

**République Algérienne Démocratique et Populaire**

**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**



**Université El-Hadj Lakhdar – Batna**

**Faculté des Sciences de l'Ingénieur**

**Département d'Informatique**

**Mémoire**

**Présenté par :**

Mme. Leïla BOUSSAAD

En vue de l'obtention du diplôme de Magister en Informatique

Option : Informatique Industrielle

**THEME :**

**RECHERCHE DANS LES BASES  
DE DONNEES IMAGES**

**Soutenu le 31/01/2009 devant le jury composé de :**

<b>Dr. A. Bilami</b>	Université de Batna	Président de jury
<b>Pr. M. Benmohammed</b>	Université de Constantine	Rapporteur
<b>Dr. A. Zidani</b>	Université de Batna	Examineur
<b>Dr. B. Belattar</b>	Université de Batna	Examineur

## *Remerciements*

*Tous mes remerciements s'adressent tous d'abord à tout puissant ALLAH, d'avoir guidé mes pas vers le chemin du savoir.*

*Je tiens à remercier Monsieur M. Benmohammed, Professeur à l'université de Constantine, pour avoir accepté d'encadrer ce mémoire et aussi pour sa disponibilité et ses précieux conseils qui m'ont permis d'aboutir dans ce travail.*

*Je tiens à remercier également Monsieur A. Bilami, Docteur à l'université de Batna, pour l'honneur qu'il m'a fait en acceptant la présidence de ce jury.*

*Je remercie vivement Monsieur B. Belattar, Maître de conférence à l'université de Batna, et Monsieur A. Zidani, Maître de conférence à l'université de Batna, pour avoir accepté la lourde charge d'évaluer ce mémoire et d'en être examinateurs.*

*Enfin, que toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail trouvent ici l'expression de mes sincères remerciements*

## *Dédicaces*

*Je tiens à dédier ce modeste travail à :*

*Mes DEFUNTS PARENTS, que Dieu leur accorde une place au Paradis.*

*Mon EPOUX MOHAMED, ainsi que mon fils MOUSLIM.*

*Mes Frères et sœurs.*

*Ma très chère cousine AMEL.*

*Mes collègues*

*Toute personne que je connais de près ou de loin.*

## **RESUME**

Ce mémoire s'intéresse aux systèmes de recherche d'images fixes par le contenu. Dans de tels systèmes, les images peuvent être décrites au niveau pixel par des descripteurs visuels globaux de couleur, de texture ou de forme. La recherche par le contenu exploite et combine alors ces descripteurs dont le coût de calcul est d'autant plus important que la taille de la base d'images est grande. Les résultats de la recherche sont ensuite classés en fonction de leur similarité à la requête soumise et présentés à l'utilisateur sous forme de liste ordonnée. Un sous-ensemble de descripteurs pourrait cependant suffire à répondre à une recherche par similarité beaucoup plus rapidement, tout en gardant une qualité acceptable des résultats de recherche.

Pour améliorer les performance d'un système de recherche par le contenu, nous avons présenté un schéma d'indexation, le principe de base de ce schéma est de partitionner l'ensemble des descripteurs en un nombre fixe de clusters, chaque cluster est représenté par son centre de gravité. Lors d'une recherche, seulement un petit nombre de clusters qui ont des représentants proches aux descripteurs requête qui sont considérés pour construire l'ensemble des résultats, et cela dans le but de réduire efficacement la quantité de données à traiter.

Nous présentons, également, une technique qui vise à améliorer l'efficacité d'un système de recherche, celle-ci est basée sur une combinaison pondérée de plusieurs descripteurs MPEG-7. Dans cette approche, un algorithme d'optimisation par essaim de particules est utilisé comme un outil pour ajuster les poids des descripteurs selon différents aspects et cela dans le but d'obtenir un meilleur résultat.

### **Mots clés**

Recherche d'images par le contenu, Image requête, Combinaison de descripteurs globaux , Optimisation par essaim particulaires, Indexation d'images.

## **ABSTRACT**

This thesis is concerned with content-based image retrieval (CBIR) systems, in such systems, still images can be retrieved by similarity searching on global visual features such as color, texture, or shape at the pixel level. Content-based retrieval systems then use these available low-level features whose computing cost can be prohibitive and they rank the images according to how well they match the submitted query-by-example. Finally, they return the best few matches to the user in a ranked result list. But, a subset of features could be sufficient enough to answer very quickly while offering an acceptable quality of results. Moreover, the administration of large collections of images accentuates the classical problems of indexing and efficiently querying information.

