

Faculté des Sciences Exactes et de l'Informatique
Département de Mathématiques et d'Informatique
Filière : Informatique

MEMOIRE DE FIN D'ETUDES
Pour l'Obtention du Diplôme de Master en Informatique
Option : **Ingénierie des Systèmes d'Information**

THEME :

Extraction de nouvelles connaissances d'une image-
vidéo indexée par le contenu: Implémentation d'un
nouvel algorithme du Data Mining

Etudiante : « Hadjer BENYAMINA »

Encadrant: « Karim Abdelkader HENNI »

Abstract

After the big progress and development of technology, data process and storage reached its highest level of difficulties of management due to human limited abilities and the existed weak systems. And with the appearance of research and retrieving systems, new challenges appeared especially with the method known as data indexing, in particular: images and videos. In order to create or improve the current systems, developers found that data mining techniques could give super services in collaboration with indexing and segmentation processes.

In this work, we studied some proposed ideas and systems that serve indexing, research and retrieving image and video data. Then, we propose our approach for data indexing using the SIFT descriptors.

Keywords : Image indexing, Video indexing, Image mining, Video mining, CBIR, CBVR, Segmentation.

Résumé

Après le grand progrès et le développement de la technologie, le traitement des données et de stockage atteint son plus haut niveau de difficultés de gestion en raison de capacités humaines limitées et les systèmes faibles existants. Et avec l'apparition des systèmes de recherche et de récupération, de nouveaux défis apparurent notamment avec la méthode dite Indexation des données. en particulier, les images et les vidéos. Afin de créer ou améliorer les systèmes actuels, les développeurs ont découvert que les techniques de fouille de données pouvaient fournir de super services en collaboration avec des processus d'indexation et de segmentation.

Dans ce travail, nous avons étudié quelques idées et systèmes proposés qui servent à l'indexation, la recherche et la récupération des données d'image et de vidéo. Ensuite, nous avons proposé notre approche pour l'indexation des données en utilisant les descripteurs SIFT.

Mots clé : Indexation des images, Indexation des vidéos, Fouille de donnée d'image, Fouilles de données de vidéo, CBIR, CBVR, Segmentation.

Dédicace

Je dédie ce travail A mes chers parents

Pour leur amour , leur soutien indéfinie et leur encouragement.

A ma sœur "Halla" et mon frère " Mehdi "

Et toute ma grande famille maternelle et paternelle.

A mes amis proches, surtout " BILALA Akila & TAÏBI Hadjer"

*Sans oublier tous les enseignants que ce soit du primaire, du moyen, du
secondaire ou de l'enseignement supérieur spécialement à*

"Mr. S. KHELIFA"

*A tous mes amis du face book, Oria club et à toutes l'équipe de la NASA
Apps challenge d'Oran.*

Et à toute personne qui me sent chère.

BENJAMINA Hadjer

Remerciements

Nous remercions Allah le tout puissant de nous avoir donné la force et le courage pour la réalisation de ce travail.

A mon encadreur: " Mr. HENNI Karim Abdelkader " pour son suivis, sa patience indéfinie sa gentillesse et ses précieuses orientations.

Ainsi qu'à ma: "Zoubida SALAH", "Kada BENTATA", "BENABBOU Nouredine Aissa" pour leur aide et conseils.

A: "Moundir DAWADJI" et "HARACHIF Abdelhak" pour leur soutien et leur conseils aussi.

Enfin, à toute personne ayant contribué de pré ou de loin à la réalisation de ce travail.

Sommaire

Résumé/Abstract.....	I
Dédicace.....	II
Remerciement.....	III
Liste des tableaux.....	i
Liste des Figures.....	ii
Introduction générale.....	1
Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu	2
1. Introduction.....	3
2. Définition d'une image.....	3
2.1. Niveau de gris.....	3
2.2. Image numérique.....	3
2.3. Espace de couleur.....	3
2.3.1. Modèle RGB.....	4
2.3.2. Modèle HSV.....	4
3. Définition d'une vidéo.....	5
4. Indexation.....	5
4.1. Indexation des images.....	5
4.2. Indexation des vidéos.....	6
4.2.1. Segmentation temporelle.....	7
4.3. Extraction des caractéristiques.....	7
4.3.1. Caractéristique de couleur.....	7
4.3.2. Caractéristique de texture.....	7
4.3.3. Caractéristique de forme.....	9
4.3.4. Points d'intérêt.....	10
4.3.5. Caractéristiques de Mouvements.....	12
5. Indexation et recherche d'image par le contenu.....	12
6. Indexation et recherche de vidéo par le contenu.....	12
7. Conclusion	14
Chapitre II: Data Mining.....	15
1. Introduction.....	16
2. Définition du Data Mining.....	16

3. Avantages du Data Mining.....	16
4. Méthodes du Data Mining.....	16
4.1. Méthodes prédictives.....	17
4.2. Méthodes descriptives.....	21
5. Mesure de distance de similarité.....	23
6. Image Mining.....	24
7. Video Mining.....	25
8. Conclusion.....	25
Chapitre III: Conception et mise en oeuvre.....	27
1. Introduction.....	28
2. Kit de développement.....	28
2.1. Visual Basic Express.....	28
2.2. Librairie EMGU CV.....	28
2.3. SQL Server 2008.....	28
3. Notre Approche.....	29
4. Expérimentations.....	31
5. Conclusion.....	46
Conclusion générale.....	47
Bibliographie.....	48

Liste des Tableaux

Tableau 1: Différentes structures d'une vidéo.....	6
Tableau 2: Quelques caractéristiques des algorithmes de classification.....	20

Liste des figures

Figure 1: Exemple d'une image niveau de gris.....	3
Figure 2: Espace de couleur RGB.....	4
Figure 3: Espace de couleur HSV.....	5
Figure 4: Représentation simple de l'hierarchie d'une vidéo.....	6
Figure 5: Exemple d'une caractéristique de texture.....	8
Figure 6: Graphe représentant les méthodes du Data Mining.....	23
Figure 7: Système d'Image Mining.....	25
Figure 8: Interface initial de l'application.....	31
Figure 9: Fenêtre de dialogue pour choisir un fichier vidéo à lire.....	32
Figure 10: Au cours du processus d'échantillonnage.....	32
Figure 11: Echantillons d'une séquence d'image (a), (b), (c), (d), (e) et (f) d'une vidéo du soleil avec une haute définition lancé par la NASA. (Vidéo N°1).....	35
Figure 12: Fin du processus d'échantillonnage.....	36
Figure 13: Image0 jusqu'à Image232 représentant le 1er plan.....	36
Figure 14: Image232 et Image233 représentant le 1er changement du plan.....	37
Figure 15: Représentation de quelques fichiers texte qui contient les SIFTs des images extraites dans l'étape d'échantillonnage.....	37
Figure 16: Echantillon du contenu d'un descripteur de points d'intérêt.....	38
Figure 17: Interface de la classification par cluster et extraction des images-clé.....	39
Figure 18: Images-clé extraites de la vidéo N°1.....	40
Figure 19: Interface initiale de la recherche.....	41
Figure 20: Exemple de recherche par requête image.....	42
Figure 21: Echantillonnage de la vidéo N°2.....	43
Figure 22: Fin du processus d'échantillonnage de la vidéo N°2.....	44
Figure 23: Clusters et images-clé représentant de la vidéo N°2.....	45
Figure 24: Détection d'un objet contenu dans une image.....	46

Introduction général

Avec le développement de la technologie et les services qu'elle offre, on aura toujours le besoin de conserver l'information pour continuer à servir l'utilisateur dans tous les domaines. Et puisque l'information de nos jours prend différentes formes, elle peut être un texte, une image, un son ou une vidéo qui est une forme complexe contenant les autres formes ensemble.

Les bases de données de nos jours se caractérisent de la quantité et la qualité des données.

De cette façon, on reste toujours face à de larges bases de données qui contiennent tout type d'information. Et avec ces grandes quantités d'information en stocke, il sera très difficile de faire une recherche rapide et efficace. Par conséquent, la recherche de donnée prend du temps significatif. En particulier, la recherche des vidéos qui est devenue un domaine de recherche très intéressant en cours de ces dernières années.

Dans le but d'améliorer la qualité de service offert (moteurs de recherche, application web ..), plusieurs systèmes et solutions informatiques ont été proposés pour résoudre le problème de la perte du temps ce qui intéresse le plus n'importe quel client ou utilisateur.

Ce travail est composé de trois chapitres:

Le premier chapitre concerne des généralités sur les images et les vidéos et leur indexation par le contenu. On cite quelques méthodes proposées pour l'indexation et la recherche par le contenu.

Le deuxième chapitre met le point sur les différentes méthodes du Data Mining et leurs techniques.

Et le dernier chapitre concerne quelques expérimentations de notre approche suggéré avec les résultats obtenus, pour un future système d'indexation et de recherche de vidéos par le contenu.

Chapitre I :

**Notions générales
sur les images et les
vidéos et leur
indexation par le
contenu**

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

1. Introduction

Les images et les vidéos représente une large quantité d'information stockée dans des bases de données immenses où l'utilisateur est toujours face à un problème de trouver la donnée recherché. Dans ce but, les systèmes d'indexation ont été créé. Dans ce chapitre, on va présenter des généralités sur les images et les vidéos ainsi que quelques méthodes qui permet d'indexer ces derniers.

2. Définition d'une image

L'image est un signale de 2D, souvent, cette image représente une réalité 3D. C'est une fonction à 2D $f(x,y)$ où x et y sont des coordonnées en espace et l'amplitude f en chaque paire (x,y) est appelé *intensité* ou *niveau de gris* de l'image en ce point.[1]

2.1. Niveau de gris

Dans un niveau de gris, chaque pixel est associé par une valeur unique. On considère une image de niveau de gris si le niveau de couleur est unidimensionnelle.[2]



Figure 1: Exemple d'une image niveau de gris.

2.2. Image numérique

C'est une matrice de $n*m$ pixel correspondant à l'échantillonnage et la quantification d'un signal acquis avec un dispositif d'acquisition.[1]

2.3. Espace de couleur

L'espace de couleur est une représentation mathématique 3D d'un ensemble de couleurs où chaque pixel d'image peut être représenté comme un point dans cette espace. Les espaces les plus utilisés dans les CBIR(CBIR: Content-Based Image Retrieval) sont: l'espace RGB(RGB: Red, Green et Blue), l'espace HSV(HSV: Hue, Saturation et Value ou HSL: Hue, Saturation et Lightning ou HSB: Hue, Saturation and Brightning).[3]

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

Les modèles les plus connus sont:

2.3.1. Modèle RGB

C'est un format de codage des couleurs. Ces trois couleurs sont les couleurs primaires en synthèse additive. Elles correspondent en fait à peu près aux trois longueurs d'ondes auxquelles répondent les trois types de cônes de l'œil humain. Elles sont aujourd'hui utilisées en vidéo, pour l'affichage sur les écrans et dans les logiciels d'imagerie. Le système RGB est une façon de décrire une couleur en informatique. Ex: le triplet {255, 255, 255} donnera du blanc.[4]

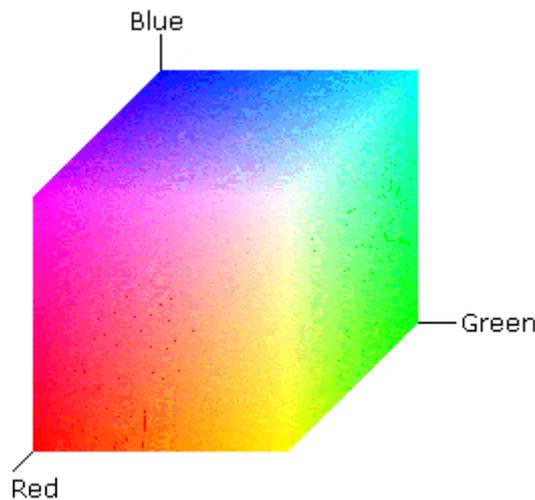


Figure 2: Espace de couleur RGB.

2.3.2. Modèle HSV

Le système HSV est défini par un cylindre qui représente la teinte, la saturation et la valeur d'une couleur. Cette espace est souvent utilisé en analyse d'image parce qu'il permet de séparer les informations relatives à la saturation, l'intensité et la teinte qui est invariante aux conditions d'éclairage et de prise de vue.[3]

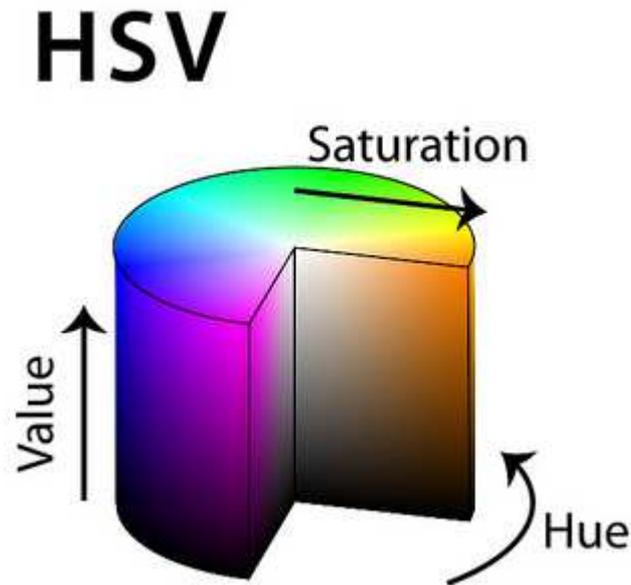


Figure 3: Espace de couleur HSV.

3. Définition d'une vidéo

C'est une technologie de capture électronique, traitement, stockage, transmission et reconstruction d'une séquence d'images accompagné ou non, par des textes, son ou mouvements. C'est un ensemble de scènes animés tel que : Les documentaires, match de sport, Films, télésurveillance..

4. Indexation

L'indexation est un processus associée à la recherche de la donnée dans la mesure où la pose d'index permet d'améliorer la recherche d'un document en répondant à une requête donnée.

Le développement d'Image Mining est basé sur le système de recherche d'images basée sur le contenu. La couleur, la texture, le modèle, la forme des objets et les emplacements de l'image sont à la base du contenu visuel de l'image et sont indexés.[5]

4.1. Indexation des images

On peut indexer une image en se basant sur l'extraction des caractéristiques visuel comme la couleur, la forme, la texture.[6]

L'image peut être indexée par le contexte dans un documents structuré(Manuelle d'utilisation, Journaux, web, rapports...) ou par le contenu où un système idéal doit être robuste aux différentes variations (Luminosité, Orientation, Pose, Occlusions, Expressions du visage et ses variétés).

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

4.2. Indexation des vidéos

Nous savons que les vidéos sont composées d'une séquence d'images individuelles en passant par une hiérarchie détaillé dans la figure 4, ce qui permet de les décrire de la même façon que les images. C'est à dire qu'une vidéo puisse être décrite par l'ensemble des caractéristiques d'une image (Couleur, forme, texture) en ajoutant l'aspect temporel, les mouvements et l'audio.[6]

L'hiérarchie d'une vidéo est présenté comme suit:

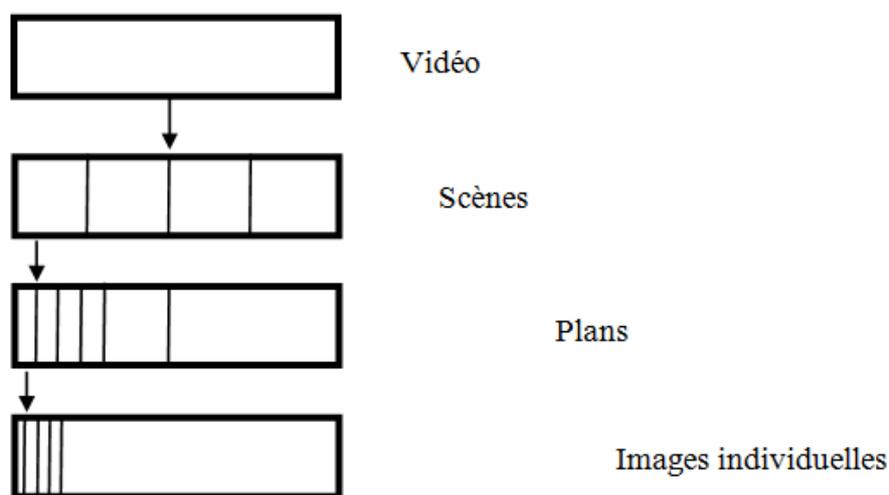


Figure 4: Représentation simple de l'hiérarchie d'une vidéo.

On présente une explication de l'hiérarchie d'une donnée vidéo sous une forme d'un tableau comme suit:

Structure	Définition
Scène	Un ensemble d'événement ou une partie d'une pièce de théâtre, film, programme télévisé... Ex: Scène de crime.
Plan	Un ensemble d'images qui représente le même objet ou appartient au même intervalle spatio-temporel.
Image individuelle	Unité de base d'un plan. Chaque plan est composé d'un ensemble d'images individuelles.

