School of Engineering and Information Technology

Using Rough Set Theory to Improve Content Based Image Retrieval System

Maryam Shahabi Lotfabadi

This thesis is presented for the Degree of

Doctor of Philosophy of

Murdoch University

September 2016

Declaration

I declare that this thesis is my own account of my research and contains as its main content work, which has not previously been submitted for a degree at any territory education institution.

Maryam Shahabi Lotfabadi

Acknowledgement

Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Mohd Fairuz Shiratuddin for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the times of research and writing of this thesis.

I would also like to give many thanks to my co-supervisor, Associate Professor Dr. Kevin Wong, for his helpful advice and useful comments at all times since I started the study. I could not have imagined having a better advisor and mentor for my PhD study.

I would like to express my very profound gratitude to my husband, Amir, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing the thesis. This accomplishment would not have been possible without you. Thank you.

I would like to thank my beautiful Mum for the support she provided me through my entire life. She was always there cheering me up and stood by me through the good times and bad. I would also like to thank my brother, Shahram, my sister, Jenos, and my lovely niece, Sajedeh, without whose love, encouragement and support I would not have finished this thesis. I am so blessed to have you all.

I thank my officemates Adzira, Sal, Ziba, Mehrnaz and Nurulhuda for all the fun we have had in the last four years. We were not only able to support each other by deliberating over our problems and finding, but also happily by talking about things other than just our papers.

In conclusion, I recognise that this research would not have been possible without the financial assistance of Australian Postgraduate Award (APA) scholarship and Murdoch University Excellence Award (MUREX) scholarship, and express my gratitude to those agencies.

Abstract

Each image in a Content Based Image Retrieval (CBIR) system is represented by its features such as colour, texture and shape. These three groups of features are stored in the feature vector. Therefore, each image managed by the CBIR system is associated with one or more feature vectors. As a result, the storage space required for feature vectors is proportional to the amount of images in the database. In addition, when comparing the similarities among images, the CBIR needs to compare these feature vectors. Nonetheless, researchers are still facing problems when working with a huge image database. Much time is needed when comparing huge feature vectors, as a large amount of memory is required to run the CBIR system. Due to this problem, feature reduction and selection techniques are employed to alleviate the storage and time requirements of large feature vectors. There are many feature reduction techniques, including linear projection techniques such as Principal Component Analysis (PCA), Linear Discriminate Analysis (LDA) and metric embedding techniques (both linear and non-linear). However, these methods have limitations in the CBIR system and cannot improve CBIR performance (retrieval accuracy) and reduce semantic gap efficiently. Therefore, we need a feature selection method that can deal with image features efficiently and has the ability to deal with uncertainties.

This research proposes an improved approach to select significant features from the huge image feature vector. The concept behind this research is that it is possible to extract image feature relational patterns in an image feature vector database. After which, these relational patterns are used to generate rules and improve the retrieval results for a CBIR system. In addition, this research proposes a CBIR system utilising the Rough Set instead of deterministic and crisp methods. In this research, Rough Set rules are evaluated with noisy images. Also, in order to have a more accurate classifier in the CBIR system, the classifier is proposed to be based on the Rough Set and Support Vector Machine (SVM) in this research.

The significance of this research is firstly, proposing an improved pre-processing phase to solve CBIR problems. Secondly, proposing an integrated framework of using Rough Set with one-versus-one (1-v-1) Support Vector Machine and Rough Set with one-versus-all (1-v-r) Support Vector Machine classifiers in CBIR systems. The Rough Set theory, as a feature selection method in this pre-processing phase, could solve huge amounts of image features problems by narrowing the search space. Also, this theory could deal with vague and incomplete areas by its upper and lower approximations and solve the incomplete and vague areas in image descriptions. As such, the accuracy of the CBIR system can be improved. This proposed approach also gives the confidence and deviation of the estimation (that traditional methods cannot provide before) when compared with historical systems. Finally, the semantic gap problem can be reduced by the Fuzzy Rough Set semantic rules.

The performance of the proposed CBIR system is assessed using 2000 images from the Corel image dataset. The images were divided into 10 semantic groups, as well as a number of features. They were then compared to other techniques such as Gain Ratio, Genetic Algorithm, Information Gain, Isomap, Kernel PCA, OneR, Principal Component Analysis (PCA) and Relief-F. The results from the experiment conducted in this thesis show that the proposed feature selections and classifiers will improve the semantic performance results in the proposed CBIR systems. Retrieval accuracy results for Fuzzy Rough feature selection is 91.06% for Normal images, and the results are 90.31%, 91.28% and 90.42% with Gaussian Noise, Salt & Pepper Noise and Poisson Noise respectively.

Moreover, comparing the Rough Set with 1-v-1 SVM and the Rough Set with 1-v-r SVM classifiers to other classifiers (Decision Tree (C5.0), K-nearest neighbour, neural network, and Support Vector Machine) show that the retrieval accuracy has increased to 91.4% for Rough Set 1-v-r SVM and 92% for Rough Set 1-v-1 SVM.

List of Publication Related to this Thesis

The following seven publications reported the results during the course of this research.

Journal

P1. Maryam Shahabi lotfabadi, Mohd Fairuz Shiratuddin, Kok Wai Wong. “Evaluation of Combining Support Vector Machine with Rough Set for Content Based Image Retrieval System.” International Journal of Computational Intelligence Research, Vol. 11, no. 1, pp. 37-47, 2015.

Conference Proceedings

P2. Maryam Shahabi lotfabadi, Mohd Fairuz Shiratuddin, Kok Wai Wong. “Content Based Image Retrieval Systems with a Combination of Rough Set and Support Vector Machine.” New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering Lecture Notes in Electrical Engineering, Springer International Publishing, Vol. 312, 2015, pp. 157-163.

P3. Maryam Shahabi lotfabadi, Mohd Fairuz Shiratuddin, Kok Wai Wong. “Utilising Fuzzy Rough Set based on Mutual Information Decreasing Method for Feature Reduction in an Image Retrieval System.” Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering Lecture Notes in Electrical Engineering, Springer International Publishing, Vol. 313, 2015, pp. 177-184.

P4. Maryam Shahabi lotfabadi, Mohd Fairuz Shiratuddin, Kok Wai Wong. "Evaluation of Fuzzy Rough Set Feature Selection for Content Based Image Retrieval System with Noisy Images." In Proceedings of the 22nd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2014), Plzen, Czech Republic, 2014, pp. 95-102.

P5. Maryam Shahabi lotfabadi, Mohd Fairuz Shiratuddin, Kok Wai Wong. "Feature Decreasing Methods Using Fuzzy Rough Set based on Mutual Information." In Proceedings of the Eighth IEEE Conference on Industrial Electronics and Applications (ICIEA 2013), Melbourne, Australia, 2013, pp. 1141-1146.

P6. Maryam Shahabi lotfabadi, Mohd Fairuz Shiratuddin, Kok Wai Wong. "Using Fuzzy-Rough Feature Selection for Image Retrieval System." In Proceedings of the 2013 IEEE Symposium On Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP 2013), Singapore, 2013, pp. 42-48.

P7. Maryam Shahabi lotfabadi, Mohd Fairuz Shiratuddin, Kok Wai Wong. "Using Rough Set Theory to Improve Content Based Image Retrieval System." In Proceedings of the Eleventh Postgraduate Electrical Engineering and Computing Symposium (PEECS2012), Perth, Australia, 2012.

Contributions of this Thesis

The contributions in this thesis, in which some have already been published and reported, are described below and summarised in Table 1.

A review of the different types of feature selection methods and a survey of various Content Based Image Retrieval Systems which use feature selection in their methodology have been completed. This work forms the basis of Chapter 2. Different parts of the work have been published in papers P1-P7.

The development of the improved pre-processing stage in the Content Based Image Retrieval system using feature selection forms a part of Chapter 3. Several proposed feature selection methods have been explored and compared. The results of this work have been published in conference proceedings papers P2 and P3, as well as conference papers P5 and P6.

The contribution in Chapter 4 is the development of the proposed pre-processing stage to handle noisy images in Content Based Image Retrieval Systems. The progress of the work, which includes algorithms, experimental results, comparison results and discussions, has been reported in conference paper P4.

In Chapter 5, two classifiers have been successfully developed to handle the Content Based Image Retrieval System problems. The experimental results in this chapter have shown significant improvements in terms of retrieval accuracy after these two classifiers have been implemented. Journal paper P1 has presented this chapter outcome.

