans la base d'image. Elle doit aussi garantir la pertinence, c'est-à-dire la représentation maximale du contenu informationnel de l'image.

Depuis la définition du problème de fossé sémantique, le domaine est divisé en deux niveaux d'indexation. Le premier niveau dit bas niveau sémantique (Low Level Semantic Indexing) utilise les caractéristiques de base niveau extraites de l'image pour l'indexer directement. Le deuxième niveau dit de haut niveau sémantique (High Level Semantic Indexing) tente d'appliquer des modèles d'indexation plus sophistiqués sur les caractéristiques extraites pour produire une indexation plus précise qui reflète la sémantique de l'image.

Le domaine de la recherche en indexation et recherche d'images couvre à son tour plusieurs directions de recherches (façades d'intérêt (Datta et al, 2008)). Les questions clés résumant les problématiques du domaine ont été présentées par (Datta et al, 2008) comme suit:

- Comment réduire le gap sémantique?
- Comment concevoir et implémenter des systèmes utiles et performants?
- Comment proposer des nouveaux modèles d'indexation d'images?
- Comment standardiser le domaine et proposer des techniques d'évaluation et de comparaison des approches?

Ces questions introduisent les quatre axes de recherche principaux illustrés dans la figure 4. Ces axes sont :

- a) **Techniques et approches:** Cette direction du domaine s'articule autour de l'étude des différentes techniques liées à la construction de la signature

visuelle, la classification automatique, les paradigmes de recherche ainsi que les mesures de similarité utilisées.

- b) **Domaines d'application** : Dans cette direction, la recherche est concentrée sur l'investigation des possibilités d'application des techniques d'indexation pour des domaines tels que l'annotation, la sécurité, l'art, le web et l'apprentissage automatique.
- c) **Généraliser pour des bases réelles** : L'étude de cette direction fait appel à quatre aspects importants qui sont:
 - L'étude de l'intention d'utilisateur.
 - L'étude du domaine des données.
 - L'étude du traitement de requêtes.
 - L'étude la visualisation des résultats.

Tous ces aspects interviennent dans la réalisation des systèmes dans la partie d'interaction avec les utilisateurs.

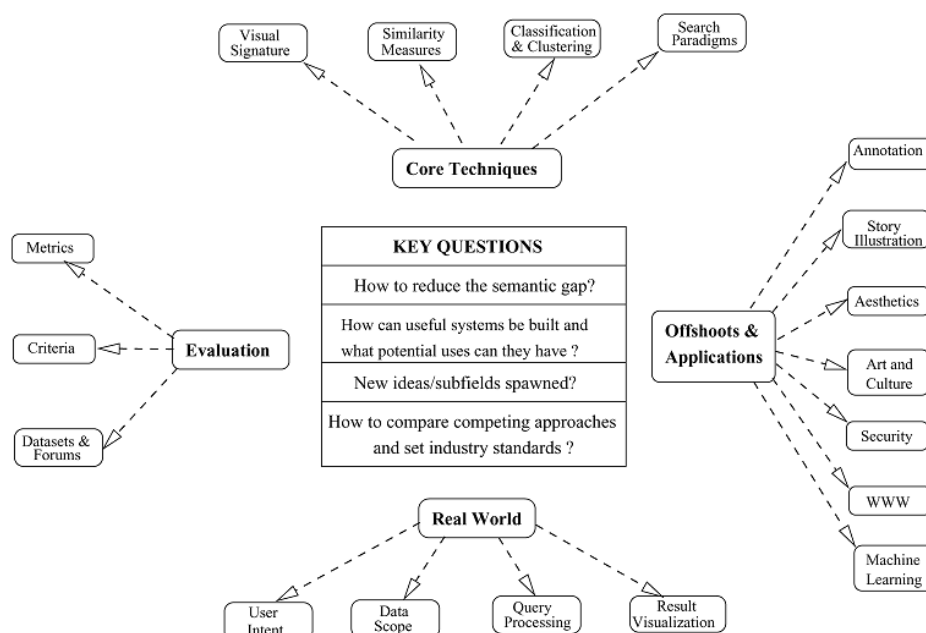


Figure 4. Les différents axes de recherche dans le domaine de l'indexation et de la recherche d'images (Datta R. et al., 2008)

- d) **Evaluation** : Les recherches en évaluation se concentrent sur l'étude des métriques et de mesures d'évaluation et de comparaison des différentes propositions. Cela nécessite la conception et la mise en place des bases

standards (benchmark) pour permettre aux chercheurs de standardiser les testes de différentes propositions.

L'indexation est l'étape préliminaire du processus de recherche d'images. Une vue de ce processus à été présentée dans l'architecture des systèmes de recherche d'images par le contenu (CBIRs) voir figure 2 dans le chapitre 1.

La figure 5 présente l'enchainement des étapes d'un processus d'indexation et de recherche d'images.

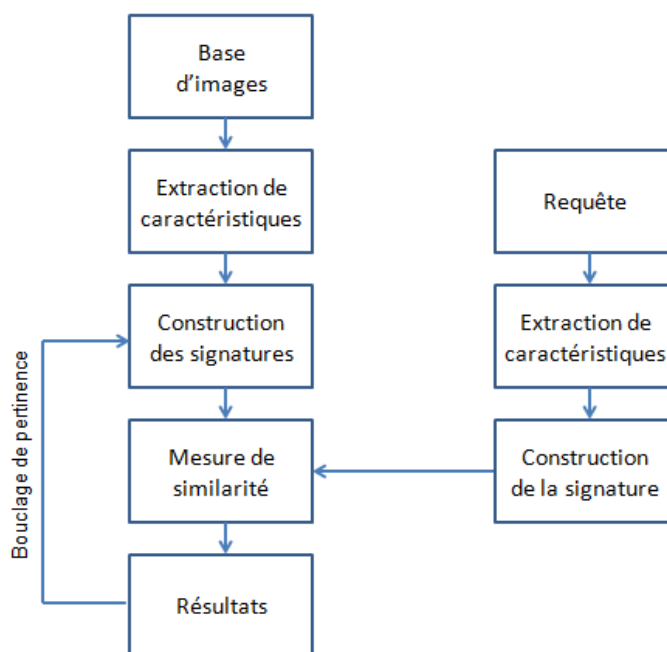


Figure 5. Le processus d'indexation et de recherche d'images par le contenu visuel

Ce chapitre est dédié à la présentation des concepts de base du domaine de l'indexation et de la recherche d'images. Ces concepts couvrent les bases d'images et les caractéristiques extraites pour représenter l'image.

La construction de la signature ainsi que les concepts avancés en liaison avec l'indexation de haut niveau sémantique (la réduction du fossé sémantique) vont être présentés dans le chapitre III.

2.2. Bases d'images

Les bases d'images peuvent être construites à partir de différentes sources. Néanmoins, comprendre la nature des images traitées joue un rôle important dans la conception et la réalisation des systèmes de recherches d'images performants.

Les bases d'images peuvent être généralement classées en deux catégories:

- **Bases spécifiques** : ce sont des bases avec un contenu spécifique d'un domaine précis (base médicale, ...). Ces bases sont homogènes et de petite taille.
- **Bases généralistes** : les bases généralistes contiennent des images diverses, hétérogènes issues de plusieurs domaines. Elles sont d'une taille très importante.

Les auteurs dans (Datta et al, 2008) ont également classés d'une manière plus détaillée les images pouvant constituer la source d'une base dans les catégories suivantes:

- **Les collections personnelles** : la taille de ces collections est généralement petite. Elle est accessible et utilisée par ses propriétaires à partir des supports de stockage locaux, les images de ces collections sont très homogènes.
- **Les collections des domaines** : elles contiennent des images homogènes accessibles par un public spécifique et contrôlé. De ce fait, les images peuvent être stockées d'une manière distribuée. Une telle catégorie peut contenir par exemple des collections des images médicales, biomédicales et satellitaires.
- **Collection d'entreprises** : ce sont des collections d'images hétérogènes accessibles par des utilisateurs d'une certaine entreprise via l'intranet. Le stockage peut être dispersé sur plusieurs sites, et l'accès dépend fortement de l'architecture de l'intranet.
- **Archives** : ils contiennent des images à intérêt historique, avec un accès ouvert sur internet mais à usage contrôlé. Les images sont homogènes et structurées ou semi-structurées sous des sujets spécifiques avec une masse très importantes.

- **Web** : l'accès et l'utilisation des images sur le web est pratiquement ouvert aux utilisateurs. Les moteurs de recherche d'images dans cette catégorie possèdent des robots de collection et de mise à jour de la base d'images, à cause de la nature dynamique du web. Les images de cette collection sont hétérogènes, semi-structurés avec une masse très importante.

2.2.1 Bases d'images standards (Benchmark):

Ces bases d'images sont des bases standards conçue et mis en place pour être utiliser par des chercheurs dans le domaine de l'indexation et la recherche d'images. Ces bases servent pour le teste des différentes propositions et contributions. Nous allons citer quelques bases d'images standards.

- **La base Corel 5K**

Cette base d'images est présentée par (Fei Fei et al., 2005). Elle constitue une base importante utilisée pour la recherche d'image par mots clef et l'annotation d'image. Elle contient 5000 images annotées manuellement avec 1 à 5 mots. Le vocabulaire contient 260 mots. La figure 6 montre des images issues de cette base.

- **La base Wang**

C'est une collection de 1,000 image sélectionnée à partir de la base Corel, elle est constituée de 10 classes avec 100 images pour chacune. L'ensemble de requêtes de teste est sélectionné à partir de ces classes, et les images réponse sont jugées pertinentes si elles sont de la même classe de la requête.



Figure 6. Des Images issues de la base Corel 5K (Web6, 2016)

- **La base IAPR TC12 :**

Cet ensemble d'image est composé d'environ 20 000 images accompagnées de descriptions dans plusieurs langues et il a été initialement publié pour la recherche documentaire multilingue. Elle peut être transformée en un format comparable aux autres bases via l'extraction des noms communs en ayant recours à des techniques de traitement de langage naturel, nous avons utilisés une collection issue de cette base dans le cadre de l'évaluation des travaux de la thèse. La figure 7 montre des images issues de la base IAPR TC12.



Figure 7. Des images issues de la base IAPR TC 12

- **La base UW dataset :**

C'est une base de l'université de Washington composée de 1109 images, et elles sont partiellement annotées. Les images sont réparties en 18 classes générales des pays, de la nature...etc. un ensemble de 352 mots clés sont utilisée. La figure 8 montre des images issues de cette base.

Il existe d'autres bases pour la recherche dans des domaines liés à l'indexation d'images tels que la reconnaissance d'objets.



Figure 8. Des images issues de la base UW data-set

Plus d'informations sur les bases benchmark peuvent être trouvées dans (Web6, 2016).

2.3. Extraction des caractéristiques des images

L'indexation d'image est basée principalement sur les différentes caractéristiques extraites et organisées pour présenter le contenu des images et y utiliser durant la recherche de celles-ci.

Généralement, ces caractéristiques appartiennent à deux grandes familles: les caractéristiques textuelles, et les caractéristiques visuelles. Nous allons présenter ces deux modalités dans ce qui suit.

2.3.1 Caractéristiques textuelles

Les descripteurs textuels sont composés des données textuelles. Ces données sont utilisées pour effectuer une indexation textuelle des images. Cette indexation peut être manuelle ou automatique.

➤ Indexation textuelle manuelle

L'indexation textuelle manuelle est réalisée par des opérateurs humains. Ils peuvent être des simples utilisateurs qui cherchent à identifier et classer leurs collections personnelles, comme ils peuvent être aussi des documentalistes spécialisées en indexation des collections d'images pour des organismes professionnels tels que les agences de presse. La tâche de cette indexation consiste à classer les images en les associant des groupes de mots décrivant certaines catégories. Deux niveaux d'indexation manuelle sont abordés, le premier est lié à ce que l'indexeur voit dans l'image, tandis que le deuxième s'intéresse à la signification de l'image. Un thésaurus peut être utilisé pour faciliter la tâche.

Les difficultés liées à ce type d'indexation peuvent être résumées dans les points suivants:

- **Subjectivité** : le choix des termes qui décrivent le contenu de l'image est fortement subjectif. Les termes choisis par l'indexeur et les utilisateurs durant la recherche ne sont pas forcément les même. Ces termes peuvent

avoir plusieurs sens (polysémie). La probabilité pour que le même terme soit choisi par deux individus différents pour décrire une entité quelconque est inférieurs à 20% (Tollari , 2006).

- **Volume important des données :** malgré le problème de la subjectivité, l'indexation textuelle manuelle reste efficace pour donner une sémantique à l'indexation. Mais cette efficacité n'est plus assurée vue que la taille des bases à indexer est de plus en plus volumineuse et donc difficile et voir impossible à l'indexer avec ce type.

➤ **Indexation textuelle automatique**

C'est le type d'indexation où les images sont annotées automatiquement sans aucune intervention des humains. On trouve dans ce type d'indexation les deux approches suivantes:

- **Indexation textuelle automatique à partir du texte associée :** cette approche est basée sur l'exploitation du texte associé aux images (voir figure 9). Ce texte peut constituer une source utile pour décrire le contenu de l'image. Les images annotées automatiquement par cette approche sont issues des catalogues, encyclopédies, web etc.

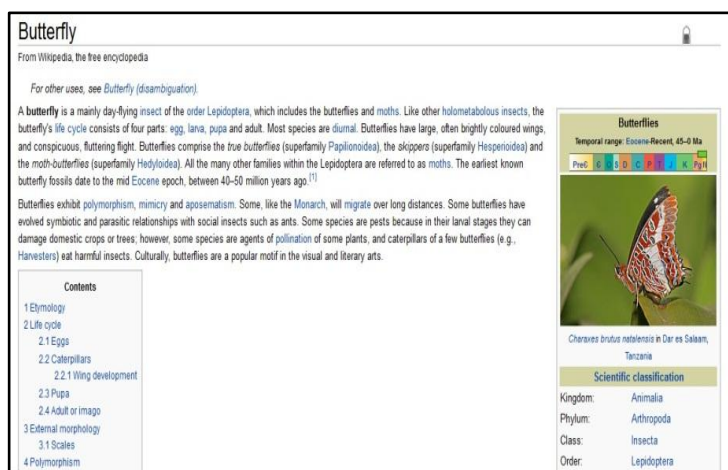


Figure 9. Exemple d'une image de web avec le texte associé (Web 7, 2009)

Dans le cas des images des pages web, le texte présent dans la page HTML constitue la source des mots associés aux images. Ces mots sont choisis à partir des différentes balises de la page et donc leurs importances varient d'une balise à l'autre. Les mots présents dans les titres ainsi que les balises avec l'attribut ALT d'images sont jugés les plus importants (pertinents).

L'inconvénient majeur de cette approche et le bruit souvent rencontrés dans les textes sources. En effet, la majorité des mots présents dans ceux-ci ne reflètent pas le contenu des images ce qui engendre une mal indexation.

- **Indexation textuelle automatique à partir du contenu visuel**

Dans cette approche, aucun texte n'est disponible pour l'utiliser en indexation des nouvelles images. Les chercheurs tentent à concevoir et réaliser des systèmes capable à apprendre à classer les nouvelles images dans des classes décrites chacune avec un mot. Une phase d'apprentissage sur des images manuellement annotées est nécessaire. Cette approche rencontre aussi le problème du bruit car les propositions n'arrivent pas encore à donner des résultats efficaces et donc l'indexation est souvent mal et bruitée. Les deux approches de l'indexation textuelles automatique peuvent être combinées. L'approche décrit un domaine de recherche qui se distingue de l'indexation d'images à base de contenu visuel. Ce domaine est l'auto annotation des images.

2.3.2 Caractéristiques visuelles

Les caractéristiques visuelles (visual features) sont des caractéristiques extraites à partir du contenu visuel des images. L'extraction de ces caractéristiques nécessite les deux premières étapes illustrées dans la figure 10.

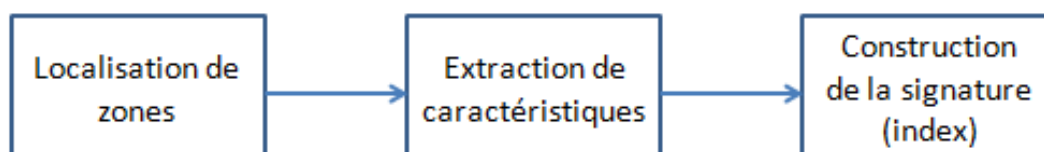


Figure 10. Etapes de la construction de la signature visuelle d'une image (index visuel)

2.4. Localisation de régions d'intérêt

Les caractéristiques visuelles des images sont extraites suivant deux approches. Ces approches concernent la localisation des régions d'intérêt pour lesquelles les caractéristiques vont être calculées, ces approches sont:

2.4.1. Approche local

Dans cette approche. Les caractéristiques visuelles sont extraites pour des points spécifiques dans l'image. Il existe plusieurs méthodes permettant la localisation de ces points clés (points d'intérêts). Nous citons le détecteur basé sur la différence gaussienne (DoG) (Lowe, 2004) utilisé avec la technique SIFT (Salient Invariant Feature Transform) et le détecteur basé sur les filtres carrés utilisé avec la technique SURF (Speeded Up Robust Features) (Bay et al., 2008).

2.4.2. Approche globale

Dans cette approche, les caractéristiques sont extraites soit pour l'image entière, soit pour une partie de l'image appelée régions. Dans le cas où elles sont calculées pour des régions dans l'image, celle-ci doit être tout d'abord divisée pour spécifier ces régions. Cette division de l'image en partie est appelée segmentation de l'image et, elle peut être régulière ou irrégulière (selon les formes produites).

2.5. Segmentation des images

La segmentation de l'image est une opération qui consiste à décomposer l'image en régions connexes selon un certain critère d'homogénéité (couleurs,...). L'union des régions doit reproduire l'image initiale, la figure 11 montre un exemple de la segmentation de deux images en utilisant l'algorithme JSEG (Liu et al., 2007).



Figure 11. Résultats de la segmentation avec l'algorithme de segmentation JSEG pour deux images (Liu et al., 2007)

Les régions produites peuvent être régulières ou irrégulières. La segmentation régulière de l'image est le résultat d'un découpage de l'image en blocs carrés de même taille (ex : découpage de l'image en 16 blocs 4*4). La segmentation irrégulière résulte de l'application d'un algorithme de segmentation.

Les algorithmes de la segmentation d'images sont généralement classés en trois catégories:

2.5.1. Segmentation basée sur l'histogramme

Dans cette approche, les histogrammes de couleurs sont la base de la segmentation. L'histogramme est divisé en clusters de couleurs dont la projection sur l'image engendre sa segmentation. Cette approche suppose que chaque cluster représente une région de l'image, mais en pratique il faut parfois séparer deux régions du même cluster de couleur parce qu'elles représentent deux entités différentes dans l'image.

2.5.2. Segmentation basée région

Pour produire une segmentation de l'image, cette approche traite directement les pixels suivant l'une des trois techniques suivantes :

- **Agrégation :** C'est une technique ascendante qui à partir de petites régions composées (d'un ou de quelques pixels), regroupe d'une manière itérative les pixels adjacents de même intensités.

- **Décomposition** : C'est une technique descendante qui découpe l'image itérativement en parties qui sont à leurs tours découpés en partie plus petite tant qu'il existe des parties hétérogènes en intensités.
- **Diviser pour réunir** : C'est une technique hybride des deux techniques précédentes. Un exemple d'un algorithme de cette technique est l'algorithme CSC (Color Structure Code) (Hartmann G, 1987).

2.5.3. Segmentation basée contours

Dans cette approche, les algorithmes cherchent à trouver les différents contours dans l'image. Des filtres de détection de contours de contour sont appliqués pour découvrir les parties de l'image avec un changement brusque de niveaux d'intensité et cela pour définir les contours des différentes régions. Le problème dans cette approche est que les contours produits sont rarement continue ce qui implique la nécessité d'appliquer d'autres algorithmes pour compléter la tâche de fermeture de contours.

2.6 Extraction de caractéristiques locales

Nous avons mentionnés dans la section précédente deux approches de localisation de zones d'intérêts, les caractéristiques extraites sont calculées selon l'approche choisie. Nous allons citer quelques caractéristiques des deux approches locale et globale.

Dans cette approche, les caractéristiques des images sont calculées dans des points spécifiques, la vue locale appliquée pour ce calcul comporte un ensemble restreint de pixels. Généralement, pour extraire ses caractéristiques locales deux phases sont appliquées. La figure 12 montre un exemple d'imagettes sélectionnées pour calculer les caractéristiques locales de l'image.

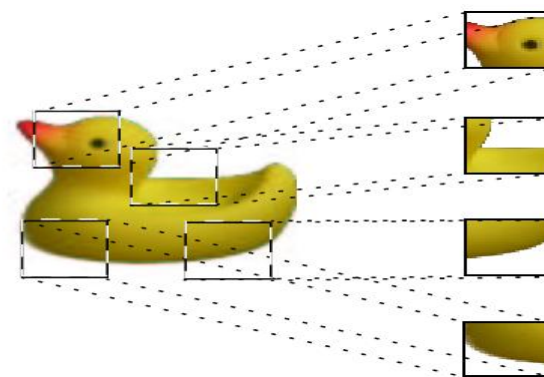


Figure 12. Exemple de régions d'intérêts choisies pour l'extraction des caractéristiques locales (Tollari S., 2006)

La première est une phase de détection des points locaux de l'image durant laquelle un filtre détecteur de point de fort contraste est appliqué. Un exemple de tel filtre est le détecteur DoG (Difference of Gaussian).

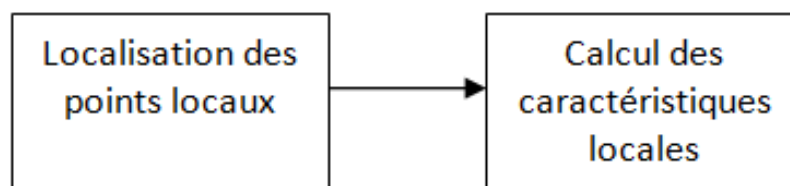


Figure 13. Les phases d'extraction des caractéristiques locales

SIFT (Salient Invariant Feature Transform) (Lowe, 2004) et SURF (Speed Up Robuste Features) (Bay et al., 2006) sont deux techniques d'extraction caractéristiques locales des images. SIFT calcule des histogrammes des orientations du gradient comme caractéristique, et SURF utilise les approximations d'ondelettes de Haar. Nous allons présenter la technique SIFT.

2.6.1. SIFT (Salient Invariant Features Transform)

Le SIFT utilise un détecteur DoG (Difference of Gaussian) pour identifier les points locaux d'intérêts dans l'image. DoG est un détecteur invariant à l'échelle, il commence par une phase de détection des points d'intérêt puis passe à une phase de filtrage pour garder seulement les points stables sous un certain niveau de bruits ajouté.

Pour identifier les points clés, l'image est scannée suivant la location et l'échelle. Les emplacements et les échelles des points clés sont détectés comme l'extrême de la fonction $D(x,y,\sigma)$ qui représente la différence gaussienne convolutionnée avec l'image $I(x,y)$ (voir la formule 1).

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \quad (1)$$

Avec k un facteur constant multiplicatif et $G(x,y,\sigma_i) = \frac{1}{2\pi\sigma_i^2} e^{-(x^2+y^2)/2\sigma_i^2}$ un noyau gaussien.

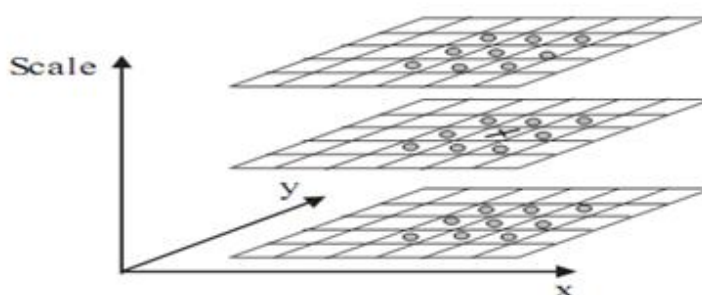


Figure 14. Détection des extrêmes par comparaison du pixel d'intérêt avec ses voisins du niveau courant ainsi que les niveaux adjacents (Lowe, 2004)

Les extrêmes locaux sont détectés en comparant chaque pixel avec ses 26 voisins, 8 du niveau courant et 9 des deux niveaux inférieur et supérieur (voir figure 14). Un point est sélectionné s'il est le plus grand ou le plus petit parmi tous ses voisins.

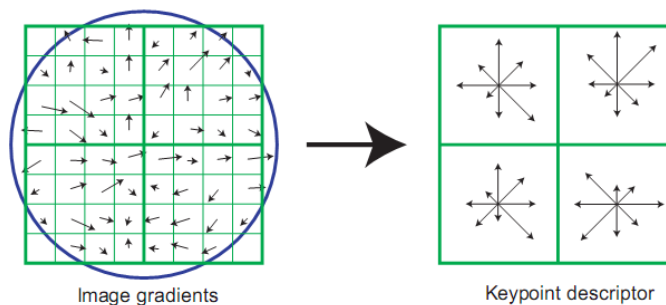


Figure 15. 2x2 vecteurs descripteurs calculés à partir d'un échantillon de 8x8 (Lowe, 2004)

Le calcul des caractéristiques des points clés détectés par DoG est réalisé avec la technique SIFT. Une orientation, une échelle et un emplacement sont assignés aux points clés. L'échelle et l'emplacement sont issues du détecteur DoG, tandis qu'une ou plusieurs orientations sont assignées suivant l'orientation dominante du gradient de la fenêtre qui entoure le point. Un histogramme d'orientations est calculé pour identifier les directions des gradients dominants. Le vecteur caractéristique final est formé de 128 éléments et il est invariant par rapport aux différentes transformations.

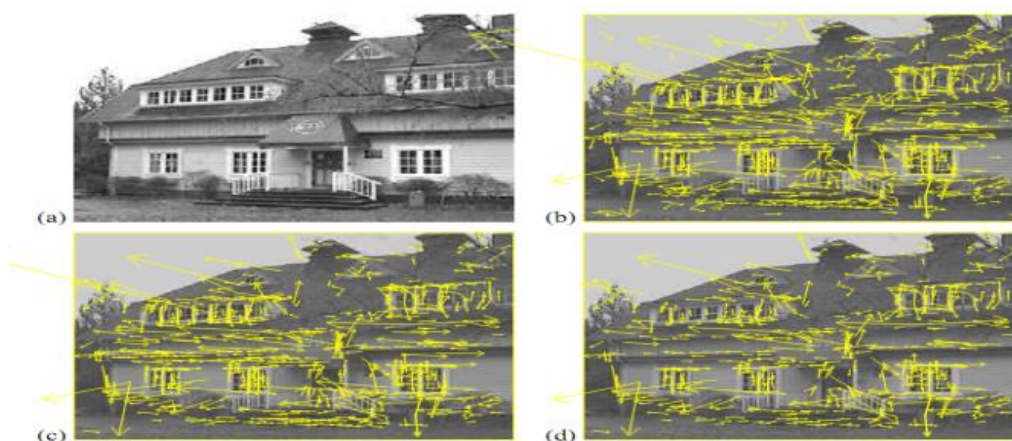


Figure 16. Une image et les points de caractéristiques locales extraites (Lowe D.G., 2004)

2.7 Extraction de caractéristiques globales

Cette catégorie comporte essentiellement les caractéristiques de la couleur, la texture et de la forme.