To improve content-based image retrieval systems. We have proposed an indexing scheme for content-based image retrieval, The basis for the proposed method is the partitioning of the feature dataset into clusters that are represented by single prototype. Upon submission of a query request, only a small number of clusters whose prototypes are close to the query object are considered for the approximate query result, effectively cutting down the amount of data to be processed enormously, and speed up the content-based retrieval on large still image databases.

In this thesis, we also present a method for obtaining a far advanced retrieval function by using a plurality of weights. In order to realize such functions, MPEG-7 is employed with the aim to provide fundamental tools for describing image contents and particles swarm optimization algorithm is used to optimize the retrieval feature spaces so as to make the retrieval universal and to get the best result set.

### **Keywords:**

Content-based image retrieval (CBIR), Query image, Combination of global features, Particles swarm optimization, Images indexing.

---



---

## Liste des figures

---



---

Figure 1.1	architecture d'un système de recherche d'images par le contenu-----	P. 14
Figure 2.1	Exemple d'image en niveaux de gris et d'image couleur-----	P. 19
Figure 2.2	Exemple de requête globale d'image avec Ikona-----	P. 20
Figure 2.3	Exemple de description R.V.B-----	P. 22
Figure 2.4	Exemple de description H.S.V-----	P. 22
Figure 2.5	Exemples d'espaces de couleurs-----	P. 23
Figure 2.6	Images retournées pertinentes/non pertinentes-----	P. 29
Figure 2.7	Courbe de rappel et précision du système Ikona-----	P. 30
Figure 4.1	Principe de classification-----	P. 54
Figure 4.2	Exemple d'une distribution de données en 3D et un regroupement Associé-----	P. 57
Figure 4.3	Exemple d'utilisation des nuées dynamiques pour grouper un espace 2D en deux clusters-----	P. 60
Figure 4.4	Traitement des points isolés-----	P. 63
Figure 5.1	Images similaires-----	P. 71
Figure 5.2	Amélioration de la qualité de la recherche par similarité des objets 3D à travers la combinaison de deux descripteurs 3D-----	P. 72
Figure 5.3	Trois topologies différentes-----	P. 77
Figure 6.1	Exemples d'images des bases avec vérité terrain COIL-20 (images noir et blanc) et COIL-100 (images couleur)-----	P. 90
Figure 6.2 :	Exemples d'images de la base avec vérité terrain Wang-----	P. 90
Figure 6.3 :	Exemples d'images de la base avec vérité terrain CURET-----	P. 91
Figure 6.4 :	Exemples d'images de la base avec vérité terrain FeiFei-----	P. 91
Figure 6.5:	Temps de réponse en fonction du nombre de clusters-----	P. 96
Figure 6.6:	Temps de réponse en fonction du nombre d'images-----	P. 97
Figure A.1:	Schéma de description MPEG-7 d'un document multimédia-----	P. 106
Figure A.2:	Exemple d'image à décrire avec MPEG-7-----	P. 107
Figure A.3:	Exemple de descripteur MPEG-7 de type Scalable Color-----	P. 108
Figure A.4:	Exemple de descripteur MPEG-7 de type Color Layout-----	P. 109
Figure A.5:	Exemple de descripteur MPEG-7 de type Edge Histogramme-----	P. 111

---

---

## Liste des algorithmes

---

---

- Algorithme 4.1 Pseudo algorithme du voisinage dense----- P. 58
- Algorithme 4.2 Pseudo algorithme de regroupement hiérarchique----- P. 59
- Algorithme 4.3 Pseudo algorithme des centres mobiles----- P. 59
- Algorithme 4.4 Pseudo algorithme de l'algorithme k-means modifié-- P. 64
- Algorithme 5.1 Algorithme de base (Sans voisinage) ----- P. 78
- Algorithme 5.2 Algorithme avec un voisinage en étoile----- P. 79

---

---

## Liste des tableaux

---

---

Tableau 6.1 : Les types de descripteurs de la collection d'images Corel utilisés---	P. 92
Tableau 6.2 : Les types de descripteurs MPEG-7 utilisés-----	P. 93
Tableau 6.3 : F-mesure par type de descripteur-----	P. 94
Tableau 6.4 : F-mesure du système avec combinaison linéaire non pondérée de plusieurs types de descripteurs-----	P. 95
Tableau 6.5 : F-mesure du système avec combinaison linéaire pondérée de plusieurs types de descripteurs-----	P. 95



