University of Kasdi Merbah, Ouargla

Faculty of New Information and Communication Technologies

Department of Computer Science and Information Technology

Dissertation Submitted to the Department of Computer Science and Information Technology in Candidacy for the Degree of Master LMD in Fondamental computer.

Presented by:

BENYAZA Yasmina.

Title:

Automatic Image Annotation (AIA)

Publically defended:

On:

Before the jury:

Dr. AIADI Oussama	Supervisor	UKM Ouargla
Mr.	President	UKM Ouargla
Mr.	Examiner	UKM Ouargla

Academic Year: 2017 – 2018

Dedication

I dedicate this modest work

To my beloved and dear parents for them supports and encouragement during all my years of studies and without them I would have never succeeded,
God gives them good health and the paradise after a long life.
To all my family, To all my brothers and sisters,
To baklawati.
Without forget all my friends.
To the special person in my life"Αγάπη μου".

To every person having contributed to this work,

closely or remote.

Acknowledgments

In the Name of Allah, the Most Merciful, the Most Compassionate all praise be to Allah, the Lord of the worlds; and prayers and peace be upon Mohamed His servant and messenger.

First and foremost, praise and thanks to ALLAH for giving me the ability and helping me in accomplishing this thesis.

Thereafter,I would like to acknowledge the assistance of my supervisor Dr. Oussama Aiadi for his continuous support, patience and guidance throughout the course of this research study.Ideeply appreciate his immense knowledge, assistance, constructive feedback, to accomplish this thesis.

My sincere thanks also go to Ramla BENSACI for her encouragements, motivation, advice and insight.

I would also like to extend my special thanks to the honorable members of the jury who accepted to examine and evaluate this study.

I'm particularly thankful to my colleague Seggar ibtissam and Hafiane manel for their contribution which made this study feasible through their cooperation and helpful advices.

Last but not least,I gratefully acknowledge all the professors within the Department of computer science and information technology, especially whom teaching me during my curriculum. I’m also very grateful to all my colleagues.

Abstract

Image annotation an important research topic in the image retrieval field, where it is used to assign a specific words to image that describing their content. However, manual image annotation is no longer effective because it is tired and costly, so automatic image annotation becomes an alternative.

In this thesis, we propose an approach for automatic image annotation based on probabilities. For a given concept, we associate the concept with a set of similar images. Then, we extracted a set of features from training images, thereafter we applied the Gaussian Mixture Model (GMM) to model the concept. We assign each image with 5 labels.

The experimental results showed the effectiveness of the proposed approach as well as its superiority over other methods.

Keywords: Automatic image annotation (AIA), Gaussian Mixture Model (GMM), feature extraction.

Résumé

L'annotation d'image est un sujet de recherche important dans le champ de recherche d'images, dans lequel elle est utilisée pour attribuer un mot spécifique à une image décrivant son contenu. Ce pendant,l'annotation d'image manuelle n'est plus efficace car elle est fatigante et coûteuse, de sorte que l'annotation automatique des images devient une alternative.

Dans cette thèse, nous proposons une approche pour l'annotation automatique d'images basée sur des probabilités. Pour un concept donné, nous associons le concept à un ensemble d'images similaires. Ensuite, nous avons extrait un ensemble de caractéristiques à partir d'images d'entraînement, puis nous avons appliqué le modèle de mélange gaussien (GMM) pour modéliser le concept. Nous assignons chaque image avec 5 étiquettes.

Les résultats expérimentaux ont montré l'efficacité de l'approche proposée ainsi que sa supériorité sur les autres méthodes.

Mots-clés: annotation automatique d'image , modèle de mélange gaussien , extraction de caractéristiques.

ملخص

التعليق التوضيحي للصورة هو موضوع بحث مهم في مجال استرجاع الصور، حيث يتم استخدامه لتعيين كلمات للصورة تصف محتواها. ومع ذلك، لم يعد التعليق التوضيحي اليدوي للصور فعالًا لأنه متعب ومكلف، لذا يصبح التعليق التلقائي للصور بديلاً.

في هذه الأطروحة، نقترح نهجًا للتعليق التلقائي للصور استنادًا إلى الاحتمالات. بالنسبة لمفهوم معين، فإننا نربط هذا المفهوم بمجموعة من الصور المشابهة. بعد ذلك، استخرجنا مجموعة من الميزات من صور التدريب، وبعد ذلك طبقنا نموذج خليط غاوس لنمذجة المفهوم. نقوم بتعيين كل صورة مع 5 تسميات.

أظهرت النتائج التجريبية فعالية المنهج المقترح وكذلك تفوقه على الطرق الأخرى.

الكلمات المفتاحية: التعليق التوضيحي للصورة التلقائية ،نموذج خليط غاوس، استخلاص الميزات.

Table of Contents

List of Figures

List of Tables

General Introduction

Recently, the world of technology has received developed and expanded significantly, as this development has led to the emergence of modern technologies and advanced devices for taking and saving images. Due to their low prices, large storage capacity and ease of use, this allows many to own them. The common use of these devices to document important and commemorative moments results in large collections of images. The large increase in the amount of images raises questions about how to index and retrieve images?
The answer to these questions was to develop appropriate systems for managing these images effectively, allowing for accurate and easy image retrieval.

Among these systems, is CBIR which supports image retrieval based on content characteristics but he faces the problem of the semantic gap betweenlow-level features and high-level user semantics.Several solutions to this problem have been suggested, including a manual annotation that aims to add captions to image that describe their content manually.

Since the manually annotating images is very tedious and costly, as well as being tired and taking a lot of time in the large image dataset, automatic image annotation has become a important research topic in the last years.

Hence, the problematic in this work is how to annotate images automatically?

In this thesis, we propose an optimal approach to solve this issue of automatic image annotation (AIA). The proposed approach uses automated learning techniques to reach the objective of this thesis.

CHAPTER I.

CONTENT-BASED IMAGE RETRIEVAL (CBIR): PRINCIPLES,MAJOR COMPONENTS, APPLICATIONS AND ISSUES.

INTRODUCTION

With technological advancements, data storage and image acquisition technologies have created large databases. In order to manage and deal with these base, it is necessary to develop effective systems to do so. Image search and retrieval is one of the most important services that support these systems. The first research in this domain was Text-Based Image Retrieval (TBIR), which relies on text descriptions, but with the absence of descriptions in some of the modern applications became TBIR unable to perform the task of retrieving images, plus it requires difficult human explanations and take a long time.

After this turning point in the TBIR system, recent research focused on the indexing of images by Content-Based Image Retrieval(CBIR), which aims to support image retrieval based on image features that are extracted using image processor algorithms. One of the advantages of CBIR is the ability to automatically retrieve images instead of the traditional keyword-based approach.

In this chapter, the CBIR system is defined, in addition to its major components. We also show CBIR application and issues, with reference to related works.

1 PRINCIPLE OF CBIR SYSTEM

CBIR is a system that works on searching for similar images according to the content of the query image .This is done by using low-level extracted image features, such as color, texture, shape etc and similarity measures.

CBIR consists of :

- Query interface for acquisition of query image.
- Databases for storing indexing data and measurements.
- System of similarity and retrieval.

To perform the search and retrieval, CBIR follows successive stages according to the pro scheme:

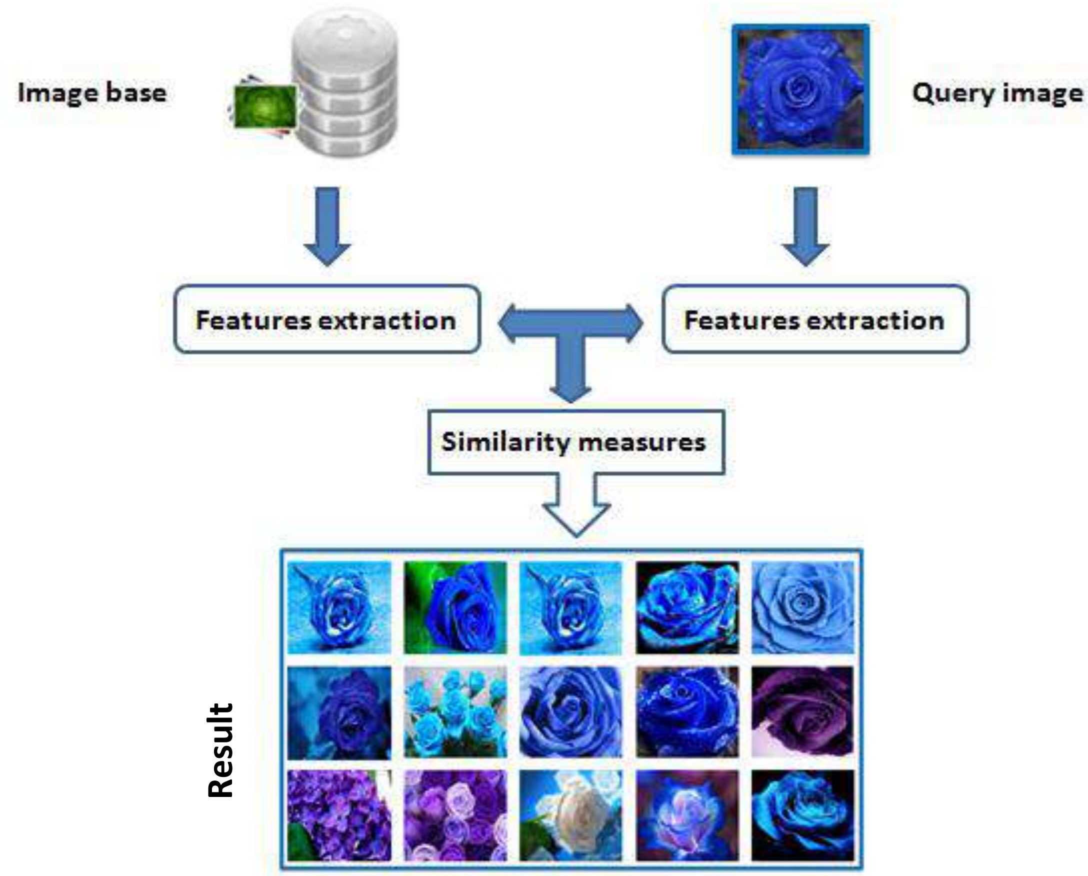

Figure 1.1 CBIR system.

Explain the steps of CBIR system:

- The first step is to make a query image.
- Second, extract the features for both the query image and the image base.
- Third, to measure the similarity between the features.
- Finally extract the results of CBIR search.

2 MAJOR COMPONENTS OF A CBIR SYSTEM

2.1 Features extraction

In order to search images based on their contents, the image in CBIR system represented by low-level features. As the image is a set of pixels, CBIR first step in the semantic understanding is to extract the visual features from these pixels.Features extraction are a digital representation or encoding of an image describing its content, such feature include texture, shape, color ect. As these features help to compare the image query and database images to get the desired search results.

Given the importance of extracting features, many research has been carried out[1]:
Interested in color feature (Swain and Ballard 1991), (Flickner, Sawhney et al. 1995).
Interested in texture feature (Haralick and Shanmugam 1973),(Ojala, Pietikainenet al.2002).
Interested in shape feature (Zhang and Lu 2002),(Tieng and Boles 1997).
And many others.

	Feature1	feature2	feature3	feature4
Image1	66.7687377...	24.3033447...	73.0283813...	3.59037937...
Image2	63.1855468...	4.86319469...	24.3033447...	3.79956662...
Image3	85.4724731...	37.6781005...	45.1207275...	1.72917442...
Image4	89.9975585...	24.3033447...	33.2885742...	25.2975463...
Image5	75.9944458...	5.24542098...	25.2975463...	63.1855468...
Image6	73.0283813...	36.8306884...	45.1207275...	1.00019104...
Image7	24.4333496...	5.23242622...	70.4543457...	85.4724731...
Image8	34.3601684...	4.07676228...	66.7687377...	34.3601684...
Image9	2.67061784...	24.6279907...	73.0283813...	5.46539952...

Features extraction

Figure 1.2 Features Extraction.

Since the extraction of features is an essential element of this work, the features used are mentioned in Chapter 2.