Table 1: Summary of the Contribution of the Thesis

Chapter	Contributions	Paper No
Chapter 2: A Review of Feature Selection in Content Based Image Retrieval Systems	Presents a literature survey on previous research related to Content Based Image Retrieval Systems which used feature selection.	P1, P2, P3, P4, P5, P6, P7
Chapter 3: Fuzzy Rough Set Feature Selection in Content Based Image Retrieval Systems	Successfully developed a Content Based Image Retrieval System to reduce the semantic gap and work with vague and incomplete image features efficiently.	P2, P3, P5, P6, P7
Chapter 4: Fuzzy Rough Set Feature Selection in Content Based Image Retrieval Systems with Noisy Images	Successfully developed Content Based Image Retrieval Systems to handle noisy image features.	P4
Chapter 5: Evaluation of Combining Support Vector Machine with Rough Set for the Content Based Image Retrieval System	Successfully enhance the performance of the two kinds of Support Vector Machine Classification by Rough Set in the Content Based Image Retrieval System.	P1

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Overview

In recent years, the Content Based Image Retrieval (CBIR) system has become a focus of research in the area of image processing and machine vision (Cerra & Datcu, 2012). General CBIR systems automatically index and retrieve images with visual features such as colour, texture and shape (Z. Chen, Hou, Zhang, & Qin, 2012). However, current research found that there is a significant gap between visual and semantic features used by humans to describe images (Penatti, Valle, & Torres, 2012). In order to bridge the semantic gap, some researchers have proposed methods for managing and decreasing image features, as well as extracting useful features from a feature vector (Li, Fan, Wang, & Liu, 2012; Xing-yuan Wang, Chen, & Yun, 2012).

This chapter provides information about the CBIR system definition and its basic components, followed by a discussion of the CBIR systems problems. After that, the aims and objectives of this thesis are described. Next, the significance and contribution of the thesis are presented. At the end of this chapter, the outline of the thesis is shown.

1.2 Content Based Image Retrieval System

In a typical Content Based Image Retrieval (CBIR) system, a user submits an image based query, which is then used by the system to extract visual features from the images (Bird, Elliott, & Griffiths, 1996). The visual features may include shape, colour or texture, depending upon the type of CBIR system being used. These visual features are examined to search and retrieve similar images from an image database or various databases. The similarity of the visual features between the queried image and each image in a database is calculated based on their distance, by comparing the feature vectors of the two images (Cerra & Datcu, 2012). As a result of the image query, the CBIR system will then display the images which have the closest similarity, according to the pre-defined threshold value in the system. The pre-defined threshold value is usually set to restrict the number of results that the CBIR system displays (Iqbal, Odetayo, & James, 2012). A general CBIR system is shown in Figure 1.1 (adapted from (Penatti et al., 2012) and (Fanjie, Baolong, & Xianxiang, 2012)).

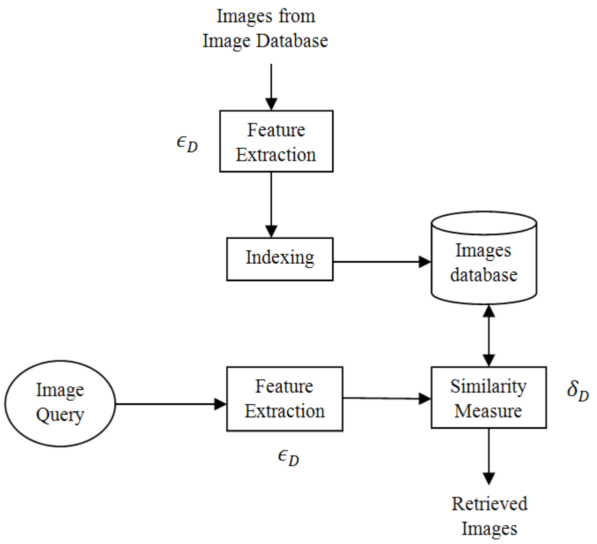


Figure 1.1: A typical Content Based Image Retrieval system.

Although CBIR systems have been widely researched on, there are still many challenges that need to be addressed, especially with the increasing amounts of images available. Different researchers propose various algorithms in order to address the problems of the CBIR system using Content Based approaches (Feng, Xiao, Zha, Zhang, & Yang, 2012; Iqbal et al., 2012; X.-y. Wang et al., 2012). Both the effectiveness and efficiency of the CBIR systems are dependent on the image descriptors that are being used. The image descriptor is responsible for characterising the image properties and computing their similarities. In other words, the image descriptor makes it possible to rank images according to their visual properties (Penatti et al., 2012).

1.2.1 Image Descriptor

The image descriptor is responsible for quantifying how similar two images are (Yildizer, Balci, Hassan, & Alhajj, 2012). An image descriptor D can be defined as a pair (ϵ_D, δ_D), where ϵ_D is a feature extraction algorithm and δ_D is a function suitable to compare the feature vectors generated (Penatti et al., 2012).

- ϵ_D encodes an image's visual properties into feature vectors (Figure 1.1). A feature vector contains information related to the image visual properties like colour, texture, shape and spatial relationship of objects.
- δ_D compares the two feature vectors. As shown in Figure 1.1, given two feature vectors, the function computes a distance or similarity value between these vectors. The distance or similarity between the vectors is considered as the distance or similarity between the images, from which the vectors were extracted.

1.2.2 Similarity Measure

Similarity measures are important for image descriptors. Their choice has a huge impact on the descriptor performance. The most common distance functions are L1, L2 and the Canberra Distance (Grana & Veganzones, 2012). L1 is also known as the Manhattan or City-Block Function, while L2 is also known as the Euclidean Distance (Grana & Veganzones, 2012). These three common functions (or variations of them) are widely used. Moreover, there are more complex functions like the Earth Mover's Distance (EMD), Angular Distance (Acharya & Devi, 2012), Czekanowski, Fu, Mahalanobis and the χ^2.

1.2.3 Feature Extraction

In a CBIR system, search and retrieval are carried out based on the visual contents of the image, instead of the text attributes such as Tags and Metadata. The important visual contents include colour, texture and shape features (Gavves, Snoek, & Smeulders, 2012).

The colour feature is a commonly used visual feature for CBIR. Colours play a major role in human perception (Yildizer, Balci, Jarada, & Alhajj, 2012). Some of the colour models available that can be used to represent images are HSI, HSV, LAB, LUV and YCrCb. The most commonly used colour model is RGB, where each component represents the colours red, green and blue respectively. The colour models, such as HSI and YCrCb, represent colour and illumination separately. There are different ways to use colour for CBIR purposes, namely by using a colour histogram, colour moment and colour coherence. The most effective method is by using a colour histogram (Iqbal et al., 2012). The colour histogram provides meaningful information for measuring the similarity between two images as it is robust against object distortion and the scaling of the object (Subrahmanyam, Maheshwari, & Balasubramanian, 2012a). Additionally, high effectiveness, simplicity, low storage requirements and real-time application possibility make it the best among others. Due to these characteristics, many researchers have started to use histogram-based colour image retrieval methods (Iqbal et al., 2012).

Texture is another important feature of an image that can be extracted for the purpose of image retrieval. Image texture refers to the surface patterns which show granular details of an image (Krishnamoorthi & Sathiya devi, 2012). It also gives information about the arrangement of different colours. For instance, the different patterns that can be seen in grass fields and block walls make them different from each other.

Two main approaches for texture features analysis, namely the structural and statistical approaches exist (Penatti et al., 2012). In the structural texture approach, the surface pattern is repeated (such as a floor design that contains the same pattern). Conversely, in the statistical texture approach, the surface pattern is not regularly repeated (such as different flower objects in a picture that normally contains similar properties, but are not exactly the same).

A co-occurrence matrix is a popular representation of the texture feature of an image. The texture of an image is an illustration of the spatial relationship of the grey level image (Iqbal et al., 2012). The co-occurrence matrix is constructed based on the orientation and distance between image pixels. Texture information can be extracted from an image using a co-occurrence matrix (Wang et al., 2012). There are many texture features that can be extracted from an image using a co-occurrence matrix such as entropy, contrast, energy and homogeneity. These features are represented as texture features and can be used for image retrieval purposes (Penatti et al., 2012). Texture characteristics that are visually useful for texture analysis include contrast, line-likeness, coarseness, directionality, roughness and regularity. These texture features are used in many CBIR systems (Krishnamoorthi & Sathiya devi, 2012).

Signal processing and wavelet transform methods are used in texture analysis. The wavelet transform is used for image classification based on the multi-resolution decomposition of images (Iqbal et al., 2012). Among the different wavelet transform filters, Gabor filters were found to be very effective in texture analysis. The Gabor filter is used in various types of applications due to its effectiveness in the area of texture-based image analysis. Two-dimensional Gabor filters are a group of wavelets. Many researchers have used the Gabor wavelet filter to extract texture features from an image. The Gabor filter is normally used to capture energy at a certain scale and orientation. Scale and orientation are the two most important and useful features that are used for texture analysis. The Gabor filter is also known as the scale and rotation invariant (Iqbal et al., 2012).