2.7.1. Les caractéristiques de la Couleur

La couleur est une caractéristique importante et largement utilisé pour caractériser les images dans le domaine. Les caractéristiques de la couleur sont calculées pour un système choisi, les systèmes de couleurs les plus utilisés sont le RGB (Red, Green, Bleu) et le HSV (Hue, Saturation, Value), plus de détails sur les systèmes de couleur sont inclus dans (Datta et al., 2008). La couleur de l'image peut être caractérisée par la matrice de covariance, l'histogramme de couleur, les moments de couleurs, et les vecteurs de cohérence de couleur.

a- Les systèmes de couleurs

Les systèmes de représentation des couleurs sont nombreux, ils sont appelé également espaces de couleur et disposent de diverses propriétés et par conséquent diverses applications. La figure 17 montre la perception de l'œil humain. Nous allons présenter dans cette section les principaux systèmes de couleurs.

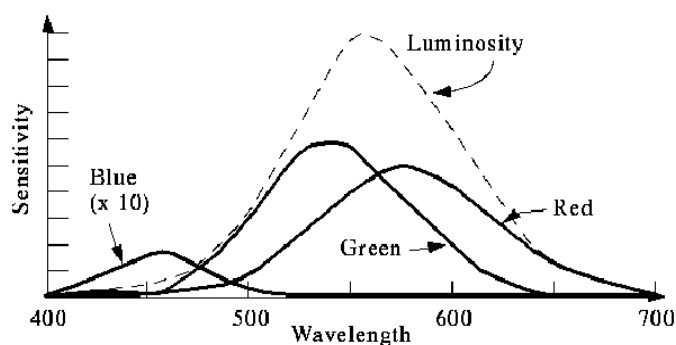


Figure 17. La perception de l'œil humain (Web 4, 2016)

La couleur est le résultat de la perception de la lumière par le système visuel humain. On distingue généralement entre deux types de couleur, la couleur physique et la couleur perçue. **La couleur physique** est caractérisée par le spectre de la lumière qui arrive sur l'œil. Le spectre d'un rayonnement lumineux est l'ensemble des longueurs d'ondes qui le composent, associées à leurs intensités respectives. Le spectre peut être mesuré par un spectromètre. **La couleur perçue** reflète l'impression visuelle. La couleur perçue est différente selon chaque individu. On ne peut pas la mesurer. Deux couleurs physiques différentes peuvent produire la même couleur perçue.

Les systèmes de représentation des couleurs ont pour objectif de représenter la couleur perçue en utilisant des valeurs. Tous les systèmes représentent (à l'exception du système CMJN), la couleur par 3 valeurs. Un espace à trois dimensions peut donc être utilisé pour représenter les couleurs comme de points de cet espace. Nous allons présenter dans ce qui suit les deux systèmes de couleurs utilisés durant la préparation des travaux de la thèse. Pour plus de détails sur les systèmes de couleurs veuillez consulter (Web 3).

- **Espace RVB / RGB**

Le système de couleurs RVB (Rouge Vert Bleu) RGB (Red Green Blue) est l'espace le plus utilisé pour représenter la couleurs. Les exemple de son utilisation sont divers. La télévision, les caméras et les moniteurs des ordinateurs utilisent ce système lors de l'affichage et de l'acquisition des images. Chaque couleur est représentée par 3 valeurs qui sont les quantités de rouge, de vert et de bleu qu'elle contient. L'espace RGB est basé sur la synthèse additive.

Les couleurs peuvent être codés sur 8 bits avec des valeurs comprises entre 0 et 255. La normalisation de ces valeurs (valeurs comprises entre 0 et 1) est également utilisée. Le tableau 1 montre des exemples de couleurs et leurs valeurs codées dans le système RGB.

Valeurs de RGB (normalisée)	Couleur représentée
(0,0,0)	noir
(1,1,1)	blanc
(1,0,0)	rouge
(0,1,0)	vert
(0,0,1)	bleu
(0,1,1)	cyan
(1,0,1)	magenta
(1,1,0)	jaune

Tableau 1. Exemple de valeurs de RGB avec les couleurs correspondantes

L'espace RGB est représenté graphiquement par cube de dimensions $1 \times 1 \times 1$ (voir figure 18), placé à l'origine d'un repère à trois dimensions dont les axes indiquent la quantité de Rouge, de Vert et de Bleu.

Chaque couleur est représenté par à un point C de coordonnée (R_c, V_c, B_c) situé dans le cube. Les couleurs Rouge, Vert, Bleu, Cyan, Magenta, Jaune , Noir et Blanc sont situées aux sommets du cube.

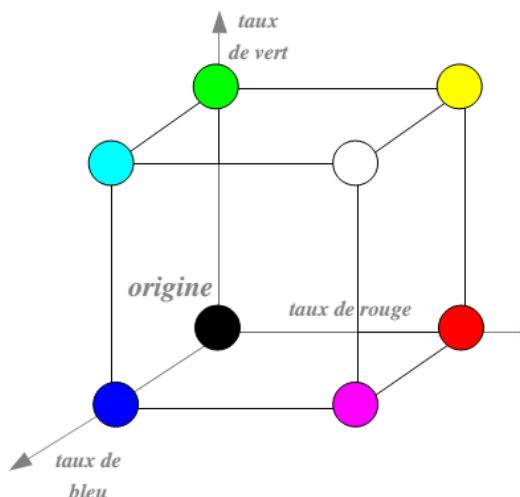


Figure 18. Cube des couleurs RGB (Manzanera A., 2008)

La couleur Noir occupe l'origine du repère. Les couleurs primaires sont situées sur les 3 axes à la distance 1 (voir figure 19).

La diagonale principale du cube est une droite tracée entre l'origine (0,0,0) et le point (1,1,1) représentant du blanc. Elle contient les points qui représentent les couleurs ayant des quantités égales de Rouge, de Vert, et de Bleu.

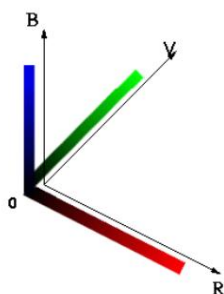


Figure 19. Dimensions de l'espace RGB (Web5, 2016)

Ces couleurs sont les différents niveaux de gris. Une couleur est choisie en fixant les trois valeurs de rouge, vert et bleu.

- **Espace HLS / HSV**

Le système de couleur HSV (Hue, Saturation, Value) est basé sur trois composantes utilisées pour la représentation de la couleur. Ces composantes sont la teinte, la saturation et la luminance.

L'espace est appelé TLS (Teinte, Luminance, Saturation), HSV (Hue, Saturation, Value) ou HLS (Hue, Luminance, Saturation). Cet espace décompose la couleur selon des caractéristiques plus intuitives, proches du vocabulaire courant pour décrire une couleur. Il se base sur la décomposition de la couleur en une couleur pure du spectre de l'arc en ciel à laquelle est ajoutée un gris (addition de R,V,B à parts égales).

➤ **Teinte (Hue)**

Les termes couramment utilisés pour désigner la couleur d'un objet (Rouge, Violet, Rose ...) expriment la notion de teinte. La teinte indique la couleur pure qui est prépondérante dans la couleur. Les teintes peuvent être représentées par un angle sur un cercle (voir figure 20).

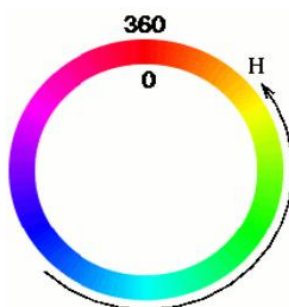


Figure 20. Cercle de la teinte (Web5)

Idéalement associé à une longueur d'onde, donc à une position sur le cercle de Newton. La figure 21 montre la composante de la teinte pour une image couleur.



Figure 21. La composante teinte d'une image couleur (Manzanera, 2008)

➤ **Saturation**

Plus la saturation est faible plus la couleur est pastel ou délavée. Plus la saturation est élevée, plus la couleur est pure (vive). La saturation correspond à la proportion de

couleur pure. Elle varie entre 0 et 1. $S=0$: gris $S=1$: couleur pure (voir la figure 22).



Figure 22. Variation de la saturation (Web5)

La saturation représente le taux de pureté de la couleur, qui doit varier entre la pureté maximale (couleur éclatante) et l'achromatisme (niveau de gris). La figure 23 montre la composante de saturation pour une image couleur.



Figure 23. La composante de saturation d'une image couleur (Manzanera, 2008)

➤ **Valeur (Value)**

La luminance caractérise l'intensité totale du point lumineux coloré. Elle doit varier entre le noir absolu et le blanc (voir figure 24).



Figure 24. Variation de la luminance (Web5)

La figure 25 montre la composante de luminance pour une image couleur. L'image est la même dans les figures 21 et 23 qui montrent les deux autres composantes.



Figure 25. La composante de luminance d'une image couleur. (Manzanera, 2008)

L'ensemble des couleurs représentables dans l'espace HLS est délimité par deux cônes ayant la même base. Pour une luminance donnée les couleurs sont situées sur un disque. Si on fixe une teinte, et qu'on fait varier la saturation, on se déplace sur un rayon du disque (voir la figure 26).

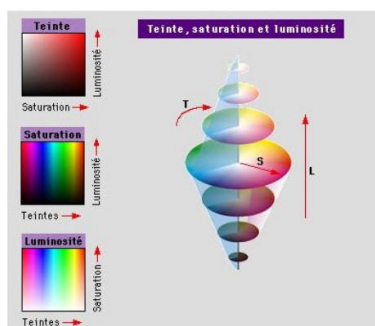


Figure 26. Représentation de l'espace HSV (Web5)

La surface du cône contient l'ensemble des couleurs pures. L'axe du cône contient les gris du noir au blanc. Voici l'ensemble des formules de passage entre l'espace HLS et l'espace RVB (Web5):

$$L = \frac{1}{3}(R + V + B)$$

$$S = 1 - \frac{3}{R + V + B} \min(R, V, B) \quad (1)$$

$$H = \arccos\left(\frac{0.5(R - V) + 0.5(R - B)}{\sqrt{(R - V)^2 + (R - B)(V - B)}}\right) \quad H = 360^\circ - H \text{ si } B > V$$

2.7.1.3. Moments de couleur

Les moments de la couleur sont la moyenne, la variance et l'écart type. Elles sont définies dans les équations suivantes (pour une image de taille $n \times m$):

$$\text{Moyenne} = \frac{\sum_{i=1}^n \sum_{j=1}^m x_{ij}}{nm} \quad (2)$$

$$\text{Variance} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \text{Moyenne})^2 \quad (3)$$

$$\text{écart - type} = \sqrt{\text{Variance}} \quad (4)$$

Avec x_{ij} la valeur du pixel de la $i^{\text{ème}}$ ligne et de la $j^{\text{ème}}$ colonne. L'asymétrie peut remplacer l'écart-type.

2.7.1.3. Histogramme de couleur

L'histogramme des couleurs représente la distribution de la couleur au sein de l'image (une région de l'image). Pour construire un histogramme de couleur, la première phase sert à quantifier les couleurs de l'image (ou de la région). Puis pour chaque niveau de couleur, les occurrences de pixel ayant ce niveau de couleur sont comptées pour être dessinées. Chaque couleur avec son intensité est considéré comme un bin. La figure 27 montre des exemples des histogrammes de couleurs.

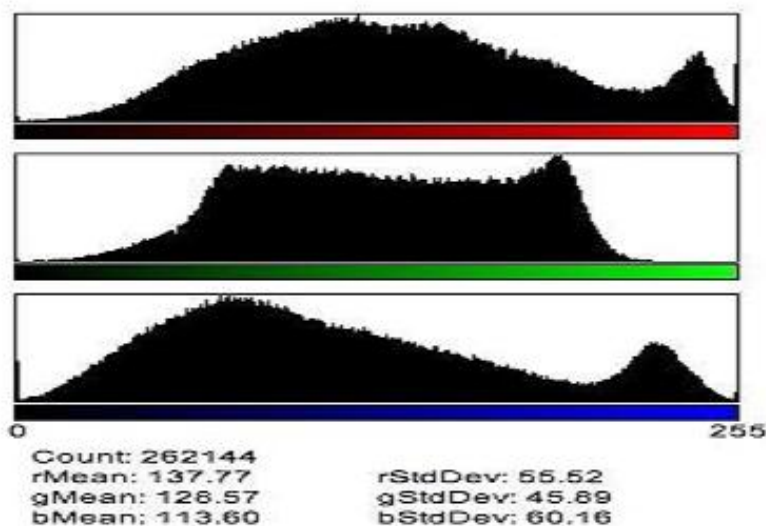


Figure 27. Histogrammes de couleur RGB (Meskaldji et al., 2009)

L'histogramme de la couleur sous la façon que nous venons de décrire ne capture pas l'information spatiale entre les pixels.

2.7.2. Les caractéristiques de la texture

La texture peut être définie en tant qu'une forme visuelle ayant des caractéristiques d'homogénéité qui sont absentes dans le cas d'une seule intensité.

La texture est une caractéristique utile pour plusieurs types d'images. La perception humaine utilise la texture pour la reconnaissance et l'interprétation du contenu des images. La figure 28 présente des images exemples pour des propriétés de la texture.



Figure 28. Exemple de textures naturelle (web7)

La texture est calculée pour un ensemble de pixels. Plusieurs techniques ont été proposées pour l'extraction des caractéristiques de la texture, elles sont généralement classées en deux catégories selon le domaine d'extraction. Ces deux grandes catégories sont l'extraction spatiale des caractéristiques de la texture et l'extraction spectrale des caractéristiques de la texture. Dans la première approche, les caractéristiques sont extraites en calculant des statistiques des pixels dans le domaine spatial de l'image. Dans la seconde approche l'image est transformée dans l'espace fréquentiel qui constitue la source de calcul des caractéristiques.

Parmi les techniques d'extraction de caractéristiques, les fonctions de Haralick et les filtres de Gabor et les caractéristiques de Tamura ont été largement utilisées.

2.7.2.1 Les caractéristiques d'Haralick

Les caractéristiques de Haralick (Haralick R. M., 1979) sont calculée sous la forme d'un ensemble de fonctions appelée fonctions d'Haralick. Ces caractéristiques sont basées sur une matrice de cooccurrence des niveaux de gris, la figure 29 montre la structure de la matrice de cooccurrence.

$$\begin{array}{cccc}
 0 & 0 & 0 & 1 \\
 0 & 0 & 1 & 1 \\
 0 & 2 & 2 & 3 \\
 2 & 2 & 3 & 3
 \end{array}
 \quad
 P_{0^\circ, 1} = \begin{bmatrix} 6 & 2 & 1 & 0 \\ 2 & 2 & 0 & 0 \\ 1 & 0 & 4 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

Figure 29. Matrice de cooccurrence calculée pour 4*4 pixels pour la direction 0°.

La matrice de cooccurrence des niveaux de gris est une matrice carrée N*N avec N le nombre de niveaux de gris. Les éléments x_{ij} sont représentent le nombre d'occurrences des pixels avec le niveau i en adjacence avec les pixels de niveau de gris j, la matrice est ensuite normalisée par le nombre de comparaison réalisé. Les éléments de la matrice sont considérés comme la probabilité que le pixel ayant un niveau de gris i soit adjacent avec le pixel ayant un niveau de gris j. Le voisinage du pixel peut être calculé selon quatre directions ce qui donne une possibilité de calcul de quatre matrices de cooccurrences (voir la figure 29).

Les cinq premières caractéristiques sont les plus utilisées car elles sont visuellement significatives. Nous avons utilisés les 14 caractéristiques de Haralick dans le cadre de cette thèse.

2.7.2.2. Les caractéristiques de Tamura

Les auteurs de (Tamura et al., 1978) ont proposés six caractéristiques pour la texture qui correspondent à a perception humaine, ces caractéristiques représente : grossièreté (coarseness), contraste (contrast), direction (directionality), ressemblance à la ligne (ligne-likeness), régularité (regularity) et la rugosité (roughness). Les auteurs ont trouvés que les trois premières caractéristiques sont très importantes car elles ont un impact implicite sur la perception humaine.

Grossièreté: cette caractéristique donne l'information sur la taille des éléments de la texture. Une grande valeur de cette caractéristique reflète une texture grosse. La grossièreté est calculée comme suit :

- 1- pour chaque point (n_0, n_1) , la moyenne sur les voisins est calculée. La taille de voisinage est puissance de deux :

$$A_k(n_0, n_1) = \frac{1}{2^{2k}} \sum_{i=1}^{2^{2k}} \sum_{j=1}^{2^{2k}} X(n_0 - 2^{k-1} + i, n_1 - 2^{k-1} + j) \quad (5)$$

- 2- pour chaque point (n_0, n_1) , la différence entre les voisinages non chevauchés dans les directions opposées horizontales et verticales est calculée comme suit :

$$E_k^h(n_0, n_1) = |A_k(n_0 + 2^{k-1}, n_1) - A_k(n_0 - 2^{k-1}, n_1)| \quad (6)$$

et

$$E_k^v(n_0, n_1) = |A_k(n_0, n_1 + 2^{k-1}) - A_k(n_0, n_1 - 2^{k-1})| \quad (7)$$

- 3- pour chaque point (n_0, n_1) , la taille qui donne la différence maximale est sélectionnée :

$$S(n_0, n_1) = \operatorname{argmax} \max E_k^d(n_0, n_1) \quad k = 1 \dots 5 \quad d = h, v \quad (8)$$

- 4- l'étape finale consiste à prendre la moyenne sur 2^S comme une mesure de grossièreté de l'image et cela comme suit :

$$F_{gross} = \frac{1}{N_0 N_1} \sum_{n_0=1}^{N_0} \sum_{n_1=1}^{N_1} 2^{S(n_0, n_1)} \quad (9)$$

Contraste : le contraste de l'image peut refléter sa qualité. Il est influencé par les quatre facteurs suivants :

- La dynamicité des niveaux de gris.
- La polarisation de la distribution du noir et blanc dans l'histogramme de niveau de gris.
- Les contours pointus.
- La période de répétition des formes.

Le contraste d'une image est calculé comme suit :

$$F_{con} = \frac{\sigma}{\alpha_4^z} \quad \text{avec} \quad \alpha_4 = \frac{\mu_4}{\sigma^4} \quad (10)$$

$\mu_4 = \frac{1}{N_0 N_1} \sum_{n_0=1}^{N_0} \sum_{n_1=1}^{N_1} (X(n_0, n_1) - \mu)^4$ est le quatrième moment de la moyenne μ , σ^2 est la variance de niveau de gris de l'image et $z = \frac{1}{4}$ un facteur empirique.

Direction : l'orientation est pertinente pour la définition de la direction. Pour calculer la direction, les deux dérivés horizontale et verticale Δ_H et Δ_V sont calculées en effectuant une convolution à l'image $X(n_0, n_1)$ respectivement avec les deux opérateurs suivants :

-1	0	1
-1	0	1
-1	0	1

-1	-1	-1
0	0	0
1	1	1

Et pour chaque élément (n_0, n_1)

$$\theta = \frac{\pi}{2} + \tan^{-1} \frac{\Delta_V(n_0, n_1)}{\Delta_H(n_0, n_1)} \quad (11)$$

Un histogramme est calculé avec ces valeurs et la direction est déterminée par la somme des seconds moments des sommets.

2.7.2.3. Les caractéristiques globales de la Texture

Les auteurs de (Terhorst B., 2003 ; Deselaers T., 2003) ont présentés un descripteur de texture globale pour l'image entière. Ce descripteur est composé de cinq parties chacune caractérise un aspect de la texture de l'image. Ces caractéristiques sont :

Dimension fractale : c'est une mesure de rugosité et de la surface.

Grossièreté : elle mesure la taille des grains de l'image. Elle dépend de la variance de l'image.

Entropie : pour mesure l'absence d'uniformité dans l'image.

Différence de niveau de gris spatial : elle est utilisée pour mesurer la relation de luminosité des pixels avec leurs voisinages.

Fonction d'auto-corrélation circulaire de Moran : pour mesurer aussi la grossièreté de l'image en utilisant un ensemble de fonctions d'auto-corrélation.

2.7.3. Les caractéristiques de la forme

La forme peut être considérée comme un élément important utilisé pour identifier et différencier les différents objets. Les techniques pour l'extraction des caractéristiques de la forme peuvent être généralement classées en deux catégories (Zhang and Lu, 2004).

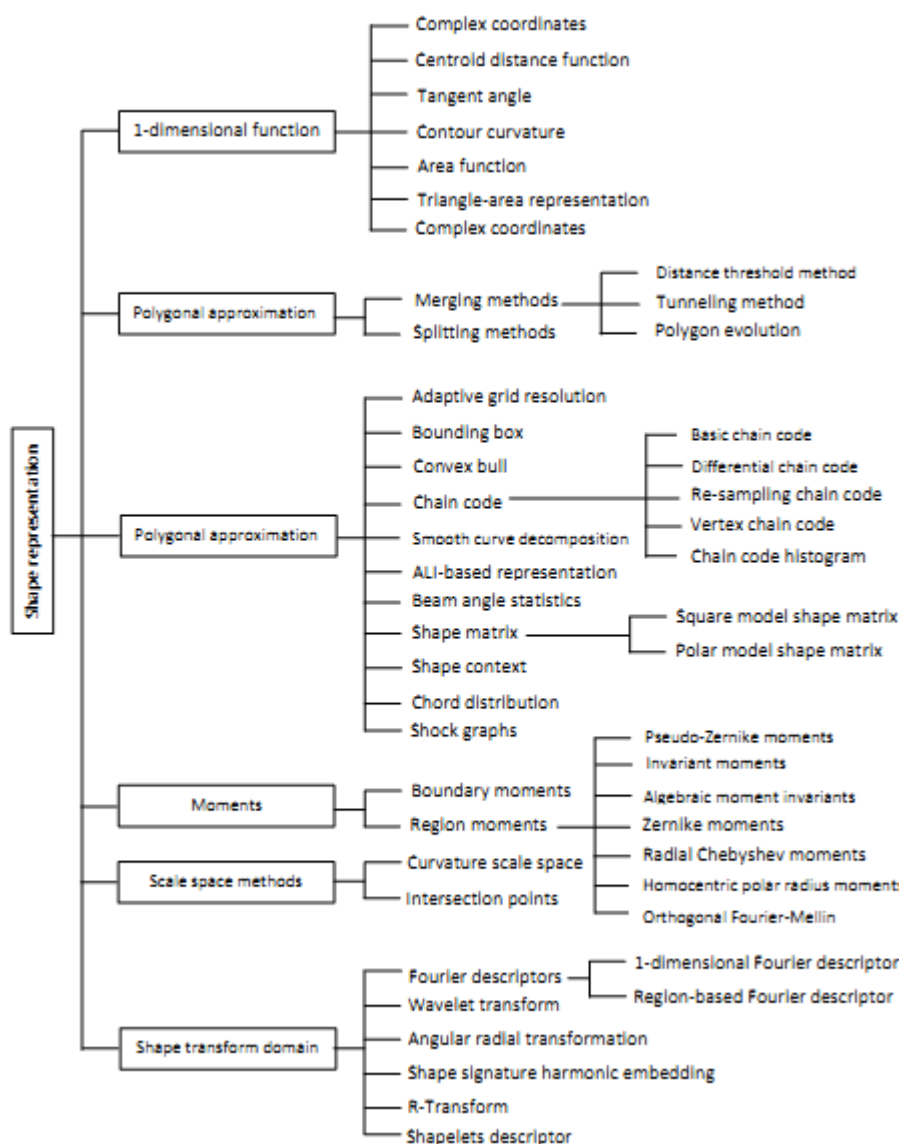


Figure 30. Classification des techniques d'extraction des caractéristiques de la forme (Yang et al., 2008)

La première catégorie des techniques est basée contours ou les caractéristique de la forme sont extraite seulement à partir des contours, tandis que la deuxième est basée régions avec une extraction des caractéristiques à partir de la région toute entière.

La figure 30 montre une classification des approches d'extraction des caractéristiques de la forme. Cette classification est donnée par (Yang et al., 2008) qui présente plus de détails sur le sujet.

2.8 Conclusion

Nous avons présentés dans ce chapitre les concepts de base liés au domaine de l'indexation et la recherche d'images. Ces concepts incluent les différentes phases d'extraction de caractéristiques de bas niveaux qui peuvent caractériser généralement la couleur, la texture et la forme de l'image. Nous allons présenter dans le chapitre suivant les concepts avancés liés à la représentation de ces caractéristiques et le traitement des différents problèmes de haut niveau de la recherche d'images.

CHAPITRE – III

INDEXATION ET RECHERCHE

D'IMAGES

CONCEPTS AVANCES

3.1 Introduction

L'avancement de la recherche dans le domaine de l'indexation et la recherche d'images a conduit à l'apparition de la problématique du fossé sémantique. Le fossé sémantique est définie par (Smeulders, 2000) comme étant *'the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation'*. Les travaux depuis ce temps tentent à réduire ce fossé en traitant plus la pertinence des caractéristiques et en proposant des modèles d'indexation plus sophistiqués. Les autres modules des systèmes CBIRs ont été également étudiés avec plus de concentration.

Nous allons présenter dans ce chapitre les notions avancés des systèmes CBIRs. Ces notions sont liées d'une manière ou d'une autre avec les problèmes du fossé sémantique. Les sujets traités par ce chapitre couvrent la construction de la signature (index), les mesures de similarité, les différents aspects en relation avec l'interaction entre l'utilisateur et le système. Une section est dédiée à la présentation des différentes approches de réduction de fossé sémantique. L'évaluation des systèmes est également abordée.

3.2 Construction de la signature

La signature de l'image désigne la manière avec laquelle les caractéristiques extraites à partir de l'image vont être représentées et par la suite exploitées.

La construction de la signature visuelle à partir des caractéristiques des images peut suivre différentes formulations, la figure 31 montre ces différentes formulations. La première formulation est mathématique et selon laquelle deux mode de représentation des caractéristiques peuvent être utilisé, les vecteurs et les distributions.

Les histogrammes et les caractéristiques des régions peuvent être considères comme vecteurs et distributions discrètes respectivement. Cependant, une distribution continue peut être construite à partir des caractéristiques locales.

