

**LA REPUBLIQUE ALGERIENNE DEMOCRATIQUE
POPULAIRE**

LA FACULTE DES SCIENCES

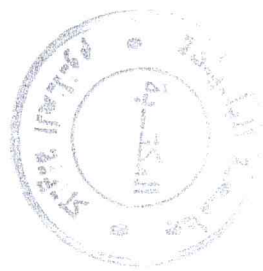
LE DEPARTEMENT DE MATH INFORMATIQUE

**Conception Et Réalisation D'un Système
D'auto Annotation D'image Par Apprentissage
Pour La Recherche MultiMedia**

Présenter par : EDDALIA Asma & RABET Amina

Promotrice : BENBLIDIA Nadjia

Encadreur : BAKALEM Mahdia



Les jurés : M^r chef Zahar
M^{me} Reguig

Année universitaire : 2008/2009

MA-004-07-1

Remerciements

Nous remercions avant tout le bon Dieu qui nous a donnée la force et la volonté pour pouvoir réaliser ce modeste travail.

Nous tenons à remercier vivement notre promotrice Mlle BENBLIDIA Nadjia et nous exprimons toutes nos gratitudees et nos sincères remerciements pour leur coopération, aide et patience et pour nous avoir donnée l'honneur de travailler avec elle.

Priant Dieu de lui donner toute la force pour continuer sa noble mission.

Un grand remerciement va également à Mlle BAKALEM Mahdia, d'avoir accepté de nous encadrer.

Nous profitons l'occasion de remercier tous les enseignants du département d'informatique, pour nous avoir formés tout au long de notre cursus, et nous espérons être à la hauteur de leur espérance dans le futur proche.

Merci à tous ceux qui de prêt ou de loin, ont fait que l'aboutissement de ce travail soit possible.



Dédicaces

« Louange à Dieu, le tout puissant »

A ma chère mère et mon cher père ;

En reconnaissance de leur soutien indéfectible et ininterrompu,

Que Dieu les garde !

A mes chères frères, mes très chères sœurs,

A mes très chères amies,

A tous mes collègues de la première

Promotion en Informatique (LMD)

Je dédie ce modeste travail avec l'expression de tous mes sentiments

D'affection et de respect !

A tous ceux que je n'ai pas cités par oubli, qu'ils me pardonnent.

PFE 2009

ASMA

Dédicaces

« Louange à Dieu, le tout puissant »

A ma chère mère et mon cher père ;

En reconnaissance de leur soutien indéfectible et ininterrompu,

Que Dieu les garde !

A mes chères frères, mes très chères sœurs,

A mes très chères amies,

A tous mes collègues de la première

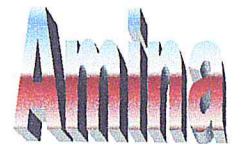
Promotion en Informatique (LMD)

Je dédie ce modeste travail avec l'expression de tous mes sentiments

D'affection et de respect !

A tous ceux que je n'ai pas cités par oubli, qu'ils me pardonnent.

PFE 2009



Résumé

L'annotation d'image est l'opération qui consiste à attribuer à une image des mots clés qui décrivent son contenu. C'est l'outil principal pour associer une sémantique à une image. Il est clair que cette tâche ne peut pas être confiée à un humain surtout que les images constituent le média le plus stocké avec le texte. On imagine bien le temps et le coût financier que cela peut représenter. L'idée d'automatiser cette tâche semble alors légitime et tout à fait logique.

Dans ce travail, nous présentons une approche pour réaliser un système d'auto-annotation en utilisant un modèle d'espace latent qui a donné des résultats encourageants dans le domaine de la recherche d'information sur des documents textuels : l'Analyse de la Sémantique Latente (LSA). Cette méthode, comme nous allons le voir dans la suite, constitue une piste prometteuse pour améliorer les performances dans ce domaine.

L'approche proposée se compose de deux étapes principales :

Dans la première partie l'apprentissage :

- Segmentation ;
- Extraction des caractéristiques ;
- Classification ;

La deuxième partie le traitement : consiste en le système d'aide automatique d'annotation de propagation pour une nouvelle image.

Liste des figures

Figure1.1 : Composants du multimédia.....	4
Figure1. 2. Schéma d'un système classique de recherche d'information.....	8
Figure1.3. Schéma Système de recherche d'image(SRIm).....	9
Figure2.1. La Décomposition en Valeurs Singulières.....	27
Figure2.2. Réduction des dimensions des matrices.....	28
Figure2.3.Auto-annotation des images avec LSA.....	29
Figure3.1. Architecture du système d'auto- annotation – LSA.....	32
Figure 3.2. Processus d'apprentissage.....	33
Figure3.3.schéma d'annotation.....	37

Sommaire

Introduction Générale

Introduction Générale.....	1
----------------------------	---

Chapitre I : Généralité sur multimédia et les systèmes de recherche

I. Multimédia	4
II. Les composants multimédia.....	5
II.1. Le texte.....	5
II- 3. Le Son.....	5
II-4. La vidéo.....	5
II.2. L'image.....	5
II.2.1. Image numérique.....	5
II.2.2. Les types d'images.....	6
II.2.3. Les formats d'images.....	7
II.2.4. Caractéristiques d'image.....	7
III-Les systèmes de recherche	8
III.1. Système de recherche d'information.....	8
III.2. Système de recherche d'image(SRIm).....	9
III.2.1. Recherche d'information basée sur les annotations.....	10
III.2.2. Recherche d'information basée sur le contenu d'image.....	10
III. Quelques systèmes de recherche d'images.....	13

Chapitre II: Etat de l'art sur annotation d'image par l'apprentissage

II.1 Indexation.....	14
II.2. Indexation et recherche textuelle d'images.....	15
II.2.1. Indexation textuelle manuelle.....	15
II.2.2 Indexation textuelle automatique.....	16
II.2.3.Les modèles classiques de la recherche textuelle.....	17
II.3 Indexation et recherche d'image par le contenu visuel.....	19
II.3.1.Les types des systèmes.....	19
II.3.2.Les bases d'images.....	19

IV. La Segmentation.....	20
V. Mesure de similarité.....	21
VI. Apprentissage automatique.....	23
VI.1. Techniques d'apprentissage automatique.....	24
VII. Annotation d'image.....	25
VII.1. Définition d'annotation d'image.....	25
VII.2. Types d'annotation d'image.....	25
1. L'annotation manuelle d'images.....	25
2. L'annotation automatique d'images	25
VII.3 Un modèle par analyse de la sémantique latente LSA	26
VII.3.1. La description théorique de la Méthode LSA.....	26
VII.3.2. Principe de la Méthode LSA.....	26
VII.3.3. Moyens mathématiques mis en œuvre : la SVD.....	27

Chapitre III : Conception

III.1. Présentation du sujet.....	30
III.2. Description du corpus.....	31
III.3. Architecture du système d'auto- annotation – LSA.....	32
III. 4. Le processus d'annotation d'image.....	33
III.4.1. La partie d'apprentissage.....	33
III.4.2. La partie de traitement.....	37
III.5. La modélisation du système D'auto- annotation- LSA	38
III.5.1. Définition de l'UML.....	38
III.5.2. Le diagramme de classe.....	39
III.5.3. Les diagrammes de cas d'utilisation.....	40
III.5.3.1.Diagramme de cas d'utilisation pour l'étape d'apprentissage.....	40
III.5.3.2.Diagramme de cas d'utilisation pour l'étape de Traitement.....	43

Chapitre IV: IMPLEMENTATION

IV.1. Présentation du MATLAB.....	45
IV.2. Les résultats du système d'auto-annotation-LSA.....	45
IV.2.1.La segmentation.....	45
IV.2.2. Extraction des caractéristiques (la texture).....	46
IV.2.3. La classification des caractéristiques (variance, uniformité).....	48
IV.2.4.LSA méthode LSA.....	49
Conclusion générale.....	53
Référence.....	54



Introduction générale

Introduction Générale

Une image vaut mieux que beaucoup de mots » ! Une image en effet de décrire un coucher de soleil, le sourire d'un vainqueur olympique, un repas de famille ... bien mieux que beaucoup de mots .C'est pourquoi les médias, les encyclopédies les simples particuliers utilisent les images ont donc souvent plus de sens que de longues phrases. Or, du fait du développement des performances de transfert des réseaux et de stockage des systèmes, on assiste ces dernières années à une véritable explosion du nombre d'image auxquelles un utilisateur souhaite accéder. Ce dernier a donc besoin d'outils pour pouvoir retrouver rapidement et efficacement les images qu'il recherche parmi la grande masse le cadre de cette thèse, nous considérons seulement les systèmes de recherche d'images(SRIm).

Les systèmes de recherche d'images par le contenu visuel (CBIR) ont fait ces dernière années de progrès pour ce qui est de rechercher des images visuellement proches d'une certaine image requête, ou bien pour retrouver un objet spécifique dans une image .Cependant, les systèmes de recherche d'images actuels sont toujours peu performants en ce qui concerne la recherche sémantique d'images sur des bases généraliste. Une des raisons à cela vient de la façon dont les images sont décrites sur les systèmes informatique .Le contenu d'une image peut être décrit à deux niveaux :

- Au niveau « numérique », une image contient des pixels colorés desquels on peut extraire des descripteurs de couleurs, de textures et de formes....,
- Au niveau « sémantique », une image peut être interprétée, elle au moins une signification.

Mais malheureusement, dans les systèmes d'information actuels, les

images sont décrites au niveau numérique alors que les utilisateurs sont intéressés par leur contenu sémantique, et il est actuellement difficile de trouver des correspondances entre le niveau numérique et le niveau sémantique, c'est ce que l'on appelle le « fossé sémantique ».

Pour comprendre la difficulté que les systèmes on peut trouver des correspondances entre le niveau numérique et le niveau sémantique.

Une manière d'ajouter un niveau sémantique à des images est de les décrire avec des mots. De plus, les mots possèdent la propriété d'avoir une signification hors contexte, et de posséder un sens en contexte qui correspondent plus ou moins au sens hors contexte de correspondances entre ces images. Malgré certains problème de subjectivité lors du choix des mots, annoter manuellement des mots à des images devient très vite fastidieux pour l'utilisateur.

Malheureusement, les méthodes d'indexation textuelle automatique (auto annotation) d'images à partir du contenu visuel ne sont pas efficaces. Mais, par contre, nous disposons de bases d'images telles que les encyclopédies, les revues ou le web où les images sont associées à du texte. Ainsi les moteurs de recherche d'images sur le web permettent de rechercher des images à l'aide de mots-clés grâce aux mots des pages web où se situent les images. Cependant, les mots de ces textes n'ont pas tous un rapport avec le contenu visuel des images : les images sont mal annotées.

Nous voyons que l'information fournie par le niveau numérique (description visuel) et celle fournie par le niveau sémantique (mots clés) apportent toutes les deux une information différente et complémentaire: l'information visuelle indique le contenu. Dans cette thèse, nous cherchons à trouver des liens entre ces deux informations dans le but d'améliorer les systèmes de recherche d'images.

En effet, les méthodes de traitement d'images ne font qu'associer à chaque image un vecteur (ou des vecteurs) de caractéristiques calculées à la base des caractéristiques d'images dites de bas niveau (couleur, texture, forme etc.).

L'interrogation d'une base de données d'images se fait alors en introduisant une image requête dans le système et la comparaison des

caractéristiques ainsi calculées en utilisant une mesure des similarités. Aucune sémantique n'est donc associée à ce processus.

L'une des manières possibles pour l'assignation d'une sémantique à une image est l'annotation des images, cette tâche consiste à attribuer, à chaque image un mot clé ou une liste de mots clés permettant de décrire son contenu sémantique. Autrement dit c'est de trouver un mécanisme permettant de faire une sorte de correspondance entre les aspects visuels et l'aspect sémantique de l'image.

Dans ce travail, nous allons tenter de répondre à la question suivante : ayant un ensemble d'images annotées, de quelle façon faut-il procéder afin d'annoter une nouvelle image introduite sans annotations. Pour cela, nous allons utiliser la méthode LSA qui a fait ses preuves sur des documents textuels.

Notre étude est organisée de la manière suivante :

Chapitre I : Ce chapitre présente généralité sur la multimédia et les systèmes de recherche

Chapitre II : Présente l'état de l'art sur les modèles actuels proposés pour l'annotation automatique d'images.

Chapitre III : Dans ce chapitre nous présente les composantes principales de notre système auto-annotation-LSA. Présente également un examen sur les techniques de traitement d'images utilisées dans les systèmes d'annotation, les techniques utilisées pour apprendre la sémantique à partir d'images à des fins d'annoter. On présente le système d'annotation qui consiste à associer des mots à la nouvelle image.

Chapitre IV : Présente nos résultats expérimentaux sur l'ensemble de données d'images. Nous présentons ici une méthode pour évaluer la performance d'annotation ainsi que quelques exemples illustratifs générés par notre système d'annotation.

En fin, une conclusion résume notre travail.

CHAPITRE I

La MultiMedia Et Les Systèmes De Recherche

I. La multimédia

Les ordinateurs peuvent être reliés aujourd'hui à des caméras, des appareils photos, des scanners, etc... Ils deviennent ainsi des appareils de saisie de l'information autant que de traitement. Cette information est de formes diverses : image, son et vidéo. Selon le type de données, le stockage de l'information nécessitait des formats de données différents. Pour pouvoir en disposer en même temps, on a introduit dans les années 1980 le concept de multimédia [1].

Le « Multimédia » désigne le fait de gérer en même temps du son, de l'image, de la vidéo et des données. Multimédia désigne donc d'une part l'équipement et les procédés de management sonores, visuels etc... , et d'autre part la caractéristique d'une machine composée de ces éléments. Le mot multimédia est apparu vers fin années 1980 [2]. Il désigne alors les applications qui pouvaient générer, utiliser ou piloter interactive (IHM). On peut citer parmi elles : Animation (ex : dessins animés) ; Editeur WEB ; Jeu vidéo.

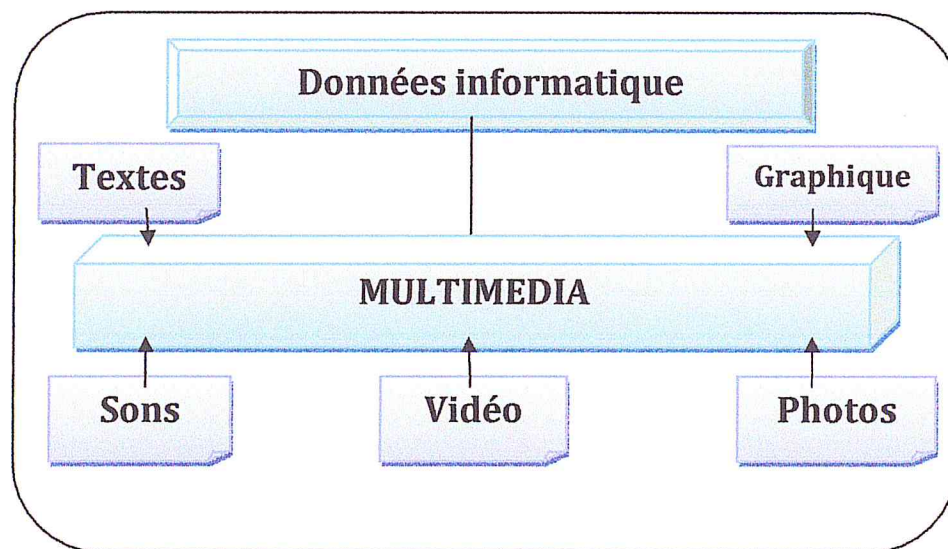


Figure1.1 : Composants du multimédia [2]

II. Les composants multimédia

Les différents types de media sont: le texte, l'image, le son et la vidéo.

II.1. Le texte

Le texte est considéré comme une donnée multimédia. La donnée texte est souvent réduite à une représentation d'une liste de chaîne de caractères. En plus de représentation de la structure logique du texte (titre, auteurs, ...), on considère souvent une représentation descriptive de l'information textuelle (mise en page) [3].

II.2. Son

Le son est une onde produite par la vibration mécanique d'un support fluide ou solide, et propagée grâce à l'élasticité du milieu environnant sous forme d'ondes « longitudinales ». Par extension physiologique, le son désigne la sensation auditive à laquelle cette vibration est susceptible de donner naissance [2].

II. 3. La vidéo

La vidéo regroupe l'ensemble des techniques, technologies, permettant l'enregistrement ainsi que la restitution d'images animées, accompagnées ou non de son. Un flux vidéo est composé d'une succession d'images, 25 par seconde en Europe et 30 par seconde aux USA, composant l'illusion du mouvement [2].

II. 4. L'image

L'image constitue l'un des moyens les plus importants qu'utilise l'homme pour communiquer avec autrui. C'est un moyen de communication universel dont la richesse du contenu permet aux humains de tout âge et de toute culture de se comprendre.

Couramment, une image désigne la représentation d'un objet par différents moyens ou supports, dessin, image numérique, peinture, photographie, etc... [4].