The shape feature plays a vital role in object detection and recognition of an image. In order to identify and recognise objects in an image, the object shape features provide robust and efficient information of the object (Iqbal et al., 2012). Shape features are also used to describe and differentiate objects in an image. There are two methods where shape features can be extracted from an image. They are the contour-based and region-based methods. Contour-based methods are normally used to extract the boundary features of an object's shape. Such methods will completely ignore the important features inside the boundaries (Cerra & Datcu, 2012; Iqbal et al., 2012). Region-based methods that rely on shape descriptors are able to extract both boundary and region features (Yildizer, Balci, Hassan, et al., 2012). Region-based methods use a moment-based theory such as the Hu, Legendre and Zernike moments (L. Chen, Huang, Tian, & Fu, 2014; Z. Liang, Zhuang, Yang, & Xiao, 2013). These moment-based theories provide valuable information to represent the shape of an image for feature extraction (Penatti et al., 2012).

The objectives of this research are to develop an improved approach to reducing image features and preserve significant ones from a huge amount of features and apply them to the CBIR system. In addition, this improved approach can reduce the semantic gap and improve the CBIR performance. It can also work in vague and uncertain areas. As seen in Figure 1.1, a modified feature selection step is applied after feature extraction. A complete definition of this feature selection step and the way it works is described in Chapter 3.

1.3 Problem Statements

Through the review of the literature, some of the problems identified with respect to CBIR systems are outlined below. For each problem, the scope of the research in this thesis will be highlighted. The research scopes are by no means definite, but they will serve as starting points to achieve the main aims and objectives of this research.

P1: Semantic gap

The fundamental problem of CBIR lies in capturing the concept of similarity. While users understand the meaning (semantics) of an image and evaluate similarity with respect to it, the search systems work with the low-level visual descriptors. The discrepancy between these two perspectives is referred to as the semantic gap. "The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.” (Z. Chen et al., 2012; D.-x. Li et al., 2012). This research will thus look for ways to improve the pre-processing stage, so as to reduce this problem.

P2: Huge amount of image features

Some CBIR systems extract many features such as colour, shape and texture from the images, and there is a problem to manage this huge amount of features. If there is a method that can extract the more significant features, it will be valuable (Iqbal et al., 2012; Krishnamoorthi & Sathiya devi, 2012). In this research, improved methods will be investigated in order to reduce the huge amount of features. This thus reduces the problem of space to work with.

P3: Incomplete and vague area in image description

Both the effectiveness and the efficiency of CBIR systems are very dependent on the image descriptors that are being used. The image descriptor is responsible for characterising the image properties and computing their similarities. In other words, the image descriptor makes it possible to rank images according to their visual properties. If there is a vague and incomplete area in the image descriptor, it needs to be handled properly (Penatti et al., 2012). In this thesis, improved methods will be investigated into, in order to deal with the vague and incomplete areas.

P4: Missing data in some image features

Bright and dark blotches that represent areas of missing pixels are the types of degradation that frequently appear on images. Often, the damage is so bad that the original image data is completely obliterated (Yangxi Li, Geng, Yang, Xu, & Bian, 2012; Yildizer, Balci, Jarada, et al., 2012). In this research, using improved methods of semantic rules could reduce this problem.

The aim of this research is to reduce or better still, solve the problems mentioned above. This research proposes an improved approach to the CBIR system by using the Rough Set. Rough Set is used to extract features from huge feature vectors so that fewer features (instead of all the features) will be used.

1.4 Aim and Objectives

The main aim of this research is to develop an improved approach for the CBIR system, apart from the deterministic and crisp methods.

The objectives of the research are as follows:

O1: Develop an algorithm in the pre-processing stage to enhance the handling of the semantic gap and at the same time, improve time and memory efficiency.

O2: The developed algorithm will reduce the retrieval cost by narrowing the search space and by reducing the dimensionality of the feature vector.

O3: Investigate the use of the Rough Set theory to deal with vague and incomplete image features.

O4: Develop a CBIR system using Rough Set for handling the uncertainty of the image feature vector. Uncertainty means cannot clearly define and not distinct an object from others. For example, when a system is trying to clearly define the exact border of cloud in the grey areas, the system cannot understand whether this area is the object or the background.

O5: Extract human understanding rules using the Rough Set theory for improving the handling of the interpretability of the CBIR system.

1.5 Significance and Contributions

The main contribution of this research is to improve the CBIR performance and reduce the semantic gap. Another contribution is the selection of a small but efficient and representative subset of the collected huge image features based on the Rough Set method. One of the important results of this research will be to introduce an appropriate framework for representing and processing, in the vague and incomplete areas with missing data. With the Rough Set, we can manage image features and extract semantic rules that can improve the CBIR performance. In addition, the accuracy of decision rules is important. The more accurate the decision rules, the higher the quality of the retrieval. However, the traditional retrieval techniques have not provided an accurate reflection of the decision rules, especially in the incomplete and vague areas.

The significance of this research is the creation of the novel framework for designing a successful retrieval system. The research shows the contributions in the retrieval domain, in that it presents new knowledge for image feature selection. These improved techniques borrow the concepts from the Rough Set theory and apply them to the image features. In addition, this research will attempt to prove that the Rough Set is another paradigm, besides deterministic and crisp methods, which can solve huge image feature problems. Consequently, because feature reduction is used in several fields (for example, image processing, data mining and image analysing), the improved methods can be applied broadly.

By employing a pre-processing phase, an improved CBIR system is proposed. In the pre-processing phase, the most important image features are selected by using Fuzzy Rough Set feature selection. Semantic rules are then generated with these features. After that, the Support Vector Machine (SVM) classifier is built using these semantic rules. Experimental results illustrate the effectiveness of this pre-processing phase.

Usually, a CBIR system is not comprehensive enough to deal with both noisy and non-noisy images. However, with a pre-processing phase, this improved CBIR system has good retrieval results with noisy images as well. In the testing phase, the user feeds the noisy queried image, instead of the normal query image, to the system. The system extracts the noisy queried image features and gives these features to the SVM classifier, which is built into the training phase. This classifier will then extract the relevant images based on the noisy queried image provided. Experimental results with the three kinds of noise (Gaussian noise, Poisson noise and salt and pepper noise) presented the effectiveness of this pre-processing phase with the noisy image features.

The two classifiers based on a combination of firstly, 1-v-1 (one-versus-one) Support Vector Machine (SVM) and Rough Set and secondly, 1-v-r (one-versus-all) Support Vector Machine and Rough Set, are presented. In the experiment, 10 semantic groups from the Corel image dataset were used, and two new classifiers were compared with the Decision Tree (C5.0), K-nearest Neighbour, Neural Network and

Support Vector Machine. It is shown that the Rough Set can enhance the overall performance in terms of retrieval accuracy. Also, it reduces the storage requirements for 1-v-1 SVM, training time and works better with noisy images for 1-v-r SVM.

1.6 Outline of the Thesis

The thesis is organised as follows:

Chapter 2 covers the existing CBIR systems which use feature selection in their methodology. This chapter also highlights the different feature selection methods and their limitations.

Chapter 3 discusses the proposed algorithm which performs feature selection utilising the Rough Set. It gives different comparisons of the proposed methods with other existing methods, so as to highlight the success of the work.

Chapter 4 presents the proposed algorithm for noisy query images. In addition, the chapter explains the different consequences of the proposed methods, discusses the evaluation results and highlights the merits of the proposed methods.

Chapter 5 investigates the Rough Set with two kinds of Support Vector Machine (SVM) classifier. These two kinds of SVM are one-versus-one SVM and one-versus-r SVM. These improved classifiers are used in the CBIR systems. The experimental results for the improved classifiers are compared to other classifiers given.

Finally, Chapter 6 gives the total consequences of the work and explains achievements attained during the execution of the work. It also explains future work which can be continued in the same domain.

Chapter 2: A Review of Feature Selection in Content Based Image Retrieval Systems

2.1 Introduction

The selection of relevant features and removal of non-relevant ones is one of the main problems in Content Based Image Retrieval (CBIR) systems (Dharani & Aroquiaraj, 2013). The success of CBIR systems is usually related to the quality of the data on which they work on. If the data contains redundant or irrelevant features, most CBIR systems may produce a less accurate result. Prior to this, feature selection tries to identify and remove as much irrelevant and redundant data as possible. This is important as working on a reduced number of features often benefits in terms of classification accuracy and learning speed (Vasconcelos, 2003).

The "curse of dimensionality" is another motivation for the selection of important features for images (Luo, Zhang, Fan, & Deng, 2001). It turns out that any two randomly picked feature vectors (independent of each other) in a high dimensional space will tend to have a fixed distance from each other, no matter the distance measured. (Dharani & Aroquiaraj, 2013; Luo et al., 2001). This means that even if most of the features are not associated, the task of a classifier could be complicated by the fact that the distances between positive examples are quite similar to the distances between the positive and negative examples. Feature selection could help to resolve this problem. Finally, there is a practical observation in which running most of the classifiers, such as support vector machine and neural network on a full feature set, causes the software to run out of memory (Najjar, Ambroise, & Cocquerez, 2003).