2.2 Similarity measures:

Measurement of similarity between images is one of the key issues in CBIR. After extracting database features and query image, the similarities between these features are measured in order to identify the images related to the query image. There are many similarity measurements that perform this process, below, we present some of these measures[2]:

Table 1.1 Similarity measures.

Measure	Formula
Minkowski distances	$D_p(I_1, I_2) = \left(\sum_{i=1}^{M} \lvert I_1(i) - I_2(i) \rvert\right)^{\frac{1}{p}}$
Histogram intersection	$D_{Hist}(H_1, H_2) = 1 - \frac{\sum_{i=1}^{M} min(H_1(i),\ H_2(i))}{\sum_{i=1}^{M} H_2(i)}$
Quadratic distance	$D_Q(I_1, I_2) = \sqrt{(I_1 - I_2)^T A (I_1 - I_2)}$ $A_{i,j} = 1 - \frac{D_2(i,j)}{d_{max}}$
EarthMovers Distance (EMD)	$D_{EMD}(I_1, I_2) = \frac{\sum_{i,j} d_{i,j} g_{i,j}}{\sum_{i,j} g_{i,j}}$
Cosine Distance	$d_{\cos}(\mathbf{Q}, \mathbf{T}) = 1 - \cos\theta = 1 - \frac{\mathbf{Q}^t \mathbf{T}}{\lvert \mathbf{Q} \rvert \cdot \lvert \mathbf{T} \rvert}$
χ 2 Statistics	$d_{\chi^2}(\mathbf{Q}, \mathbf{T}) = \sum_{i=0}^{N-1} \frac{(Q_i - m_i)^2}{m_i}$
Mahalanobis Distance	$d_{mah} = [(\mathbf{X_Q} - \mathbf{X_T})\ \Sigma^{-1}\ (\mathbf{X_Q} - \mathbf{X_T})]^{\frac{1}{2}}$

3 APPLICATION FIELDS OF CBIR:

CBIR is the most important system that helps manage a wide range of applications effectively[3], some of them are listed below:

➢ Medical Applications:

Since hospitals are among the institutions that produce a large number of images daily, as these images are very important in the diagnosis of diseases, medical research and education e.g. CT, MRI, Ultrasound, The Visible Human . To process these images requires high quality effective techniques and CBIR is one such technique

➢ Biodiversity Information Systems:

Image retrieval techniques can help biologists in many search for many data for studies on biodiversity, as an example take image of a plant and ask the system search for images of plants that have the same paper form.

➢ Digital Libraries:

There are many digital libraries that support CBIR services. For example, a digital library that supports geographic image retrieval. The system manages air images that can be retrieved through fabric descriptors. The place names associated with retrieved images can be viewed by satellite signal with a Geographical Name Information System.

The CBIR technology has been used in many other applications such as Scientific Databases(e.g. Earth Sciences), Art Collections(e.g. Fine Arts Museum of San Francisco), fingerprint identification crime prevention, historical research, Photograph archives,Architectural, engineering design and others.

4 CBIR ISSUE:

CBIR system offers several services and advantages to the world of technology, which helps in accomplishing many researches and facilitates the search process. It also contributes to the building of new and effective technologies.But despite his potential, he suffers from several problems that hamper his work and reduce its effectiveness. One of these obstacles is the semantic gap.Below we will address the problem

➢ Semantic gap:

As his name implies, CBIR is interested only in the content of the image (visual features) and does not pay attention to the semantic concepts associated with images, which leads to the reduction of the performance of CBIR

The results obtained from the image retrieval process have the same visual features of the query image, however, these results may include images unrelated to the content of the query image.

For CBIR, these results are consistent, contrary to the human perspective, which confirms the lack of a relationship between them .This contradiction is called the semantic gap.

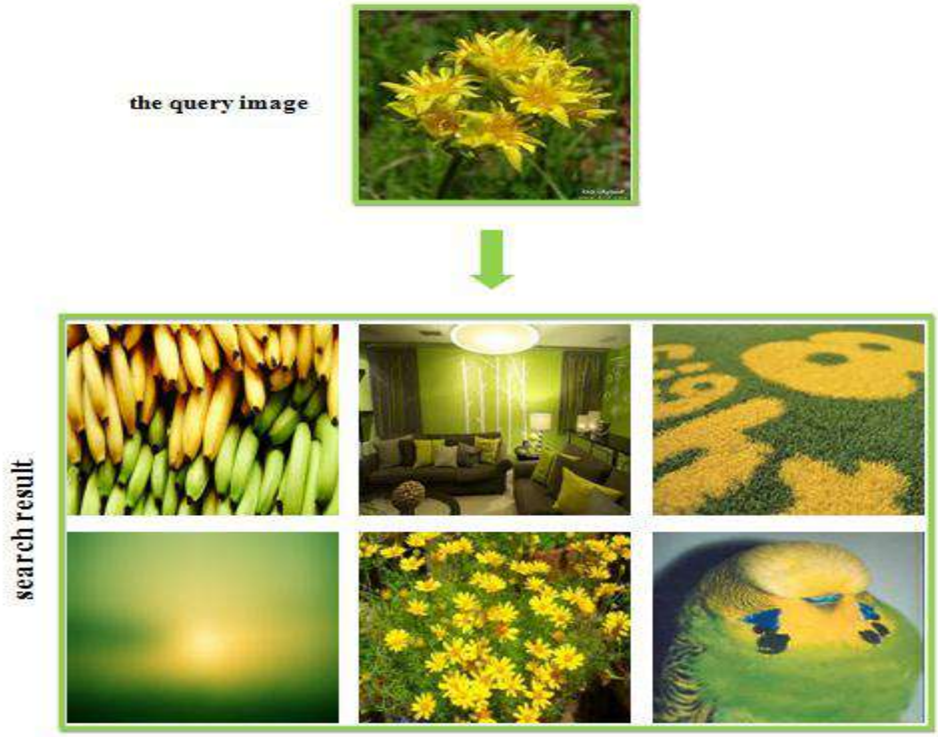

Figure 1.3 Images representing the semantic gap concept.

In this example, the concept of the semantic gap is clarified. Also note that the search result for the query image was not satisfactory,this indicates that there is no correspondence between the information that machines can draw from images and interpretations understood by humans.

Figure 1.4 Semantic gap between low-level image features and high-level semantics.

In order to eliminate and treating the problem of the semantic gap, and to get good performance for the CBIR, We suggest automatic image annotation as solution to this problem.

5 AUTOMATICIMAGE ANNOTATION(AIA):

With the development of multimedia and network technology, image data has been becoming more common rapidly.This made traditional manual annotation a noncompliant technique, and Automatic image annotation became an alternative.

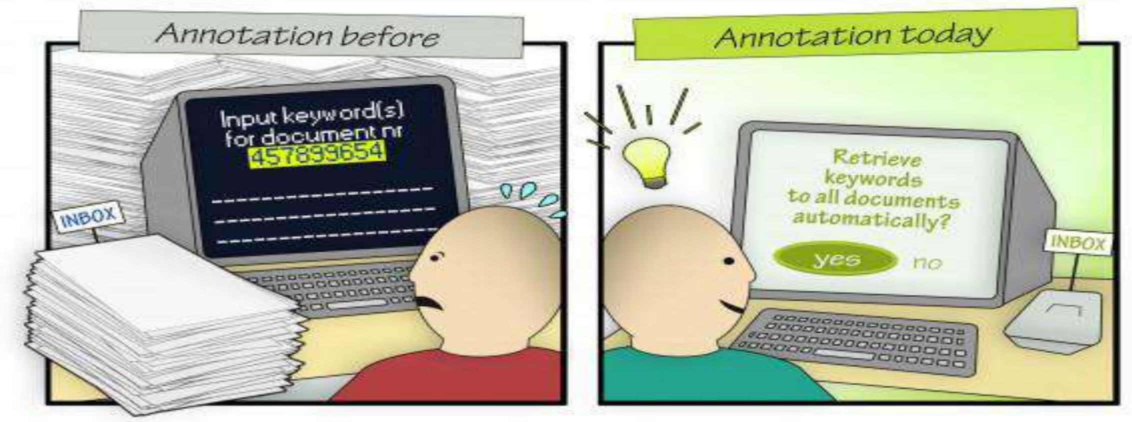

Figure 1.5 Manual annotation and automatic annotation.

- **Definition of AIA**

Automatic image annotation (AIA) is a technique works on assign images with labels i.e., annotations, that describe the image content .According to recent publications, they are an active topic of research[4, 5, 7]. Where it could help in the content-based image retrieval (CBIR) process in a large scale image database more rapidly and precisely.

Figure 1.6 Automatic image annotation.

AIA is an effective method to resolve the problem of “Semantic Gap”, where it erects bridge between high level semantic and low level features.

5.1 Automatic image annotation in machine learning :

Automatic image annotation methods in solving some problems such as the semantic gap problem is based on the use of learning techniques that seek to make the machine capable of learning automatically. Machine learning algorithms are classified into three class: supervised, semi-supervised and unsupervised[1].

5.1.1 Automatic image annotation based on supervised learning:

Supervised learning uses labeled training data to forecasting response gauge(i.e., result) for new unlabeled data (Jain, Murty et al. 1999).

The main goal of supervised learning is to use available techniques to improve a particular function. Ideal drafting for the task of supervised learning,are classification algorithms including parametric and non-parametric ones.

5.1.2 Automatic image annotation based on unsupervised learning:

Unsupervised learning is used to select styles in unlabeled data i.e., select the input data structure with shortage of response gauge (i.e., result)(Arbelaitz,Gurrutxaga et al. 2013).

Among the unsupervised models are Relevance models, where it has the ability to predict labelsof images that are unlabeled(Zhou, Cheung et al. 2011).

Unsupervised learning aims to findstylesand relationships between data. Clustering and dimension reduction are ideal examples of unsupervised learning.

5.1.3 Automatic image annotation based on semi-supervised learning:

Semi-supervised take place between supervised and unsupervised learning

In this kind of learning,each of labeled and unlabeled data are used, where the quantity of labeled data is comparatively much less than the labeled data.

Unlike supervised and unsupervised learning in Semi-supervised there are find a few methods.

5.2 Automatic image annotation techniques:

- **Single labeling annotation using binary classification:**
  - Image annotation using support vector machines.
  - Image annotation using artificial neural network.
  - Image annotation using decision tree.

- **Multi-labeling annotation using Bayesian methods:**
  - Non-parametric approach.
  - Parametric approach.
- **Image annotation incorporating metadata.**

