2020

Construction of Statistical Visual Descriptors for CBIR Exploiting Time-inhomogeneous Markov Chain Model

Islam, Md. Saiful

University of Rajshahi, Rajshahi

Construction of Statistical Visual Descriptors for CBIR Exploiting Time-inhomogeneous Markov Chain Model

by

Md. Saiful Islam

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

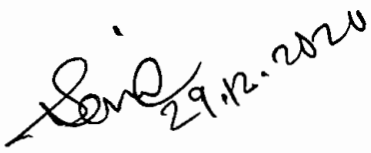
UNIVERSITY OF RAJSHAHI

Dec, 2020

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



Md. Saiful Islam
Associate Professor & PhD Researcher
Department of Computer Science and Engineering
Rajshahi University, Rajshahi, BANGLADESH
Dec, 2020

CERTIFICATE

I have the pleasure to certify that the thesis entitled "Construction of Statistical Visual Descriptors for CBIR Exploiting Time-inhomogeneous Markov Chain Model" has been prepared by Md. Saiful Islam, Associate Professor, Department of Computer Science & Engineering, Rajshahi University, Bangladesh under my supervision and guidance. The entire thesis is the achievement of the candidate's own work.

I also certify that I have thoroughly gone through the final draft of the thesis and found it satisfactory for submission to the Rajshahi University, in fulfilment of the requirements for the award of the degree of Doctor of Philosophy (PhD).

To the best of my knowledge this work or any of its part has not been previously submitted to any other institution for any other degree or diploma.

Dr. Md. Ekramul Hamid
Supervisor and Professor
Department of Computer Science & Engineering
University of Rajshahi, Rajshahi, Bangladesh

The thesis is dedicated

To

My Respected Parents:
Late Abdul Hai and Begum Samsun Nahar

Construction of Statistical Visual Descriptors for CBIR Exploiting Time-inhomogeneous Markov Chain Model

Abstract

Markov stationary features (MSF) based on homogeneous Markov chain model for content-based image retrieval (CBIR) is getting popularity nowadays. It not only considers the distribution of colors that the histogram method does, but also characterizes the spatial co-occurrence of histogram patterns. However, handling a large scale database of images with large degree of heterogeneity, a simple MSF method is not sufficient to discriminate the images. The method does not capture sufficient spatial co-occurrence information as required for large databases. To overcome the shortcoming in this research, two extended methods namely multi-channel nonhomogeneous MSF (MCN-MSF) and multi-resolution nonhomogeneous MSF (MRN-MSF) based on original MSF are proposed. In both cases, the concept of nonhomogeneous Markov chain model is exploited to construct the features. For the first method, we incorporate spatial co-occurrence structures of different color channels of an image by applying the time inhomogeneous (nonhomogeneous) Markov chain model. For the second method, by exploiting the similar nonhomogeneous model, we incorporate the spatial co-occurrence information more consistently by mapping the image with different resolution. Without compromising effectiveness and robustness, the methods proposed in this paper keeps the features level simplicity. Widely recognized WANG1000 and Corel10800 databases are used to evaluate and compare the

✓

performances of the proposed algorithms with the existing methods. The experimental results show that both methods significantly improve the performances compare to the existing methods. The results also prove that second method is more effective for large databases.

Keywords: Markov stationary features, Multi-channel Nonhomogeneous MSF, Multi-resolution Nonhomogeneous MSF, Content-based Image retrieval.

Acknowledgements

At first, I would like to give my earnest gratitude to Almighty Allah for giving me long-term patience and all the opportunities to accomplish my doctoral thesis.

I sincerely wish to thank my supervisor Professor Dr. Ekramul Hamid for his sincerely guidance in the research. His constant encouragement has been fruitful for entire period of research activities.

I especially would like to thank my departmental Chairman, Professor Dr. Bimal Kumar Pramanik for his generous support in my difficult times.

Thanks are also given to all my colleagues including Professor Dr. A. K. M. Akhtar Hossain, Professor Dr. Somlal Das, Md. Rokanujjaman, Md. Tohidul Islam, Subrata Pramanik for their co-operative attitude during my research works.

I would like to give my special thanks to Professor Dr. Md. Emdadul Haque for his all kinds of cooperation. His enthusiastic engagement in my research and his never-ending stream of ideas have been absolutely essential for the results presented here.

I am grateful to all my family members especially to my mother, Begum Samsun Nahar, for her prayer during the years of PhD course.

Finally, special thanks go to my beloved wife, who always supports and encourages me with her great love and patience whenever and wherever.

I greatly appreciate the financial support given me by the Information and Communication Technology Division, Ministry of Posts, Telecommunication and Information Technology of Bangladesh, under the ICT-fellowship program.

Contents

:

List of Figures

List of Tables

List of Acronyms

CBIR	Content-based Image Retrieval
CAC	Color Auto Correlogram
DB	Database
DBMS	Database Management System
HOG	Histogram of Gradient
HSV	Hue Saturation Value
LBP	Local Binary Pattern
MAA	Mean Average Accuracy
MAP	Mean Average Precision
MAR	Mean Average Recall
MCN-MSF	Multi-channel Nonhomogeneous MSF
MRN-MSF	Multi-resolution Nonhomogeneous MSF
MSF	Markov Stationary Feature
RGB	Red Green Blue
SIFT	Scale-invariant Feature Transform
SURF	Speeded up Robust Features
TAA	Total Average Accuracy
TAP	Total Average Precision
TAR	Total Average Recall
TBIR	Text-based Image Retrieval

Chapter 1

Introduction

Recent years have seen a rapid increase in the size of digital image collections together with the fast growth of the Internet. Digital images have found their way into many application areas, including Geographical Information System, Office Automation, Medical Imaging, Computer Aided Design, Computer Aided Manufacturing, Robotics, etc. There are currently billions of web pages available on the Internet using hundreds of millions (both still and moving) images. However, we cannot access or make use of the information in these huge image collections unless they are organized so as to allow efficient browsing, searching, and retrieval over all textual and image data.

The straightforward solution to managing image databases is to use existing keyword or text-based techniques i.e., retrieving images by their textual labels or surrounding text. Due to the advances of textual information retrieval, text-based image retrieval has been the most successful image retrieval strategy for decades. Text and alphanumeric symbols are used to describe images in the text-based approach. The main advantage of this approach is the potential use of data modelling, multidimensional indexing and query evaluation methods, which have matured in the text-based databases. However, this retrieval paradigm is sufficient to meet most users' information needs if images are well-annotated by textual information. With the growing popularity of social networks, people are now generating and sharing image content at a much faster rate. Many of these images are without informative text annotations. Moreover, users are now able to easily snap anything they see by using their mobile devices; and they would like to use the images they snapped as queries to immediately search for relevant images. On the other hand, the rich content of images makes the annotation to be a conceptually difficult task. For example, describing a landscape or a country map by keywords is hard and underlying visual meaning could not be conveyed

completely. Besides, manual annotation is an exhausting and protracted task specifically in large-scale databases. Another problem is that, different people may perceive a specific image differently. Human perceptual judgment is essential for many applications and textual approaches may cause unexpected and unrecoverable mismatches in the final retrieval stage, because language mismatch can occur when the user and the domain expert use different keywords. If the image database is very large and to be shared globally, linguistic barriers will render the use of text-based approach impractical and ineffective. These difficulties have led the researchers to pay attention to an alternative approach [1-4], a fully automated Content-based Image Retrieval (CBIR).

1.1 Research Background

Images from database are indexed by summarizing their visual contents through automatically extracted quantities of features such as color, texture, shape and spatial relationship according to user's visual requirement. Color is one of the most flexible and reliable visual features used in image retrieval or other image classifications systems. The feature is almost independent of image size and orientation, which is robust to its background complication.

Image histogram is widely used for visual representation for its robustness and simplicity to image variations. Histogram representations, *e.g.*, color histogram [5], histogram of Local binary patterns [6], and Bag-of-Words based on SIFT features [7], have been widely used in computer vision and multimedia communities for visual recognition, content based image retrieval, and video content analysis. The histogram describes the global visual content distribution in the image. These methods are based on first-order image global statistics. However, the histogram comparison saturates the discrimination when the database contains a large number of images. Thus, it was inevitable to integrate local spatial information of an image with its global image information. Layout histograms and multi-resolution histograms [8] are the pioneering attempts to incorporate spatial information for improving the discriminating capability of the

histogram features. In Coherence Vector method [9], proposed by G. Pass *et al.*, the histogram bins are divided into two types: coherent and incoherent. According to [9], a bin is coherent if it belongs to large color region. Otherwise, the bin is incoherent. The coherence vector method shows better performance than histograms when the images in the database have mostly uniform colors and the image is texture dominated. Auto-correlogram method, introduced by Huang *et al.* [10-11], is used to characterize both the color distribution of image pixels and the spatial correlation of pairs of colors. Instead of the indirect use of spatial information, the auto-correlogram method encodes local spatial structure information directly into histograms. However, this method only take cares the information of within-histogram bins. An effective content-based image feature called Markov Stationary Feature (MSF) was introduced by Li *et al.* [12] for image representation. It characterizes the spatial co-occurrence of histogram patterns by Markov chain models. The methods are generally based on the second-order spatial co-occurrence statistics of histogram bins, and their effectiveness in characterizing spatial structure are still limited.

1.2 Motivation

In general, local spatial relationship modeling (in an image) is receiving more and more attention in the community of computer vision arena because of its wide prospective applications in image retrieval [13-15], classification [12], video analysis [16], etc. It has appealed an increasingly larger group of scholars to research on this topic. Specifically, the local spatial structure of an image encodes two aspects of information named local appearance and local geometric structure [17].

A number of state-of-the-art methods for the local spatial relationship modeling of an image considers co-occurrence matrix which is actually the co-occurrence properties of the image local features [14, 18, 19]. A co-occurrence matrix or co-occurrence distribution is a matrix or distribution that is defined over an image to be the distribution of co-occurring values at a given offset. For visual

classification tasks, the co-occurrence matrix can measure the texture of the image by considering the image local features, *e.g.*, intensity or grayscale values of the image [14] or various dimensions of color, as well as other local image features such as edges [19]. The original co-occurrence matrices are typically large and sparse, therefore, various algorithmic extensions and development have been proposed to get a more compact and useful set of features. Spatial co-occurrence matrix based Markov chain model introduced by Li *et al.* [12] to encode the intra-histogram-bin and inter-histogram-bin relationships into histograms, where the initial and stationary distributions of the Markov chain model are combined to form the so-called *Markov stationary features* (MSF). The MSF approach achieves a more compact feature representation from the original image feature co-occurrences locally (*i.e.*, pixel pairs) [18] for encoding local spatial information in image classification jobs [12, 17].

Despite the successes of improving image discriminating power by the MSF approach, there are still many essential issues and problems unresolved in both theory and practice. The simple MSF may not be useful to model the local spatial information with more than two pixels (3^{rd}-order or higher order moment of distribution) [17]. However, the higher-order spatial structure can convey much richer and more descriptive information which can be collected by exploiting the concept of nonhomogeneous Markov chain model. In this context, this paper proposes two extended MSF approaches based on the nonhomogeneous scheme to the cope the problems to some extent.

1.3 Contribution

The existing content-based image retrieval systems does not contain sufficient spatial co-occurrence structures to distinguish images fairly for a huge database with large number of categories and greater degree of heterogeneity. In this thesis we eliminate the limitations of the existing methods by introducing nonhomogeneous schemes. However, the contributions of these research is as follows:

i) We formulate the nonhomogeneous Markov chain models to collect significant amount of spatial information to improve the image classification performances.

ii) Multi-channel nonhomogeneous MSF algorithm is proposed in which an image is mapped into a number of color channels. From the color channels the co-occurrence spatial information are collected and combined for the MCN-MSF.

iii) Multi-resolution nonhomogeneous MSF algorithm is proposed to capture more consistent spatial information to distinguish the images of larger databases.

1.4 Organization

The organization of this thesis is as follows. Chapter 2 contains the introductory concepts of text-based and content-based image retrieval systems. We illustrate details about the components of CBIR systems.

Chapter 3 illustrates the state-of-art formulation of Markov stationary features (MSF) based on homogeneous Markov chain model. Then nonhomogeneous scheme is introduced and formulate with all necessary mathematical derivations and examples.

Chapter 4 describes the multi-channel nonhomogeneous MSF algorithm. This chapter contains key performance metrics including recall, precision and accuracy with proper mathematical notations to evaluate the performances of the proposed algorithm over the existing methods. The experimental results of the methods using two recognized databases named WANG1000 and COREL10800, and discussions are given with different figures and tables in the chapter.

Chapter 5 discuss the proposed multi-resolution nonhomogeneous MSF algorithm. The performances of the algorithm is evaluated with the same databases. The comparative results are given with a number of figures and tables. It also contains the summary of comparisons all methods including the proposed two algorithms.

Chapter 6 contains concluding remarks with the direction of future works.

Chapter 2

Brief Overview on Image Retrieval

2.1 Introduction

Since 1970s, both Database Management and Computer Vision community has been working on image retrieval from different perspectives [20]. While the former is interested in text-based image retrieval in which the retrieval system is performed by employing the information retrieval based on the surrounding text or annotation text of images, while the latter explores the content-based image retrieval in which the image is represented by its salient features depending on some representations of visual contents of images (such as color, texture shape, objects). Thanks to the maturity of textual information retrieval techniques, text-based image retrieval has been well-studied, leading to several successful commercial systems like Google Images search. In this chapter, we summarize these two approaches by emphasizing the second one.

2.2 Text-based Image Retrieval

In text-based image retrieval (TBIR) systems, an operator, more specifically a human indexer, manually annotates each image in the collection by text. In other words, images are described as a set of keywords or free text [21]. Then, such annotations are stored in a traditional Database Management System (DBMS) [22, 23].

Generally, images are retrieved from DBMS by using conventional queries that are executed on exact or probabilistic match of the query text. Query text can be either a single keyword or a description of an object depicted in the image. The DBMS cannot retrieve the desired images, unless the images in the database are

correctly and sufficiently described. Retrieval performance is directly related to congruence between the vocabularies of the operator(s) and the user(s) of a TBIR system.

2.2.1 Difficulties with TBIR

TBIR has two main problems. The first problem emerges due to the human subjectivity on complex images. Since "a picture is worth a thousand words", it is not always possible to define or describe wide variety of images just by using some textual information. Also, different people or the same person in different situations describe or judge the same image differently, due to human perception subjectivity. Fig. 2.1 shows such subjectivity by example. Some people see a rabbit in Fig. 2.1.a while the others see a duck. Again, some people see nine human faces and a dog lying on street in front of a haunted house if they carefully examine the image in Fig. 2.1.b while the others see immediately an old bearded man with his hand resting on one flap of his jacket.

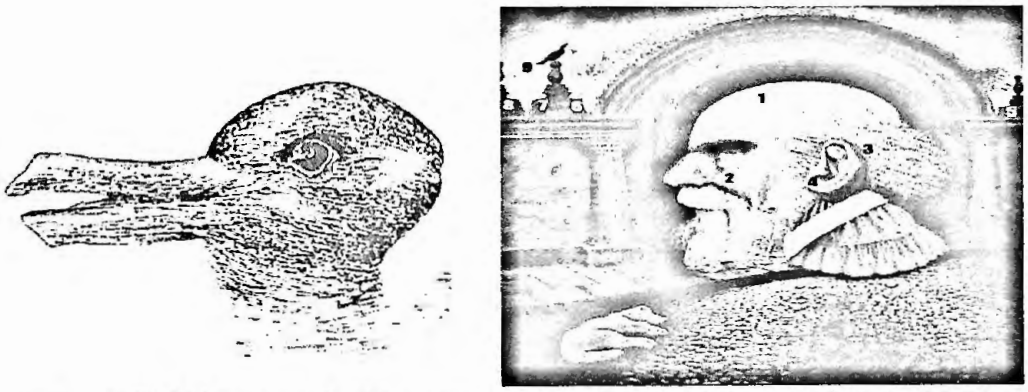
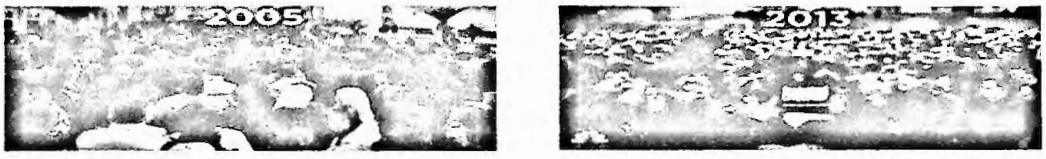


Fig. 2.1. a) Rabbit or duck? b) Can you see 9 faces or an old bearded man with his hand resting on one flap of his jacket?

The second problem lies in the difficulty of manual indexing. Thanks to the advancement of mobile devices and internet, the number of images gets larger and the total amount of time spent in manual image annotation is increased. This results two congenital defects. The first defect is that images have to speak for themselves since the nature of image is beyond words. Compared to words, it is more inherent for users to express their intents by images. Of late, people are more willing to snap photos and search directly from mobile devices. This triggers the

demand of CBIR once again (see Fig. 2.2.a). The second defect is the prohibitive labor cost in obtaining accurate textual description for the vast amount of images. As illustrated in Fig. 2.2.b, unlike the previous decades when images on the Web were well-annotated by experts like news press or product vender, a large number of today's images are posted by casual users with little or no informative annotations. These two defects of text-based image retrieval prompts the emergence of CBIR as a key technology for image retrieval on the Web, especially in the social network and mobile search environment [24-25].



(a) The Pope inauguration in 2005 (left) and 2013 (right)

(b) Surrounding text of images about Cavenagh Bridge o Singapore River posted in a BBS forum in 1996 (left) anc Facebook in 2010 (right).

Fig. 2.2. The development of the images on the Web: (a) The advances of mobile devices privilege us taking photos anywhere and anytime; (b) However, users are less cooperative to annotate images as before. Images are more difficult to be retrieved by the associated key words.

Although a number of surveys about TBIR exist in the literature [22, 23, 26, 27, 28], the most interesting remark was done by Berrut *et al.* [26]. Their survey found that despite the above difficulties and disadvantages, users of TBIR systems seemed generally satisfied with their system due to high expressive power of the keyword indexing [29].

2.3 Content-Based Image Retrieval

The content-based approach was proposed in the early 1990's and research interests in this area have grown rapidly. This approach represents a promising and cutting-edge technology to address the aforementioned difficulties.

The earliest use of the term content-based image retrieval in the literature seems to have been by Kato [30]. In CBIR, images are indexed by their own visual contents such as color, texture, and shape. Visual contents are extracted from the images as automatically as possible [28]. Thus, CBIR systems have two main advantages over TBIR systems. First, they minimize the human effort. Second, due to reduced people intervention, subjectivity is also reduced. This feature makes CBIR systems more useful in many areas, such as search and browse large image collections.

The fundamental idea of this approach is to generate automatically image descriptions directly from the image content by analyzing the content of the images. Given a query image, a content-based image retrieval system retrieves images from the image database which are similar to the query image. In a typical situation, all the images in the database are processed to extract the selected features that represent the contents of the images. This is usually done automatically once when the images are entered into the database. This process assigns to each image a set of identifying descriptors which will be used by the system later in the matching phase to retrieve relevant images. The descriptors are stored in the database, ideally in a data structure that allows efficient retrieval in the later phase.

Next a query is posted in the matching phase. Using the same procedures that were applied to the image database the features for the query image are extracted. Image retrieval is then performed by a matching engine, which compares the features or the descriptors of the query image with those of the images in the database. The matching mechanism implements the retrieval model adopted according to the selected metric, or similarity measure. The images in the

database are then ranked according to their similarity with the query and the highest ranking images are retrieved. Efficiently describing the visual information of images and measuring the similarity between images described by such pre-computed features are the two important steps in content-based image retrieval.

2.3.1 Applications of CBIR

Detailed applications for CBIR technology can be found in [31]. Some of them are listed below:

Web searching: A large number of digital images are accessed by the Internet users. CBIR systems can help the users to effectively find what they are looking for.