Unlike other dimensionality reduction methods, feature selectors preserve the original meaning (or semantics) of the features after reduction (Prasanna, Ramakrishnan, & Bhattacharyya, 2003). This has been found in applications that involve datasets containing huge numbers of features (i.e. in the order of tens of thousands), which would be impossible and difficult to process. In addition, feature selection methods have been applied to small and medium sized datasets, so as to locate the most informative features for later use. Most datasets will contain a certain amount of redundancy that will not help in the image retrieval process and may, in fact, mislead the process. Therefore, the aim of the feature selection is to find useful features to represent the data and remove non-relevant ones, which could also save the processing time.

Some other reviews on image database systems, image retrieval or multimedia information systems have been published in (Datta, Joshi, Li, & Wang, 2008; Dharani & Aroquiaraj, 2013; Remco C. Veltkamp & Tanase, 2002). However, none of them focused on providing a thorough investigation to just feature selections used in CBIR systems. The purpose of this chapter is to provide a review and overview of the feature selection methods and datasets used in CBIR. This chapter will also provide a related area in feature transformation methods in CBIR. CBIR systems that use feature selection have been reviewed, and a comparison between them has been carried out.

The remainder of the chapter includes the following. Section 2.2 presents an overview of the feature selection criteria; Section 2.3 presents different CBIR systems that use feature selection; Section 2.4 provides a summary of the limitations of the current methods; Section 2.5 provides a discussion and comparison of these systems and Section 2.6 summarise this chapter.

2.2 Overview of Feature Selection

The improvement of computational efficiency without losing the accuracy of CBIR systems can be executed by selecting the best features and decreasing the length of the feature vector (Alattab & Kareem, 2013). We can separate dimensionality reduction methods into two main groups: feature transform and feature selection (Guldogan & Gabbouj, 2008). The feature transform method, such as Principle Component Analysis (PCA) (Nikhil Naikal, Allen Y. Yang, & Sastry, 2011) and Independent Component Analysis (ICA) (Hoffmann., 2007), maps the original feature space into the lower dimensional space and constructs new feature vectors. The problem of feature transform algorithms is their sensitivity to noise and that the resultant features convey no meaning for the user (Turcot & Lowe, 2009). On the other hand, the feature selection method is robust against noise (Ganivada, Ray, & Pal, 2013) and the selected features are meaningful. The objective of feature selection is to pick up a subset of features, to reduce the length of feature vectors with the lowest information loss.

Feature selection is one of the important stages in CBIR systems and is used for enhancing semantic image retrieval results by decreasing the retrieval process complexity and improving the overall system efficiency. The overall procedure for any feature selection method is as shown in Figure 2.1 (Haiyu Song, Xiongfei Li, & Wang, 2010).

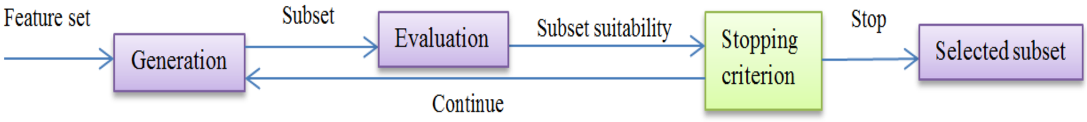


Figure 2.2: Feature selection

Referring to Figure 2.1, the generation procedure implements a search method that produces subsets of features for evaluation. It may start with no feature, all features, a selected feature set or some random feature subsets. Those methods that start with an initial subset usually select these features heuristically beforehand. In the first two cases, features are added (forward selection) or removed (backward elimination) iteratively. Features are either iteratively added or removed or in the last case, produced randomly. An alternative selection strategy is to choose instances and examine the differences in their features. The evaluation function computes the suitability of a feature subset produced by the generation procedure and compares this with the previous best candidate, replacing it if the existing subset is found to be better than the previous one.

Still referring to Figure 2.1, a stopping criterion is tested at every iteration to specify whether the feature selection process should continue or not. For example, such a criterion may halt the feature selection process, when a certain number of features have been selected based on the generation process. A common stopping criterion centred on the evaluation procedure is to halt the process when an optimal subset is reached. Once the stopping criterion has been met, the loop terminates.

There are a number of frameworks or models which can be employed for the feature selection task, and these can be broadly divided into three types as shown in Figure 2.2: filter, wrapper and embedded models respectively (Hammami, Ben Jemaa, & Ben-Abdallah, 2012).

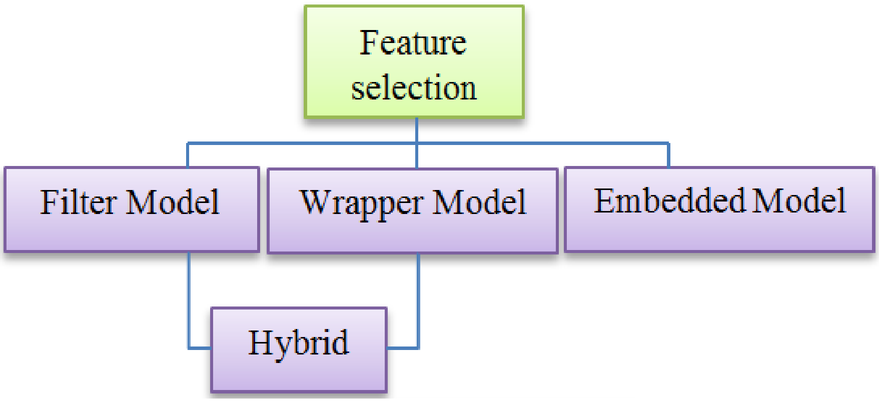


Figure 3.2: Feature selection models

The *Filter model* selects the best features according to some prior knowledge (independent measure). Each generated subset is then evaluated by an independent measure and compared with the current best subset. If, as a result of the evaluation, the generated subset offers an increase in the value of the independent measure, it will become the new best subset. The search continues until a pre-defined stopping criterion has been reached. After this, the best current subset is presented as an input to the classification algorithm (Hopfgartner, Urruty, Lopez, Villa, & Jose, 2010). The Filter feature selection model is shown in Figure 2.3. The Filter model is faster than the Wrapper model.

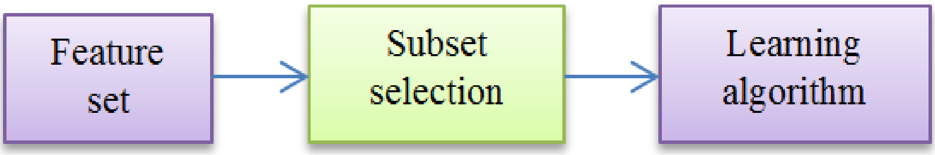


Figure 2.4: Filter feature selection model

The *Wrapper model* uses a search procedure in the space of possible feature subsets using some search strategy such as Sequential Forward Selection (SFS) or Sequential Backward Selection (SBS). In addition, various subsets of features are generated and evaluated. The evaluation of a specific subset of the features is obtained by a specific learning algorithm. The Wrapper model is used in conjunction with a learning or data mining algorithm, where the learning algorithm forms a part of the validation process. The Wrapper feature selection model is shown in Figure 2.4. This method is time-consuming but has better results compared to the Filter model (Kiktova-Vozarikova, Juhar, & Cizmar, 2013).

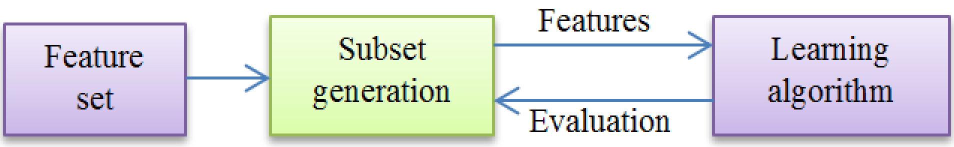


Figure 2.5: Wrapper feature selection model

The *Embedded model* tries to take advantage of the previous two models (filter and wrapper). Figure 2.5 shows the Embedded feature selection model. This model uses both a measure and learning algorithm to evaluate the feature subset. The measure is used to decide which subset is the best for a given cardinality. Moreover, the learning algorithm is used to select the final and best overall feature subset from a pool of feature subsets of different cardinalities (Hopfgartner et al., 2010). Decision trees are an example of the Embedded model (Kiktova-Vozarikova et al., 2013). Some direct advantages of feature selection are (Rashedi, Nezamabadi-Pour, & Saryazdi, 2013; Yi, Yihua, & Haozheng, 2012): (a) more rapid data mining algorithms convergence and (b) more accurate outcomes in the classification, clustering and similarity searching processes.