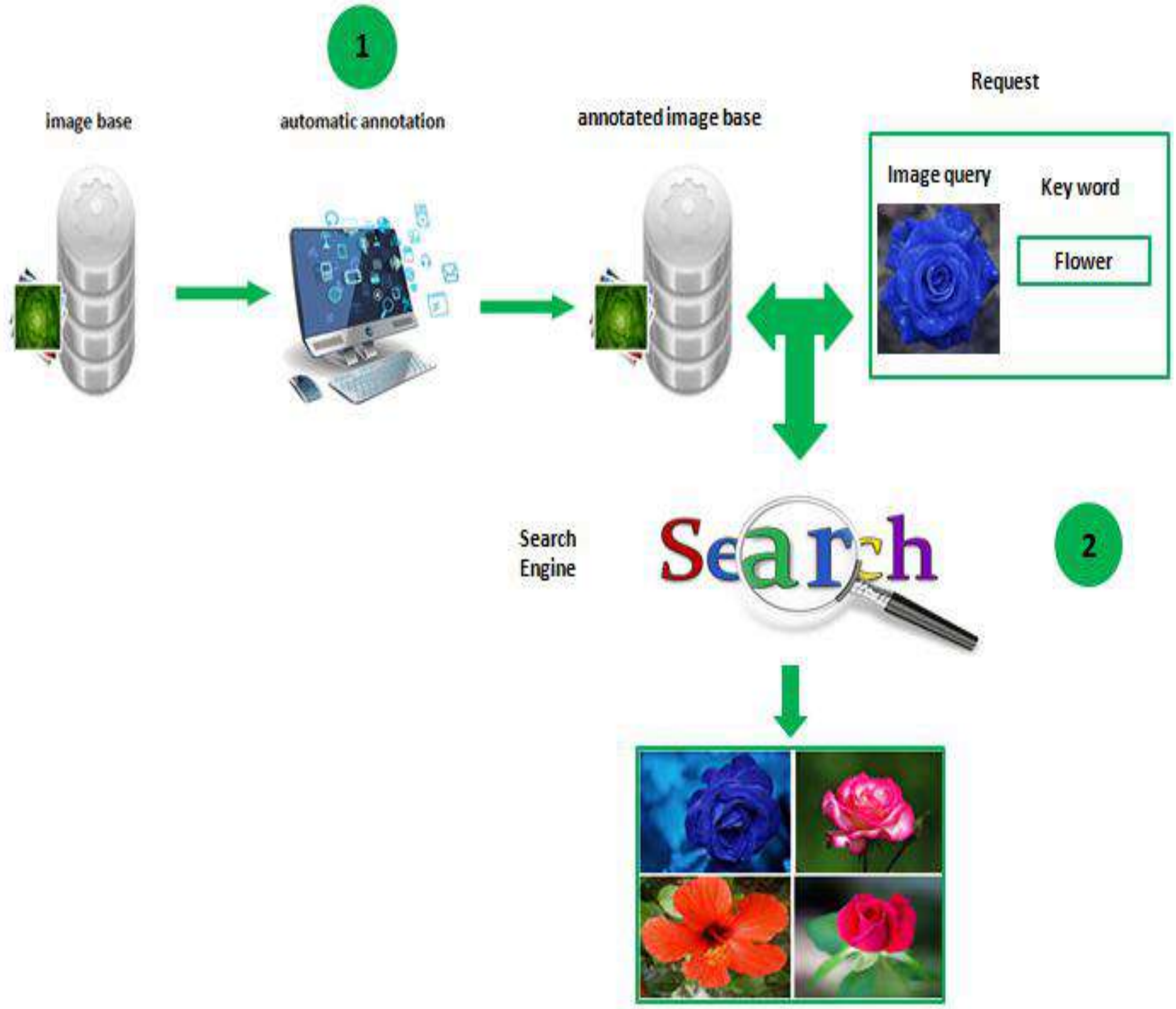

Figure 1.7 Example of image retrieval using AIA and CBIR.

6 RELATED WORK:

Due to the importance of the automatic image annotation, researchers in this field have proposed many works around it[1]. We mentioned some of them in the following table:

Table 1.2 Related work of automatic image annotation.

Work	Date	Method	Dataset
Makadia, Pavlovic et al.	2008	A baseline method	. Corel5K . IAPR TC-12 . ESP Game
Zhao, Lu et al.	2013	semantic relevance	. Corel5K
Liu, Xu et al.	2011	Textual query of personal photos facilitated by large-scale web data	. Kodak . Flickr
Wang, Zhou et al.	2008	Improving automatic image annotation with Google semantic link	.Yahoo! image search querylog . Flickr
Ballan, Uricchio et al.	2014		. Corel5K . IAPR-TC12 . MIRFlickr-25K
Murthy, Maji et al.	2015		. Corel-5k . ESP-Game . IAPRC-12
Tsoumakas and Zhang	2009	Multi-label learning for image categorization	. Scene . Mediamill
Sun, Tang et al.	2014		. Flickr Natural Scene Image Dataset . MediaMill Video Dataset
Duygulu, Barnard et al.	2002	Machine Translation Model (TM)	. Corel
Jeon, Lavrenko et al.	2003	Cross Media Relevance Model (CMRM)	. Corel
Lavrenko, Manmatha et al.	2003	Continuous Relevance Model (CRM)	. Corel
Gong, Li et al.	2010	Methodsincorporatingsemanticrelatedness	. Corel 5K
Liu, Wang et al.	2007		.Corel Dataset . Web Dataset
Wu, Jin et al.	2013	Annotation refinement and completion for image retrieval method	.Corel dataset .Labelme photo collection .Flickr photo collection .TinyImg image collection

CONCLUSION

CBIR is one of the most important systems adopted in the field of research and retrieval of images because of its benefits in technological progress.

We have demonstrate in this chapter the importance of CBIR,where we discussed the basic principle of this system, with its major components and steps of the process from extraction features to similarity measures. Including the problem of the semantic gap facing this system, and our proposal as a solution was **automatic image annotation,** which is our topic for research in this work.

CHAPTERII.

METHDOLOGY OF AUTOMATIC IMAGE ANNOTATION

Introduction

Since the automatic image annotation is a hot research topic, a several ways to solve this problem have been suggested as mentioned in the previous chapter. In this chapter we present the proposed approach by us for automatic image annotation, where we provide an overview of the approach as well as the important techniques used.

Overview on the proposed approach

In this thesis, we examine the task of image annotations as a supervised learning problem. Where the annotation system is trained on a set of images that are categorized according to the concepts they contain,with the aim of defining each concept with its common visual features. After training, the task of the system is to predict a set of keywords to annotation on the set of test images.

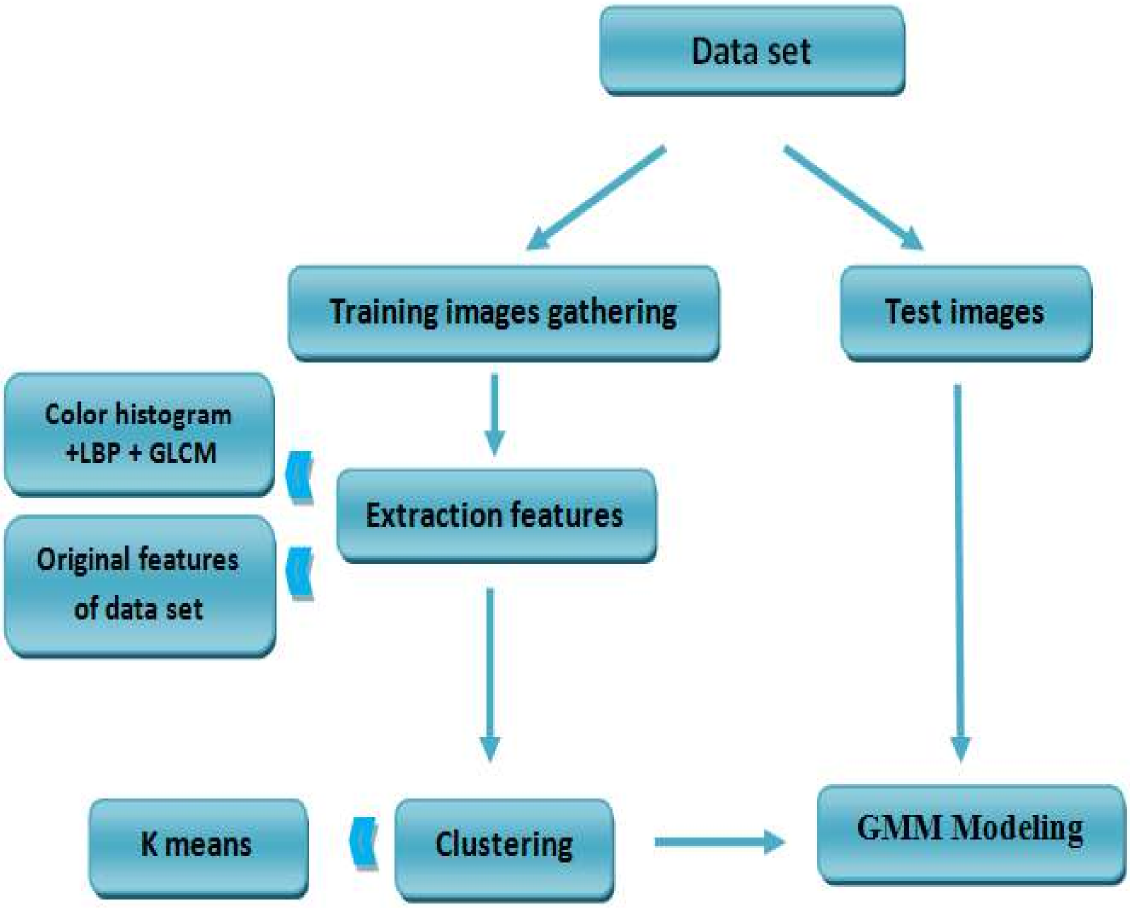

Figure 2.1. Methodology chart.

1 Training images gathering :

In order to conduct training,we assign to each concept a set of images that contain this concept.

Figure 2.2 Gathered images for certain semantic concepts

2 Extraction features:

In the previous chapter, we discussed the extraction features and their importance. As a key element in creating the automatic image annotation application and as the first step we take after dividing the database for training and testing, the extracted features are described below.

2.1 Gray Level Co-occurrence Matrix(GLCM) :

The co-occurrences matrix of an image is a square matrix of coefficients C (i, j) and of size nxn where n is the number of grey levels of the image. The elements of the GLCM represent the probability of occurrences of the couple of grey levels (i, j) distant of d and directed by an angleθ.

To convert the probability of occurrences in relative frequencies, it is necessary to normalize the elements of the matrix by dividing them by the total number of elementary points pairs of the image separated by the distance d in the considered direction.

Many indications can be determined from the matrix C (i, j) of co-occurrences GLCM (Haralick, 1979). Only four indications were determined because they describe the heterogeneous character and the correlation between the textures of the used images[8].

The analysis of the image textures was realized further to the determination of signatures textural vectors in four indications extracted from the matrix GLCM. Four indications are: the homogeneity the values of which vary meanwhile [0 1], the contrast which is nil for the images at grey levels constant, the correlation which varies meanwhile [1 1] it is of infinite value for the images with value of grey levels constant, the energy which varies meanwhile [0 1] it is 1 for an image with values of constant grey levels. Two parameters bound to the method GLCM were adjusted or the distance of connectivity of pixels $d \in \{1,2\}$ and the angle of neighborhood pixel orientation$\theta \in \{0, \pi/4, \pi/2, 3\pi/4\}$.

The gray level co-occurrence matrixes (GLCM) were determined for four textures of the image:

Homogeneity:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} C(i, j / d, \theta)$$

Contrast:

$$i.j.(i-j)^2 . \sum_{i=1}^{n}\sum_{j=1}^{n} C(i, j)$$

Correlation:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} (i - m_x).(j - m_y) C(i, j)$$

Where

$$m_x = \frac{1}{n}\sum_a\sum_b a.C(a,b) \qquad m_y = \frac{1}{n}\sum_a\sum_b b.C(a,b)$$

Energy:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} C(i, j)^2$$

2.2 Color Histogram:

A color space is a model for representing color in terms of intensity values[9].There are many different color spaces such as RGB, HSV, HSL, Grayscale and CMYK. A color space defines from 1 to 4 dimensional space, any point within this space represents a color.A color dimensional space (i.e. one dimension per pixel) represents the gray-scale space. Sometimes a alpha-channel is used, this is a mask that indicates how transparent or how well the color merges when the color overlays another color.RGB, HSVmodels are commonly used in color image retrieval system.

The RGB color model is composed of the primary colors Red, Green, and Blue where each primary color represents a dimension. The three colors are added together to produce the desired color. Assign a value for each channel range from 0 to 255Which gives 256^3= 16777126 colors. This model is the most commonly used because it’s perfectly suited for hardware applications, and the human eye strongly responds to red, green and blue.This color space is used in (CRT) monitors and color raster graphics.

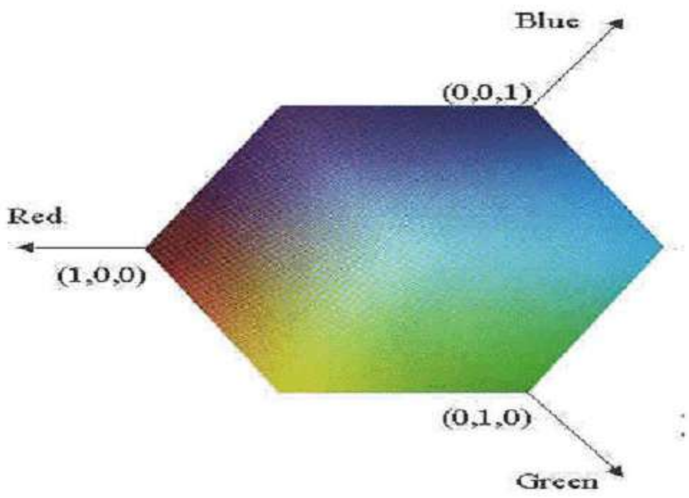

Figure 2.3RGB color.