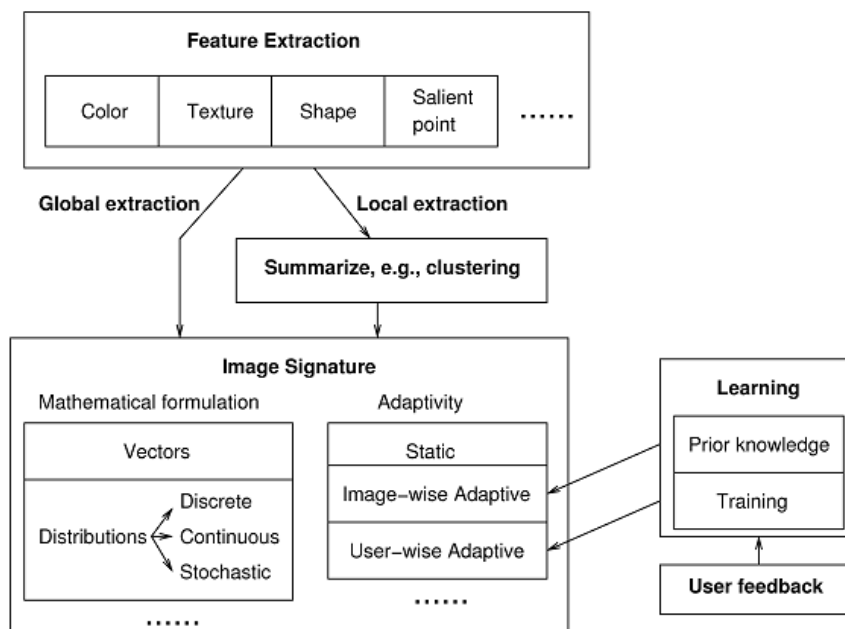


Figure 31. Les différentes formulations de la signature visuelle (Datta et al., 2008)

Des modèles stochastiques peuvent présenter une forme des signatures, ils sont construits à partir des densités continues en tenant compte de la dépendance spatiales des vecteurs de caractéristiques locales. En effet, l'utilisation de ces derniers modèles est limitée à cause de la complexité de leur construction et de leur aspect moins intuitives (Datta et al., 2008).

La deuxième formulation pour les signatures est une formulation adaptative. C'est une construction hybride qui exploite la classification des images pour choisir le mode de représentation de la signature. Les auteurs de (Datta et al., 2008) ont classé les signatures de cette formulation en trois catégories : statique, image wise adaptative et user wise adaptative.

Les signatures statiques sont générées uniformément pour toutes les images. Les signatures sont de la deuxième catégorie (image wise adaptative, voir la figure 31) si elles sont générées et varient suivant la classification des images, cette approche est aussi caractérisée par le terme d'approches sensible à la sémantique (semantic-sensitive) (Wang et al., 2001). les travaux de (Wang et al., 2001 ; Ng et al., 2005) sont des exemples de cette dernière catégorie des techniques de construction de la signature. Cette adaptation de la signature peut être également basée sur le bouclage de pertinence des utilisateurs (user wise adaptative).

Les auteurs de (Datta et al., 2008) déclarent que l'utilisation de cette approche adaptative doit faire attention pour que la hétérogénéité des caractéristiques n'affecte pas la performance de l'indexation et de la recherche. Plus de détails sur ces approches adaptatives de la construction des signatures visuelles peuvent être trouvées dans (Datta et al., 2008).

3.3 Mesure de similarité.

Pour répondre à une requête de l'utilisateur, le système de recherche d'image par le contenu calcule la distance entre le vecteur de caractéristiques de la requête et les vecteurs caractéristiques des images de la base. Cette distance est calculée en utilisant une mesure de similarité.

Il existe plusieurs mesures de similarités utilisées pour définir la distance entre deux descripteurs d'images. Une mesure de similarité est qualifiée comme étant une distance si elle dispose des quatre propriétés suivantes : non négativité, réflexivité, symétrie, et l'inégalité triangulaire.

Nous allons décrire dans cette section la distance euclidienne, la distance cosinus et la mesure de Kullback Leibler (KL) pour mesurer la distance entre les distributions de probabilités.

a) **Distance Euclidienne** : La distance Euclidienne est une distance géométrique dans un espace multidimensionnel. La distance euclidienne entre deux vecteurs i et j est calculée comme suit :

$$\text{Dis}(i, j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (12)$$

Où x_{ik} et x_{jk} représentent les k ème éléments des vecteur i et j respectivement.

b) **Distance Cosinus** : La similarité cosinus (ou mesure cosinus) permet de calculer la similarité entre deux vecteurs à n dimensions en déterminant le cosinus de l'angle entre eux. Pour deux vecteurs A et B , la distance est calculée en utilisant la formule suivante:

$$\text{Cosine}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (13)$$

- c) **Mesure de Kullback Leibler** : c'est une fonction utilisée pour calculer la similarité entre deux distributions de probabilité. Cette mesure calcule la divergence entre deux distributions p et q comme suit :

$$D(p, q) = p \log_2 \frac{p}{q} \quad (14)$$

Une forme symétrique de cette mesure est utilisée et calculée selon la formule suivante :

$$\text{KL}(p, q) = \frac{1}{2} [D(p, q) + D(q, p)] \quad (15)$$

La divergence de Jensen-Shannon peut être aussi utilisée et elle est calculer comme suit :

$$\text{JS}(p, q) = \frac{1}{2} \left[D \left(p, \frac{p+q}{2} \right), D \left(q, \frac{p+q}{2} \right) \right] \quad (16)$$

3.4 Indexation multimodale

L'indexation multimodale est basée sur l'utilisation de plusieurs modalités de caractéristiques combinées. Les informations textuelles associées aux images, sont actuellement la source utilisée pour cette indexation multimodale des images. La majorité des travaux qui abordent le sujet de l'utilisation des différentes modalités de données visent à appliquer cette combinaison pour la recherche en auto annotation des images, des travaux de ce domaine sont illustrés dans (Datta et al., 2008). La fusion des modalités pour l'indexation d'images par le contenu reste un domaine récent ouvert pour les nouvelles propositions de recherche. (LaCasia et al. 1998) constitue une des première contributions qui proposent l'utilisation de l'information textuelles avec l'information visuelle mais sans combinaison et cela dans le but d'une indexation d'images du web par le contenu.

L'une des approches de combinaison des caractéristiques multimodale est basée sur la transformation de la représentation d'une modalité en une forme similaire à celle

de l'autre modalité. Un exemple de cette transformation est le sac de mots visuels (bag of visual words). Cette unification des représentations permet d'étendre l'application des techniques de modélisation des différentes modalités à la combinaison aussi, de ce fait l'inspiration des modèles proposés pour les informations est possible et leurs application aussi sur des données de nature différentes (appliquer les modèles textuels sur des données visuels et vice versa et aussi sur la combinaison).

La combinaison des différentes modalités constitue aussi une source de recherche pour le domaine de l'indexation et la recherche des vidéos. Pour ce dernier, les informations textuelles et audio-visuelles peuvent être combinées.

La figure 32 montre une classification des modalités de caractéristiques.

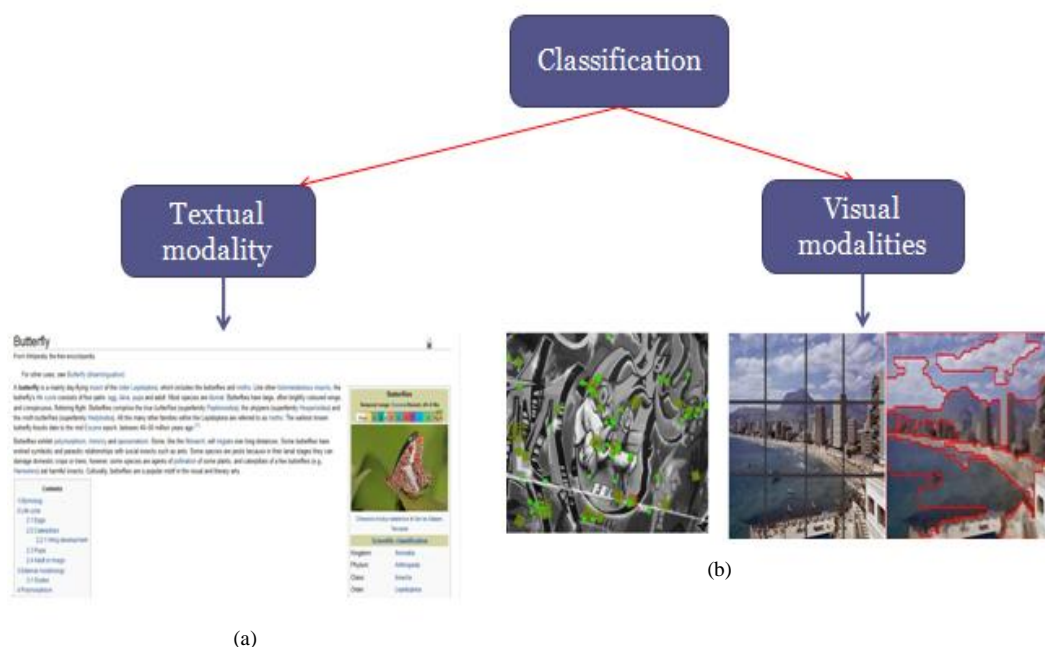


Figure 32. Classification des modalités de caractéristiques des images. (a) représente un exemple de modalité textuelles et (b) deux exemples de modalités visuelles

3.5 Recherche d'images

La recherche des images est la phase en ligne du schéma général des systèmes CBIRs (voir figure 3 du chapitre 1). Cette phase est initiée par l'utilisateur qui fournit une requête généralement sous la forme d'une image dans le but de récupérer les

images similaires à celle-ci à partir de la base d'image. Les requêtes peuvent prendre aussi d'autres formes. Dans le web, la requête est souvent sous la forme des **mots clés** (keywords), une autre forme des requêtes est le **texte libre** (free text) comme des petites phrases, des phrases complexes et des questions. La figure 33 montre des exemples de requêtes.

(Datta et al., 2008) présente aux utilisateurs la possibilité de raconter une histoire (petit paragraphe) de ce qu'ils veulent chercher dans le système. Une forme de requête souvent visées par la recherche d'image par le contenu est celle des requêtes images.

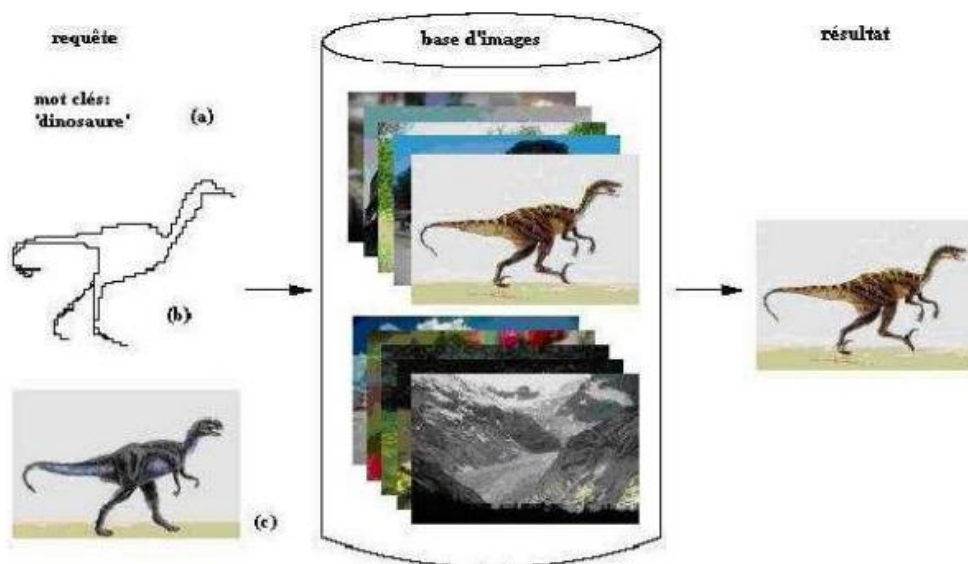


Figure 33. Exemples de formes de requêtes

Cette forme peut être considérée comme une forme standard (par défaut) des systèmes CBIRs. L'utilisateur fournit une image requête pour le système qui est chargé de répondre avec les images similaires à partir de la base. Les requêtes peuvent être aussi sous une forme de **graphique** dessinée manuellement par les utilisateurs ou bien générées avec l'ordinateur.

Toutes les modalités de requête mentionnées précédemment peuvent être utilisées toutes séparées ou bien combinées dans un même système.

L'intensité des utilisateurs a constitué un sujet de recherche important. La définition de cette intention implique l'étude des phases de recherche et de bouclage de pertinence. Les auteurs dans (Datta et al. 2008) ont proposé de diviser l'ensemble d'utilisateurs des systèmes CBIRs en trois catégories, navigateur, surfer et chercheur.

Navigateur : cette catégorie comporte des utilisateurs qui n'ont pas de buts finale claire. De ce fait les différentes sessions de recherche effectuées sont indépendante et explore des sujets multiples avec des requêtes incohérentes.

Surfer : dans cette catégorie, les utilisateurs ont une certaine clarté moyenne de leurs buts. Il commence leurs sessions comme navigateurs mais converge vers une amélioration de leurs précisions de recherche.

Chercheur : cette catégorie présente des utilisateurs très précis dans leurs demandes de recherche. Ils effectuent des recherches avec de très clairs buts, des sessions courtes ainsi que des requêtes très cohérentes.

3.6 Visualisation des résultats

La présentation des résultats de la recherche d'images dans un CBIR est l'une des caractéristiques importante de sa réputation. Les auteurs de (Datta et al. 2008) ont résumé les différents schèmes pouvant être suivis pour l'affichage des résultats, les cinq schèmes suivants ont été cités :

- **Tri par pertinence** :

C'est le schème le plus adopté pour l'affichage des résultats. Dans ce modèle, les images sont triées suivant une certaine mesure de pertinence par rapport à la requête.

- **Tri par chronologie** :

Dans cette catégorie, les images sont affichées suivant leur chronologie plutôt que la pertinence. C'est une option dans le système Google Picasa (Datta et al., 2008).

- **Tri hiérarchique** :

Ce schème est possible à adopter si les images sont accompagnées avec des métadonnées hiérarchiques (exemple des sujets de WordNET). Ces métadonnées

peuvent faciliter l'organisation hiérarchique de la visualisation et par cette effet son exploration. Ce mode est très conseillé pour les archives éducatives.

▪ **Tri par Clustering :**

Dans ce modèle, les techniques de clustering peuvent être appliquées pour afficher les résultats en forme de classes selon leurs métadonnées ou bien leurs caractéristiques visuelles. Ce modèle semble utile pour améliorer la performance des systèmes (Datta et al., 2008).

▪ **Tri composé :**

L'utilisation de plusieurs modes de visualisation reste toujours une option possible pour augmenter la personnalisation des systèmes et donc l'interactivité avec les utilisateurs.

La figure 34 montre un exemple de visualisation des résultats de la recherche du mot 'Tigers' par le moteur de recherche Google

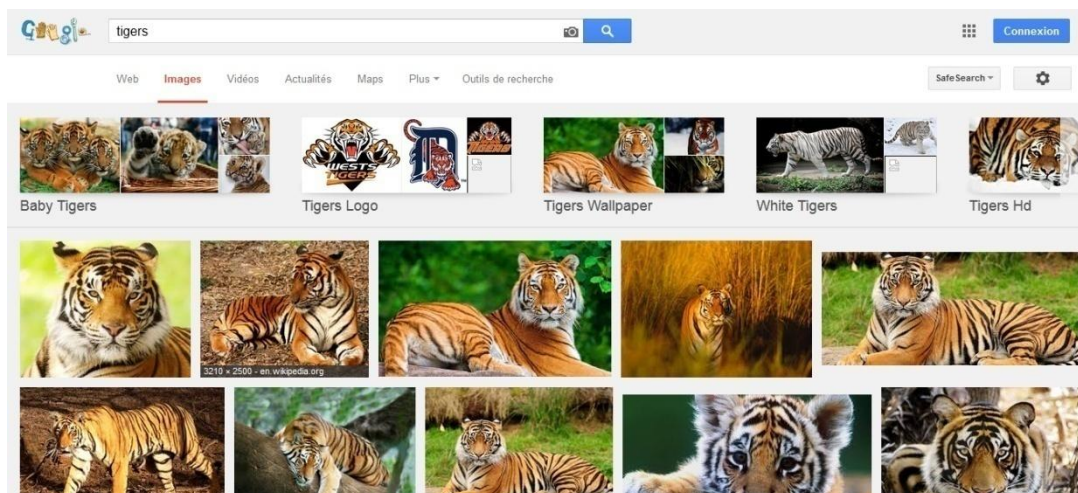


Figure 34. Exemple de visualisation des résultats de la recherche du mot 'Tiger' par le moteur de recherche Google (google image, Juin 2014)

Le module de visualisation des résultats fait partie de l'interface utilisateurs du système, de ce fait la conception de telle interface nécessite l'étude des besoins des utilisateurs, de leurs spécificités ainsi que les caractéristiques du matériel utilisé.

3.7. Bouclage de pertinence

La phase de bouclage de pertinence est une phase éventuelle des systèmes de recherche d'images par le contenu visuel. Cette phase implique l'intégration des utilisateurs dans une boucle de recherche composée de plusieurs sessions. Le fonctionnement général du système durant cette phase commence avec la fourniture du premier ensemble des résultats suite à la requête de l'utilisateur. Le système ensuite offre à l'utilisateur une possibilité d'évaluation positive (l'image résultat est semblable aux images cherchées par lui) ou négative des résultats retournés.

Le bouclage de pertinence offre la possibilité d'apprendre la sémantique de la requête et l'intention des utilisateurs. (Rui et al., 1998) présente la première contribution sur le BP pour les systèmes CBIRs. Les auteurs dans (Datta et al., 2008) ont produit une catégorisation des contributions qui traitent le BP suivants leurs nature:

3.7.1 Contributions basée sur l'apprentissage automatique

Cette catégorie contient des contributions qui utilisent l'apprentissage automatique pour modifier les paramètres des caractéristiques ou de la requête. Le problème lié à l'utilisation de cette approche est que l'ensemble de BP offre un nombre limité des données étiquetées et donc insuffisant pour l'apprentissage. Les auteurs de (Huang et al., 2002 ; Tian, 2000) ont proposés d'utiliser un algorithme D-EM (Discriminant EM) pour résoudre se problème en élargissant le cercle des images d'apprentissage à des images non étiquetées de la base. D'autres algorithmes ont été proposés pour réaliser un apprentissage performant, (MacArthur et al., 2000 ; Low and Chua ,1998) ont utilisés les arbres de décision. Les auteurs dans (Chen et al., 2003) ont proposé un apprentissage basé sur les SVM. L'apprentissage actif à été employé par (Rui and Huang, 2001 ; Tong and Chang, 2001), ce dernier utilise des SVM adaptés à l'apprentissage actif. Les auteurs dans (Goh et al., 2004 ; He et al., 2004) ont travaillés sur des extensions de l'apprentissage actif, le dernier est appelé un apprentissage basé sur les manifolds. Plus de détails sur le sujet est dans (Zhou and Huang, 2003).

3.7.2 Contributions basée sur les spécifications du BP

Cette catégorie d'approches contient des contributions qui traitent le BP avec un nouveau paradigme de spécification des requêtes. Un BP sémantique à été proposé par (Yang et al., 2005), cette contribution demande des utilisateurs de caractériser les images par des étiquettes textuelles. (Hoi and Lyu, 2004) propose la réutilisation des anciens BP sauvegardés pour construire un requête composé de plusieurs images. Les auteurs de (Wu et al., 2004) utilisent un score de pertinence de BP pour les différentes sessions.

3.7.3 Contribution basées sur l'utilisateur

Dans cette catégorie d'approches, les contributions concentrent sur le point de vue des utilisateurs pour le BP plutôt que d'intéresser à la capacité des systèmes à apprendre à partir de ses BP. Le travail de (Hoi and Lyu, 2004) peut être classé dans cette catégorie aussi. (Jaimes et al., 2004 ; Nagamine et al., 2004) proposent d'offrir des clés aux utilisateurs pour reformuler leurs requêtes. Les auteurs de (Fang and Geman, 2005 ; Fang et al., 2005) ont appliqués un Framework bayésien pour apprendre des distributions d'images de la base qui représentent l'image mentale des utilisateurs et ensuite utilise ces distributions pour la recherche. (Lin et al., 2005) ont présentés l'idée de l'utilisation d'un manifold sémantique pour capturer les préférences des utilisateurs.

3.7.4 Contributions probabilistes

Cette catégorie d'approches est basée sur l'utilisation des modèles probabilistes pour modéliser l'intention d'utilisateur. Un exemple des contributions de cette catégorie est les travaux de (Cox et al., 2000) qui utiliser une distribution probabiliste pour modéliser la connaissance incertaine du but de l'utilisateur. (Su et al., 2003) utilise un classifieur bayésien pour le tri des résultats de BP.

3.7.5 Contributions basées régions

Cette catégorie contient des approches liées à la recherche basée sur l'utilisation des régions des images. (Carson et al., 2002 ; Wang et al., 2001 ; Ko and Byun, 2002 and Jing et al., 2004) ont introduit le facteur d'importance de la région dans le processus de BP.

Plus de détails sur d'autres contributions de BP sont inclus dans (Datta et al., 2008; Liu et al., 2007).

3.8 Réduction du fossé sémantique

Nous avons parlés dans la section d'introduction du problème de fossé sémantique qui consiste en une absence de relation entre les informations de bas niveau extraites à partir des images et l'interprétation sémantique de ce contenu par l'utilisateur dans une situation précise. Plusieurs contributions traitant la réduction du fossé sémantique ont été proposées dès sa définition par (Smeulders et al., 2000). Les auteurs de (Liu et al., 2007) ont donné la catégorisation suivante des différentes approches de réduction du fossé sémantique, la figure 35 résume ces approches.

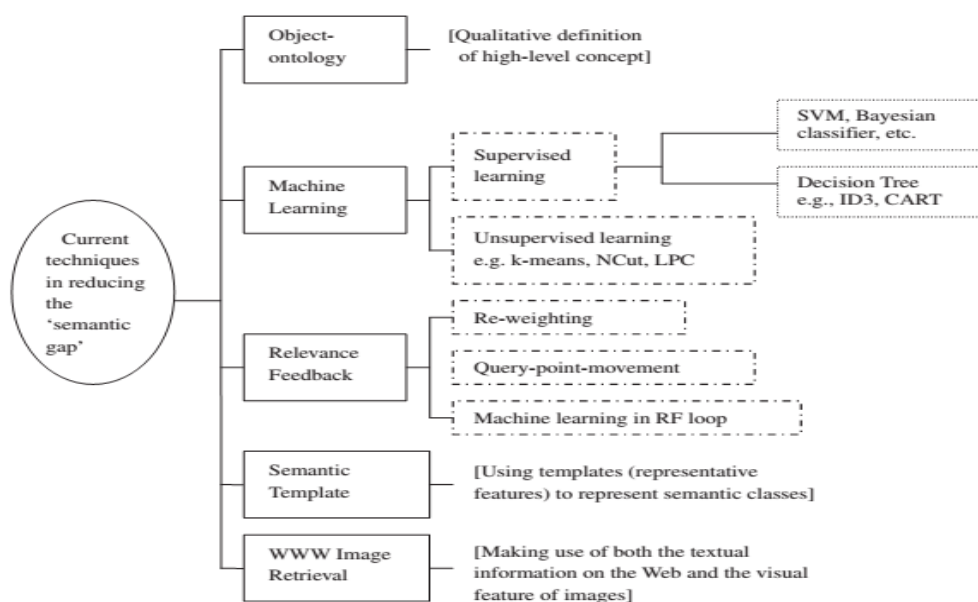


Figure 35. Résumé des approches de la réduction du gap sémantique (Liu et al., 2007)

3.8.1 Les ontologies d'objets

Les ontologies peuvent être utilisées pour définir des caractéristiques de niveau moyen de sémantique. L'utilisation des intervalles sur l'ensemble des valeurs des caractéristiques visuelles permet d'associer une étiquette sémantique qui caractérise l'intensité des couleurs (claire, moyenne, foncée) ou bien aussi d'autres étiquettes sémantiques posées sur les régions (uniforme, en haut, avec de telle couleur). Le

vocabulaire formé par ces étiquêtes est appelé ontologie d'objets avec une définition qualitatives de haut niveau pour les requêtes. La figure 36 montre une ontologie construite et utilisé par (Mezaris et al, 2003), elle contient la description de la couleur moyenne, la position spatiale, le volume et la forme des objets.

Pour construire des ontologies d'objets, la quantification des couleurs et texture est nécessaire. Les auteurs dans (Conway, 1992 ; Berk et al., 1982, Stanchev et al., 2003) ont présenté les contributions qui s'intéressent de la description sémantique des valeurs numériques de la couleur.

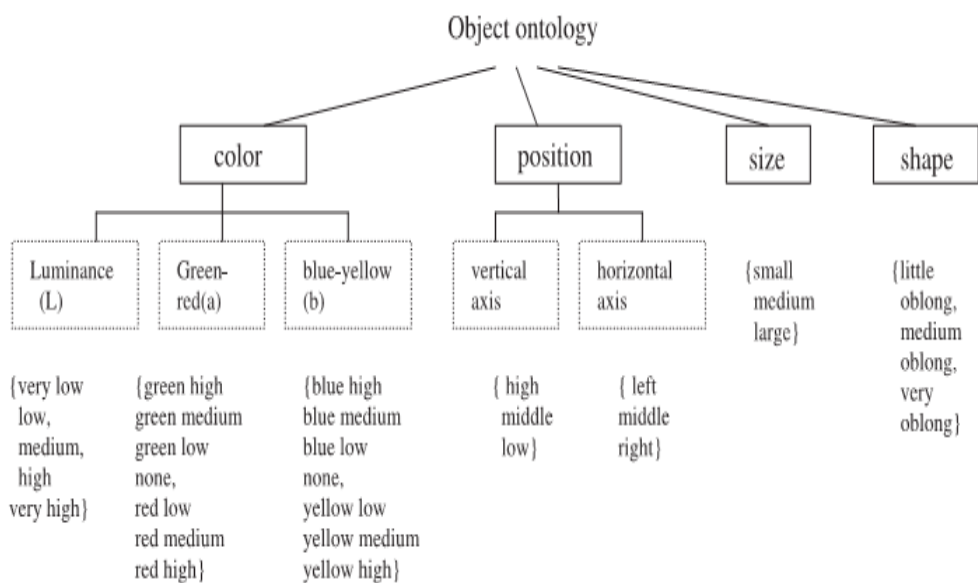


Figure 36. Ontologie d'objets (Ying et al., 2007)

Ces contributions ont proposés l'utilisation d'un CNS (Color Naming System) qui utilise des noms naturels pour décrire les valeurs et les caractéristiques des couleurs. La contribution de (Datta et al., 2008) a étudié les couleurs dominantes, d'autres travaux comme (Datta et al., 2008) ont contribué à la proposition des TNS (Texture Naming Systems) pour décrire sémantiquement la texture. Plus de détails sur des contributions similaires peuvent être trouvés dans (Liu et al., 2007).