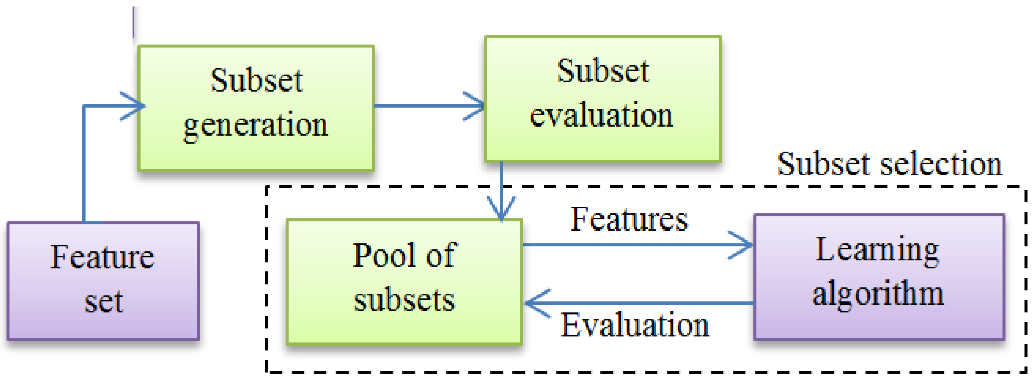


Figure 2.6: Embedded feature selection model

2.3 CBIR Systems using Feature Selection Methods

In this section, a number of CBIR systems that use feature selection or feature transform are reviewed. However, the main focus is feature selection. The stages of how each feature is selected or transformed are described for each CBIR system. This section is divided into two main sub-sections, CBIR systems which use: (1) feature transform methods and (2) feature selection methods.

2.3.1 Feature Transformation

In this sub-section, some CBIR systems which use feature transformation in their stages are described.

The Query By Image Content (QBIC) system implemented a method of retrieving images based on a rough user sketch (Carlton W. Niblack et al., 1993). QBIC was one of the first systems that applied multi-dimensional indexing to enhance the speed performance of the system. The average colour and texture features (both 3-dimensional vectors) are indexed using R*-trees. The 18-dimensional moment-based shape feature vector is first reduced using the Karhunen–Loeve Transform and then indexed by using R*-trees.

In (Buijs & Lew, 1999), edge maps of the images collected by web crawlers are obtained using the Sobel operator and a Gaussian blurring filter. A frequency histogram of the 3×3 binary pixel patterns occurring in the edge image, which is called the trigram vector, is computed for all images. This vector is subjected to a dimensionality reduction using a band-pass filter. Various other features, used in object matching, are taken at the pixel level: colour, laplacian, gradient magnitude, local binary patterns, invariant moments and fourier descriptors.

Another paper used Karhunen–Loeve Transform and band-pass filter as long as rotation, mirroring and intensity for achieving better results in CBIR system. In (Michael Lew Nies & Lew, 1996), there are three options for the pixel domain: the intensity image, the gradient image (obtained by Sobel operators) and the threshold gradient image. One feature vector is the horizontal/ vertical projection vector. For an image with $m \times n$ pixels, this vector has $m + n$ components computed as the average of the row/ column pixel values in the selected space. A second feature is the trigram vector, a frequency histogram of the 3×3 binary pixel patterns in the threshold gradient image. This 512-length vector can be subjected to a dimensionality reduction and to a component weighing scheme (low weights are applied on either end of the sorted frequency range). One way of reducing the length of the trigram vector is by forming groups of rotation, mirroring and/ or intensity invariant binary patterns. By using the Rotation and Mirroring (RM) group, the dimension of the feature vector is reduced to 102 and by forming the Rotation, Intensity and Mirroring (RIM) group, it is reduced to 51. Another method used consists of suppressing the contributions of the black and white (which are among the most common patterns) and rare patterns. This is called a band pass filter. A Karhunen-Loeve Transform can also be used for feature vector length reduction. However, this method has a problem with high dimensional data.

In (Sclaroff, Taycher, & La Cascia, 1997), the features used for querying are colour and texture orientation. The system computes distributions of colour and orientation over 6 sub-images (the whole image and 5 sub-regions: the central and corner regions). The result is an image index vector made of 2×6 sub-vectors. This dimension is subject to a reduction via a Principal Component Analysis (PCA) for each of the sub-vector spaces. Image colour histograms are computed in the CIE Luv colour space, and each histogram quantizes the colour space into 64 bins. The texture direction distribution is calculated using steerable pyramids. At each of the four levels of the pyramid, texture direction and strength for each pixel is calculated, resulting in an orientation histogram, quantized to 16 bins.

The paper by (Haiyu Song et al., 2010) discussed Region Based Image Retrieval System. The way that the system works is shown in Figure 2.6.

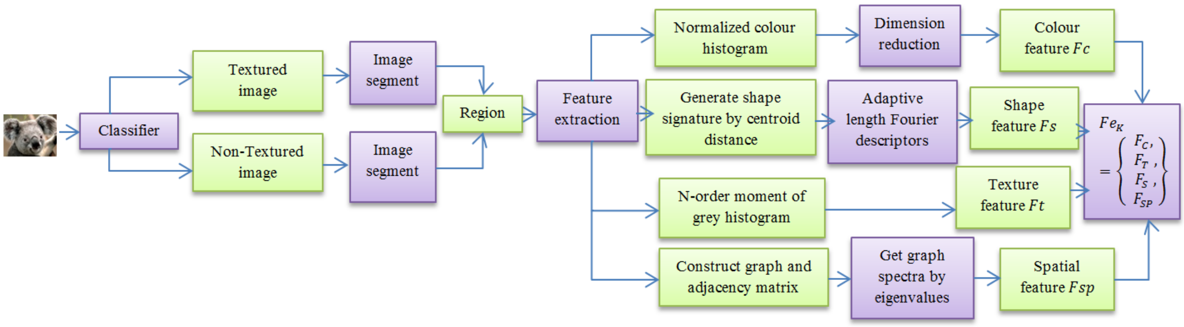

Classifier
Textured image
Non-Textured image
Image segment
Image segment
Region
Feature extraction
Normalized colour histogram
Dimension reduction
Colour feature Fc
Generate shape signature by centroid distance
Adaptive length Fourier descriptors
Shape feature Fs
N-order moment of grey histogram
Texture feature Ft
Construct graph and adjacency matrix
Get graph spectra by eigenvalues
Spatial feature Fsp
$Fe_K = \{F_C, F_T, F_S, F_{SP}\}$

Figure 2.7: Overview of Region Based Image Retrieval System (Haiyu Song et al., 2010)

The Isomap is used as a dimensionality reduction of the colour feature vector. The Isomap is a non-linear dimensionality reduction technique that uses MDS techniques with geodesic inter-point distances. Geodesic distances represent the shortest paths along the curved surface of the manifold. Unlike the linear techniques, the Isomap can discover the non-linear degrees of freedom that underlie complex natural observations. The 72-dimension colour feature can map to an 8-dimension by the Isomap. However, the weakness of the Isomap is that it is topology unstable.

2.3.2 Feature Selection

In this sub-section, the CBIR systems which use feature selection in their methodology are described. As described earlier in Section 2, feature selections are classified in three models: filter, wrapper and embedded. Every CBIR system is classified in one of these models according to their methodology.

2.3.2.1 Filter Model

Most of the CBIR systems which use feature selection in their methodology use the Filter model. The reason is that the Filter model is simpler and can thus be executed faster (Dharani & Aroquiaraj, 2013; Hopfgartner et al., 2010).

Some papers under the Filter model used soft computing methods like Genetic Algorithm (Valliammal & Geethalakshmi, 2012; Yi et al., 2012) and Mutual Information (Fei, Qionghai, & Wenli, 2006; Guldogan & Gabbouj, 2008; Maryam Shahabi Lotfabadi, Shiratuddin, & Wong, 2012b) as feature selections in their methodology. These papers are described below.

In (Fei et al., 2006), adaptive mixture models based on mutual information theory are adopted to determine the codebook size. In addition, a new method which can select combined feature axes is proposed. This paper addressed some problems mentioned in (Wei Jiang, Guihua Er, Qionghai Dai, & Jinwei Gu, 2006).

The technique in (Wei Jiang et al., 2006) can only select feature axes parallel to the original ones. A new algorithm which can select the combined feature axes effectively is needed. This paper suggests the following feature selection algorithm to solve this problem. The steps of this feature selection are as follows: Let N be the appropriate codebook size calculated before, the similarity of the "relevant" and "irrelevant" sets along the $i-th$ original axis be S_i and the $i-th$ column of the $N \times N$ identity matrix be $e_i = ([0 \dots 010 \dots 0]_{l\times N})^T$, $i = 1,2, \dots, N$. Then, the axes parallel to the original feature can be denoted as vectors:

$$a_i = e_i, i = 1,2, \dots, N$$

The combined axis a_{N+1} can be computed as follows:

$$a_{N+1} = \frac{1}{Z}\sum_{i=1}^{N} a_i \exp(-S_i)$$

where Z is a normalization factor to make $\|a_{N+1}\| = 1$. Consequently, the similarity of the relevant and irrelevant sets along each original axis is calculated and the first combined axis according to the above two equations is constructed.