The histogram is Important in many computer vision applications, which uses to obtain a statistical picture of the basic distribution of data.Hence, color histogram is a statistic for the continuous distribution of three color values. And refers to the probability mass function of the image intensities.

Color histogram is constructed by calculating the number of pixels for each color in the image. It also be possible to use partial information of the Image, this by use advanced techniques of data extraction in RGB color space.

This technique of representing color data shown to have good results for image indexing and retrieval tasks [10]. The color histogram is defined as:

$h_{R,G,B}(r,g,b)=N.Prob(R=r,G=g,B=b)$

Where R, G and B represent the three color channels (RGB) and Nis the number of pixels in the image.

The color histogram can be considered as a group of vectors. For images of the gray scale, these two-dimensional vectors. one of the dimensions is the gray level and the other is number of pixels in the gray level. For color images, color charts consist of 4d vectors.

There are many ways to represent the color histograms, one of the easier it is to display histograms color channels separately. This type of perception some of salient features of color histograms.

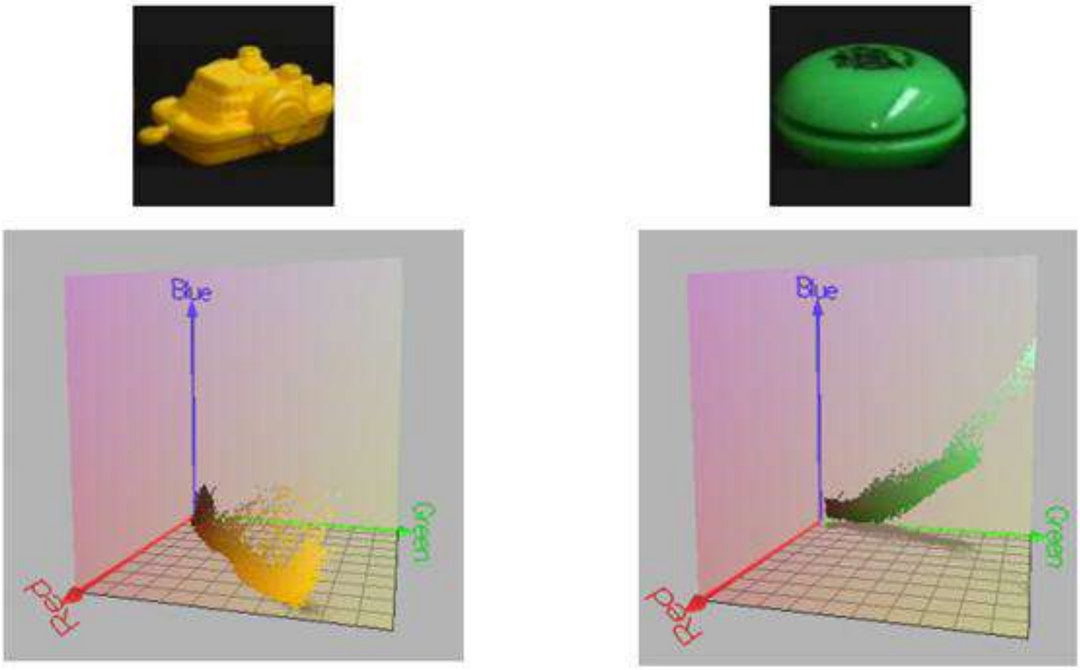

Figure 2.4 Example of color histogram.

2.3 Local Binary Patterns (LBP):

Local Binary Patterns (LBP) was proposed by (Ojala, Pietikainen et al. 2002), the idea of this texture operator is to assign a binary code of eight bits to each pixel depending on the gray levels Neighboring.The gray level of the central pixel (ic) of coordinates (xc, yc) is compared with that of its neighbors (in) according to equation:

$$LBP(x_c,y_c) = \sum_{m=0}^{p-1} s(i_n - i_c) \times 2^n$$

$$s(i_n - i_c) = \begin{array}{lll} 1 & si & i_n - i_c \geq 0 \\ 0 & si & i_n - i_c < 0 \end{array}$$

With p is the number of neighboring pixels. We thus obtain, a matrix containing LBP values between 0 and 255 for each pixel. A histogram is calculated based on these values to form the LBP descriptor.

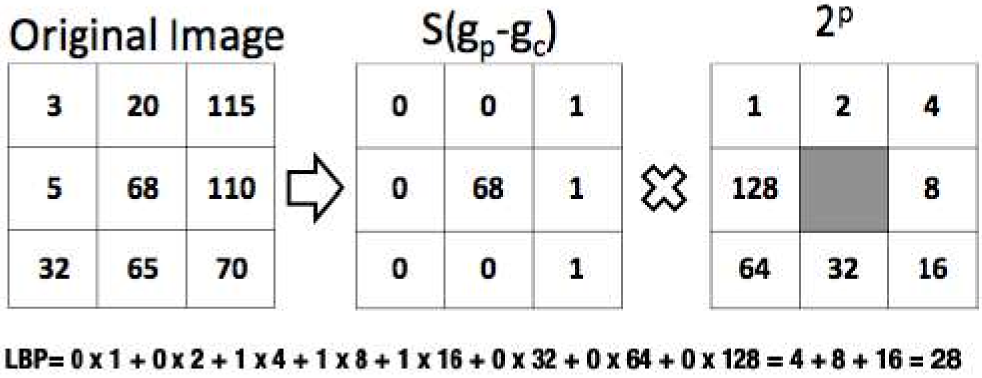

Figure 2.5 Example of LBP feature.

3 Normalization features:

After extracting the features, we noticed that the values of color histogram were greater than the values of the GLCM. This large difference leads to the focus of the treatment process on color histogram feature. To avoid this difference we normalize this features.

The concept of normalization was first developed and documented by E. F. Codd (1972),and is used informally in statistics. The term normalized data have multiple meanings. When you normalize data you eliminate the units of measurement for data, enabling to more easily compare data. The normalization is a scaling technique, as can be considered a pre-processing stage of any type problem[11]. Is takes important role in soft computing, cloud computing. Where it is scale down or scale up the range of data. The result of normalization is a consistentdata.

There are a many normalization techniques Min-Max, Z-score, Standardizing residuals,Moments,**vectors (in linear algebra) to a norm of one**and Decimal scaling normalization[12-13].

- **Rescaling data to have values between 0 and 1(**Min-Max normalization)**:** This is usually called *feature scaling.*Is a simple technique keeps relationship among original data where it gives new range of values between 0 and 1 from an existing one range.

  $$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

  Where X=(X1,...,Xn) the old data and X_{new} is now normalized data.

  This is the technique we used in our approach.

- **z-score or t-score**. is a technique gives natural values or group data from non-indigenous structural data using concepts such as mean and the standard deviation

  $$v_i' = \frac{v_i - \bar{E}}{std(E)}$$

  Where v_i' is Z-score normalized one values and v_i is value of the row E of ith column

  $$\text{std (E)} = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} (v_i - \bar{E})^2} \qquad \bar{E} = \frac{1}{n} \sum_{i=1}^{n} v_i$$

- **Standardizing residuals:** Ratios used in regression analysis can force residuals into the shape of a normal distribution.
- **Normalizing Moments:** using the formula μ/σ.
- **Normalizing vectors (in linear algebra) to a norm of one:** Normalization in this sense means to transform a vector so that it has a length of one.
- **Decimal scaling normalization:** Normalized by moving the decimal point of values of feature X.

  $$DS(X_{ij}) = \frac{X_{ij}}{10^c}$$

  C is smallest integer, DS is the newfeature.

4 Dimensionality reduction

Since the dimensions of the features used are relatively high and in order to reduce them and determine their most distinctive, we use the PCA method.

Principal Component Analysis is a statistical algorithm used to extract and view important information contained in a multivariate data table. The CPA synthesizes this information into only a few new variables called main components.

These new variables correspond to a linear combination of the original variables. The number of principal components is less than or equal to the number of original variables.The pro algorithm show the principal of PCA.

Algorithm: Principal Component Analysis (PCA)[1].

Input: *x*: Raw data of *NxM*dimensions, *D:* new dimension,

Output: *Xnew*: New data with a reduced dimension

Begin

Calculate the mean of*X*

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Centralize the data on $\overline{X}$

$X' = X - \overline{X}$

Calculate the covariance matrix *C*, covariance of two variables *A*and *B*from $\overline{X}$having $\overline{A}$and $\overline{B}$ as mean, is given by

$$Cov(A,B) = \frac{1}{N-1}\sum_{i=1}^{N}(A_i - \bar{A})(B_i - \bar{B})$$

Calculate the eigenvalues of*C*by solving the following equation

$$P(\lambda) = |C - \lambda I|$$

Calculate the eigenvector *V*corresponds to each λ by solving the following the equation

$$Cv = \lambda v$$

Sort the eigenvalues in decreasing order and keep the top *D*ones having the highest values,and sort the eigenvectors accordingly, where *V*denotes the set of top eigenvectors. Project$\overline{X}$ in the new reduced space

$$Xnew = X' \times V$$

End

5 Clustering:

Clustering is an important field of application for a variety of areas, including the extraction of data, and analysis of statistical data, pressure, the muzzling of the data carriers. The clustering was formulated in different ways in the machine learning, and pattern recognition, optimization. The basic objective of the clustering is the collection of data elements that are similar to each other.

There are many clustering algorithms among them K means,Fuzzy k-means,Minimum Spanning Tree(MST), Mulual Neighborhood, Signal-Link, Complete-Link, Mixture Decomposition. We used k-means algorithm to implement this work.

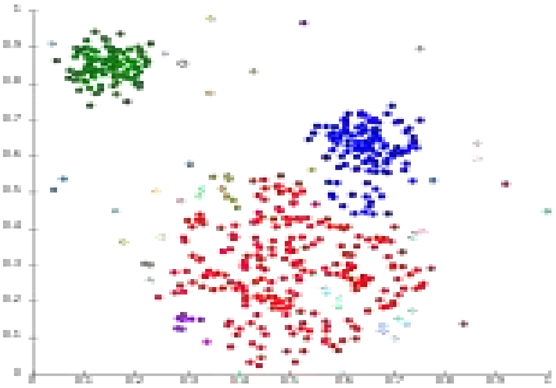

Figure 2.6 Example of clustering.

- **K-means:**

One of the popular clustering algorithm that has been used in a variety of application domains, such as image segmentation (Marroquin &Girosi, 1993) and information retrieval (Bellot& El-Beze, 1999).it is defined either globally (totaldesign) or locally (on the subset from thedesigns), is the K-means technique (Vaishali and Rupa, 2011).

K-means clustering is one of the older predictive n observations in d dimensional space ,and it work to is determine a set of c points to minimize the mean squared distance from each data point to its nearest center with which each observation belongs. The K-means clustering requires less effort.

The aim of clustering is to figure out commonalities and designs from the large data sets by splitting the data into k groups.

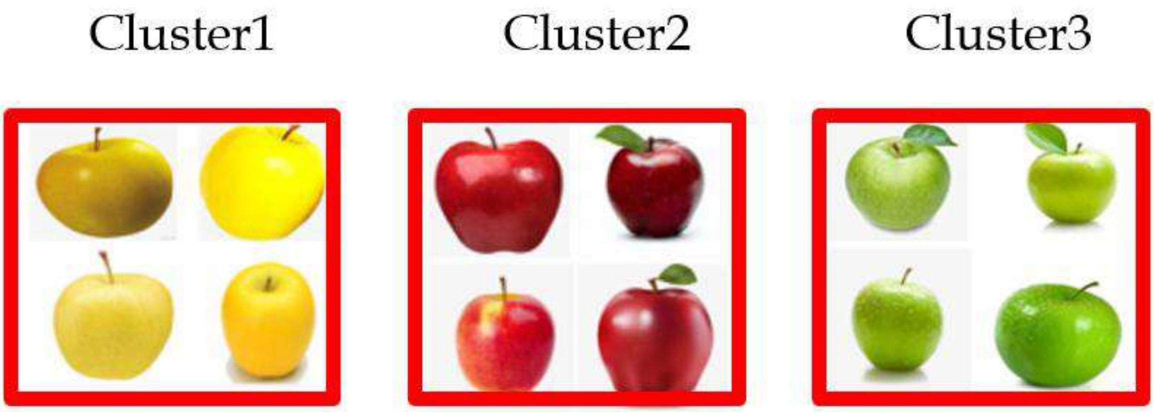

Figure 2.7 Example of K means clustering.