3.8.2 L'apprentissage automatique

Les techniques d'apprentissage automatique ont été utilisées pour améliorer la performance des CBIRs. L'apprentissage automatique permet à la machine d'apprendre la classification d'images, et des caractéristiques ainsi que la prédiction des besoins d'utilisateurs. Plus de détails sur les techniques d'apprentissage automatiques peuvent être trouvés dans le chapitre 3.

3.8.3 Le bouclage de pertinence

Le bouclage de pertinence a présenté une approche pour apprendre l'intention des utilisateurs et donc améliorer la satisfaction de leurs besoins. L'analyse de l'intention des utilisateurs permet d'ajouter un niveau sémantique à leur recherche et de rapprocher de leurs descriptions visées. Plus de détails sur le bouclage de pertinence sont trouvés dans la section 3.7 de ce chapitre.

3.8.4 Les modèles sémantiques

Les modèles sémantiques sont des caractéristiques représentatives d'un concept calculées à partir d'une collection d'images exemples (Smith and Li, 1998 ; Zhuang Y et al., 1999). Les auteurs dans (Chang et al., 1998) ont utilisés des icône et des images exemples et à partir lesquelles l'utilisateur construit un SVT (semantic visual template), ce modèle sémantique visuel est utilisé comme requête, le système à été proposé pour la recherche de la vidéo. La contribution de (Zhuang et al., 1999) utilise le bouclage de pertinence pour définir automatiquement des modèles sémantique pour les concepts textuels fournies par l'utilisateur comme requête initiale de sa recherche. Plus de détails sur cette approche de modèles sémantique peuvent être trouvés dans (Liu et al, 2007).

3.8.5 Recherche d'image dans le web

La recherche d'images sur le web a été considérée par (Liu et al., 2007) comme étant une direction des techniques qui abordent le sujet de la recherche sémantique de haut niveau et donc la réduction du fossé sémantique. L'avantage de cette direction de recherche est la présence des données complémentaires qui peuvent constituer une source d'informations de haut niveau. La hiérarchie des URL, les différentes balises HTML ainsi que le texte associé aux images sont tous la source de ces informations

utile à l'indexation et la recherche d'images de haut niveau. Plus de détails sur les travaux d'indexation et de recherche d'images sur le web peuvent être trouvés dans (Liu et al, 2007).

3.9 Critères d'évaluation

Le domaine de l'indexation et recherche d'image par le contenu utilise les métriques d'évaluation de la recherche d'information. Les deux mesures les plus utilisées sont le rappel et la précision. Le rappel et la précision sont calculés selon les formules suivantes :

$$\text{Precision} = \frac{A}{B} \quad (17)$$

$$\text{Recall} = \frac{A}{C} \quad (18)$$

Avec

A: Nombre d'images pertinentes retournées par le système.

B: Nombre total des images résultats.

C: Nombre d'images pertinentes dans la base.

Les systèmes sont souvent évalués en calculant le MAP (Mean Average Precision), cette mesure est calculée sur plusieurs requêtes et représente la moyenne arithmétique des différentes moyennes de précision de chaque requête. Le MAP permet d'évaluer la précision des réponses du système en tenant compte la position de la réponse pertinente dans l'ensemble des réponses.

Plus de détails sur l'évaluation de la recherche d'images peuvent être trouvés dans (Liu et al., 2007; Datta et al., 2008).

3.10 Conclusion

Nous avons donné dans ce chapitre une vue sur les concepts avancés du domaine de l'indexation et la recherche d'images. C'est un domaine intéressant qui n'a pas cessé à se développer à partir des années 80s. Les travaux actuels de recherche dans le domaine sont devenu de plus en plus spécialisés, le développement d'un système CBIR complet et fonctionnel n'est plus le but, mais plutôt la contribution à l'étude et la recherche des solutions d'une des directions de la recherche dans le domaine.

L'étude de l'intention des utilisateurs est la direction la plus importante qui nécessite une étude approfondie. Dans le cadre de cette direction, la recherche doit concentrer sur la construction des modèles de compréhension des utilisateurs pour personnaliser sa recherche et la sélection des caractéristiques d'images les plus pertinentes à sa situation. Le profil utilisateur semble très utile comme notion élémentaire de recherche à aborder dans le but de l'intégrer dans les futurs travaux. Dans ce point, nous pensons que l'inspiration du domaine de e-Learning est possible et elle peut constituer une base des ces perspectives de la recherche sémantique.

Nous allons présenter dans le chapitre suivant une famille de modèles statistiques et probabilistes qui contient des techniques utilisées dans la préparation des contributions de la thèse. Cette famille de modèle est appelé les modèles de sujets (ou les modèles à concepts, en anglais topic models).

CHAPITRE – IV

MODELES DE SUJETS

5.1. Introduction

Ce chapitre contient une présentation d'un ensemble de modèles à concepts. Cette famille comporte plusieurs modèles dont le LSA (Latent Semantic Analysis) (Deerwester et al., 1990), le pLSA (probabilistic LSA) (Hofmann, 2001), le LDA (Latent Dirichlet Allocation) (Blei et al., 2003), le CTM (Correlated Topic Model) (Blei and Lafferty, 2007), et le PAM (Pachinko Allocation Model) (Li and McCallum, 2006). Le dernier modèle PAM est celui utilisé dans le cadre de cette thèse.

Les modèles à concepts ont été proposés à l'origine dans le domaine de traitement automatique de textes, leur but principale été de fournir un moyen de modélisation et de recherche de textes dans des grands corpus. Ils peuvent être classés en deux catégories : modèles à concepts déterministes et modèles à concepts probabilistes.

Tous ces modèles utilisent au départ une matrice de cooccurrence termes-documents. Etant donné que la base contient M documents d_1, \dots, d_M , le contenu des documents est issu d'un vocabulaire composé de N mots w_1, \dots, w_N . Les lignes de la matrice contiennent les mots de vocabulaire, et les colonnes contiennent les documents de la base. Chaque élément de la matrice représente la fréquence du mot de la ligne dans le document de la colonne, voir figure 37.

$$\begin{pmatrix} f(w_1, d_1) & \cdots & f(w_1, d_M) \\ \vdots & \ddots & \vdots \\ f(w_N, d_1) & \cdots & f(w_N, d_M) \end{pmatrix}$$

Figure 37. Matrice termes-documents, chaque élément de la matrice spécifie la fréquence d'apparition du mot de la ligne dans le document de la colonne (Steyvers and Griffiths, 2007)

Les modèles de concepts tentent d'extraire à partir de la matrice de cooccurrence l'espace latent qui contient les concepts cachés responsable de la présence (génération pour le cas probabiliste) des mots dans les documents, ces concepts représentent les sujets (ou thèmes) inclus dans les documents d'où le nom de modèle

de sujets utilisé aussi pour la description de ces modèles. L'espace latent est utilisé pour l'indexation et la recherche des documents au lieu d'utiliser la matrice initiale. Cette utilisation des espaces latents présente les avantages suivants :

La réduction de l'espace de la recherche car le nombre de concepts est très réduit par rapport au nombre des mots de vocabulaire. Cette réduction permet une recherche plus rapide dans la base.

L'utilisation des concepts de l'espace latent permet de résoudre des problèmes classiques de la recherche d'informations qui sont: la polysémie, la synonymie et l'absence. En effet, les mots sont regroupés sous les différents concepts latents en se basant sur leurs cooccurrences mutuelles avec d'autres mots, c'est à dire deux mots sont associés au même concept s'ils co-occurrent avec les mêmes mots et donc ayant le même contexte. La figure 38 montre un exemple de quatre concepts latents extraits à partir du corpus TASA (Steyvers and Griffiths, 2007).

Le premier modèle à concepts que nous allons présenter est le modèle LSA (Latent Semantic Analysis), il a été le premier à proposer pour l'extraction des concepts latents. Nous allons présenter par la suite sa version probabiliste, le pLSA (probabilistic Latent Semantic Analysis), qui a introduit la deuxième catégorie probabiliste des modèles à concepts.

Les parties qui suivent dans ce chapitre présentent d'autres modèles de concepts y compris le modèle PAM (Pachinko Allocation Model) utilisé dans le cadre de la thèse.

5.2. Modèles à Concepts déterministes

Cette catégorie contient les modèles à concepts qui utilisent des techniques déterministes pour extraire les concepts à partir des documents. L'analyse de la sémantique latente est un exemple de ces techniques que nous allons présenter dans la section suivante.

5.2.1 Latent Semantic Analysis (LSA)

L'analyse de la sémantique latente LSA a été proposée par (Deerwester et al., 1990) comme une technique statistique pour l'analyse des grands corpus de documents.

Topic 247		Topic 5		Topic 43		Topic 56	
word	prob.	word	prob.	word	prob.	word	prob.
DRUGS	.069	RED	.202	MIND	.081	DOCTOR	.074
DRUG	.060	BLUE	.099	THOUGHT	.066	DR.	.063
MEDICINE	.027	GREEN	.096	REMEMBER	.064	PATIENT	.061
EFFECTS	.026	YELLOW	.073	MEMORY	.037	HOSPITAL	.049
BODY	.023	WHITE	.048	THINKING	.030	CARE	.046
MEDICINES	.019	COLOR	.048	PROFESSOR	.028	MEDICAL	.042
PAIN	.016	BRIGHT	.030	FELT	.025	NURSE	.031
PERSON	.016	COLORS	.029	REMEMBERED	.022	PATIENTS	.029
MARIJUANA	.014	ORANGE	.027	THOUGHTS	.020	DOCTORS	.028
LABEL	.012	BROWN	.027	FORGOTTEN	.020	HEALTH	.025
ALCOHOL	.012	PINK	.017	MOMENT	.020	MEDICINE	.017
DANGEROUS	.011	LOOK	.017	THINK	.019	NURSING	.017
ABUSE	.009	BLACK	.016	THING	.016	DENTAL	.015
EFFECT	.009	PURPLE	.015	WONDER	.014	NURSES	.013
KNOWN	.008	CROSS	.011	FORGET	.012	PHYSICIAN	.012
PILLS	.008	COLORED	.009	RECALL	.012	HOSPITALS	.011

Figure 38. Exemple de quatre concepts latent extraits à partir du corpus TASA (Steyvers and Griffiths, 2007)

Cette technique permet l'extraction de la relation entre les mots à base de leurs contextes. Etant donné un vocabulaire de mots de taille N et un corpus de document de taille M. La LSA utilise une décomposition en valeurs singulière SVD pour écrire la matrice de cooccurrence A de taille N×M comme produit des trois matrices W, Σ et Dt, comme dans la formule 2.1. La SVD est illustrée dans la figure 39 :

$$A = W\Sigma D^t \tag{19}$$

La matrice orthonormale W de taille N×r tel que r est le rang de la matrice A, contient les vecteurs singuliers gauches de A. La matrice D est une matrice orthonormale de taille M×r contenant les vecteurs singuliers droits de A. Σ est une matrice diagonale de taille contenant les valeurs singulières de A.

Si on suppose que les valeurs singulière de la matrice Σ sont ordonnées, alors une approximation $\hat{A} = \tilde{W} \tilde{\Sigma} \tilde{D}^t$ peut être calculée, les k plus grandes valeurs singulières sont gardées et les autres seront remet à zéro et les lignes et la colonne nuls de la matrice Σ peuvent être éliminées ainsi que les lignes et les colonnes correspondantes dans les matrices W et D.

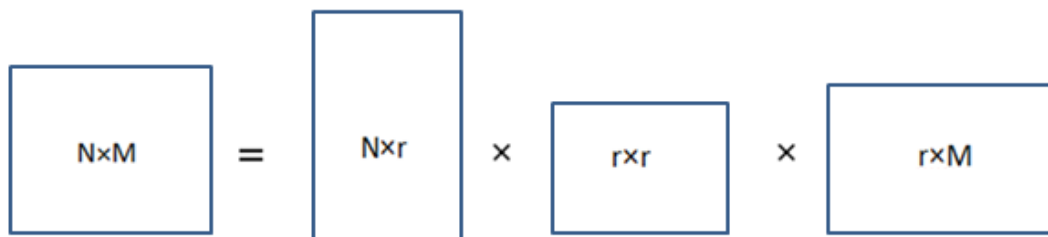


Figure 39. Factorisation de la matrice pour le modèle LSA

Les nouvelles dimensions des matrices \tilde{W} , $\tilde{\Sigma}$ et \tilde{D} sont respectivement $N \times K$, $M \times k$, et $k \times k$.

Pour comparer un nouveau document requête d_q avec les documents de la base, elle doit être transformée en une représentation adéquate \tilde{d}_q en utilisant la formule suivante :

$$\tilde{d}_q = d_q^t \tilde{W} \tilde{\Sigma}^{-1} \quad (20)$$

Ensuite, une mesure de similarité peut être appliquée pour calculer la similarité entre cette requête et les documents de la base. Un exemple des mesures est le cosinus. Une représentation peut être aussi calculée pour les mots offrant la possibilité de comparer des mots de la base avec des mots requête. La formule utilisée est la suivante :

$$w_q = \tilde{\Sigma}^{-1} \tilde{D}^t w_q^t \quad (21)$$

Une variante probabiliste du LSA a été proposée par (Hoffmann, 2001), c'est le modèle pLSA, le sujet de la section suivante. Plus de détails concernant LSA sont trouvés dans (Tollari, 2006; Hörster, 2009).

5.3. Les modèles à concepts probabilistes

La liste des modèles à concepts que nous allons présenter dans le reste de ce chapitre contient des modèles à concepts probabiliste qui diffèrent dans leur nature du modèle LSA présenté dans la section précédente. Nous allons présenter dans cette section les modèles pLSA, LDA, CTM et PAM. Ces modèles représentent les documents par un

modèle de mélange de concepts probabiliste. Les modèles à concepts probabilistes sont basés sur l'idée que les documents peuvent être modélisés par un mélange de concepts ou les qui représentent des distributions sur les mots. Les modèles à concepts probabilistes sont des modèles génératifs spécifiant des procédures probabilistes permettant de générer les documents, en d'autre terme les règles probabilistes décrivant comment les mots peuvent être générés à la base des variables latentes (concepts). Des techniques d'inférence statistiques peuvent être utilisées pour inverser le processus et inférer les concepts qui ont été responsable de la génération des mots.

Comme avec le modèle LSA, les autres modèles à concepts probabilistes utilisent comme entrée la matrice de cooccurrence termes-documents. Considérons une base de données contenant M documents d_i définis sur un vocabulaire de taille N . chaque document d_i est représenté par une liste de mots $w_i = \{w_1, w_2, \dots, w_{N_i}\}$ ou w_n représente le $n^{\text{ème}}$ mot dans la liste. Une liste des variables latentes $z_i = \{z_1, z_2, \dots, z_K\}$ responsables de la génération des mots dans le corpus.

5.3.1 Probabilistic Latent Semantic Analysis

Le modèle pLSA est la variante probabiliste du modèle LSA. Le processus génératif du pLSA pour un document d_i est le suivant :

- Choisir un document d_i avec une probabilité à priori $P(d_i)$.
- Pour chaque mot de la liste des mots du document d_i :
 - Choisir un concept latent z_n avec la probabilité $P(z_n|d_i)$.
 - Générer un mot w_n avec une probabilité $P(w_n|z_n)$.

La figure 40 donne le modèle graphique du PLSA. La probabilité du mot w_i dans le document d_i est présentée dans la formule suivante :

$$P(w_j, d_i) = P(d_i) \sum_{k=1}^K P(w_j|z_j = k)P(z_j = k|d_i) \quad (22)$$

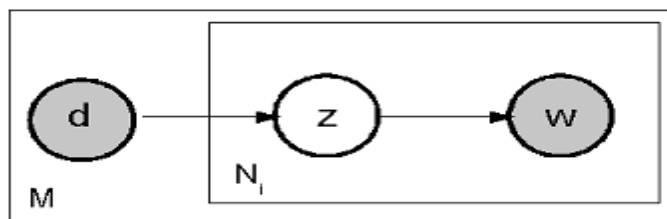


Figure 40. La représentation graphique du modèle pLSA (Hörster, 2009)

5.3.2 Latent Dirichlet Allocation

Le modèle LDA est un modèle à concepts générative similaire au pLSA mais différents dans ses propriétés statistiques. Il spécifie aussi un modèle de mélange sur les concepts dont chaque concept est caractérisé par une distribution sur les mots, et chaque occurrence de mots dans un document est associée à un concept latent. Le LDA représente le mélange de concepts comme variable latente et pose un paramètre à priori de dirichlet sur celle-ci, la figure 41 montre le modèle graphique du LDA.

Le processus de génération des documents par le modèle LDA comporte les étapes suivantes :

- Choisir une variable aléatoire à K-dimensions $\theta_i \sim Dir(\alpha)$, K est le nombre de concepts dans la base.
- Pour chaque mot de la liste N_i du document i
 - Choisir un concept $z_n \sim Multinomial(\theta_i)$
 - Générer la valeur w_n du $n^{\text{ème}}$ mot avec la probabilité conditionnelle $P(w_n | z_n, \beta)$.

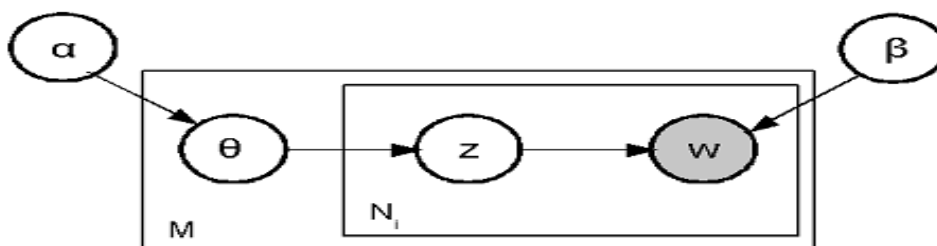


Figure 41. Modèle graphique du LDA (Hörster, 2009)

La probabilité d'un document w_i est donnée avec la formule suivante :

$$P(w_i|\alpha, \beta) = \int P(\theta_i|\alpha) \prod_{j=1}^{N_i} (\sum_{k=1}^K P(z_j = k|\theta) P(w_j|z_j = k, \beta)) d\theta \quad (23)$$

5.3.3 Correlated Topics Model (CTM)

Le CTM est un modèle à concepts qui se diffère des autres modèles par ses propriétés statistiques. Le CTM choisit les proportions des concepts à partir d'une distribution normale logistiqu, la représentation graphique du modèle est illustrée dans la figure 42. Le processus génératif des documents pour le modèle CTM comporte les étapes suivantes:

- Choisir $\Pi_i | \{\mu, \Sigma\} \sim N(\mu, \Sigma)$, avec μ come vecteur de moyennes à K dimensions et Σ une matrice de cooccurrence de taille $K \times K$, Π une variable de mélange de concepts.
- Pour chacun des N_i mots dans le document i :
 - Choisir assignement des concepts à partir d'une loi multinomiale $f(\Pi_i)$.
 - Choisir un mot w_n à partir la probabilité multinomiale conditionnée avec les concepts z_n comme $P(w_n|z_n, \beta)$.

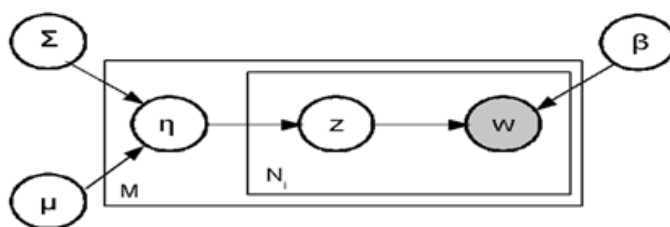


Figure 42. La représentation graphique du modèle CTM (Hörster, 2009)

L'algorithme d'espérance maximisation peut être utilisé pour inférer les concepts du modèle CTM. Plus de détails sur les paramètres du modèle ainsi que l'inférence peuvent être trouvés dans (Steyvers and Griffiths, 2007).

5.3.4 Pachinko Allocation Model

Le modèle PAM (Li and McCallum, 2006) diffère des autres modèles probabilistes par l'extraction de plusieurs niveaux des sujets. Ils s'intéressent non seulement par la capture de la corrélation entre les mots mais aussi entre les sujets eux même via l'extraction d'une couche supplémentaire des super-sujets. La figure 43 montre la représentation graphique du modèle PAM.

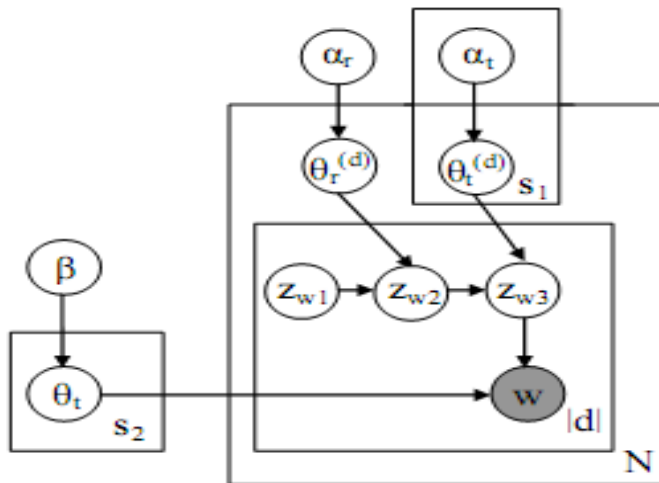


Figure 43. La représentation graphique du modèle PAM (Li and McCallum, 2006)

Le processus génératif du modèle peut être décrit comme suit :

- Choisir $\theta_{t_1}^{(d)}, \theta_{t_2}^{(d)}, \dots, \theta_{t_s}^{(d)}$ à partir de $g_1(\alpha_1), g_2(\alpha_2), \dots, g_s(\alpha_s)$, avec $\theta_{t_i}^{(d)}$ comme distribution multinomiale de sujet t_i sur ces fils.
- Pour chaque mot w dans le document
 - Choisir le chemin de sujet z_w of d'une longueur $L_w: \langle z_{w1}, z_{w2}, \dots, z_{wL_w} \rangle$. z_{w1} est la racine et z_{w2} à z_{wL_w} sont les neuds des sujets dans l'ensemble de sujets T . z_{wi} est le fils of $z_{w(i-1)}$ et choisir selon une distribution multinomiale $\theta_{z_{w(i-1)}}^{(d)}$.
 - Choisir le mot w à partir de $\theta_{z_{wL_w}}^{(d)}$.

Notre travail de cette thèse été basé sur l'étude de l'impact de l'utilisation du modèle PAM pour l'indexation et la recherche d'images par le contenu. Cette étude est

absente des travaux de recherches traitant l'application des modèles de sujets pour le traitement d'images.

5.4. Extraction des concepts

Les techniques employées pour l'extraction des concepts peuvent être classées en catégories : déterministe et probabiliste. LSA utilise une technique algébrique pour l'extraction des concepts, elle les représente sous forme d'espaces engendrés par l'application d'une décomposition en valeurs singulières (SVD).

Les modèles à concepts probabilistes utilisent des techniques d'inférences statistiques pour inférer les sujets. Les deux techniques les plus utilisées sont : l'expectation maximization (EM) et le Gibbs Sampling.

5.4.1 SVD

Le procédé d'algèbre linéaire de décomposition en valeurs singulières (ou SVD, de l'anglais : singular value decomposition) d'une matrice est un outil important de factorisation des matrices rectangulaires réelles ou complexes. Le théorème spectral énonce qu'une matrice normale peut être diagonalisée par une base orthonormée de vecteurs propres. On peut voir la décomposition en valeurs singulières comme une généralisation du théorème spectral à des matrices arbitraires, qui ne sont pas nécessairement carrées.

5.4.2 Expectation Maximization

Le problème de la maximisation de la vraisemblance consiste à trouver les paramètres qui la maximisent. Soit $p(x|\theta)$ une fonction de densité qui dépend des paramètres θ . L'ensemble $X = x_1, x_2, \dots, x_n$ un ensemble de N données extraites à partir de cette distribution. La densité de l'échantillon peut être écrite sous la forme :

$$p(X|\theta) = \prod_{i=1}^N p(x_i|\theta) = V(\theta|X) \quad (24)$$

Cette fonction est une fonction de vraisemblance qui exprime la vraisemblance des paramètres selon les données. Le log-vraisemblance est souvent utilisé car il est plus facile à calculer, et il s'exprime de la façon suivante :

$$L(X|\theta) = \ln(V(\theta|X)) \quad (25)$$

Le problème revient alors à un problème d'optimisation de ces paramètres par rapport à la distribution.

$$\theta^{\text{optimal}} = \text{Max } L(X|\theta) \quad (26)$$

L'algorithme EM et les algorithmes itératifs qui lui ressemblent sont très utiles également pour résoudre les problèmes des échantillons avec des données manquantes. La figure 44 montre l'algorithme EM.

L'algorithme d'espérance-maximisation comporte :

- une étape d'évaluation de l'espérance (E), où l'espérance de la vraisemblance est calculée en tenant compte des dernières variables observées,
- une étape de maximisation (M), où le maximum de vraisemblance des paramètres est estimé en maximisant la vraisemblance trouvée à l'étape E.

Les paramètres trouvés à l'étape M vont être ensuite utilisés comme point de départ d'une nouvelle phase d'évaluation de l'espérance, et les itérations continuent jusqu'à la convergence à un maximum local.

Plus de détails concernant l'algorithme d'expectation maximisation et les techniques d'inférence peuvent être trouvés dans les références suivantes (Tollari, 2006; Steyvers and Griffiths, 2007; Uribe, 2007; Hörster, 2009).