Medical diagnosis: A large number of medical images have been stored by hospitals. Thus, CBIR systems can be used to aid diagnosis by identifying similar past cases.

Journalism and advertising: Articles, photographs, videos of the newspapers, journals or televisions are queried by using CBIR systems.

Military: Databases of all images in military applications; such as remotely sensed data, weapons, aircrafts, automatic target recognition, *etc.*

Intellectual property: Most of the companies have their own trademark image. Whenever a new trademark image is to be registered, it must be compared with existing marks to eliminate duplications.

Crime prevention: After a serious crime, law enforcement agencies search their archives for visual evidence. Such archives include photographs, fingerprints, shoeprints, and *etc.* of the past occasions. Thus, a CBIR system may help those agencies in finding related evidence.

2.3.2 Principle of CBIR

Content-based retrieval uses the contents of images to represent and access the images. Atypical content-based retrieval system is divided into off-line feature extraction and online image retrieval [32]. A conceptual framework for content-based image retrieval is illustrated in Fig. 2.3. In off-line stage, the system automatically extracts visual attributes (color, shape, texture, and spatial information) of each image in the database based on its pixel values and stores them in a different database within the system called a feature database. The feature data (also known as image signature) for each of the visual attributes of each image is very much smaller in size compared to the image data, thus the feature database contains an abstraction (compact form) of the images in the image database. One advantage of a signature over the original pixel values is the significant compression of image representation. However, a more important reason for using the signature is to gain an improved correlation between image representation and visual semantics [33].

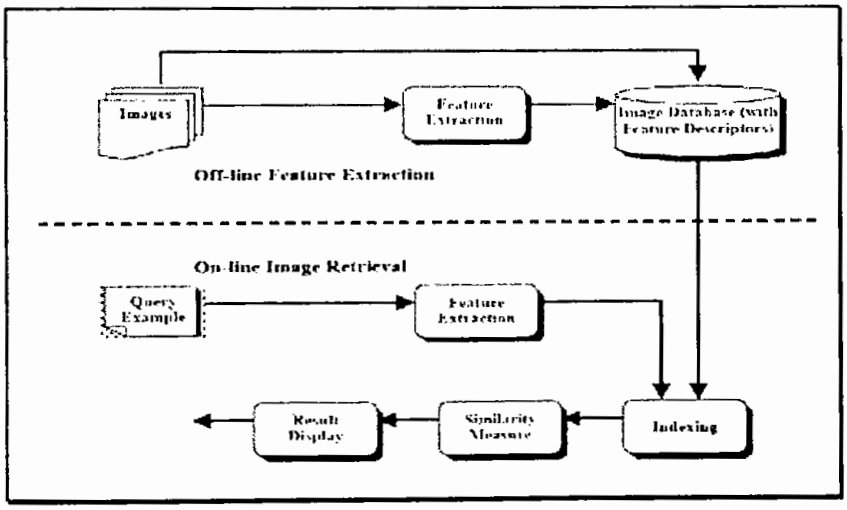


Fig. 2.3 A conceptual framework for context-based image retrieval

In search of desired images in a CBIR system, users can submit a query example. With a feature vector, the system can represent this example. The similarities in terms of distances between the feature vectors of the query image and those in

the feature database are computed and then ranked. To provide an efficient way of searching in the database the retrieval system is usually conducted by using an indexing scheme. Finally, the CBIR system ranks the results and then returns the search results which are most similar to the query example. In the following section, we introduce these fundamental techniques for content-based image retrieval.

2.3.2.1 Visual Content Descriptors

Image content may include both visual and semantic content. Visual content [34] can be very general or domain specific. *General visual content* include color, texture, shape, spatial relationship, etc. *Domain specific visual content*, like human faces, is application dependent and may involve domain knowledge.

A visual content descriptor can be either global or local. A global descriptor uses the visual features of the whole image, whereas a local descriptor uses the visual features of regions or objects to describe the image content. Global features describe the image as a whole to generalize the entire object whereas the local features describe the image patches (key points in the image) of an object. These features include contour representations, shape descriptors, and texture features. Local features represent the texture in an image patch. SIFT, SURF, LBP are some examples of local descriptors.

To obtain the local visual descriptors, an image is often divided into parts first. The simplest way of dividing an image is to use a partition, which cuts the image into tiles of equal size and shape. A simple partition does not generate perceptually meaningful regions but is a way of representing the global features of the image at a finer resolution. A better method is to divide the image into homogenous regions according to some criterion using region segmentation algorithms that have been extensively investigated in computer vision.

Global features are based on configurations of spatial scales and are estimated without invoking segmentation or grouping operations. Shape Matrices, Invariant Moments (Hu, Zerinke), Histogram Oriented Gradients (HOG) and

Co-HOG are some typical examples of global descriptors that provide clues to image retrieval.

Semantic content in content-based image retrieval is obtained either by textual annotation or by complex inference procedures based on visual content. This chapter concentrates on general visual contents descriptions, especially on spatial relationships.

Regions or objects with similar color and texture properties can be easily distinguish by imposing spatial constraints. For instance, regions of blue sky and ocean may have similar color histograms, but their spatial locations in images are different. Therefore, the spatial location of regions (or objects) in an image is very useful for searching images [35-36].

2.3.2.2 Similarity Measures and Indexing Schemes

Similarity between images can be defined by image features such as color, texture, or shape, and on the composition of these features in an image. The discrepancy between the semantic query that the user has in mind and the syntactic features used to describe it makes it hard both for the user to specify the query, and for the system to return the correct images. Until semantic image interpretation can be done automatically, image retrieval systems cannot be expected to find the correct images. Instead, they should strive for a significant reduction in the number of images that the user needs to consider, and provide tools to view these images quickly and efficiently.

CBIR finds visual similarities between a user's query and images in a database rather than exact matching (*i.e.* semantic). Therefore, the retrieval result is not a single image but a number of images that are ranked by their similarities with the user's query. There are several similarity/distance measures have been developed in recent years depending on experiments and distribution of features. The measures (similarity/dissimilarity) affect the performances of an image retrieval system remarkably. Some of the commonly used similarity measures are introduced [37] here, where D (I, J) is denoted as the distance measure

between the user's query I and the image J in the database; and f_i (I) as the number of pixels in bin i of I.

Minkowski-form distance

If each dimension of image feature vector is independent of each other and is of equal importance, the Minkowski-form distance [38] is appropriate for calculating the distance between two images. This distance is defined based on the L_p norm as:

$$d_{L_p}(I,J) = \left(\sum_i |I_i - J_i|^p\right)^{1/p}. \qquad (2.1)$$

Histogram intersection

The Histogram intersection is used by Swain and Ballard [39] to compute the similarity between color images. The intersection of the two histograms of I and J is defined as:

$$d_H(I,J) = 1 - \frac{\sum_i min\,(I_i, J_i)}{\sum_i k_i}. \qquad (2.2)$$

It is attractive because of its ability to handle partial matches when the area of one

histogram (the sum over all the bins) is smaller than that of the other. It is shown in [37] that when the areas of the two histograms are equal, the histogram intersection

is equivalent to the (normalized) L_1 distance.

Mahalanobis Distance

The Mahalanobis distance metric [37-38] is appropriate when each dimension of image

feature vector is dependent of each other and is of different importance. It is defined

as:

$$d_M(I,J) = \sqrt{(F_I - F_J)^T C^{-1}(F_I - F_J)}. \quad (2.3)$$

where F_I and F_J are the features of query and database images, respectively, and **C** is the covariance matrix of the feature vectors.

Kullback-Leibler divergence and Jeffrey divergence

The Kullback-Leibler (K-L) divergence [40] is defined as:

$$d_{KL}(I,J) = \sum_i I_i log \frac{I_i}{J_i}. \quad (2.4)$$

From the point of view of theory of information, the equation (2.4) measures how inefficient on average it would be to code one histogram using the other as the code-book.

Chi-square statistics

The Chi-square statistics is often used in computer vision to compute distances between some bag-of-visual-word representations of images. It is defined as

$$d_{Chi-square}(I,J) = \sum_i \frac{(I_i - m_i)^2}{m_i}, \quad (2.5)$$

where $m_i = \frac{I_i + J_i}{2}$. This quantity measures how unlikely it is that one distribution was drawn from the population represented by the other.

2.3.2.3 Performance Evaluation Measures

Measurement is being described [41] as a way of determining effectiveness among various systems. In our context, the main purpose of performance measurement is a means of comparison between various systems. More specifically, the evaluation measure works as a simulation of an operational setting to which the system may be deployed. There are different types of measures with different priorities in the simulation of the operational setting. The types of performance measures are precision, recall, accuracy, precision-recall, average precision, average recall, mean average precision, mean average recall, and so on. Among them the first

three are basic metrics and the rest of the measures are just derived from the three basic metrics.

The principal quality of a CBIR system is its search effectiveness. This is an assessment of a number of relevant images which are retrieved by a selected query.

In the first era of retrieval system, the system partitioned the collection of images into two sets, those images that were retrieved by a user's query and those that were not. Combining the sets from a test collection of the relevant documents to queries, the contingency table can be created as shown in the following table.

Table 2.1: A 3X3 Contingency Table

Relevant/Actual (A+C)	Non-Relevant (B+D)	Total Observation (A+B+C+D)
True Positive (A)	False Positive (B)	Retrieved/Predicted (A+B)
False Negative (C)	True Negative (D)	Non-Retrieved (C+D)

Precision (P), recall (R) and accuracy (Ac) are defined as:

$$P = \frac{A}{A + B}, \tag{2.6}$$

$$R = \frac{A}{A + C}, \tag{2.7}$$

$$Ac = \frac{A + D}{A + B + C + D}. \tag{2.8}$$

Precision assesses the fraction of retrieved documents that are relevant. In other words, precision assesses the proportion of correct positive classification from the positive predicted (*i.e.* retrieved) class.

Recall assesses the fraction of relevant documents retrieved. In other words, recall assesses the proportion of correct positive classification from the actually positive (*i.e.* relevant) class.

Accuracy assesses the proportion of correct classification (*i.e.* true positive and negative) from the overall documents.

Chapter 3

Markov Stationary Features Revisited

3.1 Introduction

Markov stationary features (MSF) is a general framework that basically extends the primitive histogram based image features characterizing the spatial co-occurrence of histogram patterns by Markov chain models. The Markov chain models can give us a deeper understanding about the content based visual features. These models also provide a powerful tool to investigate theoretical results, such as the uniqueness and convergence of stationary visual features.

Before going to discuss about Markov chain related MSF feature, it will be logical to discuss briefly about Markov chain model. In this chapter initially, we derive time homogeneous model of Markov chains. Then we will switch to time inhomogeneous model of Markov chains. We focus on a special class of Markov chains, namely regular Markov chain over a finite set of states.

3.2 Markov Chain Model

In probability theory, a Markov model is a stochastic model used to model randomly changing systems through time [42]. The simplest Markov model is the Markov chain. It is assumed that future states depend only on the current state, not on the events that occurred before it.

When the outcome of a given experiment in a randomly changing systems affect the outcome of the next experiment, the process is called a Markov chain. The term "Markov chain" refers to the sequence of random variables such a process moves through, with the Markov property defining serial dependence only

between adjacent periods (as in a "chain"). It can thus be used for describing systems that follow a chain of linked events, where what happens next depends only on the current state of the system [42].

Pictorially, shown in Fig. 3.1, a Markov chain can be illustrated by means of a state transition diagram, which is a diagram showing all the states and transition probabilities [43].

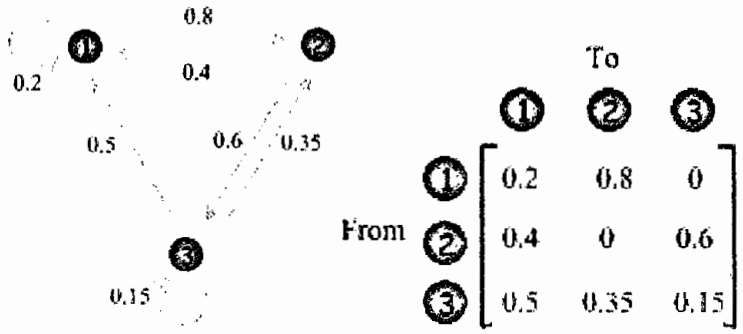


Fig. 3.1. A Markov chain's: a) State transition diagram and b) Transition probability

3.2.1 Characteristics of Markov Chain

The Markov chain process is characterized by a state space, a transition matrix describing the probabilities of particular transitions, and an initial state (or initial distribution) across the state space. By convention, we assume all possible states and transitions have been included in the definition of the process, so there is always a next state, and the process does not terminate [44].

The changes of state of the system are called transitions. The probabilities associated with various state changes are called transition probabilities.

In a mathematical point of view, let us consider, a sequence of random observed process $(X_1, X_2, \ldots)$ related with finite state space, $S = \{s_1, s_2, \ldots, s_K\}$, the Markov chain can be defined formally as,

$Pr[X_{t+1} | X_1 \ldots X_t] = Pr[X_{t+1} | X_t]$, that is, X_{t+1} the state of the system at time t+1 depends only on the state of the system at t. Thus, Markov process must

possess a property that is usually characterized as "memory less", also known as the Markov property: to make the best possible prediction of what happens "tomorrow" (time $n+1$), we only need to consider what happens "today" (time n), as the "past" (times $0, \ldots, n-1$) give no additional useful information.

3.2.1.1 Transition Matrix

If the conditional probabilities are well defined, i.e. if $Pr[X_1 \ldots X_t] > 0$, then, the possible values taken by the random variable X_i form a countable set S called the state space of the chain. The probability, $Pr[X_{t+1} | X_t]$ is known as transition probability and the Markov chain can be modeled as a probability transition matrix, simply transition matrix. For the sake of notational simplicity, say the conditional probability of moving from i (*i.e.* X_t) to j (*i.e.* X_{t+1}) in one time step is $Pr[X_{t+1} | X_t] = Pr(j|i) = p_{ij}$, the transition matrix **P** is given by using p_{ij} as the i^{th} row and j^{th} column element, *e.g.*,

$$P = \begin{pmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,j} & \cdots & P_{1,K} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,j} & \cdots & P_{2,K} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{i,1} & P_{i,2} & \cdots & P_{i,j} & \cdots & P_{i,K} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{K,1} & P_{K,2} & \cdots & P_{K,j} & \cdots & P_{K,K} \end{pmatrix} \tag{3.1}$$

3.2.1.2 Properties of Transition Matrix

Any transition matrix P must satisfy the conditions below. Conversely, any |K|×|K| matrix that satisfies these conditions can be interpreted as a Markov chain transition matrix:

$$p_{ij} \geq 0 \; for \; all \; i, j \in \{1, \ldots, K\} \text{ and} \tag{3.2}$$

$$\sum_{j=1}^{K} p_{ij} = 1 \; for \; all \; i \in \{1, \ldots, K\} \tag{3.3}$$

3.2.1.3 Initial Distribution

Besides the state space and transition matrix, the Markov chain has another characteristic, namely the *initial distribution,* which tells us how the Markov chain starts. This is the probability distribution of the Markov chain at time *0*. The initial distribution is represented as a row vector $\mu^{(0)}$ given by

$$\mu^{(0)} = \left(\mu_1^{(0)}, \mu_2^{(0)}, \ldots, \mu_K^{(0)}\right). \tag{3.4}$$

Since $\mu^{(0)}$represents a probability distribution, we have

$$\sum_{i=1}^{K} \mu_i^{(0)} = 1. \tag{3.5}$$

3.2.2 Chapman-Kolmogorov Equation

When the probability distribution on the state space of a Markov chain is discrete and the Markov chain is homogeneous, the Chapman–Kolmogorov [47] equations can be expressed in terms of matrix multiplication.

For a Markov chain $(X_1, X_2, \ldots)$ with state space $\{s_1, s_2, \ldots, s_K\}$, and once we know the initial distribution $\mu^{(0)}$ and the transition matrix P, we can compute all the distributions $\mu^{(1)}, \mu^{(2)}$ ... at times 1, 2 ... respectively, of the Markov chain, by the following equation:

$$\mu^{(n)} = \mu^{(0)} \mathrm{P}^n, \tag{3.6}$$

where P^n is for the n^{th} power of the matrix P, that means, the result is simply a matter of matrix multiplication. By the induction hypothesis, we have

$$\mu^{(m+1)} = \mu^{(m)} P = \mu^{(0)} P^m P = \mu^{(0)} P^{m+1}, \tag{3.7}$$

what suggest that the state distribution of the Markov chain at time $m+1$ can be obtained by simply multiplication of initial distribution and the $(m+1)^{th}$ power of the transition matrix. And by this way, we have

$$p^{m+n} = p^m p^n. \qquad (3.8)$$

Here, the equation (3.8) is known as *Chapman-Kolmogorov equation.*

3.2.3 Stationary Distribution

Perhaps, the most important feature of Markov chain is its *stationary distribution,* the marginal distribution of all states at any time will always be the stationary distribution. Assuming irreducibility, the stationary distribution is always unique if it exists, and its existence can be implied by positive recurrence of all states. The stationary distribution has the interpretation of the limiting distribution when the chain is ergodic.

A Markov process $\{X_t : t \geq 0\}$ is stationary if the joint distribution of $(X_1, X_2, \ldots X_t)$ is the same as the joint distribution of $(X_{1+m}, X_{2+m}, \ldots X_{t+m})$, where $m \geq 0$. For a Markov chain with finite state space $S = \{s_1, s_2, \ldots, s_K\}$ and transition matrix P, a row vector $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$ is said to be *stationary*, if it satisfies,

(i) $\pi_i \geq 0 \; for \; i = 1, \ldots, K, and \; \sum_{i=1}^{K} \pi_i = 1$, means that π should describe a probability distribution on $\{s_1, s_2, \ldots, s_K\}$, and

(ii) $\pi P = \pi$, meaning that $\sum_{i=1}^{K} \pi_i P_{i,j} = \pi_j, for \; j = 1, \ldots, K$, implies that if the initial distribution $\mu^{(0)} = \pi$, then the state distribution $\mu^{(1)}$ of the chain at time 1 satisfies, $\mu^{(1)} = \mu^{(0)} P = \pi P = \pi$, and, iteratively we see that $\mu^{(n)} = \pi$ for every n.

Thus, according to *Chapman-Kolmogorov equation* (3.8), we have, $\pi = \pi P = \cdots = \pi P^n$. Hence, stationary distribution is also known as invariant distribution of a Markov chain.

To get an invariant distribution, it has to be evidence of the existence, uniqueness and convergence of any Markov chain. Otherwise, the chain will amount to be trivial. Hence, a Markov chain can have a stationary distribution if the Markov chain is nontrivial. For a Markov chain to be nontrivial, the chain should be regular. The transition matrix of a regular Markov chain should satisfy two conditions: *irreducibility* and *aperiodicity.* In other words, any irreducible and aperiodic Markov chain has exactly one stationary distribution. Hence, *irreducibility* and *aperiodicity* are the central importance in our study of stationary distributions.

3.2.3.1 Irreducibility, Periodicity, and Recurrence

A Markov chain $(X_1, X_2, \ldots)$ with state space $S = \{s_1, s_2, \ldots, s_K\}$ and transition matrix P, is said to be *irreducible* if for all s_i, $s_i \in S$ we have that $s_i \leftrightarrow s_j$. In other words, a state s_i communicates with another state s_j, if the chain has the positive probability of ever reaching s_j when we start from s_i. Mathematically, s_i communicates with s_j if there exists an n such that [45].

$$\Pr(X_{m+n} = s_j \mid X_m = s_i) > 0. \tag{3.9}$$

Since the states s_i and s_j communicate each other, they also knows as *recurrent* states. A state s_j being recurrent means it will be visited over and over again, an infinite number of times. For an irreducible Markov chain, if all states are recurrent, then we say that the Markov chain is *positive recurrent*, otherwise the chain is *transient* (*i.e.* not over and over again) or *null recurrent* (*i.e.* never). More specifically, with a finite state space, an irreducible Markov chain is always positive recurrent, that is, each state s_j will be visited infinitely many times, regardless of the initial state $X_1 = s_i$.