6 Modeling of semantic concepts using Gaussian Mixture Models:

The probability density function(pdf) makes it possible to determine areas of high and low probability for the values of a random variable.

Since the data in each cluster is normally distributed, the probability density function (pdf) is calculated for each cluster representing which makes it possible to determine areas of high and low probability for the values of a random variable according to the following equation:

$$G(x_i|\theta_k) = \frac{1}{\sqrt{|\sum_k|\ (2\pi)^M}} e^{-\frac{1}{2}(x_i - \mu_k)^T \Sigma^{-1} (x_i - \mu_k)}$$

Where x:M-dimensional data vector that belongs to the set$X$

θ_k : denotes the parameters of the Gaussian distribution that corresponds to the k^{th}cluster

μ_k:The mean vector of the points belonging to the cluster

$\sum_k$:The covariance matrix

Then, pdf is merged with Gaussian Mixture Model (GMMmodel of classification)that represents the visual form of the concept. Which is a statistical model expressed in a mixed density. It is usually used to parametrically estimate the distribution of random variables by modeling them as a sum of several Gaussians (called kernels). It is then necessary to determine the variance, the average and the amplitude of each Gaussian. These parameters are optimized according to a maximum likelihood criterion in order to get as close as possible to the desired distribution. This procedure is done most often iteratively via the expectation-maximization (EM) algorithm. It is given:

$$P_{Cp}(x_i|\theta_{Cp}) = \sum_{k=1}^{k'_{Cp}} w_k\, G\,(x_i/\theta_{Cp_k})$$

$$\theta_{Cp} = \{\theta_{Cp_k}, k = 1, \dots, k'_{Cp}\}$$

Where w_k:denotes the weight of thek^{th}distribution.

θ_{cp}: The parameters of Gaussian component densities of the mixtureP_{cp},

The probability of the data that belong to the concept is given by:

$$P = \prod_{i=1}^{N_{Cp}} P_{Cp}\,(x_i|\theta_{Cp})$$

WhereN_{Cp} denotes the number of data points that belong tothe concept..

Here's an example of how to annotate an image.

Class 1 and class 2 are two concept, where Class 1 represent Apple and is contain there 3 clusters Red Apple(Ra), Green Apple(Ga) and Yellow Apple(Ya), and class 2 represent Horseand is contain 2 clusters White horse dotted with black(Hbl), and White horse dotted with brown(Hbr).

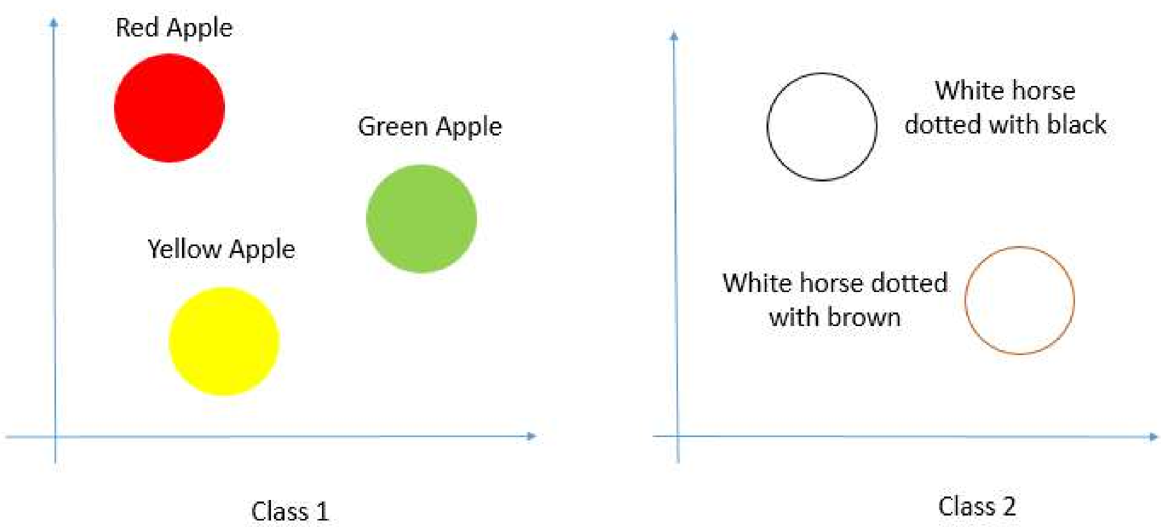

Figure 2.8 Example of tow concept with clustering.

To find the appropriate annotation for the image X we must calculate probability of belonging the image to each concept according to the following equation:

$$P(X|Ci) = \sum(W_{ji} * P(X/Cij))$$

Where ***P(X|Ci)*** is probability of belonging **X** to concept C_i, ***W*** is weight of the cluster *j* in C_i (Weight of each cluster is the number of image in this cluster divide by number of image in the concept), *P(X|Cij)* is probability of belonging **X** to cluster ***j*** in C_i.

Probability of belonging the image X to each concept in our example:

P(X|Apple*)=W(Ra)*P(X|Ra)+W(Ga)*P(X|Ga)+W(Ya)*P(X|Ya).*

P(X|Horse*)=W(*Hbl*)*P(X*|Hbl*)+W(*Hbr*)*P(X*|Hbr*).*

Suppose that *P(X*|Apple*)=0.75 and P(X*|Horse*)=0.63,* and consider that the annotation is in one keyword only. The class that gives the greatest probability is the appropriate annotation, this means that the image is annotated by Appel.

Conclusion:

In this chapter, we have introduced the methodology used to solve the problem of automatic image annotation. From extracting image features to annotate an image, through clustering, model used and other methods.In the next chapter, we will see the results obtained from this methodology.

Chapter III.

Experimental setup, results and discussions

Introduction

In this chapter, we will present the various results obtained from the proposed approach. The chapter contains two sections, experimental setup and experimental results. The first part, we present the experiences of our experiment, including the data sets, performance measures. In the second part, we show the results we obtained with analysis and discussion.

1 Experimental setup:

1.1 Datasets

Datasets are an integral part of the field of machine learning. IAPR TC-12 is considered as a popular image annotation dataset.

➢ Our dataset:

We use the IAPR TC-12 to carry out as our experimental data set, the widely used image data set that was first used in [1] for cross-language retrieval, and has since become a standard benchmark for annotation performance.It's associated with 255 categories with a total of 19627, images taken from sites around the world, this includes photos of various sports, photos of people, animals, cities, landscapes and many other aspects of contemporary life .17665 image from them used as training data, and the rest of 1962 as test data. It contains 291 labels with an average of 5.7 concepts per image. Fig. 3.1 shows typical samples from the IAPR TC-12.

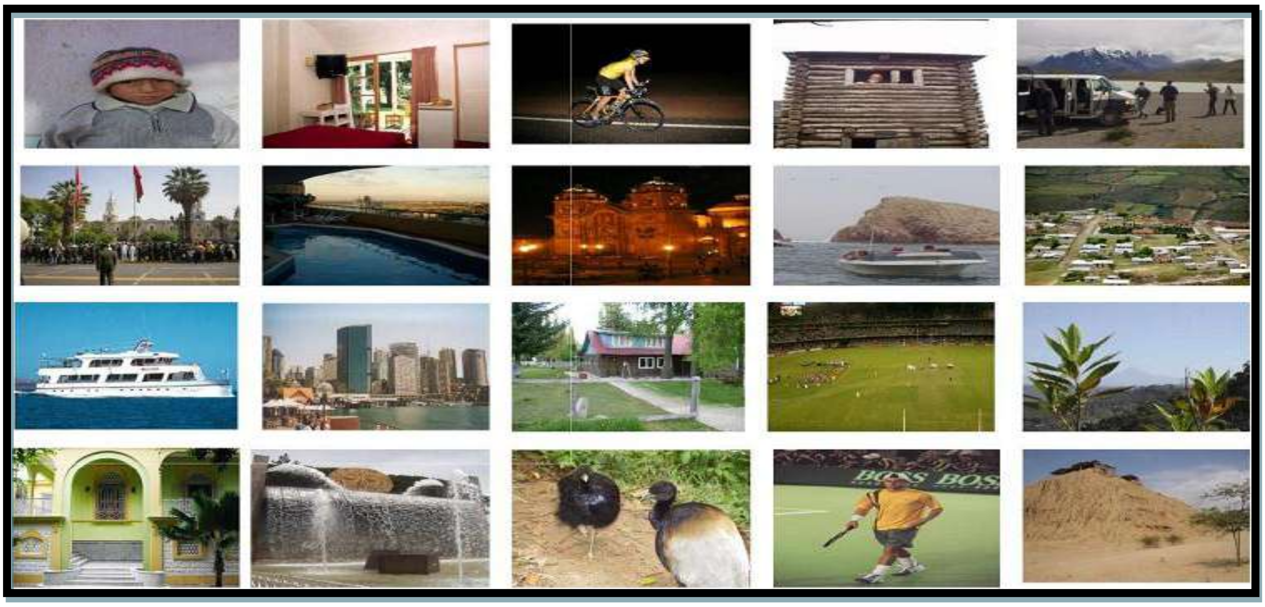

Figure 3.1 Typical samples fromthe IAPR TC-12.

1.2 Performance metrics:

Image annotation performance has been evaluated using a variety of tools. In this work, we take into account two metrics for evaluating the performance for supervised image annotations namely N^+ andmean precision given a concept, the two metrics are given by the following formulas:

$$\boldsymbol{Precision\ c = \frac{Number\ of\ assigned\ image\ to\ c}{number\ of\ correctly\ assigned\ image\ to\ c}}$$

$\boldsymbol{Mean\ Precision = \frac{\sum_{i=1}^{N} Precision\ c}{N}}$ **,In our case N=291.**

$$\mathbf{N}^{+} = \boldsymbol{Number\ of\ concepts\ correctly\ assigned\ to, at\ least, one\ image}$$

1.3 Development tools:

- **Matlab:**

Matlab is a high-level language that allows the execution of tasks requiring a lot of computing power and whose implementation will be much simpler and faster than with traditional programming languages such as C, C ++ . It has several toolboxes in particular that of image processing "Image Processing ToolBox" which proposes a set of algorithm and graphical tools of reference for the treatment, the analysis, the visualization and the development of algorithms image processing.To perform our experience we have used Matlab R2016a version.

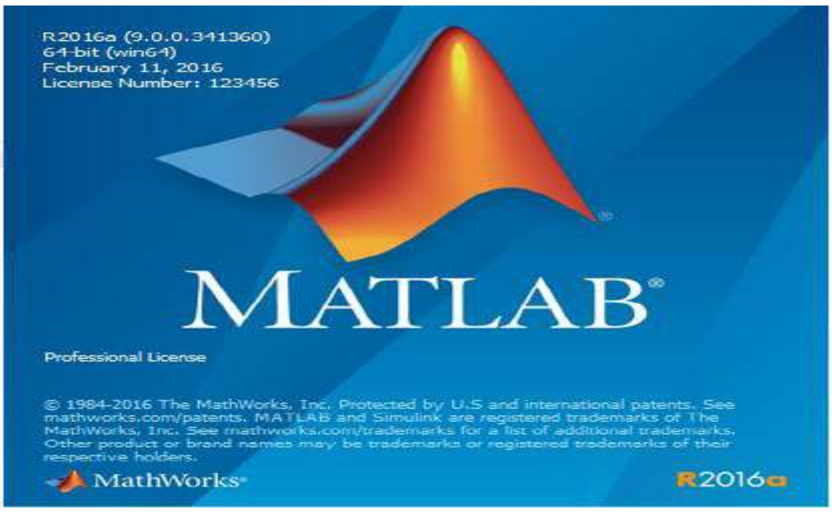

Figure3.2. Matlab R2016a.

2 Experimental result:

In this section we assess the performance of the proposed approach in several aspects. The first is the strength of the original features of the dataset. The second is the power of each feature combination, as for the third is comparing the performance of our approach with that used in previous work.

2.1 First experiment:

We experiment the proposed approachby following the steps in the previous chapter using the original features of the dataset.Figure3.3 shows the precision per concept yielded by feature combination. Figure3.4 shows a sample from the automatic annotation results.