---

---

## Table des matières

---

---

<b>Chapitre 1: Introduction générale</b> -----	<b>P. 09</b>
1.1 Présentation du problème-----	P. 10
1.2 Architecture d'un système de recherche d'image par le contenu	P. 12
1.3 Objectifs du mémoire-----	P. 14
1.4 structure du mémoire-----	P. 15
1.5 Références bibliographiques-----	P. 17
<b>CHAPITRE 2: Principe de la description d'images par le contenu</b> -----	<b>P. 18</b>
2.1 Introduction-----	P. 19
2.2 Description des images fixes -----	P. 21
2.2.1 Recherche sur la couleur-----	P. 21
2.2.2 Recherche sur la texture-----	P. 24
2.2.3 Recherche sur la forme-----	P. 25
2.3 Combinaison des types de descripteurs globaux-----	P. 27
2.4 Les descripteurs MPEG-7-----	P. 27
2.5 Mesures pour évaluer un système de recherche d'images par le contenu-----	P. 28
2.6 Conclusion-----	P. 31
2.7 Références bibliographiques-----	P. 32
<b>CHAPITRE 3: Etat de l'art des méthodes d'indexation Multidimensionnelles</b> -----	<b>P. 35</b>
3.1 Introduction-----	P. 36
3.2 Recherche par similarité-----	P. 37
3.3 Index multidimensionnels-----	P. 38
3.3.1 Partitionnement de données-----	P. 38
3.3.2 Partitionnement de l'espace-----	P. 39
3.3.3 Malédiction de la dimension-----	P. 41
3.4 Nouvelles approches-----	P. 42
3.4.1 Les techniques de filtrage-----	P. 42
3.4.2 Recherche approximative des plus proches voisins----	P. 44
3.4.2.1 Approximation de la représentation des données	P. 44
3.4.2.2 Approximation de la technique de recherche-----	P. 45

3.5 Conclusion-----	P. 47
3.6 Références bibliographiques-----	P. 48

#### **CHAPITRE 4: Schéma d'indexation des bases d'images fixes P. 51**

4.1 Introduction-----	P. 52
4.2 Classification et regroupement (clustering) -----	P. 53
4.3 Distances et mesures de similarité-----	P. 55
4.4 Clustering: principe et applications-----	P. 56
4.4.1 Méthodes de clustering-----	P. 58
4.4.1.1 Approche basée sur la densité-----	P. 58
4.4.1.2 Approche hiérarchique-----	P. 58
4.4.1.3 Approche centres mobiles-----	P. 59
4.4.2 Application à l'indexation des images-----	P. 60
4.5 Schéma d'indexation proposé-----	P. 61
4.6 Amélioration proposée-----	P. 63
4.6.1 Traitement des points isolés-----	P. 63
4.6.2 Fixation de la taille des clusters-----	P. 64
4.7 Mise à jour de l'index-----	P. 65
4.8 Etapes d'organisation des images-----	P. 65
4.9 Conclusion-----	P. 66
4.10 Références bibliographiques-----	P. 67

#### **CHAPITRE 5: La recherche par similarité dans un espace multi-métriques P. 68**

5.1 Introduction-----	P. 69
5.2 Métrique simple vs. Métrique multiple-----	P. 70
5.3 Combinaison de plusieurs métriques-----	P. 73
5.4 Combinaison dynamiques de plusieurs métriques-----	P. 74
5.4.1 Construction de l'espace composé-----	P. 74
5.5 Sélection des poids appropriés basée sur un algorithme d'optimisation par essaim de particules-----	P. 76
5.5.1 Optimisation par essaim de particules-----	P. 76
5.5.1.1 Notion de voisinage-----	P. 76
5.5.1.2 Algorithmes-----	P. 77
5.5.2 Comparaison entre un algorithme génétique et un OEP	P. 80
5.5.3 Application de l'OEP pour la sélection des poids-----	P. 81
5.6 Etapes de traitement d'une requête par l'exemple-----	P. 82
5.7 Conclusion-----	P. 83

5.8 Références bibliographiques-----	P. 84
<b><u>CHAPITRE 6: Evaluations expérimentales</u></b> -----	<b>P. 86</b>
6.1 Introduction-----	P. 87
6.2 Environnement expérimental-----	P. 88
6.3 Choix des bases de données-----	P. 89
6.4 Descripteurs-----	P. 92
6.5 Résultats expérimentaux-----	P. 93
6.5.1 Précision de la recherche-----	P. 93
6.5.2 Influence du nombre de clusters sur les performances du système-----	P. 95
6.5.3 Influence de la taille de la base des descripteurs sur les performances du système-----	P. 97
6.6 Conclusion-----	P. 98
6.7 Références bibliographiques-----	P. 99
<b><u>Conclusion générale</u></b> -----	<b>P. 101</b>
<b><u>Annexe A Les descripteurs MPEG-7</u></b> -----	<b>P. 105</b>
<b><u>Annexe B VIZIR</u></b> -----	<b>P. 113</b>