Tableau 1: Différentes structures d'une vidéo.

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

4.2.1. Segmentation temporelle

La segmentation temporelle de la vidéo est une étape nécessaire dans la définition d'une telle structure [Tableau 1] [7]. La segmentation vidéo ou la détection des frontières implique l'identification des changements qui se produit entre les plans vidéos.[8]

Cette étape permette de faciliter le processus d'extraction des caractéristiques de la vidéo.

4.3. Extraction des caractéristiques

L'extraction des caractéristiques est une méthode de capture de contenu visuel d'images et vidéos pour le but d'indexation et de la recherche par le contenu.[9]

4.3.1. Caractéristique de couleur

La couleur est la matière la plus utilisé pour l'extraction des caractéristiques visuel en se basant sur l'utilisation des descripteurs de couleur comme: L'histogramme de couleur; c'est la moyenne la plus commune pour extraire ce genre des caractéristiques.

- **Histogramme de couleur:**

Un histogramme de couleur est une entité 3D où chaque pixel d'une image est placé dans un bin. L'espace de couleur est divisé en bins volumétriques dont chacun représente une gamme de couleurs. Chaque axe de l'espace de couleur peut être divisé indépendamment des autres. Cela permet aux axes d'avoir des résolutions différentes.[10]

Il y'a différentes type d'histogrammes de couleur, on cite quelques un:

- **Conventional Color Histogram:**

L'histogramme de couleur conventionnel d'une image indique la fréquence d'occurrence de toutes les couleurs de l'image. D'un point de vue probabiliste, Il capture les probabilités communes des intensités des canaux de couleurs. En matière du calcul, il est construit en comptant le nombre de pixels de chaque couleur.

- **Fuzzy Color Histogram:**

Dans l'approche d'histogramme de couleur floue, un pixel appartient à tous les bins d'histogramme avec différents degrés d'appartenance à chaque bin.[11]

4.3.2. Caractéristique de texture

La texture est une caractéristique qui fournit des informations significatives sur l'arrangement spatial de la couleur ou des intensités dans une image ou l'identification d'objets ou de régions d'intérêt dans une image ou vidéo. Elle peut être définie techniquement comme une occurrence répétitive d'un même motif.[8]

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.



Figure 5: Exemple d'une caractéristique de texture.

On mentionne que les filtres de Gabor, les matrices de cooccurrence et les ondelettes sont très utilisés en indexation pour la description de la texture.[3]

- **Matrice de cooccurrences**

Elle contient les fréquences spatiales relatives d'apparition des niveaux de gris selon quatre directions :

$$\theta = 0, \theta = \frac{\pi}{4}, \theta = \frac{\pi}{2}, \theta = 3\frac{\pi}{4}$$

Une matrice de cooccurrences est définie au moyen d'une relation géométrique \mathfrak{R} entre deux pixels (x_1, y_1) et (x_2, y_2) par :

$$P_{d,t}(i, j)$$

$$\text{avec } t = \theta$$

Où les indices de la matrice de cooccurrences sont donc les niveaux de gris de la texture étudiée.

$P_{d,t}(i, j)$ représente la répétition où une pair de points séparés par la distance d dans la direction θ ayant les niveaux de gris I_i et I_j . [3]

- **Gray Level Co-occurrence Matrix:**

Dans l'analyse statistique de la texture, les caractéristiques de texture sont calculées à partir de la distribution statistique des combinaisons observées d'intensités à des positions spécifiées l'une par rapport à l'autre dans l'image. Selon le nombre de points d'intensité (pixels) dans chaque combinaison, les statistiques sont classées en statistiques du premier ordre, du second ordre et d'ordre supérieur.

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

La méthode de la matrice de cooccurrence du niveau de gris (GLCM) est un moyen d'extraire des caractéristiques de texture statistique d'ordre secondaire.[9]

- **Filtre de Gabor**

Les filtres de Gabor peuvent servir des filtres passe-bande pour les signaux unidimensionnels. Un filtre de Gabor est obtenu en modulant une sinusoïde avec un gaussien pour le cas de signaux unidimensionnels 1D. Pour les signaux 2D tels que les images, la sinusoïde est combinée avec une fonction gaussienne et la réponse d'un filtre de Gabor à une image est obtenue par une opération de convolution 2D.[12]

- **Gabor Wavelet Transform:**

Dans le domaine spatial, une ondelette de Gabor est une modulation exponentielle complexe modulée par la fonction gaussienne.

En sélectionnant différentes fréquences centrales et orientations de l'ondelette Gabor; il est possible obtenir une famille de noyaux de Gabor, qui peut être utilisé pour extraire des éléments d'une image.[13]

4.3.3. Caractéristique de forme

La forme est une propriété fondamentale d'un objet. La description de la forme consiste en une partie clé de la description de l'image dans le CBIR.[14]

L'extraction des caractéristiques de forme appuie sur des descripteurs de forme que peuvent être utilisés pour une variété de tâches d'enregistrement, à la comparaison, l'analyse de forme et la recherche.

Les descripteurs de forme sont utilisés dans la comparaison entre les formes 3D afin de trouver des points uniques sur la surface, puis font correspondre ces points sur différents modèles. Ils peuvent également être utilisés pour accélérer les recherches en réduisant le modèle à un petit nombre de caractéristiques facilement comparables.

Il se trouve généralement 2 types de descripteurs de forme:

- **Descripteurs de forme basé sur l'approche "Contour":**

Ils décrivent une région au moyen des pixels situés sur son contour "Edges". Ils sont basé sur la détection des frontières de la forme dans une image.

Cette approche fait classiquement référence aux descripteurs de Fourier ou la transformée en ondelettes.

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

○ **Transformée en Ondelettes:**

La transformée d'ondelettes pour un signal 2D comme les images est considéré comme une décomposition du signal par une cascade de filtres en utilisant une paire de filtres pour chaque niveau de résolution :

- *Filtre passe-bas*: Il laisse passer les basses fréquences et atténue les hautes fréquences.
- *Filtre passe-haut*: Il laisse passer les hautes fréquences et atténue les basses fréquences.[1]

Le choix initial d'ondelette est alors remplacé par le choix du filtre. Pour calculer une transformé en ondelettes, on réalise un produit de convolution du signal avec ces filtres.

Une des transformées en ondelettes connues est la transformée de *HAAR*.

– **Transformé de HAAR**

Il s'agit d'une fonction constante par morceaux pour simplifier la compréhension d'ondelettes ainsi que son implémentation. Elles est généralement utilisé pour la détection des personnes et des visages en temps réel.[3]

● **Descripteurs de forme basé sur l'approche "Région":**

Les descripteurs de forme sont dérivés en utilisant toutes les informations de pixels dans une région de forme [15]. Ils fait référence aux moments invariants et est utilisé pour caractériser l'intégralité de la forme d'une région.

○ **Moments de région:**

Les moments de région sont linéairement relatives aux caractéristiques de l'image et sont invariants à l'échelle et la rotation. La détection d'une structure d'image qui peut apparaître dans différentes région avec des tailles et orientations inégales est faite par la construction d'un descripteur d'apparition qu'on nomme: Descripteur de moments de région.

– **Descripteur de moments de région:**

Le calcul d'un descripteur de moments de région est rendu efficace en concevant une procédure rapide basée sur la représentation intégrale de l'image dans n'importe quel domaine rectangulaire.[16]

4.3.4. Points d'intérêt

Le point d'intérêt est le point d'ancrage et fournit souvent les attributs d'invariance à l'échelle, la rotation et l'illumination pour le descripteur. Le descripteur ajoute plus de détails et plus d'attributs d'invariance. Les groupes de points d'intérêt et les descripteurs décrivent ensemble les objets.[17]

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

On cite quelques uns:

- **Scale Invariant Feature Transform (SIFT):**

C'est une méthode d'extraction des caractéristiques très distinctives et invariante à partir d'images qui peuvent être utilisées pour effectuer une correspondance fiable entre différentes vues d'un objet ou d'une scène. Ces caractéristiques sont invariantes à l'échelle et à la rotation de l'image mais ils sont partiellement invariants aux changements des caméra de point de vue 3D et d'illumination.

La bonne localisation de ces caractéristiques réduit la probabilité de la perturbation causé par l'occlusion, le désordre et le bruit.

Le coût d'extraction de ces caractéristiques est minimisé en adoptant une approche de filtrage en cascade.

La génération de ces caractéristiques est basé sur 4 étapes principales:

- ✓ La détection des zones extrêmes et les emplacements d'image. Il est mis en œuvre en utilisant une fonction de différence de Gaussien pour identifier les points d'intérêt potentiels qui sont invariables à l'échelle et à l'orientation.
- ✓ La Localisation des points d'intérêt où à chaque emplacement candidat, un modèle détaillé permet de déterminer l'emplacement et l'échelle. Les points d'intérêt sont sélectionnés en fonction des mesures de leur stabilité.
- ✓ L'affectation d'orientation: Une ou plusieurs orientations sont attribuées à chaque emplacement du point central en fonction des directions des gradients de l'image locale. Toutes les opérations futures sont effectuées sur des données d'image qui ont été transformées par rapport à l'orientation, l'échelle et l'emplacement assignés pour chaque caractéristique, ce qui permet d'inverser ces transformations.
- ✓ Descripteur de points d'intérêt: Les gradients d'image locaux sont mesurés à l'échelle sélectionnée dans la région autour de chaque point d'intérêt. Ceux-ci sont transformés en une représentation qui permet des niveaux significatifs de distorsion de forme locale et de changement d'illumination.[18]

- **Speed-Up Robust Features (SURF):**

Le SURF a été conçu pour assurer une vitesse élevée des trois étapes du SIFT : La détection des points d'intérêt, la description et la correspondance.[19]

Cette approche peut être divisée en trois étapes principales:

- ✓ Les points d'intérêt sont sélectionnés dans des emplacements distincts dans l'image tels que les coins, les blobs et les jonctions.

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

- ✓ Le voisinage de chaque point d'intérêt est représenté par un vecteur de caractéristiques. Ce descripteur doit être distinctif. Au même temps, il devrait être robuste au bruit, aux erreurs de détection et aux déformations géométriques et photométriques.
- ✓ Les vecteurs des descripteurs sont adaptés aux différentes images.[20]

4.3.5. Caractéristiques de mouvement

Les caractéristiques de mouvement d'une séquence vidéo fournissent l'accès le plus simple à sa dimension temporelle et sont donc d'une importance majeure dans l'indexation des vidéos. Un volume énorme d'informations est nécessaire pour exprimer le mouvement d'une séquence typique de vidéo. [21]

5. Indexation et recherche d'image par le contenu

Les approches de la recherche des images indexés par le contenu ont été proposées depuis les années 1990's. Selon [22], La recherche d'image basée sur le contenu (CBIR: Content-based Image Retrieval) est une technique qui permet à un utilisateur d'extraire une image basée sur une requête, à partir d'une base de données contenant une grande quantité d'images. Une question très fondamentale dans la conception d'un système de recherche d'image basé sur le contenu consiste à sélectionner les caractéristiques d'image qui représentent le mieux le contenu de l'image dans une base de données.

La recherche d'image basée sur le contenu est un processus permettant de trouver des images similaires dans le contenu visuel à une requête donnée à partir d'une base de données d'images. Il est habituellement réalisé sur la base d'une comparaison de caractéristiques de bas niveau, telles que la couleur, la texture et les caractéristiques de forme, extraites des images elles-mêmes.[22]

Le contenu des images et de vidéos représente un point très intéressant pour le processus d'indexation et de recherche par le contenu où on a trouvé plusieurs approches qui ont démontré l'utilité d'extraction des caractéristiques visuelle qui donnent de bons résultats.

6. Indexation et recherche de vidéo par le contenu

Récemment, de nombreuses applications sont créées pour la catégorisation, l'indexation et la recherche de contenu vidéo. Ces applications sont utilisées pour gérer la grande quantité de contenu vidéo.

Les recherches se sont concentrées sur des efforts supplémentaires dans la recherche de vidéo qui contiennent certaines informations visuelles plutôt que l'image de leur intérêt.

Une telle recherche est facilitée par les méthodes de recherche de vidéo basée sur le contenu (CBVR: Content-Based Video Retrieval). En particulier, la segmentation de la vidéo est

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

l'étape la plus importante puisque les résultats extraits sont basés sur les limites de segmentation.

La détection de limite de prise de vue peut être effectuée en utilisant diverses techniques différentes comme l'histogramme, le vecteur de mouvement ...etc.[23]

- *Systèmes proposés pour l'indexation et la recherche des images et des vidéos par le contenu:*

L'approche de détection des humains avec le HOG (HOG: Histogrammes de Gradients Orientés) dans [24] représente une étude d'un ensemble de caractéristiques pour la reconnaissance robuste des objets visuels, où une séparation est effectuée sur une base de données de piétons (Détection des personnes les plus visibles dans des poses plus ou moins verticales) en tant que cas de test contenant des images humaines indexé avec une large gamme de variantes de poses et d'arrière-plans en utilisant la SVM linéaire comme classeur de référence pour simplifier et accélérer le processus de détection.

L'idée de base est que l'apparition et la forme des objets locaux peuvent souvent être caractérisée plutôt par la distribution des gradients d'intensité locale ou des directions des arêtes, même sans une connaissance précise du gradient ou des positions de contours correspondantes. Ceci est implémenté en divisant l'image en petites régions spatiales (cellules β), pour chaque cellule accumulant un histogramme 1D local de directions de gradient ou d'orientations de bord sur les pixels de la cellule. Les entrées combinées d'histogramme forment la représentation. Pour une meilleure invariance à l'illumination et l'ombre, il ont normalisé les réponses locales avant de les utiliser. Cela peut se faire en accumulant une mesure de l'histogramme local " l'énergie " sur des régions spatiales un peu plus grandes "blocks" et en utilisant les résultats pour normaliser toutes les cellules dans le bloc. Il est référé aux blocs de descripteurs normalisés en tant que descripteurs d'histogramme de gradients orienté (HOG). Le carrelage de la fenêtre de détection avec une grille dense (superposée) de descripteurs HOG et l'utilisation du vecteur des caractéristiques combinés dans une classe de fenêtre conventionnelle SVM donne la chaîne de détection humaine.

L'utilisation d'histogrammes d'orientation atteint son meilleur performance lorsqu'elle est combinée à l'histogramme et à la normalisation spatiale locale dans l'approche des SIFTs pour l'appariement d'image de base large dans lequel elle fournit le descripteur d'échantillon d'image sous-jacent pour l'appariement des points d'intérêt invariables à l'échelle. Il est mentionnée dans [25] que les approches de SIFTs se comportent remarquablement bien dans cette application.

La représentation HOG / SIFT présente plusieurs avantages. Elle capture une structure de gradient qui est très caractéristique de la forme locale et elle le fait dans une représentation locale avec un degré d'invariance facilement contrôlable aux transformations géométriques et photométriques locales.[25]

Chapitre I: Notions générales sur les images et les vidéos et leur indexation par le contenu.

L'approche proposé dans [9] sert à extraire les caractéristiques de texture d'image à l'aide du descripteur de matrice de cooccurrence de niveau de gris(GLCM). En extrayant les caractéristiques d'une image par une approche GLCM, le temps de compression de l'image peut être considérablement réduit dans la conversion de l'image du RGB vers le niveau gris.

Ces caractéristiques sont utiles dans l'estimation de mouvement de vidéos et en applications de reconnaissance de formes en temps réel comme les applications militaires et médicales. [9]

7. Conclusion

Différentes méthodes ont été proposés afin de faciliter le processus d'indexation des données vidéos et la classification en analysant le contenu des vidéos. A partir de cette petite étude, on conclure qu'il y a toujours une possibilité de proposer de nouvelle méthode pour servir l'utilisateur et la technologie ensemble.

Chapitre II :

Data Mining

1. Introduction

L'exploration de données est l'étude de la collecte, du nettoyage, du traitement, de l'analyse et de l'acquisition de données utiles à partir des données. C'est une partie d'un processus complexe qui est interactive et itérative où les étapes du processus de la découverte des connaissances dans des bases de données KDD (Knowledge Discovery in Databases) doit être faites séquentiellement. Dans ce chapitre, on est concerné par l'étape du Data Mining.

2. Définition du Data Mining

Le Data Mining est la découverte automatique de connaissances non évidentes et de nouveaux faits et relations dans une vaste collection de données.

C'est un processus non triviale d'identification de nouveaux modèles utiles et compréhensibles dans les données. Il se base sur le concept d'extraction des informations précédemment inconnues, valides et actionnables de grandes bases de données, puis l'utilisation de l'information pour prendre des décisions d'affaires cruciales.