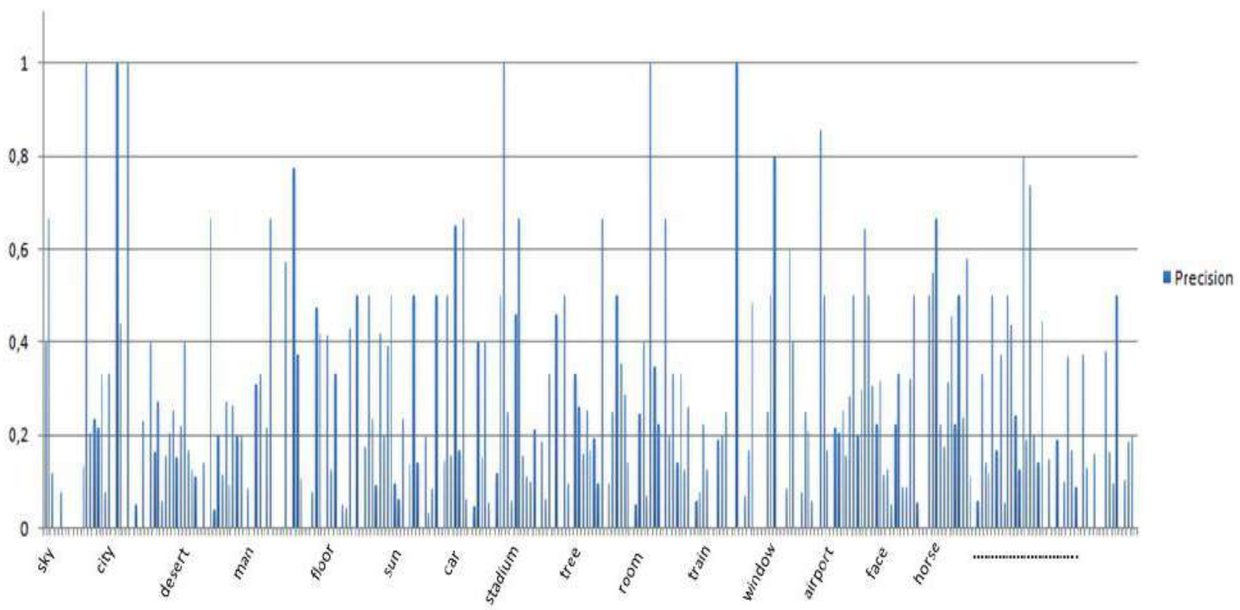

Figure 3.3 Precision per concept yielded by feature .

According to **Figure 3.3,**we note that there is a significant difference in the precision resulting from each concept and its values between 0 and 1.Where it achieve high precision for concepts that can be easily distinguished, Such as the stadium because it has a fixed shape (rectangular and green space). In contrast, some of the concepts achieve lowprecision, this is because it is difficult to identify, Such as man may be difficult to identify because it may be in different situations or performs a movement. We get amean precision of22.71% and N^+ equal to 227.

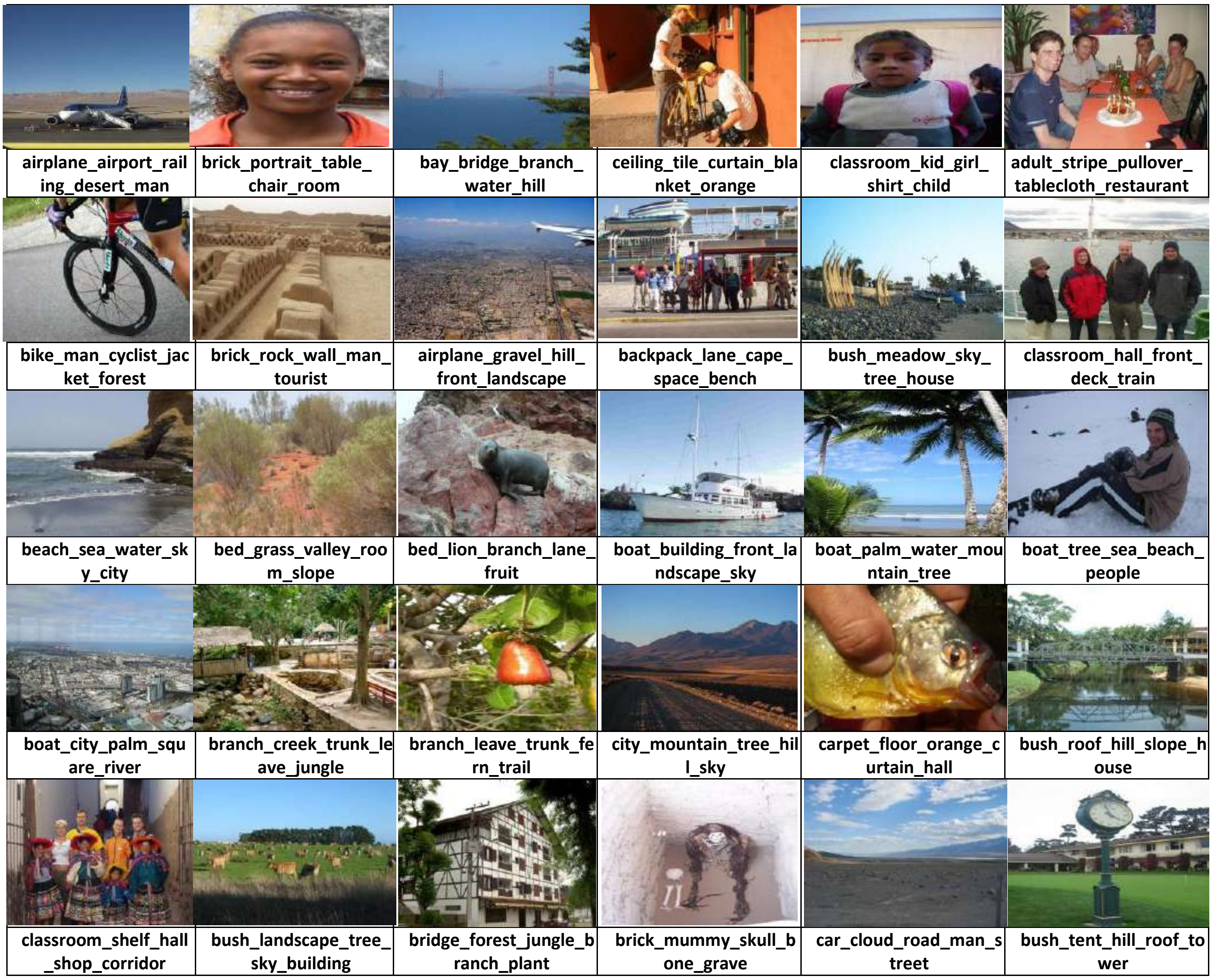

Figure3.4 Result of automatic image annotation using feature combinations.

2.2 Second experiment:

We experience our approach using certain other features namely color histogram (HIST), Local Binary Patterns (LBP), and Gray-Level Co-occurrence Matrix (GLCM). Figure3.5 show the precision per concept yielded by the different feature combination.

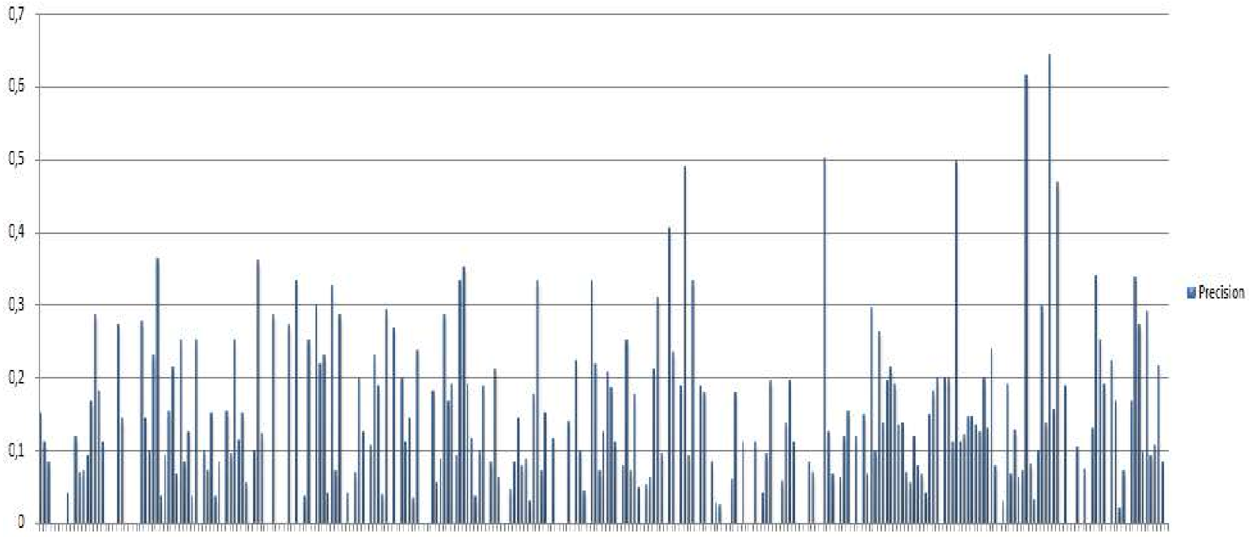

Figure 3.5.precisionper concept yielded by HIST+LBP+GLCM

According to **Figure 3.5**, there is a difference in precision between concepts and their value is limited to 0 and 0.7, due to the ease or difficulty of identifying these concepts

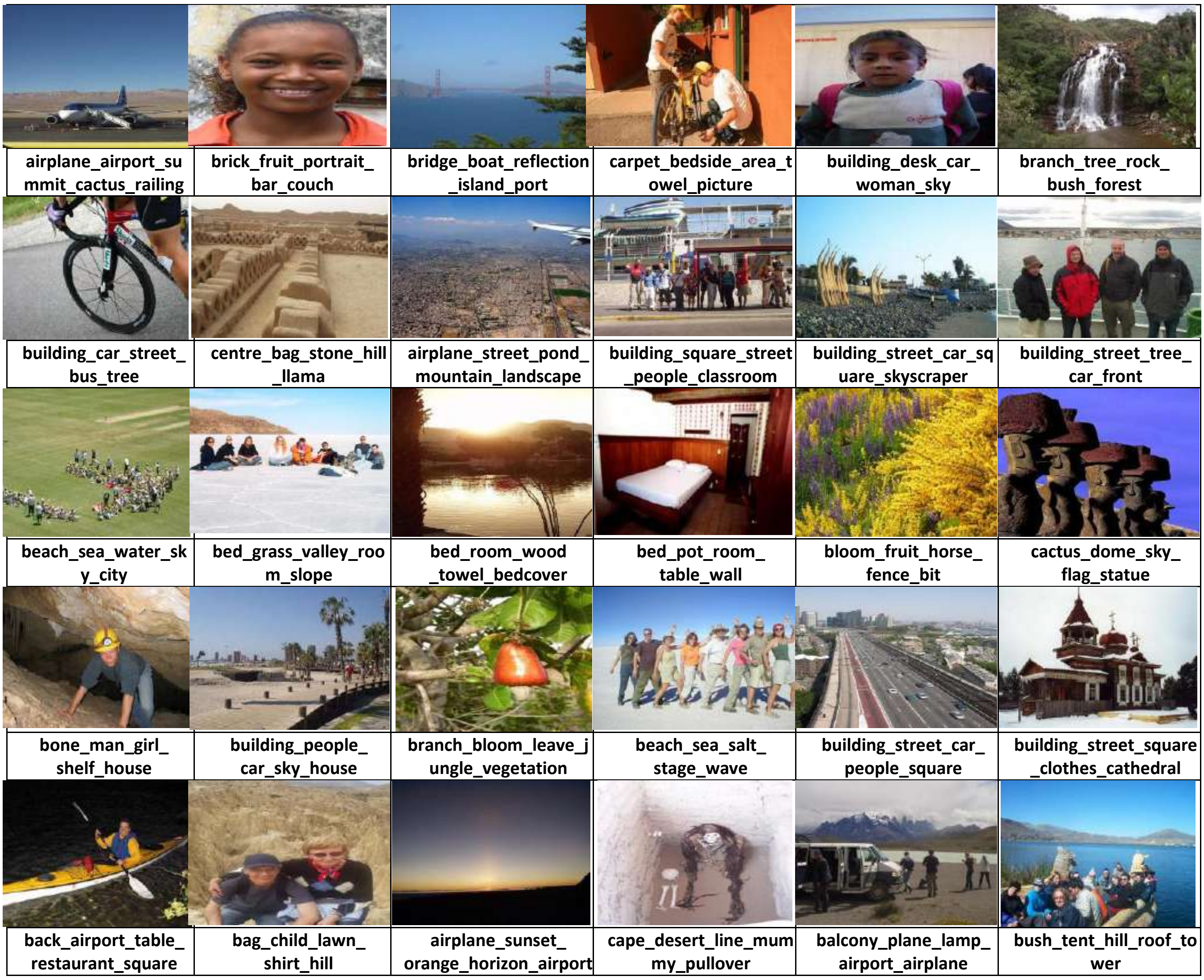

Figure3.6. Result of automatic image annotation using feature combinations.