However, a *positive* recurrent Markov chain can be either *periodic* or *aperiodic.* A Markov chain to be *periodic* if at least one state is existed where the process will return periodically with a fixed step. The chain (mainly for irreducible chain)

is said to be *aperiodic* if all of its states are aperiodic. For a finite set $\{a_1, a_2, \ldots\}$ of positive integers, we write $gcd\{a_1, a_2, \ldots\}$ for the greatest common divisor of $a_1, a_2, \ldots$. The period $d(s_i)$ of a state $s_i \in S$ is defined as

$$d(s_i) = \gcd\{n \geq 1 : (\mathrm{P}^n)_{i,i} > 0\}. \qquad (3.10)$$

In words, the period of s_i is the greatest common divisor of the set of times that the chain can return (i.e., has positive probability of returning) to s_i, given that we start with $X_0 = s_i$. If $d(s_i) = 1$, then we say that the state s_i is *aperiodic*, means the chain is aperiodic. However, for an irreducible Markov chain it is necessary to have only one aperiodic state to imply all states are aperiodic.

3.2.3.2 Convergence to a Stationary State

For an irreducible Markov chain [44] it is necessary to have only one aperiodic state to imply all states are aperiodic. In a nutshell, a Markov will be *regular* [12] if the chain with *positive recurrent* is irreducible and it will have exactly one stationary distribution. That is, the regular chain has a unique stationary distribution π, and the distribution $\mu^{(n)}$ of the chain at time n approaches π as $n \to \infty$, regardless of the initial distribution $\mu^{(0)}$. For more simplicity, with a regular Markov chain we have

$$\lim_{n\to\infty} P^n = \pi, \qquad (3.11)$$

where all the rows $\vec{\pi}$ of the matrix $\mathbb{\Pi}$ are same and strictly positive which satisfies $\vec{\pi}\mathbf{P} = \mathbf{P}$. In other words, a *regular* Markov chain converges to $\boldsymbol{\pi}$ directly via the equation (3.11), if and only if the chain is *aperiodic*.

For a quantized image, the state space (*i.e.* color space) is definitely finite. As every pixel of an image is surrounded in connection by adjacent pixels (*i.e.* there is no isolated pixel), the Markov chain of this state space is obviously irreducible. However, the important question is that whether the Markov chain is positive recurrent in other words aperiodic. That is, we cannot insure the chain is regular.

If the Markov chain is irregular, the n-th step transition matrix $\mathbf{P}^n$maybe periodic, *i.e.* after every r-th steps, $\mathrm{P}^{n+r} = \mathrm{P}^n$. Thus, the convergence of the equation (3.11) to the stationary distribution $\boldsymbol{\pi}$ does not hold, hence the equation is useless. Therefore, we should explore another way to solve the general case of Markov chain which also includes the irregular one (i.e. not regular) containing the *periodicity* problem.

In computing stationary probability (*i.e.* distribution, one can resort to the fundamental limit theorem of a Markov chain [46] to get rid of the periodicity of the chain. For a transition matrix P, the periodic chain can be handled with following averaging formula

$$\mathbf{A} = \lim_{n\to\infty} \frac{1}{n+1} \sum_{0}^{n} \mathrm{P}^n. \qquad (3.12)$$

Here, $\mathbf{P}^0$ is zero a matrix and each row of the resultant matrix in the equation (3.12) is the required stationary distribution [12].

3.3 Markov Stationary Feature for Homogeneous Scheme

The MSF extends the histogram features by characterizing the spatial co-occurrence of histogram patterns utilizing the Markov chain models. It improves the distinguishable capability of histogram to the extra-bin distinguishable level. The MSF exploits co-occurrence matrix what is constructed directly from image contents by accumulating spatial co-occurrence of colors among the neighborhood pixels.

Suppose, an image I is quantized into K levels, thus the set of histogram bins of the image is $S = \{s_1, s_2, \ldots, s_K\}$. The co-occurrence matrix containing spatial information is defined as $\mathbf{C} = \{c_{ij}\} \epsilon \mathbb{R}^{K\times K}$ with each element

$$c_{ij} = \#(x_1 = s_i, x_2 = s_j \,||x_1 - x_2| = d), \qquad (3.13)$$

where d indicates L_1 distance between two (adjacent if d=1, that is in our case) pixels x_1 and x_2. Each element c_{ij} accumulates the number of co-occurrence for bin s_i and s_j. When the patterns s_i and s_j have large spatial co-occurrence, the possibility that s_i transit to s_j is high. Note here the co-occurrence matrix **C** is a nonnegative symmetric matrix and can be interpreted from a statistical view [12]. Markov chain model is adopted to characterize the spatial relationship of histogram bins, which treats bins as states in that model. The transition matrix, essential component of the chain, is statistically derived from the spatial co-occurrence matrix, defined as $\mathbf{P} = [p_{ij}] \in \mathbb{R}^{K\times K}$, where

$$p_{ij} = \frac{c_{ij}}{\sum_{j=1}^{K} c_{ij}}. \tag{3.14}$$

Here, the p_{ij} is precisely the $(ij)^{th}$ element of the transition matrix **P**.

It can be noted that, every individual image can be represented by a Markov chain what is modeled as an individual transition matrix. Thus, comparing two Markov chains (i.e. two images) corresponds to the comparing two transition matrices. However, the transition matrix is space expensive because it requires M (K^2) spaces for K states i.e. bins. At the same time, time complexity of comparing two transition matrices is O (K^2) what is impractical for a large image database. Thus, it is desirable to build up a compact yet robust feature representation from the transition matrix rather than to become the required feature itself (i.e. the full transition matrix).

It said to be consistent to find the expected (unique) stationary distribution applying the equation (3.12) in which the Markov chain may be regular or irregular. To reduce the time complexity of computing **A**, it is desirable to limit the parameter n (e.g. $n<50$). By definition, it is provided that each row of the matrix **A** will be same for the sake of the uniqueness of stationary distribution. However, in practice, the uniqueness of the rows of matrix **A** may not be the case

due to small n. Therefore, to mitigate the approximation error due to small n, it is a good idea to average the rows of the matrix **A** further as below

$$\pi \approx \frac{1}{K}\sum_{i=1}^{K} \vec{a}_i \text{, where } A = [\vec{a}_1, \ldots, \vec{a}_K]^T. \qquad (3.15)$$

However, in order to be well-defined Markov chain, there must have an initial distribution $\boldsymbol{\pi}(0) = \boldsymbol{\mu}^{(0)}$. It is also noted from the extension theorem by Tulcea [48-49] that a Markov chain is uniquely represented by its transition matrix and initial distribution. In practice, it is reasonable to incorporate the intra-bin transition of an image colors which is essentially the initial distribution of our MSF feature. The initial distribution can be computed directly from the co-occurrence matrix, say **C**, defined as

$$\boldsymbol{\pi}(0) = \frac{c_{ii}^3}{\sum_{i=1}^{K} c_{ii}^3}. \qquad (3.16)$$

After computing the initial distribution and stationary distributions, the required MSF is defined as the combination of the initial distribution $\boldsymbol{\pi}(0)$ (*i.e.* auto-correlogram) and the stationary distribution $\boldsymbol{\pi}$ as

$$\vec{h}_{\text{MSF}} = [\boldsymbol{\pi}(0), \boldsymbol{\pi}]^T. \qquad (3.17)$$

3.4 Non-homogeneous Markov Chain Scheme

With a finite state space, if a Markov chain has only one transition matrix (i.e. having stationary transition probabilities), then the chain is called *time homogeneous* or simply *homogeneous* Markov chain [44]. So far we have discussed the homogeneous Markov chain where the transition matrix is **P**, which was stationary in time. With an initial distribution $\mu^{(0)}$, already we have shown previous section, the distribution of the chain at n time is

$$\mu^{(n)} = \mu^{(0)}P^n \text{ and } \quad P^n = P^{l+m} = P^l P^m, \qquad (3.18)$$

where $n = l + m$.

With a finite state space, if a Markov chain has m different transition matrices $P_1, P_2, \ldots, P_m$ then it is called *time inhomogeneous* or simply *non-homogeneous* Markov chain. For any m, we have

$$\mu^{(m)} = \mu^{(0)} P_1 . P_2 \ldots . P_m . \qquad (3.19)$$

The existence of unique stationary distribution in non-homogeneous Markov chain depends on the two issues namely merging (total variation) and stability of the transition probabilities of the chain. By combining the m transition matrices by the following way, we can merge the total variation as

$$\mathrm{P^m} = \mathrm{P_1 . P_2 \ldots . P_m} . \qquad (3.20)$$

Thus we have, $\mu^{(m)} = \mu^{(0)} \mathrm{P^m}$, where $\mu^{(m)}$ is a state distribution at time m. If all the transition matrices of the Markov chains are irreducible, and the sizes of the state spaces are same, then stability of the chain is not a question at all (in our case). If the sizes of state spaces are different, say q is for $\mathrm{P_1}$ and r is for $\mathrm{P_2}$, then the one way to ensure the stability is

$$\mathrm{P^2} = \mathrm{P_1 . P_2}, \quad \text{where } q < r \text{ and } \mathrm{P_1 . P_2} \neq \mathrm{P_2 . P_1} . \qquad (3.21)$$

That is, before multiplying, the transition matrices should be sorted first in ascending order considering their sizes. Then, the transition matrices of small sizes are to be modified according to the size of the largest matrix. To do so, the additional rows should be added to the smaller matrices without violating the basic properties of transition probability (described in Section 3.2.1.2), where the numerical elements of a row should be distributed uniformly. As a result, the modified matrices turns into rectangle matrices which should be square again, because by definition the transition matrices are always square matrices. To solve the problem, extra columns are to be added to the rectangle matrices with padding zeros. As an example, say two transition matrices of different sizes are

$$\mathrm{P_1} = \begin{pmatrix} .75 & .25 \\ .25 & .75 \end{pmatrix}, \text{and } \mathrm{P_2} = \begin{pmatrix} .4 & .3 & .3 \\ .25 & .25 & .5 \\ .4 & 0 & .6 \end{pmatrix}. \qquad (3.22)$$

Hence, the smaller matrix P_1 should be turned into

$$P_1 = \begin{pmatrix} .75 & .25 & 0 \\ .25 & .75 & 0 \\ .5 & .5 & 0 \end{pmatrix}. \tag{3.23}$$

And then apply the equation (3.21) on the resultant matrices maintaining the sorted order. Thus, we have

$$P^2 = \begin{pmatrix} .75 & .25 & 0 \\ .25 & .75 & 0 \\ .5 & .5 & 0 \end{pmatrix} \begin{pmatrix} .4 & .3 & .3 \\ .25 & .25 & .5 \\ .4 & 0 & .6 \end{pmatrix}. \tag{3.24}$$

After combining all the transition matrices (*i.e.* merging the total variation with stability), the rest of the process of finding stationary distribution π of the non-homogeneous Markov chain is similar to the homogeneous one. Thus, the relation of homogeneous and non-homogeneous chain can be established via

$$\mu^{(n)} = \mu^{(m)} P^{n-m}, \text{where } n \gg m. \tag{3.25}$$

Chapter 4

Image Retrieval Using Multi-channel Nonhomogeneous MSF

4.1 Introduction

In the original MSF scheme, pixels of an image (with K level quantization) in all directions (*i.e.* 8-neighborhood) are counted to generate a single dimensional co-occurrence matrix (K-*by*-K in size). The two components (*i.e.* initial and stationary distribution) of the MSF are then derived based on the co-occurrence matrix. Practically, the MSF shows better retrieval performance on a large scale image database ranging from homogeneity to limited heterogeneity of image storages. However, when the underlying image database contains moderate or higher degree heterogeneous images (*i.e.* histogram-level distinguishable, inter-bin distinguishable, extra-bin distinguishable and histogram undistinguishable images with different size, orientation, color and light condition) the single dimensional MSF method does not show the desirable performance. This is because, with an original MSF, it is possible to capture only limited (*i.e.* up to 2^{nd}-order) spatial information that is not sufficient to discriminate the images accurately. To overcome the difficulties, we have introduced here the time inhomogeneous Markov chain model that can capture higher-order spatial information. We have demonstrated the model with two different schemes of image descriptors discussed in this current and next chapters.

In this chapter, we have discussed the implementation of the proposed extended Markov chain model for content based image retrieval. We have applied the method to construct multi-channel nonhomogeneous (*i.e.* time inhomogeneous) MSF (MCN-MSF) feature for the image retrieval.

A channel in a digital image is the grayscale image of the same size as a color image, made of just one of these primary colors. In a RGB color model, for

example, each pixel of an image has a red, green and blue components which are referred to as channels. Note that, a grayscale image is just a single-channel image. Among several color models, we have used the HSV color model in which the three components H, S, and V represent three different channels. The transition matrices are computed for each of the channels. These three matrices are combined using the time inhomogeneous Markov chain model. Here, the sizes of the matrices are unique. The demonstration of the proposed method with different sizes of transition matrices will be discussed in the next chapter.

4.2 HSV Color Space

Among several color models, perhaps HSV is the most prominent color space for low level image processing. For instance, using HSV color space, the CBIR system results in better agreement with color perception than using RGB. This is because RGB is not perceptually uniform; thereby the small perceptual differences between some pairs of colors are equivalent to larger perceptual differences in other pairs. Therefore, it is fair to convert the RGB images in the database to corresponding HSV images before any computation.

In HSV model, the luminance (*i.e.* brightness) component of a color pixel is separated from its chrominance components (*i.e.* Hue and Saturation). Hue represents the attribute that describes pure color. "Hue" slightly differs from "color" because a color can have saturation or brightness as well as hue. Hue is perceived with the incident of sufficient illumination of light containing single wavelength. Saturation is the colorfulness of a color (hue) relative to its own brightness. A pure color is fully saturated which is diluted by mixing of white light. Brightness is an attribute of visual perception in which a source appears to reflecting light. In other words, brightness is the perception elicited by the luminance of a visual target. Actually, the colorfulness is a function of brightness and saturation. More specifically, the brightness is the perceived intensity of a light whereas saturation is the relative color intensity of a light. In our literature,

the term hue, saturation and brightness are treated separately to compute the individual transition matrices.

4.3 The Proposed Algorithm

4.3.1 MCN-MSF Feature Extraction

The existing schemes related to the spatial co-occurrence of information are based on color (*e.g.* hue) [50-52], intensity (*e.g.* brightness or saturation) [53] or both [54-55]. In the color and intensity based method, the color and intensity of an image are treated alone or at most separately, and the individual features are then concatenated. Thereby the size of the resultant feature space is usually grown geometrically with very limited spatial co-occurrence information. While the feature spaces are giant, the local variations of some color components (e.g. S and V for HSV) of an image are still ignored. This paper proposes MCN-MSF technique which is essentially a generalization of color MSF (*i.e.* regarding hue) and grayscale MSF (*i.e.* regarding brightness and saturation) in which the contribution of both color and intensity is recognized. This integrated approach extracts the features capturing the spatial co-occurrence of histogram patterns of color as well as intensity around every pixel of an image using the HSV color space. To do so, some of the useful properties of HSV color space (discussed earlier) are utilized. The main advantage of our proposed technique, in contrast to the other schemes, is that, the color and intensity variations are contained in a single resultant feature. Thereby, the compactness of the feature space dimension can be assured without compromising robustness and efficiency.

The fig. 4.1 shows the block diagram of the proposed method. In our algorithm, at first the RGB image is converted into HSV image. The individual H, S, V images (channels) are then quantized into K levels, where H represents color (*i.e.* hue) information, S represents saturation of the color and V represent intensity of the color. Three individual co-occurrence matrices (that is, $\mathbf{C_1}, \mathbf{C_2}, \mathit{and}\ \mathbf{C_3}$,respectively) are counted from the corresponding H, S, V images

using the equation (3.14) among the neighborhood pixels (*i.e.*, $d = 1$). Equation (3.14) is then used to compute three corresponding transition matrices $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$. Using the three transition matrices, the unique composite transition matrix $\mathbf{P}^3$ is then calculated exploiting the idea of non-homogeneous Markov chain model explained in Chapter 3. This composite transition matrix $\mathbf{P}^3$is similar to the targeted transition matrix $\mathbf{P}$ of homogeneous Markov chain for finding the required unique stationary distribution. Finally, equation (3.15) is applied to compute the desired stationary distribution $\boldsymbol{\pi}$.

The resultant vector $\boldsymbol{\pi}$ is the required stationary distribution that encodes the extra-bin transitions of three channels of the image that is the first part of our concerned MCN-MSF feature. The other part of the MCN-MSF is initial distribution $\boldsymbol{\pi}(0)$ which is computed by the equation (3.16), as shown in the Fig. 4.1.

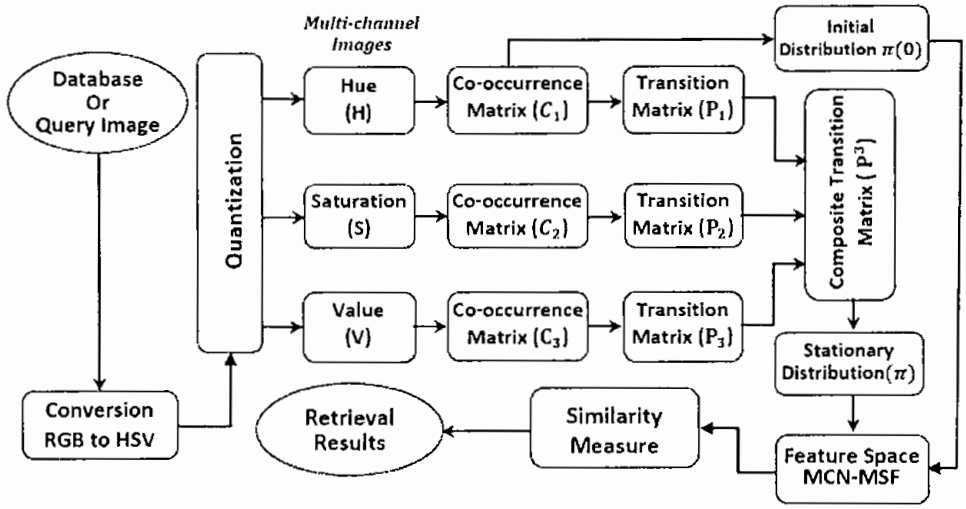


Fig. 4.1 Block diagram for MCN-MSF Method

4.3.2 Similarity Measurement and Image Retrieval

We assume that no textual captions or other manual annotations of the images are given. Therefore, the proper representation of the visual features of an image will be the proper description of the image content, such as MCN-MSF. In order to find images in an image retrieval system that are visually similar to the given

query a measure is required. The measure determines how similar or dissimilar [37] the different images are from the query. For matching between the MCN-MSF features of the query and database images the well-known distance/similarity measures such as Euclidean distance, Mahalanobis distance, *Chi* square distance etc. can be used. In the previous section we have seen that, each MCN-MSF feature contains two K dimensional vectors where each vector is treated as a histogram due to the non-negative nature of its elements. Hence we apply *Chi* square distance (equation (2.5)) in our algorithm. For two histograms $\mathbf{h}_q$ and $\mathbf{h}_d$ of the images I_q and I_d, respectively, the *chi* square distance can be defined as:

$$\mathrm{D}(I_q, I_d) = \frac{1}{2}\sum_{k=1}^{K}\frac{[\mathbf{h}_q(k) - \mathbf{h}_D(k)]^2}{\mathbf{h}_q(k) + \mathbf{h}_D(k)} \qquad (4.1)$$

The above distance formula is used to compute the differences between the query image and database images for both the initial and stationary distribution individually. These two distances are combined to find the aggregated distance. The aggregated distance is computed using weighted summation of the two distances. The aggregated distances for all the database images are then stored in an array with ascending order. Theoretically, the image that corresponds to the first element of the array indicates the most similar and highest ranked image. Similarly image that corresponds the last element of the array indicates the least similar and lowest ranked image of the database. Then required number of images from the sorted array is selected (top matched) for image retrieval. The performance of the proposed method is evaluated using the following performance metrics.