Il s'agit d'un domaine interdisciplinaire regroupant des techniques de l'apprentissage automatique, de la reconnaissance des formes, des statistiques, des bases de données, de la visualisation et des réseaux de neurones.[25]

3. Avantages du Data Mining

Le Data Mining a attiré beaucoup d'attentions dans l'industrie de l'information et dans la société en raison de la grande disponibilité d'énormes quantités de données et du besoin de transformer ces données en informations et connaissances utiles. Les informations et les connaissances acquises peuvent être utilisées pour des applications allant de l'analyse du marché, la détection de fraude, et la rétention des clients, au contrôle de la production et l'exploration scientifique.

Le fouille de données peut être considérée comme le résultat de l'évolution naturelle de la technologie de l'information.[26]

4. Méthodes du Data Mining

Les méthodes du fouille de données permettent de découvrir ce que contiennent les données utilisées. Deux grandes familles de méthodes se distinguent:

- Méthodes prédictives.
- Méthodes descriptives.

4.1. Méthodes prédictives

Elles visent à expliquer ou prévoir plusieurs phénomènes observables et effectivement mesurés. [27][28]

- **Classification**

C'est une technique de fouille de données pour prédire la valeur d'une variable catégorisé en construisant un modèle basé sur différents critères (types de données, modèles de données ..)

La classification selon le Data Mining se fait en utilisant des données déjà classifiées pour trouver les modèles prédits.[29]

La classification peut s'appliquer en utilisant différentes algorithmes de classification. Parmi eux:

- **Réseaux de neurones artificielles(ANN):**

Les réseaux de neurones artificiels(ANN: Artificial Neural Network) sont traités comme un outil standard du Data Mining et utilisés pour de nombreuses tâches du fouille des données, telles que la classification des modèles, l'analyse des séries temporelles, la prédiction et le regroupement. Ils sont des modèles informatiques pour le traitement de l'information et sont particulièrement utiles pour identifier les relations fondamentales entre un ensemble de variables ou modèles similaire aux réseaux de neurones biologique, dans les données.[30]

En raison de l'énorme flux de données concernant les objets astrophysiques et la scène astronomique, les astronomes découvrent que l'avenir des données astronomiques consiste à entraîner les machines à reconnaître lorsqu'un télescope a saisi quelque chose de nouveau et intéressant.

Il est mentionné dans [31] qu'un télescope ramènerait 1,5 million de candidats transitoires (phénomènes astronomiques) ce qui rend difficile à l'œil humain de tout observer. Ainsi, les astronomes au Royaume-Uni ont développé des algorithmes basés sur les réseaux de neurones artificiels qui peuvent analyser les relations complexes entre différentes variables telles que la forme, la taille et la couleur des objets astrophysiques, classer les galaxies, puis présenter le type de galaxie approprié(spirale, elliptique..)

Les astronomes ont d'abord entraîné l'algorithme sur 75 000 objets astrophysiques avant de tester ses capacités de classer 1 million d'objets.

La classification des types de galaxies représente la première étape vers la compréhension de l'origine et de l'évolution des galaxies.

Les résultats se sont révélés assez fiables pour se mettre d'accord avec les classifications humaines des galaxies avec un pourcentage de 90%. [31][32]

○ Support Vector Machine(SVM):

Les machines du vecteurs à supports sont un ensemble de méthodes connexes pour l'apprentissage supervisé, applicable à la fois aux problèmes de la classification et de régression. Une SVM choisit un hyperplan de marge maximale qui se trouve dans un espace d'entrée transformé et divise les classes d'exemples tout en maximisant la distance avec les exemples les plus proches.[30]

– Pseudo-code du SVM

1: CandidatSV= { la paire plus proche des classes opposées } (SV: Vecteur du Supports)

2: Tantque il y'a des points de violation Faire

3: Trouver le violateur

4: CandidatSV = candidat du vecteur du supports U violateur

5: Si quelque soit $\alpha_p < 0$ en raison de l'ajout de c à S Alors

6: CandidatSV = candidatSV / p

7: Répéter jusqu'à tous ces point soient élaguées

8: Fin Si

9: Fin Tantque

[33]

○ K-Plus Proche Voisins(KNN):

C'est une méthode d'apprentissage machine non paramétrique utilisé pour la classification. Le k-plus proche voisin(KNN: K- Nearest Neighbour) est un algorithme simple que stocke tous les cas disponible et classifie les nouveaux cas en se basant sur une mesure de similarité comme le calcul d'une certaine distance. L'entrée consiste du k plus proches exemples dans l'espace des caractéristiques utilisé.[34]

– Pseudo-code du KNN

✓ Soit (X_i, C_i) où $i = 1, 2, \dots, N$ sont des points de données. X_i désigne les valeurs de caractéristiques et C_i désigne les étiquettes pour X_i pour chaque i .

✓ En supposant le nombre de classes comme c , $C_i \in \{1, 2, 3, \dots, c\}$ pour toutes les valeurs de i

✓ Soit x un point pour lequel l'étiquette n'est pas connue et on veut trouver la classe d'étiquettes à l'aide de l'algorithmes du KNN.

1: Calculer $d(x, x_i)$ $i = 1, 2, \dots, n$ Où d désigne la distance euclidienne entre les points.

2: Disposer les distances euclidiennes calculées dans un ordre non décroissant.

3: Soit k un entier , prenez les k -premières distances de cette liste triée.

Chapitre II: Data Mining.

- 4: Trouvez ces k-points correspondant à ces k-distances.
- 5: Soit k_i le nombre de points appartenant à la i ème classe parmi les k-points: $k \geq 0$
- 6: Si $k_i > k_j, \forall i \neq j$ Alors
- 7: Mettez x dans la classe i. [35]

On a trouvé dans [29] que SVM et KNN sont suggérés comme les algorithmes du Data Mining les plus prédominants dans la classification des images.

o **Arbre de décision**

L'arbre de décision est l'une des techniques de classification qui se fait selon les critères de fractionnement. L'arbre de décision est un organigramme comme une structure arborescente qui classe les instances en les tri en fonction de la valeur d'une caractéristique (attribut). Chaque nœud d'un arbre de décision représente une caractéristique d'une instance à classer. Toutes les branches désignent un résultat du test, chaque nœud de feuille détient l'étiquette de classe. Les instances sont classées à partir du démarrage en fonction de leur valeur de caractéristique. L'arbre de décision génère la règle pour la classification de l'ensemble de données. Les algorithmes largement utilisés sont ID3 et le C4.5 qui est une extension de ce dernier.

ID3 est une dichotomie itérative 3. C'est un algorithme d'arbre de décision plus ancien. Le concept de base est de faire un arbre de décision en utilisant l'approche gourmande descendante.

C4.5 est un algorithme de l'arbre de décision. L'algorithme C4.5 est largement utilisé en raison de sa classification rapide et de sa haute précision dans les domaine d'apprentissage machine et du Data Mining.[36]

– *Pseudo-code du C4.5:*

Entrée: Base de données d'attributs valorisés D

- 1: Arbre = { }
- 2: Si D est "pure" ou autres critères d'arrêt ont rencontrés Alors
- 3: Mettre fin
- 4: Fin Si
- 5: Pour tout attribut $a \in D$ Faire
- 6: Calculer le critère d'information théorique si on devise sur a
- 7: Fin Pour

Chapitre II: Data Mining.

8: a_{meilleur} = Meilleur attribut selon le critère calculé au-dessus.

9: Tree = Créer un noeud de décision qui teste a_{meilleur} dans la racine.

10: D_v = Sous ensembles induits de D en se basant sur a_{meilleur}

11: Pour tout D_v Faire

12: $\text{Tree}_v = \text{C4.5}(D_v)$

13: Attacher Tree_v à la branche de l'arbre correspondante

14: Fin Pour

15: Retourner Tree [37]

Chaque algorithme de la méthode de classification diffère de l'autre par quelques caractéristiques[30]. On les résume dans le tableau suivant:

Algorithme de classification	Caractéristiques
SVM	Plus utilisé pour la classification. Capture des caractères essentiels de la donnée.
KNN	Simple, effective, non paramétrique et simple à implémenter. Taux d'erreur faible dans le processus d'apprentissage.
ANN	Elle fournit de bons résultats dans les domaines complexes. Meilleure technique pour les données continues. Processus du test est plus rapide.
C4.5	Préférable pour les problèmes du monde réel. Gestion des valeurs manquantes.

Tableau 2: Quelques caractéristiques des algorithmes de classification.

- **Prévision (Prediction)**

En utilisant des techniques du Data Mining, la prévision implique l'analyse des tendances, la classification et l'association. L'analyse des instances ou événements passés rend possible de prédire un autre événement.[38]

Par exemple: Supposons que le responsable du marketing doit prédire combien de clients vont dépenser pendant une vente dans sa compagnie. Dans cet exemple, on prédit une valeur numérique. Par conséquent, la tâche d'analyse des données est un exemple de prédiction numérique. Dans ce cas, un modèle ou un prédicteur sera construit et qui prédit un point de valeur continue ou ordonné.[39]

Chapitre II: Data Mining.

- **Estimation**

L'estimation est similaire à la classification, sauf que la variable cible est numérique plutôt que catégorique. Les modèles sont construits en utilisant des données qui fournissent la valeur de la variable cible.

On prend l'exemple de la pression artérielle d'un patient d'hôpital. En se basant sur son âge, son sexe, son indice de masse corporelle, et le taux de sodium, la relation entre la pression artérielle et le prédicateur variable de l'ensemble d'apprentissage donne un modèle d'estimation. Ce modèle peut être appliqué à de nouveaux cas où on peut estimer la pression artérielle de nouveaux patients.[40]

Les techniques des méthodes prédictives sont basées sur l'apprentissage supervisé.

- *Apprentissage supervisé:*

Les méthodes supervisées sont des méthodes qui permettent de découvrir la relation entre les données d'entrée qui peuvent être des variables dépendantes ou indépendantes. La relation découvert est représenté selon une structure nommé "Un model".[30]

La construction d'un model est le but d'apprentissage machine supervisé. Il fait des prédictions basées sur des preuves en présence d'incertitude. La machine "apprend" à partir des observations. Lorsqu'il est exposé à plus d'observations, la machine améliore sa performance prédictive.

Plus précisément, un algorithme d'apprentissage supervisé prend un ensemble connu de données d'entrée et des réponses connues aux données de sortie et forme un modèle pour générer des prédictions raisonnables pour la réponse aux nouvelles données.[41]

4.2. Méthodes descriptives

Ces méthodes permettent d'organiser, de simplifier et d'avoir une compréhension synthétique d'un ensemble de données. On peut trouver plusieurs types de méthodes de description:

- **Description ou visualisation**

La description et la visualisation peuvent être mono ou multidimensionnelles. Il s'agit de rendre visible des objets ou des concepts qui se trouvent dans des espaces de description trop riches. Il existe plusieurs manières afin de représenter les données à visualiser:

- *Les statistiques élémentaires:*

Les indicateurs les plus utilisés généralement pour la synthèse sont la moyenne, l'écart type le mode et la médiane. [27]

Chapitre II: Data Mining.

- *L'histogramme:*

L'histogramme est une structure de données qui peut être utilisée pour rapprocher la distribution de fréquence des valeurs d'élément dans un flux de données. [42]

- *Nuage des points:*

Il utilise une collection de points placés à l'aide des coordonnées cartésiennes pour afficher les valeurs à partir de deux variables. En affichant une variable dans chaque axe, on peut détecter si une relation ou une corrélation entre les deux variables existe.[43]

- **Regroupement ou Clustering**

C'est le processus de regroupement des objets semblables au sens d'une métrique donnée. Le cluster est une collection des données des objets qui sont similaires dans le même cluster est dissimilaire aux objets contenant dans d'autres clusters.

Le regroupement est également appelé "Segmentation" dans certaines applications car les partitions en clusters regroupent de grandes données en fonction de leur similitude. Le regroupement peut également être utilisé pour la détection des valeurs aberrantes, où les valeurs aberrantes peuvent être plus intéressantes que les cas communs. Les applications de la détection des limites comprennent la détection de la fraude par carte de crédit et le suivi des activités criminelles dans le commerce électronique.[42]

Les méthodes de structuration ont pour objet de repérer ces structures de groupes invisibles à l'œil nu. Il est préférable de catégoriser les objets en classes en fonction de certains critères données ou propriétés communes et ensuite leur attribuer un nom.[27]

- **Association**

L'association (ou la relation) est probablement la technique du fouille des données la plus connue et la plus simple. On fait une corrélation simple entre deux éléments ou plus, souvent du même type, pour identifier les modèles. Par exemple, lors du suivi des habitudes d'achat des gens, on peut identifier qu'un client achète toujours de la crème lorsqu'ils achète des fraises, ce qui suggère que, la prochaine fois qu'ils achètent des fraises, ils pourrait aussi acheter de la crème.[38]

Ces techniques des méthodes descriptives sont classifié sous le type d'apprentissage non supervisé.

- **Apprentissage non supervisé:**

L'étiquette de classe de chaque tuple d'apprentissage n'est pas connue et le nombre ou un ensemble de classes à apprendre ne peut pas être connu à l'avance. Il est référé à l'apprentissage non supervisé par le mot "Clustering" en raison de l'absence d'une base de données d'apprentissage.

Chapitre II: Data Mining.

D'autre façon, l'apprentissage non supervisé n'est pas basé sur des modèles d'apprentissage ou classes prédéfinis.[42][44]

On résume les méthodes du Data Mining par le graphe suivant:

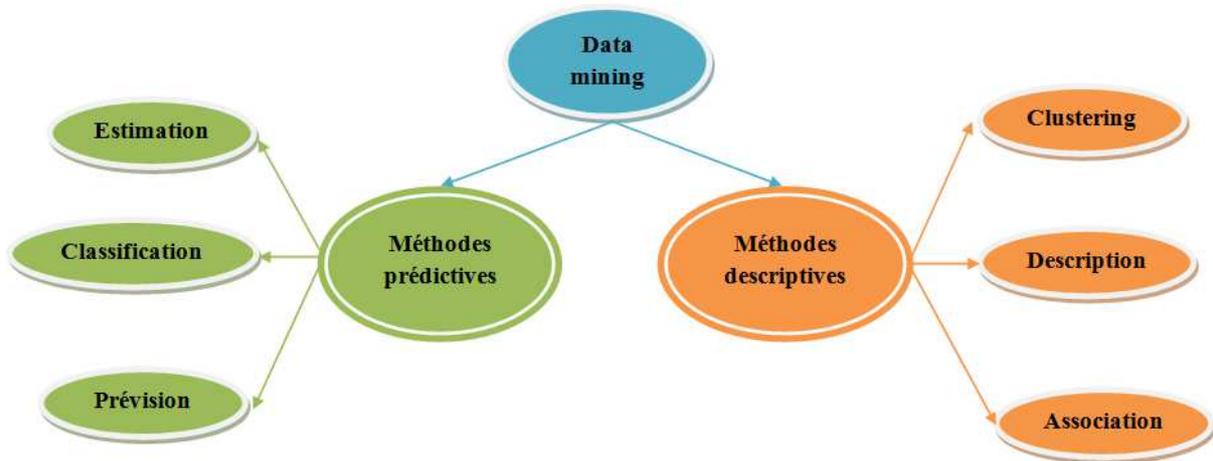


Figure 6: Graphe représentant les méthodes du Data Mining.

5. Mesure de distance de similarité

La mesure de la similarité ou de la distance entre deux points d'une donnée est une exigence fondamentale pour plusieurs tâches dans différents domaines, en particulier, dans le fouille de données et de découverte de connaissances qui impliquent un calcul de distance.[45]

La similitude dans un contexte de fouille de données est généralement décrite comme une distance avec les dimensions représentant les caractéristiques des objets. Une petite distance indiquant un degré élevé de similitude et une grande distance indiquant un faible degré de similarité.[46]

Les mesures de distance les plus connues sont:

- **Distance Euclidienne:**

La distance Euclidienne est la distance entre deux objets sur une ligne droite. Sur 2 dimensions, la formule est définie comme suit:

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}$$

- **Distance Manhattan:**

Ou la distance de bloc, on peut seulement déplacer sur la direction des axes. La formule de distance de Manhattan sur 2-D est :

$$d = |x_1 - x_2| + |y_1 - y_2|$$

- **Distance de corrélation de Pearson:**

C'est une mesure de distance d'attribut numérique qui normalise les valeurs des données d'un objet particulier. Par conséquent, il s'agit d'une métrique de distance utile pour les attributs numériques où un objet particulier peut avoir un biais dans ses attributs.[46]

$$d = 1 - r \text{ avec } r = Z(x).Z(y)/n$$

- r est le produit point des z-scores des vecteurs x et y.
- Z-score de x est construit en soustrayant de x sa moyenne et en se divisant par son écart-type.