II. 4.1. Image numérique

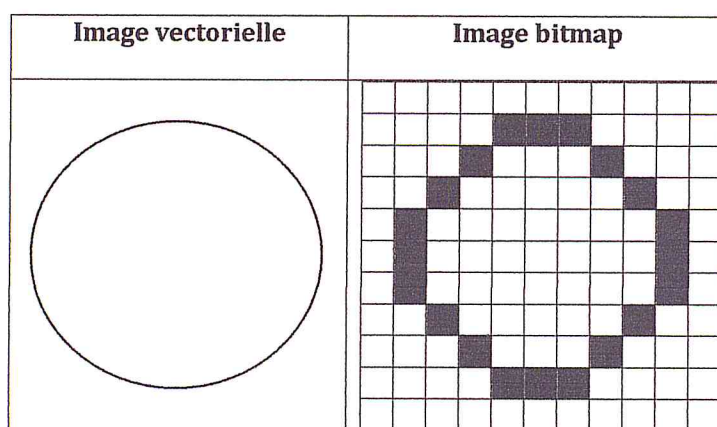
L'image numérique est l'image dont la surface est divisée en éléments de taille fixe appelée cellule, ou pixel, ayant chacune comme caractéristique un niveau de gris ou de couleur prélevé à l'emplacement correspondant dans l'image réelle, ou calculée à partir d'une description interne de la scène à représenter. On désigne par pixel la plus petite unité adressable dans une image.

La numérisation d'une image est la conversion de celle-ci de son état analogique (distribution continue d'intensité lumineuse dans un plan xOy) en une image numérique représentée par une matrice bidimensionnelle de valeurs numériques $f(x, y)$ où x, y sont des coordonnées cartésiennes d'un point de l'image et $f(x, y)$ est le niveau de gris en ce point [6].

4.2. Les types d'images

On distingue généralement deux grandes catégories d'images :

- **Les images bitmap** (appelées aussi image raster) : il s'agit d'images pixellisées, c'est-à-dire un ensemble de points (pixels) contenus dans un tableau, chacun de ces points possédant une ou plusieurs valeurs décrivant sa couleur.
- **Les images vectorielles** : les images vectorielles sont des représentations d'entités géométriques telles qu'un cercle, un rectangle ou un segment. Ceux-ci sont représentés par des formules mathématiques (un rectangle est défini par deux points, un cercle par un centre et un rayon, une courbe par plusieurs points et une équation). C'est le processeur qui sera chargé de « traduire » ces formes en information interprétable par la carte graphique [6].



Deux types d'image

II.4.3. Les formats d'images

Le format d'image est la description du fichier image en général. Le format comprend deux parties : une partie contenant les informations sur l'image (largeur, hauteur, nombre, de couleur, etc...) et une autre qui contient la base données de l'image, dite (niveau de gris ou couleur). Parmi les formats standards de fichier existants : **TEFF, GIF, JPEG, BMP** etc... [5].

II.4.4. Caractéristiques d'image

1. La texture

La texture d'image est une primitive visuelle importante. L'herbe, les fleurs, un tissu, sont des exemples de texture dans les images. Beaucoup d'objet peut être distingués par leur texture. S'il est simple d'expliquer ce qu'est une texture, la définir de manière formelle est plus complexe [6].

2. La couleur

La couleur est une caractéristique chromatique de l'image. L'extraction de la couleur à besoin d'un modèle, car les distances entre les couleurs. Généralement, la couleur d'une image peut être trouvée en accordance avec le sens qu'elle transport et la sensation qu'elle provoque [6].

3. La forme

La forme d'un objet identifie et représente sa silhouette. Contrairement aux modèles géométriques standard comme le carré, le cercle, le rectangle ..., il est difficile de décrire sous forme de texte un objet d'une forme complexe, d'où la nécessité d'introduire d'autre méthode de description de formes [6].

III. Les systèmes de recherche

III.1. Systèmes de recherche d'information

Depuis une décennie, les bases de documents multimédia se multiplient et leur taille augmente de manière spectaculaire. Parmi les médias les plus stockés à l'heure actuelle, on trouve en première position le texte et les images.

Ce dernier a donc besoin d'outils pour pouvoir retrouver rapidement et efficacement les images qu'il recherche parmi la grande masse d'information disponible. Ces outils sont regroupés dans les systèmes de recherche d'information (SRI). Dans le cadre de ce travail, nous considérons seulement les systèmes de recherches d'images (SRIm) [7].

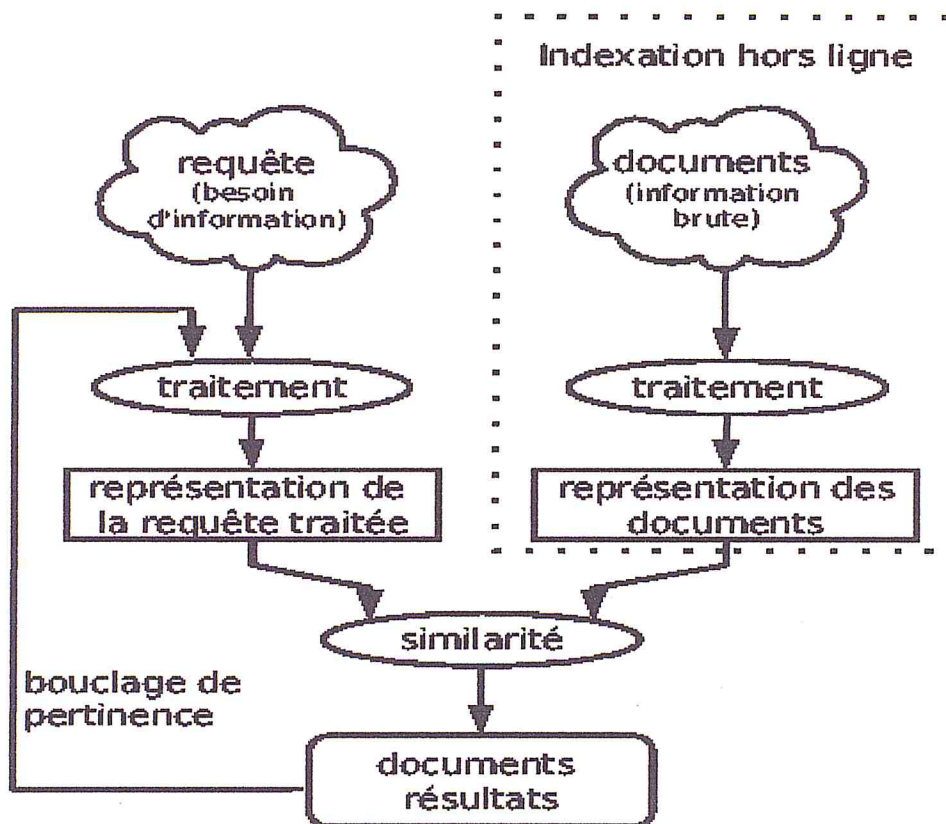


Figure 1. 2. Schéma d'un système classique de recherche d'information [7]

III. 2. Système de recherche d'image(SRIm)

Les systèmes de recherche d'image classiques comportent plusieurs phases :

- La phase d'indexation qui consiste à attribuer aux données de la base des représentations manipulable par le système.
- La phase de recherche, cette étape consiste à appliquer une fonction de similarité à fin de trouver les documents pertinents.
- Éventuellement, une phase de bouclage, si les réponses trouvées ne sont pas satisfaisantes [7]

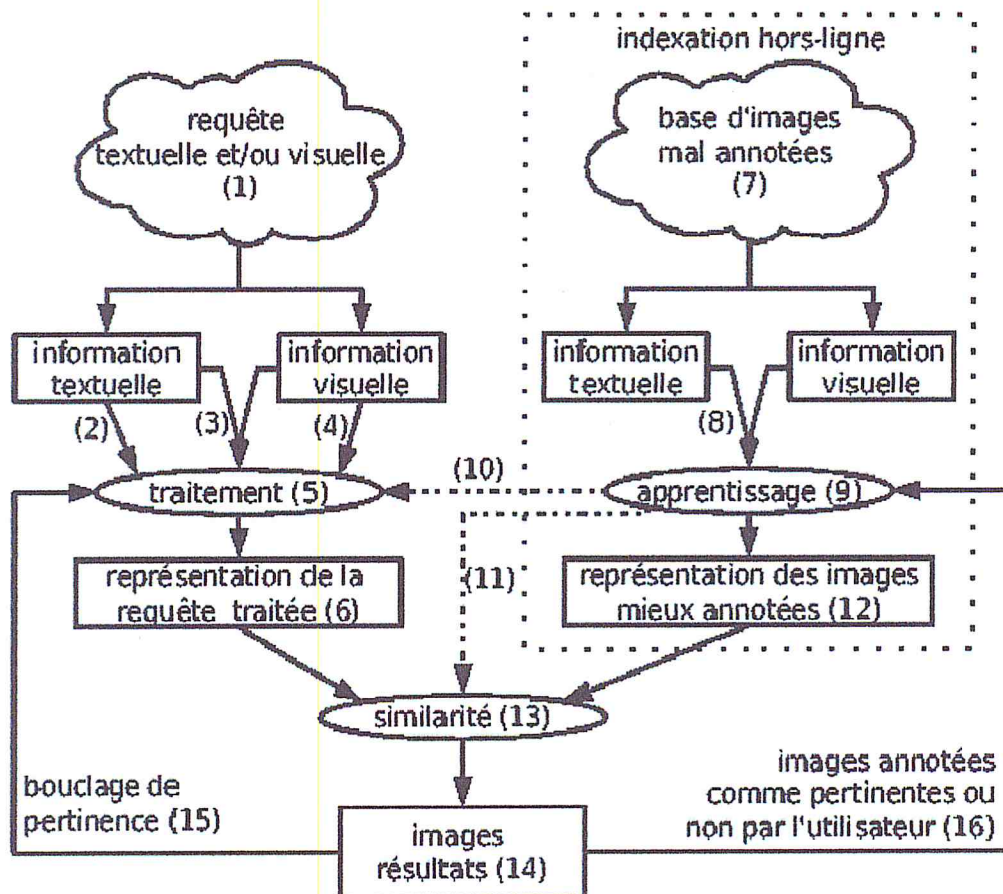


Figure 1.3. Schéma Système de recherche d'image(SRIm) [7]

Dans le domaine de recherche d'images on trouve principalement 2 classes :

Systèmes de recherches basés sur l'annotation,

Système de recherches basé sur le contenu visuel.

1. Recherche d'information basée sur les annotations

Dans ce type de système les images sont annotées et l'utilisateur fournit sa requête sous forme textuelle, la recherche devient une simple recherche textuelle entre la requête et les mots associés aux images (annotation). L'avantage de ce type de recherche est dû par leur facilité d'utilisation et leur efficacité. Une recherche de ce type portera donc sur le sens des images recherchées. Cependant, rien ne garantit que les annotations de l'image sont sémantiquement proches du contenu visuel de l'image. La performance du système est donc entièrement dépendante de la qualité des annotations (d'où l'importance de l'annotation dans ce type de recherche). C'est un risque que l'on peut prendre dans le cas d'images personnelles, mais qui est risqué dans le cas d'internet ou des images illégales mal indexées pourraient ne pas être filtrés et diffusées de cette manière. Exemple de ce type de système on trouve **Google Image** [7].

2. Recherche d'information basée sur le contenu d'image

Les systèmes issus de ce type de recherche s'appuient essentiellement sur les caractéristiques visuelles (comme la texture, la couleur, la forme) contenues dans les images.

On peut citer parmi ces systèmes QBIC5 d'IBM, Visual SEEK de l'université de Columbia, Photo book de MIT Media Lab. Le principe consiste à extraire les descripteurs visuels des images puis calculer les similarités entre ces descripteurs afin de trouver les images de la base les plus similaires visuellement bien sûr et pas forcément sémantiquement de l'image requête. Dans ce type de recherche, la requête est forcément composée de descripteurs visuels souvent fournis sous la forme d'une image [7].

L'inconvénient majeur de ce type de recherche est dû au faible niveau sémantique des descripteurs utilisés dans la comparaison pour trouver les images pertinentes, et qui rend l'approche peu efficace.

Remarque : Il existe une troisième classe de type de recherche d'images qui consiste à combiner les deux approches.

Pour pallier les faiblesses dues au fossé sémantique de ces deux types de système, nous proposons de construire un système appartenant à une troisième catégorie de système de recherche d'image qui combine les informations textuelles et visuelles en différentes phases du SRI. La FIG.2. Montre un schéma de ces différentes phases.

Nous proposons pour ce type de système de base de document (7) contient les images annotées automatiquement à partir du texte associé à l'image et provenant des pages du web, d'encyclopédie ou de revues. ce texte n'est donc pas forcément pertinent pour l'image : l'image sont mal annotées.

Dans la phase d'indexation, les informations visuelles et textuelles peuvent être combinées (8) afin d'utiliser les techniques de fouilles d'image ou d'apprentissage (9) Pour trouver les liens (motifs fréquents, zones visuelles caractéristiques d'un mot....) entre informations textuelles et visuelles, des images mais également de trouver des correspondances caractéristiques contre les informations contenues dans les requêtes de l'utilisateur (10) ou bien encore d'apprendre les descripteurs visuels pertinents pour un certain concept (11) la représentation (12) des documents obtenue après l'apprentissage peut être de type textuelles, visuelles ou les deux.

Dans la phase de recherche, trois types de requête (1) peuvent être exprimés :

	Système textuelle de recherche d'image	Système de recherche d'image par le contenu visuel	Système de recherche d'image fusionnant texte et Images
<i>Document contenant l'information</i>	Texte	Images	Image(*), texte (*) ou les deux
<i>Indexation</i>	Textuelle	Visuelle	Textuelles, visuelles ou les deux
<i>Requête</i>	Textuelle	Visuelle	Textuelles, visuelles ou les deux
<i>Document résultat</i>	Image	Images	Image, texte (**) ou les deux

Tab1.comparaison les caractéristiques de trois types de système de recherche d'image.

(*) Même s'il est possible d'avoir dans la base des documents uniquement textuelles ou uniquement visuelles, il est important pour, la phase d'apprentissage, que le nombre de document bimodaux soit conséquent.

(**) le système peut fournir comme résultat des textes, même dans le cas où la requête est une image, mais dans ce cas ce n'est plus un système de recherche d'image.

L'utilisateur fournit une requête visuelle (4) .(par exemple une image requête ,dans ce cas la ,le système peut se soit rechercher les images similaire visuellement comme dans les systèmes classique CBIR et éventuellement récupérer les mots clés associées aux images similaire pour annoter l'image avec des mots clés grâce aux liens appris par apprentissage(auto annotation),puis rechercher les images qui sont annotée avec les même mots

L'utilisateur fournit une requête textuelles(2),(par exemple ensemble de mots clés).Le systèmes recherche les images qui ont été annotée avec les même mots clés comme dans les systèmes textuelles classiques visuelles et sémantique communes appris lors de la phase d'apprentissage par combinaisons des informations textuelles et visuelles.

L'utilisateur fournit un ou plusieurs mots clés ainsi qu'une requête visuelle (3).Le système peut alors rechercher des images qui possèdent même genre de caractéristiques sémantiques et visuelles par exemple en réalisant la similarité entre les vecteurs représentant concaténant les espaces textuelles et visuelles .il peut également rechercher les images similaire seulement par les dimension visuelles qui sont les plus discriminantes par les concepts représentées par les mots clés de la requête en fonction ce qu'il a appris lors de la phase d'apprentissage.

Eventuellement, dans la phase de bouclage de pertinence, l'information textuelle Et/ou visuelles fournit par l'utilisateur sont prise en compte pour améliorer la requête (15) ainsi comme données entrantes de la partie apprentissage (16). (Par exemple les images sélectionnées par l'utilisateur sont annotées avec les mots de requêtes de l'utilisateur).

Le tableau1 résume les caractéristiques des trois types de système de recherche d'image,nous voyant que nous pouvant combiner les informations textuelles et visuelles en nombreux phases d'un système de recherche d'image .la question qui se pose maintenant est l'efficacité et la performance d'un tel système notant déjà qu'il n'est pas nécessaire quel e

système combine les informations textuelles et visuelles dans toutes les phases de systèmes qui peut considérablement réduire sa complexité. Se pendant le choix des phases de système qui combineront l'information textuelles et visuelles dépend de la tâche pour laquelle il est utilisé.

IV. Quelques systèmes de recherche d'images

- L'un des premiers systèmes de recherche d'image par le contenu visuel est le système **QBIC**. QBIC utilise des traits de couleur, de texture et de forme. Il permet des requêtes par image-exemple, directes ou par croquis. Il permet également de faire de la navigation avec bouclage de pertinence(BP). Le texte associé aux images, comme le nom de l'auteur ou du média, est seulement utilisé pour faire un filtrage des résultats (par exemple, seules les images associées à tel auteur sont affichées), et non pas pour améliorer sémantiquement les résultats[8].

- Le système **PICTION**, semble être le premier système à vraiment proposer de combiner texte et information visuelle. Le texte est extrait de la légende associée aux images d'articles de presse. Une hiérarchie visuelle combinant à la fois des informations de couleur, de texture et de forme est également proposée [8].

- Le système de **NETRA**, utilise des descripteurs de couleur, de texture, de forme, une segmentation en région est mesurée par deux rectangles (intérieur et extérieur) il n'utilise pas d'information textuelle [8].

- Le système **A-LIP**, est un module du système **SIMPLICITY** ; il propose d'annoter les images à l'aide de modèles de **MARKOV** cachés. A partir des descripteurs des images d'apprentissages correspondant à un certain concept, le système construit pour chaque concept forment un dictionnaire de concept. Lorsqu'une nouvelle image doit être annotée, la vraisemblance entre les descripteurs de image et le modèle est calculée les mots associés aux concepts les plus significatifs sont associés à l'image pour permettre une recherche textuelle ultérieure [8].