Then, these lines are repeated for $k = 1,2,\ldots,K$

a) Select the optimal axis among the $N + 1$ axes.

b) Remove the worst axis among the $N + 1$ axes.

c) Construct a new combined axis according to second equation. Here, $a_i, i = 1,2,\ldots,N$ in the equation stands for the N feature axis left.

d) Update the sample weights and the obtained similarity results.

Another paper (Guldogan & Gabbouj, 2008) worked on two criteria for feature evaluation and a method for feature selection.

• A new criterion based on categorised member relations within the same cluster is used to label training data to better understand the description power of the feature for each cluster.

• A new criterion based on the discrimination power of the features calculated by using Pearson’s Product–Moment Correlation (PPMC) to define the correlations between different classes is also used.

The majority of votes on the results of these criteria, as a decision mechanism to select the appropriate features, is applied using mutual information, inter-cluster and inner-cluster relations.

Mutual Information (MI) calculates how much knowledge two variables carry about each other. It is the difference between the sum of the marginal entropies and their joint entropies. Two independent items always have zero mutual information. In particular, mutual information with Shannon's entropy is used in this paper.

A better understanding of the existing pattern in a given data space is the main objective of the inner-cluster analysis. The inner-cluster information is used as a criterion for feature selection in this research. The feature is considered descriptive for the cluster when the cluster is tight and compact, or the elements of a cluster are close to each other in the represented feature space. Inner-Cluster Relation (ICR), which represents the inner-cluster scatter information using the principal component information of the cluster, is proposed as a new measure for inner-cluster information. It is also related to the closeness of cluster elements similar to compactness as follows:

$$ICR = \frac{\Delta}{\frac{2}{N(N-1)}\sum_{i=0}^{N-2}\sum_{j=i+1}^{N-1} d(x_i, x_j)}$$

Assume d is the Euclidean distance between cluster members and N is the number of items within the cluster. x_i shows the feature vector corresponding to the $i-th$ item in the cluster. Δ is the distance between the best representative feature vector and mean vector.

In addition, majority voting is adopted in the feature selection process. Majority voting chooses the candidate with the biggest amount of votes. The voting method for sorting and selecting of features is used. Different from the categorization problem, the output of the decision-making black box is a list of features, which are sorted according to corresponding votes in descending order. The most important and powerful feature discriminating the associated data is shown by the first feature in the output vote list. The disadvantage of this work is the failure to achieve meaningful results. An overview of the feature selection system used in (Guldogan & Gabbouj, 2008) is shown in Figure 2.7.

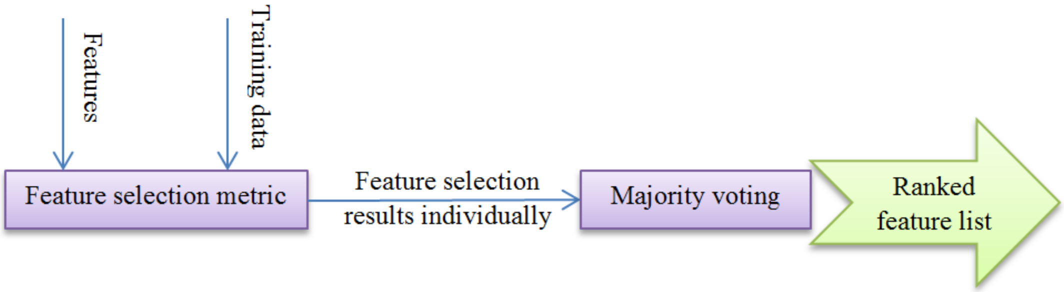


Figure 2.8: Feature selection system

The development and evaluation of a feature selection method for Content Based Image Retrieval (CBIR) computer-aided detection scheme based on multi-dimensional feature K-Nearest Neighbour (KNN) algorithm was the aim of (Yi et al., 2012). After examining the problems in tradition Genetic Algorithm (GA), it is found that there are usually different feature subsets when running GA for feature selection in different times; the reason for it is that the initial values for genes in GA are always generated randomly. The answer as to which feature subset could be selected as the optimal one still remains. As a result, this paper proposes a method for feature selection which is called Frequency-GA (F-GA).

An overview of CBIR CAD scheme based on multi-dimensional feature KNN algorithm is shown in Figure 2.9.

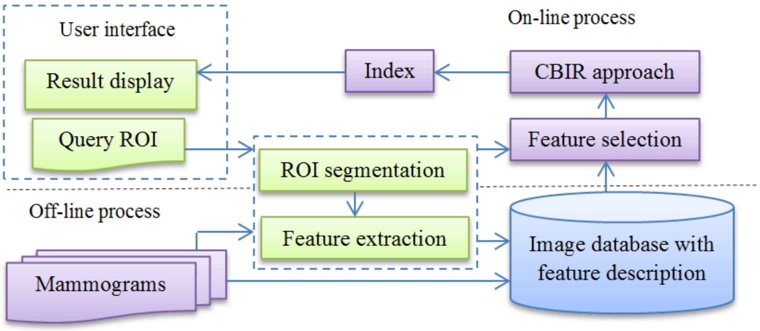


Figure 2.9: Diagram of CBIR CAD scheme based on multi-dimensional feature KNN algorithm

The feature selection step in Figure 2.9 is described as follows. Firstly, let the feature space be represented by F. The traditional GA is run m times, so that the m feature subset will be obtained $(F_1, F_2, \dots, F_m)$. After that, the emergence frequency of each feature which has been selected in $F_1, F_2, \dots, F_m$ would be accounted for. Those features which have the highest frequency (i.e. %*p*, *p* is a threshold) were selected to form the ultimate feature subset. Then, this feature subset is applied to the CBIR scheme which has been mentioned above.

A new system known as Plant Leaf Image Retrieval (PLIR) is developed for digitised images of plant leaves in (Valliammal & Geethalakshmi, 2012). The steps for this system are as follows: boundary extraction using median filter, feature fusion (shape, colour and texture feature) extraction, feature selection using Genetic Algorithm (GA) and an efficient classification using the Support Vector Machine (SVM).

GA optimises the parameters of all the shape, texture and colour features. GA is used as a search algorithm because the normal process for searching the features is computationally expensive. This paper does not provide information on how to use GA and the evaluation function used in GA.

In addition to Genetic Algorithm and Mutual Information, other researchers used statistical methods (Manikandan & Rajamani, 2008; Tsun-Wei, Yo-Ping, & Sandnes, 2009; Vendrig, Worring, & Smeulders, 1999) for their feature selection stage. Three papers are listed below.

The retrieval process in (Vendrig et al., 1999) is as follows: Let I_S be the image set after s reductions (filtering) and let F denote the set of 10 colour features described. The image set is clustered based on an automatically selected feature subset F_S of F. The images from I_S are ranked independently for each feature in F_S and each such ranking is divided into 4 clusters (corresponding to a reduction of 25%). Each cluster centroid is chosen as the cluster representative. The union of these representatives for all rankings forms the representative set of I_S, which will be shown to the user for the next reduction. The choice of feature subset F_S at stage s in the retrieval process is based on statistical analysis. For each feature, the variance of the feature values of all images is computed and F_S consists of the features with the highest variances that do not highly correlate with each other. This system fails to achieve meaningful results.

In (Manikandan & Rajamani, 2008), the CBIR system was constructed from two subsystems namely, the enrolment and query subsystems. The enrolment subsystem acquired the information that will be stored in the database for later use, while the query subsystem retrieved similar images from the database based on the user's query image.

The statistical tool (T-test) was calculated for all database and query images in this CBIR system. Besides, the power value is also computed for each image from this test. The images with power values less than 0.05 are considered as the selected images in the selection process. Otherwise, the images which have power values greater than 0.05 are rejected.

The T-test assumes that both distributions have identical variance and makes no assumptions as to whether the two distributions are discrete or continuous. In the case of the T-test, the null hypothesis is $\mu_1 = \mu_2$, indicating that the mean of feature values for class 1 is the same as the mean of feature values for class 2. The test determines if the observed differences are statistically significant and return a score representing the probability that the null hypothesis is true. Thus, the features can be ranked using either of these statistics, according to the significance score of each feature. However, the T-test has a problem with small datasets.

An entropy-based feature selection method for finding images of interest from the database is proposed in (Tsun-Wei et al., 2009). Six visual features (body, beak, flying, walking, colour and foot) are used to indicate birds and hence used to formulate search queries. According to the bird information ontology, the most relevant matches are retrieved.

A feature with larger entropy is more likely to reduce the result set. The entropy-based features elimination starts with the full feature set. The features with higher entropy are less relevant to the retrieval target, and they will be removed from the candidate features.