4.4 Performance Metrics

The general performance of histogram, Color Auto Correlogram (CAC), original MSF with our proposed MCN-MSF method regarding image retrieval system is

measured in terms of average, mean average and total average of recall, precision and accuracy which are described in below.

4.4.1 Average Recall

Recall for the first n retrieved images is defined as

$$R(I_q, n, G) = \frac{1}{N_G}\sum_{i=1}^{n}\left[\gamma\left(f(I_i), f(I_q)\right) \mid \text{Rank}(I_q, I_i) \leq n\right], \tag{4.2}$$

where,

$$\gamma\left(f(I_i), f(I_q)\right) = \begin{cases} 1 & \text{if } f(I_i) = f(I_q), \\ 0 & \text{Otherwise} \end{cases}$$

and $f(I)$ stands for the category of image I, Rank(I_q, I_i) returns the rank of image I_i according to the similarity metric for query image I_q, and N_G indicates the number of relevant images of a G category in the database DB.

We defined *average recall* for the k-th category of the image database are given by

$$R_{av}^{k}(n) = \frac{1}{N_{kG}}\sum_{i=1}^{N_{kG}} R(I_i, n, k)|_{n<|DB|}. \tag{4.3}$$

where, N_{kG} indicates the number of relevant images of G category in the database DB.

Mean Average Recall (MAR) for M randomly selected queries from database image, is defined by

$$MAR(n) = \frac{1}{|M|}\sum_{i=1}^{|M|} R(I_i, n, k_i)|_{n<|DB|}. \tag{4.4}$$

Total average recall (TAR) for the entire image database is defined by

$$TAR(n) = \frac{1}{|DB|}\sum_{i=1}^{|DB|} R(I_i, n, k_i)|_{n<|DB|}, \tag{4.5}$$

where k_i indicates the kth category where I_i image belongs to.

4.4.2 Average Precision

Suppose, n indicates the number of retrieved (top matched) images by a system then *Precision* is defined as

$$P(I_q, n) = \frac{1}{n}\sum_{i=1}^{n}\left[\gamma\left(f(I_i), f(I_q)\right) \mid \text{Rank}(I_q, I_i) \leq n\right]. \quad (4.6)$$

It should be noted from the equations (recall and precision), the number of retrieved documents should be different with the number of relevant documents. Again, it should be careful that the number of retrieved documents must be less than number of relevant documents, otherwise false positive results would be accumulated by the growing number of retrieved documents.

We defined *average precision* for the *k*th category of the image database are given by

$$P_{av}^{k}(n) = \frac{1}{N_{kG}}\sum_{i=1}^{N_{kG}} P(I_i, n)|_{n<|DB|}. \quad (4.7)$$

Where, N_{kG} indicates the number of relevant images of a G category in the database DB.

Mean Average Precision (MAP) for M randomly selected queries from database image, is defined by

$$MAP(n) = \frac{1}{|M|}\sum_{i=1}^{|M|} P(I_i, n)|_{n<|DB|}. \quad (4.8)$$

Total average precision (TAP) for the entire image database is defined by

$$TAP(n) = \frac{1}{|DB|}\sum_{i=1}^{|DB|} P(I_i, n)|_{n<|DB|}. \quad (4.9)$$

4.4.3 Average Accuracies

Accuracy for the first n retrieved images is defined as

$$A(I_q,n)=\frac{1}{|DB|}\sum_{i=1}^{|DB|}\left[\left\{\gamma\left(f(I_i),f(I_q)\right)\mid \text{Rank}(I_q,I_i)\le n\right\}\right.$$
$$\left.\cup\left\{\delta\left(f(I_i),f(I_q)\right)\mid \text{Rank}(I_q,I_i)>n\right\}\right], \qquad (4.10)$$

where,

$$\delta\left(f(I_i),f(I_q)\right)=\begin{cases}1 & \text{if } f(I_i)\neq f(I_q)\\ 0 & \text{Otherwise}\end{cases}.$$

Mean Average Accuracy (MAA) for M randomly selected queries from the database is defined as

$$MAA(n)=\frac{1}{|M|}\sum_{i=1}^{|M|}A(I_i,n)|_{n<|DB|}. \qquad (4.11)$$

Total average accuracy (TAA) for the entire image database is defined by

$$TAA(n)=\frac{1}{|DB|}\sum_{i=1}^{|DB|}A(I_i,n)|_{n<|DB|}. \qquad (4.12)$$

4.5 Experiment and Discussion

The Multi-channel Nonhomogeneous Markov Stationary Feature (MCN-MSF) can be exploited in the various fields of image processing and pattern recognition problems. Here, content-based image retrieval application is regarded as a field of study to evaluate the effectiveness of the MCN-MSF features. The whole evaluation process followed the diagram illustrated in Fig. 4.1.

In this paper, we use two different worldwide recognized databases namely: WANG1000 (also known as CORELI1000) and COREL10800 for our

experiments. The databases contain large number of images in various conditions and types ranging from animals and outdoor sports to natural images. Because of the heterogeneity of size, orientation, color and lighting, it is axiomatic that the Corel databases meet all the requirements to evaluate the performance of any CBIR system. However, for each database, we evaluate the performance applying the existing algorithms and our proposed two algorithms, MCN-MSF and MRN-MSF. The performances are compared in terms of different metrics including average, mean average and total average of precision, recall and accuracy.

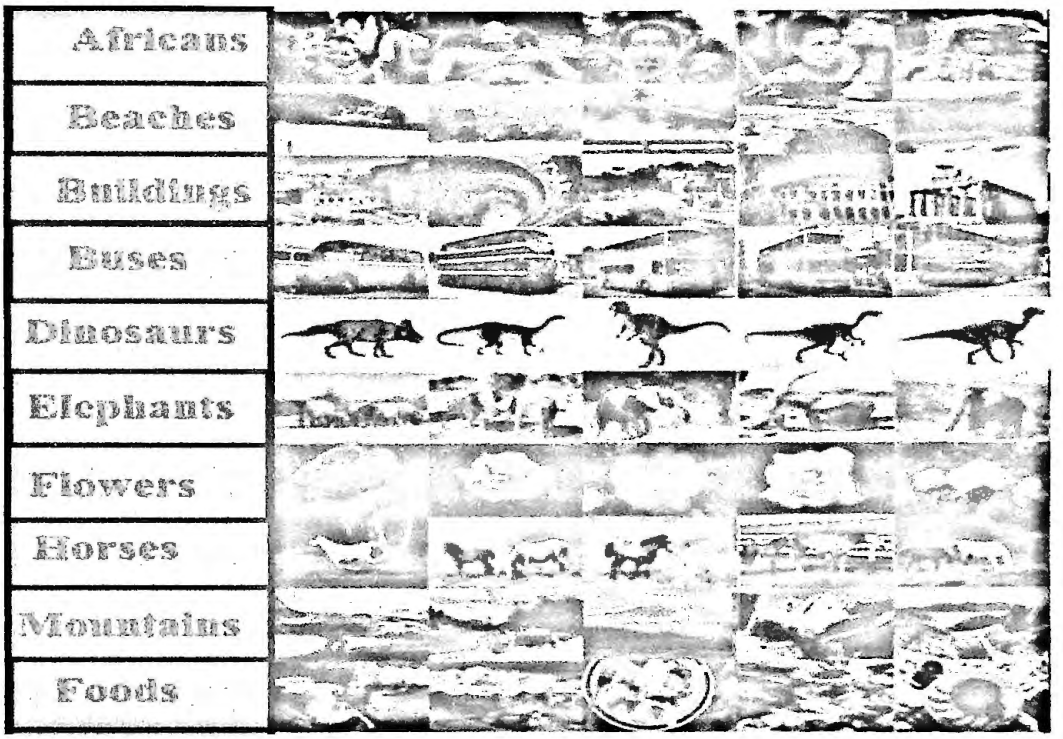


Fig. 4.2 Sample images of WANG1000 database for different categories: African, Beach, Building, Bus, Dinosaur, Elephant, Flower, Horse, Mountain, and Food.

4.5.1 Database I: WANG1000db

For the WANG1000 [56], the images of the database are pre-classified into 10 different categories each with 100 in size. These images are collected from ten different domains, namely, *Africans*, *beaches*, *buildings*, *buses*, *dinosaurs*,

elephants, *flowers*, *horses*, *mountains*, and *food*. The Fig. 4.2 shows the some sample images of each categories of the database.

4.5.1.1 Single Query Image Retrieval

In this work, at first a single image is used as a query image. From the database image no. 21 is used as query image which is the image of an African category. The image retrieval result with our proposed method is shown in Fig. 4.3, where the top left image is a query image and the rest are the retrieved top 19 matches of images. The retrieved images are arranged from left to right and then from top to bottom. From the figure we see that the first retrieved image is essentially the same as the query image. The next few images are similar to the query image. When the image rank is decreased the similarity is also decreased. Thus, the last few images are least similar. From the figure we see some results that are not relevant to the African category.

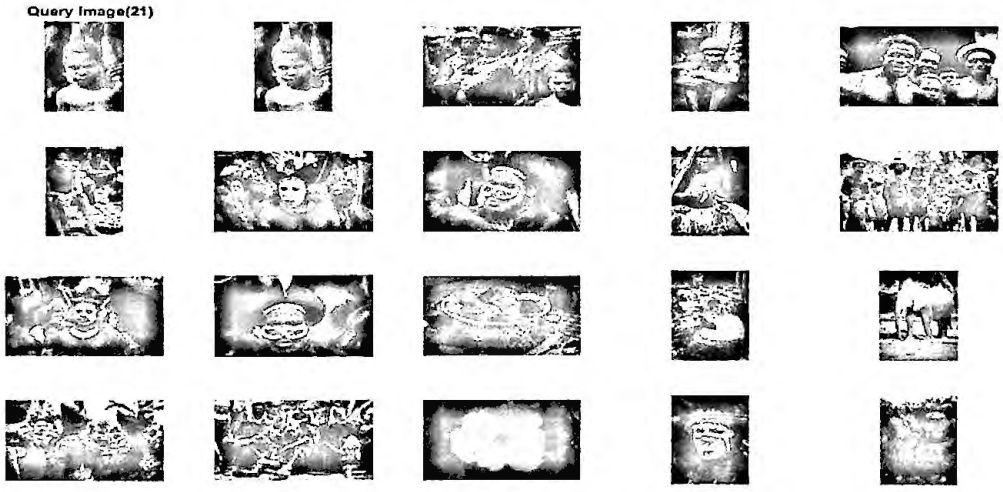


Fig. 4.3 Retrieval result of the proposed MCN-MSF method obtained for the query image-21 in WANG1000db.

In order to show the comparative results, we find recall-precision graph. We find the precisions using equation (4.6) over different recalls using equation (4.2) for Histogram, CAC, original MSF and our proposed method (MCN-MSF). The Fig. 4.4 shows the comparative results of for query image no. 21. The figure indicates that for recall 0.01 to 0.02 the precision is 100% for all the methods. From recall 0.06 the proposed method outperforms the existing methods. We also see the significant improvement of performance with the proposed method up to recall 0.7. After that the performance is almost similar to the other methods.

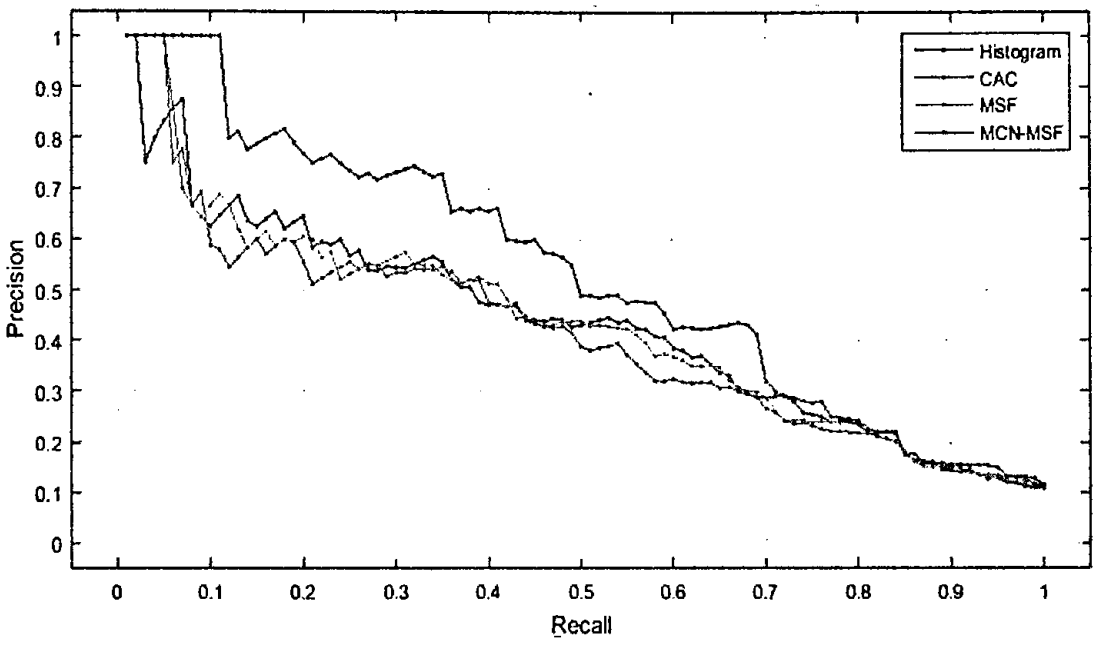


Fig. 4.4 Recall-precision curves of Histogram, CAC, original MSF and proposed MCN-MSF methods for query image-21 in WANG1000 database.

4.5.1.2 Category-wise Query Image Retrievals

To show the category-wise performances with large number of query images we select images from all the 10 categories. 20 images are selected randomly from each category. For all the categories and retrievals we retrieve 50 images to estimate the performances. Fig. 4.5.a, and Fig. 4.5.b show the category-wise recall and precision using the equations (4.3) and (4.7) respectively, for 50 top matches (*i.e.*, n=50). Beach images are background-dominated which are very similar to other background-dominated images in the database. Thus, beach

category shows the least performances for all the methods as shown in the figures. One the other hand, horse category shows the best performances because the images of the category are object-dominated.

In all the figures the MCN-MSF method outperforms the existing methods in terms of average recall and precision for all the categories except category 4 (Buses). The histogram performance of category 4 is better than the MCN-MSF method. Bus images are color-dominated. Usually, colors in the color-dominated images dominates other visual features such as texture, shape. Note that in texture-dominated images spatial information is more influential. Thus, histogram yields the better results for the color-dominated images as it ignores spatial information of images.

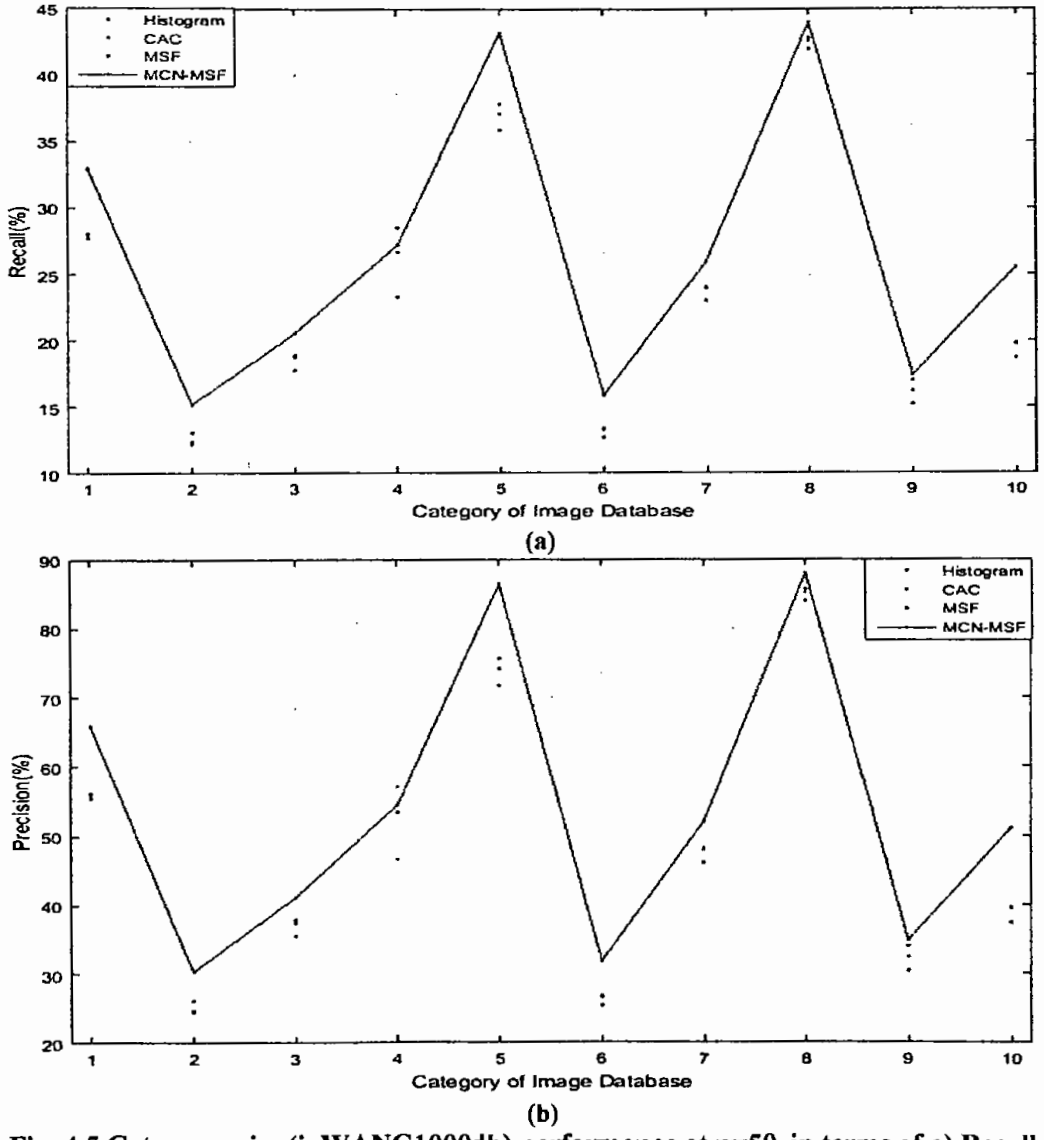


Fig. 4.5 Category-wise (inWANG1000db) performance at n=50, in terms of a) Recall and b) Precision for Histogram, CAC, MSF and MCN-MSF.