CHAPITRE II

**Etat De L'art Sur L'annotation D'image
Par L'apprentissage Automatique**

II.1 Indexation

Comme on avait vu dans l'introduction les systèmes classiques de recherche d'image peuvent être classés en deux catégories :

- Les systèmes textuels de recherche d'image,
- Les systèmes de recherche d'image par le contenu visuel.

II. Ils possèdent tout les deux une phase d'indexation et une phase de recherche. Le terme « **indexation** » est parfois ambigu, car il est utilisé pour deux problèmes distincts :

- Le processus d'indexation des informations à partir des documents.
- La représentation de cette information.

L'indexation a donc pour but d'extraire et de représenter le sens d'un document de manière à ce qu'il puisse être retrouvé par l'utilisateur. On utilise indifféremment les deux emplois de ce terme.

Dans ce chapitre, on détaille comment les informations, textuelles puis visuelles, sont extraites des documents, et comment ces systèmes les utilisent pour rechercher l'information. On cherche surtout à mettre en évidence les difficultés rencontrées pour construire de tel système. Puis on décrit quelques systèmes de recherche d'image par le contenu de l'état de l'art, et pour chaque système on regarde s'il utilise et si oui de quelle manière l'information textuelle [9].

II.2. Indexation et recherche textuelle d'images

II.2.1. Indexation textuelle manuelle

L'indexation textuelle manuelle d'images est le plus souvent réalisée par un documentaliste appelé iconographe .son rôle est de classer et d'indexer les images on les associant à des catégories et à des groupes de mots ,souvent extrait à des thésaurus permettant de retrouver facilement les images .son travail est très utile pour les agences de presse ,les centres de documentation ,les musées ...du fait de l'accroissement du nombre de photographie personnelles ce travail est aussi souvent réaliser par les utilisateurs qui souhaitent décrire leur image personnelle [9].

Ce pendant les iconographes comme les utilisateurs rencontrent différents problèmes d'indexation manuelle d'image. Le principal problème est le choix des termes. En effet, la personne qui indexe l'image et la personne qui recherche une image ne choisirons pas forcément le même terme pour décrire le même concept .en effet la probabilité pour que le même terme soit choisis par deux individus différents pour décrire une entité quelconque est inferieur à 20% .une même image peut avoir plusieurs sens, contenir plusieurs thèmes [9].

Il existe deux types d'indexation d'image .la première **le hard indexing** correspond a ce que l'indexeur voit dans l'image : le portrait d'une femme est « la Joconde » de Leonard de Vinci.

Malgré sa subjectivité, l'indexation manuelle reste une méthode efficace pour associer un sens à des images .cependant lorsque l'on a un grand volume d'image à indexer, ce travail devient vite fastidieux, voir impossible, ce qui n'est pas le cas pour l'indexation automatique. de plus ,en ce qui concerne les photographies personnelles, une étude [4] sur la façon dont les utilisateurs gèrent leurs photographies conclue que ceci ne sont pas intéressés par annoter leurs photographies du fait des efforts nécessaires, et du fait qu'il ne reste plus alors au système qu'a transcrire en texte, pour obtenir une indexation automatique[9].

II.2.2 Indexation textuelle automatique

L'indexation textuelle automatique d'image consiste à associer des mots à une image au moyen d'un système informatique sans aucune intervention humaine. Il existe deux approches : l'indexation textuelle automatique à partir du texte associé à l'image, l'indexation textuelle automatique à partir du contenu visuel de l'image [9].

La première approche n'est possible que lorsque les images sont associées à du texte. C'est le cas des images des encyclopédies, des catalogues de vente, des manuels techniques ... et aussi du web. L'indexation textuelle des images sur le web peut s'effectuer à partir des mots présents dans le titre de la page ou des mots les plus fréquents ou pertinents de cette page. Cependant, toutes les images présentes sur une même page web ne devraient pas être indexées avec les mêmes mots. Beaucoup de moteurs de recherche utilisent aussi l'URL et le nom de l'image, mais la plus part des images ne sont pas nommées de façon pertinente, mais bien souvent par des noms génériques comme image_001.jpg qui ne porte pas de sens. D'autres techniques considèrent les mots associés à l'attribut ALT de la balise IMG d'une image ou bien le texte proche de l'image, ou bien une fusion de toutes ces informations [19]. Mais dans la pratique, il est difficile d'indexer les images sur le web de cette façon, car cela nécessite une indexation manuelle que l'on sait être très coûteuse en temps. De plus le texte proche de l'image n'est pas forcément celui que l'on associerait à l'image [9].

La deuxième approche est souvent appelée auto-annotation par le contenu. Annoter une image avec des mots seulement à partir du contenu visuel est impossible. C'est pourquoi la plus part du temps les méthodes d'auto-annotation sont en effet des méthodes de classification supervisées multi-classes (une classe par mot). Elles utilisent un ensemble d'apprentissage où les images sont associées aux classes de mots pour apprendre à prédire les mots sur de nouvelles images. Cependant comme nous le verrons l'auto-annotation d'image est un problème toujours ouvert, même les meilleurs systèmes ne sont pas capables d'apporter un gain de plus de 40% pour le modèle empirique basée sur la distribution a priori qui a lui-même un score de classification très faible [9].

L'indexation textuelle automatique d'image, qu'elle s'effectue à partir du texte associé à l'image ou du contenu de l'image, comme et donc beaucoup d'erreurs les images sont mal annotées.

II.2.3. Les modèles classiques de la recherche textuelle

Une fois indexées textuellement, les images peuvent être recherchées avec les modèles classiques de recherche dans les documents textuels (modèle booléen), modèle vectoriel, modèle probabiliste [10].

- **Modèle booléen**

Dans le modèle booléen, un document d_i est représenté par une conjonction de termes indépendants s que l'on représente sous la forme d'un ensemble : $d_i = \{w_1, w_2, \dots, w_n, w_d\}$ sans pondération. La requête q est une expression logique ET, OU, et NON. Un document sera jugé pertinent par le système si l'expression logique de la requête est satisfaite par ce document. Ce modèle ne permet pas de retrouver les documents qui ne correspondent que partiellement à la requête [10].

- **Modèle vectoriel**

Dans le modèle vectoriel, un document est représenté sous la forme d'un vecteur à n_w dimension : $d_i = (w_{i,1}, w_{i,2}, \dots, w_{i,n_w})$ où chaque $w_{i,j}$ est la pondération associée au terme w_j dans le document d_i . Ce modèle suppose que les vecteurs sont des points dans un espace où les termes forment une base orthogonale. Les termes sont supposés indépendants. La requête est exprimée selon le même formalisme $q = (w_{q,1}, w_{q,2}, \dots, w_{q,j})$ pour évaluer la pertinence d'un document par rapport à une requête, le système calcule une valeur de similarité classique est le cosinus, la formule de Dice et la formule de Jaccard, les pondérations $w_{i,j}$ tiennent compte de la fréquence du terme dans le document (term frequency tf) du nombre de documents dans lesquels apparaît le terme (document frequency df), de la longueur du document, et l'apparition des termes d'indexation dans les parties logiques du document, comme le titre, le résumé, etc. [10].

- **Le modèle LSI**

Indexation en sémantique latente LSI (Latent Semantic Indexing LSI) est une technique d'indexation s'inspirant de la LSA (latent semantic analysis) développée au début des années 90. Ce modèle se fait conçu comme une nouvelle technique d'analyse statique qui vise à décrire de manière économique les cooccurrences de termes qui surviennent au sein d'un ensemble de documents, puis d'en déduire des proximités sémantiques entre termes. En effet, cette analyse se base sur la matrice d'occurrence terme X document A qui est considérée comme une représentation des documents, dans un espace de grande dimension par le biais d'une décomposition en valeurs singulières (SVD), par la décomposition de la matrice A en trois matrices [10].

$$A = USV^T$$

Où U est une matrice orthogonale de la taille $M \times N$ de description de terme est matrice orthogonale de taille $N \times N$, à partir d'un certain nombre $k < N$ nous nous apercevons de l'existence de la valeur singulière très faible et qui peut être négligée dans la matrice.

Il y a une meilleure proximité donnée par :

$$A_k = U_k S_k V_k^T$$

Cette réduction va permettre de ne garder que les k termes les plus significatifs afin de calculer la similarité document _ document. Les documents sont représentés dans un espace vectoriel de dimension k . Les colonnes de produit de matrice $S \times V$ sont les coordonnées de vecteur document, par la suite la LSI, s'appuie sur une mesure basée sur le cosinus de l'angle formé par deux vecteurs, celui-ci représentant la similarité [10].

II.3 Indexation et recherche d'image par le contenu visuel

II.3.1. Les types de systèmes

Les systèmes de recherche d'image par le contenu visuel peuvent être construits avec différents objectifs en vue. certains systèmes recherche à reconnaître un objet en particulier, on parle alors de reconnaissance de forme ou d'objet. parmi ceux-ci, certains systèmes sont spécialiser pour travailler sur des images médicales ou des photos satellite. D'autre cherche à classer les images en fonction du type de scène qu'elle représente (naturel /artificiel, extérieur...). on parle alors de reconnaissance de scène ou de classification de scène. une troisième catégorie de système traite des images généraliste et cherche à obtenir de bonne performance en moyen sur toutes les images. dans le cadre de cette thèse, on s'intéresse surtout au troisième type de système[11].

III.3.2. Les bases d'images

On peut distinguer deux catégories de base d'image :

- **La base d'image spécialiste** qui contient des images de même type : image médicale, image géographique, portrait de personne
- **La base d'image généraliste** qui contient des images de nature variées : photographies personnelles, image des encyclopédies, de Corel...

Les stratégies de recherche d'image pour ces deux catégories de base sont très différentes. Pour la première, on connaît à priori le type d'image que l'on peut y rencontrer, ainsi que le type de recherche que l'on va y mener. par exemple rechercher une personne parmi d'autres même si elle porte des lunettes, un chapeau ou non. cette connaissance a priori permet de développer des techniques d'indexation et de recherche très efficaces. pour la seconde catégorie, par contre, on ne sait pas ce que contiennent les images. et du fait du faussé sémantique, on ne sait pas sémantiquement ce que recherche l'utilisateur.

Dans cette thèse nous nous intéressons uniquement aux bases d'image généraliste [11].

IV. La Segmentation

La segmentation d'image consiste à séparer en régions homogènes les divers composants visibles dans une image. L'humain sait naturellement séparer des objets dans une image. Pour cela, il se base notamment sur des connaissances de haut niveau qui lui permettent de détecter ce qui l'intéresse dans l'image. En traitement du signal, on caractérise une région comme étant un ensemble de points (pixels) ayant des propriétés communs d'intensité, de texture, de couleur..., qui la différencient des régions voisines. Il y a de nombreuses méthodes de segmentation, on distingue cependant deux familles d'algorithmes :

La segmentation par approche « contour » ou « frontière » (edge-based segmentation) [12].

Un contour est une frontière entre deux milieux différents (2couleurs, 2 niveaux de Indexation et recherche d'images par le contenu visuel).

On utilise souvent l'image dérivée pour mettre en évidence les variations de niveau. Quant il y a une forte variation, cela est souvent du à un changement de teinte, d'éclairement ou de texture, et de façon générale à un changement de région.

La segmentation par approche « région » (region-based segmentation). Le principe est de trouver les régions en regroupant les pixels ayant des caractéristiques similaires et en séparant ceux qui sont différents (technique division-fusion, par accroissements de régions, par statistiques bayésienne ...).

Une des difficultés de la segmentation est de savoir en combien de régions doit être découpée une image, car certaines images peuvent être segmentée en peu de régions (par exemple, un portrait sur fond uniforme), alors que d'autre nécessitent plus de régions.

A quel niveau une segmentation doit-elle s'arrêter ?

On comparant des méthodes classiques de segmentation pour l'indexation d'images (clustering, mean shift, approche morphologique ou multi résolution).Leurs résultats montrent qu'il n'existe pas de segmentation générique ou meilleure que les autre, mais que en fonction de la tache à accomplir certaines méthodes sont plus efficaces que d'autres.

Il existe également des approches *par segmentation ou les régions* d'images sont déterminées a priori. Elles supposent par exemple que l'objet intéressant est présent au centre de l'image ou au contraire impose une grille ou toutes les régions de l'image ont la même

aire. Pour une recherche sémantique d'image, ce type de grille ne nous paraît pas être intéressant, car plusieurs objets peuvent se retrouver dans la même région. De plus, les descripteurs de formes et de textures ne peuvent pas être utilisés[12].

V. Mesure de similarité

Pour rechercher les images les plus similaires à une image-exemple ou pour les regrouper, il faut pouvoir mesurer la similarité(ou la dis similarité) des images. Idéalement, les mesures doivent être capable de mesurer la similarité sémantique de deux images, dans la pratique elles ne sont capables que de mesure la similarité visuelle. Cependant, beaucoup de travaux essayent de s'inspirer du système visuel humain pour proposer des mesures plus efficaces.

Nous allons décrire maintenant des mesures de similarité dans le cadre de la recherche d'images similaire ; Ces mesures sont également utilisables pour le cadre de la recherche de régions similaires, mais aussi dans le cas de la classification supervisée ou non-supervisée d'images (ou de donnée au sens général).

Il existe un grand nombre de mesures de similarité. Certaines sont des distances, c'est-à-dire des mesures qui ont les propriétés de non-négativité, réflexivité, symétrie et qui respectent l'inégalité triangulaire. Certaines mesures sont spécifiques aux histogrammes ou aux distributions. Une étude comparative de 9 mesures est proposée. Nous donnons ci-après quelques unes des mesures classiques [13].

➤ **Distances euclidienne** Une distance classique est la L_p -norme (ou distance de Minkowski) :

$$L_p\text{-norme}(\vec{x}, \vec{y}) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

Où p est le paramètre de la norme. La L_1 - norme est appelée distance de Manhattan ou distance City-Block. La L_2 - norme est distance euclidienne.

➤ Distance quadratiques

L'inconvénient des mesures telles que la distance euclidienne ou l'intersection d'histogrammes est qu'elles comparent les composantes des vecteurs une par une, sans prendre en compte les autres composantes. Pour remédier à ce problème, on peut utiliser la distance quadratique. Elle est définie ainsi :

$$\delta_Q(\vec{x}-\vec{y}) = \sqrt{(\vec{x} - \vec{y})^T A (\vec{x} - \vec{y})}$$

La matrice A permet de pondérer le poids des composantes voisines en fonction de leur distances à la composantes considérée. Cependant cette mesure de complexité quadratique.

➤ Similarité cosinus

Image d'apprentissage pour l'annotation sémantique automatique de l'image Similarité cosinus La plupart des systèmes de récupération de texte représente une image et d'interrogation comme un ensemble de mots clés. Deux images, V_q et V_d peuvent être considérés comme similaires si l'angle entre de leurs vecteurs est faible. Le produit scalaire normalisé est utilisé pour mesurer similitude:

$$Sim(\vec{D}, \vec{V}) = Cos(\vec{D}, \vec{V}) = \frac{\sum_{i=1}^n (d_i \times v_i)}{\sqrt{\sum_{i=1}^n d_i} \times \sqrt{\sum_{i=1}^n v_i}}$$

Une similitude cosinus de 1 implique que deux images sont identiques, un a similitude de 0 implique qu'ils ne sont pas liés. Nous utilisons cette mesure pour calculer la similitude entre un vecteur d'image et chaque vecteur extrait de mot-clé - Blob matrice. Les résultats obtenus montrent également que la similitude cosinus fonctionne un peu mieux que la distance euclidienne normale. Les résultats sont alors classés en utilisant les valeurs de similarité[14].

Discussion

L'inconvénient de la distance euclidienne est que des certains points qui sont visuellement proches peuvent être aussi distants que des points visuellement éloignés.

L'inconvénient de la distance quadratique est que des vecteurs visuellement différents peuvent être similaires [13].

VI. Apprentissage automatique

La faculté d'apprendre est essentielle à l'être humain pour reconnaître une voix, une personne, un objet... on distingue en générale deux types d'apprentissage : l'apprentissage « par cœur » qui consiste à mémoriser telles quelle des informations, et l'apprentissage par généralisation ou l'on apprend à partir d'exemple un modèle qui nous permettra de reconnaître les nouveaux exemples pour les systèmes informatique, il est facile de mémoriser un grand nombre de données (texte, image, vidéo...), mais difficile de généraliser. par exemple, il est difficile de construire un bon modèle d'un objet et d'être ensuite capable de reconnaître efficacement cet objet dans de nouvelles images. L'apprentissage automatique est une tentative de comprendre et de reproduire cette faculté d'apprentissage dans des systèmes artificiel. Il nous semble donc appropriée d'utiliser des techniques issue de domaine pour découvrir et modéliser des connaissances liant texte et image. et pouvoir ainsi réduire le fossé sémantique[15].

Dans le domaine de l'indexation et la recherche d'image, l'apprentissage peut jouer les rôles suivants :

- Sélectionner les descripteurs visuels les plus pertinents,
- Associer les descripteurs de bas niveau à des concepts,
- Regrouper les images de manière hiérarchique ou non par similarité visuelle et/ou conceptuelle,
- Apprendre à fur et à mesure des interactions avec les utilisateurs,

Le grand volume de données multimédia et le nombre important de dimensions nécessaires pour les décrire imposent d'utiliser des techniques de réduction de dimension afin de pouvoir, d'une part extraire et résumer les connaissances souvent inconnues des données, et, d'autre part indexer et retrouver rapidement les documents. on peut séparer les techniques d'analyse de données en deux grandes catégories :

- La classification qui permet de réduire la taille de l'ensemble des individus en regroupant ce qui se ressemblent.