The feature selection strategy used in (Xu & Zhang, 2007) is as follows. The features used in the categorization model are generated from two stages: salient patch selection and feature extraction. The salient patch selection consists of three steps:

a) Salient patch detection: In this step, the local salient feature detector can detect salient patches. This detector finds regions that are salient over both scale and location. Features were detected and represented by intensity information. Once the regions are recognised, they are cropped from the image and rescaled to the size of a small pixel patch. Principal Component Analysis (PCA) is performed on the patches from all images because a high dimensional Gaussian is difficult to manage. Then, each patch is represented by a vector of the coordinates within the first 15 principal components.

b) Visual keywords construction: The visual keyword vocabulary is constructed by the 15-dimensional feature vectors. The vector quantization is performed on the vectors of all the images within one category, conducted by K-means clustering. Clusters of the vectors are the visual keywords for the category. A cluster histogram of salient patches shows its distribution, in which each bin corresponds to a visual keyword. Those visual keywords with large numbers of salient patches over a pre-determined threshold are selected because they are regarded as the most important features for the image category.

c) Noise exclusion: Two types of noise can be excluded. First, the most similar noise can be excluded by the Region Of Dominance (ROD). ROD is defined as the maximal distance between two patch clusters in the feature space. Second, the most non-common noises can be excluded by the salient entropy.

The Integrated Patch (IP) model is used to describe and categorise images based on the selected visual keyword. The appearance of the combination of the visual keywords, considering the diversity of the object or the scene, is showed by the IP model. This three-step feature selection method is shown in Figure 2.10.

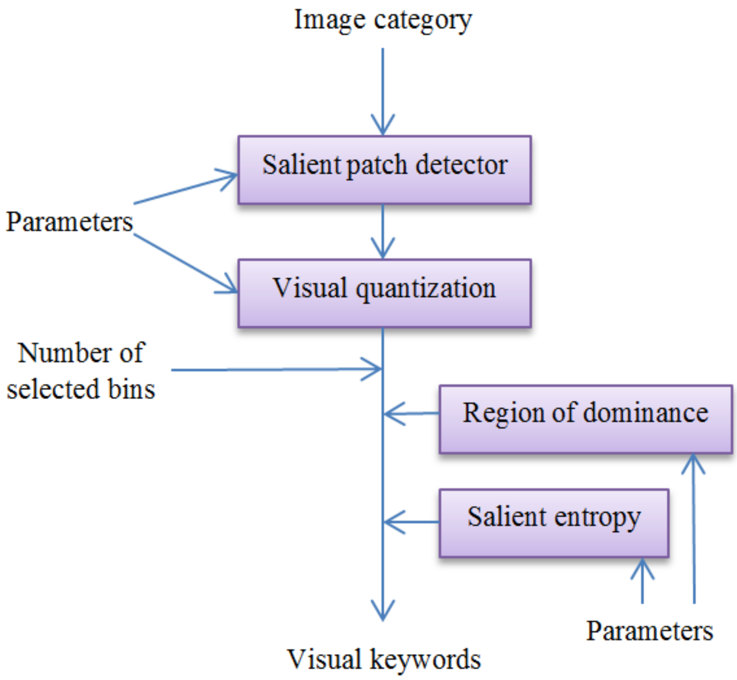


Figure 2.10: Three-step feature selection method.

In the feature extraction part, a 64-dimensional feature is extracted for each selected salient patch. For each new test image, the posterior probability is calculated for each category; after which, the label with the biggest probability is being assigned to it.

The image category can be chosen through the feature selection method and the IP model.

The aim of (Turcot & Lowe, 2009) is to select useful image features which are robust, using an unsupervised pre-processing phase. This step uses the Bag-Of-Words (BOW) framework and $tf - idf$ ranking. Firstly, image descriptors are extracted and quantized into visual words. Then, images in the database are matched against one another via $\mathrm{tf} - \mathrm{idf}$ ranking. Using epipolar geometry, the best $tf - idf$ matches are geometrically validated. After validation, geometrically consistent descriptors are labelled and retained, while all other descriptors are discarded. The validated image matches are stored in the form of an image adjacency graph, where matched image pairs are joined by an edge.

Scale Invariant Feature Transform (SIFT) is one of the local feature extraction methods. The disadvantage of this algorithm is the generation of hundreds of thousands of features per image, which seriously affects the application of SIFT in the CBIR system. For this reason, (Han-ping & Zu-qiao, 2011) proposed a method to select salient and distinctive local features using the integrated visual saliency analysis. According to this method, all the SIFT features in an image are ranked by their integrated visual saliency and only the most distinctive features will be reserved. The problem with this work is that it cannot work with vague and incomplete image features.

In (Abdolhossein Sarrafzadeh, Habibollah Agh Atabay, Mir Mosen Pedram, & Shanbehzadeh, 2012), the Relief-F feature selection is used in CBIR systems. The steps of this CBIR system are as follows. Firstly, a feature vector (for every image of the Coil-20 grey scale image dataset) using Legendre moments-based shape features is constructed. Then, a weight for each feature using the Relief-F algorithm is calculated, and the *k* top features as the effective subset are selected. After that, this subset using the accuracy of the SVM classifier is evaluated (one against one SVM used). Finally, the best subset is selected.

In (Alattab & Kareem, 2013), for semantic features selection and representation, a new method was described by the user directly, through appropriate verbal descriptions using the natural language concepts. A total of 100 respondents participated in selecting and weighing the semantic features of the human face, based on the level of importance of each feature. The semantic features were also integrated directly with the Eigen faces and colour histogram features for facial image searching and retrieval, so as to enhance retrieval accuracy for the user. The Euclidean distance was used for features-classes integration and classification purposes. The problem with this system is the rank features that leave the user to choose their own subset.

2.3.2.2 Wrapper Model

Besides the Filter model, many researchers used the feature selection-based Wrapper model for their CBIR systems. The main reason is the CBIR systems using the wrapper feature selection have better results (Acharya & Devi, 2012); however, they are more time-consuming (Datta et al., 2008).

A new hierarchical approach to Content Based Image Retrieval called the “Customized-Queries Approach” (CQA) is presented in (Dy, Brodley, Kak, Broderick, & Aisen, 2003). By the CQA, the query is classified as one of the disease categories using level 1 (described below) features and classifiers. After that, CQA retrieves n similar images within the query's class, using level 2 (described below) feature subset customised to the query's classified disease class. Euclidean distance is used as a dissimilarity measure for retrieval with each customised feature standard to variance one.

In level 1, those features that discriminate more accurately between the disease classes and achieve the highest classification accuracy are selected. C5.0 is chosen for this level.

In level 2, those features that have the most similarity within a single disease class are selected. Feature Subset Selection using Expectation Maximization clustering (FSSEM) is used at this level. Feature subset selection wraps around the clustering algorithm instead of a classifier. The basic idea of this approach is to search for the feature subset space, evaluating each subset F_t, by first clustering in space F_t, using Expectation Maximization (EM) clustering and then, evaluating the resulting clusters in space F_t, using the feature selection criterion. The result of this search is the feature subset that optimizes the criterion function.

The CBIR system used in (Kien-Ping, Chun-Che, & Kok-Wai, 2005) is shown in Figure 2.11. The feature selection method in Figure 2.11 is a simple statistical discriminant framework, which consists of the feature selection process, as well as the relevant feedback from the users. The discriminant ability of each image feature is analysed individually after the relevance feedback. The similarity measurement is calculated using the selected features in the next retrieval cycle. Using the ratio formula (as shown below), the discriminant ability of each image feature can be computed separately.

$$ratio = \frac{\left(\sum_{n=1}^{N_y} d'_n \Big/ N_y\right)}{\max(dp)} \quad d'_n = \begin{cases} \max(d_p), d_n > \max(d_p) \\ d_n, d_n \leq \max(d_p) \end{cases}$$

Assume d_p and d_n are the calculated distances of the respective positive and negative label samples from the positive centroid. The total number of negative samples is N_y. Firstly, the system calculates the ratio using the training samples. Image features are selected when their calculated ratio value is over the threshold value (pre-set by system). The selected image features are then cascaded into a flat feature vector once again and ready for another retrieval cycle.

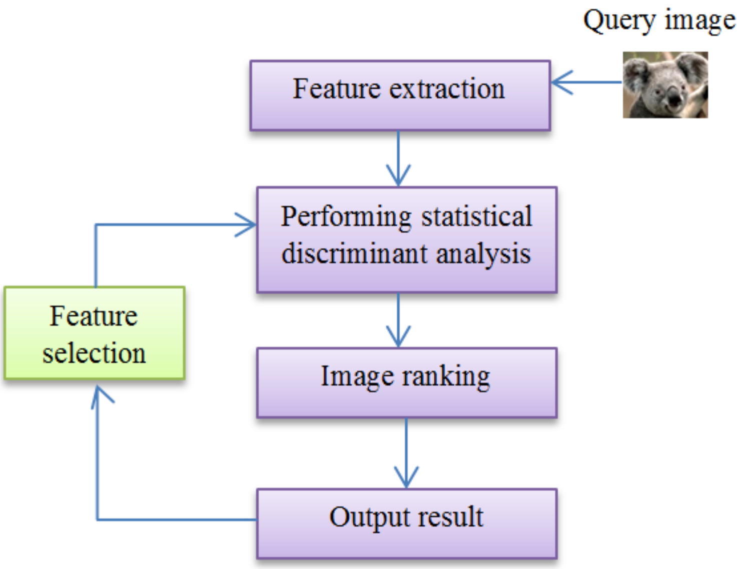


Figure 2.11: CBIR system used in (Kien-Ping et al., 2005)

Image data usually have many dimensions. Traditional clustering algorithms cannot deal with these dimensions appropriately. Therefore, the paper by (L. Wang & Khan,

2006) proposed a weighted feature selection algorithm as a solution to this problem. For a given cluster, the relevant features based on histogram analysis are determined. Furthermore, the relevant features are assigned greater weights as compared to less relevant features.