Les avantages du Data Mining et la diversité des méthodes du Data Mining qui sont applicable dans de nombreux domaine de vie allant de la médecine, l'économie vers l'astronomie et l'astrophysique, ont permet de l'apparition de deux nouvelle disciplines qui sont l'Image Mining et le Video Mining.

6. Image Mining

L'Image Mining est le processus de recherche et de découverte d'informations et de connaissances précieuses dans de grands volumes de données image.

Certaines des méthodes utilisées pour recueillir des connaissances sont la recherche d'images, l'exploration de données, le traitement d'images et l'intelligence artificielle.[47]

Les étapes les plus importantes d'un systèmes d'indexation et de recherche des images basé sur le Mining sont simplifiées dans le schéma suivant:

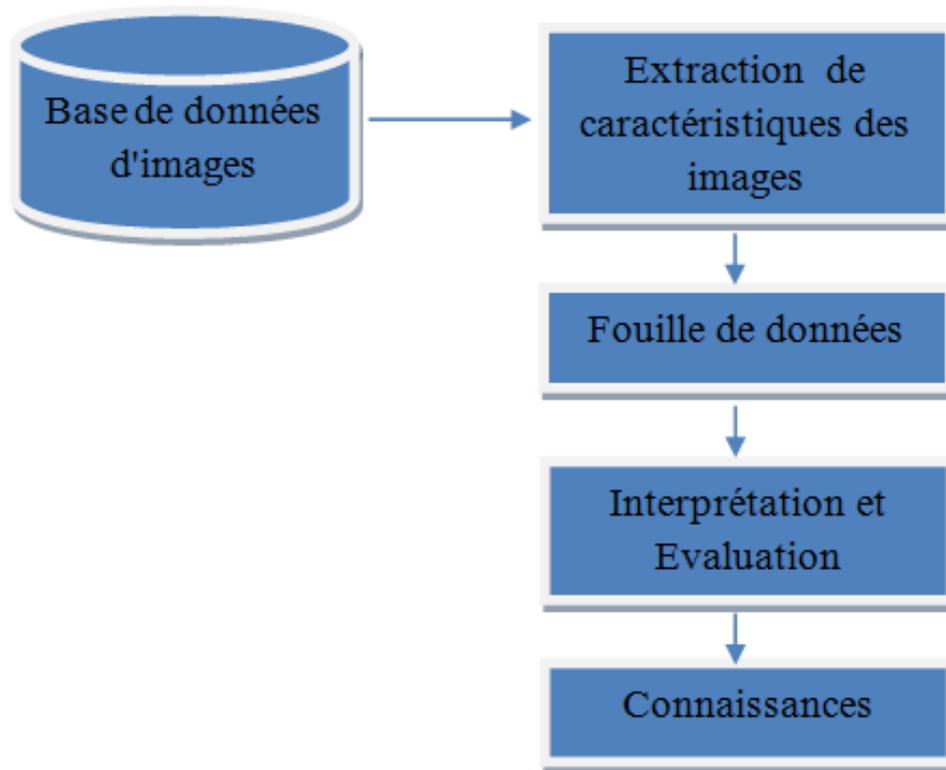


Figure 7: Système d'Image Mining.

7. Video Mining

Le fouille de donnée vidéo peut être définie comme la découverte non supervisée de motifs dans un contenu audio-visuel. Les caractéristiques temporelles (mouvement) et spatiaux (couleur, texture, formes et régions) de la vidéo peuvent être utilisées pour le Mining. L'objectif de l'exploration des vidéos est de découvrir et de décrire des modèles intéressants à partir de la quantité énorme de données vidéo, car il est l'un des principaux domaines problématiques de la communauté de recherche en Data Mining. [42]

8. Conclusion

Dans ce chapitre, on a découvert que le Data Mining est un domaine très vaste et qu'il permet d'appliquer de multiples méthodes en collaboration avec les techniques d'apprentissage machine afin de miner la donnée et extraire de nouvelles connaissances qui serve la recherche scientifiques avec excellence. Chaque méthode du Data Mining a ces caractéristique et façon d'application qui lui diffèrent des autres méthodes; ça permet une meilleur utilisation. Ces méthode ensemble représentent la force du ce domaine.

Dans le domaine d'indexation et de recherche des images et des vidéos par le contenu, plusieurs caractéristiques ont été utilisées(Couleurs, formes, textures ...).

Chapitre II: Data Mining.

Ces caractéristiques ont donnée de bons résultats. Avec l'apparition du Data Mining et leur avantages, on pense qu'on peut les appliquer afin d'extraire de nouvelles connaissances qui puisse grandement améliorer ce concept d'indexation et de la recherche des vidéos par le contenu.

Chapitre III:

**Conception et mise
en œuvre**

1. Introduction

Dans ce chapitre, on va détailler les étapes de notre approche proposé pour l'indexation par le contenu des images-vidéos on citant notre pseudo-algorithme. Puis, on présente quelques expérimentations avec leurs résultats.

Pour la réalisation de notre application, on a choisie le langage de programmation Visual basic en intégration avec la bibliothèque EMGU CV et le SQL Server pour le stockage des données.

2. Kit de développement

Ici, on cite les outils de développement qu'on a utilisé pour réaliser notre travail.

2.1. Visual Basic Express

C'est une version de Visual Basic qui est facile à apprendre. Il s'agit d'un outil de développement complètement fonctionnel pour les programmeurs qui n'ont pas besoin de la version complète de Visual Basic. Il permet également d'apprendre à programmer en Visual Basic.

Visual Basic Express est un outil puissant capable de créer des applications complètement fonctionnelles et des composants que pouvant être partagés.[48]

2.2. Librairie EMGU CV

Emgu CV est une plate-forme croisée .Net. C'est un package de la bibliothèque de traitement d'image OpenCV. Elle permet aux fonctions d'OpenCV d'être appelées à partir du langages compatibles .NET tels que C#, VB, IronPython...

Le package peut être compilé par Visual Studio et autre outils de développement. Il peut fonctionner sous Windows, Linux, Android, Windows Phone...[49]

2.3. SQL Server 2008

SQL Server Management Studio (SSMS) est un environnement intégré pour accéder, configurer, gérer, administrer et développer tous les composants de SQL Server.

SSMS combine un large groupe d'outils graphiques avec un certain nombre d'éditeurs de scripts riches pour fournir l'accès à SQL Server aux développeurs et aux administrateurs de tous les niveaux de compétences.[50]

3. Notre Approche

L'idée de notre proposition est décrite selon 3 étapes:

- **Echantillonnage:**

Cette étape sert à capturer séquentiellement et temporellement des images individuelles de l'entrée vidéo.

- **Calcul des SIFTs:**

Pour chaque image capturée dans la première étape, on calcul le descripteur de points d'intérêt SIFT.

- **Stockage des données:**

Dans cette étape:

- On stocke les données d'images (Identifiant de l'image, Nom de l'image et la taille des SIFTs) dans une base de données créer sur SQL Server 2008.
- On stocke les images échantillonnées sur des fichiers (.jpg).
- On stocke les descripteurs SIFTs sur des fichiers (.txt)

- **Extraction des images-clé:**

Selon les SIFTs extraits dans la 2eme étape, on définit une mesure de distance d pour comparer entre les images afin de composer des regroupements des images similaires selon un seuil donnée.

Seuil: On définit la valeur du seuil entre le 0 et le 1 où :

- 0 représente une similarité total (100%).
- 1 représente une dissimilarité total.

Pour toute I_i et I_{i+1} :

$$\{I_i, I_{i+1}\} = \begin{cases} 0 & , \quad d_i < \text{Seuil} \\ 1 & , \quad d_i > \text{Seuil} \end{cases}$$

On indique que l'image de l'indice $i+1$ est l'image représentant la 1ère image du plan suivant et l'image de l'indice i représentant la dernière image du plan précédent.

En comparant entre chaque I_i et I_{i+1} , on peut trouver l'image qui représente le changement du plan, celle-ci est l'image-clé I_{i+1} qui représente un certain plan.

L'extraction des images-clé est une réduction considérable de la base de données initiale, cela facilite beaucoup le processus de la recherche qui est le but de tout processus d'indexation.

- Et voila le bloc d'instruction de notre Pseudo-algorithme:

Pseudo-Algorithm:

$i = 0$ (Indice de l'image)

$Pl = []$ (Vecteur des plans)

$P = 1$; (Indice du plan)

Entré: Fichier video V

Sortie: Image-clé kf

1: **Lire**(V)

2: **Tanque** (durée (V) $\langle \rangle$! fin(durée(V)))) **Faire**

$I_i \leftarrow \text{Capture}(V)$ (I: Image capturée)

$\text{jpg_file} \leftarrow \text{Sauvegarder}(I_i)$

$S \leftarrow \text{Extract SIFT}(I_i)$

$\text{txt_file} \leftarrow S$

$\text{Vecteur}[i] \leftarrow \text{Taille}(S)$

$i = i+1$

Fin Tanque

3: **Lire** seuil

$\text{kf} \leftarrow I_0$

$Pl[1] \leftarrow I_0$

4: **De** $i=1$ **à** Count(I_i)

Si ($|\text{Vecteur}[i] - \text{Vecteur}[i+1]| > \text{seuil}$) **Alors**

$\text{kf} \leftarrow I_{i+1}$

$P++$

Sinon $Pl[P] \leftarrow I_i$

Fin si

4. Expérimentations

Notre proposition a été testé sur des courtes vidéos de différents domaines : Nature, Espace... etc. Depuis nos expérimentations, on a obtenu des résultats considérable selon différentes valeurs du seuil.

Dans la 1 ère fenêtre, on a utilisé un composant de Lecteur Windows Media par défaut pour lire un fichier vidéo stocké dans le même ordinateur que l'application.

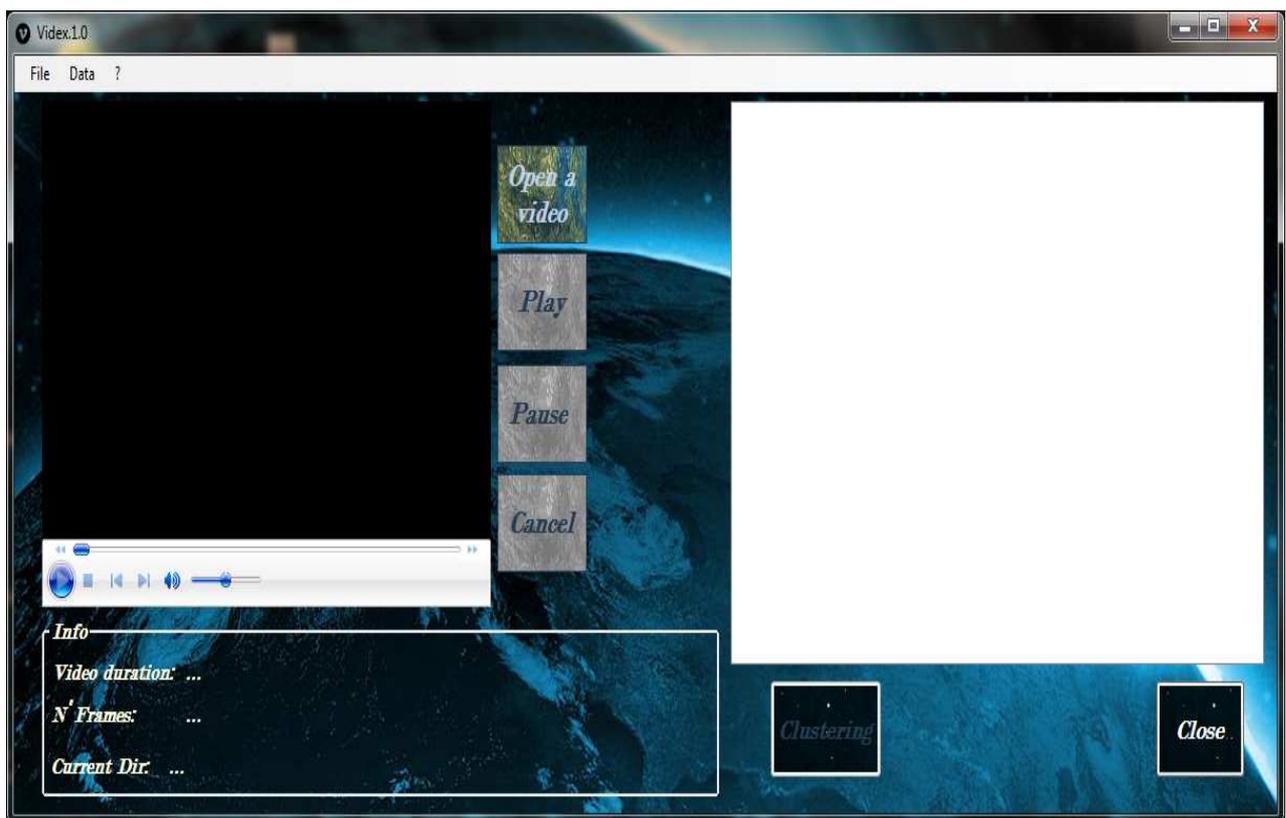


Figure 8: Interface initial de l'application.

- ✓ **Open a video:** Ce bouton sert à choisir une vidéo.
- ✓ **Pause:** Il sert à arrêter la vidéo et le processus de la segmentation.
- ✓ **Play:** Il sert à reprendre le processus arrêté par le bouton Pause.
- ✓ **Cancel:** Il sert à annuler le processus.
- ✓ **Video duration:** Cette étiquette affiche la durée de la vidéo par seconde.
- ✓ **N°Frames:** Elle affiche le nombre total des images capturées.
- ✓ **Current Dir:** Elle affiche le lien de sauvegarde des fichier .jpg des images.
- ✓ **Close:** Ce bouton permet de sortir de l'application.

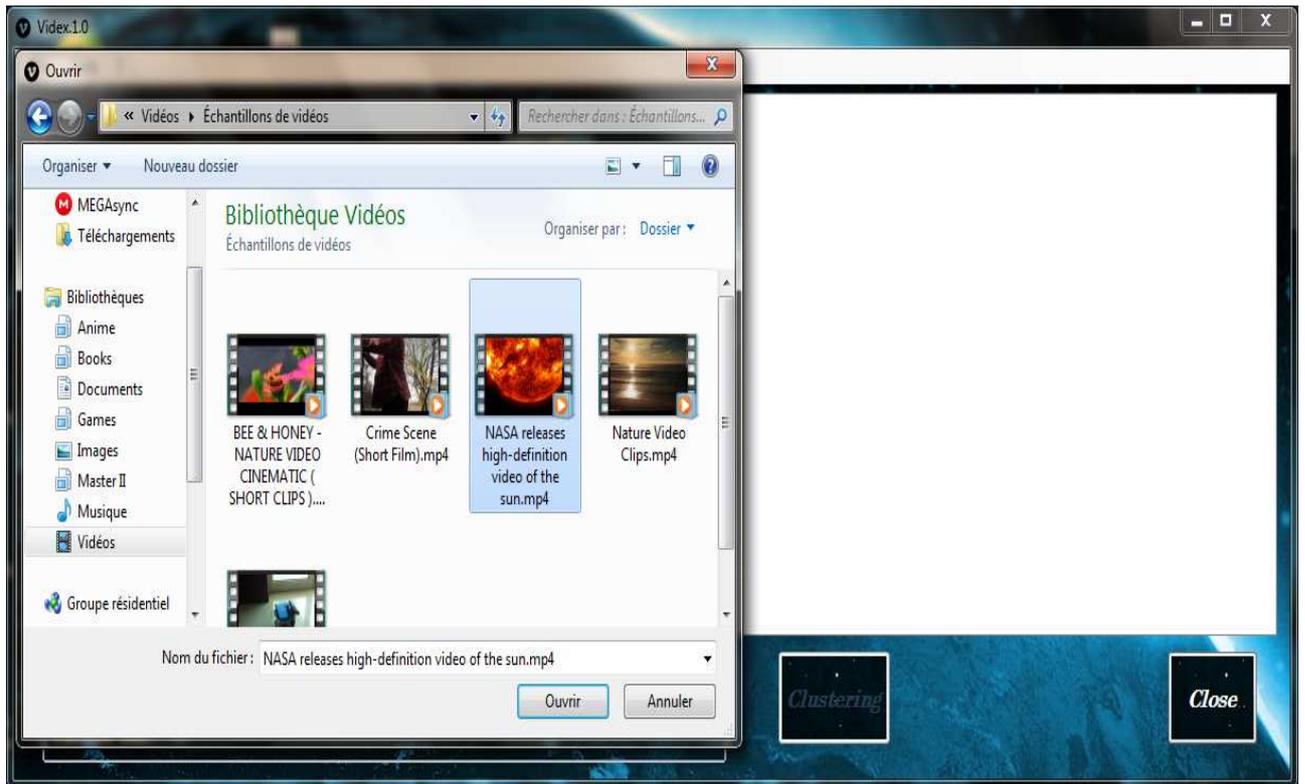


Figure 9: Fenêtre de dialogue pour choisir un fichier vidéo à lire.