4.5.1.3 M-Randomly Selected Query Image Retrievals

To show the extensive performances of proposed method we choose 50 random query images from the database. We show the performance results using the equations (4.4),

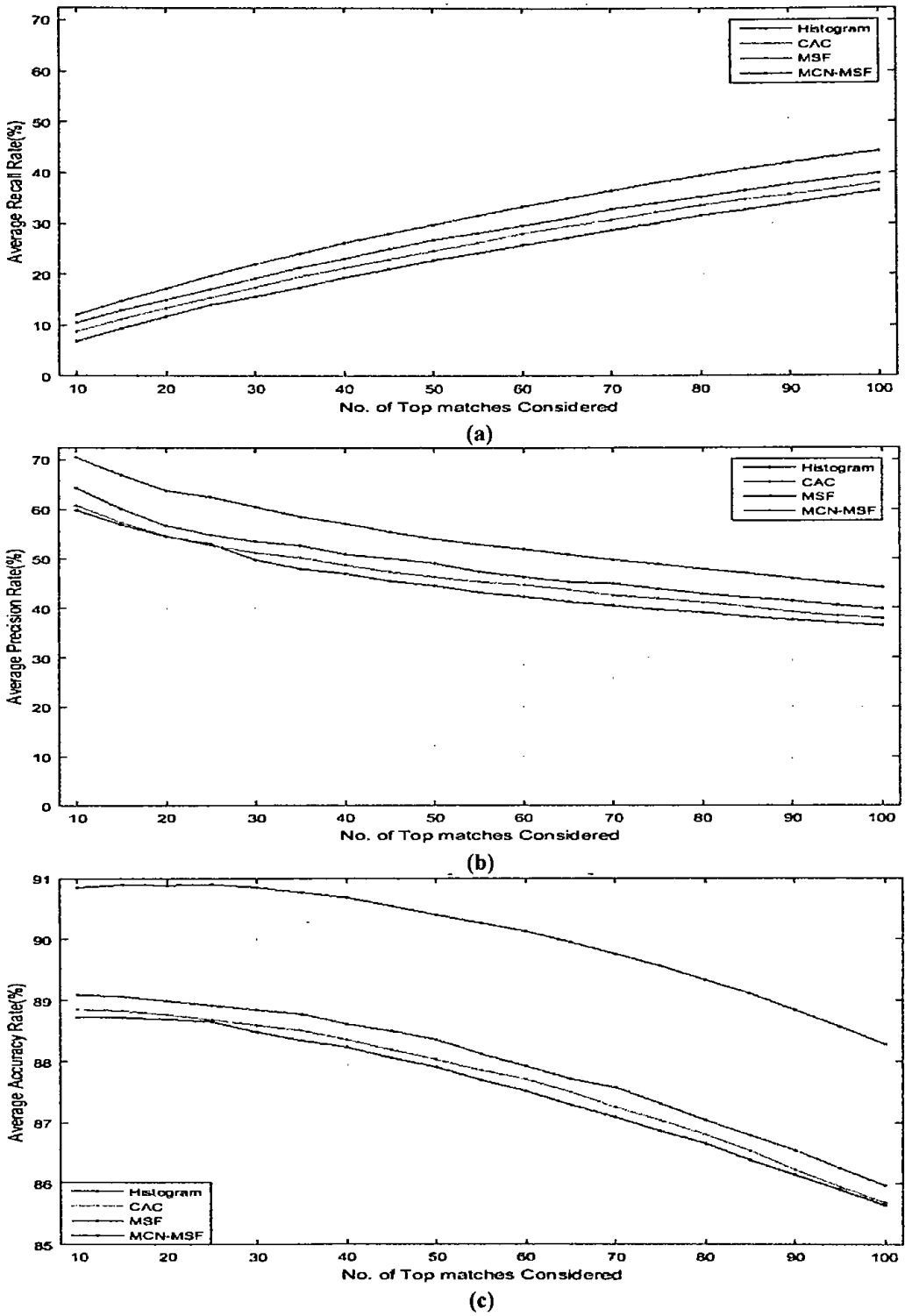


Fig. 4.6 Mean average a) Recall, b) Precision and c) Accuracy curves of Histogram, CAC, original MSF and proposed MCN-MSF methods for randomly 50 query images in WANG1000 database.

(4.8) and (4.11) varying the number of top matches from 10 to 100 with the interval of 10. For a particular top matches value we compute the average result of 50 queries. The mean average recall, precision and accuracy of the 50 query images are shown in the Fig. 4.6.a, Fig. 4.6.b and Fig. 4.6.c, respectively. In each case, we compare the results of our proposed method with Histogram, CAC and MSF. For all the cases, the proposed method shows the significant improvements than the existing methods as shown in the figures. For the case of MAR the improvement of the proposed method is gradually increasing over the top matches as shown in Fig. 4.6.a. For the case of MAA the improvement of the proposed method is more promising and constant progressing over the number of top matches as shown in Fig. 4.6.c.

4.5.1.4 Entire Database Search Results

To show overall performance of the entire image database, we select every image of the database as a query image. Similar to the previous subsection, we show the performance results using equations (4.5), (4.9) and (4.12) varying the number of top matches from 10 to 100. For a particular top matches value we compute the average result for all the queries. The total average recall, precision and accuracy of all the query images are shown in the Table 4.1.a, Table 4.1.b and Table 4.1.c, respectively. The last rows of all the tables show grand averages of the corresponding performance measures for all the methods including Histogram, CAC, MSF and MCN-MSF. For all the measures the proposed method shows better performance as shown in the tables.

From the tables, we see that the grand averages of recall, precision and accuracy of the histogram are the lowest among all the methods. The grand averages of the CAC and MSF show the better performances than the histogram. But the performances of these three methods are very close to each other. Only the MCN-MSF method shows the promising difference from other methods.

Table 4.1: Comparison of total averages between the MCN-MSF and the existing methods in terms of a) Recall, b) Precision and c) Accuracy for WANG1000db.

Number of Top	Histogram	CAC	MSF	MCN-MSF
10	9.34	11.18	12.81	**15.70**
20	13.76	15.57	17.31	**20.81**
30	17.77	19.56	21.37	**25.37**
40	21.51	23.19	25.06	**29.46**
50	24.82	26.49	28.34	**33.15**
60	27.87	29.51	31.46	**36.41**
70	30.75	32.34	34.31	**39.48**
80	33.40	34.93	36.96	**42.21**
90	35.87	37.36	39.36	**44.69**
100	38.35	39.78	41.77	**47.11**
Grand Average	**25.34**	**26.99**	**28.87**	**33.44**

(a) Total average Recall

Number of Top	Histogram	CAC	MSF	MCN-MSF
10	56.59	57.54	59.87	**67.85**
20	52.04	53.28	55.74	**63.74**
30	48.90	50.31	52.69	**60.40**
40	46.57	47.87	50.18	**57.53**
50	44.31	45.71	47.84	**54.95**
60	42.33	43.78	45.97	**52.52**
70	40.67	42.12	44.25	**50.47**
80	39.11	40.56	42.69	**48.51**
90	37.71	39.16	41.20	**46.70**
100	36.31	37.77	39.71	**45.03**
Grand average	**44.45**	**45.81**	**48.01**	**54.77**

(b) Total average Precision

Number of Top	Histogram	CAC	MSF	MCN-MSF
10	88.72	88.84	89.06	**90.99**
20	88.60	88.71	88.96	**91.01**
30	88.40	88.51	88.77	**90.92**
40	88.15	88.24	88.51	**90.74**
50	87.81	87.90	88.17	**90.48**
60	87.42	87.50	87.79	**90.13**
70	87.00	87.07	87.36	**89.75**
80	86.53	86.59	86.89	**89.29**
90	86.02	86.07	86.37	**88.79**
100	85.52	85.56	85.85	**88.26**
Grand Average	**87.42**	**87.50**	**87.77**	**90.04**

(c) Total average Accuracy

4.5.2 Database II: COREL10800db

For the COREL10800 [57], the number of images is 10800 which are pre-classified into 80 concept groups, e.g., autumn, aviation, bonsai, castle, cloud, dog, elephant, iceberg, primates, ship, stalactite, steam-engine, tiger, train, waterfall and so on. In fact, it was collected and reorganized from the Corel Photo Gallery, because 1) many images with similar concepts were not in the same group and 2) some images with different semantic contents were in the same group in the original database. In the reorganized database, each group includes more than 100 images and the images in the group are category-homogeneous.

4.5.2.1 Single Query Image Retrieval

In this work, at first a single image is used as a query image. From the database image no. 2287 is used as query image which is the image of bus category. The image retrieval result with our proposed method is shown in Fig. 4.7, where the top left image is a query image and the rest are the retrieved top 29 matches. The retrieved images are arranged from left to right and then from top to bottom. From the figure we see that the first retrieved image is essentially the same as the query image. The next few images are similar to the query image as like the Database I result. When the image rank is decreased the similarity is also decreased. Thus, the last few images are hardly similar. From the figure we see some results that are not relevant to the bus category.

In order to show the comparative results, we find recall-precision graph. We compute the precisions over different recalls as like the previous section for Histogram, CAC, original MSF and our proposed method (MCN-MSF). The fig. 4.8 shows the comparative results of the methods of for query image no. 2287. The figure indicates that for recall 0.01 and 0.02 the precision is 100% for all the methods. From recall 0.04 the proposed method outperforms the existing methods. At recall 0.1 the precision of the method is 0.84 which is almost double to the other methods. We also see the significant improvement of performance

with the proposed method is up to recall 0.63. After that the performance is almost similar to the other methods.

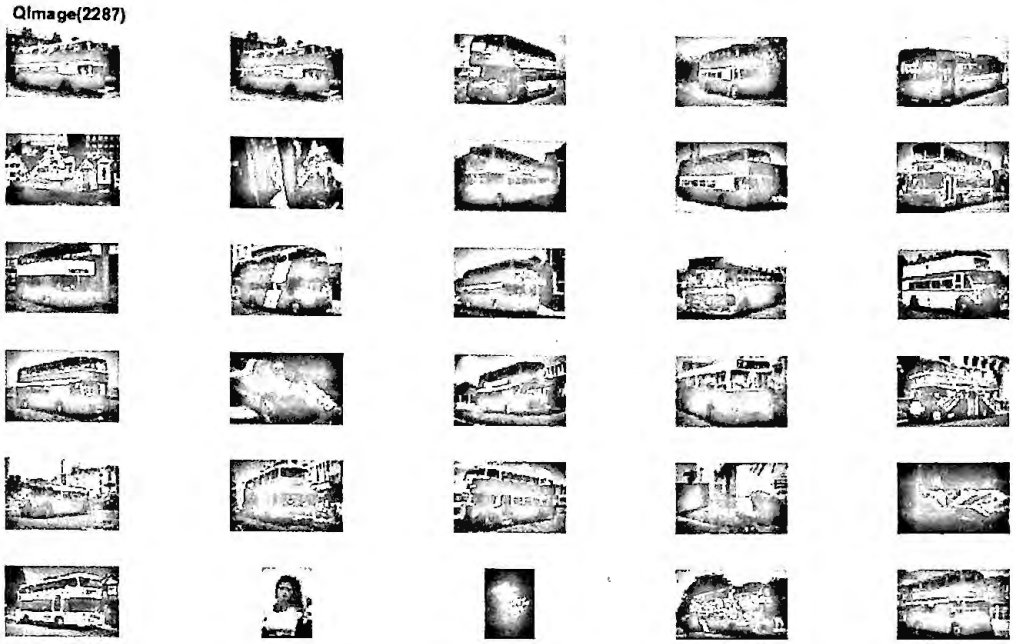


Fig. 4.7 Retrieval result of the proposed MCN-MSF method obtained for the query image-2287 in COREL10800db.

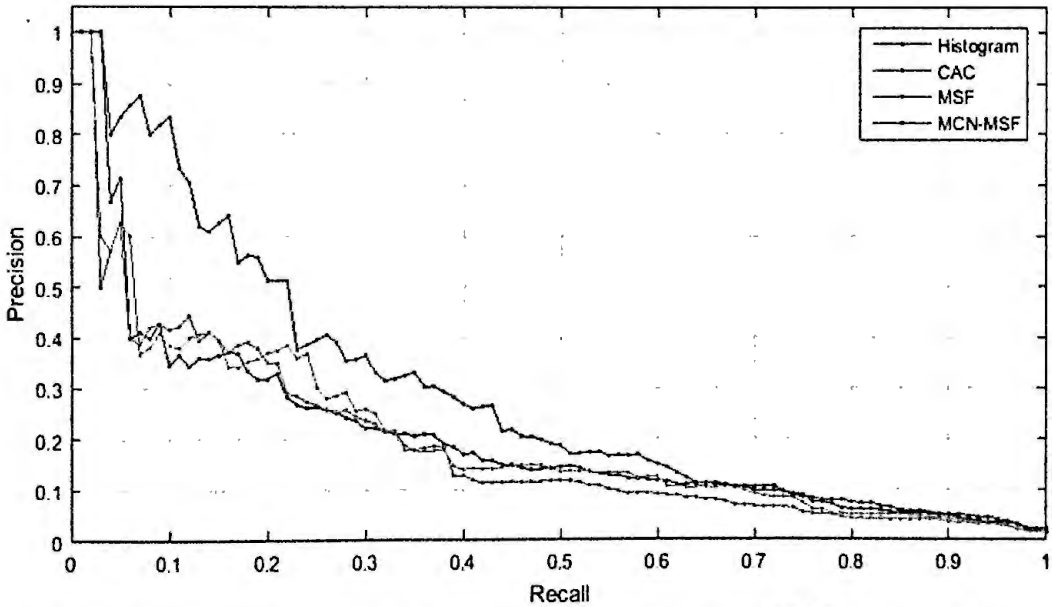


Fig. 4.8 Recall-precision curves of Histogram, CAC, original MSF and proposed MCN-MSF methods for query image-2287 in COREL10800 database.

4.5.2.2 Category-wise Query Image Retrievals

To show the category-wise performances with large number of query images we select images from all the 80 categories. 50 images are selected randomly for query from each category. For all the categories and retrievals we retrieve 150 images to estimate the performances. Fig. 4.9.a and Fig. 4.9.b show the category-wise recall and precision, respectively for 150 top matches (*i.e.*, n=150) using the same equations as previous sections. Category 14 (aviation) images are background-dominated and the foreground

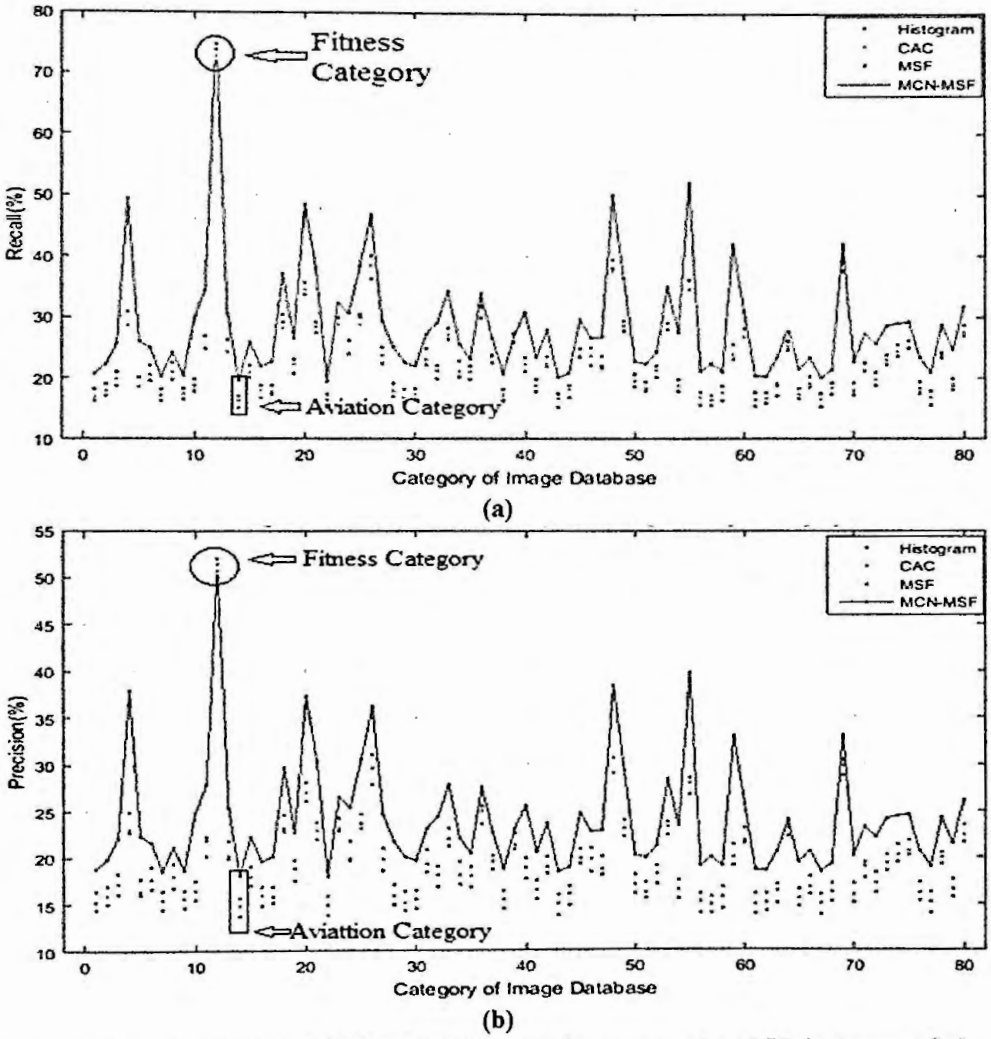


Fig. 4.9 Category-wise (in COREL10800db) performance at n=150, in terms of a) Recall and b) Precision for Histogram, CAC, MSF and MCN-MSF.

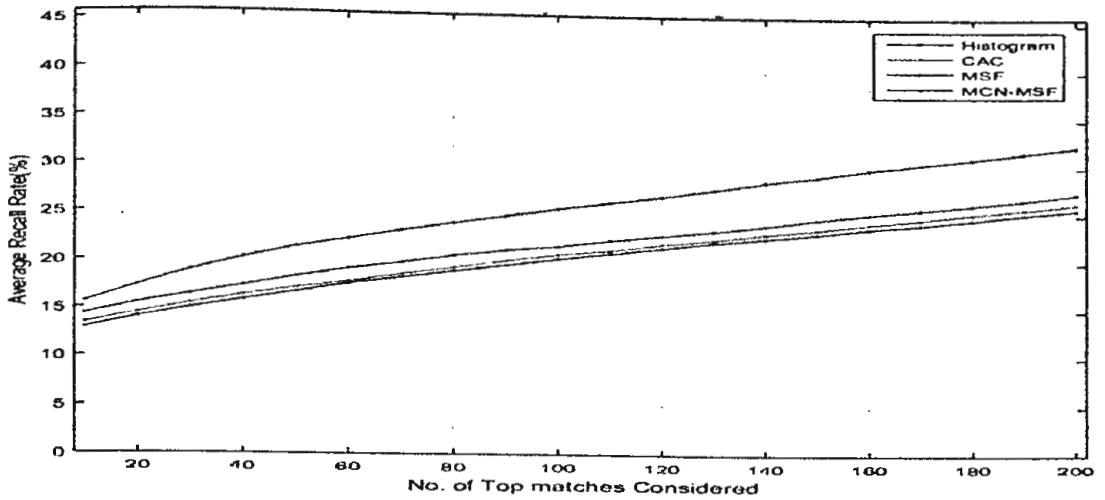
(*i.e.* the expected object) of those images are coarse in size, shape and color. Thus they are more similar to other background-dominated images in the

database than that of the category. As a result, aviation category shows the least performances for all the methods as shown in the figures. One the other hand, fitness category (*i.e.* category no. 12) shows the best performances because the images of the category are object-dominated.

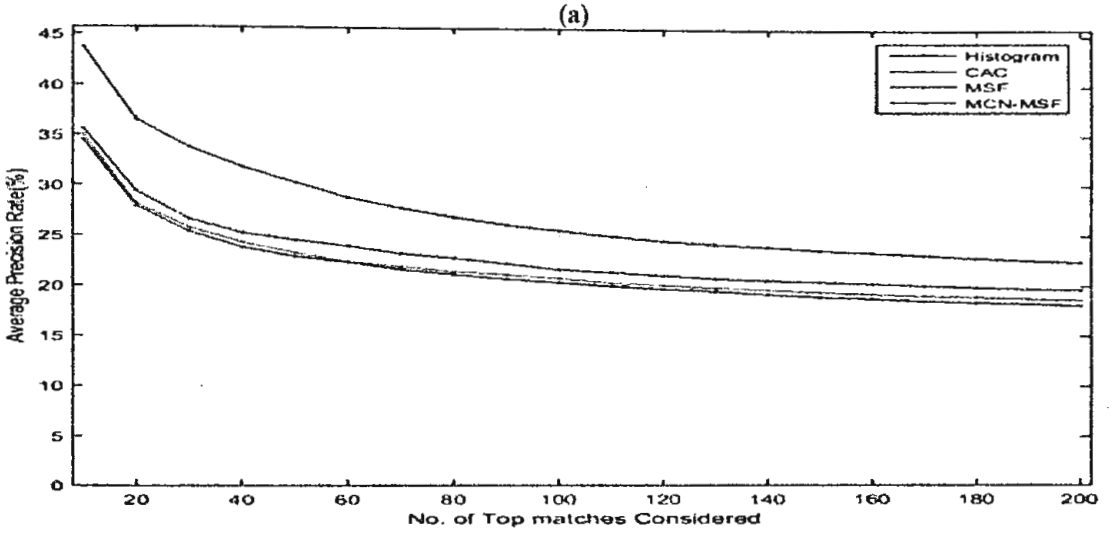
In all the figures the MCN-MSF method outperforms the existing methods in terms of average recall, precision and accuracy for all the categories except category 12 (Fitness). The histogram performance of category 12 is better than the MCN-MSF method. Fitness images are the most color-dominated in the database. Not only that there are the fewest colors in this category images compare to other category. Hence, the images contain very limited spatial information. Thus, histogram yields the better results for the color-dominated images as it ignores spatial information of images.