- Les analyses factorielles qui permettent de réduire le nombre de descripteurs [15].

VI.1. Techniques d'apprentissage automatique

L'apprentissage automatique se trouve au carrefour de nombreux autres domaines : intelligence artificielles, statistique, science cognitive, théorie de probabilité, de l'optimisation, du signal et de l'information. Il est donc bien difficile de donner une taxinomie des techniques d'apprentissage .Après avoir donné quelques définitions, nous présentons brièvement les deux principaux d'apprentissage : d'apprentissage supervisé et non –supervisé [15].

➤ Apprentissage supervisé

L'apprentissage supervisé suppose qu'un oracle fournit les étiquettes de chaque données d'apprentissage .on distingue en générale trois types de problèmes auxquels l'apprentissage supervisé est appliqué : la classification supervisée, la régression, et les séries temporelles , des trois types de problèmes différencient en fonction du type d'étiquette fournit par l'oracle [15].

➤ Apprentissage non supervisé

Contrairement a l'apprentissage supervisé dans l'apprentissage non –supervisé il n'y a pas d'oracle qui explicite les étiquettes .l'utilisation de ce type d'algorithme permet de trouver des structures, des dépendances entre descripteurs.

Le plus connue de problèmes non supervisé est la classification non supervisé ou clustering. La classe que, nous appellerons clusters, sont fournées par regroupement des documents qui certains caractéristiques en commun .un autre problème non supervisé est l'estimation. de densité.son objectif est de modéliser la distribution des données .le modèle doit pouvoir fournir une bonne estimation d'un document de test issue de la même distribution (inconnue)que les données d'apprentissage .nous pouvons également citer le problème de la réduction de la dimensionnalité .son objectif est de résumer l'information importante [15].

Le clustering est un outil important pour l'analyse de données .il visent à trouver les structures intrinsèque des données .on les organisant en groupe homogène et distincts(les clusters).les objets dans un même cluster doivent être similaire entre eux et différents des objets des autres clusters [15].

VII. Annotation d'image

VII.1 Définition d'annotation d'image

C'est un processus qui permet d'attribuer à une image un mot ou une liste de mots qui décrivent la sémantique de cette image. Généralement utilisée dans les systèmes de recherche d'images pour retrouver des images pertinentes dans une base de données d'image. Il existe deux types d'annotation d'images : manuelle et automatique [16].

VII.2. Types d'annotation d'image

Il existe deux types d'annotation d'images : manuelle et automatique.

1. L'annotation manuelle d'images

Comme sa dénomination l'indique, l'annotation est effectuée manuellement par un humain chargé d'attribuer à chaque image un ensemble de mots clés. Cette méthode est couteuse en temps et en argent, et devient impossible vue que la taille des bases de données d'images actuelles ne cessent pas de croitre[16].

2. L'annotation automatique d'images

C'est le système informatique qui réalise l'annotation de façon automatique, l'avantage de cette méthode c'est qu'elle augmente la productivité par contre la précision est actuellement moins fiable que celle de l'annotation manuelle.

Remarque : il existe un troisième type d'annotation qu'on nome *semi automatique* c'est comme l'annotation automatique mais elle nécessite l'intervention d'un utilisateur pour valider les décisions de système[16].

VII.3. Un modèle par analyse de la sémantique latente LSA

VII.3. 1. La description théorique de la Méthode LSA

LSA (pour *Latent Semantic Analysis* ou *Analyse de la Sémantique Latente*) est une théorie et une méthode pour extraire et représenter la signification contextuelle des mots par des calculs statistiques appliqués à un corpus volumineux. L'idée fondamentale est que l'agrégat de tous les contextes dans lesquels un mot donné apparaît ou n'apparaît pas fournit un ensemble de contraintes mutuelles qui détermine en grande partie la similitude de la signification des mots et des ensembles de mots entre eux[17].

LSA a été développée dès 1990 en vue d'améliorer le modèle vectoriel. Depuis, elle a été appliquée dans des domaines divers : la recherche d'informations multilingue, l'évaluation automatique de dissertations, l'acquisition de connaissance [17].

VII.3.2. Principe de la Méthode LSA

LSA définit le sens d'un mot statistiquement à partir de l'ensemble des Contextes dans lesquels ce mot apparaît et n'apparaît pas. Par exemple le mot université sera très proche des mots laboratoire, professeur... même si ces mots n'apparaissent pas ensemble, et très éloignés des mots cuisine, plage... même s'ils apparaissent ensemble dans le même document, car LSA possède un mécanisme d'enchaînements de liens de cooccurrence à plusieurs niveaux qui permettent une représentation correcte des sens des mots [17]. On peut dire alors que LSA se repose sur deux règles principales :

- Deux mots sont similaires s'ils apparaissent dans des contextes similaires.
- Deux contextes sont similaires s'ils comportent des mots similaires.

Cette récursivité croisée ne peut pas se faire à l'aide des fréquences de contiguïté ou d'un simple compte de cooccurrence, mais exige la mise en œuvre d'une analyse mathématique puissante qui est capable d'impliquer correctement des relations beaucoup plus profondes [17].

VII.3.3. Moyens mathématiques mis en œuvre : la SVD

On construit la matrice termes x documents A on mettant les images sur les colonnes et les mots clefs sur les lignes, chaque case A_{ij} de la matrice correspond à la fréquence du mot clef i dans le l'image j [2].

Toute matrice A_{mn} de rang r avec m, n peut se décomposée en produit de trois matrices par la Décomposition en Valeurs Singulières (SVD) :

$$A_{mn} = U_{mr} S_{rr} V^t_{rn} \quad (1)$$

Ou :

U et V définissent les vecteurs propres orthonormés associés aux r valeurs propres non nulles de AA^t et A^tA respectivement.

S est une matrice diagonale contenant les racines carrées non nulles des n valeurs propres de AA^t . On les nomme les valeurs singulières de A .

On range les éléments de S dans l'ordre décroissant et on élimine les valeurs singulières les plus proche de zéro de S et les vecteurs qui leurs correspondent dans les matrices U et V comme illustré dans la figure5. Le fait de réduire les dimensions des matrices U , S et V permet de représenter les structures d'associations mot clef-image les plus importantes, d'ignorer celles de moindre influence et d'éliminer le bruit du au hasard dans le choix du vocabulaire. Or il n'y a pas de règle pour déterminer le nombre de valeurs singulières à éliminé. Le nombre de dimension optimal de l'espace réduit pour la langue anglaise a été estimé empiriquement à 300 dimension [18]

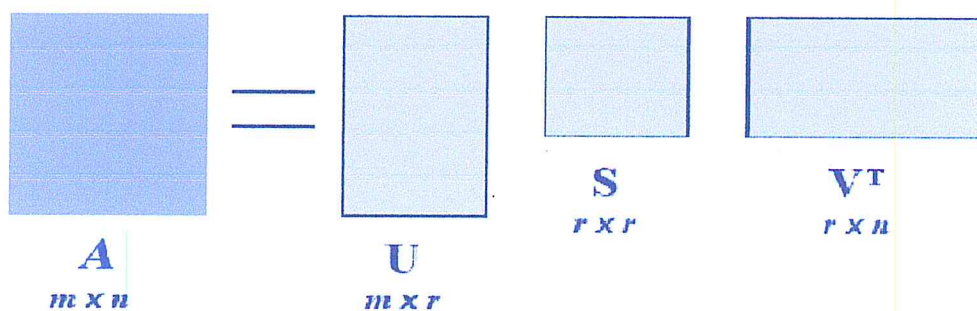


Figure.2.1. La Décomposition en Valeurs Singulières [18]

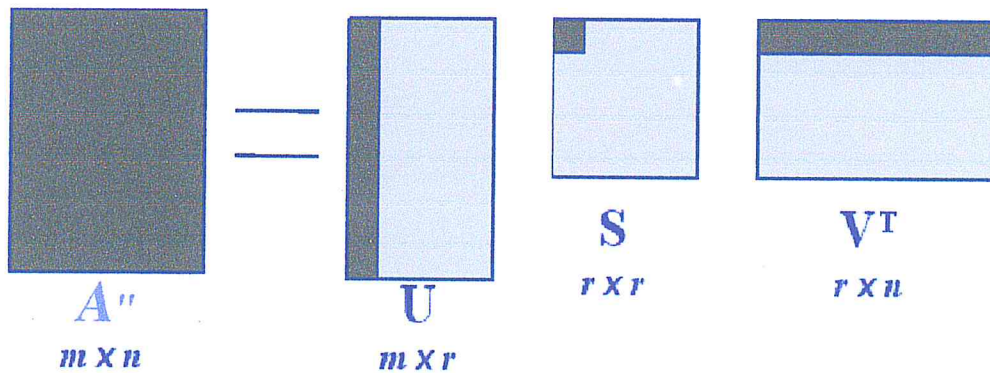


Figure 2.2. Réduction des dimensions des matrices [18]

Maintenant on dispose d'un espace latent où chaque document (resp. mots) est représenté par un vecteur ligne de V (resp. U). Pour comparer deux documents i et j (resp. mots) il suffit de calculer la distance (le cosinus par exemple) entre les deux vecteurs i et j de la matrice $V_{ni}S_u$ (resp. $U_{mi}S_u$). Pour résoudre une requête (trouver les documents similaires à un document d), on construit le vecteur associé au document requête en comptant les occurrences des mots du vocabulaire choisi qui apparaissent dans le document d , puis on le transforme en un pseudo-document en le projetant dans l'espace latent :

$$D = d^t_q \quad U_{mi} S^{-1}_u \quad (2)$$

Ensuite on utilise la distance entre D et chaque document de l'espace latent construit pour déterminer les documents les plus pertinents [2].

Dans la recherche d'information sur les documents images on se confronte avec le problème suivant : une même couleur peut décrire deux objets différents et deux couleurs différentes peuvent décrire le même objet.

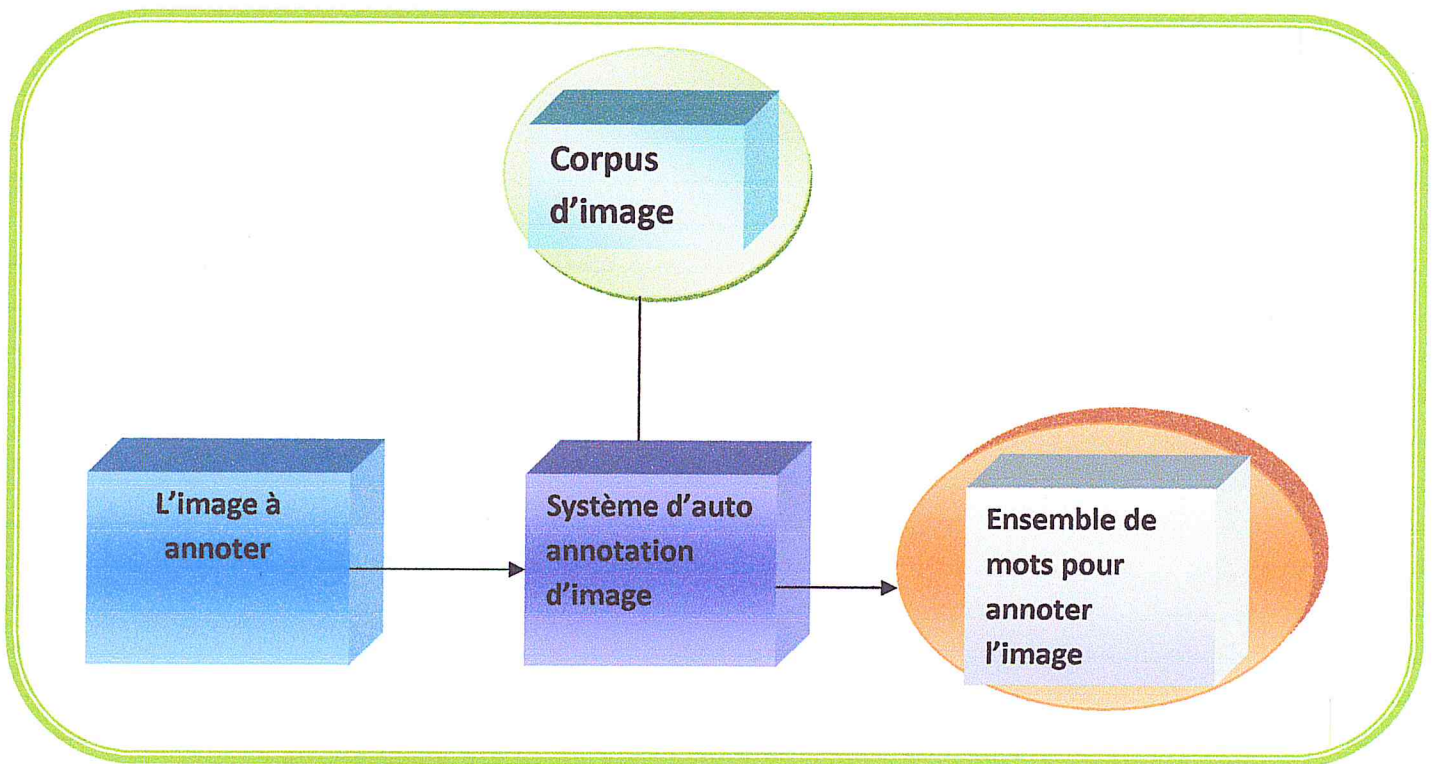


Figure2.3.Auto-annotation des images avec LSA [19]

Donc pour représenter la sémantique d'une image, on ne peut pas se contenter des descripteurs visuels et des histogrammes qui ne représentent que les caractéristiques de l'image (couleurs, textures, ...) et non pas son contenu.

Dans ce travail, on a essayé d'utiliser la méthode LSA pour annoter automatiquement des images en prenant en compte les caractéristiques (couleurs, textures, ...) et le contenu sémantique, en utilisant un corpus volumineux (figure2.3).

L'objet de la section suivante est de décrire l'implantation et le fonctionnement du système, ainsi que les résultats obtenus [19].

CHAPITRE III

Conception Et Réalisation

III. 1. Présentation du sujet

La recherche d'image sur le Web est un problème plus complexe. On a besoin d'outils pour pouvoir rapidement et efficacement des images parmi une grande masse d'information disponible. Ces outils sont regroupés dans les systèmes de recherche d'images (SRIM).

Le problème majeur qui se repose dans la recherche d'image est la sémantique d'une image, la solution permettant d'assigner la sémantique à une image, consiste à ajouter des mots clés à l'image : c'est l'opération d'annotation.

➤ **Problématique**

Dans ce travail, nous allons tenter de répondre à la question suivante : ayant un ensemble d'images annotées, de quelle façon faut-il procéder afin d'annoter une nouvelle image introduite sans annotations.

➤ **Solution**

Il s'agit ici d'attribuer aux images des termes textuels (mots clés) correspondant aux sens véhiculés par ces images (décrivant son contenu sémantique) d'une manière automatique.

Les objectifs du projet sont les suivants :

- Concevoir et réaliser un système d'auto annotation d'image.
- Evaluation des résultats à travers un corpus adéquat.

Dans ce travail on a proposé un système d'auto annotation basé sur la méthode LSA qui est une amélioration des anciens systèmes d'annotation, dans le souci d'améliorer les résultats, on a utilisé deux matrices, la matrice de mots clés / images, et la matrice de blobs / images au lieu la matrice d'image/mots clés et région, ainsi on a travaillé avec les blobs au lieu les régions, pour réduire car les blobs regroupent les caractéristiques visuels des régions de l'image segmenté de manière générale.

III.2. Description du corpus

La base d'images de COREL, contient **9883** images, chaque image est annotée manuellement de **1 à 5 mots**, et segmentée de **3 à 10 régions**, chaque région possède un vecteur visuel.

On partage la base d'images en deux ensembles

- le premier ensemble contient un peu moins de 5000 images, et sera utilisé par le système comme une base d'apprentissage.
- le deuxième servira pour faire des tests sur les performances de notre système.

Mais dans notre travail on a travaillé avec 10 images de corpus et chaque image est annotée de **3 à 5 mots**, et segmentée de **5 à 10 régions**, chaque région possède un vecteur visuel de deux composant (variance, et uniformité).