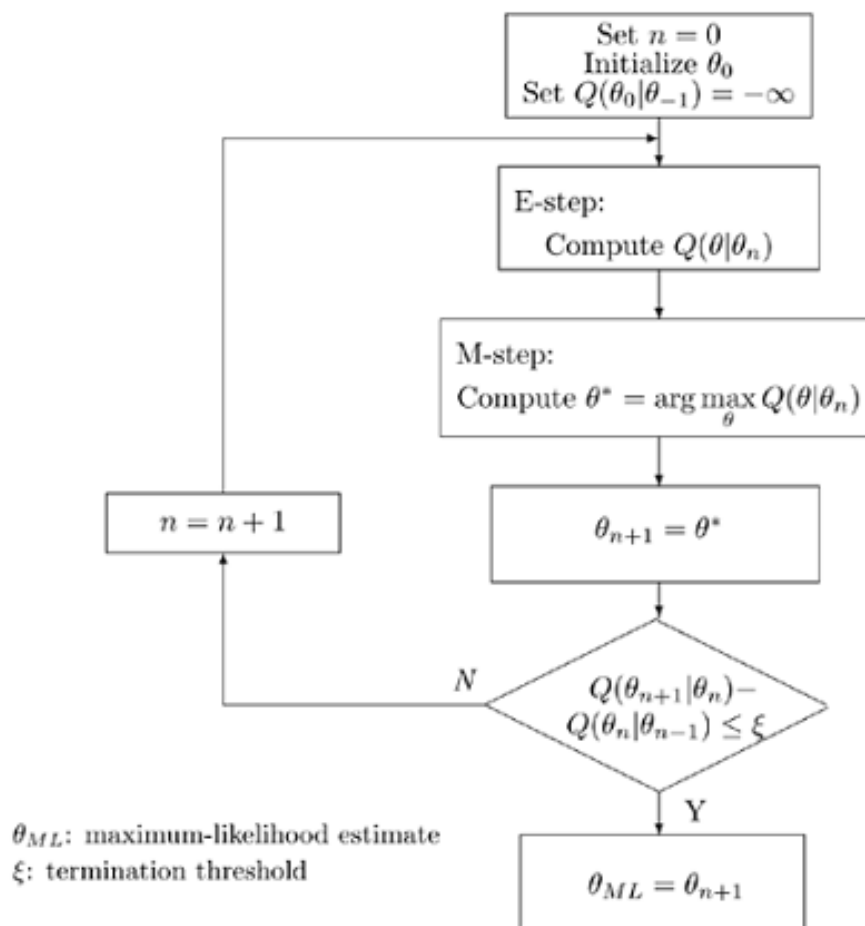


Figure 44. Algorithme EM (Urli, 2007)

5.4.3 Gibbs Sampling

Le gibbs samplig (Geman and Geman, 1984) est une méthode MCMC (Markov Chain Monte Carlo). Cette méthode à été introduit pour la première fois dans le domaine de la restauration des images.

Étant donnée une distribution de probabilité π sur un univers Ω , cet algorithme définit une chaîne de Markov dont la distribution stationnaire est π .

MCMC contient un ensemble de techniques d'approximation itératives. Gibbs sampling est une forme spéciale des MCMC utilisée pour simuler des distributions à haute dimension en sous ensembles des variables à dimension réduite, chaque sous ensemble est conditionné par les valeurs de tous les autres.

Les auteurs de (Steyvers and Griffiths, 2007) ont mentionnés que les variables d'intérêt à estimer pour les modèles à concepts probabilistes (modèles de sujets) sont ϕ et Θ qui représentent respectivement la distribution des sujets-mots et la distribution de chaque document. L'estimation directe de ces paramètres en utilisant l'algorithme EM souffre du problème de maximum local. L'estimation indirecte en utilisant la distribution à posteriori qui peut être réalisée par le gibbs sampling, améliore l'estimation de ces variables (Steyvers and Griffiths, 2007).

Le gibbs samling pour l'extraction des sujets a été présenté par (Steyvers and Griffiths, 2007). La procédure prend en considération à chaque itération, et estime la probabilité d'assignement du mot courant à chaque sujet conditionnée par l'assignement du sujet à tous les autres mots. La formule suivante illustre le calcul :

$$P(z_i = j | \mathbf{z}_{-i}, w_i, d_i, \cdot) \propto \frac{C_{w_i j}^{WT} + \beta}{\sum_{w=1}^W C_{w j}^{WT} + W\beta} \frac{C_{d_i j}^{DT} + \alpha}{\sum_{t=1}^T C_{d_i t}^{DT} + T\alpha} \quad (27)$$

Avec C^{WT} et C^{DT} des matrices de fréquences de dimensions respectives $W \times T$ et $D \times T$. La matrice $C_{w_j}^{WT}$ contient (28) ore de fois le mot w a été assigner au sujet j en exclu l'instance courante. $C_{d_j}^{DT}$ Contient le nombre de fois le sujet j a été assigné à un mot pris du document d en exclu l'instance courante.

Les auteurs de (Steyvers and Griffiths, 2007) montrent également comment calculer ϕ et Θ à partir des matrices par les deux formules suivantes :

$$\phi_i^{(j)} = \frac{C_{ij}^{WT} + \beta}{\sum_{k=1}^W C_{kj}^{WT} + W\beta} \quad (28)$$

$$\theta_j^{(d)} = \frac{C_{dj}^{DT} + \alpha}{\sum_{k=1}^T C_{dk}^{DT} + T\alpha} \quad (29)$$

Plus de détails sur l'extraction des sujets peuvent être trouvés dans (Steyvers and Griffiths, 2007; Hörster , 2009).

5.5. Application des modèles à concepts en traitement d'images

L'analogie entre l'image et le texte a permis l'inspiration des techniques et modèles de modélisation textuelle pour les appliquer avec les problèmes du traitement d'image. Comme le texte qui est un ensemble de mots, l'image peut être aussi considérée comme ensemble de régions, de points, de parties.... Cette approche permet de définir des mots visuels et les relier avec les images et donc appliquer sur eux les différents modèles textuels. Nous allons référer dans cette partie quelques contributions du domaine de traitement d'images qui ont utilisés des modèles textuels et plus précisément des modèles à concepts.

Le premier modèle de sujet apparue est le modèle LSA. Il été également le premier modèle de sujet à appliquer pour le traitement d'images. LSA à été appliqué par (La Cascia et al., 1998) pour implémenter un moteur de recherche d'image sur le web, il à été appliqué pour indexer les textes associés aux images. (Pham et al., 2007 ;Hare and Lewis, 2005; Chandrika et al.,2010 ; Fergus et al., 2005) ont utilisés LSA ou pLSA pour l'indexation d'images par le contenu, ces deux références ont étudiés l'impact de l'utilisation de ces deux modèles avec une combinaison de caractéristiques visuelles et textuelles. pLSA à été également employé par (Quelhas et al., 2005) pour la modélisation des scènes en utilisant les caractéristiques locales. (Fei Fei et al., 2007) ont utilisés le CTM pour la catégorisation des scènes naturelles. Un algorithme basé sur les modèles génératives a été utilisé par (Winn and Jojic, 2006) pour traiter le problème d'apprentissage des classes d'objets.

Une approche hybride discriminante-générative à été proposée par (Bosch et al., 2008) pour la classification des scènes.

Une étude détaillée sur l'utilisation des modèles de sujets pour la recherche d'images sur des bases d'images du monde réel à été conduit par (Horster, 2009) dans le cadre de sa thèse, cette contribution à étudiée pLSA, LDA, et le CTA. Les auteurs de (Sivic J et al., 2008) ont présentés un modèle hiérarchique basé sur la LDA pour l'apprentissage des classes d'objets visuels. La contribution de (Cao L and Fei Fei L, 2007) propose un modèle à concept appelé LTM (Latent Topic Model) pour la

segmentation et la classification des objets. Les auteurs de (Li Y et al., 2007) ont utilisés le PAM pour la reconnaissance d'objets.

Plus de détails sur l'utilisation des modèles de sujets pour l'extraction et la détection des catégories visuelles peuvent être trouvés dans (Fritz and Schiele, 2008).

5.6. Conclusion

Nous avons présentés dans ce chapitre une branche du domaine de l'apprentissage automatique, cette branche comporte la famille des modèles de sujets (ou modèles à concepts).

La section suivante illustre les contributions réalisé dans le cadre de la préparation de cette thèse. Ces contributions sont basées sur l'utilisation de quelques modèles de sujets pour l'indexation et la recherche d'images par le contenu.

CHAPITRE – V

CONTRIBUTIONS

6.1 Introduction

La thématique de la thèse s'articule autour de la fusion d'approches pour l'indexation d'images par le contenu.

Les approches de l'indexation d'images peuvent être classés suivant plusieurs critères. Le modèle d'indexation constitue une base de classification, les systèmes utilisent des modèles basés sur différents principes (statistiques, graphique, ... etc). Le type de caractéristiques utilisées peut être considéré aussi comme critère de classification des approches (Couleur, Texture, ...). Les différents axes de recherches mentionnés dans la section 1 du chapitre 2 de la thèse constituent aussi une base de classification des travaux suivant le centre d'intérêt de leurs contributions (signature, bases d'images, bouclage de pertinence, ... etc).

Dans ce cadre, nous avons concentrés notre étude sur la partie (module/façade) de l'indexation en étudiant l'utilisation des modèles statistiques de sujets (concepts). Ces modèles constituant le cœur de l'indexation permettent une fusion des différentes modalités de caractéristiques et la production d'une représentation unifiée (concepts/sujets) pour indexer et rechercher les images.

Le chapitre présente principalement deux contributions. La première est liée aux images généralistes et elle est basée sur l'utilisation du modèle PAM (Pachinko Allocation Model) pour l'indexation de ces images. Tandis que la deuxième contribution est basée sur l'utilisation du modèle LSA (Latent Semantic Analysis) pour l'indexation des images médicales de tumeurs mammaires. Cette deuxième contribution contient également une partie pour la reconnaissance de type de tumeur basée sur les résultats de l'indexation et la recherche et utilise un algorithme de KNN (K nearest neighbors).

La figure 45 illustre la première contribution tandis que la deuxième contribution est illustrée par la figure 56.

1ère Contribution

PAM pour l'indexation

des

images généralistes

Cette première partie est dédiée à la présentation de la contribution principale de la thèse. Dans cette dernière, nous avons visé l'étude de l'utilité du modèle PAM pour l'indexation et la recherche des images généralistes. L'étude de la combinaison de modalités est également visée par cette contribution. La figure 45 montre l'architecture de l'étude réalisée.

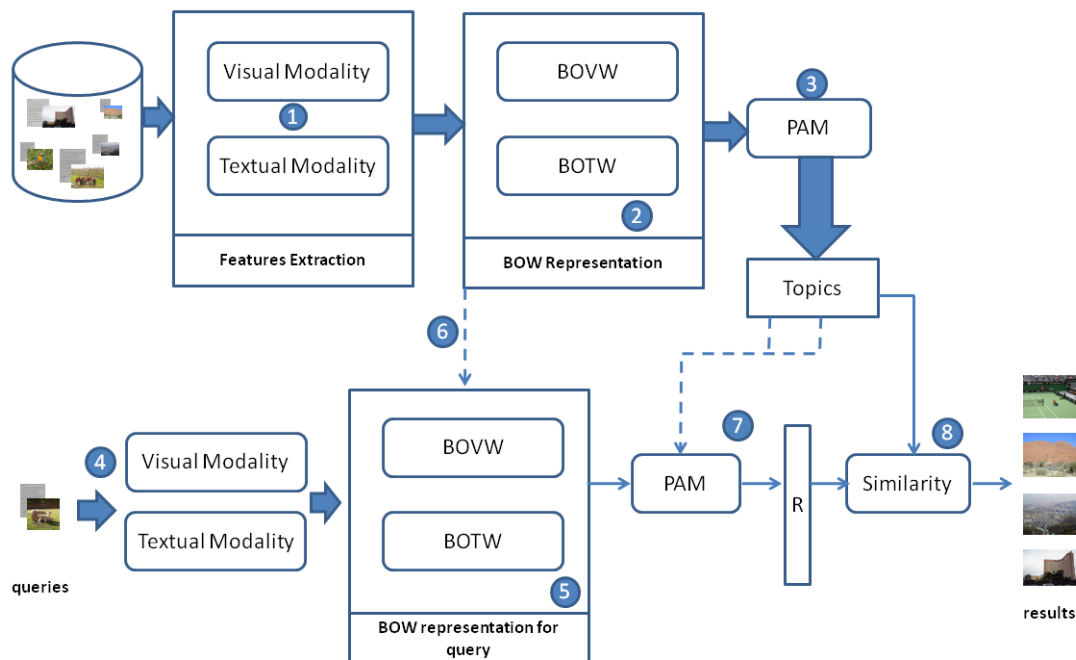


Figure 45. Architecture de l'étude de la première contribution de la thèse

L'étude de ces modèles inclut leur application avec un seul type de modalité des caractéristiques d'images (c.à.d. visuelles ou textuelle) et aussi une combinaison de ces deux types de caractéristique.

6.2 Bases d'images utilisées dans les expérimentations

Les images utilisées dans les expérimentations de la première contribution de la thèse sont issues de la base IAPR TC 12 (IAPR TC-12 Benchmark, 2012). Celle-ci consiste en une collection d'images accompagnées avec des fichiers textes contenant une description de l'image. La figure 46 montre un exemple d'image et sa description dans le fichier texte associé.



```

<DOC>
<DOCNO>annotations/00/25.eng</DOCNO>
<TITLE>The Plaza de Armas</TITLE>
<DESCRIPTION></DESCRIPTION>
<NOTES>The Plaza de Armas is one of the
most visited places in Cochabamba. The
locals are very proud of the colourful
buildings</NOTES>
<LOCATION>Cochabamba,
Bolivia</LOCATION>
<DATE>1 February 2002</DATE>
<IMAGE>images/00/25.jpg</IMAGE>
<THUMBNAIL>thumbnails/00/25.jpg</THU
MBNAIL>
</DOC>

```

Figure 46 Un exemple d'une image de la base IAPR TC 12 et le fichier contenant sa description textuelle

Dans la deuxième contribution, les images sont issues de la base de données MIAS (Web 8). Cette base est constituée une source pour les chercheurs intéressés par le traitement des images mammographiques.

6.3 Modalités des caractéristiques

Les caractéristiques d'images utilisées dans le cadre des travaux de la thèse sont de deux types de modalité, la modalité visuelle et la modalité textuelle. Le premier type contient deux catégories de caractéristiques visuelles, les caractéristiques visuelles locales extraites à partir des images en utilisant la technique SIFT (voir la section 2.6.1 du chapitre 2 de la thèse). La deuxième catégorie de la modalité visuelle constitue les caractéristiques globales, et dans laquelle nous avons calculés un ensemble composé des moments de couleurs et des fonctions d'Haralick pour la texture (voir la section 2.7.2.1 du chapitre 2 de la thèse).

Le deuxième type de modalité est la modalité textuelle. Le calcul des caractéristiques d'images à partir cette modalité se fait en utilisant des textes associés aux images comme sources d'indexation, l'image est indexé à partir ces texte en utilisant des mots clés extrait qui représentent des indexes textuels.

L'étude est basée sur l'utilisation de ces caractéristiques avec les modèles seules et combinées (avec une seule catégorie, et avec une combinaison de deux ou plus de catégories). Dans le cadre de la première contribution, toutes les caractéristiques mentionnées ont été étudiées. La deuxième contribution qui étudie le modèle LSA, utilise la modalité visuelle seulement avec ses deux catégories mentionnées auparavant.

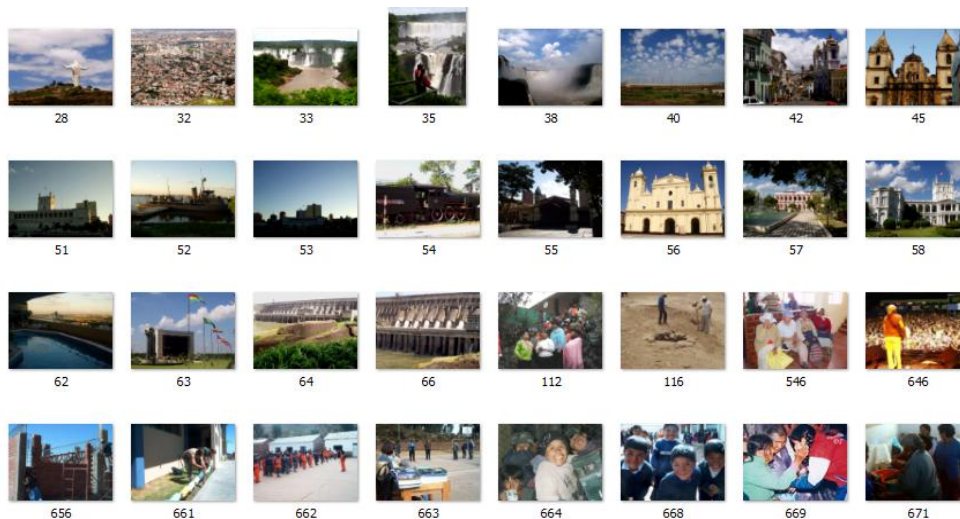


Figure 47. Exemple des images issues de la base IAPR TC 12

6.4 Préparation des données

La phase de préparation des données est la phase qui précède l'application des modèles de sujets. Elle consiste à l'exécution de traitement d'images de la base pour fournir en sortie un ensemble de caractéristique sous une forme qui permet l'application de ces modèles.

Cette phase est composée de deux parties : l'extraction des caractéristiques des images et la construction des matrices de cooccurrence termes-documents (termes-images). La construction de cette matrice pour les caractéristiques visuelles nécessite une partie supplémentaire de travail consiste à l'adaptation de ces caractéristiques en construisant toute d'abord des mots visuels semblable aux mots textuels. La sortie est donc une matrice de termes-visuels-images.

6.4.1 L'extraction des caractéristiques textuelles

Les fichiers associés aux images de la base IAPR TC 12 constituent la source des indexes textuels, ils sont extraits après avoir appliqué des algorithmes de traitement de texte. Les balises présentes dans les textes donnent une information utile pour le calcul de la fréquence des mots extraits.

Un poids qui reflète l'importance est associé à la balise, et la fréquence du mot clé est calculée selon la formule suivante :

$$\text{freq}(\text{index}_{\text{text}}) = \rho \times \text{nbr_occu} \quad (30)$$

Où ρ représente le poids d'importance associé à la balise contenant le mot et nbr_occu est le nombre d'occurrence du mot dans la balise. Si le mot est présent dans plusieurs balises alors la somme de toutes les fréquences calculées est retenue.

Le tableau 2 illustre les valeurs utilisées comme poids d'importances des balises présentes dans les documents textuels

TAG	Poids associé γ
TITLE	9
DESCRIPTION	6
NOTES	3

Tableau 2 poids d'importance associés aux différentes balises du fichier textuel

6.4.2 Caractéristiques visuelles locales

L'extraction des caractéristiques visuelles locales se fait par le moyen de la technique SIFT (voir chapitre 2). Le package de (Web 7) est utilisé pour réaliser l'extraction des points clés et le calcul de leurs caractéristiques. La figure 48 montre les caractéristiques locales extraites avec ce package pour une image de la base utilisée.

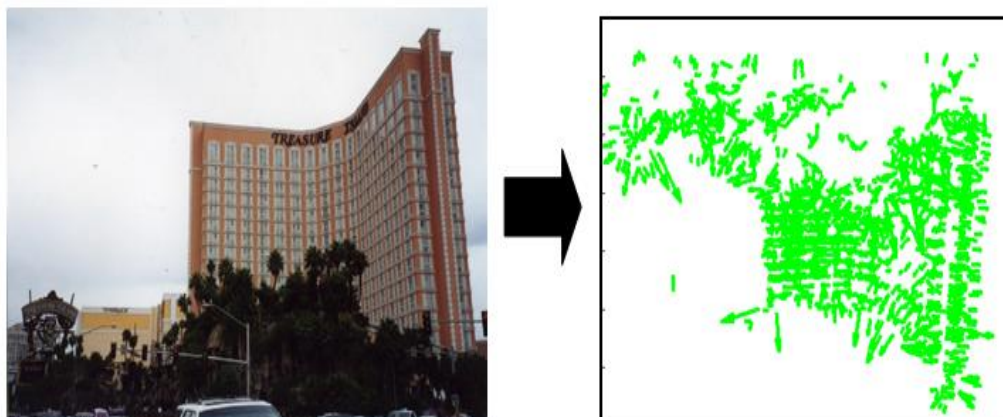


Figure 48. Une image de la base IAPR-TC 12 et les points de caractéristiques locales extraites en utilisant le package de (Web7)

6.4.3 Caractéristiques visuelles globales

L'extraction des caractéristiques globales se fait pour 16 régions de l'image, ces imagerie sont issue de l'application d'une décomposition de 4*4 des images. La figure 49 montre la décomposition de l'image en 16 régions.

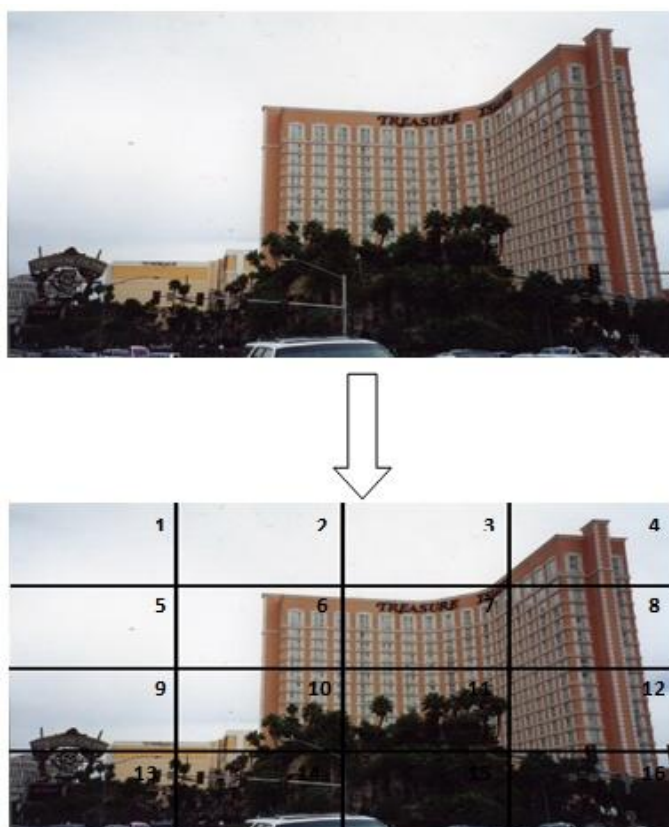


Figure 49. Un exemple d'une image divisée en 16 régions. Chaque région est étiquetée par numéro et elle est considérée comme un mot visuel lors de la construction des mots visuels

Pour chacune de ces régions, la moyenne, la variance et l'écart-type sont calculés pour caractériser la couleur de la région dans les deux espaces RGB et HSV.

Pour la texture, les fonctions d'Haralick sont calculées et constitue les descripteurs de texture de chaque région. A la fin de cette partie, l'image dispose de 16 vecteurs de caractéristiques de couleur et de texture.

6.4.4 Construction des matrices de cooccurrences

La matrice de cooccurrence termes-document est le point de départ pour tous les modèles à concepts. Cette matrice regroupe les sacs de mots de tous les documents, en ligne les termes et en colonne les documents et les éléments de la matrice représentent la fréquence du terme dans le document (voir chapitre 4).

La construction des matrices termes-documents pour les caractéristiques textuelles est simple, après avoir extraire les mots clés et calculer leurs fréquences (voir section 3.1.1 de ce chapitre), ils sont sauvegardé dans la matrice.

Le cas des caractéristiques visuelles nécessite un travail supplémentaire qui consiste à définir des termes visuels à la base de ces caractéristiques. Après avoir construire ces mots visuels, les fréquences sont calculées pour compléter la construction de la matrice.

➤ Construction des mots visuels :

La construction des mots visuels basés sur les caractéristiques locales est réalisée comme suit :

- Appliquer un clustering avec kmeans pour tous les points clés de chaque image. Les centroids sont considérés comme des mots visuels de l'image et la dimension du cluster comme sa fréquence.
- Définir le vocabulaire de la base, cette étapes consiste à un deuxième clustering appliqué sur les centroids de l'étape précédente. Les centroids du même cluster vont représenter le même mot dans toute la base (exemple mots_20).

Le résultat de ces étapes transforme les images en textes (textes visuels), la figure 50 montre la description textuelle construite pour l’image.

Pour le cas des mots visuels construits à la base des caractéristiques globales, la construction est effectuée comme suit :

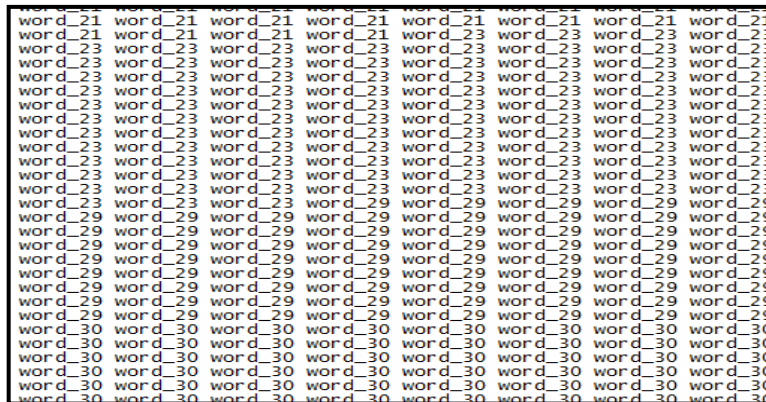


Figure 50. Un fragment de texte généré à la base des caractéristiques visuelles locales. Les mots sont des mots visuels

- Considérer chaque région des 16 régions de l’image comme mot visuel de l’image.
- Appliquer un clustering kmeans de toutes les régions des images de la base pour construire le vocabulaire. Les régions du même cluster représentent le même mot dans toute la base.

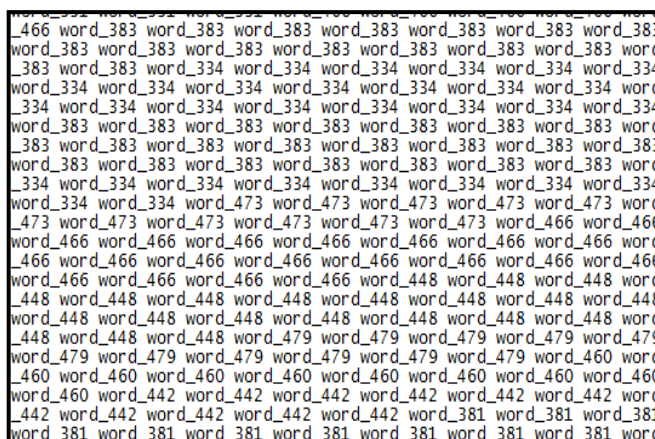


Figure 51. Un fragment de texte généré à la base des caractéristiques visuelles globales. Les mots sont des mots visuels

Le calcul des fréquences des mots visuels dans le cas des caractéristiques visuelles globales passe par le processus suivant :

1. Diviser les 16 régions de l'image en trois classes. Chaque classe est associée à un facteur d'importance Imp_i distribué sur ses régions avec i est le numéro de la région (voir figure 49).

La class "*up*" contient les régions 1,2,3 et 4.

La class "*center*" contient les régions 6,7,10,11,14 et 15.

Class "*right*" contient les régions 5,9 et 13.

Class "*left*" contient les régions 8,12 et 16.

Les deux classes "*right*" et "*left*" constituent une classe "*sides*".