4.5.2.3 M-Randomly Selected Query Image Retrievals

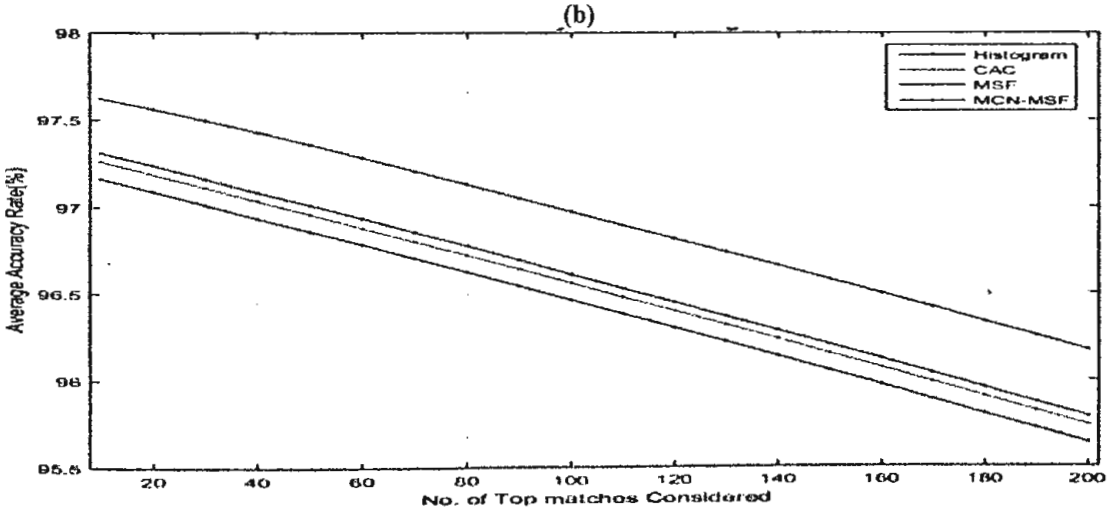
To show the extensive performances of proposed method we choose 100 random query images from the database. We show the performance results varying the number of top matches from 10 to 200 with the interval of 10. For a particular top matches value we compute the average result of 00 queries. The mean average recall, precision and accuracy of the 100 query images applying the equations of the Section 4.4 are shown in the Fig. 4.10.a, Fig. 4.10.b and Fig. 4.10.c, respectively. In each case, we compare the results of our proposed method with Histogram, CAC and MSF. For all the cases, the proposed method shows the significant improvements than the existing methods as shown in the figures. For the case of MAR the improvement of the proposed method is gradually increasing over the top matches as shown in Fig. 4.10.a. For the case of MAA the improvement of the proposed method is promising and constant progressing over the number of top matches as shown in Fig. 4.10.c.



(a)



(b)



(c)

Fig. 4.10 Mean average a) Recall, b) Precision and c) Accuracy curves of Histogram, CAC, original MSF and proposed MCN-MSF methods for randomly 100 query images in COREL10800 database.

4.5.2.4 Entire Database Search Results

To show overall performance of the entire image database, we select every image of the database as a query image. Similar to the previous subsection, we show

Table 4.2: Comparison of Total averages between the MCN-MSF and the existing methods in terms of a) Recall, b) Precision and c) Accuracy for COREL10800db

Number of Top	Histogram	CAC	MSF	MCN-MSF
20	14.55	15.51	16.48	**19.41**
40	16.45	17.19	18.31	**21.63**
60	17.98	18.85	19.95	**23.45**
80	19.37	20.19	21.39	**25.24**
100	20.66	21.43	22.54	**26.45**
120	21.79	22.64	23.71	**27.89**
140	22.93	23.67	24.80	**29.15**
160	23.94	24.71	25.76	**30.54**
180	24.88	25.71	26.62	**31.73**
200	25.83	26.63	27.74	**32.85**
Grand average	**20.84**	**21.65**	**22.73**	**26.83**

(a) Total average Recall

Number of Top	Histogram	CAC	MSF	MCN-MSF
20	28.73	29.55	30.40	**36.95**
40	24.63	24.96	26.28	**31.85**
60	22.63	23.41	24.58	**29.07**
80	21.46	22.23	23.49	**27.71**
100	20.66	21.43	22.54	**26.45**
120	19.99	20.86	21.92	**25.78**
140	19.52	20.34	21.43	**25.18**
160	19.09	19.94	20.98	**24.95**
180	18.71	19.62	20.57	**24.83**
200	18.42	19.31	20.37	**23.91**
Grand average	**21.38**	**22.17**	**23.25**	**27.67**

(b) Total average Precision

Number of Top	Histogram	CAC	MSF	MCN-MSF
20	96.95	97.05	97.15	**98.27**
40	96.80	96.90	97.00	**98.13**
60	96.65	96.75	96.85	**97.97**
80	96.49	96.58	96.69	**97.82**
100	96.33	96.42	96.52	**97.66**
120	96.16	96.26	96.36	**97.50**
140	96.00	96.09	96.20	**97.34**
160	95.83	95.93	96.03	**97.18**
180	95.66	95.76	95.86	**97.02**
200	95.50	95.59	95.69	**96.85**
Grand average	**96.24**	**96.33**	**96.44**	**97.57**

(c) Total average Accuracy

the performance results varying the number of top matches from 20 to 200 with the interval of 20. For a particular top matches value we compute the average result for all the queries. The total average recall, precision and accuracy of all the query images are shown in the Table 4.2.a, Table 4.2.b and Table 4.2.c, respectively. The last rows of all the tables show grand averages of the corresponding performance measures for all the methods including Histogram, CAC, MSF and MCN-MSF. For all the measures the proposed method (MCN-MSF) shows better performance as shown in the tables.

From the tables, we see that the grand averages of recall, precision and accuracy of the histogram are the lowest among all the methods. The grand averages of the CAC and MSF show the better performances than the histogram. But the performances of these three methods are very close to each other. Only the MCN-MSF method shows the promising difference from other methods.

Chapter 5

Image Retrieval Using Multi-Resolution Nonhomogeneous MSF

5.1 Introduction

In the previous chapter we discussed the implementation of time inhomogeneous Markov chain model by applying the multi-channel nonhomogeneous MSF (MCN-MSF) feature for content based image retrieval. In this chapter we have discussed the implementation of the same model constructing muti-resolution nonhomogeneous MSF descriptor named MRN-MSF feature for the image retrieval.

Based on colors, digital images can be generally three types such as: i) *binary images* which are produced in black and white and represented by pixels consisting of one bit each, ii) *grayscale images* are composed of pixels represented by a number of bits of information, usually ranging from 2 to 8 bits yielding from 4 to 256 color levels and iii) *color images* are usually represented by a greater number of bits ranging from 8 to 24 or higher, which can represent up to 16.7 million color values. For 24-bit images, the bits are often grouped into three parts: 8 for red, 8 for green, and 8 for blue.

Color resolution (known as color depth or bit depth) in a digital image refers to the ability of each pixel to represent a number of color tones (known as color levels or color value). It is normally measured by the level of ability to express the intensity of the three primary components (*e.g.* hue, saturation and value in HSV color model). The number of levels (i.e. $K=2^n$) available to represent each color depends on the number of bits (i.e. n). In multi-resolution nonhomogeneous MSF scheme, an image of a database is represented by a number of different K levels quantized images. The transition matrices are computed for each of the

quantized images. Here, the sizes of the transition matrices are different unlike the MCN-MSF scheme. These matrices are then combined using the time inhomogeneous Markov chain model.

5.2 The Proposed Algorithm

In this paper we propose the Multi-Resolution Nonhomogeneous Markov Stationary Feature (MRN-MSF) algorithm to address the problem of database heterogeneity by populating more spatial information in the feature components. In this method, an image I is quantized into nK (e.g. K=10) levels, thus the set of histogram bins are $S_{nK} = \{s_{1K}, s_{2K}, \ldots, s_{NK}\}$, where $n = 1,2, \ldots, N$. Here, N indicates the dimension of our proposed MRN-MSF method. The corresponding co-occurrence matrices are calculated as (for, K=10)

$$c_{ij}^{10n} = \#(x_1 = s_i, x_2 = s_j \,|(s_i, s_j) \in \{S_{10n}\}),$$
$$n = 1,2, \ldots N. \qquad (5.1)$$

For example, if N=3, there will be three different co-occurrence matrices c_{ij}^{10}, c_{ij}^{20}, and c_{ij}^{30} of sizes 10-*by*-10, 20-*by*-20 and 30-*by*-30, respectively.

After computing the N different transition matrices (*i.e.* nonhomogeneous) from the co-occurrence matrices with different sizes, the equation (3.20) is applied to the nonhomogeneous matrices to form a single composite transition matrix. The Markov chain model (discussed in the Chapter 3) is adopted for the composite matrix to characterize the spatial co-occurrence relation of corresponding set of histogram bins. An initial distribution $\boldsymbol{\pi_0}$ (using equation (3.16)) is computed from the largest co-occurrence matrix. The stationary distribution $\boldsymbol{\pi}$ is calculated from the resultant composite transition matrix using the equation (3.15). The two resultant distributions contribute to form our proposed single dimensional MRN-MSF feature space, as

$$\mathbf{x} = [\pi_0, \pi]. \qquad (5.2)$$

5.3 Implementation of MRN-MSF Method

HSV is one of the most prominent color space for low level image processing. Using HSV color space the CBIR system results in better agreement with color perception than when using RGB. Thus, the RGB image of the underlying database is converted into HSV image first. Then, the individual H component (i.e. H image) of the HSV image is quantized into nK levels. The corresponding spatial co-occurrence matrices (e.g. c_{ij}^{10}, c_{ij}^{20}, and c_{ij}^{30}) for K=10 and N=3 are calculated using equation (5.1). The initial distribution (e.g. π_0^{30}) is then computed from the co-occurrence matrix (*e.g.*, c_{ij}^{30}) using equation (3.16). One the other hand, the equation (3.14) is employed individually for each co-occurrence matrix to compute the corresponding transition matrices (e.g. p_{ij}^{10}, p_{ij}^{20}, and p_{ij}^{30}), respectively. The nonhomogeneous transition matrices are combined into a single composite matrix (*e.g.*, P^{30}) applying the concept non-homogeneous scheme discuss in the Section 3.4. Then, a stationary distribution (e.g., π^{30}) is calculated from the composite transition matrix using the equation (3.15). The initial and the resultant stationary distributions are used to form the targeted feature space (MRN-MSF) according to the equation (5.2). The similarity measurements and the performance evaluation processes between the proposed and the existing systems (*i.e.* Histogram, CAC, and original MSF) are similar to the MCN-MSF system discussed in previous chapter. The block diagram of our proposed system is shown in Fig. 5.1.

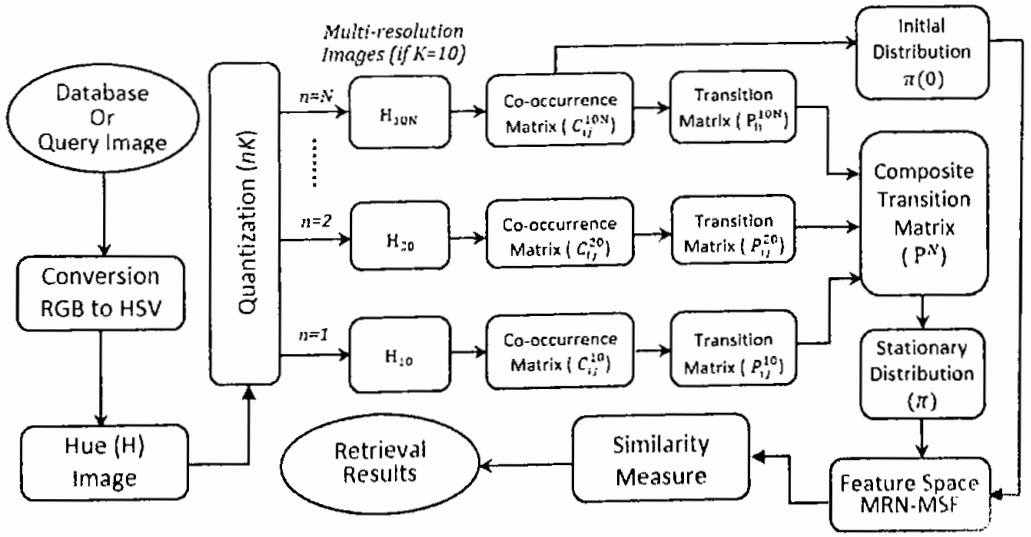


Fig. 5.1 Block diagram of the proposed MRN-MSF Method.

5.4 Experiment and Discussion

Like the MCN-MSF, Multi-Resolution Nonhomogeneous Markov Stationary Feature (MRN-MSF) can also be exploited in the CBIR application as a field of study to evaluate the effectiveness of the MRN-MSF features. The whole evaluation process followed the diagram illustrated in Fig. 5.1.

In this thesis, the very worldwide recognized databases namely: WANG1000 (also known as Corel1000) and Corel10800 are used for our experiments. The performances are compared in terms of different metrics including average, mean average and total average of precision, recall and accuracy as like the previous chapter.

5.4.1 Database I: WANG1000db

In this subsection we have discussed the implementation of the MRN-MSF method with the WANG1000. We show the results with single query, category-wise queries, randomly selected queries and entire database queries search results similar to the previous chapter.

5.4.1.1 Single Query Image Retrieval

At first we used a single image as a query image. From the database image no. 21 is used as query image which is the image of an African category.

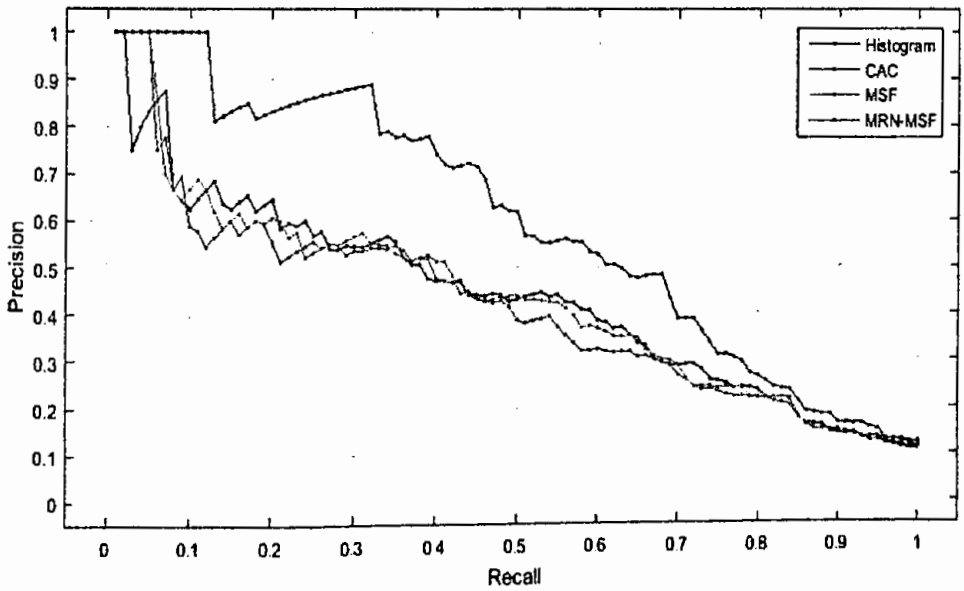


Fig. 5.2 Recall-precision curves of Histogram, CAC, original MSF and proposed MRN-MSF methods for query image-21 in WANG1000 database.

In order to show the comparative results, we find recall-precision graph. We find the precisions over different recalls for Histogram, CAC, original MSF and our proposed method (MRN-MSF). The fig. 5.2 shows the comparative results for query image no. 21. The figure indicates that for recall 0.01 and 0.02 the precision is 100% for all the methods. But up to recall at 0.12, the precision is 100% only for the proposed method. Actually, from recall 0.06 the proposed method outperforms all the existing methods. We also see the significant difference of performance with the proposed method up to recall 0.78. After that the performance is almost similar to the other methods.

5.4.1.2 Category-wise Query Image Retrievals

Like the previous chapter, we select images from all the 10 categories with large number of query images for the category-wise performances. 20 images are selected randomly from each category. For all the categories and retrievals we retrieve 50 images to estimate the performances. Fig. 5.3.a, and Fig. 5.3.b show the category-wise recall and precision, respectively for 50 top matches (*i.e.*, n=50). Beach images (*i.e.* category no. 2) are background-dominated which are very similar to other background-dominated images in the database. Thus, like the MCN-MSF method, beach category shows the least performances for all the methods as shown in the figures. One the other hand, horse category (*i.e.* category no. 8) shows the best performances because the images of the category are object-dominated.

In all the figures the MRN-MSF method outperforms the existing methods in terms of average recall and precision for all the categories. The MRN-MSF method outperform even for category 4 (Buses) unlike MCN-MSF method.

5.4.1.3 M-Randomly Selected Query Image Retrievals

To show the extensive performances of the proposed method we choose 50 random query images from the database. We show the performance results

varying the number of top matches from 10 to 100 with the interval of 10. For a particular top matches value we compute the average result of 50 queries. The mean average recall, precision and accuracy of the 50 query images are shown in the Fig. 5.4.a, Fig. 5.4.b and Fig. 5.4.c, respectively. In each case, we compare the results of our proposed method with Histogram, CAC and MSF. For all the cases, the proposed method shows the significant improvements than the existing methods as shown in the figures. For the case of MAR the improvement of the proposed method is gradually increasing over the top matches as shown in Fig. 5.4.a. For the case of MAA the difference of the proposed method is more promising and constant progressing over the number of top matches as shown in Fig. 5.4.c.