III.3. Architecture du système

Nouvelle image

Base d'image



Système d'auto- annotation - LSA

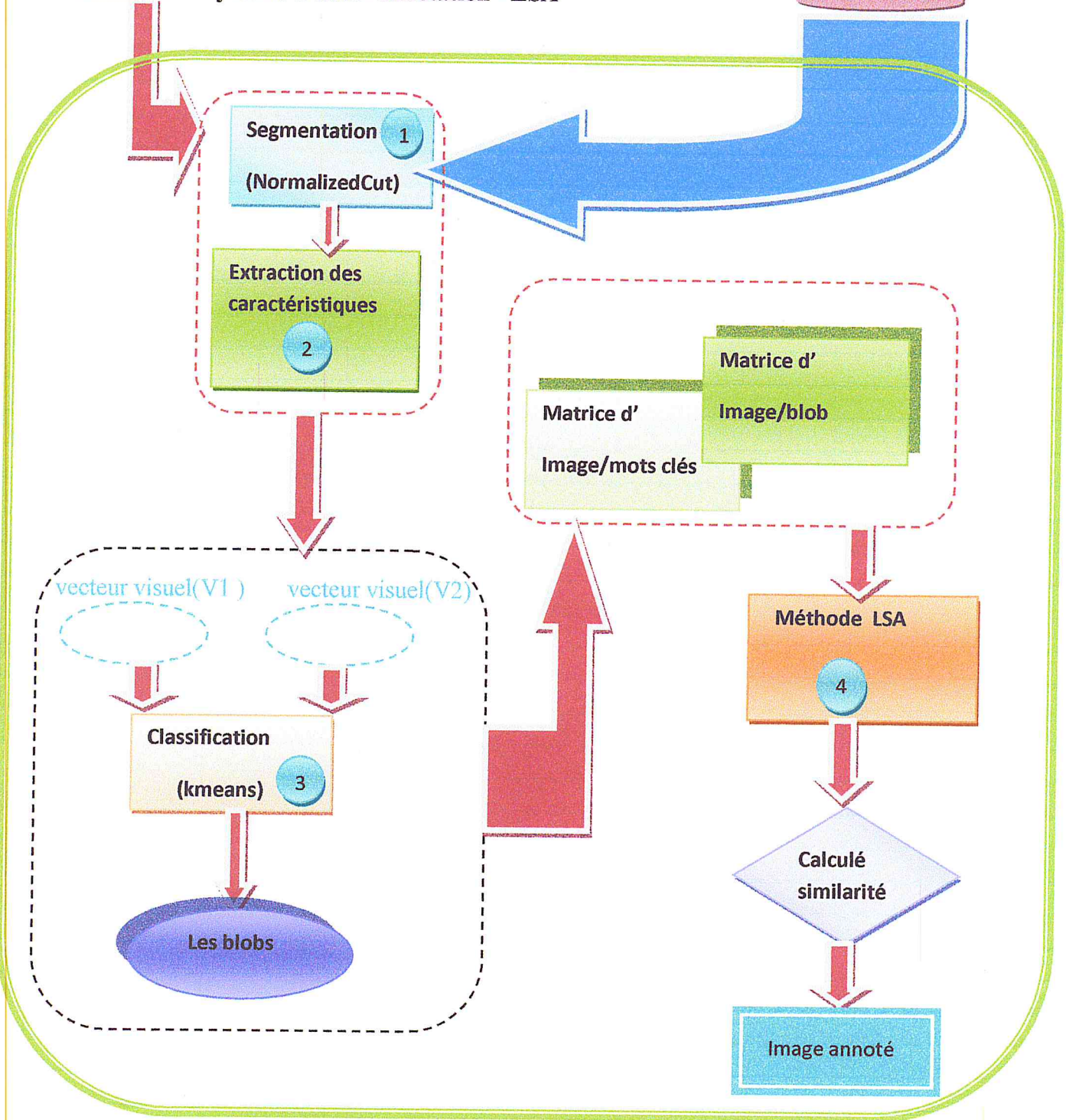


Figure3.1. Architecture du système d'auto- annotation - LSA

III.4. Le processus d'annotation d'image

Notre approche se compose de deux étapes principales : **L'apprentissage** et **le traitement**.

Dans la première partie on a segmenté les images en plusieurs régions ,ensuite on a fait l'extraction des caractéristiques de la texture des régions d'images du corpus, en effet on a classifié ces vecteurs de caractéristiques en blobs.

Dans la deuxième partie on a appliqué la méthode LSA (Latent Sémantique Analyse). On travaillé avec cette méthode pour résoudre le problème de synonymie et d'autonymie et réduire la complexité des calculs. An utilisant des expériences sont menées sur des ensembles de données d'images Corel.

III.4.1. La partie d'apprentissage

Dans cette partie l'image passe par un pré traitement suivant les étapes illustré dans la figure ci dessus.

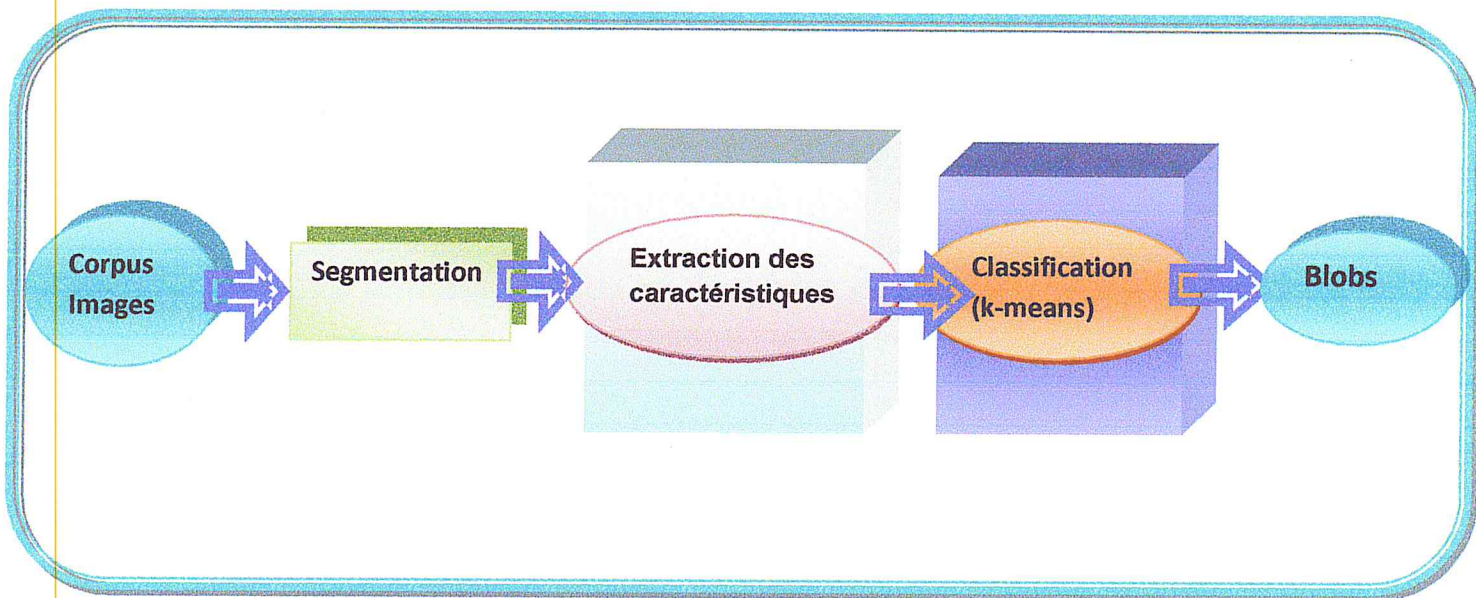


Figure 3.2. Processus d'apprentissage

1) Segmentation

Consiste à décomposer l'image en plusieurs régions, on utilise l'algorithme *NormalizedCut*, on a choisi cet algorithme à cause de leur efficacité et leur utilisation par plusieurs systèmes d'annotation dans le cadre de segmentation d'image il a donné un bon résultat.

Algorithme de Normalized Cuts

- Apporter une image de séquence i
- Construire le graphe $G=(V,E)$ où chaque nœud est un pixel d'image.
- Laisser N être le nombre des points (pixels) $e.i |V|$.

Etape 1

- Construire $N*N$ la matrice X symétrique de similarité comme :

$$w_{ij} = e^{-\frac{\|F(i)-F(j)\|_2^2}{\sigma_f^2}} * \begin{cases} c \frac{-\|X(i)-X(j)\|_2^2}{\sigma_x^2} & \text{if } \|X(i)-X(j)\|_2 < r \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Où $X(i)$ est la location spécial des nœuds i signifie les coordonnants dans l'image original i , et $F(i)$ est le vecteur valeur définit par :

1. $F(i) = 1$ pour la segmentation d'ensemble des points.
 2. $F(i) = I(i)$, la valeur d'intensité, pour la segmentation d'image grise.
 3. $F(i) = [v, u \cdot \sin(h), v \cdot \cos(h)](i)$, où h, s, v sont HSV valeur, de couleur de Segmentation.
1. $F(i) = [|I * f_1|, \dots, |I * f_n|](i)$, où f_n sont DOOG filtrer au différentes balances et orientation, de texture de segmentation.
 - laisser $d_i = \sum_j w_{ij}$ être le totale des connections des nœuds, a tous les autres nœuds.
 - Construction de $N*N$ la matrice diagonale D avec d sur, est le diagonale.

Etape 2

- Résoudre $(D-W)y = \lambda Dy$ généraliser le système de poids et avoir le vecteur de poids avec la deuxième et la plus petite valeur de poids, heureusement, le matlab possède la forme `pows, to solve` (résoudre) le système de poids.

Etape 3

- Utiliser le vecteur de division pour la bipartition du graphe, dans le cas idéal le vecteur de division ne doit prendre que deux valeurs discrètes, et les signes montre comment on partitionne le graphe ($A = \{V_i | y_i > 0\}$, $B = \{V_i | y_i \leq 0\}$).

Soit y est détendu pour prendre la valeur réel, donc, on a besoin de choisir le point terrible, il ya plusieurs façons comme telle.

1. Prendre 0.
2. Prendre la médiane
3. Cherche le point terrible ou les résultats a ce N cut (A,B) sont minimiser N cut valeur peut être trouver par la répétition de calcule de :

$$\frac{y^T(D-W)y}{y^T D y} \quad (5)$$

Ou $y = (1+x) - b(1-x)$ ou $b = k/(1-k)$ ou

$$k = \frac{\sum_{x_i > 0} d_i}{\sum_i d_i}$$

Ou x est N vecteur indicateur dimensionnel ; $x_i = 1$ nœud est dans A et -1 sinon, le point terrible optimale est généralement entour la valeur voulu obtenue du vecteur de poids obtenu.

Heureusement le matlab a la fonction `fminsearch` pour cet exprès.

Etape 4

Répéter la bipartition récursivement arrêté si N cut valeur est plus grand que la valeur perspective de bordure (seuil) (une grande N cut valeur c.-à-d. qu'il n'y a pas une partition clair du point ni plus).

On outre, arrêter si le nombre totale des nœuds dans la partition est plus petite que la valeur perspective de bordure.

2) Extraction des caractéristiques

Dans notre travail on s'intéresse à la texture on a calculée la variance et l'uniformité à partir la matrice de cooccurrence.

- La variance : $var_{t-1}(y_t) = E_{t-1}(\varepsilon_t^2) = \sigma_t^2$
- L'uniformité : $Y = f\left(\frac{x}{a}, b\right) = \frac{1}{b(a,b)} \int_0^x t^{a-1} (1-t)^{b-1} dt$

3) Classification

Consiste à classifier et regrouper les régions similaire dans des blobs, en appliquant l'algorithme **k-means**, nous avant appliqué cet algorithme sur notre processus de classification car il est efficace et performant pour le regroupement des vecteurs des régions similaire.

✚ Algorithme k-means

Note préliminaire : dans la suite, d est issue de la distance euclidienne.

Il procède ainsi (on note S la fonction qui à un point de $\{0..n\} \times \{0..m\}$ associe son niveau de gris) :

1. Choisir k niveaux (distincts) y_1, y_2, \dots, y_k (on les appelle centres des classes)
2. Construire les k classes C_i telles que

$$C_i = \{x | d(y_i, S(x)) < d(y_j, S(x)) \forall j \neq i\}$$

$$y_i = \frac{\sum_{x \in C_i} S(x)}{|C_i|}$$

3. Recalculer le centre des classes :
4. Reconstruire les k nouvelles classes en retournant à l'étape 2. Si les nouvelles classes ne sont pas différentes des anciennes, s'arrêter.

III. 4.2. La partie de traitement

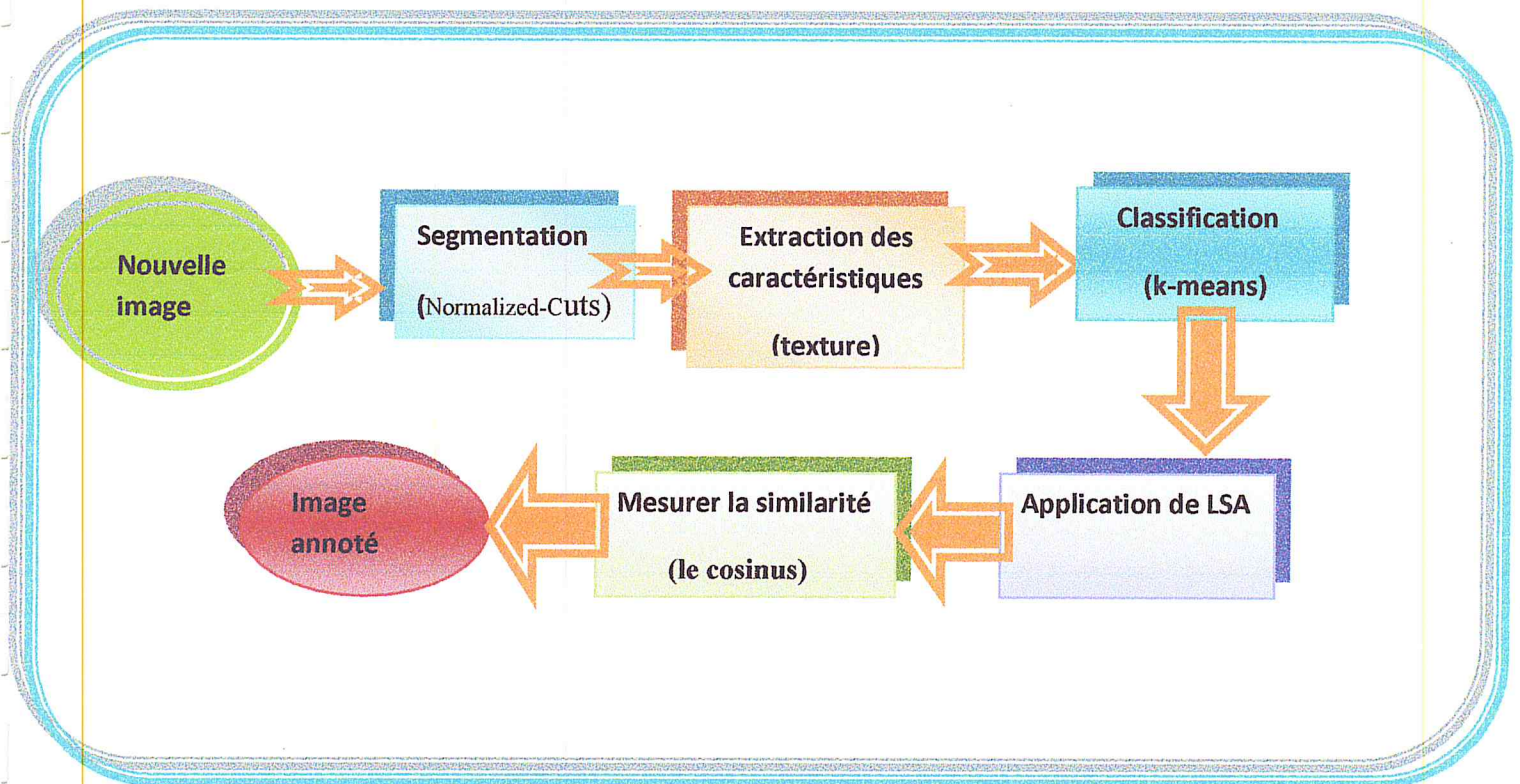


Figure3.3.schéma d'annotation

Dans cette partie on a segmenté une nouvelle image en plusieurs régions avec l'algorithme NormalizedCuts, puis on a fait l'extraction de la texture de ces dernières, en calculant la variance et l'uniformité, ensuite on les a classifiées avec l'algorithme k-means, en effet on a appliqué la méthode LSA, nous avons travaillé par cette méthode pour résoudre le problème du fossé sémantique (synonymie et autonymie).

On a mesuré la similarité. Nous avons utilisé le cosinus comme fonction de similarité. Le choix du cosinus comme mesure de similarité n'est pas entièrement aléatoire, mais dû à la simplicité du calcul, et l'exactitude des résultats par rapport à d'autres mesures comme le produit scalaire et la distance euclidienne.

➤ **Mesure de similarité**

$$Sim(\vec{D}, \vec{V}) = Cos(\vec{D}, \vec{V}) = \frac{\sum_{i=1}^n (d_i \times v_i)}{\sqrt{\sum_{i=1}^n d_i} \times \sqrt{\sum_{i=1}^n v_i}}$$

Avec $\vec{V} = v_1, v_2, v_3, \dots, v_n$

$\vec{D} = d_1, d_2, d_3, \dots, d_n$

Le cosinus de l'angle entre deux vecteurs représentant deux images est égale à 1 si les deux images sont similaires, et égale à 0 si les deux images sont entièrement dissimilaires.