- Quand on choisie la vidéo, elle est directement lu sur le Lecteur Windows Media et échantillonné au même temps.

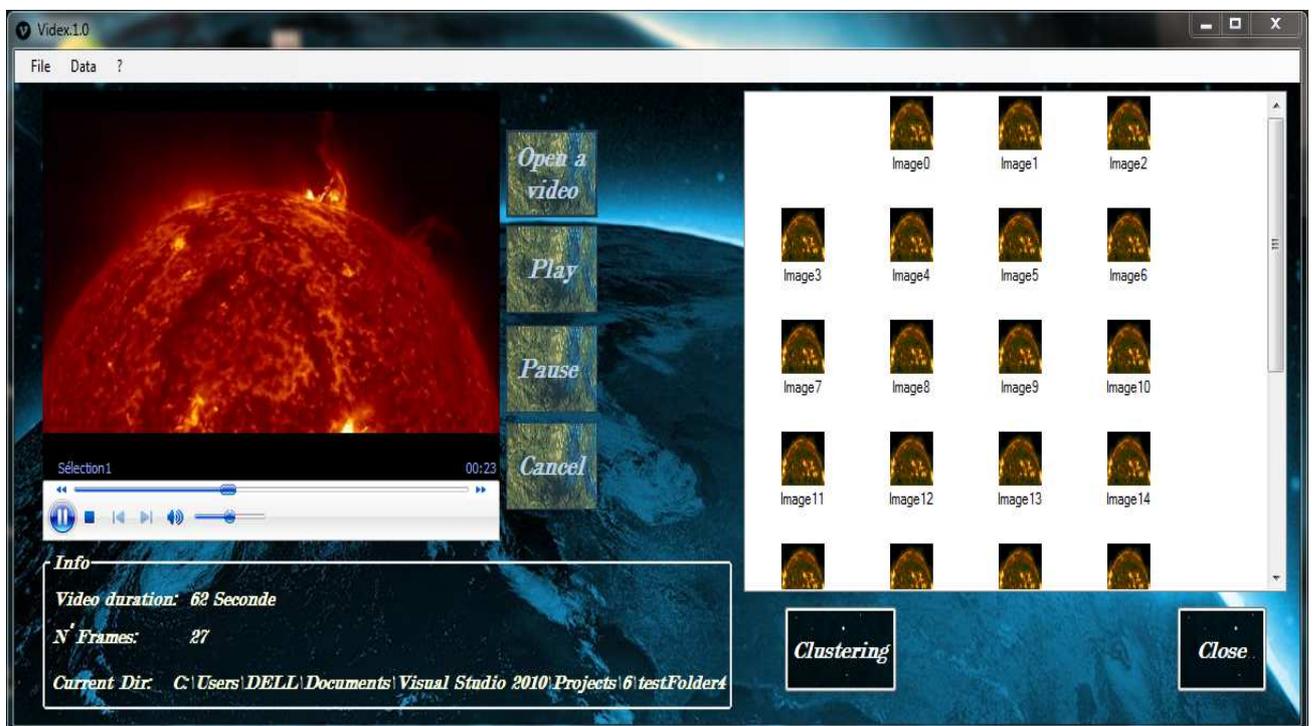
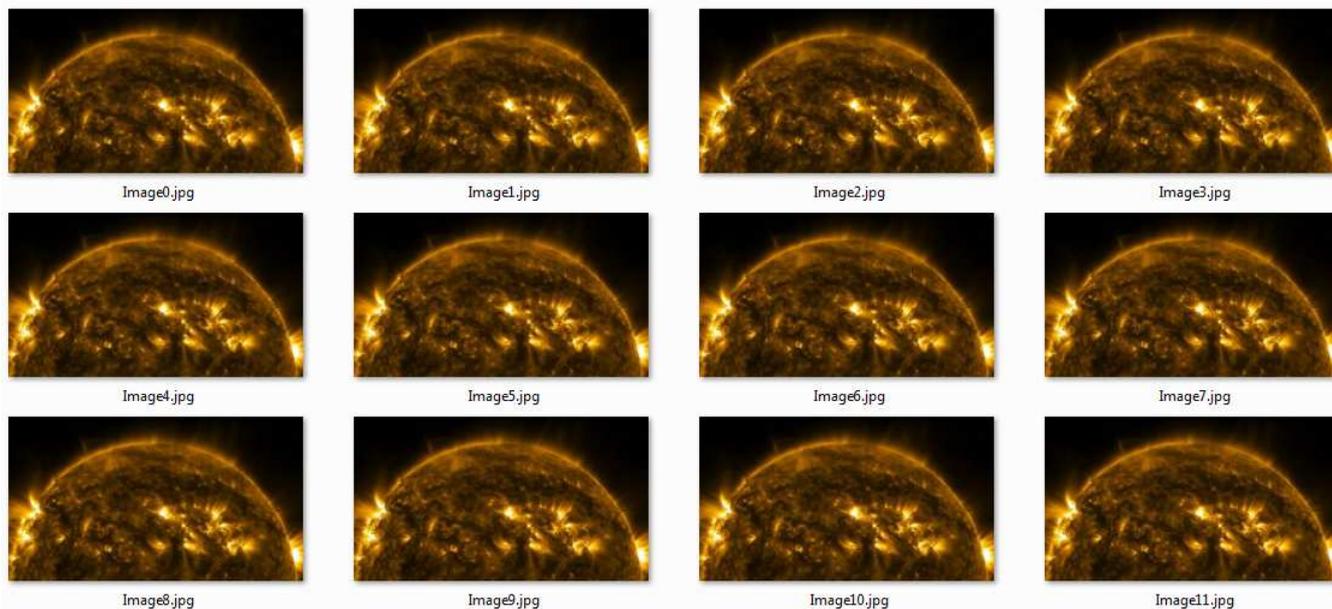


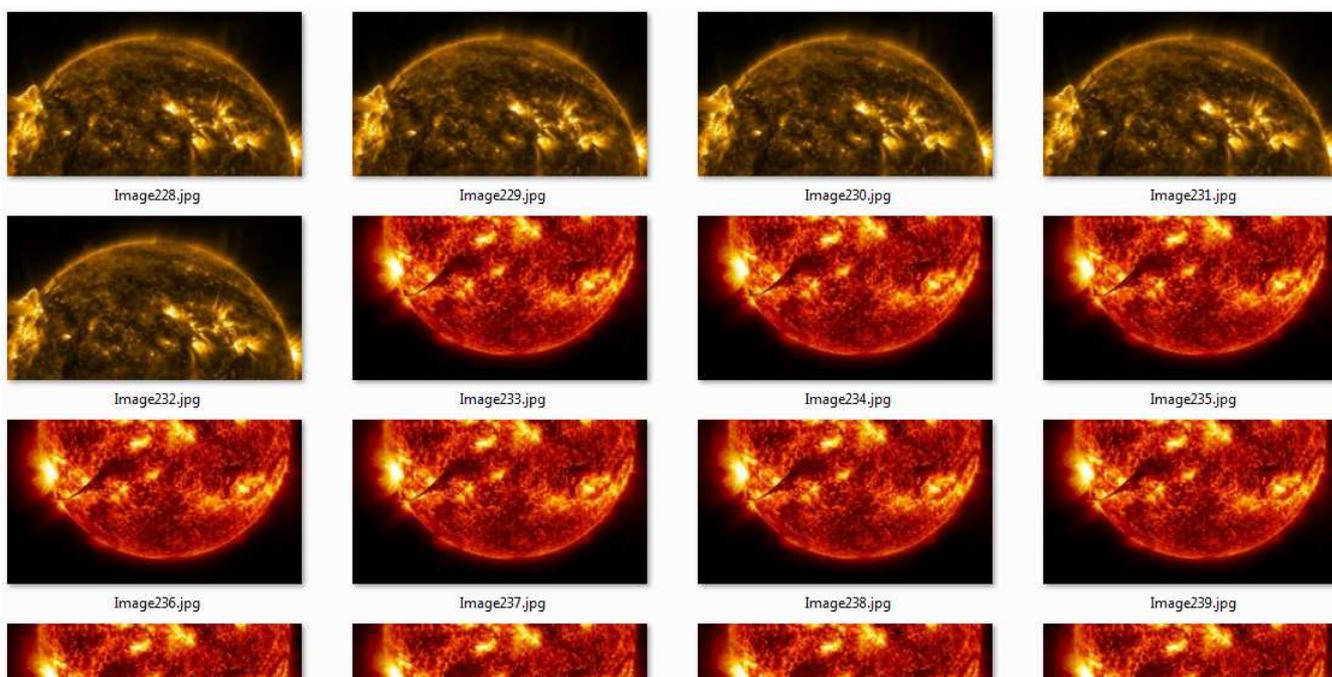
Figure 10: Au cours du processus d'échantillonnage.

Chapitre III: Conception et Mise en œuvre.

Les résultats d'échantillonnage sont affichés dans une liste-view et sauvegardés dans un répertoire créé lors du lancement du processus, sous l'extension (.jpg)

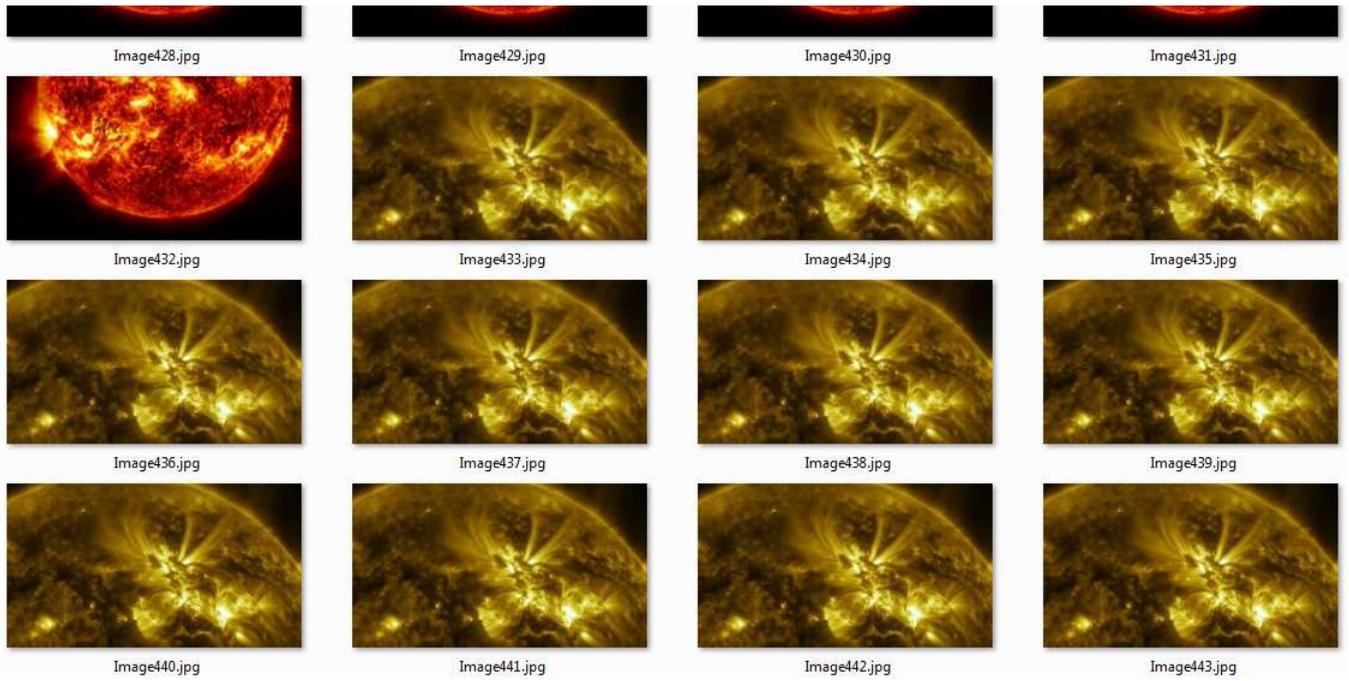


(a)

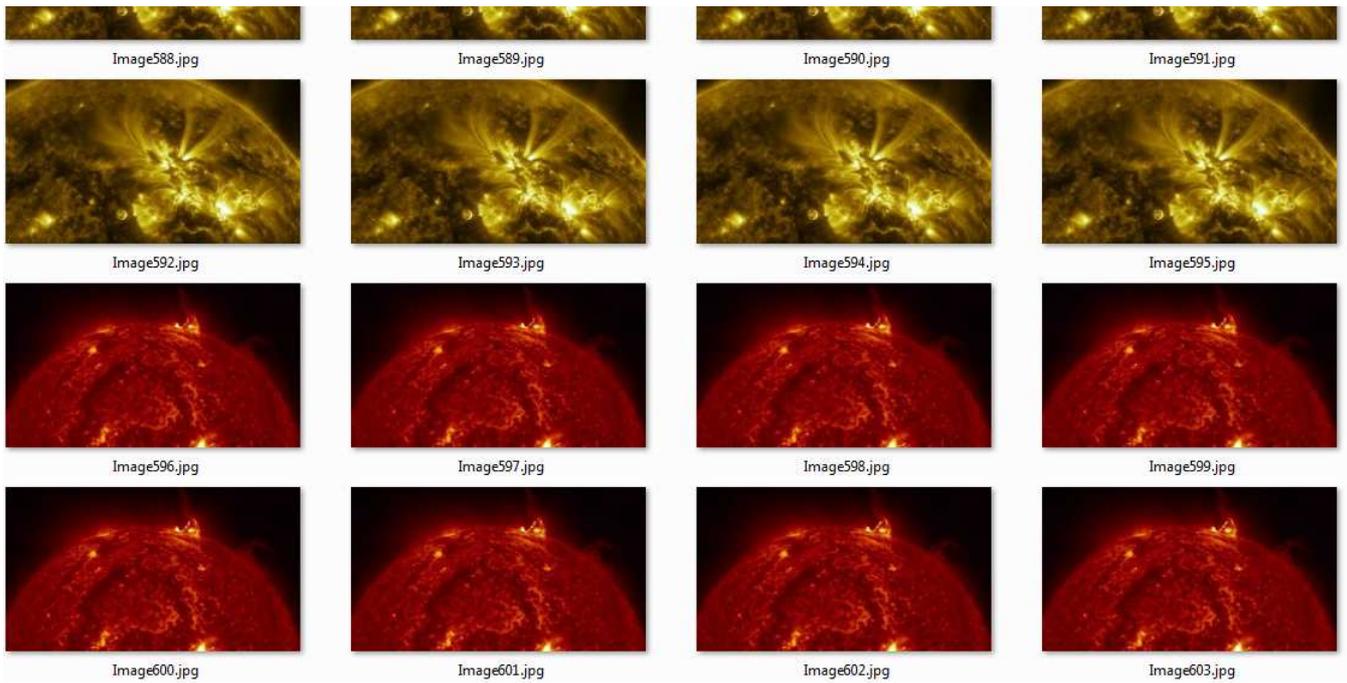


(b)

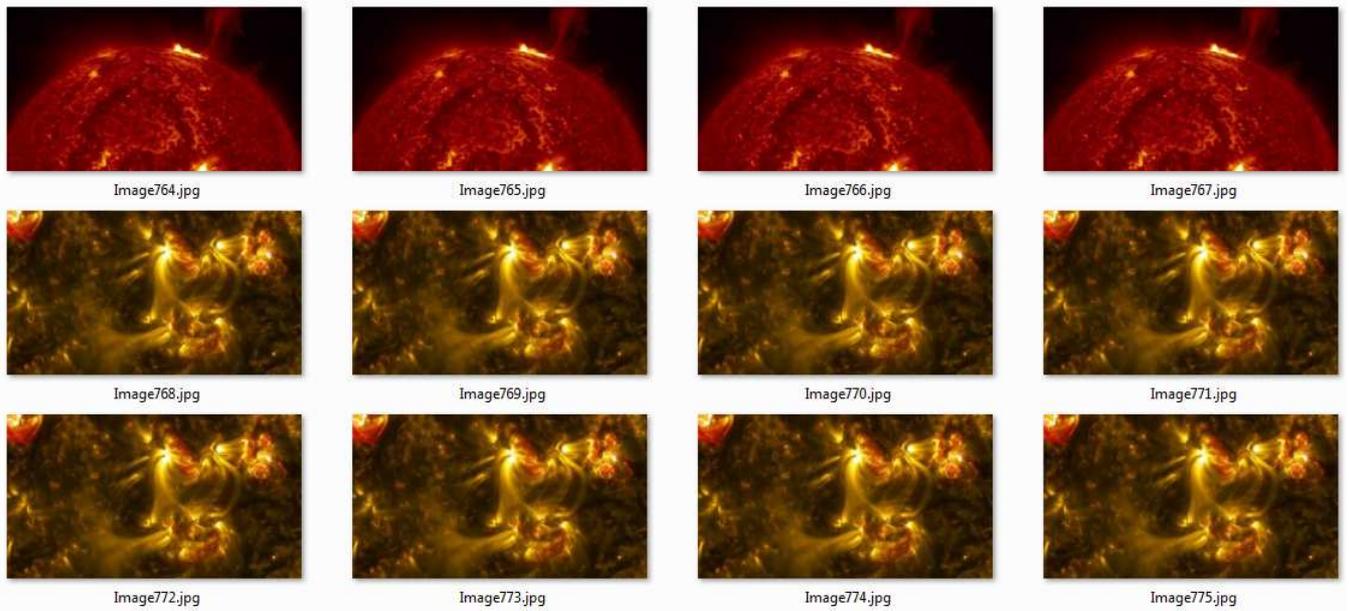
Chapitre III: Conception et Mise en œuvre.