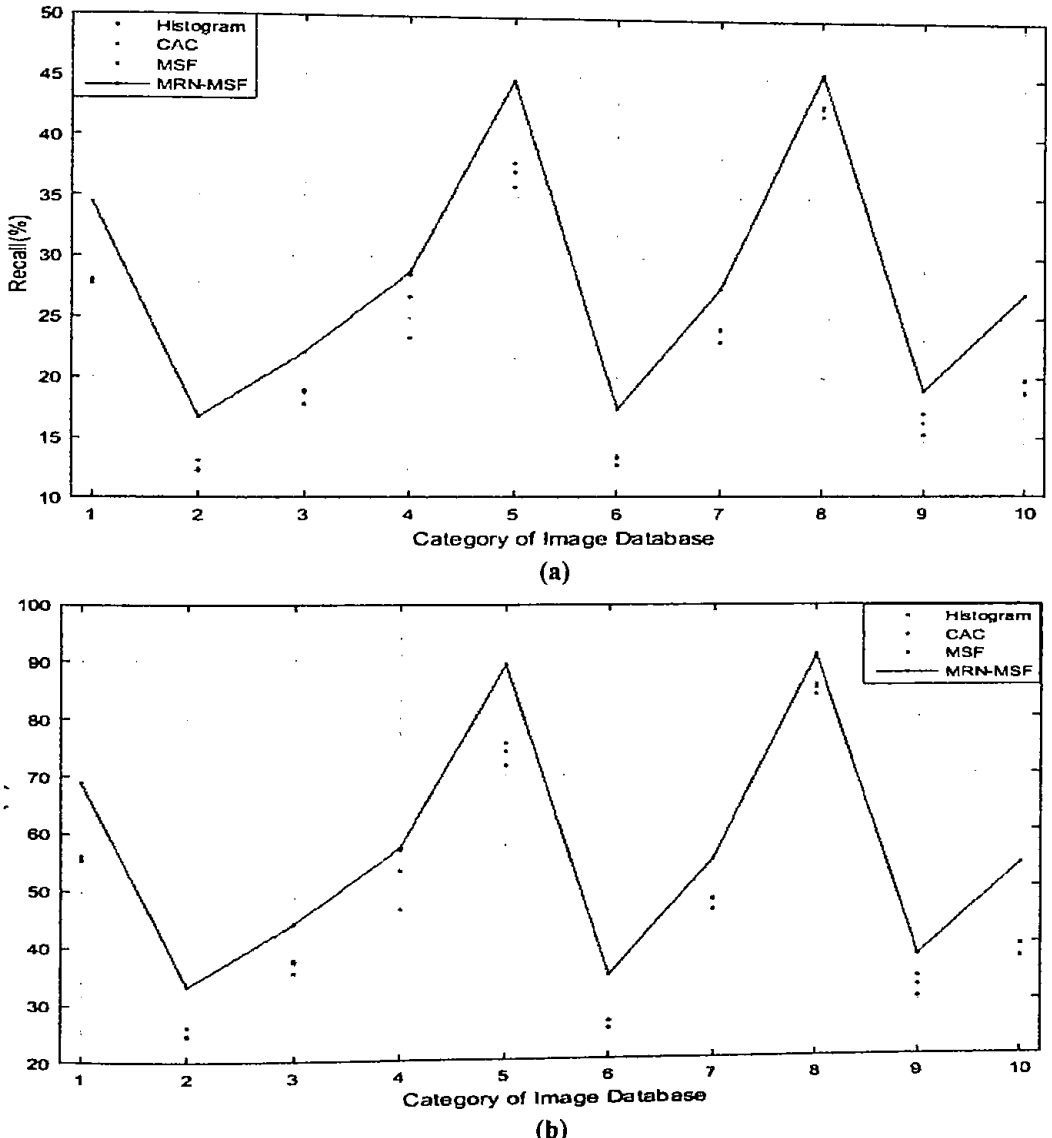
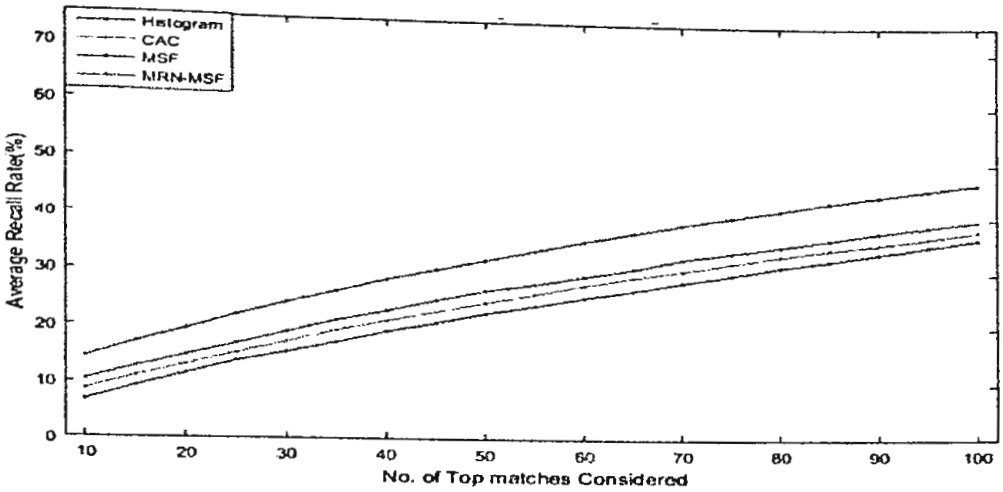
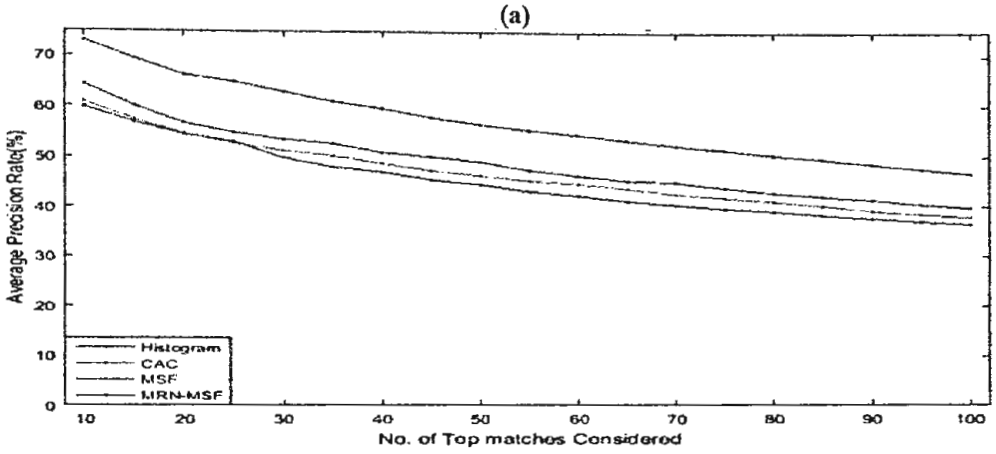


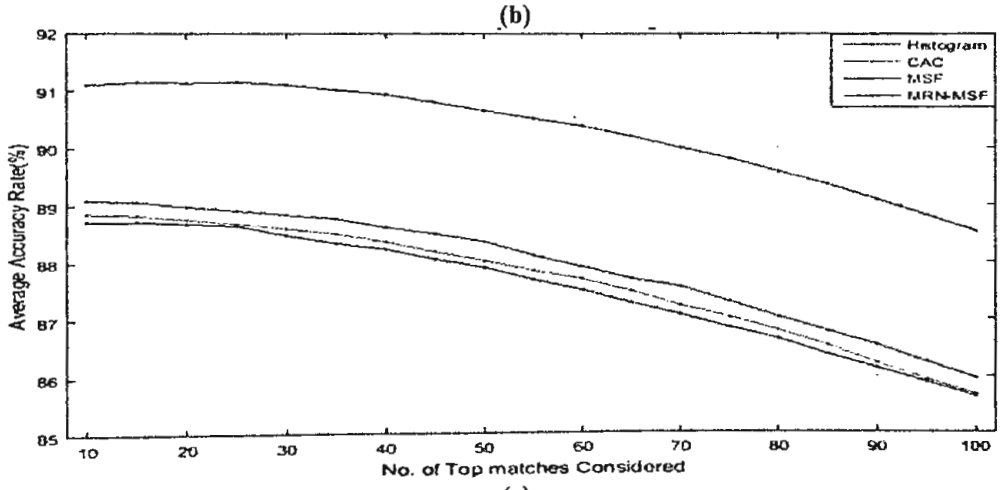
Fig. 5.3 Category-wise (inWANG1000db) performance at n=50, in terms of a) Recall and b) Precision for Histogram, CAC, MSF and MRN-MSF.



(a)



(b)



(c)

Fig. 5.4 Mean average a) Recall, b) Precision and c) Accuracy curves of Histogram, CAC, original MSF and proposed MRN-MSF methods for randomly 50 query images in WANG1000 database.

5.4.1.4 Entire Database Search Results

To show overall performance of the entire image database, we select every image of the database as a query image. Similar to the previous chapter, we show the

Table 5.1: Comparison of Total averages between the MRN-MSF and the existing methods in terms of a) Recall, b) Precision and c) Accuracy for WANG1000db

Number of	Histogram	CAC	MSF	MRN-
10	9.34	11.18	12.81	**17.20**
20	13.76	15.57	17.31	**22.31**
30	17.77	19.56	21.37	**26.87**
40	21.51	23.19	25.06	**30.96**
50	24.82	26.49	28.34	**34.65**
60	27.87	29.51	31.46	**37.91**
70	30.75	32.34	34.31	**40.98**
80	33.40	34.93	36.96	**43.71**
90	35.87	37.36	39.36	**46.19**
100	38.35	39.78	41.77	**48.68**
Grand average	**25.34**	**26.99**	**28.87**	**34.95**

(a) Total average Recall

Number of Top	Histogram	CAC	MSF	MRN-MSF
10	56.59	57.54	59.87	**69.35**
20	52.04	53.28	55.74	**65.24**
30	48.90	50.31	52.69	**61.90**
40	46.57	47.87	50.18	**59.03**
50	44.31	45.71	47.84	**56.45**
60	42.33	43.78	45.97	**54.02**
70	40.67	42.12	44.25	**51.97**
80	39.11	40.56	42.69	**50.01**
90	37.71	39.16	41.20	**48.20**
100	36.31	37.77	39.71	**46.40**
Grand average	**44.45**	**45.81**	**48.01**	**56.26**

(b) Total average Precision

Number of Top	Histogram	CAC	MSF	MRN-MSF
10	88.72	88.84	89.06	**91.14**
20	88.60	88.71	88.96	**91.16**
30	88.40	88.51	88.77	**91.07**
40	88.15	88.24	88.51	**90.89**
50	87.81	87.90	88.17	**90.63**
60	87.42	87.50	87.79	**90.28**
70	87.00	87.07	87.36	**89.90**
80	86.53	86.59	86.89	**89.44**
90	86.02	86.07	86.37	**88.94**
100	85.52	85.56	85.85	**88.44**
Grand average	**87.42**	**87.50**	**87.77**	**90.19**

(c) Total average Accuracy

performance results varying the number of top matches from 10 to 100. For a particular top matches value we compute the average result for all the queries. The total average recall, precision and accuracy of all the query images are shown in the Table 5.1.a, Table 5.1.b and Table 5.1.c, respectively. The last rows of all the tables show grand averages of the corresponding performance measures for all the methods including Histogram, CAC, MSF and MRN-MSF. For all the measures the proposed method shows better performance as shown in the tables.

From the tables, we see that the grand averages of recall, precision and accuracy of the histogram are the lowest among all the methods. The grand averages of the CAC and MSF show the better performances than the histogram. But the performances of these three methods are very close to each other. Only the MRN-MSF method shows the promising difference from other methods.

5.4.2 Database II: COREL10800db

5.4.2.1 Single Query Image Retrieval

Here, at first a single image is used as a query image. From the database image no. 2287 is used as query image which is the image of bus category.

In order to show the comparative results, we find recall-precision graph as the previous sections. We see the precisions over the different recalls for the methods including the proposed MRN-MSF. The Fig. 5.5 shows the comparative results of for query image no. 2287 of the database. The figure indicates that for recall 0.01 and 0.02 the precision is 100% for all the methods. From recall 0.04 the proposed method outperforms the existing methods. At recall 0.15 the precision of the method is 0.75 which is very high compare to the other methods. We also see the clear difference of performance with the proposed method is up to recall 0.82. After that the performance is almost similar to the other methods.

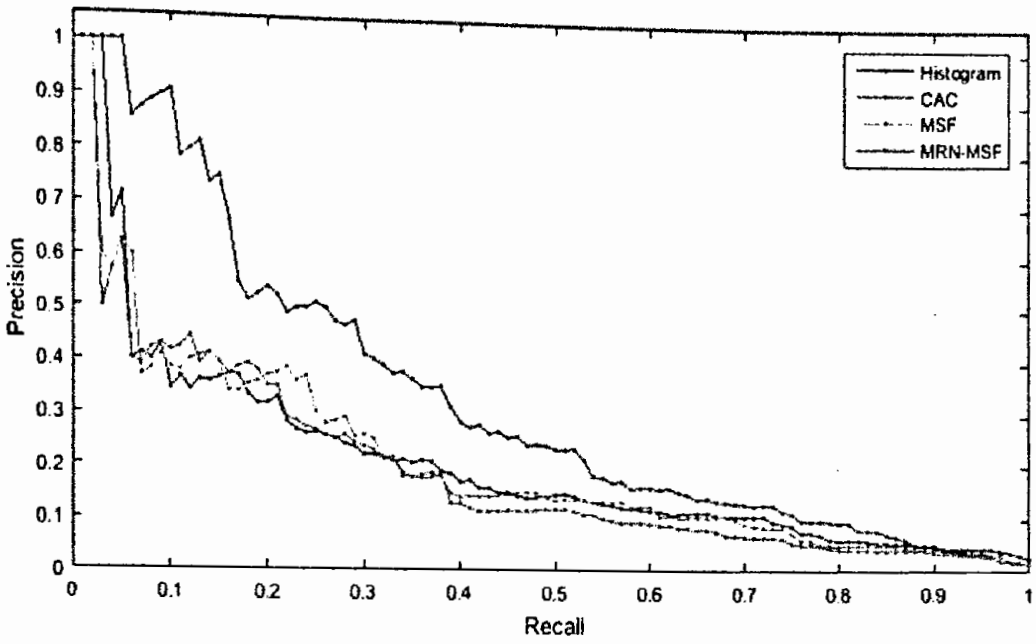


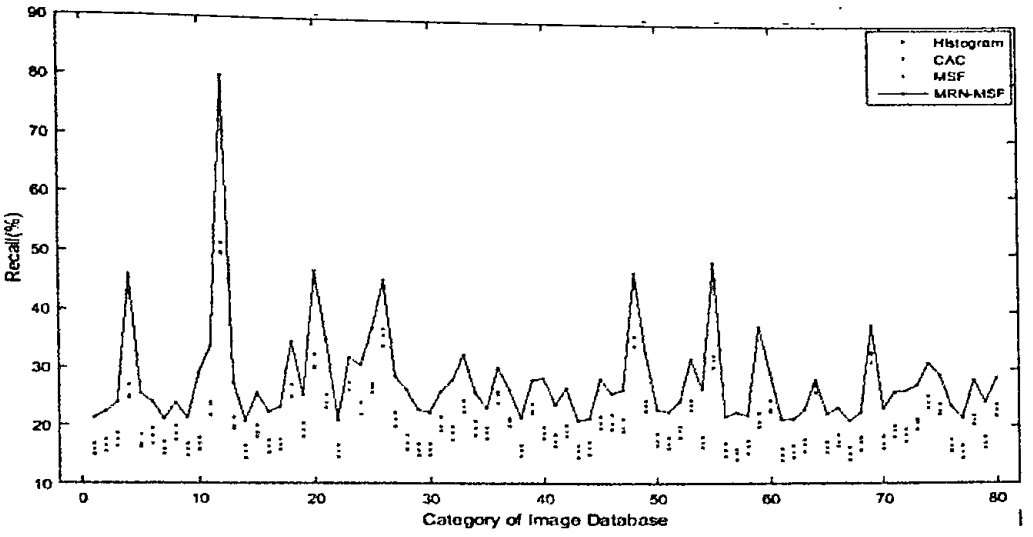
Fig. 5.5 Recall-precision curves of Histogram, CAC, original MSF and proposed MRN-MSF methods for query image-2287 in COREL10800 database.

5.4.2.2 Category-wise Query Image Retrievals

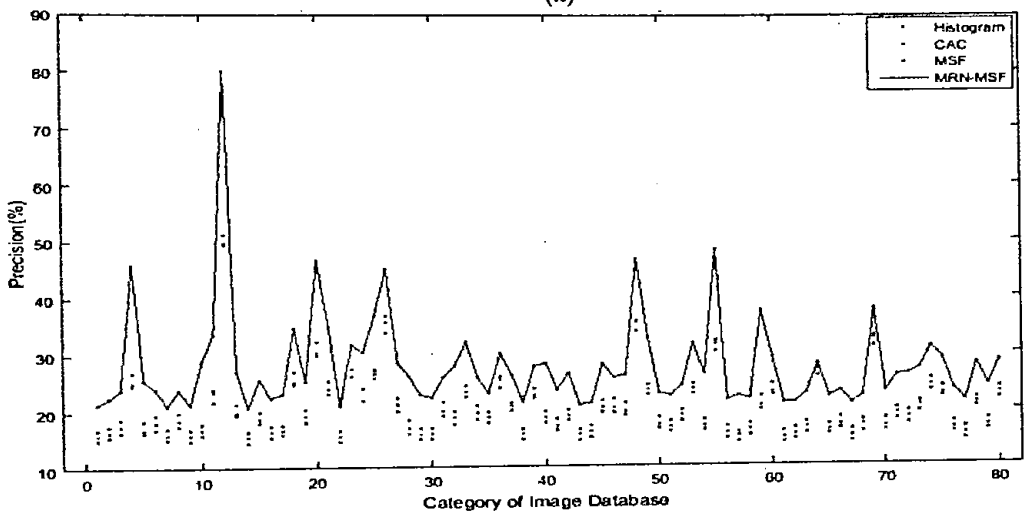
To show the category-wise performances with large number of query images we select images from all the 80 categories. 60 images are selected randomly for query from each category. For all the categories and retrievals we retrieve 100 images to estimate the performances. Fig. 5.6.a and Fig. 5.6.b show the category-wise recall and precision, respectively for 100 top matches (*i.e.*, n=100).

Category 14 (aviation) images are background-dominated and the foreground (*i.e.* the expected object) of those images are coarse in size, shape and color. Thus they are more similar to other background-dominated images in the database than that of the category. As a result, aviation category shows the least performances for all the methods as shown in the figures. One the other hand, fitness category (*i.e.* category no. 12) shows the best performances because the images of the category are object-dominated.

In all the figures the MRN-MSF method outperforms the existing methods in terms of average recall, precision and accuracy for all the categories.



(a)



(b)

Fig. 5.6 Category-wise (in COREL10800db) performance at n=100, in terms of a) Recall and b) Precision for Histogram, CAC, MSF and MRN-MSF.

5.4.2.3 M-Randomly Selected Query Image Retrievals

To show the extensive performances of proposed method we choose 100 random query images from the database similar to the subsection 4.5.2.3. We show the performance results varying the number of top matches from 10 to 200 with the interval of 10. For a particular top matches value we compute the average result of 100 queries. The mean average recall, precision and accuracy of the 100 query images are shown in the Fig. 5.7.a, Fig. 5.7.b and Fig. 5.7.c, respectively. In each case, we compare the results of our proposed method with Histogram, CAC and MSF. For all the cases, the proposed method shows the significant improvements

than the existing methods as shown in the figures. For the case of MAR the improvement of the proposed method is gradually increasing over the top matches as shown in Fig. 5.7.a. For the case of MAA the improvement of the proposed method is promising and constant progressing over the number of top matches as shown in Fig. 5.7.c.

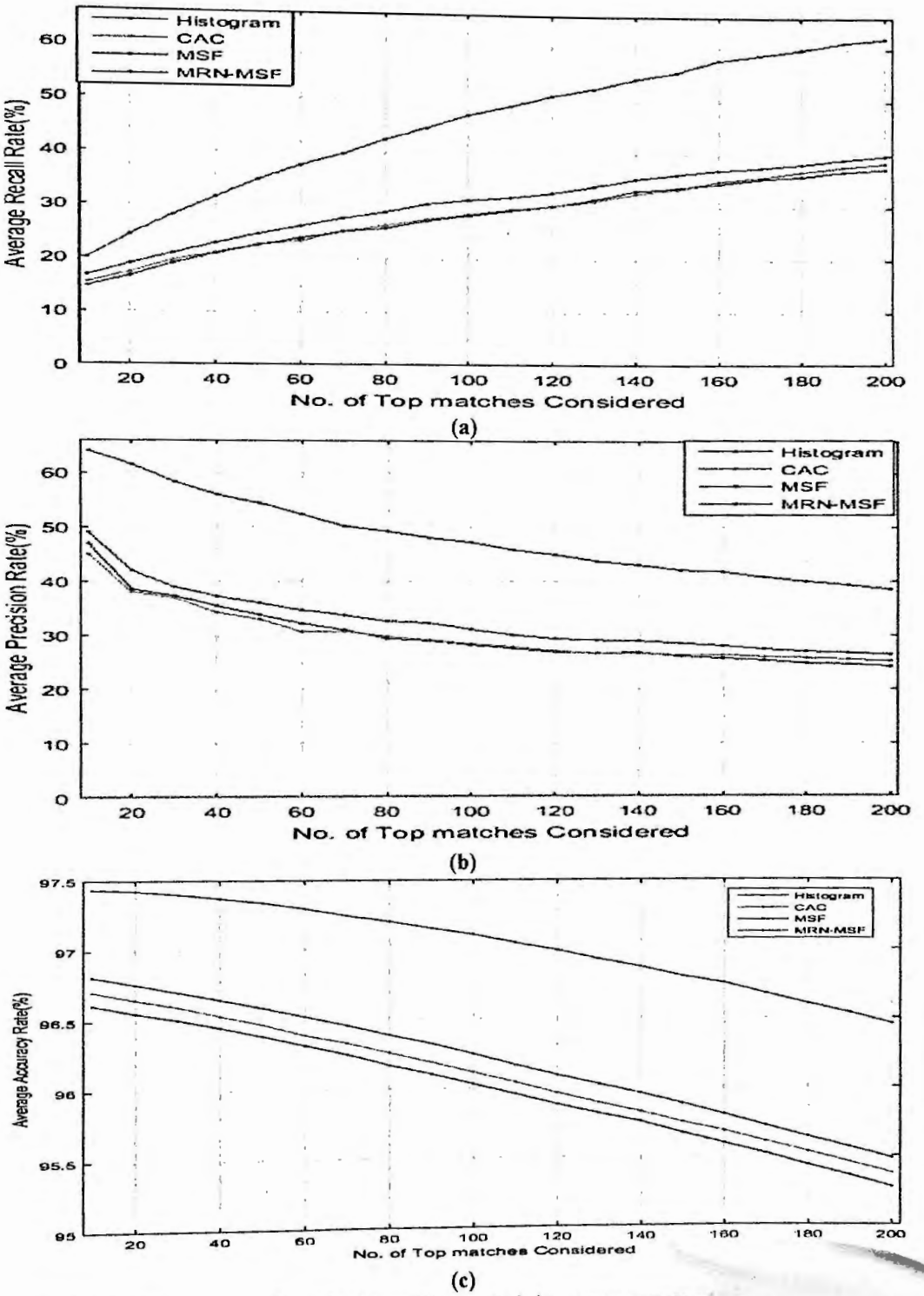


Fig. 5.7 Mean average a) Recall, b) Precision and c) Accuracy curves of original MSF and proposed MRN-MSF methods for randomly 100 queries COREL10800 database.