III. 5. La modélisation du système D'auto- annotation- LSA

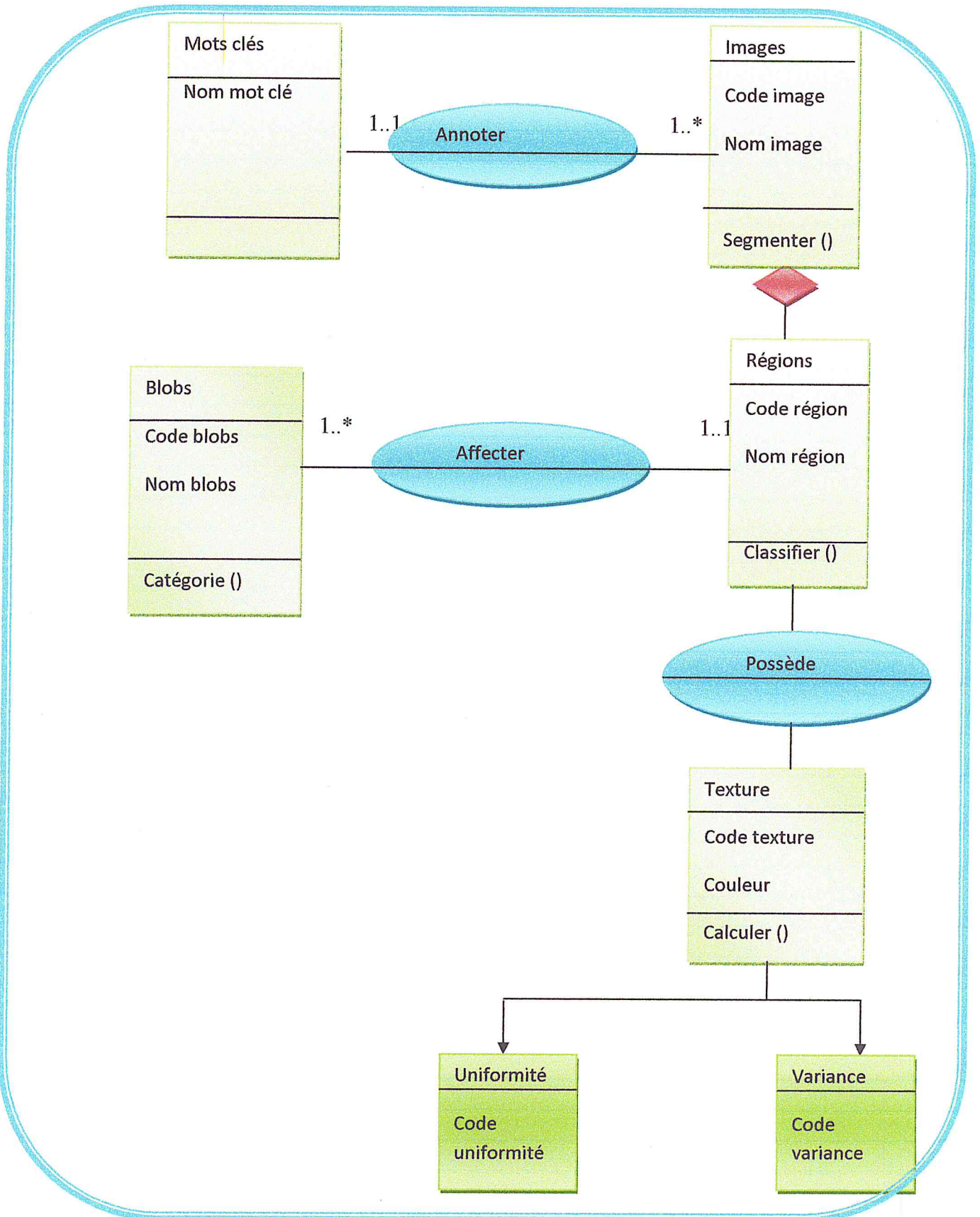
Pour la modélisation, nous allons suivre une démarche orientée objet à la base du langage de modélisation « UML » en passant par les étapes suivantes :

- Présentation du langage de modélisation UML.
- Description du processus de système d'annotation par les diagrammes de cas d'utilisation « use case ».
- Elaboration du diagramme de classe qui constitue l'aspect statique de notre système

III. 5.1. Définition de l'UML

UML (Unified Modeling Language), que l'on peut traduire par « langage de modélisation unifié » est une notion permettant de modéliser un problème de façon standard. Ce langage est né de la fusion de plusieurs méthodes existant auparavant, et est devenu désormais la référence en terme de modélisation objet, à un tel point que sa connaissance est souvent nécessaire pour obtenir un poste de développement objet [8].

III. 5.2. Le diagramme de classe



III. 5.3. Les diagrammes de cas d'utilisation

1) Diagramme de cas d'utilisation pour l'étape d'apprentissage

- Diagramme de cas d'utilisation pour la segmentation d'image

Dans ce cas d'utilisation, on a comme entrée un corpus d'image, le système va les segmenter en plusieurs régions avec l'algorithme NormalizedCut.

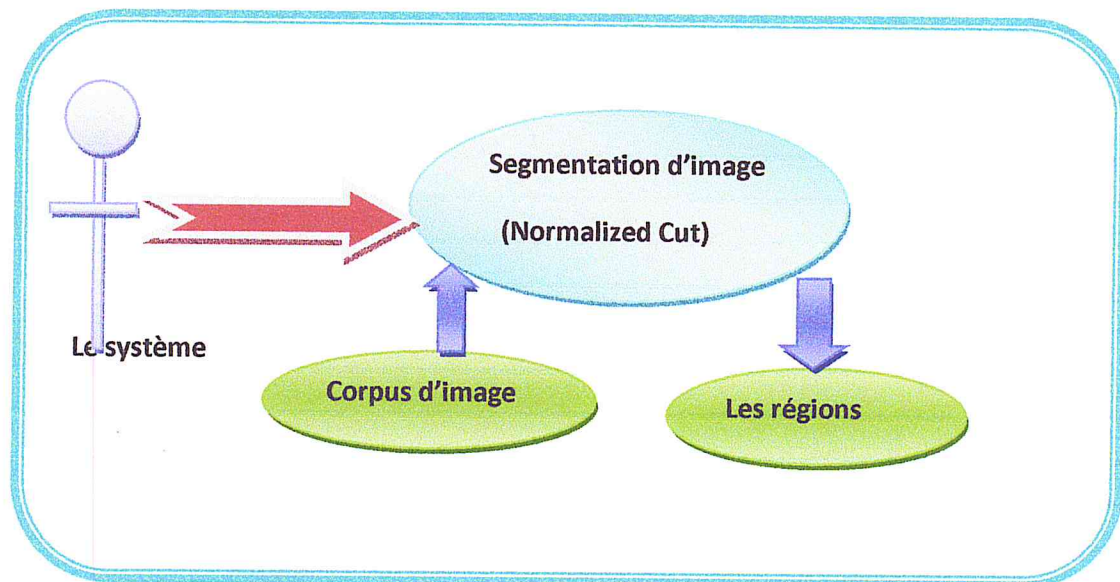


Diagramme de cas d'utilisation pour la segmentation d'image

• Diagramme de cas d'utilisation pour Extraction des caractéristiques

Ce diagramme présent l'extraction de caractéristique (la texture) des régions des images du corpus, notre système calcule pour chaque région le vecteur de la variance ainsi l'uniformité.

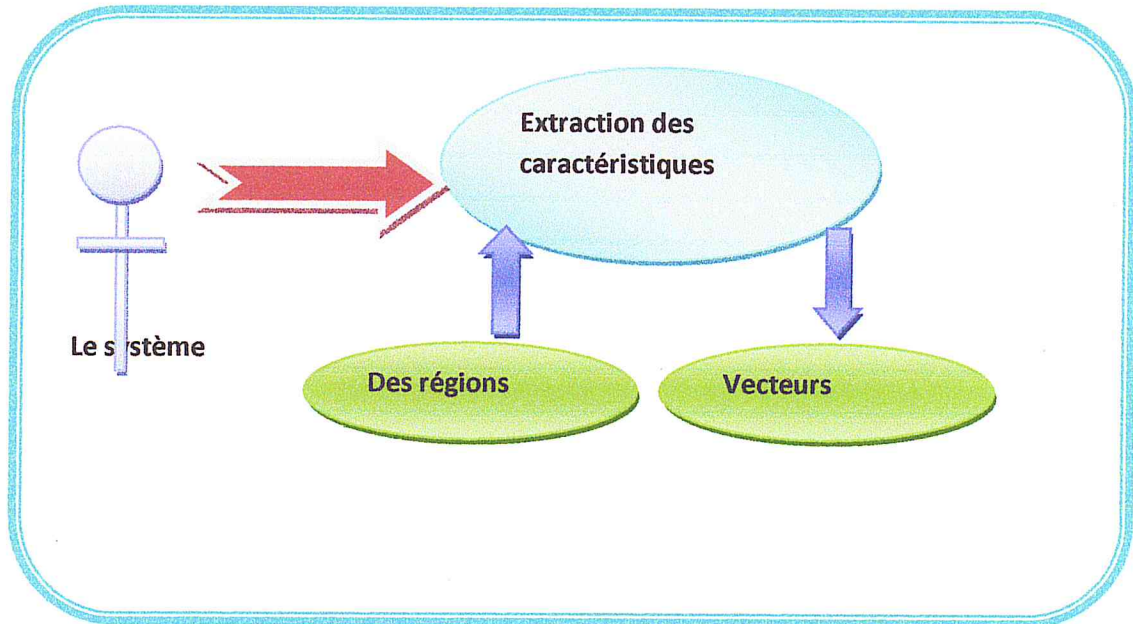


Diagramme de cas d'utilisation pour Extraction des caractéristiques

• Diagramme de cas d'utilisation pour la Catégorisation des blobs

Notre système va catégoriser les vecteurs de caractéristique (la texture) des régions des images du corpus en des blobs, c.-à-d. chaque blob contient l'ensemble des vecteurs de caractéristique similaire.

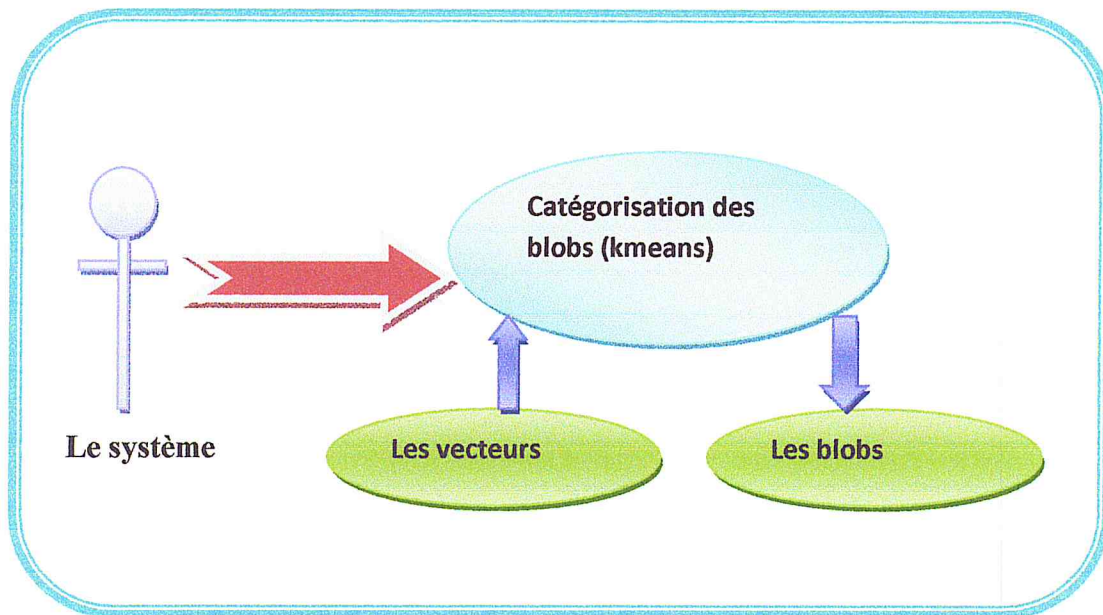


Diagramme de cas d'utilisation pour la Catégorisation des blobs

2) Diagramme de cas d'utilisation pour l'étape de Traitement

- Diagramme de cas d'utilisation pour la segmentation d'image

Dans ce cas d'utilisation, le système va segmenter la nouvelle image introduite, en plusieurs régions avec l'algorithme NormalizedCut.

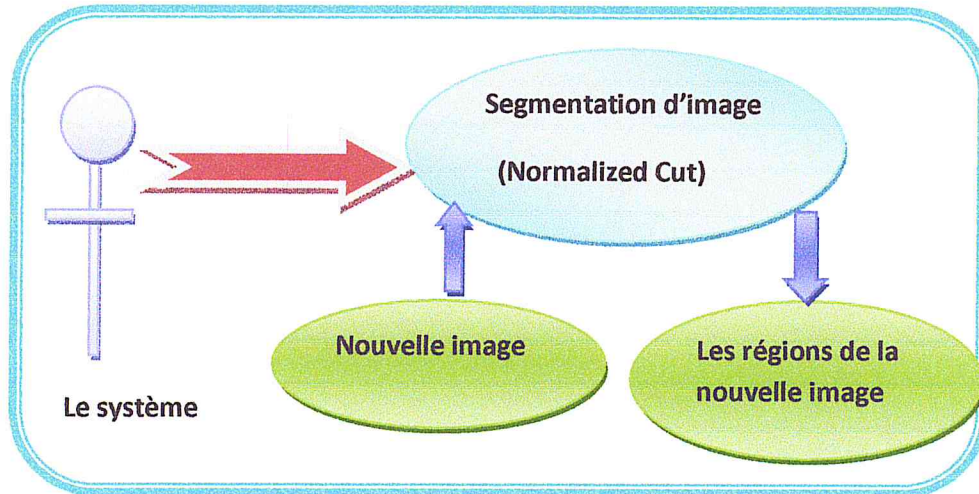


Diagramme de cas d'utilisation pour la segmentation d'image

- Diagramme de cas d'utilisation pour Extraction des caractéristiques

Ce diagramme représente l'extraction de caractéristique (la texture) des régions de la nouvelle image, notre système calcule pour chaque région le vecteur de la variance ainsi l'uniformité.

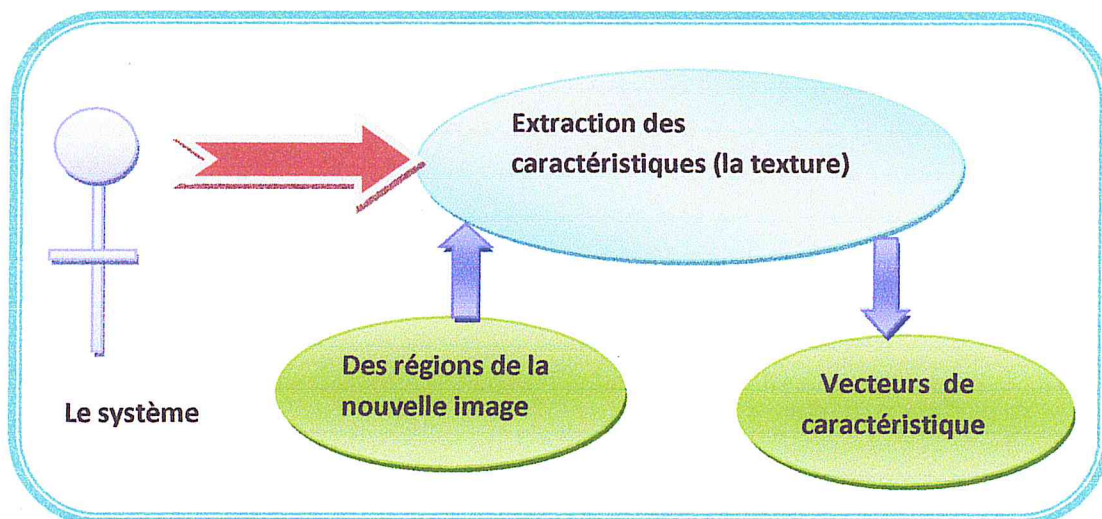


Diagramme de cas d'utilisation pour l'extraction des caractéristiques

• Diagramme de cas d'utilisation pour la Catégorisation des blobs

Notre système va catégoriser les vecteurs de caractéristique (la texture) des régions de l'image introduite en des blobs, c.-à-d. chaque blob contient l'ensemble des vecteurs de caractéristique similaire.