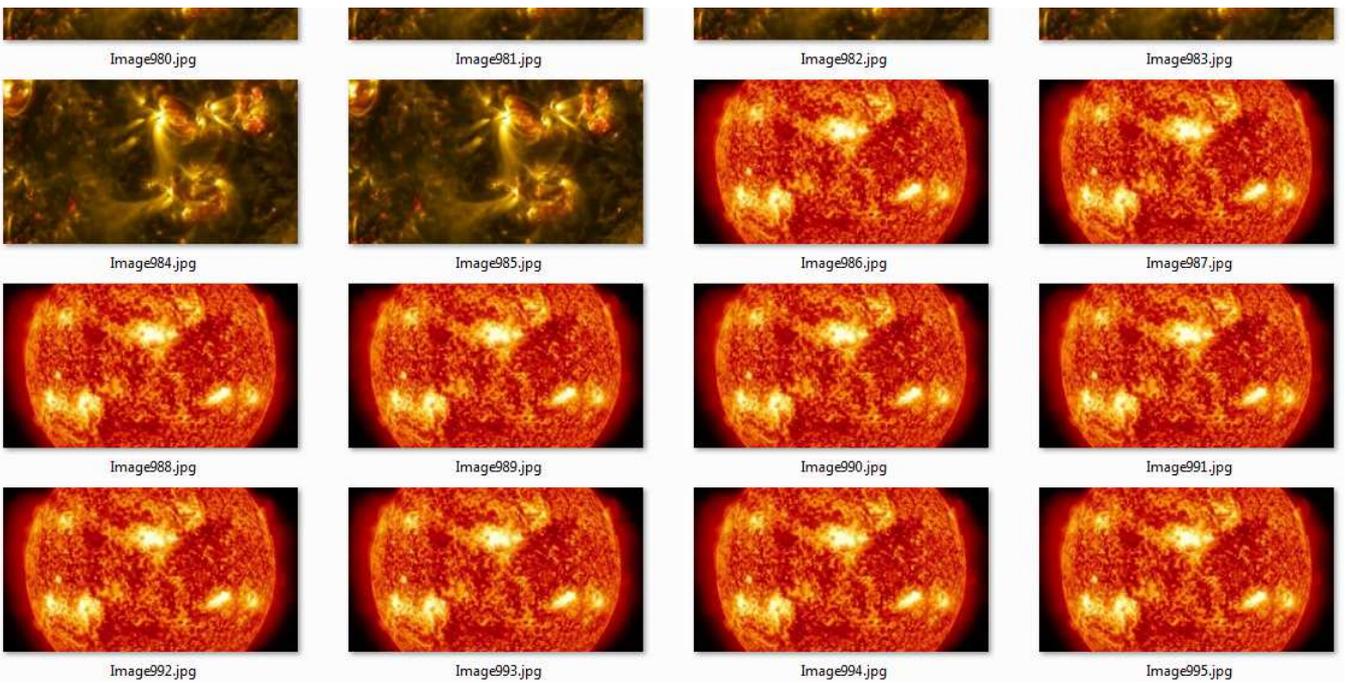
(c)



(d)



(e)



(f)

Figure 11: Echantillons d'une séquence d'image (a), (b), (c), (d), (e) et (f) d'une vidéo du soleil avec une haute définition lancé par la NASA. (Vidéo N°1)

Chapitre III: Conception et Mise en œuvre.

- On a obtenu exactement 1554 Images.

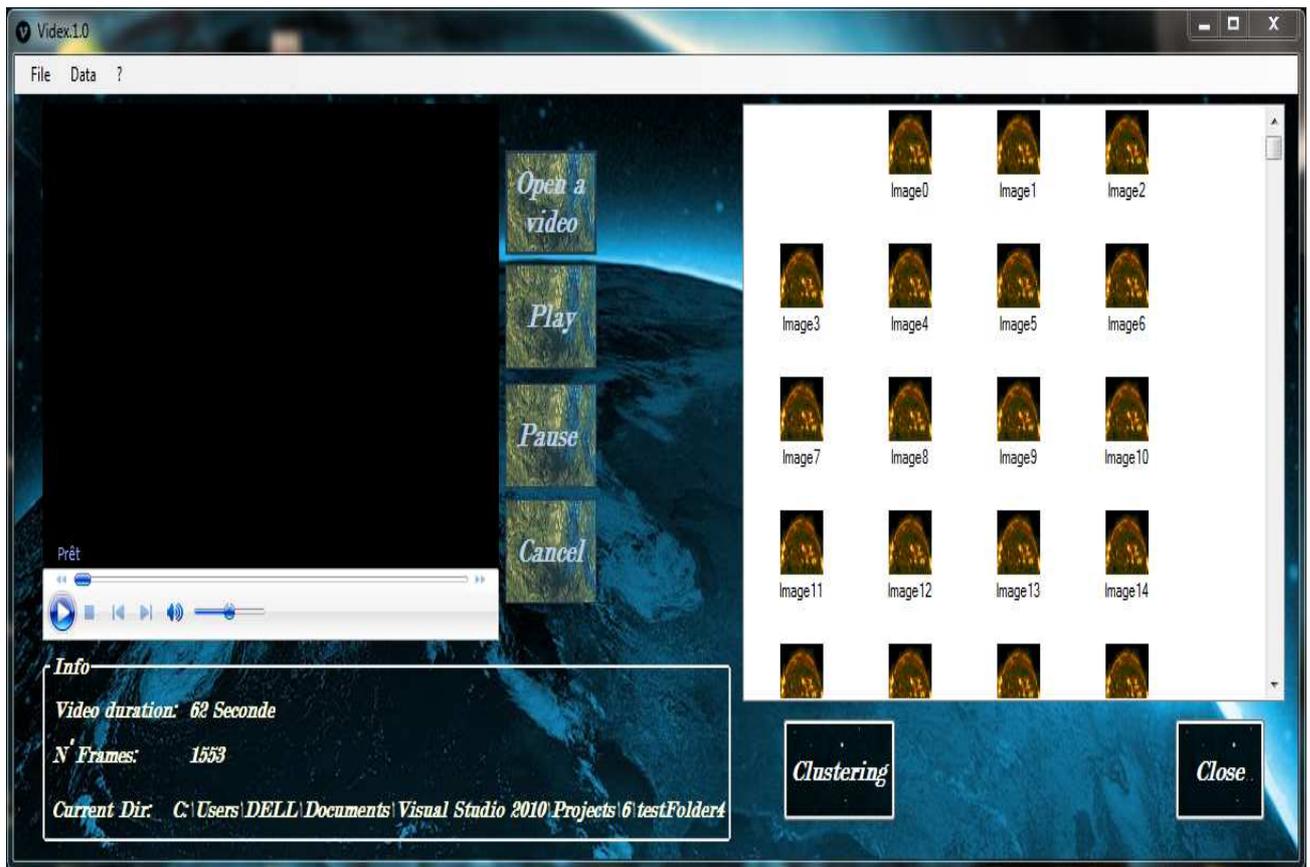


Figure 12: Fin du processus d'échantillonnage.

- La figure suivante -Figure 13- représente le début et la fin du 1er plan extrait par notre algorithme. Puis, On représente un exemple de changement du plan dans la figure 14.

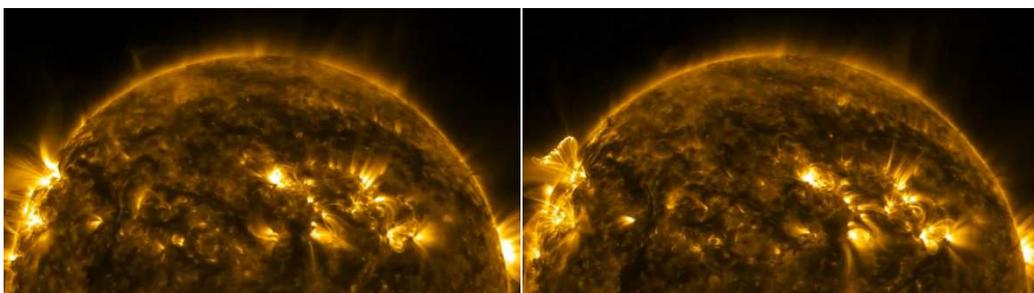


Figure 13: Image0 jusqu'à Image232 représentant le 1er plan.

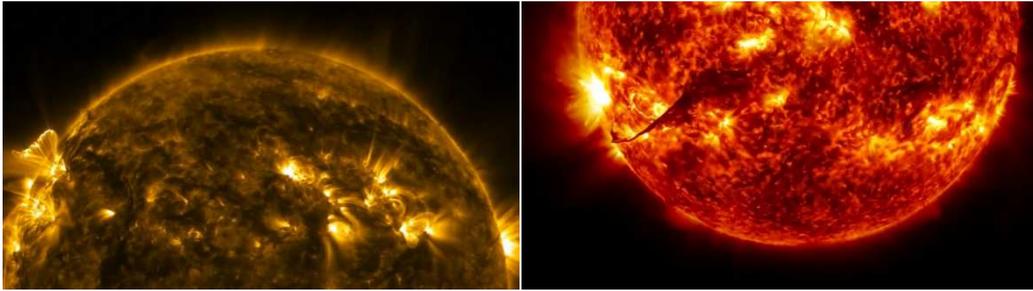


Figure 14: Image232 et Image233 représentant le 1er changement du plan.

- L'étape suivante sert à calculer les descripteurs des points d'intérêt (SIFTs) de chaque image. Puis, sauvegarder les résultats sous format (.txt) comme il est démontré dans la figure suivante:



Figure 15: Représentation de quelques fichiers texte qui contiennent les SIFTs des images extraites dans l'étape d'échantillonnage.

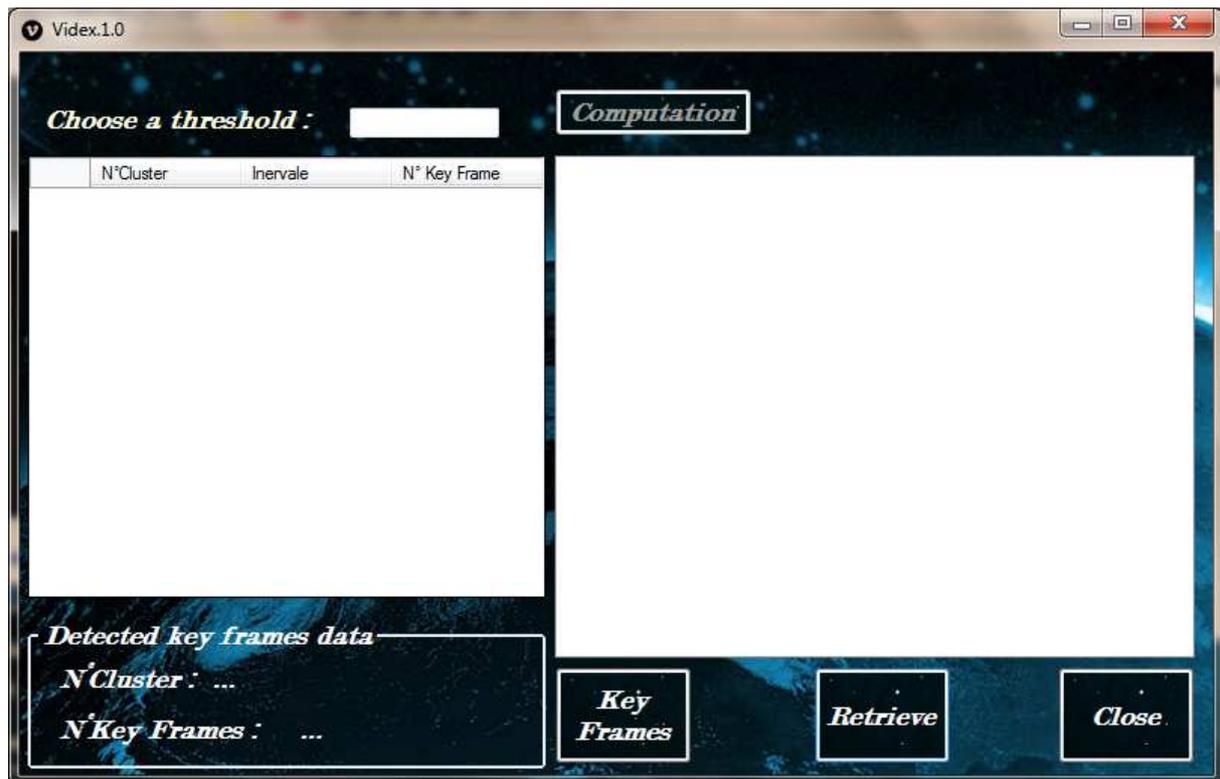


Figure 17: Interface de la classification par cluster et extraction des images-clé.

- ✓ **Computation:** Après la saisie d'un seuil qui appartient à un intervalle $[0,1]$, on appuie sur ce bouton pour obtenir le résultat démontré dans la figure 18.

Chapitre III: Conception et Mise en œuvre.

On a obtenu 9 images-clé qui représentent 9 plans selon un Seuil = 0,5

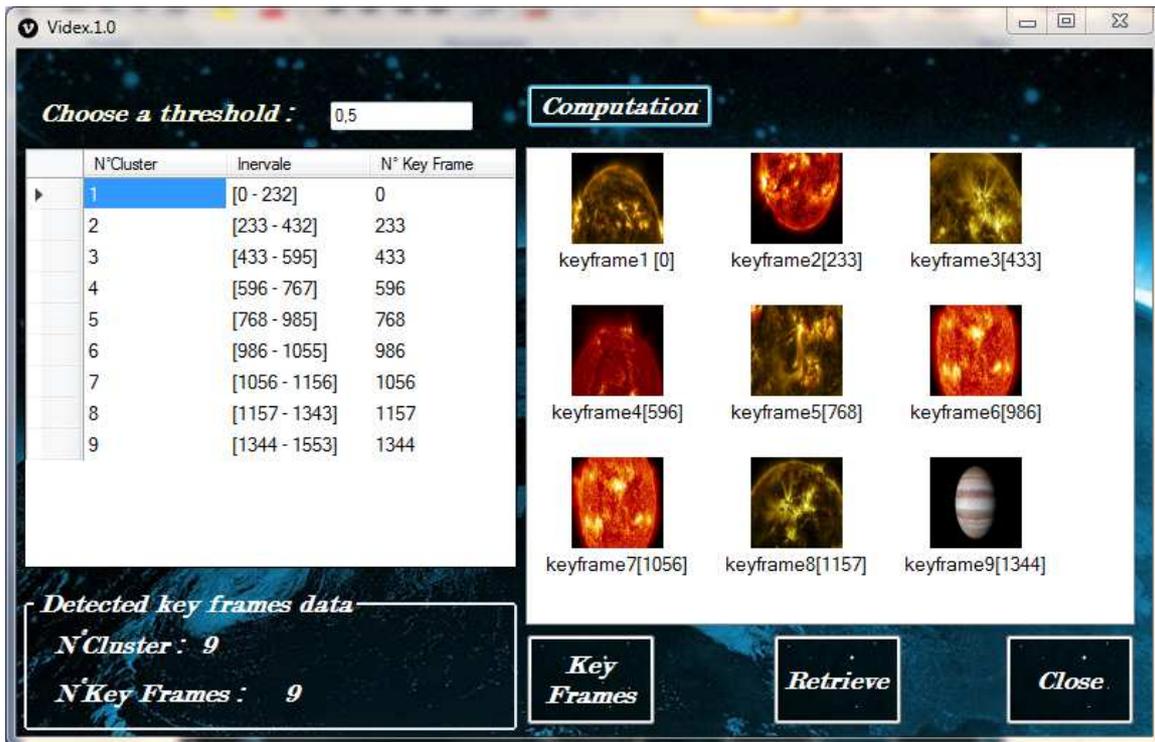


Figure 18: Images-clé extraites de la vidéo N°1.