5.4.2.4 Entire Database Search Results

To show overall performance of the entire image database, we select every image of the database as a query image similar to the previous chapter. We show the

Table 5.2: Comparison of Total averages between the MRN-MSF and the existing methods in terms of a) Recall, b) Precision and c) Accuracy for COREL18000db

Number of Top	Histogram	CAC	MSF	MRN-MSF
20	14.55	15.51	16.48	**22.39**
40	16.45	17.19	18.31	**24.73**
60	17.98	18.85	19.95	**26.45**
80	19.37	20.19	21.39	**28.19**
100	20.66	21.43	22.54	**29.55**
120	21.79	22.64	23.71	**30.89**
140	22.93	23.67	24.80	**32.17**
160	23.94	24.71	25.76	**33.50**
180	24.88	25.71	26.62	**34.76**
200	25.83	26.63	27.74	**35.82**
Grand average	**20.84**	**21.65**	**22.73**	**29.84**

(a) Total average Recall

Number of Top	Histogram	CAC	MSF	MRN-MSF
20	28.73	29.55	30.40	**39.95**
40	24.63	24.96	26.28	**34.83**
60	22.63	23.41	24.58	**32.08**
80	21.46	22.23	23.49	**30.73**
100	20.66	21.43	22.54	**29.55**
120	19.99	20.86	21.92	**28.74**
140	19.52	20.34	21.43	**28.12**
160	19.09	19.94	20.98	**27.69**
180	18.71	19.62	20.57	**27.31**
200	18.42	19.31	20.37	**26.91**
Grand average	**21.38**	**22.17**	**23.25**	**30.59**

(b) Total average Precision

Number of Top	Histogram	CAC	MSF	MRN-MSF
20	96.95	97.05	97.15	**99.07**
40	96.80	96.90	97.00	**98.98**
60	96.65	96.75	96.85	**98.76**
80	96.49	96.58	96.69	**98.66**
100	96.33	96.42	96.52	**98.47**
120	96.16	96.26	96.36	**98.25**
140	96.00	96.09	96.20	**98.15**
160	95.83	95.93	96.03	**97.99**
180	95.66	95.76	95.86	**97.83**
200	95.50	95.59	95.69	**97.64**
Grand average	**96.24**	**96.33**	**96.44**	

(c) Total average Accur

performance results as like as the subsection 4.5.2.4. For a particular top matches value we compute the average result for all the queries. The total averages are shown in the Table 5.2.a, Table 5.2.b and Table 5.2.c, respectively. The last rows of all the tables show grand averages of the corresponding performance measures for all the methods including Histogram, CAC, MSF and MRN-MSF. For all the measures the proposed method shows better performance as shown in the tables.

From the tables, we see that the grand averages of recall, precision and accuracy of the histogram are the lowest among all the methods. The grand averages of the CAC and MSF show the better performances than the histogram. But the performances of these three methods are very close to each other. Only the MRN-MSF method shows the promising difference from other methods.

5.5 Performance Summary

In this subsection, we discuss the overall comparative performances for all the methods exploiting both databases. We summarize the performances for the methods in terms of the grand average recall, precision and accuracy as shown in Table 5.3. To compute the grand averages, we average all the values for different top matches from 10 to 100 with interval 10 for Database I (WANG1000) and from 20 to 200 with interval 20 for Database II (COREL10800), individually for every methods.

As the MRN-MSF feature gather more spatial information than MCN-MSF, the MRN-MSF method always outperforms MCN-MSF as shown in Table 5.3. We also see from the table that the grand average recall and precision improvements of MRN-MSF method for large database are higher compare to the small database. Thus, the second method is more appropriate for large database.

We have already discussed that, the two methods (MCN-MSF and MRN-MSF) show always better performances than the existing methods (Histogram, CAC and MSF). From the table we see that all the performances of the methods for Database I is higher than Database II. The exception rage

accuracy. Note that, the size and degree of heterogeneity for Database II is more than Database I. Hence, the performances of the Database I is higher than the Database II. It can be noted from eq. (2.8) that true negative value plays major role in calculation of accuracy. Usually for a huge database (in size and category), the true negative value for a query image is much higher than small database. Thus, the grand average accuracies for Database II is high.

Table 5.3: Performance of all methods for all databases

Databases	Performance Metrics (Grand Average)	Methods				
		Histogram	CAC	MSF	MCN-MSF	MRN-MSF
WANG1000	Recall (%)	25.34	26.99	28.87	33.44	34.95
	Precision (%)	44.45	45.81	48.01	54.77	56.26
	Accuracy (%)	87.42	87.50	87.77	90.04	90.19
Corel10800	Recall (%)	20.84	21.65	22.73	26.83	29.84
	Precision (%)	21.38	22.17	23.25	27.67	30.59
	Accuracy (%)	96.24	96.33	96.44	97.57	98.38

Chapter 6

Conclusion and Future Works

6.1 Conclusion

In this research we propose two general frameworks to construct two image descriptors MCN-MSF and MRN-MSF to extend original MSF feature. Firstly, MCN-MSF is based on nonhomogeneous Markov chain model. Basically, this model collects more spatial information than the original MSF. It combinedly gathers inter-bin level and extra-bin level image distinguishable information similar to MSF. Initial distribution of a Markov chain is used to collect inter-bin spatial information. Stationary distribution of the Markov chain is used to find extra-bin spatial information. To compute the stationary distribution nonhomogeneous Markov chain model is used in MCN-MSF. Unlike MSF, in the nonhomogeneous model same size different transition matrices are applied to capture more spatial information. Here the multiple transition matrices are computed from multiple color channels of an image. Then the matrices are multiplied to construct composite transition matrix. The stationary distribution of the matrix along with the initial distribution contribute to form MCN-MSF descriptor. The performance of proposed method is evaluated with two recognized databases, namely, WANG1000 and COREL10800. For both cases, the proposed method significantly increases the performances in term of recall, precision and accuracy. For example, more than 4% improvements of average recall are recorded for the both databases.

Secondly, MRN-MSF is also based on nonhomogeneous Markov chain model. Unlike MCN-MSF, the composite transition matrix is computed from transition matrices with different sizes. Here the transition matrices are computed from an image with different color resolutions (*i.e.* different color levels). An image quantized with lower resolution is used to compute lower size transition matrix. One the other hand, image quantized with higher resolution is used to

higher size transition matrix. Usually, the transition matrix with higher size carries higher degree of spatial information. In addition, different transition matrices contain different degree of spatial information. These matrices are combined to form the composite matrix. The composite matrix contains more and variety of spatial information. The matrix is then used to construct the required stationary distribution for the MRN-MSF descriptor. The performance of the descriptor is evaluated with the databases similar to MCN-MSF method. Since the method contains more distinguishable spatial information it is more preferable to larger databases. For example, more than 3% improvement of average recall is recorded for the Database II, whereas around 1% improvement is recorded for the Database I.

6.2 Future Works

This proposed methods are independent from the nature of the problem and can be applied to wide range of pattern recognition problems. These concepts also can be used for feature extraction for other visual image descriptors like HOG which is generally used to detect object in images, with some modification.

The methods take cares of images with histogram, inter-bin and extra-bin level distinguishable features. By analyzing inherent spatial structure of images rigorously, considerable improvements still be possible for the image retrieval results. However, the methods do not cover histogram-undistinguishable images for image retrieval. Hence, the general image retrieval results can suffer for their large spatial variance. Thus, further investigation is needed to cover all the classes of images.

References:

[1] A. N. Tikle, C. Vaidya and P. Dahiwale, "A Survey of Indexing Techniques for Large Scale Content-Based Image Retrieval," IEEE Int. Conf. on Electrical, Electronics, Signals,Communication and Optimization (EESCO), Visakhapatnam, India, Jan. 24-25, 2015, pp. 1-5.

[2] K. Juneja *et al.*, "A Survey on Recent Image Indexing and Retrieval Techniques for Low-Level Feature Extraction in CBIR Systems," IEEE Int. Conf. on Computational Intelligence & Communication Technology (CICT), Ghaziabad, Pakistan, Feb. 13-14, 2015 , pp. 67-72.

[3] Keyuri M. Zinzuvadia1, Bhavesh A. Tanawala, and Keyur N. Brahmbhatt, "A Survey on Feature Based Image Retrieval Using Classification and Relevance Feedback Techniques," Int.l J. of Innovative Research in Computer and Communication Engineering, Jan. 2015, vol. 3, no. 1, pp. 508-513.

[4] Mussarat Yasmin, Sajjad Mohsin, and Muhammad Sharif, "Intelligent Image Retrieval Techniques: A Survey," J. of Applied Research and Technology, Feb. 2014, vol. 12, no. 1 pp. 87-103.

[5] G. Pass, and R Zabih, "Histogram Refinement for Content-Based Image Retrieval," 3rd IEEE Workshop on. Applications of Computer Vision, Sarasota, Florida, Dec. 2-4, 1996, pp. 96-102.

[6] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns," TPMAI, 24(7):971-987, 2002.

[7] D. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," International Journal of Computer Vision, 60(2):91-110, 2004.

[8] E. Hadjidemetriou, M. Grossberg, and B. Nayar, "Multiresolution Histograms and Their Use for Recognition," TPAMI, 26(7):831-847, 2004.

[9] G. Pass, R Zabih, and J. Miller, "Comparing Images Using Color Coherence Vectors," 4th ACM Int. Conf. on Multimedia, Boston, USA, Nov. 18-22, 1996, pages 65-73.

[10] J. Huang *et al.*, "Image Indexing Using Color Correlograms," IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR '97), San Juan, Puerto Rico, June 17-19, 1997, pp. 762-768.

[11] J. Huang, *et al.*, "Spatial Color Indexing and Applications," Int. J. of Computer Vision (IJCV), vol. 35, no. 3, Dec. 1999, pp. 245-268.

[12] Li, J., Wu, W., Wang, T., Zhang, Y., "One step beyond histogram: Image representation using markov stationary features," In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. (2008).

[13] Elfadel, I., Picard, R., "Gibbs random fields, cooccurrences, and texture modeling," IEEE Transactions on Pattern Analysis and Machine Intelligence. (1994) 24–37.

[14] Partio, M., Cramariuc, B., Gabbouj, M., Visa, A.: "Rock texture retrieval using gray level co-occurrence matrix," (2010).

[15] Davis, L., Johns, S., Aggarwal, J.: "Texture analysis using generalized cooccurrence matrices," IEEE Transactions on Pattern Analysis and Machine Intelligence 1 (1979) 251–259.

[16] Ni, B., Yan, S., Kassim, A.: "Contextualizing histogram," In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, (2009) 1682–1689.

[17] Bingbing Ni, Shuicheng Yan, Meng Wang, A. A.Kassim, Qi Tian. "High-Order Local Spatial Context Modeling by Spatialized Random Forest", IEEE Transactions on Image Processing, 2013.

[18] Haralick, R., Shanmugam, K., Dinstein, I.: "Textural features for image classification," IEEE Transactions on Systems, Man and Cybernetics 3 (1973)

[19] Geisler, W.S., Perry, J.S., Super, B.J., Gallogly, D.P.: "Edge co-occurrence in natural images predicts contour grouping performance," Vision Research 41 (2001) 711–724.

[20] Y. Rui, T. Huang, and S. Chang. "Image retrieval: Current techniques, promising directions and open issues," Journal of Visual Communication and Image Representation, 10(4):39-62, 1999.

[21] Peter L. Stanchev. "General image retrieval model," In Proc. 27-th Conf. of the Union of Bulgarian Math.,Pleven, Bulgaria, pages 63-71, 1998.

[22] Hideyuki Tamura and Naokazu Yokoya, "Image database systems: A survey," pattern Recognition, 17:29-43, 1984.

[23] Shi-Kuo Chang and Arding Hsu. "Image information systems: Where do we go from here?" IEEE Trans. on Knowledge and Data Eng., 4(5):431- 442, 1992.

[24] M. Wang, K. Yang, X.-S. Hua, and H. Zhang. "Towards a relevant and diverse search of social images," TMM, 2010.

[25] M. Satyanarayanan. "Mobile computing: the next decade," ACM SIGMOBILE, 2011.

[26] C. Berrut *et al.* "Status review on non-text information retrieval," Technical Report ELPUB106, European Commission, Brussels, 1995.

[27] E. Rasmussen. "Indexing images," Annual Review of Information Science and Technology, 32:169-196, 1997.

[28] Abby A. Goodrum. "Image information retrieval: An overview of current research," Informing Science, Special Issue on Information Science Retrieval, 3(2):63-67, 2000.

[29] John Eakins and Margeret Graham. "Content-based image retrieval: A report to the jisc technology applications programme," Technical report, Institute for Image Data Research, University of Northumbria, 1999.

[30] T. Kato. Database "Architecture for Content-Based Image Retrieval," In SPIE 1662, Image Storage and Retrieval Systems, pages 112-123, 1992.

[31] V. N. Gudivada and V. V. Raghavan. "Content-based image retrieval systems," IEEE Computer, 28(9):18-22, 1995.

[32] Mrs. Sheema. A. S, Mr. Allwin Stephen, 2015, "A Study and Analysis on Image Indexing and Retrieval in Texture Base using Haar Wavelets," International Journal Of Engineering Research & Technology (IJERT) NCICN – 2015 (Volume 3 – Issue 13).

[33] John Erickson, "Database Technologies: Concepts, Methodologies, Tools, and Applications". Book, University of Nebraska-Omaha, USA, 2009.

[34] David Feng, W.C. Siu, Hong Jiang Zhang, "Multimedia Information Retrieval and Management: Technological Fundamentals and Applications," Book, Springer Science & Business Media, Apr 17, 2013.

[35] S. K. Chang, and A. Hsu, "Image information systems: where do we go from here?" IEEE Trans. on Knowledge and Data Engineering, Vol.5, No.5, pp. 431-442, Oct.1992.

[36] S. K. Chang, E. Jungert, and Y. Li, "Representation and retrieval of symbolic pictures using generalized 2D string", Technical Report, University of Pittsburgh, 1988.

[37] Yossi Rubner, Carlo Tomasi, "Perceptual Metrics for Image Database Navigation," Springer Science+Business Media, New York, 2001.

[38] David Feng, W.C. Siu, Hong Jiang Zhang, "Multimedia Information Retrieval and Management: Technological Fundamentals and Applications", Book, Springer Science & Business Media, Apr 17, 2013.

[39] D. Slater and G. Healey. "The illumination-invariant recognition of 3D objects using local color invariants," IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(2):206-210, 1996.

[40] S. Kullback. "Information Theory and Statistics," Dover, New York, NY, 1968.

[41] "ImageCLEF: Experimental Evaluation in Visual Information Retrieval," edited by Henning Müller, Paul Clough, Thomas Deselaers, Barbara Caputo. 2010, Springer-Verlag Berlin Heidelberg.

[42] Gagniuc, Paul A. "Markov Chains: From Theory to Implementation and Experimentation," (2017), USA, NJ: John Wiley & Sons. pp. 1–256. ISBN 978-1-119-38755-8.

[43] Oliver C. Ibe, "Markov Processes for Stochastic Modeling", Book, Second Edition 2013, Elsevier Inc.

[44] Olle Häggström, "Finite Markov Chains and Algorithmic Applications" Book, Cambridge University Press, May 30, 2002.

[45] Karl Sigman, "Markov Chain Analysis with Stochastic", lecture notes, University of Columbia, Website: http://www.ieor.columbia.edu/ ~sigman/stochastic-I.html.

[46] Zhanfeng Li, Min Huang, Xiaohua Meng, Xiangyu Ge, "The Limit Theorems for Function of Markov Chains in the Environment of Single Infinite Markovian Systems", Mathematical Problems in Engineering, vol. 2020, Article ID 8175723, 11 pages, 2020. https://doi.org/ 10.1155/2020/ 8175723.

[47] Ross, Sheldon M. (2014). "Chapter 4.2: Chapman–Kolmogorov Equations," Introduction to Probability Models (11th ed.). p. 187. ISBN 978-0-12-407948-9.

[48] Lothar Breuer, "Introduction to Stotachtic Processes," University of Kent, Sep. 2014. Accessed Mar. 17, 2016. https://www.kent.ac.uk/smsas/personal /lb209/files/sp07.pdf

[49] R. Ash, "Real analysis and probability," Academic Press, New York, 1972, Probability and Mathemat-ical Statistics, No. 11.

[50] J. Li, *et al.*, "One Step Beyond Histograms: Image Representation Using Markov Stationary Features," IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR '08), Alaska, USA, June 23-28, 2008, pp. 1-8.

[51] Y. Song, Xiao Chen, S, Qu, "Content based Image Retrieval with Color Invariants," 2nd Int. Conf. on Computer Science and Electronics Engineering (ICCSEE 2013), Hangzhou, China, Mar. 22--23, 2013, pp. 2829-2832.

[52] B. Ni, S. Yan, and A. Kassim, "Directed Markov Stationary Features for Visual Classification," IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Taipei, Taiwan, Apr. 19-24, 2009, pp. 825-828.

[53] F. Lee, *et al.*, "Face Recognition Algorithm Using Multi-directional Markov Stationary Features and Adjacent Pixel Intensity Difference Quantization Histogram," 7th Int. Conf. on Systems and Networks Communications (ICSNC 2012), Lisbon, Portugal, Nov. 18-23, 2012, pp. 113-117.

[54] Md. Saiful Islam, Md. Emdadul Haque, Md. Ekramul Hamid, "Content based Image Searching using Multidimensional MSF", International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC^4ME^2-2016), University of Rajshahi, Bangladesh.

[55] Md. Saiful Islam, Md. Emdadul Haque, Md. Ekramul Hamid, "Multidimensional Markov Stationary Feature for Image Retrival Systems", Rajshahi University Journal of Science and Engineering, 2016.

[56] James Z. Wang, "The WANG databases used in SIMPLIcity paper for research comparison," Jun 2007. Accessed Aug. 15, 2015. http://wang.ist.psu.edu/ docs/related/.

[57] Dacheng Tao, "The COREL Database for Content Based Image Retrieval," July 2009. Accessed Dec. 16, 2015. https://sites.google.com/site/dctresearch/Home/ content-based-image-retrieval.