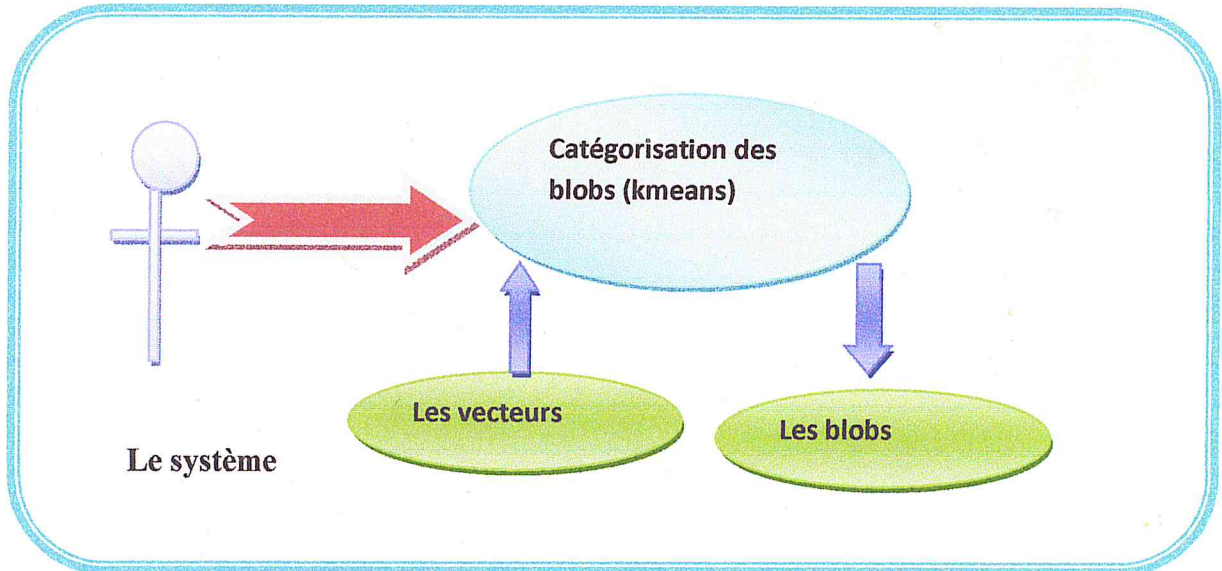


Diagramme de cas d'utilisation pour la Catégorisation des blobs

• Diagramme de cas d'utilisation pour l'application de la méthode LSA

Dans le diagramme ci dessus présente le cas ou le système applique le LSA (Latent Sémantique Analyse) sur les deux matrices, 'image/ mots clés ainsi images /blobs pour annoter la nouvelle image

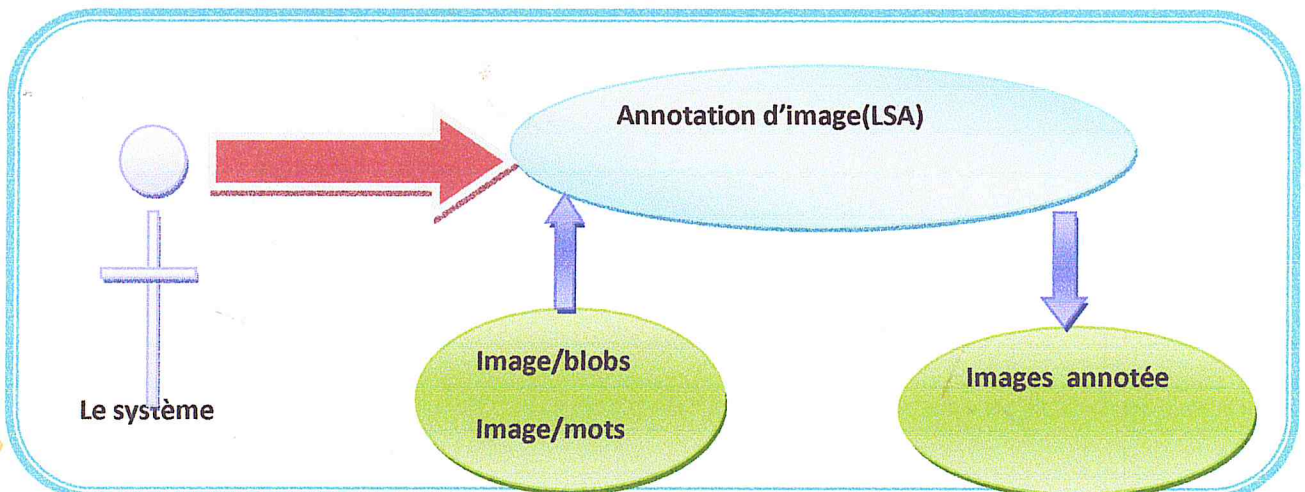


Diagramme de cas d'utilisation pour l'application de la méthode LSA

Chapitre IV

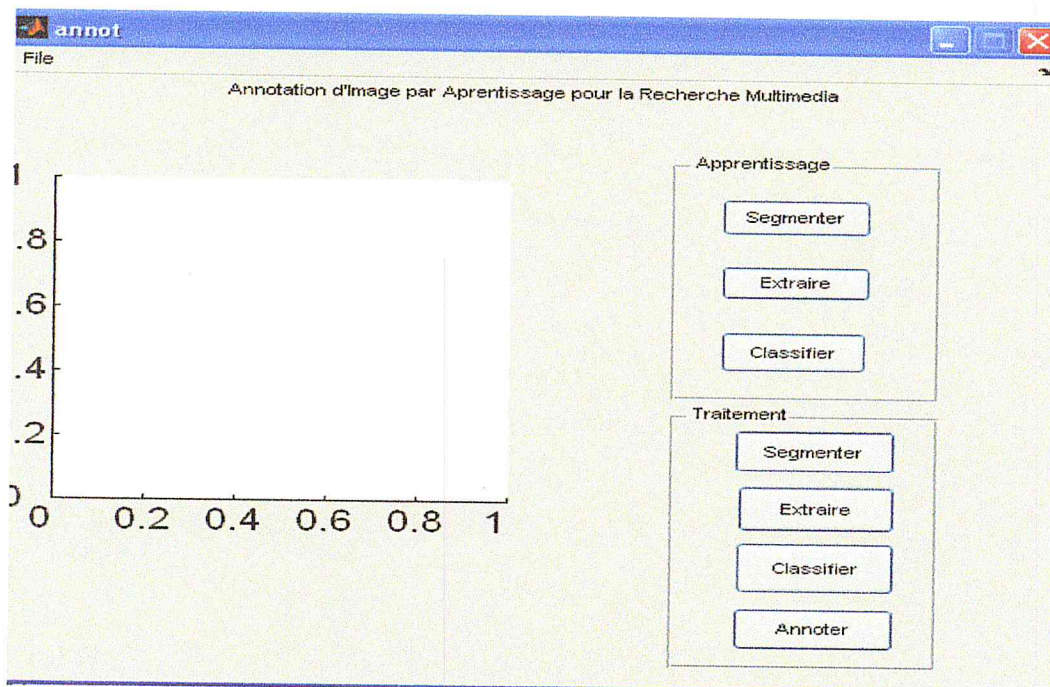
Implementation

IV. 1. Présentation du MATLAB

MATLAB est un interactif basé sur le calcul matriciel (Matrix Laboratory). Il est utilisé dans les calculs scientifiques et les problèmes d'ingénierie parce qu'il permet de résoudre des problèmes numériques complexes en moins de temps requis par les langages de programmation, et ce grâce à une multitude de fonction intégrées et à plusieurs programme outils testés et regroupés selon usage (boite à outils ou Toolbox).

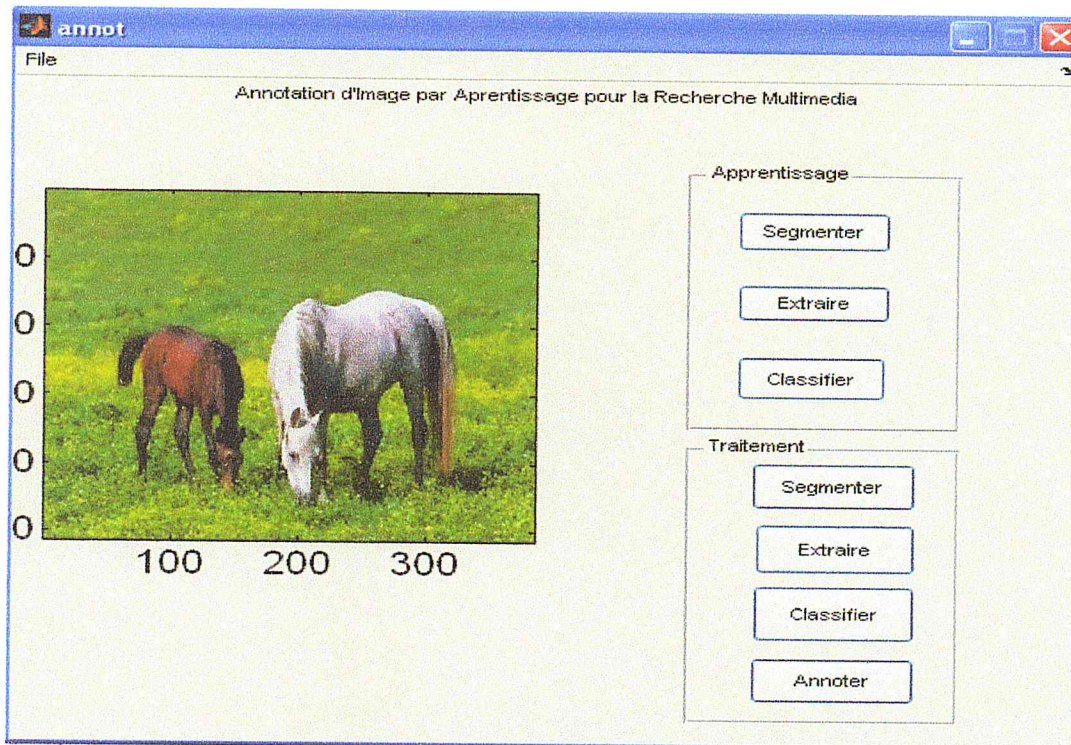
MATLAB fonctionne essentiellement avec un seul type d'objets : des matrices rectangulaire les avec des éléments pouvant être réels, complexes ou symboliques.

➤ L'interface d'Auto_Annotation_LSA



➤ Sélectionner une image dans Auto_Annotation_LSA

Dans le menu File on trouve trois boutons 'open' pour sélectionner une image, et 'print' pour imprimer l'image et, enfin le bouton 'close' pour fermer l'application.



IV.2. Les résultats du système d'auto-annotation-LSA

IV.2.1. La partie apprentissage

Dans cette partie le système fait un pré traitement sur un ensemble d'images généralistes, nous avons segmentée ses derniers ainsi extraire leurs caractéristiques visuels puis les classifier en blobs ou chaque blob contient les régions les plus similaires

➤ La segmentation

Est la première étape d'auto-annotation-LSA ou le système décompose chaque image en scinque segments différents, avec l'algorithme NormalizedCut qui fait d'abord le calcul des caractéristiques (figure 1), ensuite il limite les contours d'image (figure 2) après il affiche leurs scinques segments (figure 3).

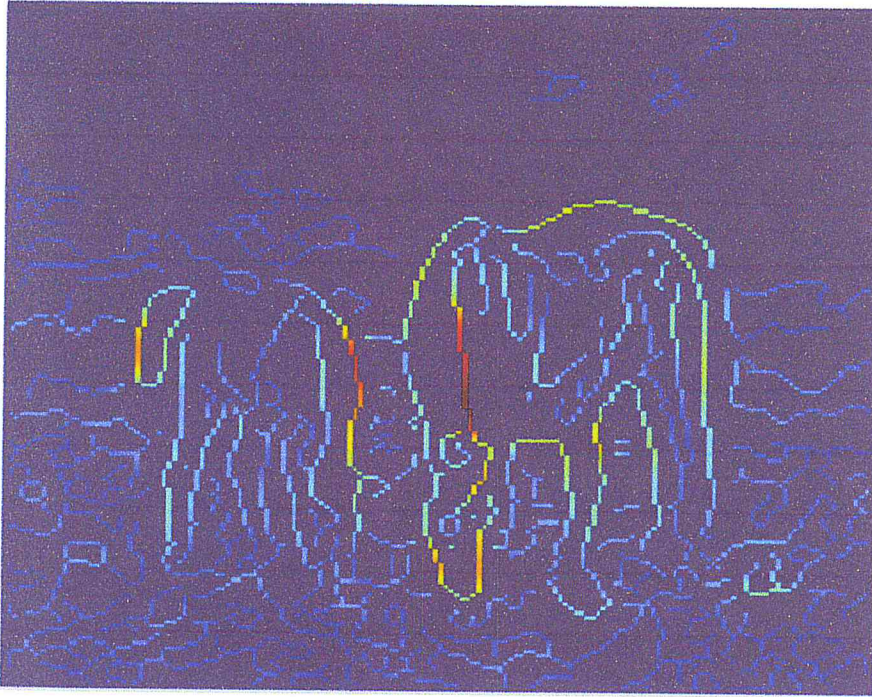


Figure 1. Calcule des caractéristiques

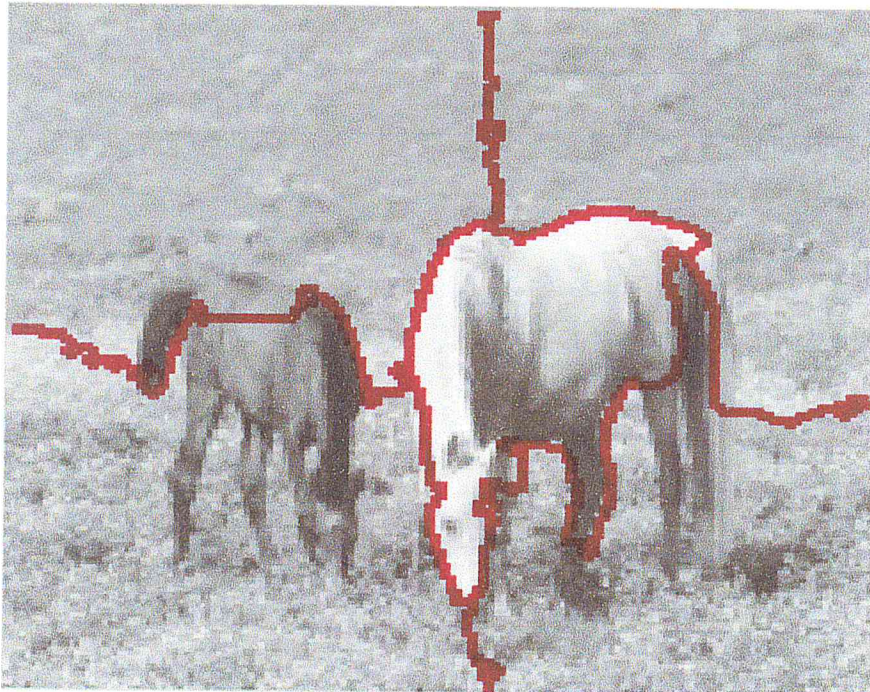


Figure 2. Limitation les contours

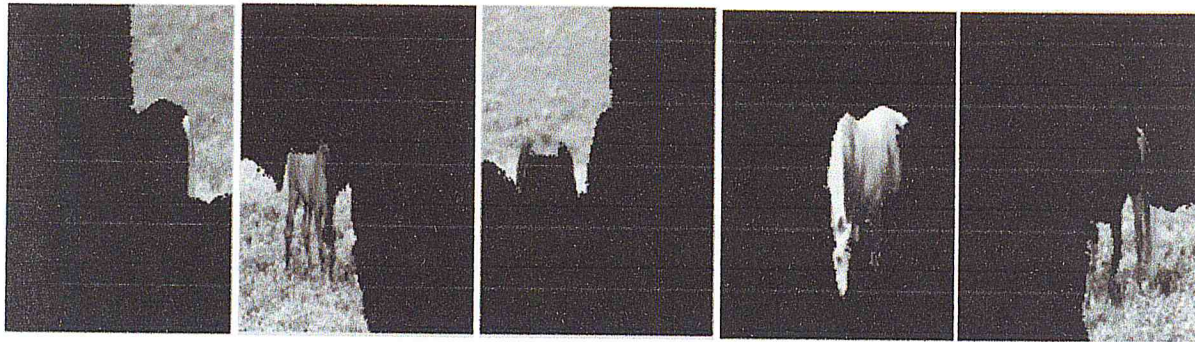


Figure3. Les segments d'image

➤ L'extraction des caractéristiques

Dans cette étape le système calcule l'uniformité (figure1) ainsi que la variance (figure 2) de chaque segment et les rassembler dans une matrice.