2. La relation entre les différents facteurs d'importance est définie comme suit:

$$Imp_i(i \in sides) = 0,75 \times Imp_i(i \in center) \quad (31)$$

$$Imp_i(i \in up) = 0,25 \times Imp_i(i \in center) \quad (32)$$

3. La fréquence de chaque mot visuel (région i) est calculée comme suit:

$$F_i = Imp_i \times \rho \quad (33)$$

avec ρ comme facteur empirique, il est définie suivant une analyse des différentes expérimentation pour assurer l'équilibrage des vecteurs caractéristiques.

Le résultat de ces étapes transforme les images en texte, la figure 51 montre la description textuelle construite pour l'image à la base de ses caractéristiques visuelles globales.

6.5 Application du modèle PAM pour l'indexation

L'application du modèle PAM sur les différentes représentations des images (mots textuels ou mots visuels) permet d'indexer les images avec l'ensemble des sujets probabilistes extraits. La distribution des sujets est considérée comme vecteur des caractéristiques des images.

Nous avons utilisés un modèle PAM à quatre niveaux. Ce modèle contient deux niveaux de sujet probabiliste, un niveau de super-sujets et un niveau de sous-sujets (voir figure 52).

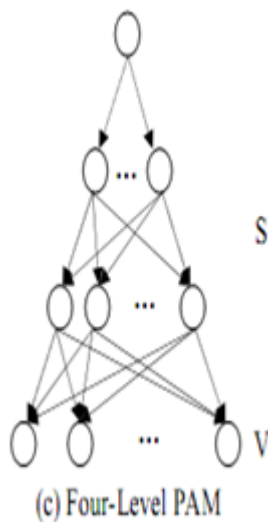


Figure 52. Structure de PAM à 4 niveaux utilisé dans la thèse (Li W and McCallum, 2006))

Les nombre des super-sujets et de sous-sujets à inférer sont précisés. En utilisant le package de (Web6), PAM permet d'appliquer un Gibbs sampling pour inférer l'ensemble des super-sujets probabilistes ainsi que des sous-sujets pour chaque images de la base.

La figure 53 montre un exemple des deux distributions de sujet (super, sous) inférées pour une image de la base. Le nombre de sujets ainsi que les paramètres des modèles de sujet reste toujours un champ d'étude.

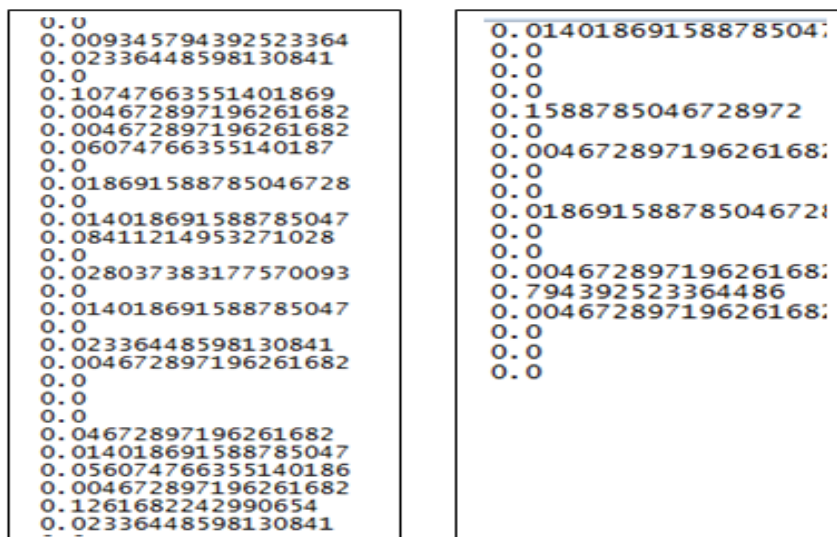


Figure 53. Exemple des deux distributions inférées pour une image de la base, à gauche les sous-sujets et à droite les super-sujets (cas des mots visuels global)

6.5.1 Résultats et perspectives.

Nous allons présenter dans cette section l'ensemble des expérimentations ainsi que les résultats obtenus. La concentration des travaux de la thèse est sur l'indexation et la recherche d'images par le contenu, ce domaine représente un axe indépendant de l'annotation automatique des images et de la reconnaissance d'objets. Nous concluons cette section avec des perspectives et des futures directions.

6.5.2 Les expérimentations de la thèse.

La première contribution dans la thèse contient un ensemble d'expérimentation, elles consistent à étudier l'impact de l'utilisation du modèle PAM avec les caractéristiques dans deux cas : une seule modalité de caractéristiques ou bien plusieurs modalités combinées. L'ensemble des expérimentations de la thèse comporte :

- L'application du modèle PAM avec les caractéristiques locales, cette partie d'expérimentation a fait l'objet des productions scientifique (A.Boulemden and Y.Tlili, 2011a ; A.Boulemden and Y.Tlili, 2011b).
- L'application du modèle PAM avec les caractéristiques globales.
- L'application du modèle PAM avec les caractéristiques textuelles.

- L'application du modèle PAM avec une combinaison des caractéristiques locales et textuelles.
- L'application du modèle PAM avec une combinaison des caractéristiques globales et textuelles.
- L'application du modèle PAM avec une combinaison des caractéristiques locales et globales, cette partie d'expérimentation a fait l'objet des productions scientifique (A.Boulemden and Y.Tlili, 2012a ; A.Boulemden and Y.Tlili, 2012b).

Les expérimentations qui couvrent l'utilisation du PAM avec les caractéristiques globales et textuelles (seules et combinées) font l'objet de la production scientifique de (Boulemden A. et al, 2016).

6.5.3 Les bases d'image utilisée

L'ensemble des images utilisé pour la réalisation des expérimentations de la première contribution est constitué de 504 images d'apprentissage et 61 images de teste. Ces images sont sélectionnées aléatoirement de la base IAPR TC 12. Elles sont de 7 classes montrées dans le tableau 3.

Animals	46	9
City buildings	72	8
Indoor	100	10
Landscape	65	9
People	150	13
Ski	21	5
Tennis	50	7
Nombre total	504	61

Tableau 3 Les classes des images utilisées dans les expérimentations

6.5.4 Extraction des caractéristiques

L'extraction des caractéristique visuelles locale est réalisée en utilisant le package de (Veldaldi, 2008). Pour la construction des mots visuels locaux, nous avons choisie 6 mots par image pour le premiers clustering et 150 mots pour le vocabulaire générale.

Pour la construction des mots visuels globaux, 16 mots par image selon le nombre de régions et 150 mots du vocabulaire.

6.5.5 Application du modèle PAM.

PAM été appliqué pour inférer 18 sous-sujets et 5 super-sujets. Le package du (Mallet, 2011) à été utilisé.

6.5.6 Evaluation et Résultat.

Le choix des valeurs empiriques des différentes phases de préparation des données est basé sur l'étude de quelques travaux utilisant les modèles de sujets.

Pour évaluer les expérimentations, le MAP (Mean Average Precision) à été calculé, les différents résultats vont être présentés.

Les résultats ont été évaluée en utilisant les 3 mesures de similarité euclidienne, cosine et KL divergence (voir chapitre 2 de la thèse).

	Mean Average precision		
	Image global features	Text	Image+text
Animals	0.9336	0.1209	0.2462
City buildings	0.4001	0.1495	0.0864
Indoor	0.2794	0.1511	0.5156
Landscape	0.1288	0.0801	0.0695
People	0.2011	0.3856	0.1673
Ski	0.0246	0.8790	1.0000
Tennis	0.0552	0.6807	0.4953
Global mean average Precision	0.2889	0.3496	0.3686

Tableau 4 Résultats de PAM avec les caractéristiques globales, textes et fusion

	Mean Average precision		
	Image Local features	Text	Image+text
Animals	0.9802	0.1209	0.0639
City buildings	0.4071	0.1495	0.1004
Indoor	0.2796	0.1511	0.1113
Landscape	0.1271	0.0801	0.7823
People	0.1993	0.3856	0.4576
Ski	0.0246	0.8790	1.0000
Tennis	0.0524	0.6807	0.4953
Global mean average precision	0.2958	0.3496	0.4301

Tableau 5 Résultats de PAM avec les caractéristiques locales, textuelles et fusion

	local-global
Animals	0.9889
City buildings	0.4013
Indoor	0.2778
Landscape	0.1266
People	0.1989
Ski	0.0245
Tennis	0.0525
Global mean average Precision	0.2974

Tableau 6 Résultats de PAM avec Combinaison des caractéristiques locales et globales

La combinaison des distances s'effectue selon 4 schémas : distance entre les super-sujets, distances entre les sous-sujets, minimum, moyenne.

Le MAP est une mesure d'évaluation qui prend en considération la position de la réponse pertinente dans la liste des résultats triés. Cela présente un avantage par rapport au calcul simple du rappel et précision de la recherche. La figure 54 présente un exemple de requête utilisée et les 5 résultats triés. La précision directe dans ce cas est égale à 80%, Tandis que le MAP est égale à 74,20%.



Figure 54. 5 premiers résultats pour une image requête. Cas de fusion locales-texte

Les résultats de chaque classe d'images dans les tableaux 4,5 et 6 présentent la moyenne des MAP de toutes les requêtes appliquées pour la classe. Le MAP globale est la moyenne de tous les MAP des différentes classes.

Nous résumons ce qui a été montré par les expérimentations comme suit :

- Au niveau du MAP globale, la combinaison des caractéristiques locales avec le texte a donné le meilleur score et améliore considérablement la performance, voir le tableau 5.
- La combinaison des caractéristiques globales avec le texte permet d'améliorer les résultats mais reste faible par rapport à la première combinaison, voir tableau 5 et 6.
- A partir des deux remarques précédentes, nous pouvons conclure que l'utilisation des caractéristiques locales est plus adéquate pour le cas des bases d'images généralistes.
- Cette dernière conclusion peut être confirmée par les résultats de la combinaison locale-globales et de toutes les caractéristiques.

Nous concluons en général que l'utilisation du PAM est souhaitable pour conduire à une amélioration des résultats dans le cas de l'utilisation d'une combinaison des caractéristiques visuelles locales avec les caractéristiques textuelles, cette conclusion est pour les bases d'images généralistes.

Il faut remarquer aussi que les expérimentations sont sensibles aux différents paramètres qui sont le nombre des super-sujet et sous-sujets, les paramètres interne des modèles, les itérations de l'algorithme d'inférence des sujets. Tous ces paramètres nécessitent une étude supplémentaire.

6.5.6.1 Discussion détaillée des résultats

Nous avons donné dans la section précédente une vue globale sur les résultats du MAP obtenu dans cette première partie de contribution. Nous continuons dans cette section avec une présentation et une discussion détaillée de ces résultats. Nous allons discuter les différents cas d'expérimentation pour chaque classe d'images. ces cas sont:

- L'utilisation de la modalité visuelle.

- L'utilisation de la modalité textuelle.
- Les différentes formes de combinaison appliquées.

a) L'utilisation de la modalité visuelle

Les résultats concernant l'utilisation d'un seul type de caractéristiques visuelles sont présentés dans les tableaux 4 et 5. Tandis que les résultats de la combinaison de ces deux caractéristiques sont illustrés dans le tableau 6.

Tout d'abord, commençons avec le cas présenté dans les tableaux 4 et 5 où les caractéristiques globales et locales ont été appliqués indépendamment. Nous remarquons que cette manipulation a conduit à des résultats très proche et ce pour les mêmes classes d'images.

Suivant cette évaluation qualitative, la classe "*Animals*" présente le meilleur MAP et qui est un MAP très élevé. Le MAP de la classe "*City Buildings*" présente un MAP élevé. Concernant les résultats des MAP pour les classes "*Indoor*", "*People*" et "*Landscape*" respectivement, ils peuvent être considérés comme de grade moyen. Les faibles MAP ont été rencontrés avec les classes "*Tennis*" et "*Ski*" respectivement.

Nous pensons que le degré de diversité des images influence les résultats de l'application des caractéristiques visuelle. On peut exprimer cette diversité avec trois facteurs. Le premier facteur est la diversité en couleur et texture, le deuxième concerne la diversité des différentes formes dans l'image et le troisième facteur est celui du contraste de l'image dans un sens de degré de distinction entre les différents objets et régions de l'image.

Une évaluation visuelle des images de la base basée sur ces facteurs peut exprimer les résultats obtenus. Cette évaluation humaine peut être subjective mais elle peut plus au moins fournir une vue sur la diversité de chaque classe. Le tableau 7 présente les résultats d'évaluation des 3 critères par quatre utilisateurs.

Les classes avec un niveau "High" de contraste et "Medium" pour les autres facteurs présente les MAP les plus élevés. L'élargissement de cette évaluation est souhaitable pour confirmer mieux les résultats.

	MAP Global	MAP local	Color-Texture	Shape	Contrast	Overall
Animals	0.9336	0.9802	57.75%	74.50%	83.00%	71.75%
City Buildings	0.4001	0.4071	66.25%	58.00%	83.00%	69.08%
Indoor	0.2794	0.2796	91.50%	49.75%	83.00%	74.75%
Landscape	0.1288	0.1271	49.50%	58.00%	66.00%	57.83%
People	0.2011	0.1993	91.50%	49.50%	91.50%	77.50%
Ski	0.0246	0.0246	41.25%	41.25%	49.75%	44.08%
Tennis	0.0552	0.0524	66.00%	49.50%	49.50%	55.00%

Tableau 7 Evaluation visuelles de trois critères visuels dans l'image

b) L'utilisation de la modalité textuelle

Les résultats de MAP obtenu pour le cas de l'indexation textuelle dépendent fortement de la qualité des annotations extraites des fichiers textuels. En effet, lors de l'évaluation des fichiers contenant les index textuels, on trouve que les classes ayant les meilleurs MAP ont une indexation textuelle plus précise et moins bruitée. Les images ayant aussi plus d'index textuels en commun pour ces classes.

La performance des modèles de sujet a été prouvée pour l'indexation textuelle dans plusieurs travaux vu qu'ils sont initialement proposés pour ce type de données. Cette performance est reflétée par nos résultats obtenus. Il faut noter que la source textuelle de la base IAPR n'est pas une source riche et puissante. Donc, de ce point de vue, les résultats peuvent être généralement acceptables.

b) La combinaison des modalités textuelles et visuelles

- **La combinaison des caractéristiques locales avec le texte**

D'une manière générale, la combinaison du texte avec les caractéristiques locales permet d'améliorer les résultats de MAP pour les classes ayant un faible MAP visuel. Ces classes sont "Landscape", "People", "Ski" et "Tennis". À l'inverse, le texte affecte d'une manière négative les classes ayant un MAP visuel fort qui sont les classes "Animals", "City Building" et la classe "Indoor" du MAP visuel local moyen (Voir tableau 5).

- **La combinaison des caractéristiques globales avec le texte**

Les caractéristiques textuelles ont un impact positif lors de la combinaison avec les caractéristiques visuelles des classes ayant de faible MAP visuelle. Ces dernières sont les classes "Ski" et "Tennis". A l'inverse, les classes "Animals" et "City Buildings" ayant de MAP très fort ont été considérablement influencé par la combinaison dans le sens négatives. L'impact négative aussi est rencontré pour les classes "People" et "Landscape". La combinaison été positive pour la classe "Indoor" et qui permet d'améliorer son MAP.

On pense qu'une étude basé sur les facteurs cités précédemment dans la section peut être utile pour caractériser les classes d'images généralistes pouvant tirer avantage lors de la combinaison de leurs caractéristiques globales avec le texte (Voir tableau 4).

b) La combinaison des modalités visuelles

La combinaison des caractéristiques visuelles locales et globales n'avait pas une influence remarquable sur les résultats. Une légère amélioration est remarqué pour la classe "Animals" (Voir tableau 6). En effet, ces résultats semble cohérente avec une étude que nous avons menés dans (Boulemden A. et Tlili Y., 2010). Cette étude ayant pour objectif l'indexation des images web par LSA et combinaison de caractéristiques a montrée que l'utilisation d'une seule modalité (sous ensemble de vecteur visuel) peut garantir des résultats semblables à ceux d'une combinaison de plusieurs caractéristiques visuelles et cela dans un contexte d'images généralistes.

Nous avons choisi de comparer le meilleurs résultats obtenu avec les meilleurs résultats de (Pham et al., 2007) et (Chandrika and Jawaher, 2010). Ces deux travaux ont utilisés des modèles de sujets avec des images de la base IAPR. Le tableau 8 résume la comparaison.

	Modèle de sujet appliqué	Total dimension visuelle	MAP combinaison
Notre travail	PAM	160	43.01 %
Pham et al., 2007	LSA	224	46.96 %
Chandrika and Jawaher, 2010	LSA	Inconnu	55 %
	pLSA		59 %

Tableau 8 Comparaison des résultats avec des travaux similaires

Nous pensons qu'avec plus de manipulation de paramètres d'application du PAM comme le nombre de sujets à inférer et les itérations de l'algorithme d'inférence. Nos résultats peuvent être améliorés et rapprochés du résultat de (Pham et al., 2007) et (Chandrika and Jawaher, 2010). Cette dernière référence fournit des informations partielles concernant leur application de modèles de sujets, donc plus d'informations est nécessaire pour la comparaison. Reste que cette dernière référence montre que l'application du pLSA rencontre le meilleur score. Le gain en dimensionnalité du vecteur visuel dans notre travail peut être considéré comme un apport positif vis-à-vis le passage à l'échelle.

La section suivante présente la deuxième contribution de la thèse. Celle-ci concerne l'utilisation des modèles de sujets avec des caractéristiques multimodales pour l'indexation et la recherche des images de tumeurs mammaires.

2ème Contribution
LSA et KNN
pour
l'indexation et la reconnaissance
des
tumeurs mammaires

6.6 Indexation et reconnaissance des tumeurs mammaires par le modèle LSA

6.6.1. Introduction

Cette partie de contribution vise essentiellement à étudier l'application du modèle à concepts LSA (Latent Semantic Analysis) dans un contexte médical. Dans ce cadre, LSA est utilisé pour l'indexation et la recherche des images de tumeurs mammaires. Le travail contient également une deuxième partie qui consiste à l'utilisation de l'algorithme de classification de KNN (K nearest neighbors) sur les résultats de la recherche pour reconnaître de type de tumeur de la requête.

La première partie de cette contribution à été réalisé pour la préparation des mémoires de master de (Mouassa et al., 2015) et (Mender et al., 2016). Les buts des études menées dans cette contribution peuvent être résumés comme suit:

- 1- Etudier l'utilité des modèles à concepts pour l'indexation et la recherche d'images de tumeurs mammographie.
- 2- Explorer l'utilisation de plusieurs types de modalités de caractéristiques.
- 3- Etudier la classification des tumeurs par le moyen de l'algorithme des KNN (K nearest Neighbors) appliqué sur les résultats de la recherche.

Des travaux de recherche similaires utilisent LSA ou sa variante probabiliste pLSA dans le contexte de l'imagerie médicale. Parmi ces recherches, nous citons (Bousslimi et al., 2013) dans laquelle les auteurs ont utilisé LSA pour l'auto-annotation des images médicales. Ils ont employés une combinaison de la représentation des mots visuels construits à partir des images avec les rapports textuels pour produire l'auto-annotation des images médicales.

Pour réaliser les buts de l'étude, l'ensemble des expérimentations performées comporte:

- L'application du modèle LSA avec les caractéristiques locales de SIFT.

- L'application du modèle LSA avec les caractéristiques globales de texture d'Haralick.
- L'application de KNN sur les résultats de recherche pour classer les tumeurs.

D'autres expérimentations constituent des perspectives de cette partie et sont:

- L'application du modèle LSA avec une combinaison des caractéristiques locales et globales.
- L'application du modèle LSA avec des caractéristiques textuelles issues des rapports médicaux.
- L'application du modèle PAM avec une combinaison des caractéristiques visuelles et textuelles.
- Etudier la reconnaissance par KNN et la combinaison de toutes les caractéristiques.

Nous allons présenter dans cette section de contributions les points suivant:

- a. Les images utilisées pour la réalisation des différentes expérimentations.
- b. Les caractéristiques utilisées et leur présentation.
- c. L'application du modèle LSA pour l'indexation et la recherche des tumeurs mammaires.
- d. L'application de l'algorithme KNN sur le résultat de recherche pour reconnaître le type de tumeur.
- e. Les résultats des différentes expérimentations.
- f. La conclusion et les perspectives de cette partie de contribution.

6.6.2. Base d'images

Nous avons mentionnés dans la section 5.2 du chapitre contribution que la base MIAS a constituée la source des images utilisées dans les expérimentations de cette partie de contribution. Deux collection ont été construits à partir les images de cette base, la première consiste en une base d'apprentissage et composé de 140 images de tumeurs mammaires dont 70 pour les tumeurs de type maligne et 70 pour les tumeurs de type bénigne. La deuxième collection représente la base de teste, elle est composée de 20 images requêtes dont 10 pour des tumeurs de type maligne et 10 pour des tumeurs de type bénigne. Les images de ces deux collections contiennent

seulement la partie de tumeurs qui a été extraite manuellement. La figure 55 montre un exemple d'une image de la base d'apprentissage utilisée dans les expérimentations.

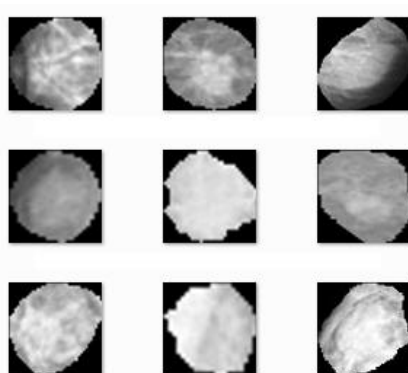


Figure 55 Exemple des images utilisées pour la contribution

6.6.3. Indexation et recherche d'images de tumeurs mammaires

La première partie de cette contribution s'intéresse par l'étude de l'application du modèle LSA pour l'indexation et la recherche des tumeurs mammaires. La figure 56 montre une architecture de cette étude.

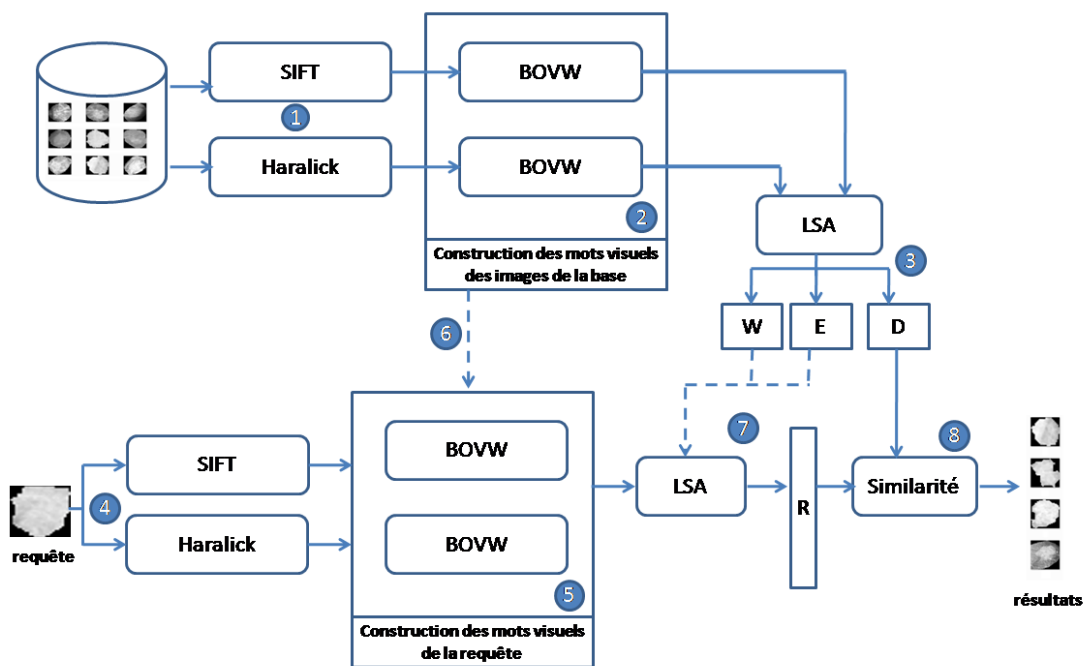


Figure 56 Architecture de l'étude de LSA pour l'indexation des tumeurs mammaires

Le modèle LSA constitue le cœur du module de l'indexation, il est appliqué sur des matrices de cooccurrence des mots visuels images. Ces matrices sont construites en utilisant les sacs de mots visuels.

Les détails de la construction des mots visuels à la base des caractéristiques visuelles locales et globales peuvent être trouvés dans la section 5.4 de ce chapitre de contribution.

Les caractéristiques visuelles locales et globales sont calculées pour les images de la base (voir partie 1 de la figure 56). Pour chacune de ces deux types de caractéristiques visuelles, le processus approprié de la construction des mots visuels est appliqué (BOVW pour bag of visual words, voir partie 2 de la figure 56). Le modèle LSA est par la suite appliqué sur chaque matrice de mot-visuels-images produite. L'application du modèle LSA sur la matrice de cooccurrences des mots-visuels-images produit 3 matrices (voir chapitre 4 de la thèse), une matrice des documents qui représente l'espace de l'indexation des images d'apprentissage. Une matrice des mots visuels et une matrice des valeurs singulières (voir la partie 3 de la figure 56).

Lors de la phase de la recherche, les caractéristiques visuelles locales et globales sont calculées pour l'image de la requête. Ces deux dernières matrices vont être utilisées pour produire la représentation du vecteur représentant les images de requêtes (voir la partie 4 de la figure 56). La construction des mots visuels est par la suite appliquée sur chaque type de ces caractéristiques (voir la partie 5 de la figure 56). Cette construction utilise les informations issues de la même phase exécutée pour les images de la base, les mots visuels de la requête vont être construits en faisant référence aux mots visuels construits pour les images de la base (voir la partie 6 de la figure 56). Le modèle LSA permet par la suite de calculer un vecteur descripteur de la requête, ce calcul est effectué en utilisant les matrices W et E issues de l'application de LSA sur les images de la base (voir la partie 7 de la figure 56). La dernière étape consiste à mesurer la similarité entre le vecteur descripteur de la requête et ceux des images de la base. Ces derniers sont trouvés dans la matrice D . La liste des résultats est finalement affichée (voir la partie 8 de la figure 56).

6.6.4. Reconnaissance de type de tumeur par l'application de KNN

La phase de l'indexation et la recherche des images est suivie par une phase de reconnaissance de type de tumeur mammaires de l'image requête. Cette reconnaissance est basée sur l'application de l'algorithme de classification de KNN (K nearest neighbors) (voir chapitre 3 de la thèse) sur l'ensemble des images résultats de la requête. La simplicité de l'algorithme KNN est avantageuse pour la rapidité.