Table 3.1 Mean precision and N^+ for all concepts the combinations.

	Color HIST	LBP	GLCM	HIST+LBP	HIST+LBP+GLCM
Mean precision	0.0948	0.0558	0.0137	0.1202	0.1195
N^+	173	42	24	203	219

Through the table, we note that there is a difference in mean precision and N^+ resulting from each feature combinations.Where both of HIST + LBP and HIST + LBP + GLCM achieve very high precision/ N^+, HIST achieves very good precision/ $N^+$and LBP achieves a bad precision/ N^+, as for the GLCM the precision/ $N^+$is very low than thatconfirms the need to integrate this feature with others feature.Figure 3.7showthe mean precision per concept yielded by each feature combinations**.**

2.3 Third experimental:

In this part of the experiment, we show the difference between the first and second experiment. Plus the mean precision and N+ yielded by different number of cluster in the best experiment.

Table 3.2 Mean precision and N^+offirst and second experimental.

	First experimental	**Second experimental**
Mean precision	0.2271	0.1195
N^+	227	219

From the Table 3.2, we observe that both mean precision and N^+ of the first experiment was higher than the second experiment. This means that annotation by the original features of the dataset was the best.

2.3.1 Effect of changing K in k-means

Since the annotation by using the original features of the dataset was best, we calculated the mean precision and N^+ as function of the number of clusters. The results are shown in the figure below.

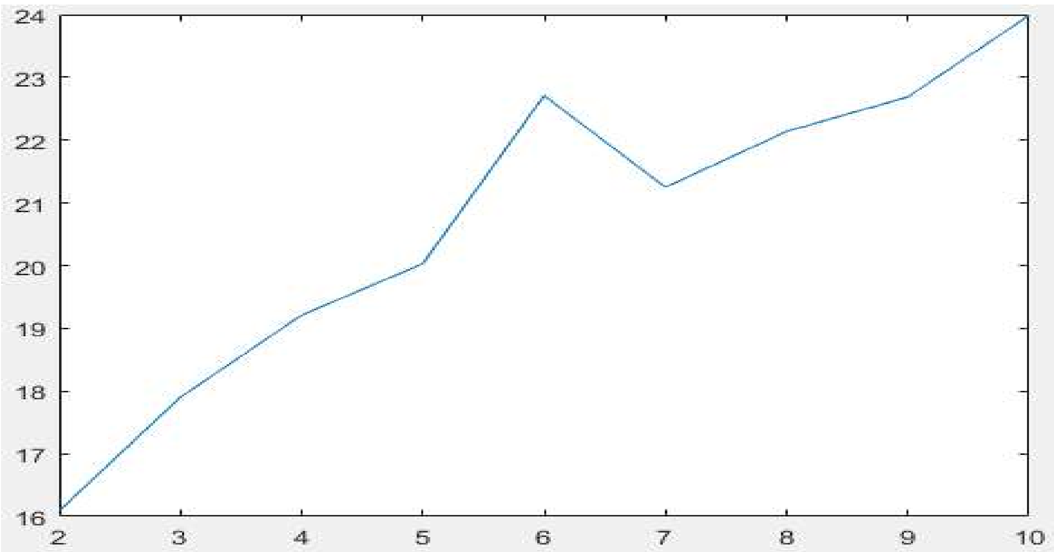

Figure 3.7 Mean precisionyielded by different number of cluster.

According to the Figure 3.7 ,we note that whenever the number of K increased the Mean precision increased as well.

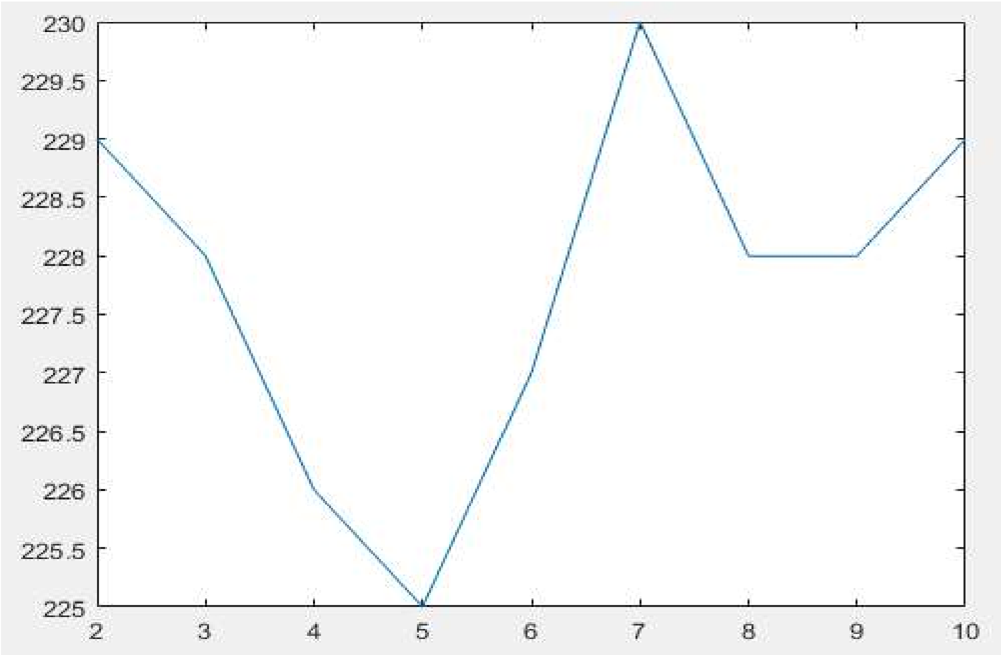

Figure 3.8N^+ yielded by different number of cluster.

From the Figure 3.7 We note that there is a fluctuation in N^+ values, higher one was k=7 and low was in k=5.

3 Comparison with state of the art:

In this part we have compared the performance of our approach with the performance of other approaches from the state of the art.

Table 3.3 Comparison of the proposed approach with other approaches

	Mean precision	**N^+**
MBRM[15]	0.24	223
Ours	0.2271	227

From the Table 3.3, we notice that the proposed approach has outperformed the previous approach in terms of N^+.

4 Interface

In this part we represent the interfaces of our system.

The first interface is for project presentation, it contain tow buttons:

GO: *is to start the program.*

EXIT:*is to close the program.*

Figure 3.9. First interface.

The second interface contain some pictures related to our work AIA. The button NEXT is to pass to the following interface.

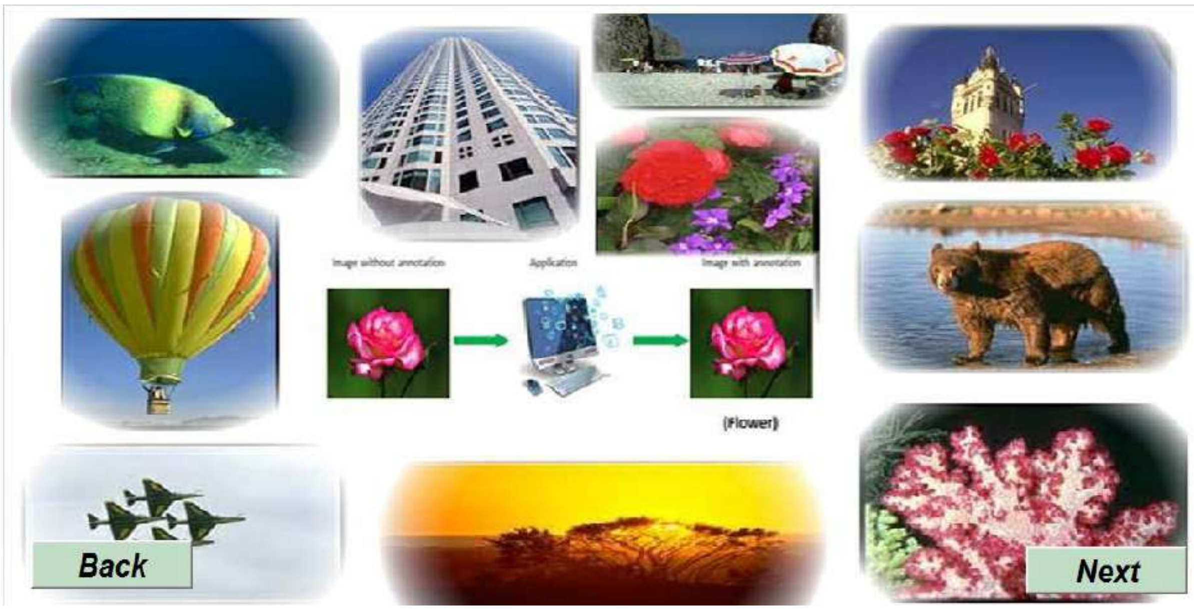

Figure 3.10 Second interface

The third interface, contain a small definition of our system. It contain also four buttons: *Load image, Annotation, Back and Close.*Each button applied a certain method as it represent in the figure bellow.

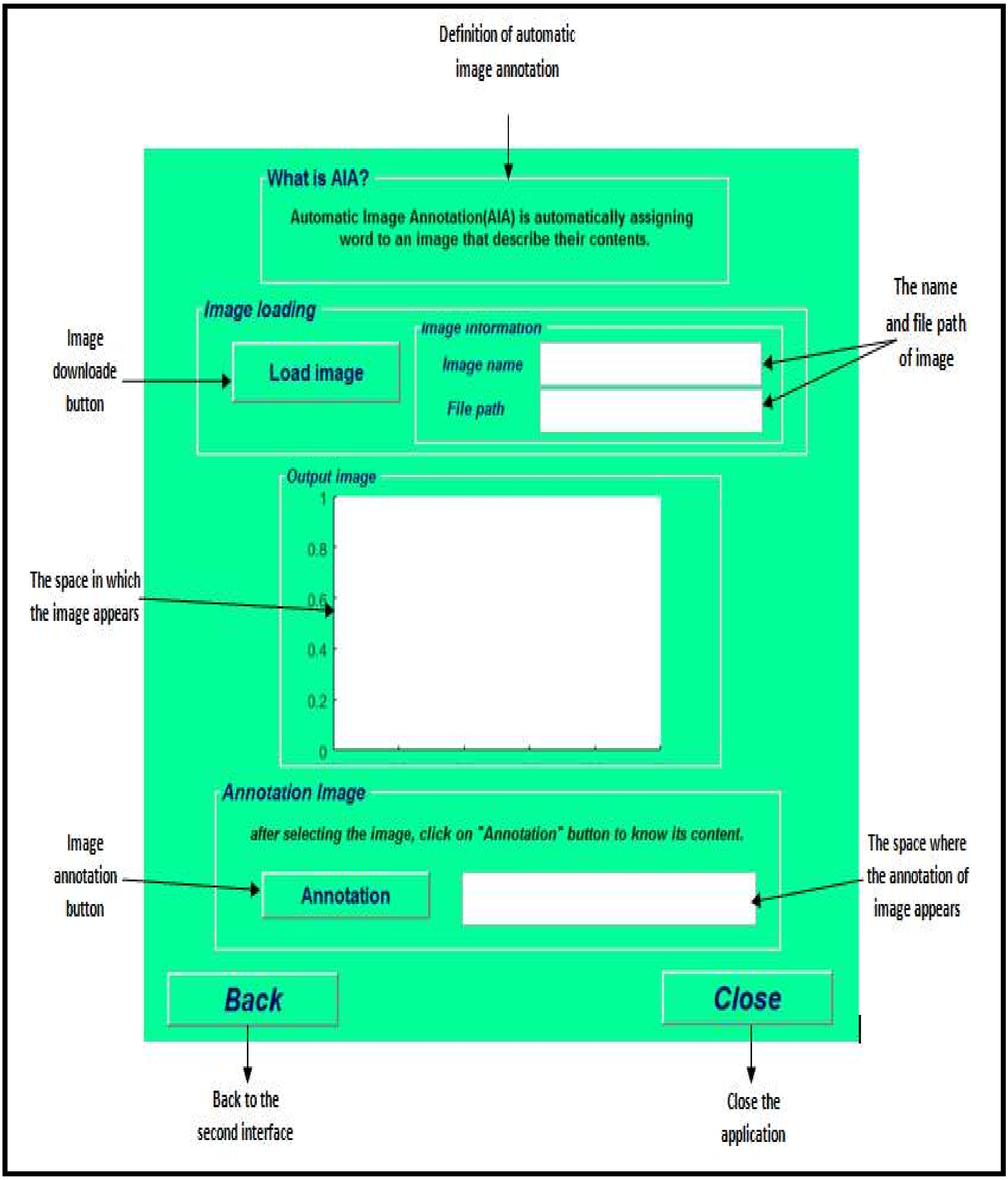

Figure 3.11 Third interface and its components.