Les plans, ici, sont considérés comme des clusters.

- ✓ **Key Frames:** Mène à un dossier indépendant pour voir les images-clé clairement.
- ✓ **Retrieve:** Mène à un interface de recherche.

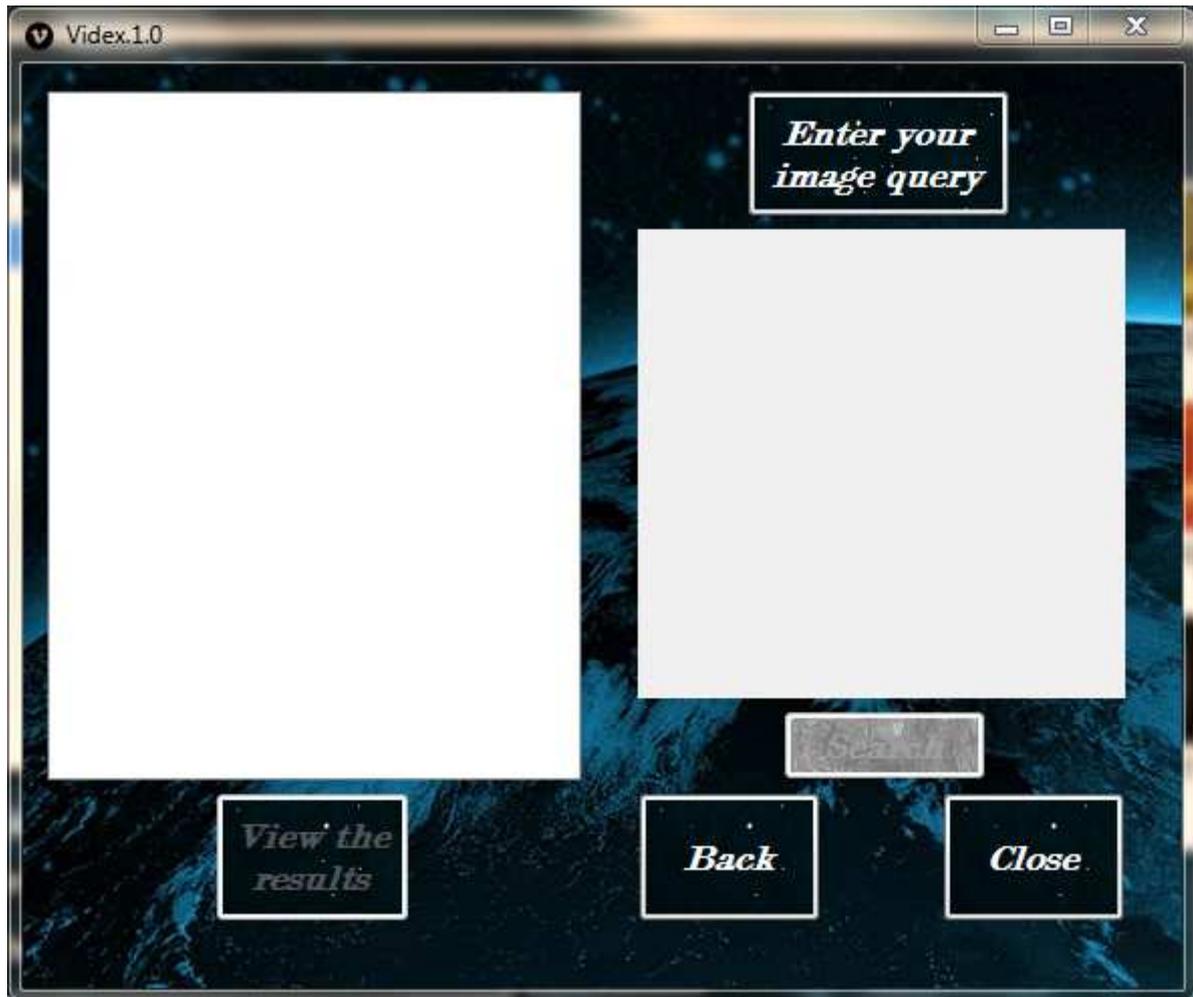


Figure 19: Interface initiale de la recherche.

- ✓ **Search:** Ce bouton permet de trouver des images similaires à une image requête entrée par l'utilisateur..

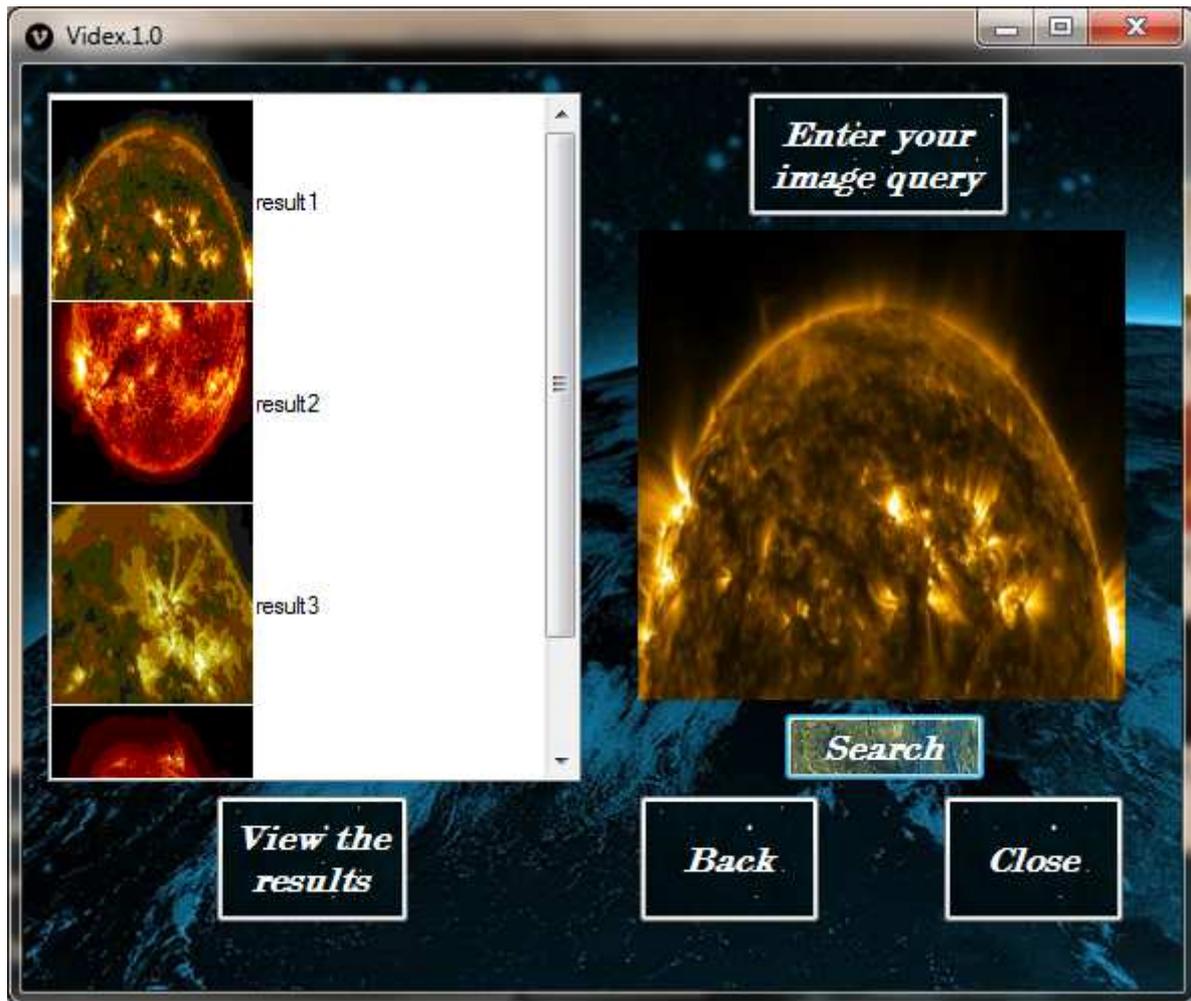


Figure 20: Exemple de recherche par requête image.

- ✓ **Back:** Ce bouton permet de revenir à la fenêtre précédente si on veut effectuer une classification avec un nouveau seuil par exemple.

Notre 2ème vidéo est concerné par un contenu de nature.

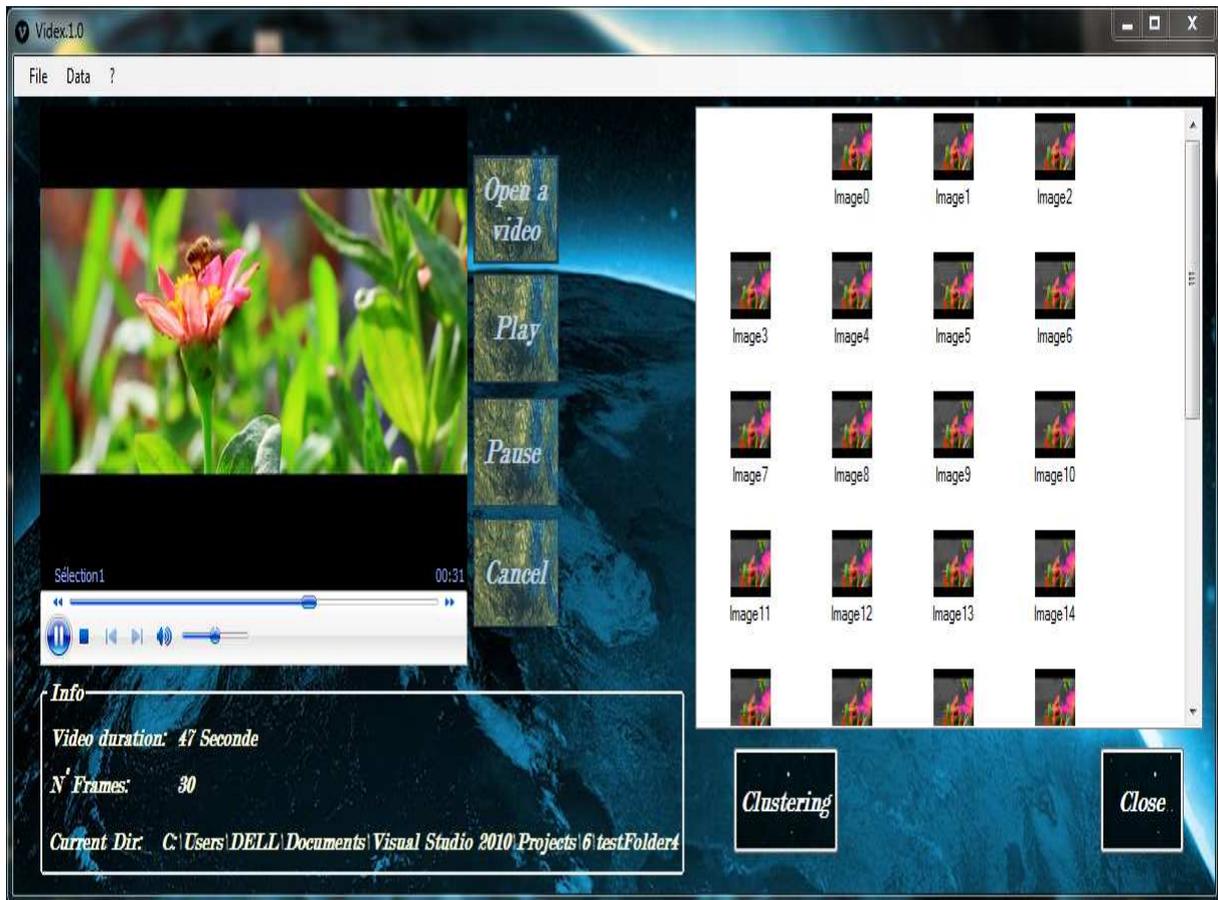


Figure 21: Echantillonnage de la vidéo N°2.

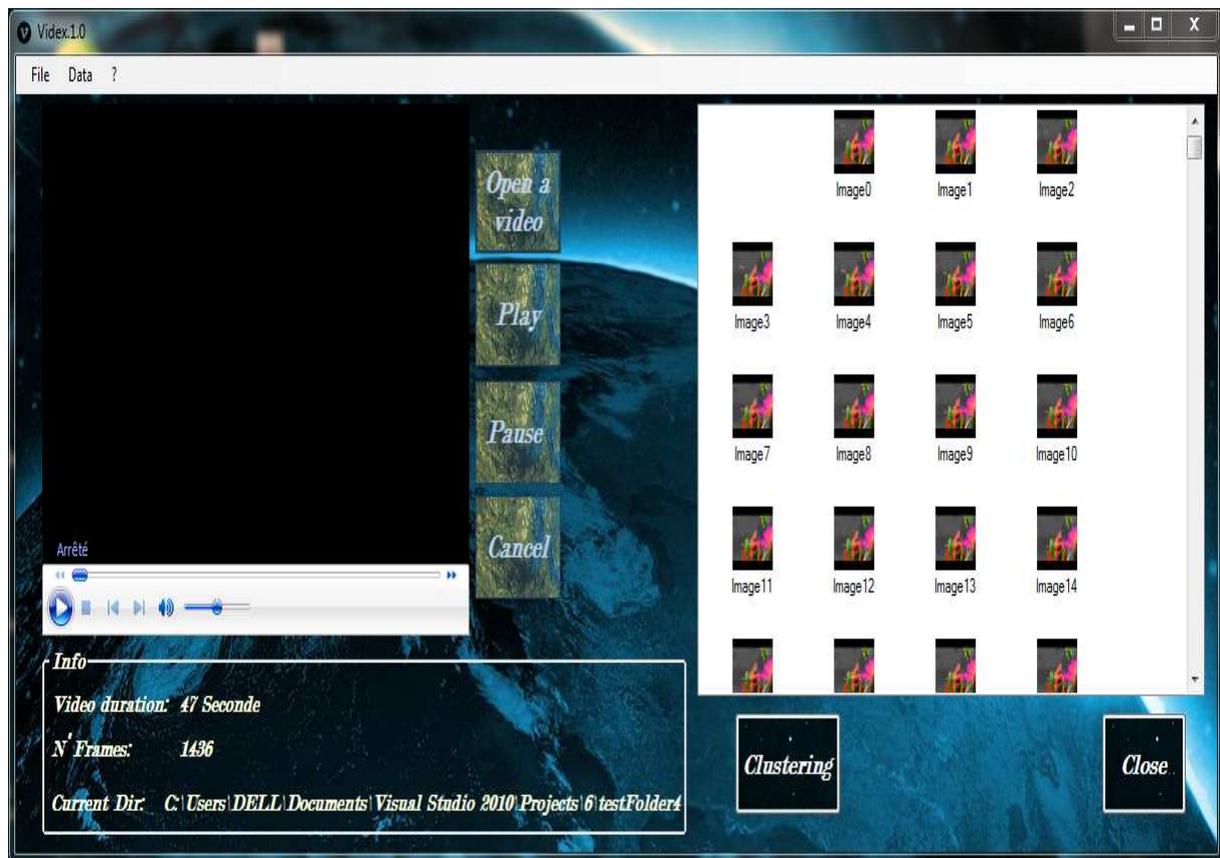


Figure 22: Fin du processus d'échantillonnage de la vidéo N°2.

- On a obtenu dans ce test 1437 images individuelles qui représente la vidéo N°2 entière.