Nous avons explorés les résultats de KNN en variant la valeur de k utilisé.

6.6.5. Résultats et discussion

Cette partie présente les différents résultats liés à l'étude menée dans cette deuxième contribution. Les résultats concernent les deux parties de la contribution, celle de l'indexation et la recherche et l'autre de la reconnaissance de type de tumeur.

Nous avons utilisé le MAP (Mean Average Précision) pour l'évaluation de l'indexation et la recherche d'images. Le MAP des requêtes de la base de teste est calculé pour les deux cas étudiés, le cas des caractéristiques locales de SIFT, et le cas des caractéristiques globales de Haralick. Le tableau 9 montre les résultats de l'évaluation du MAP. Le nombre de requêtes utilisées est 20 requêtes dont les 10 premières sont des requêtes de tumeurs malignes et les 10 dernières sont des requêtes de tumeurs bénignes. La figure 57 montrent les graphes des résultats des deux catégories malignes et bénignes.

Requête malignes	Local SIFT	Global Haralick	Requête bénignes	Local SIFT	Global Haralick
AP 1	0.4608	0.4900	AP 11	0.5839	0.5292
AP 2	0.5234	0.4866	AP 12	0.5117	0.5480
AP 3	0.4692	0.4843	AP 13	0.5227	0.5256
AP 4	0.5119	0.4707	AP 14	0.5910	0.5142
AP 5	0.5440	0.4931	AP 15	0.5665	0.5193
AP 6	0.4846	0.4793	AP 16	0.5376	0.5228
AP 7	0.4480	0.4865	AP 17	0.4830	0.5297
AP 8	0.4544	0.4758	AP 18	0.5515	0.5167
AP 9	0.4857	0.4831	AP 19	0.5600	0.5119
AP 10	0.4855	0.4862	AP 20	0.5895	0.5126
Moyenne AP malignes	0.4868	0.4836	Moyenne AP bénignes	0.5497	0.5230
MAP local	0.5183		MAP global	0,5033	

Tableau 9 Résultats de l'évaluation du MAP de LSA pour l'indexation des tumeurs mammaires

Les résultats de l'application de LSA pour l'indexation et la recherche des tumeurs mammaires montrent que son utilisation est meilleure pour les tumeurs de type bénignes.

Le MAP des requêtes bénignes pour l'utilisation de LSA avec les caractéristiques locales SIFT est mieux de 11,5% par rapport aux requêtes malignes du même cas. Dans le cas de l'utilisation des caractéristiques de Haralick avec LSA, le MAP des requêtes bénignes est meilleur de 7,54%. Le MAP globale de toutes les requêtes présente une amélioration de 2,9% pour le cas des caractéristiques locales par rapport aux caractéristiques globales.

Les résultats restent proches et peuvent être considérés acceptable. L'amélioration des résultat constitue un perspective qui doit être aborder aussi avec les reste des expérimentation non réalisée.

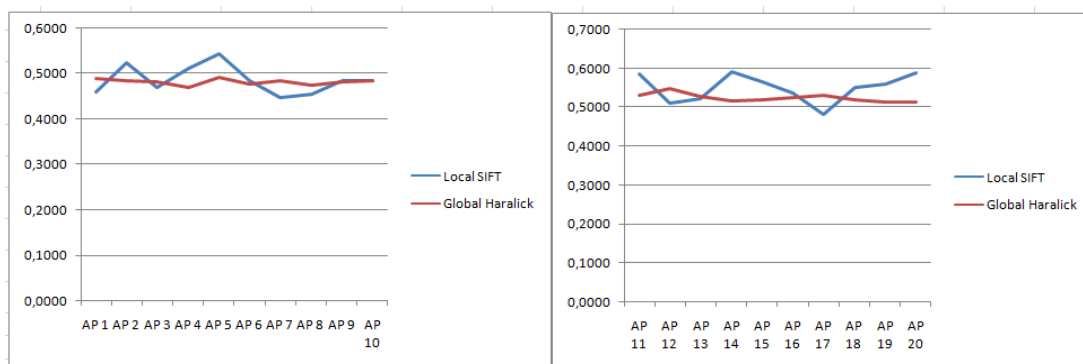


Figure 57 graphes des résultats de MAP par catégories de requêtes, à gauche les requêtes malignes et à droite les requêtes bénignes

Nous avons calculés le taux de la classification des tumeurs pour l'évaluation de la reconnaissance par KNN. Le taux de la reconnaissance est calculé pour les deux cas également, le cas des caractéristiques locales de SIFT, et le cas des caractéristiques globales de Haralick. En effet, plusieurs taux ont été calculés suivants les différentes valeurs de k utilisé par l'algorithme KNN. Les résultats de l'évaluation de la reconnaissance de type de tumeur sont illustrés aussi par les graphes de la figure 58. Elle présente un échantillon de la variation des valeurs de taux de reconnaissance en fonction de la valeur de k utilisé par l'algorithme KNN.

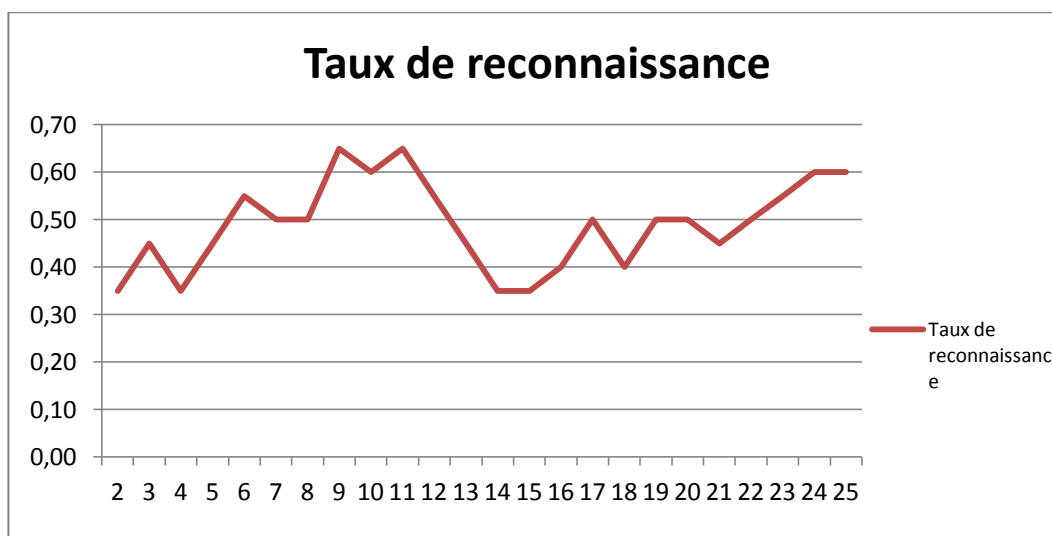


Figure 58 résultats de l'évaluation de la reconnaissance de type de tumeur

Les résultats de l'utilisation de l'algorithme de KNN pour la reconnaissance de type de tumeurs montrent une variation suivant la valeur de K. Le taux maximum rencontré été à 80% pour K=69. Tandis que le taux minimum été à 35% pour plusieurs valeurs de K. Nous jugeons le taux comme acceptable en vue la simplicité de l'algorithme de KNN. L'application d'un algorithme de classification plus sophistiqué peut donner des taux supérieurs. Nous préférons d'utiliser des algorithmes simple car ils permettent un passage à l'échelle plus efficace. Les algorithmes simples ont montré leurs efficacité dans d'autres domaines lorsque des manipulations d'amélioration sont ajoutés, cela peut être trouvé par exemple dans la détection de visage en temps réel avec le travail de (Viola and Jones, 2001).

6.7 Conclusion

Nous avons présentés via ce chapitre les différentes contributions réalisées dans la thèse. Essentiellement, elles sont classées en deux contributions majeures.

La première contribution de la thèse concerne l'étude de l'application du modèle PAM (Pachinko Allocation Model) pour l'indexation et la recherche des images généralistes. Cette étude comporte l'application du PAM avec différents types de caractéristiques visuelles et textuelles seules et combinées. Les résultats montrent l'utilité du modèle pour l'indexation et la recherche des images généralistes. Cette

utilité varie selon le type de caractéristique employée avec le modèle. L'amélioration des résultats obtenus ainsi que la réduction de la sensibilité du modèle aux paramètres de son application constituent les perspectives de cette première contribution.

L'indexation et la recherche des tumeurs mammaires ainsi que la reconnaissance de leurs types font les sujets étudiés dans la deuxième contribution. L'étude du premier sujet est basée sur l'utilisation du modèle à concepts LSA (Latent Semantic Analysis) avec les caractéristiques visuelles de l'image locales et globales. Ceci est étudié pour l'indexation et la recherche des tumeurs mammographiques.

L'étude du deuxième sujets inclut principalement l'application de la technique de classification de KNN (K Nearest Neighbors) sur les résultats de la recherche pour reconnaître le type des tumeurs requête.

Le but de l'évaluation est principalement de comparer la performance de l'indexation et de la reconnaissance avec les deux types de caractéristiques (globales et locales). Les résultats ont montrés l'utilité des approches étudiées. L'amélioration des résultats ainsi que l'étude de l'utilisation d'autres types de caractéristiques seules et combinées, présentent les perspectives de cette deuxième contribution.

Les résultats des différentes études illustrent l'utilité des ces propositions. Les perspectives concentrent essentiellement sur l'amélioration de l'application de ces différentes propositions étudiées.

CONCLUSION GENERALE ET PERSPECTIVES

7.1 Conclusion

L'indexation et la recherche d'image par le contenu est un domaine de recherche très intéressant et très dynamique. Son importance est majeure pour améliorer l'implémentation des CBIRs (Content Based Image Retrieval Systems) pouvant répondre aux besoins d'utilisateurs et ainsi récupérer les images similaires à partir des bases d'images mise à leurs dispositions.

Le problème majeur du domaine est le fossé sémantique, l'absence de liaison entre les caractéristiques numérique de bas niveau extraites à partir des images et l'interprétation sémantique visée par l'utilisateur, ce problème est plus compliquées avec les bases d'images généralistes. La réduction de ce fossé permet une amélioration de performance des CBIRs.

L'indexation multimodale qui consiste à combiner plusieurs modalités de caractéristiques d'images ainsi que l'utilisation des modèles d'indexation de haut niveau permettant de capture les liens entre les données, ces deux approches sont au cœur des travaux de réduction du gap sémantique.

Pour contribuer avec de solutions à ce problème, nous avons proposés une approche d'indexation et de recherche d'images par le contenu. Cette approche est basée sur un modèle d'indexation appelée Pachinko Allocation Model (PAM). PAM appartient à la famille des modèles de sujets probabilistes, cette famille initialement proposée pour la modélisation des connaissances dans les grands corpus de texte. L'avantage de cette famille de modèle réside dans sa capacité à capturer la corrélation des mots dans ces corpus. Cet avantage peut être reproduit lors de l'utilisation de ces modèles dans le traitement d'image. En effet, ils permettent de construire un modèle d'indexation de haut niveau. Le PAM est plus riche par rapport aux autres modèles de sujets probabilistes en termes de niveaux de concepts à extraire à partir des données. L'utilisation de PAM sur les images nécessite l'adaptation des

caractéristiques de celle-ci et cela via la construction des mots visuels similaire aux mots textuels.

Nous avons étudié l'impact de l'utilisation du modèle PAM comme modèle d'indexation et cela avec une seule modalité de caractéristiques (visuelle locale, visuelles globales et textuelles) et aussi avec une combinaison de ces modalités. Le contexte de cette étude est les bases d'images généralistes.

Les résultats de cette étude permettent de confirmer l'utilité de ce modèle comme modèle d'indexation et surtout avec une combinaison des caractéristiques visuelles locales et des caractéristiques textuelles. L'étude nous a permis aussi de confirmer que l'utilisation des caractéristiques visuelles locales est le plus adéquat pour le cas des bases d'images généralistes.

Nous avons également une contribution basée sur l'utilisation du modèle de sujet LSA (Latent Semantic Analysis) dans un contexte d'indexation d'images spécialistes. Le contexte consiste en une application du modèle LSA pour l'indexation et la recherche des tumeurs dans les bases d'images de mammographie. Une deuxième phase est appliquée pour la reconnaissance de type de tumeurs à la base des résultats de la recherche de la première phase. Cette reconnaissance implique l'utilisation de l'algorithme de classification de KNN (K Nearest Neighbors).

7.2 Perspectives

Il est possible d'affirmer que le travail mené dans le cadre de cette thèse a pour ambition de fournir un modèle d'indexation d'images par le contenu permettant d'améliorer la performance des systèmes CBIRs. Cependant, ce travail ne s'arrête pas là, et on peut dire que beaucoup de perspectives expérimentales et de recherche seront poursuivies à court et à long terme.

Nous donnons un résumé sur ces perspectives dans les points suivant :

- Etudier les paramètres d'application des modèles de sujets est une voie de recherche importante dans le but d'améliorer leur performance. L'étude aussi des caractéristiques pertinentes des images pour l'utilisation de ces modèles.

- L'étude de l'application des modèles de sujets pour l'indexation et la recherche d'images dans les bases d'images spécialistes. D'autres bases médicales peuvent être visées ainsi que des bases spécialistes de différentes natures.
- La combinaison des modèles de sujets semblent une voie de recherche importante. Nous sommes entrainés de mener des expérimentations dans ce sens.
- La proposition de nouveau modèle de sujet est une autre perspective importante.

Références

Agarwal, S. and Roth, D. (2002) 'Learning a sparse representation for object detection'. Proceedings of the 7th European Conference on Computer Vision-Part.

Ahrendt, P., Goutte, C. and J. Larsen (2005) 'Co-occurrence models in music genre classification'. Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing.

Baeza Yates, R.A. and B. Ribeiro Neto, B. (1999) 'Modern Information Retrieval'. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Barnard, K. et al. (2003) 'Matching words and pictures'. Journal of Machine Learning Research.

Bay, H., Tuytelaars, T. and Gool, L. V. (2008) 'SURF: Speeded Up Robust Features', Computer Vision and Image Understanding, Vol. 110 No. 3, pp. 346-359.

Belongie, S., Malik, J. and Puzicha, J. (2002) 'Shape matching and object recognition using shape contexts'. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, pp.509–522.

Berg, A.C., Berg, T.L. and Malik, J. (2005) 'Shape matching and object recognition using low distortion correspondences'. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition.

Berk, T. et al. (1982) 'A new color-naming system for graphics language', IEEE Computer Graphics and Applications. Vol.2 No.3, pp.37–44.

Bhattacharjya, R.K. (2013) 'Introduction to Genetic Algorithms', department of Civil Engineering, IIT Guwahati

Bichot, C.E. (2007) 'élaboration d'une nouvelle métaheuristique pour le partitionnement de graphe : la méthode de fusion-fission. application du découpage de l'espace aérien', Thèse de doctorat, Sûreté du Logiciel et Calcul à Haute Performance, École doctorale : Informatique et Télécommunications,.

Blei, D.M., et al. (2004) 'Hierarchical topic models and the nested chinese restaurant process'. S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*.

Blei, D.M. and Jordan, M.I. (2003) 'Modeling annotated data'. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*.

Blei, D.M. and Lafferty, J.D. (2007) 'A correlated topic model of science'. *The Annals of Applied Statistics*, Vol.1No.1, pp.17–35.

Blei, D.M. et al. (2003), 'Latent Dirichlet Allocation'. *Journal of Machine Learning Research*, Vol.3, pp.993–1022.

Bodenhofer, U. (2013) 'Genetic Algorithms: Theory and Applications', *Lecture Notes, Fuzzy Logic Laboratorium, Linz-Hagenberg*.

Bosch, A., Zisserman, A., and Munoz, X. (2008) 'Scene Classification Using a Hybrid Generative/Discriminative Approach', *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 30, No. 4.

Boulemden, A. and Tlili, Y. (2011) 'Pachinko Allocation Model with Image Local Features for Image Retrieval Tasks', *Proceedings of the International Arab Conference of Information Technology*.

Boulemden, A. and Tlili, Y. (2012) 'Image Indexing and Retrieval with Pachinko Allocation Model: Application on Local and Global Features', *Knowledge Management and Acquisition for Intelligent Systems*, Volume 7457 of the series *Lecture Notes in Computer Science*, pp.140-146.

Boulemden, A. and Tlili, Y. (2016) ‘Content Based Image Retrieval with Pachinko Allocation Model and a Combination of Color, Texture and Text Features’, *International Journal of Computational Vision and Robotics*, (To appear).

Bouslimi, R. et al. (2013) ‘Using A Bag Of Words For Automatic Medical Image Annotation With A Latent Semantic’, *International Journal of Artificial Intelligence & Applications (IJAIA)*, Vol.4 No.3. pp. 51-60.

Cao, L. and Fei-Fei, L. (2007) ‘Spatially coherent latent topic model for concurrent object segmentation and classification’. In *IEEE International Conference on Computer Vision*.

Carson, C. et al. (2002) ‘Blobworld: image segmentation using expectation-maximization and its application to image querying’. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24 No.8, pp. 1026–1038.

Chandrika, P. and Jawahar, C. (2010) ‘Multi modal semantic indexing for image retrieval’. In *CIVR 2010. Proceedings of the ACM International Conference on Image and Video Retrieval*, Xi'an, China, pp. 342-349.

Chang, S.F. et al. (1998) ‘Semantic visual templates: linking visual features to semantics’, *Proceeding of the Workshop on Content Based Video Search and Retrieval in the International Conference on Image Processing*.

Chemudugunta, C., Smyth, P., and Steyvers, M. (2007) ‘Modeling general and specific aspects of documents with a probabilistic topic model’. In B. Schölkopf, J. Platt, and T. Hoffman editors, *Advances in Neural Information Processing Systems*.

Chen, Y. et al. (2003) ‘An unsupervised learning approach to content-based image retrieval’, *Proceedings of the International Symposium on Signal Processing and its Applications*, pp. 197–200.

Conway, D.M. (1992) ‘An experimental comparison of three natural language color naming models’, *Proceedings of the East–West International Conference on Human-Computer Interactions*, St. Petersburg, Russia.

Crandall, D. et al. (2009) ‘Mapping the world’s photos’. *Proceedings of the 18th International World Wide Web Conference*.

Cornuéjols, A. (2002) ‘Une nouvelle méthode d’apprentissage : Les SVM. Séparateurs à Vaste marge’, bulletin de l’AFIA, N° 51, Université de Paris-Sud.

Cortes, C. and Vapnik, V. (1995) ‘Soft margin classifiers’. *Machine Learning*, Vol.20, pp. 273-297.

Cox, J. *et al.*, 2000 ‘The Bayesian Image Retrieval System, PicHunter : Theory, Implementation, and Psychological Experiments’, *IEEE Transactions on Image Processing*, Vol.9 No.1, pp.20-37.

Csurka, G., et al. (2004) ‘Visual categorization with bags of keypoints’. Workshop on Statistical Learning in Computer Vision, ECCV, pages 1–22, 2004.

Datta, R. , Joshi, D., Li, J., and Wang, J.Z. (2008) ‘Image retrieval: Ideas, influences, and trends of the new age’, *ACM Computer Surveys*, Vol.40 No. 2, pp.1–60.

Deerwester, S.C. et al., (1990) ‘Indexing by latent semantic analysis’, *Journal of the Association for Information Science and Technology*, Vol. 41 No.6, pp.391-407.

Dempster, A., Laird, N., and Rubin, D. (1977) ‘Maximum likelihood from incomplete data via the EM algorithm’. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

Dutot, A. and Olivier, D. (2005) ‘Optimisation par essaim de particules Application au problème des n-Reines’, *Laboratoire Informatique du Havre, Université du Havre*.

Fan, J., Gao, Y., Luo, H. and Xu, G. (2004) ‘Automatic image annotation by using concept sensitive salient objects for image content representation’. *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Fang, Y. and Geman, D. (2005) ‘Experiments in mental face retrieval’. *Proceedings of the International Conference on Audio-Based and Video-Based Biometric Person Authentication*.

Fang, Y. et al. (2005) ‘An interactive system for mental face retrieval’. Proceedings of the International Workshop on Multimedia Information Retrieval at the International Multimedia Conference.

Fei Fei, L., Fergus, R., and Perona, P. (2007) ‘Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories’, *Computer Vision and Image Understanding*, Vol.106 No.1, pp.59–70.

Fergus, R., Fei-Fei, L., Perona, P., and Zisserman, A. (2005) ‘Learning object categories from google’s image search’. Proceedings of the Tenth IEEE International Conference on Computer Vision, pp.1816–1823.

Fergus, R., Perona, P., and A. Zisserman, A (2003) ‘Object class recognition by unsupervised scale invariant learning’. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Fischler, M.A. and R. A. Elschlager, R.A. (1973) ‘The representation and matching of pictorial structures’. *IEEE Transactions on Computers* Vol.22 No.1, pp. 67-92.

Geman S. and Geman, D. (1984) ‘Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images’. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.6 No.6, pp.721-741.

Gevers, T. and Smeulders, A.W.M. (2004) ‘Content-based image retrieval: an overview’.

G. Medioni and S. B. Kang, editors, *Emerging Topics in Computer Vision*.

Grangier, D. and Bengio, S. (2008) ‘A discriminative kernel-based approach to rank images from text queries’. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.30 No.8, pp.1371–1384.

Griffiths, T.L. and Steyvers, M. (2004) ‘Finding scientific topics’. *Proceeding. The National Academy of Sciences*.

Griffiths, T.L., et al. (2005) ‘Integrating topics and syntax’. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems*.

- Hare, J.S and Lewis, P.H. (2005) ‘On Image Retrieval using Salient Regions with Vector-Spaces and Latent Semantics’, Image and Video Retrieval of the series Lecture Notes in Computer Science Vol.3568, pp 540-549
- Haralick, R. M., Shanmugam, K. and Dinstein, I. (1973) ‘Textural features for image classification’, IEEE Transactions on systems, man and cybernetics Vol.3 No.6, pp.610-621.
-
- Hartmann, G. (1987). ‘Recognition of Hierarchically Encoded Images by Technical and Biological Systems’, Biological Cybernetics, Vol. 57, pp.73-84.
- Goh, K.S. et al. (2004) ‘Multimodal concept-dependent active learning for image retrieval. Proceedings of the ACM International Conference on Multimedia’.
- HE, J. et al. (2004) ‘Mean version space: A new active learning method for content-based image retrieval’. Proceedings of the ACM International Workshop on Multimedia Information Retrieval (MIR).
- Hofmann, T. (2001) . ‘Unsupervised learning by probabilistic latent semantic analysis’. Machine Learning, Vol.42 No.1-2, pp.177–196.
- Hofmann, T., Puzicha, J., and Jordan, M.I. (1999) ‘Learning from dyadic data’. Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems.
- Hoi, C.H. and Lyu, M. R. (2004) ‘A novel log-based relevance feedback technique in content-based image retrieval’. Proceedings of the ACM International Conference on Multimedia.
- Hsu, W.H., Kennedy, L.S. and Chang, S.F. (2006) ‘Video search reranking via information bottleneck principle’. Proceedings of the 14th Annual ACM International Conference on Multimedia.
- Hörster, E. (2009) ‘Topic models for image retrieval on large-scale databases’. University of Augsburg, Augsburg, Germany.
- Huang, T.S. et al. (2002) ‘Learning in content-based image retrieval’, Proceedings of the International Conference on Development and Learning, pp. 155–162.

Jaimes, A. et al. (2004) ‘Memory cues for meeting video retrieval’. Proceedings of the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE) at the ACM International Multimedia Conference.

Jain, A.K. and Farrokhnia, F. (1990) ‘Unsupervised texture segmentation using gabor filters’. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics.

Jain, A.K. and Vailaya, A. (1990) ‘Image retrieval using color and shape’. Pattern Recognition, Vol.29 No.1, pp.1233–1244.

Jing, F. et al. (2004) ‘An efficient and effective region-based image retrieval framework’. IEEE Transactions on Image Processing, Vol.13 No.5, pp. 699–709.

Jing, Y. and S. Baluja, S. (2008) Page rank for product image search. Proceeding of the 17th International Conference on World Wide Web.

Jourdan, L. (2010) ‘Méta-heuristiques Coopératives : du déterministe au stochastique’, Habilitation à Diriger les Recherches de l'Université Lille I.

Ko, B. and H. Byun, H. (2002) ‘Probabilistic neural networks supporting multi-class relevance feedback in region-based image retrieval’. Proceedings of the 16th International Conference on Pattern Recognition (ICPR’02).

La Cascia *et al.*, (1998) ‘Combining Textual and Visual Cues for Content-based Image Retrieval on the World Wide Web’. *Proceeding of the IEEE Workshop on Content-based access of Image and Video Libraries*.

Larlus, D. and Jurie, F. (2006) ‘Latent mixture vocabularies for object categorization’. Proceeding of the British Machine Vision Conference.

Larochelle, H. (2009) ‘Étude de techniques d’apprentissage non-supervisé pour l’amélioration de l’entraînement supervisé de modèles connexionnistes’, Thèse de doctorat, Spécialité Informatique, Université de Montréal.

Leibe, B., Leonardis, A., and Schiele, B. (2004) ‘Combined object categorization and segmentation with an implicit shape model’. Workshop on Statistical Learning in Computer Vision.

Li, F.F. and P. Perona, P. (2005) ‘A Bayesian hierarchical model for learning natural scene categories’. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Li, J. and Wang, J.Z. (2008) ‘Real-time computerized annotation of pictures’. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.30 No.6, pp.985–1002.

Li, W. and McCallum, A. (2006) ‘Pachinko allocation: DAG-structured mixture models of topic correlations’, Proceedings of the 23rd International Conference on Machine Learning, Pennsylvania, USA, pp. 577-584.

Lienhart, R. and Effelsberg, W. (2006) ‘Automatic text segmentation and text recognition for video indexing’. Multimedia Systems, Vol.8 No.1, pp.69–81.

Lienhart, R. and Hartmann, A. (2002) ‘Classifying images on the web automatically’. Journal of Electronic Imaging, Vol.11 No.4, pp.445–454.

Lin, Y. et al. (2005). ‘Semantic manifold learning for image retrieval’. Proceedings of the ACM International Conference on Multimedia.

Liu, Y. et al. (2007) ‘A survey of content-based image retrieval with high-level semantics’, Pattern Recognition Society, Vol.40 No.1 , pp. 262 – 282.

Liu, D. and Chen, T. ‘Semantic-shift for unsupervised object detection’. Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition.

Low, W.C. and Chua, T.S. (1998) ‘Color-based relevance feedback for image retrieval, Proceedings of the International Workshop on Multimedia’, pp. 116–123.

Lowe, D. G. (2004) ‘Distinctive image features from scale-invariant keypoints’, International Journal of Computer Vision, Vol.60 No. 2, pp.91–110.