Once we loaded the image, itsinformations be displayed, then we click on annotation button for display the annotation result of image loaded.

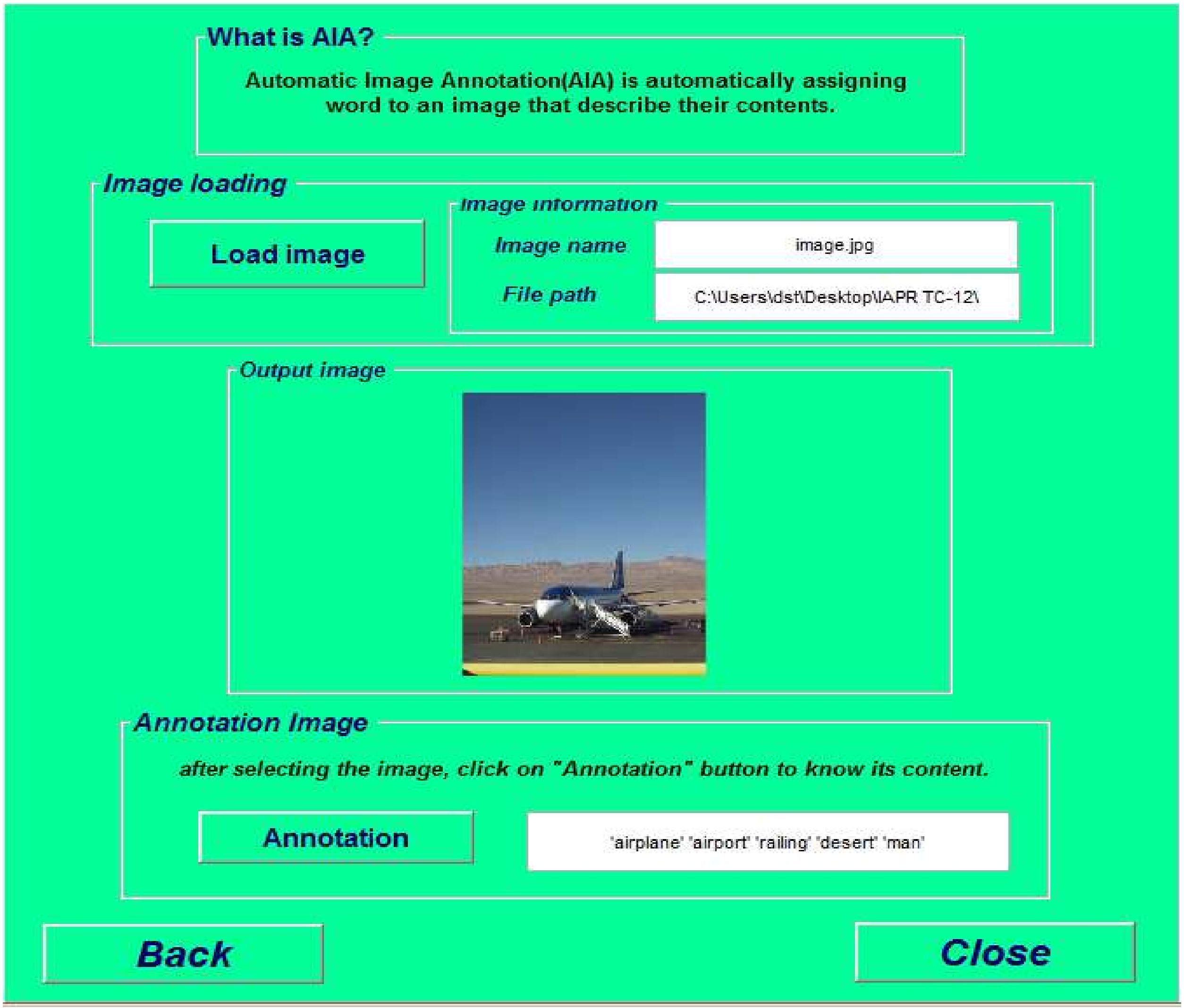

Figure 3.12 Interface ofannotation results.

Conclusion

In the last chapter of this thesis, we tested the effectiveness of the proposed approach and mentioned the experimental results obtained. The first step was to address the set of data used and performance measures. Then, our findings on the annotation on the images. After several experiments, the efficiency of our approach was clearly demonstrated. By testing multi-feature combinations, the strength of the proposed approach has been successfully demonstrated

GENERAL CONCLUSION

In this thesis we touched three chapters, the first was about the need of automatic image annotation in image retrieval, the second was about the proposed approach as for the third was the experimental result of our approach.

The proposed approach based on Gaussian Mixture Model (GMM), we experimented the approach on IARP TC-12 database, where we used the original features of dataset and external features (color histogram, LBP and GLCM). The result with original features of dataset was best then HIST+LBP+GLCM.

Through this work, we have shown the importance of AIAand its need in the world of technology.

Bibliography

[1] Aiadi.O (2017). “Bridging the semantic gap in image search via the transformation of semantic concepts into a visual representation”.

[2] Dengsheng Zhang and Guojun Lu. EVALUATION OF SIMILARITY MEASUREMENT FOR IMAGE RETRIVAL.2003.

[3] João Augusto da Silva Júnior, Rodiney Elias Marçal and Marcos Aurélio Batista. Image Retrieval: Importance and Applications 2014

[4] D. Grangier and S. Bengio. A discriminative kernelbased approach to rank images from text queries. PAMI, 30(8):1371–1384, 2008.

[5] J. Li and J. Wang. Real-time computerized annotation of pictures. PAMI, 30(6):985–002, 2008.

[7] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In ICCV, 2009.

[8] Olfa Charfi Marrakchi. « ELABORATION DE SIGNATURES DES TEXTURES D’IMAGES DETÉLÉDÉTECTION ». Teledetection, Scientifics editions GB, 2013, 11 (2), pp.337-350.

[9] James Z. Wang, "Integrated Region-Based Image Retrieval", Boston, Kluwer Academic Publishers, 2001

[10] M. Swain and D. Ballard. Indexing via color histograms. International Journal of Computer Vision, 7, 1991.

[11] Shalabi, L.A., Z. Shaaban and B. Kasasbeh, Data Mining: A Preprocessing Engine, J. Comput. Sci., 2: 735-739, 2006

[12] Sanjaya K. Panda, Subhrajit Nag and Prasanta K. Jana, “A Smoothing Based Task Scheduling Algorithm for Heterogeneous Multi-Cloud Environment”, 3rd IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC), IEEE, Waknaghat, 11th - 13th Dec 2014.

[13] Sanjaya K. Panda and Prasanta K. Jana, “A Multi-Objective Task Scheduling Algorithm for Heterogeneous Multi-cloud Environment”, International Conference on Electronic Design, Computer Networks and Automated Verification (EDCAV), IEEE, Meghalaya, 29th –30th Jan 2015.

[14] Grubinger, M.: Analysis and Evaluation of Visual Information Systems Performance. PhD thesis, Victoria University, Melbourne, Australia (2007)

[15]Feng, S., Manmatha, R., Lavrenko, V.: Multiple Bernoulli Relevance Models for Image and Video Annotation. In: CVPR. (2004)

Swain, M. J. and D. H. Ballard (1991). "Color indexing." International j ournal of computer

vision 7(1): 11-32.

Flickner, M., H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee and D. Petkovic (1995). "Query by image and video content: The QBIC system."

Computer 28(9): 23-32.

Haralick, R. M. and K. Shanmugam (1973). "Textural features for image classification." IEEE Transactions on systems, man, and cybernetics(6): 610-621.

Ojala, T., M. Pietikainen and T. Maenpaa (2002). "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Transactions on pattern analysis and machine intelligence 24(7): 971-987.

Zhang, D. and G. Lu (2002). A comparative study of Fourier descriptors for shape representation and retrieval. Proc. of 5th Asian Conference on Computer Vision (ACCV), Citeseer.

Tieng, Q. M. and W. Boles (1997). "Recognition of 2D object contours using the wavelet transform zero-crossing representation." IEEE Transactions on Pattern Analysis and Machine Intelligence 19(8): 910-916.

Jain, A. K., M. N. Murty and P. J. Flynn (1999). "Data clustering: a review." ACM computing surveys (CSUR) 31(3): 264-323.

Arbelaitz, O., I. Gurrutxaga, J. Muguerza, J. M. PéRez and I. Perona (2013). "An extensive comparative study of cluster validity indices." Pattern Recognition 46(1): 243-256.

Zhou, N., W. K. Cheung, G. Qiu and X. Xue (2011). "A hybrid probabilistic model for unified collaborative and content-based image tagging." IEEE Transactions on Pattern Analysis and Machine Intelligence 33(7): 1281-1294.

Makadia, A., V. Pavlovic and S. Kumar (2008). A new baseline for image annotation. European conference on computer vision, Springer.

Zhao, P., Y. Lu, W. Wang and W. Zhu (2013). Automatic image annotation using semantic relevance. Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service, ACM.

Wang, M., X. Zhou and T.-S. Chua (2008). Automatic image annotation via local multi -label classification. Proceedings of the 2008 international conference on Content-based image and video retrieval, ACM.

Ballan, L., T. Uricchio, L. Seidenari and A. Del Bimbo (2014). A cross-media model for automatic image annotation. Proceedings of International Conference on Multimedia Retrieval, ACM.

Murthy, V. N., S. Maji and R. Manmatha (2015). Automatic image annotation using deep learning representations. Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, ACM.

Tsoumakas, G. and M.-l. Zhang (2009). "Learning from multi -label data."

Sun, F., J. Tang, H. Li, G.-J. Qi and T. S. Huang (2014). "Multi -label image categorization with sparse factor representation." IEEE Transactions on Image Processing 23(3): 1028-1037.

Duygulu, P., K. Barnard, J. F. de Freitas and D. A. Forsyth (2002). Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. European conference on computer vision, Springer.

Jeon, J., V. Lavrenko and R. Manmatha (2003). Automatic image annotation and retrieval using cross-media relevance models. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, ACM.

Lavrenko, V., R. Manmatha and J. Jeon (2003). A model for learning the semantics of pictures. Advances in neural information processing systems.

Gong, T., S. Li and C. L. Tan (2010). A semantic similarity language model to improve automatic image annotation. 2010 22nd IEEE International Conference on Tools with Artificial Intelligence, IEEE.

Liu, J., B. Wang, M. Li, Z. Li, W. Ma, H. Lu and S. Ma (2007). Dual cross-media relevance model for image annotation. Proceedings of the 15th ACM international conference on Multimedia, ACM.

Wu, L., R. Jin and A. K. Jain (2013). "Tag completion for image retrieval." IEEE Transactions on Pattern Analysis and Machine Intelligence 35(3): 716-727.