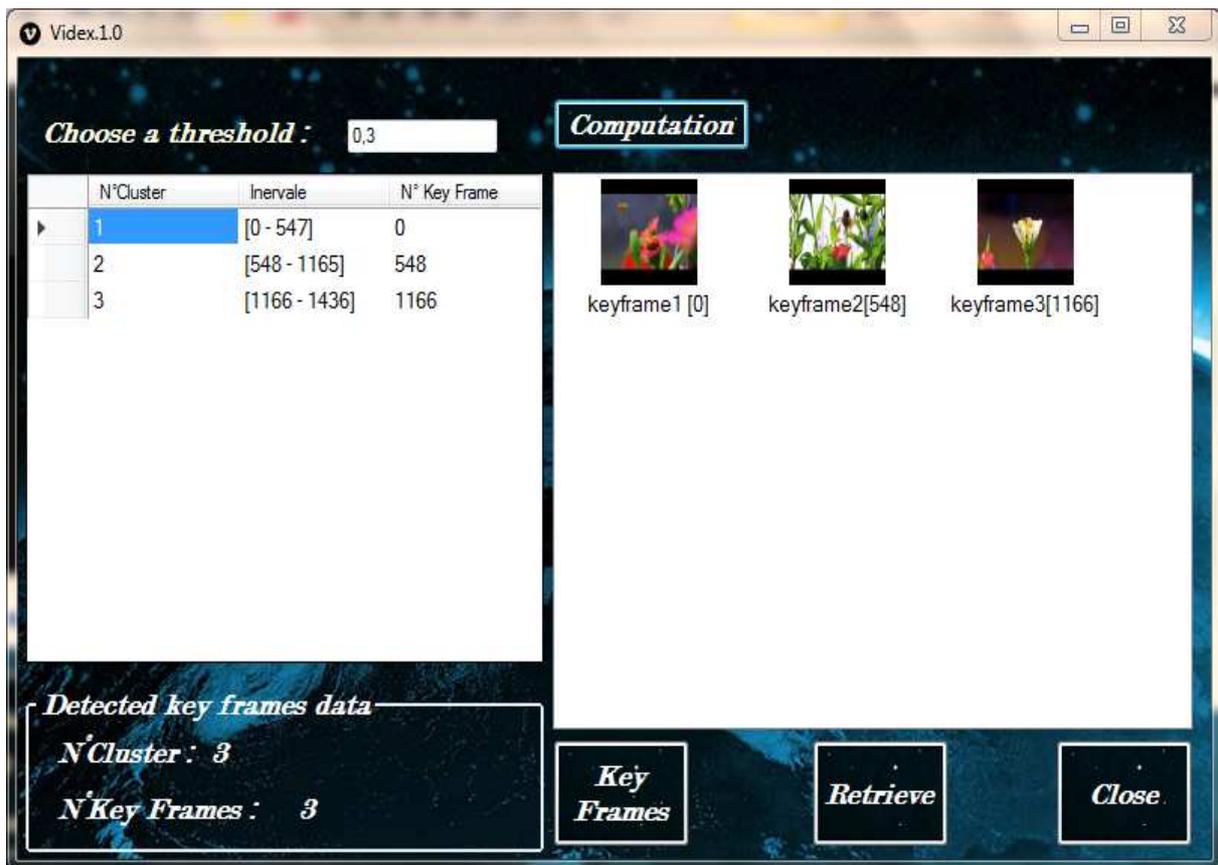


Figure 23: Clusters et images-clé représentant de la vidéo N°2.

- Pour un seuil = 0,3 on a obtenu 3 images-clé qui représente les 3 plans de la vidéos.

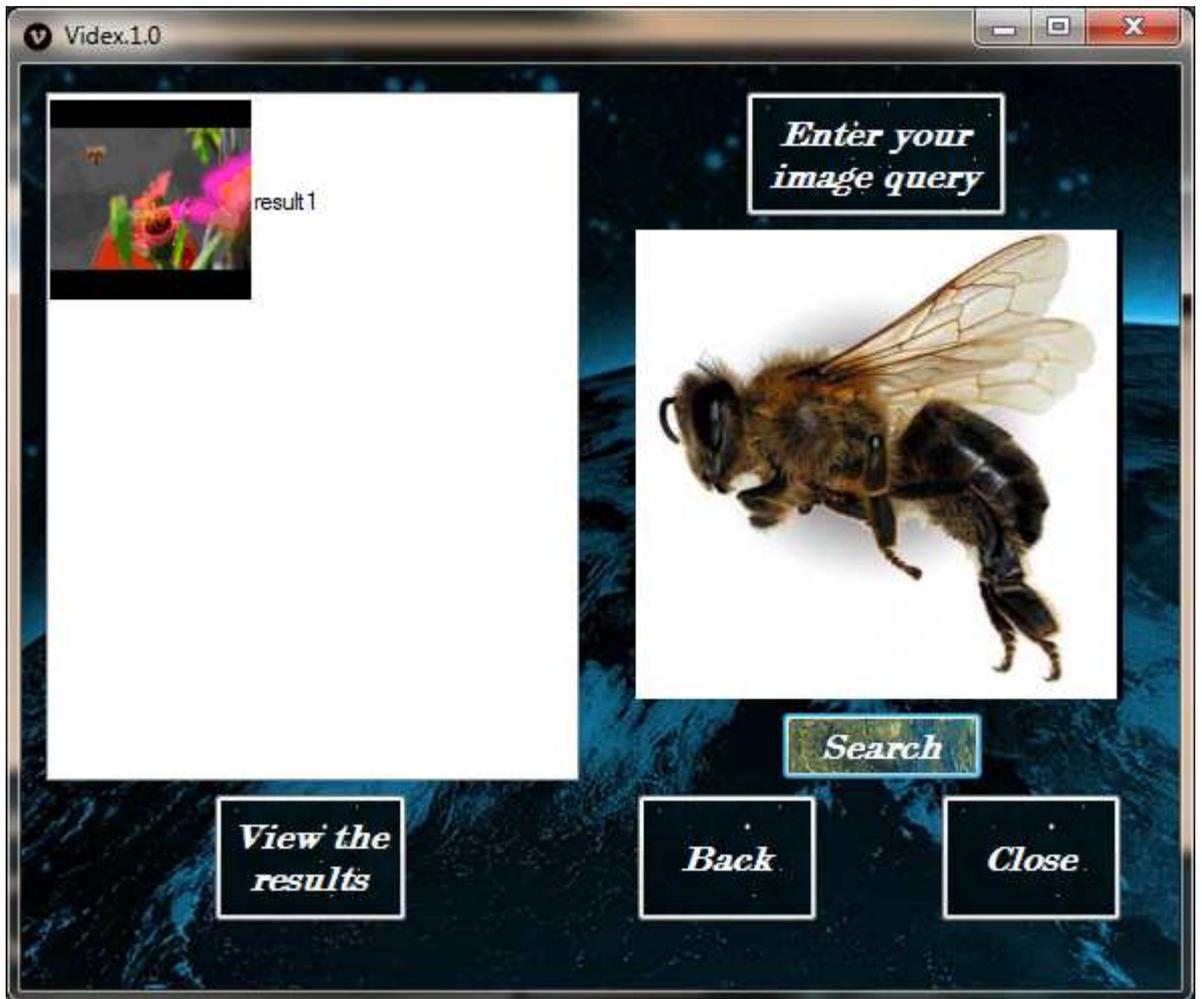


Figure 24: Détection d'un objet contenu dans une image.

Dans la figure 24, on a détecté une "Abeille" dans une image où l'avantage d'invariance à l'échelle montre son efficacité dans notre approche et supporte notre choix de descripteur SIFTs.

5. Conclusion

Dans ce chapitre, on a démontré la faisabilité de notre approche avec les résultats obtenus. C'est une approche basée sur les descripteurs de points d'intérêt SIFTs qui détecte les points d'intérêt dans un ensemble d'images échantillonnées d'une entrée vidéo ce qui permet de créer un système d'indexation visuelle automatique par le contenu en facilitant le processus de recherche de données par le contenu.

Conclusion général

Les images et les vidéos en général, ont gagné l'attention des recherches récentes en raison de la grande demande des clients surtout sur les site web pour la recherche des images et des vidéo comme "Youtube".

L'immense base de données existantes qui contient ces larges quantités d'informations et données a imposé les développeurs de trouver une solution pour faciliter l'interaction avec ces bases de données, traitement, stockage et le processus de la recherche des données, pour servir les demandes des utilisateurs au tour du monde.

On a montré que quelques systèmes proposés ont réussit de résoudre les problèmes d'indexation et de recherche d'images et de vidéo en utilisant le contenu des ces données (caractéristiques visuelles) surtout le facteur du temps, toute en utilisant des techniques de fouille de données qui ont servi d'améliorer les systèmes d'indexation et de recherche d'images et des vidéos.

De notre part, on a proposé notre approche qui a montré des résultats acceptable et il a attendu le but de notre recherche qui est l'indexation des données par le contenu visuel où on a indexé notre base de données par la taille du SIFT.

Il est toujours possible de développer cette approche en ajoutant autres indexes comme l'audio et le mouvement, pour améliorer la performance de notre système d'indexation dont le but est gagner du temps de recherche de vidéo.

Bibliographie

- [1] A. Bahmed. Notes de cours de traitement d'images. Centre Universitaire de RELIZANE - Ahmed ZABANA. 2014-2015.
- [2] S.Marchand Maillet, Y.M. Sharaiha. BINARY DIGITAL IMAGE PROCESSING A Discrete Approach. 2000.
- [3] Kamel Houari. RECHERCHE D'IMAGES PAR LE CONTENU. Thèse De Doctorat. UNIVERSITE MENTOURI CONSTANTINE. 06/2010.
- [4] <http://www.apprendre-en-ligne.net/info/images/index.html> consulté le: 13/05/2017.
- [5] Preeti Chouhan, Mukesh Tiwari. Image Retrieval Using Data Mining and Image Processing Techniques. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering. Vol. 3, Issue 12. 12/2015.
- [6] V. Vijayakumar , R. Nedunchezian. A study on video data mining. Int J Multimed Info Retr. Trends and surveys. 25/08/2012.
- [7] A. Manoury et H. Nicolas. Segmentation temporelle de vidéos numériques fondée sur l'utilisation de mosaïques. IRISA/INRIA, projet TEMICS, Campus de Beaulieu, 35042 Rennes Cedex. 13/12/2013.
- [8] Abdelkader Hamadi. Utilisation du contexte pour l'indexation sémantique des images et vidéos. Informatique [cs]. Université Joseph Fourier Grenoble -1-, 2014. Français. Thèse soumis sur HAL archives-ouvertes.fr, Id: tel-01137526. 30/03/2015.
- [9] P. Mohanaiah, P. Sathyanarayana, L. GuruKumar. Image Texture Feature Extraction Using GLCM Approach. International Journal of Scientific and Research Publications, Volume 3, Issue 5. 05/2013.
- [10] Kranti Burman, Rahul Gedam. Histogram Based Color Image Authentication By Digital Image Watermark Technique. Kranti Burman, Rahul Gedam / International Journal of Engineering Research and Applications, Vol. 3, Issue 4. 07-08/2013.
- [11] Michele Saad. Low-Level Color and Texture Feature Extraction for Content-Based Image Retrieval. Final Project Report. EE 381K: Multi-Dimensional Digital Signal Processing. 09/05/2008.

- [12] V. Shiv Naga Prasad, Justin Domke. Gabor Filter Visualization. University of Maryland. 838 Information Visualization- Spring. 2005.
- [13] N. Kumaran 1, R. Bhavani. MRI Image Retrieval Using Gabor Wavelet Based Texture Features. International Journal of Advanced Research in Science and Technology. Volume 2, Issue1. 2013.
- [14] Dengsheng Zhang, Guojun Lu. Enhanced Generic Fourier Descriptors for Object-based Image Retrieval. Copyright 2002 IEEE. Published in the 2002 International Conference on Acoustics, Speech, and Signal Processing (ICASSP). 2002.
- [15] Paul Heider, Alain Pierre-Pierre, Ruosi Li and Cindy Grimm. Local shape descriptors, a survey and evaluation. Washington University in St. Louis, Queens University. 2011.
- [16] Gianfranco Doretto, Yi Yao. Region Moments: Fast invariant descriptors for detecting small image structures. Visualization and Computer Vision Lab, GE Global Research, Niskayuna. IEEE. 2010.
- [17] Scott Krig. Computer Vision Metrics: Survey, Taxonomy and Analytics. 26/05/2014.
- [18] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision. 05/01/2004.
- [19] Maridalia Guerrero. A Comparative Study of Three Image Matching Algorithms: Sift, Surf, and Fast. 01/01/2011.
- [20] Nagham Hamid, Abid Yahya, R. Badlishah Ahmad, Osamah M. Al-Qershi. A Comparison between Using SIFT and SURF for Characteristic Region Based Image Steganography. International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3. 05/2012.
- [21] Sylvie Jeannin, Ajay Divakaran. MPEG-7 Visual Motion Descriptors. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL.11, NO. 6. 06/2001.
- [22] Kommineni Jenni, Satria Mandala, Mohd Shahrizal Sunar. Content Based Image Retrieval Using Colour Strings Comparison. 2nd International Symposium on Big Data and Cloud Computing. Procedia Computer Science 50. 2015.
- [23] D.Sudha, J.Priyadarshini. Reducing Semantic Gap in Video Retrieval with Fusion: A survey. 2nd International Symposium on Big Data and Cloud Computing. Procedia Computer Science 50. 2015.
- [24] Navneet Dalal, Bill Triggs. Histograms of Oriented Gradients for Human Detection. <http://lear.inrialpes.fr> . Consulté le: 13/05/2017.

- [25] S. Sumathi, S.N. Sivanandam. Introduction to Data Mining and Its Application. Studies in Computational Intelligence, Volume 29. Spring, 2006.
- [26] Preeti Chouhan, Mukesh Tiwari. Feature Extraction Techniques for Image Retrieval Using Data Mining and Image Processing Techniques. International Journal of Advanced Research in Computer and Communication Engineering. Vol. 5, Issue 5. 05/2016.
- [27] Mokeddem Sid Ahmed. Fouille de données. Université de Mostaganem Abdelhamid Ibn Badis. Notes de cours. Master 2 ISI, Semestre 3. 2016-2017.
- [28] <http://igm.univ-mlv.fr/~dr/XPOSE2012/datamining/datamining-demarche-methodes.html> Consulté le: 02/05/2017.
- [29] P. Thamilselvan, Dr. J. G. R. Sathiaselan. A Comparative Study of Data Mining Algorithms for Image Classification. I.J. Education and Management Engineering. Publié en ligne sur MECS (<http://www.mecs-press.net>). 06/2015.
- [30] Dr. L. Rokach, O. Maimon. Data Mining and Knowledge Discovery Handbook. 2010.
- [31] <http://www.space.com/9308-astronomy-overload-scientists-shifting-stargazing-data-mining.html> consulté le: 11/05/2017.
- [32] <http://www.space.com/8515-computer-program-learns-sort-galaxies-human.html> . Consulté le: 11/05/2017.
- [33] S.V.N. Vishwanathan, M. Narasimha Murty. SSVM : A Simple SVM Algorithm. Dept. of Comp. Sci. and Automation, Indian Institute of Science. Conference: Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on, Volume: 3. IEEE. 02/2002.
- [34] Poojia Gupta, Gurpreet kaur. A Concept of A-KNN Clustering in Software Engineering. International Journal of Computer Applications (0975 – 8887). Volume 119 – No.18, June 2015.
- [35] <https://dataaspirant.com/2016/12/23/k-nearest-neighbor-classifier-intro/> consulté le: 10/05/2017.
- [36] Seema Sharma, Jitendra Agrawal, Sanjeev Sharma. Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies. International Journal of Computer Applications. Volume 82 – No 16, 11/2013.
- [37] <http://www.otnira.com/2013/03/25/c4-5/> consulté le: 10/05/2017.
- [38] <https://www.ibm.com/developerworks/library/ba-data-mining-techniques>. Consulté le: 02/05/2017.

- [39] Hari Krishna Kanagala, Dr JayaramkrishnaiahVemula, Prof. Ramchand H Rao Kolasani. Review on Classification Techniques in Data Mining. International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 493- 497. Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh. 2014.
- [40] <http://www.petite-entreprise.net/P-2595-83-G1-principales-taches-du-data-mining.html> Consulté le: 02/05/2017.
- [41] <https://www.mathworks.com/help/stats/supervised-learning-machine-learning-workflow-and-algorithms.html> consulté le: 12/05/2017.
- [42] Jiawei Han, Micheline Kamber. Data mining: Concepts and techniques. 2nd Edition, The Morgan Kaufmann Series in Data Management Systems. 2006.
- [43] <http://www.datavizcatalogue.com/methods/scatterplot.html>, consulté le: 02/05/2017.
- [44] Charu C. Aggarwal. Data mining the Textbook. Springer International Publishing Switzerland. 2015.
- [45] Shyam Boriah, Varun Chandola, Vipin Kumar. Similarity Measures for Categorical Data: A Comparative Evaluation. 2008.
- [46] http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/bhoenes/similarity.html. Consulté le:30/04/2017.
- [47] Abdessalem BEN ABDELALI, Mohamed Nidhal KRIFA, Lamjed TOUIL, Abdellatif MTIBAA, Elbay BOURENNANE. Indexation De La Vidéo Par Le Contenu Et Besoin En Architectures Matérielles Dédiées. 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications. 03/2009.
- [48] <https://msdn.microsoft.com/> Consulté le: 02/05/2017.
- [49] <http://www.emgu.com/> Consulté le: 02/05/2017.
- [50] <https://docs.microsoft.com/> Consulté le: 02/05/2017.
- [51] Yojita Raut, Madhuri Shimpi, Mansvi Patil. Content Based Video Retrieval System. (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (2). 2016.
- [52] Suhas Athani, CH Tejeshwar. Performance analysis of Key Frame Extraction using SIFT and SURF algorithms. (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (4). 2016.

[53] Dipika H Patel. Content based Video Retrieval: A Survey. International Journal of Computer Applications. Volume 109 – No. 13. 01/2015.

[54] Jeffrey Byrne. Nested Motion Descriptors. IEEE Xplore. Computer Vision Foundation. CVPR. 2015.