MacArthur, S.D., Brodley, C.E. and Shyu, C.R. (2000) ‘Relevance feedback decision trees in content-based image retrieval’, Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 68–72

Manzanera, A., ‘Traitement et reconnaissance d’images’, ENSTA/UEI.

Marref, N. (2013) ‘Apprentissage Incrémental & Machines à Vecteurs Supports’, Mémoire de Magister, Informatique Industrielle, Université Hadj Lakhdar, Batna.

Martin, D.R., Fowlkes, C.C. and Malik, J. (2004) ‘Learning to detect natural image boundaries using local brightness, color, and texture cues’. IEEE Transactions on Pattern Analysis and Machine Intelligence , Vol.26 No.5, pp.530–549.

Matas, J. et al. (2002) ‘Robust wide baseline stereo from maximally stable extremal regions’. In British Machine Vision Conference, pages 384–393, 2002.

Mender, B., Boulemden, A. and Tlili, Y. (2016) ‘Indexation et recherche des tumeurs de mammographie par LSA et caractéristiques globales de texture’, Mémoire de Master, Université Badji-Mokhtar Annaba.

Meskaldji, K. et al. (2009) ‘Color quantization and its impact on color histogram based image retrieval accuracy’, Proceeding of the First International Conference on Networked Digital Technologies.

Metzler, D., and Manmatha, R. (2004) ‘An Inference Network Approach to Image Retrieval’. Proceedings of the International Conference on Image and Video Retrieval.

Mezaris, V. et al. (2003) ‘An ontology approach to object-based image retrieval’, Proceedings of the International Conference on Information Processing.

Mimno, D., Li, W., and McCallum A. (2007) ‘Mixtures of hierarchical topics with pachinko allocation’. Proceedings of the 24th International Conference on Machine Learning.

Mitchell, T.M. (1997) , ‘Machine Learning’. McGraw-Hill International Editions.

Monay, F. and Gatica-Perez, D. (2004) ‘Plsa-based image auto-annotation: constraining the latent space’. Proceedings of the 12th Annual ACM International Conference on Multimedia.

Moutarde, F. (2001) ‘Apprentissage artificiel’ (« Machine-Learning »), Cours, Centre Robotique, MINES ParisTech (Ecole des Mines de Paris).

Mustière, S. (2001) ‘Apprentissage Supervisé pour la généralisation Cartographique’, Thèse de doctorat, Université Pierre et Marie Curie, Paris.

Monay, F., and Gatica, P.D. (2003) ‘On image auto-annotation with latent space models’. Proceedings of ACM Multimedia.

Moreira, L. M. (2000) ‘The use of boolean concepts in general classification contexts’. Thèse de doctorat, École polytechnique fédérale de Lausanne.

Mori, Y. et al., (1999) ‘Image-to-word transformation based on dividing and vector quantizing images with words’. In: International Workshop on Multimedia Intelligent Storage and Retrieval Management.

Mouassa, S., Boulemden, A. and Tlili, Y. (2015) ‘Indexation et recherche des tumeur de mammographie par LSA et caractéristiques locales’, Mémoire de Master, Université Badji-Mokhtar Annaba.

Nagamine, T. et al. (2004) ‘A Visuo-spatial Memory Cue System for Meeting Video Retrieval’. Proceedings of the ACM International Conference on Multimedia.

Nilson, N.J. (1998) ‘Introduction to Machine Learning’, Robotics Laboratory, Department of Computer Science, Stanford University.

Nister, D. and Stewenius, H. (2006) ‘Scalable recognition with a vocabulary tree’. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Obdržálek, S. and Matas, J. (2002) ‘Object recognition using local affine frames on distinguished regions’. British Machine Vision Conference.

Oliva, A. and Torralba, A.B. (2001) ‘Modeling the shape of the scene: A holistic representation of the spatial envelope’. International Journal of Computer Vision, Vol.42 No.3, pp.145–175.

Pass, G., Zabih, R., and Miller, J. (1996) ‘Comparing images using color coherence vectors’. Proceedings of the Fourth ACM International Conference on Multimedia.

Pearl, J (1984) 'Heuristics Intelligent Search Strategies for Computer Problem Solving', ADDISON-WESLEY Publishing Company, California.

Pham, TT. et al., (2007) 'Latent semantic fusion model for image retrieval and annotation'. In CIKM 2007. Proceedings of the sixteenth ACM Conference on information and knowledge management, Lisbon, Portugal, pp. 439-444.

Philbin, J. et al. (2007) 'Object retrieval with large vocabularies and fast spatial matching'. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Philbin, J. et al. (2008) 'Geometric LDA: A generative model for particular object discovery'. Proceedings of the British Machine Vision Conference.

Phuong, L.H. (2008) 'Modèles Heuristiques et Metaheuristiques en Comet pour le problème Bin Packing', institut de la francophonie pour l'informatique.

Quelhas, P. et al. (2005) 'Modeling scenes with local descriptors and latent aspects'. In ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, pages 883 890, 2005.

Ranzato, M. et al. (2007) 'Unsupervised learning of invariant feature hierarchies with applications to object recognition '. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Rui Y. and T. S. Huang T.S. (2001), 'Relevance feedback techniques in image retrieval', Springer Verlag, London, UK.

Rui, Y. T. Huang, S.and S.-F. Chang, S.F. (1999) 'Image retrieval: Current techniques, promising directions, and open issues'. Journal of Visual Communication and Image Representation, Vol.10 No.1, pp.39–62.

Russell, B.C. et al. (2006) 'Using multiple segmentations to discover objects and their extent in image collections'. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Rodden, K. and Wood, K.R. (2003) ‘How do people manage their digital photographs?’ Proceedings of the conference on Human factors in computing systems

Rubner, Y. et al. (1998) ‘A Metric for Distributions with Applications to Image Databases’, Proceedings of IEEE International Conference on Computer Vision.

Rui, Y. (1997) ‘A relevance feedback architecture in content-based multimedia information retrieval systems’. Proceedings IEEE Workshop Content-Based Access of Image and Video Libraries.

Schmid, C. and Mohr, R. (1997) ‘Local greyvalue invariants for image retrieval’. IEEE Transactions on Pattern Analysis and Machine Intelligence , Vol.19 No.1, pp.530–535.

Shavlik, J.W. and Dietterich, G.D . (1990) ‘Readings in Machine Learning’, Morgan Kaufmann.

Shechtman, E. and Irani, M. (2007) ‘Matching local self-similarities across images and videos’. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Sivic, J. et al. (2005) ‘Discovering objects and their location in images’. Proceedings of the International Conference on Computer Vision.

Sivic, J. et al.. (2008) ‘Unsupervised Discovery Of Visual Object Class Hierarchies’. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008.

Sivic, J. and Zisserman, A. (2003) ‘Video google: A text retrieval approach to object matching in videos’. Proceedings of the International Conference on Computer Vision.

Smeulders, A. W. M. et al. (2000) ‘Content-based image retrieval at the end of the early years ’, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.22 No.12, pp.1349–1380.

Smith, J.R. and Li, C.S. (1998) ‘Decoding image semantics using composite region templates’, IEEE Workshop on Content-Based Access of Image and Video Libraries.

Stanchev, P. L. et al. (2003) ‘High level color similarity retrieval’, International Journal of Information Theories and Applications, Vol.10 No.3, pp.363–369.

Steyvers, M. and Griffiths, T. (2007) ‘Latent Semantic Analysis: A Road to Meaning’, Laurence Erlbaum, chapter Probabilistic topic models.

M. A. Stricker and M. Orengo. Similarity of color images. In Storage and Retrieval for Image and Video Databases (SPIE), pages 381–392, 1995.

Su, Z. et al. (2003) ‘Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning’. IEEE Transaction on Image Processing. Vol.12 No.8, pp. 924–937

E. Sudderth, E. et al. (2006) ‘Describing visual scenes using transformed dirichlet processes’. Y. Weiss, B. Schölkopf, and J. Platt, editors.

Tamura, H., Mori, S. and Yamawaki, T. (1978) ‘Textural Features Corresponding to Visual Perception’, *IEEE Transactions on Systems, Man, and Cybernetics*, pp.460-473.

Tian, Q., Yu, Y. and Huang, T.S. (2000) ‘Incorporate discriminant analysis with EM algorithm in image retrieval’, Proceedings of the International Conference on Multimedia and Expo, pp. 299–302.

Tollari, S. (2006) ‘Indexation et recherche d’images par fusion d’informations textuelles et visuelles’, Thèse de doctorat, Université Sud-Toulon.

Tong, S. and Chang, E. (2001) ‘Support vector machine active learning for image retrieval’, Proceedings of the Ninth ACM International Conference on Multimedia.

Van Leuken, R.H. et al. (2009) ‘Visual diversification of image search results’. In 18th International World Wide Web Conference.

Verbeek, J. and Triggs, B. (2007) ‘Region classification with markov field aspect models’. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Viola, P. and Jones, M. (2001) ‘Rapid object detection using a boosted cascade of simple features’. Proceeding of the CVPR.

Vogel, J. and Schiele, B. (2004) ‘Natural scene retrieval based on a semantic modeling step’. Proceedings of CIVR.

Wang, G., Zhang, Y. and Fei-Fei, L. (2006) ‘Using dependent regions for object categorization in a generative framework’. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1597–1604, 2006.

Wang, J. Z. et al. (2001) ‘Unsupervised multiresolution segmentation for images with low depth of field’, IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.23 No.1, pp.85–90.

Wang, J.Z., Li, J. and Wiederhold, G. (2001) ‘SIMPLIcity: semantics-sensitive integrated matching for picture libraries’. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23 No.1, pp.947–963.

Wei, X. and Croft. W.B. (2006) ‘LDA-based document models for ad-hoc retrieval’. Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

Winn, J. and Jovic, N. (2005) ‘LOCUS: Learning Object Classes with Unsupervised Segmentation’, Proceeding of the IEEE International Conference on Computer Vision .

Wouwer, G.V.D., Scheunders, P. and D. V. Dyck, D.V. (1999) Statistical texture characterization from discrete wavelet representations. IEEE Transactions on Image Processing.

Wu, H. et al. (2004) 'Willhunter: Interactive image retrieval with multilevel relevance measurement'. Proceedings of the IEEE International Conference on Pattern Recognition.

Yang, C. et al. (2005) 'Semantic feedback for interactive image retrieval'. Proceedings of the ACM International Conference on Multimedia.

Yang, M. et al., (2008) 'A Survey of Shape Feature Extraction Techniques', Pattern Recognition Techniques, Technology and Applications, Book edited by: Peng-Yeng Yin, ISBN 978-953-7619-24-4, pp. 626, I-Tech, Vienna, Austria.

Zhai, C. and Lafferty, J. 'A study of smoothing methods for language models applied to adhoc information retrieval'. Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

D. Zhang and G. Lu, (2004) 'Review of shape representation and description techniques' Pattern Recognition, Vol. 37, pp. 1-19.

Zhang, D. et al. (2000) 'Content-based image retrieval using gabor texture features'. In IEEE Transactions PAMI.

Zhou, S. Rui, Y. and Huang, T. (2003) 'Exploration of Visual Data'. Kluwer Academic Publishers.

Zhou, X.S. and Huang, T. S. (2003) 'Relevance feedback in image retrieval: A comprehensive review'. Multimedia Systems, Vol.8, pp. 536-544.

Zhuang, Y. et al. (1999) 'Apply semantic template to support content-based image retrieval', Proceedings of the Storage and Retrieval for Media Databases.

Zuo, F. et al., (2005). 'Multistage Face Recognition Using Adaptative Feature Selection and Classification. Proceedings of Advanced Concepts for Intelligent Vision Systems.

Références Web

Web1, <http://www.flickr.com/>.

Web2, <http://www.marketingvox.com/flickr-photo-count-hits-3-billion-041831/>.

B Bouzy, 2005/2008, <http://math-info.univ-paris5.fr>

Web3, http://www.tsi.telecomparistech.fr/pages/enseignement/ressources/beti/correl_couleur/lab.html

Web4, <http://people.via.ecp.fr/~remi/ecp/tpi/rapport/yuv.html>

Web5, http://www.sites.univ-rennes2.fr/arts-spectacle/cian/image_numFlash/pdf/chap3_cours33.pdf

Web6, <http://savvash.blogspot.com/2008/12/benchmark-databases-for-cbir.html>, last access: November 2016.

Web7, https://en.wikipedia.org/wiki/Image_texture, last access: 26-05-2017.

Productions Scientifiques

▪ Publications

- 1- A. Boulemden, Y. Tlili and Hamid A. Jalab. “Content Based Image Retrieval with Pachinko Allocation Model and a Combination of Colour, Texture and Text Features”. International Journal of Computational Vision and Robotics, 2016. (Forthcoming article).

▪ Communications (nationales et internationales)

- 1- A. Boulemden and Y.Tlili, Indexation et recherche d’images par fusion d’informations textuelles et visuelles, Journées Gestion Electronique de Documents et Réseaux Recherche en Science et Technologies de l’Information (GED’09), Annaba, Algérie, 2009.
- 2- A.Boulemden and Y.tlili, Using Latent Semantic Analysis for image retrieval tasks, 1^{ère} Conférence National sur la Modélisation et la Simulation Informatique des Systèmes Industriels (MSISI’2010).
- 3- A.Boulemden and Y.tlili, Pachinko Allocation Model with Local Features for Image Indexing and Retrieval, First International Conference on Signal, Image, Vision and their Applications , Guelma University, Algeria, (SIVA’11), 2011a.
- 4- A.Boulemden and Y.tlili, Pachinko Allocation Model with Image Local Features for Image Retrieval Tasks ,12th International Arab Conference on Information Technology Naif Arab University, Riyadh, Saudi Arabia, (ACIT’11), 2011b.

5- A. Boulemden and Y. Tlili, "Content Based Image Retrieval with Pachinko Allocation Model: Application on local and global features", Knowledge Management and Acquisition for Intelligent Systems (Lecture Notes in Artificial Intelligence), Vol. 7457

http://link.springer.com/chapter/10.1007%2F978-3-642-32541-0_12, 2012 a.

6- A. Boulemden and Y. Tlili, Les 2^{èmes} Journées Doctorales en Informatique. Guelma 18-19 Novembre 2012 c.

ANNEXE A

Les Distributions statistiques

Distributions statistiques

Nous présentons ici quelques rappels sur les distributions statistiques utiles pour la compréhension de cette thèse. Il existe deux types de distributions les distributions continues (uniformes, gaussiennes, multi-gaussiennes, de Dirichlet...) et les distributions discrètes (binomiales, multinomiales...). La plupart des distributions continues sont également définies dans le cas discret.

Distribution uniforme

Une distribution est uniforme lorsque la variable aléatoire X peut prendre n'importe quelle valeur entre une borne minimale et une borne maximale avec la même probabilité. Par exemple, le jet d'un dé non-truqué présente une distribution uniforme, car la probabilité d'apparition de l'une des six faces est équiprobable. Dans le cas discret, on note

$P(X|\{1, 2, \dots, m\}) \sim Unif(\{1, 2, \dots, m\})$. L'expression « $P(X|\{1, 2, \dots, m\})$ suit une distribution uniforme sur l'intervalle $\{1, 2, \dots, m\}$ ».

Distribution gaussienne (ou normale)

Une variable aléatoire X de moyenne μ et d'écart type σ suit une loi normale (ou loi normale gaussienne, loi de Laplace-Gauss) si sa densité de probabilité est :

$$P(X|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{X - \mu}{\sigma}\right)^2\right) \quad (34)$$

Les points d'une distribution gaussienne ont tendance à se regrouper autour de la moyenne. Une distribution gaussienne possède 68%, respectivement 95%, 99.7%, de sa population, dans l'intervalle $|X - \mu| < \sigma$, respectivement $|X - \mu| < 2\sigma$ et $|X - \mu| < 3\sigma$. L'intervalle $|X - \mu| < \sigma$ est appelé «plage de normalité de niveau de confiance 68%». La figure 59 donne un exemple de distribution gaussienne. La distance r d'un point X de la moyenne dans l'unité de l'écart-type est calculée par :

$$r = \frac{|X - \mu|}{\sigma}. \quad (35)$$

On adopte parfois la loi normale, sans justification autre que sa forme uni-modale, symétrique et décroissante, comme approximation d'une population concentrée autour d'un point, mais dont on ne connaît pas de modèle générateur ou de loi a priori.

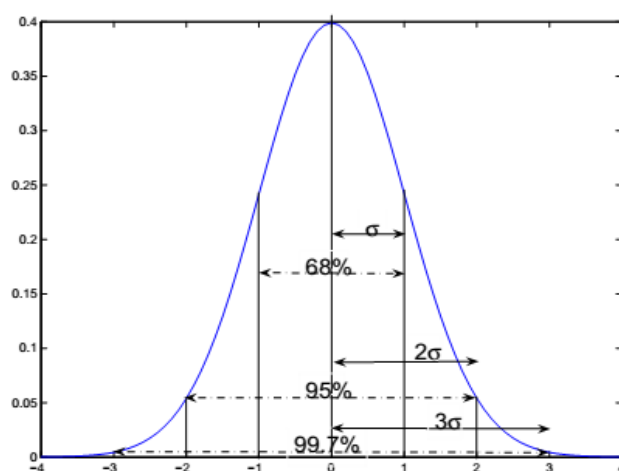


Figure 59 Distribution gaussienne (normale) d'écart-type 1 et de moyenne 0.

Distribution multi-gaussienne (multi-normale)

Soit $X^k = \{X_1, X_2, \dots, X_k\}$ un ensemble de variables aléatoires. Chaque variable X_i est associée au couple (e_i, σ_i) . Notons $\mu_{1:r} = (\mu_1, \mu_2, \dots, \mu_k)$ le vecteur des moyennes, $\sigma_{1:r} = (\sigma_1, \sigma_2, \dots, \sigma_k)$ le vecteur des écarts-types et $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$ la matrice diagonale des covariances. La distribution multi-gaussienne est de la forme :

$$P(X^k | \mu_{1:r}, \sigma_{1:r}) = \frac{1}{\det(\Sigma)^{\frac{1}{2}} (\sqrt{2\pi})^k} \exp\left(-\frac{1}{2} (X^k - \mu_{1:r})^t \Sigma^{-1} (X^k - \mu_{1:r})\right). \quad (36)$$

Distribution de Dirichlet

La distribution de Dirichlet est définie dans le cas de variables discrètes ou continues, mais nous ne considérons que le cas discret. La distribution de Dirichlet est la conjuguée *a priori* (*conjugate prior*) de la distribution multinomiale en statistique Bayésienne. Elle est de la forme :

$$P(p_1, p_2, \dots, p_m | \alpha_1, \alpha_2, \dots, \alpha_m) = \frac{1}{Z(\alpha_1, \alpha_2, \dots, \alpha_m)} \prod_{i=1}^m p_i^{\alpha_i - 1} \quad (37)$$

avec $Z(\alpha_1, \alpha_2, \dots, \alpha_m)$ une constante de normalisation telle que :

$$Z(\alpha_1, \alpha_2, \dots, \alpha_m) = \frac{\prod_{i=1}^m \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^m \alpha_i)} \quad (38)$$

et Γ est la fonction gamma.

La figure 60 montre Courbes représentant l'une des dimensions d'une distribution de Dirichlet.

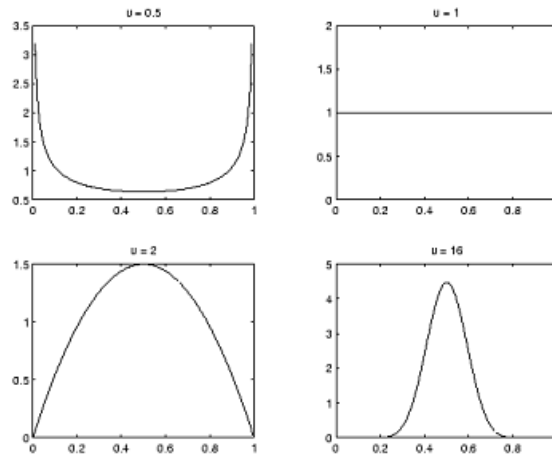


Figure 60. Courbes représentant l'une des dimensions d'une distribution de Dirichlet à deux dimensions ($m = 2$) pour différentes valeurs de u ($u = \alpha_1 = \alpha_2$).

ANNEXE B

L'implémentation du PAM

L'implémentation du PAM

Cette annexe est dédiée à la représentation d'une partie du code source java d'implémentation du modèle PAM (Pachinko Allocation Model). Cette implémentation a été réalisé par David Mimno et elle est basée sur l'implémentation du modèle LDA (Latent Dirichlet Allocation) par Andrew McCallum. Elle utilise le MLE (Maximum Likelihood Estimation) pour modéliser un PAM de quatre niveaux dont deux niveaux intérieures pour les sous-sujets et les super-sujets.

La classe est nommée PAM4L et la partie du code illustrée par la figure 61 présente ses attributs.

```
// Parameters
int numSuperTopics; // Number of topics to be fit
int numSubTopics;

double[] alpha; // Dirichlet(alpha,alpha,...) is the distribution over supertopics
double alphaSum;
double[][] subAlphas;
double[] subAlphaSums;
double beta; // Prior on per-topic multinomial distribution over words
double vBeta;

// Data
InstanceList ilet; // the data field of the instances is expected to hold a FeatureSequence
int numTypes;
int numTokens;

// Gibbs sampling state
// (these could be shorts, or we could encode both in one int)
int[][] superTopics; // indexed by <document index, sequence index>
int[][] subTopics; // indexed by <document index, sequence index>

// Per-document state variables
int[][] superSubCounts; // # of words per <super, sub>
int[] superCounts; // # of words per <super>
double[] superWeights; // the component of the Gibbs update that depends on super-topics
double[] subWeights; // the component of the Gibbs update that depends on sub-topics
double[][] superSubWeights; // unnormalized sampling distribution
double[] cumulativeSuperWeights; // a cache of the cumulative weight for each super-topic

// Per-word type state variables
int[][] typeSubTopicCounts; // indexed by <feature index, topic index>
int[] tokensPerSubTopic; // indexed by <topic index>
```

```

// [for debugging purposes]
int[] tokensPerSuperTopic; // indexed by <topic index>
int[][] tokensPerSuperSubTopic;

// Histograms for MLE
int[][] superTopicHistograms; // histogram of # of words per supertopic in documents
// eg, [17][4] is # of docs with 4 words in sT 17...
int[][][] subTopicHistograms; // for each supertopic, histogram of # of words per subtopic

Runtime runtime;
NumberFormat formatter;

```

Figure 61. Les attributs de la classe PAM4L

La classe PAM4L implémente deux constructeurs. Le premier constructeur reçoit deux paramètres qui représentent les nombres des super-sujets et des sous-sujets. En plus de ces deux paramètres, le deuxième constructeur reçoit deux paramètres du modèle qui sont la somme des α ainsi que le β . La figure 62 contient le code des deux constructeurs de la classe PAM4L.

```

public PAM4L (int superTopics, int subTopics) {
    this (superTopics, subTopics, 50.0, 0.001);
}

public PAM4L (int superTopics, int subTopics,
             double alphaSum, double beta) {
    formatter = NumberFormat.getInstance();
    formatter.setMaximumFractionDigits(5);

    this.numSuperTopics = superTopics;
    this.numSubTopics = subTopics;

    this.alphaSum = alphaSum;
    this.alpha = new double[superTopics];
    Arrays.fill(alpha, alphaSum / numSuperTopics);

    subAlphas = new double[superTopics][subTopics];
    subAlphaSums = new double[superTopics];

    // Initialize the sub-topic alphas to a symmetric dirichlet.
    for (int superTopic = 0; superTopic < superTopics; superTopic++) {
        Arrays.fill(subAlphas[superTopic], 1.0);
    }
    Arrays.fill(subAlphaSums, subTopics);

    this.beta = beta; // We can't calculate vBeta until we know how many word types...

    runtime = Runtime.getRuntime();
}

```

Figure 62. Les deux constructeurs de la classe PAM4L

L'estimation des super-sujets et des sous-sujets des documents se fait par le moyen de la méthode `estimate`. Une partie du code de cette méthode est illustrée par la figure 63. L'estimation peut être testée sur un nombre réduit d'itération (par exemple 50 itérations). Dans le cas des applications sur des corpus de données importants, le nombre minimum d'itérations conseillé par les développeurs est égale à 1000 itérations.

```

public void estimate (InstanceList documents, int numIterations, int optimizeInterval,
                    int showTopicsInterval,
                    int outputModelInterval, String outputModelFilename,
                    Randoms r)
{
    ilist = documents;
    numTypes = ilist.getDataAlphabet().size ();
    int numDocs = ilist.size();
    superTopics = new int[numDocs][];
    subTopics = new int[numDocs][];

    //      Allocate several arrays for use within each document
    //      to cut down memory allocation and garbage collection time

    superSubCounts = new int[numSuperTopics][numSubTopics];
    superCounts = new int[numSuperTopics];
    superWeights = new double[numSuperTopics];
    subWeights = new double[numSubTopics];
    superSubWeights = new double[numSuperTopics][numSubTopics];
    cumulativeSuperWeights = new double[numSuperTopics];

    typeSubTopicCounts = new int[numTypes][numSubTopics];
    tokensPerSubTopic = new int[numSubTopics];
    tokensPerSuperTopic = new int[numSuperTopics];
    tokensPerSuperSubTopic = new int[numSuperTopics][numSubTopics];
    vBeta = beta * numTypes;

    long startTime = System.currentTimeMillis();

    int maxTokens = 0;

```

Figure 63. Une partie du code de la méthode d'estimation dans la class PAM4L

La figure 64 présente un exemple d'utilisation de la classe PAM4L pour estimer les sujets (sous et super) sur 50 itérations. Il est à noter que les paramètres des constructeurs sont reçu de la part de l'utilisateur par le moyen de la ligne de commande, c'est le mode shell de l'utilisation du package Mallet. Les développeurs peuvent utilisés directement la classe dans un mode de développement.

```
// Recommended to use mallet/bin/vectors2topics instead.
public static void main (String[] args) throws IOException
{
    InstanceList  ilist = InstanceList.load (new File(args[0]));
    int numIterations = args.length > 1 ? Integer.parseInt(args[1]) : 1000;
    int numTopWords = args.length > 2 ? Integer.parseInt(args[2]) : 20;
    int numSuperTopics = args.length > 3 ? Integer.parseInt(args[3]) : 10;
    int numSubTopics = args.length > 4 ? Integer.parseInt(args[4]) : 10;
    System.out.println ("Data loaded.");
    PAM4L pam = new PAM4L (numSuperTopics, numSubTopics);
    pam.estimate (ilist, numIterations, 50, 0, 50, null, new Randoms()); // should be 1100
    pam.printTopWords (numTopWords, true);
    pam.printDocumentTopics (new File(args[0]+".pam"));
}
```

Figure 64. Exemple d'utilisation de la classe PAM4L