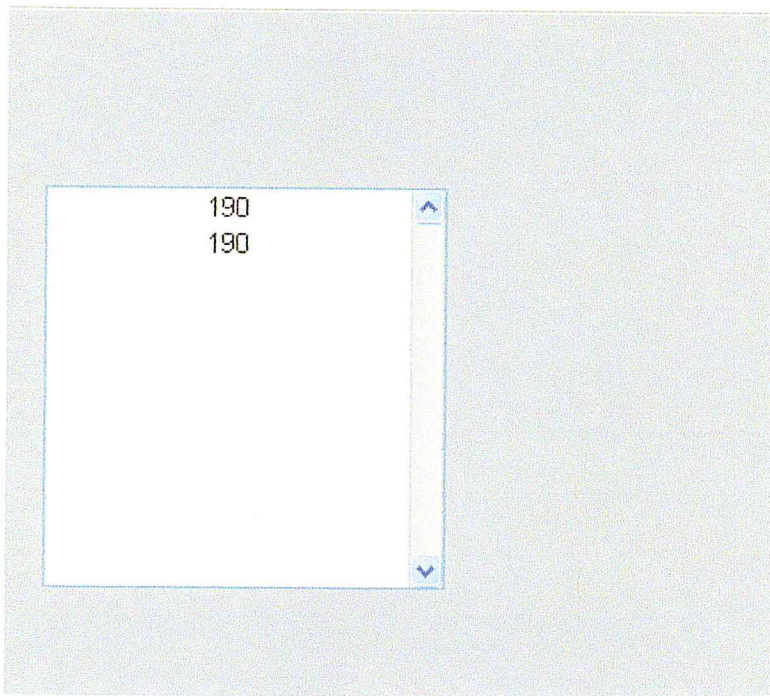


Figure 1.calcul de l'uniformité

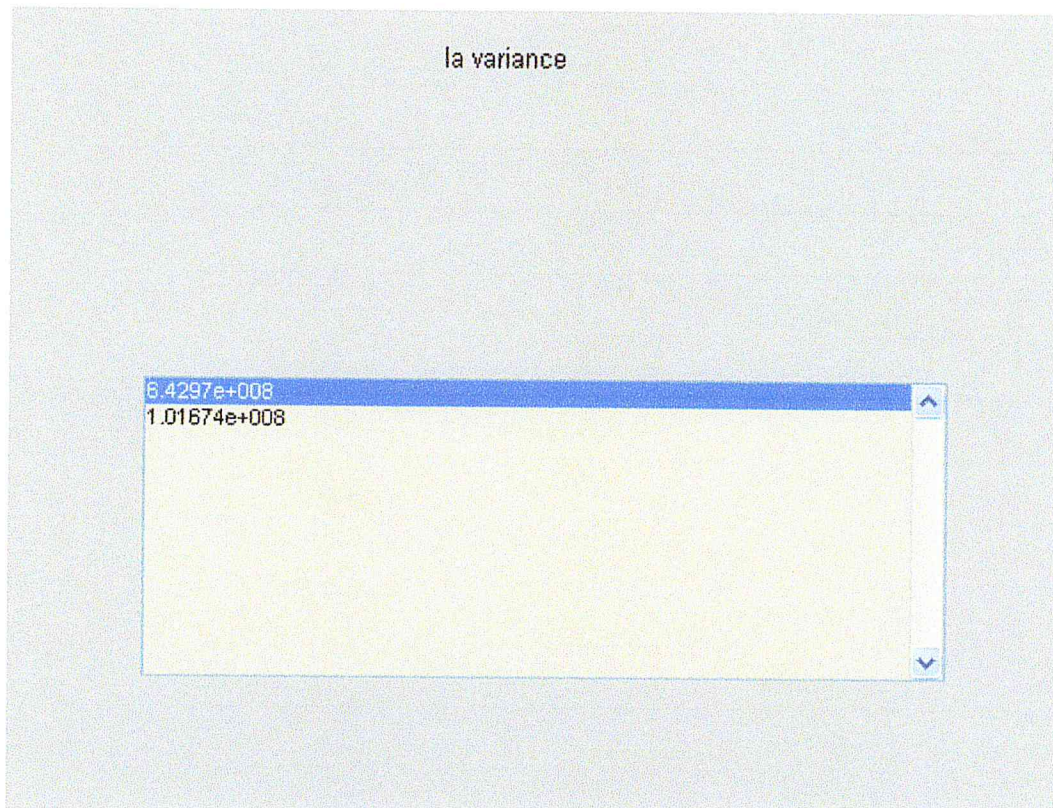
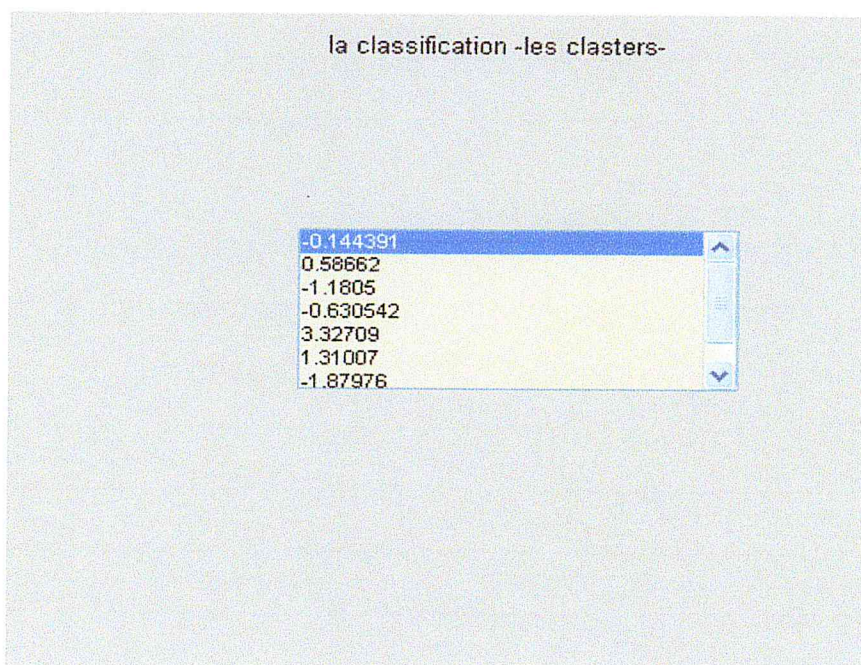


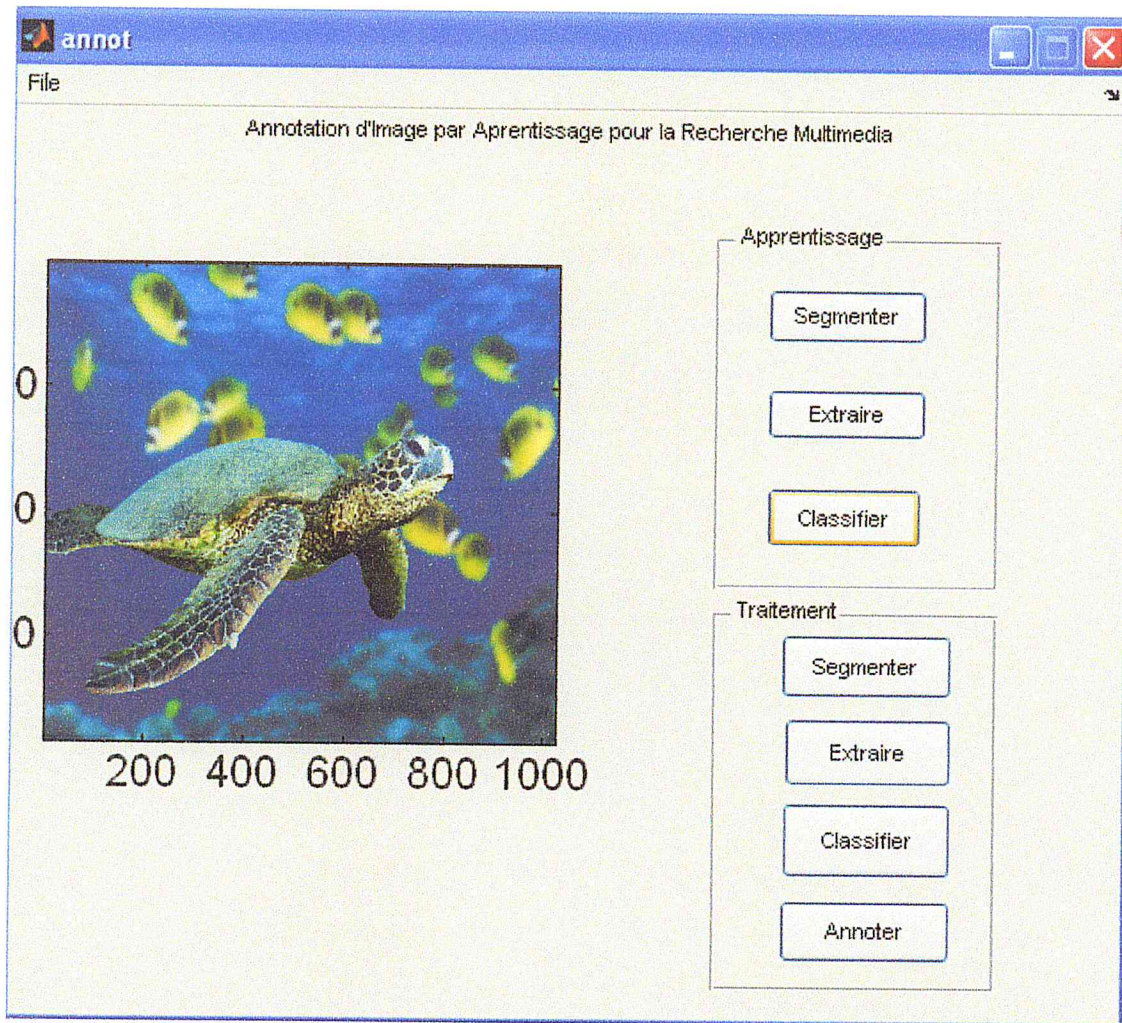
Figure 1.calcule de la variance

➤ La classification

Le système classifié les segments les plus similaires des images segmentée dans des blobs.



IV.2.2. La partie traitement



Dans cette partie le système fait le traitement sur une nouvelle image introduite, il va la segmenter ainsi extraire leur caractéristiques visuels puis la catégoriser dans les blobs ou chaque blob contient les régions les plus similaires.

➤ La segmentation

Est la première étape d'auto-annotation-LSA ou le système décompose chaque image en cinq segments différents, avec l'algorithme NormalizedCut qui fait d'abord le calcul des caractéristiques (figure 1), ensuite il limite les contours d'image (figure 2) après il affiche leurs cinq segments (figure 3).



Figure 1. Calcule des caractéristiques

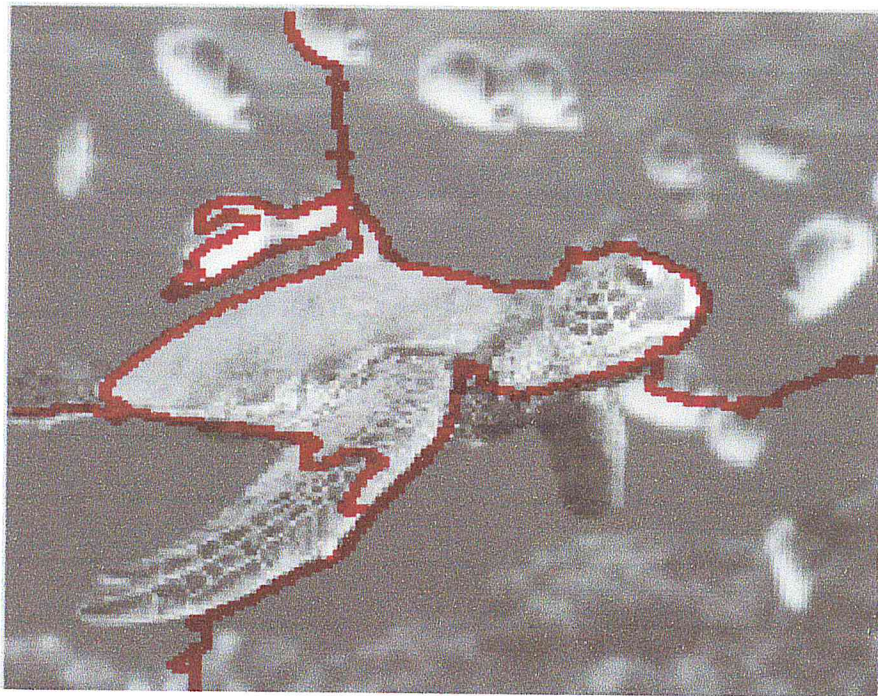


Figure 2. Limitation les contours

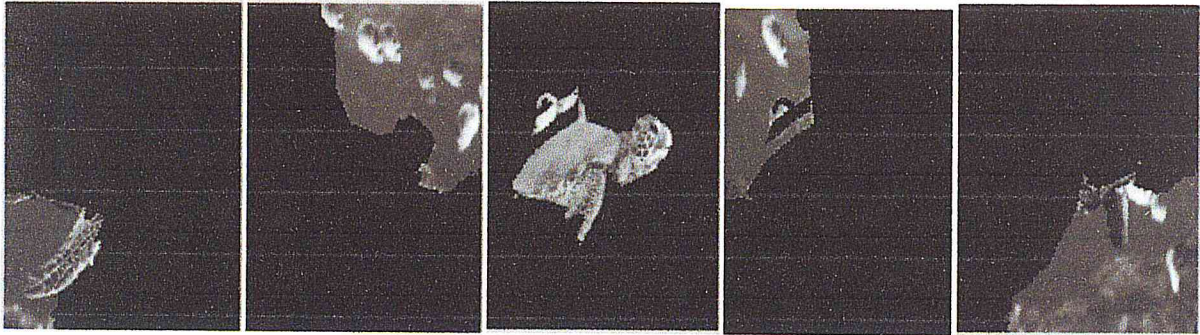


Figure3. Les segments d'image

➤ **L'extraction des caractéristiques**

Dans cette étape le système calcule la variance (figure1) ainsi l'uniformité (figure 2) de chaque segment de la nouvelle image segmentée et les rassembler dans une matrice.

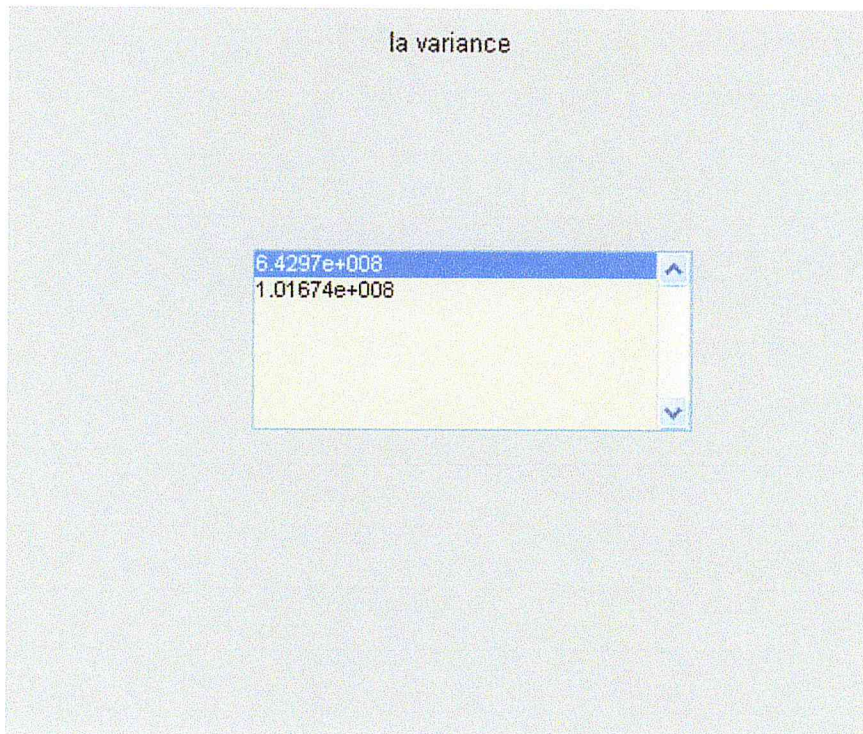


Figure 1.calcul de la variance

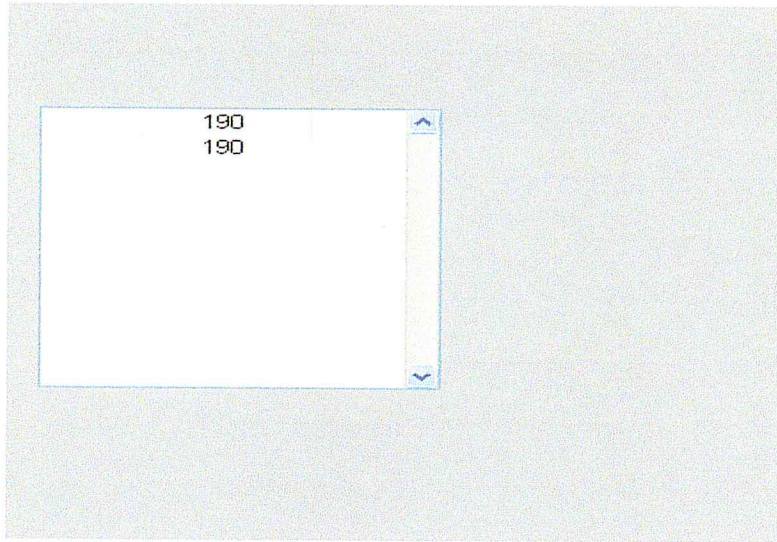
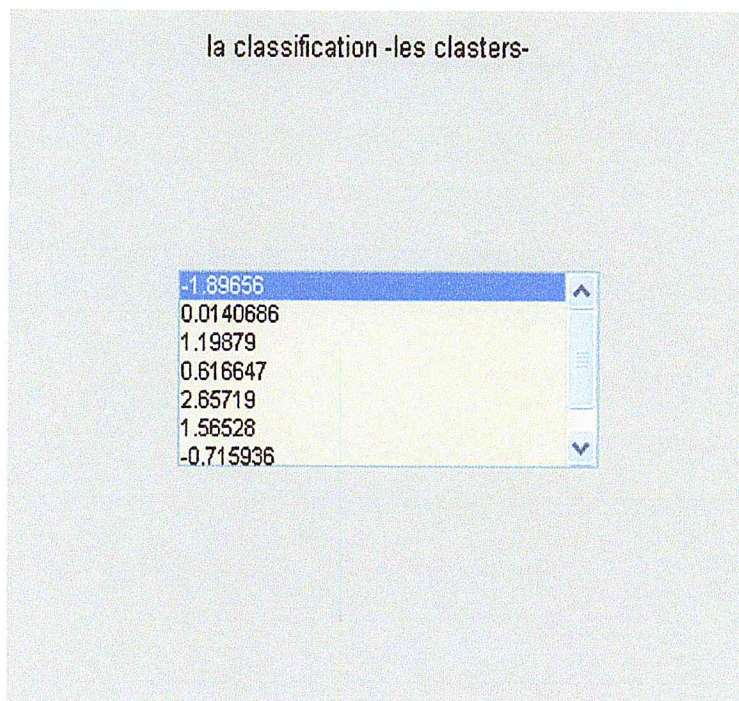


Figure 2. calcul de l'uniformité

➤ La classification

Le système classifié les segments les plus similaires des images segmentée dans des blobs



➤ L'annotation(LSA)

Dans la dernière étape le système applique le LSA sur les deux matrices image x blobs ainsi image x mots clés en inférant les mots clés de la nouvelle image introduite et c'est ca l'objectif d'auto_annotation_LSA.