The weighted feature selection method is as follows. Firstly, visual tokens using K-means, assuming equal weight, are clustered together. Secondly, visual tokens distribute into clusters and the centroids update. Thirdly, the most important features are identified for each cluster, and irrelevant features are eliminated. Finally, until the algorithm converges, the same process will be repeated. In step 3, to determine the relevancy of a feature, the weighted feature selection is applied; that means the weight of the feature is determined.

In (Xun, Xian-Sheng, Meng, Qi, & Xiu-Qing, 2007), the region-based image retrieval formulated as a Multiple-Instance Learning (MIL) problem and the MI-AdaBoost algorithm were used to solve it. Using a certain set of instance prototypes, the algorithm maps each bag into a new bag feature space. After that, it adopts AdaBoost in the algorithm to select the bag features and build the classifiers simultaneously. This algorithm uses AdaBoost to select the bag features mapped by a certain set of instance prototypes. The instance prototypes are of two types: the instances from the negative (the clustering centres) and positive bags. The name of the proposed approach is MI-AdaBoost, while the linear classifier is used as a weak classifier.

Multimedia Content Description Interface (MPEG-7) is one of the most famous multi-media metadata standards. The paper by (Tianzhong, Jianjiang, Yafei, & Qi,

2008) represented low-level image features with all the MPEG-7 feature descriptors, including colour, texture and shape descriptors. Also, it used the Genetic Algorithm (GA), taking into account K-Nearest Neighbour (KNN) classification accuracy as a fitness function for feature selection. Four schemes are assumed to do the task: weight optimisation, the selection of optimum feature descriptor subset, weight optimisation followed by the selection of optimum feature descriptor subset and the selection of optimum feature descriptor subset followed by subset weight optimisation. When optimising weights, a real-coded chromosome GA and KNN classification accuracy as fitness functions are used. In the selection of feature descriptor subset, a binary one is used, and fitness function takes into consideration KNN classification accuracy, combining with the size of feature descriptor subset. In both genetic algorithms, KNN selects close images to the query. The drawback of this method is that the user must supply the threshold that determines when the algorithm should terminate.

Dynamic Region Matching (DRM) is proposed as a new region-based retrieval framework in (Ji, Yao, & Liang, 2008). The DRM system framework is as follows. Firstly, the image is partitioned into homogeneous regions, using the clustering-based segmentation algorithm. Moreover, a probabilistic fuzzy mapping approach is used to associate two images by the assembling of region similarity. Secondly, for constructing a semantic-sensitive feature set which reveals users' retrieval perception, users' relevance feedback results are integrated into an online AdaBoost algorithm. After that, using the feedback result, the region of significance in the query image is dynamically adjusted. In addition, the Region Of Interest (ROI) in the query image can be located after limited relevance feedback operations. Finally, to predict users' retrieval targets using their former operations, a long-term learning strategy is used.

The boosting feature selection algorithm, "FeatureBoost" (for adaptive relevance feedback learning) is used in this paper. Figure 2.12 shows the "FeatureBoost" system framework.

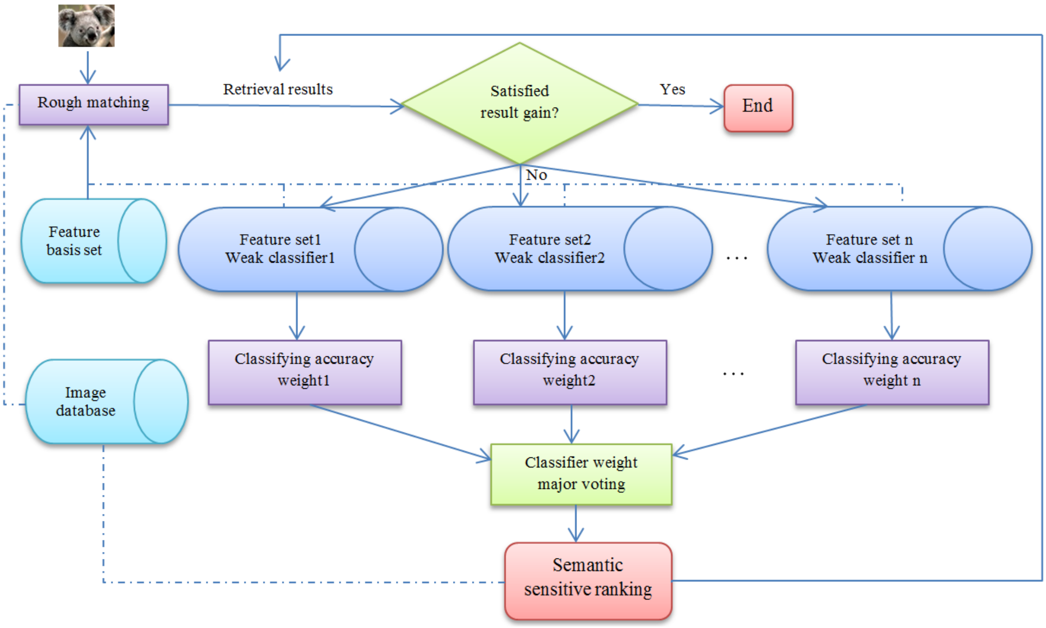

Rough matching
Retrieval results
Satisfied result gain?
Yes
End
No
Feature basis set
Feature set1 Weak classifier1
Feature set2 Weak classifier2
…
Feature set n Weak classifier n
Classifying accuracy weight1
Classifying accuracy weight2
…
Classifying accuracy weight n
Image database
Classifier weight major voting
Semantic sensitive ranking

Figure 2.12: FeatureBoost system framework

This method used users' feedback samples to construct n Eigen feature sets for n most representative features from the basis feature set. Using weighted majority voting, the features are combined together to form a sophisticated classifier.

(G. Chen & Wilson, 2008) used the Self-Organizing Map (SOM) based clustering for texture feature selection in the CBIR system. The SOM is used to group similar features and replace them by a single representative. The problem for this feature selection is that it is difficult to find the necessary parameters for turning the clustering algorithm into a specific application.

In (S. F. da Silva, Traina, Ribeiro, do E.S.Batista Neto, & Traina, 2009), an evaluation criterion (fitness function) that relies on order-based ranking evaluation functions for the feature selection ability of the genetic algorithms (GA) is presented. A ranking evaluation function provides a measure of the quality of a similarity query result. Order-based ranking evaluation functions which share the utility concept is employed in (S. F. da Silva et al., 2009). The utility of a relevant element is related to its position in the ranking, which means, the higher its utility, the higher its position in the ranking (more similar). If the utility score is set to zero, an element is not relevant (does not belong to the expected class).

In the training phase, the features extracted from the image training are sent to the feature selection process. Genetic algorithm based on the ranking evaluation function is used for feature selection process. In the test phase, the selected features by GA in the training phase will be used for indicating the images of the test set.

The selection operator in this study had two parts: selection for recombination and selection for reinsertion. The linear ranking selection method was used for the former. The best offspring $(S_p - 2)$ and two best parent-based fitness value was used for the latter selection (S_p is the population size). Uniform crossover and uniform mutation was used as crossover and mutation operations. The fitness function used in this paper is the same as mentioned in (Sérgio Francisco da Silva, Ribeiro, Batista Neto, Traina-Jr, & Traina, 2011).

In (Yu & Bhanu, 2010), the CBIR system is as follows: For a given query, the original features (colour, texture and shape) are extracted from the query image. Then, the K-Nearest Neighbour (KNN) algorithm with Euclidean metric searches the image database and retrieves the *N* top ranked images which have features closest to the query. When the user is satisfied with the retrievals, the session with this query is terminated. Otherwise, the user provides relevant feedback by labelling the retrievals as relevant (positive feedback) or non-relevant (negative feedback). A measure of inconsistency is computed based on the user feedback. Furthermore, it is given as the input to the feature selection, so as to select the feature subsets, which will guide the KNN search to obtain higher retrieval accuracy in the next CBIR iteration.

The feature selection procedure has two phases: searching the combination of feature subsets within a feature space using specified search strategy and evaluating the performance of the selected subset by a criterion. The Bayesian classifier is combined with the measure of inconsistency from relevance feedback to build the overall criterion for feature selection. The combined criterion is able to select the optimal feature subset. The feature selection diagram, along with the user feedback used in this paper, is shown in Figure 2.13.

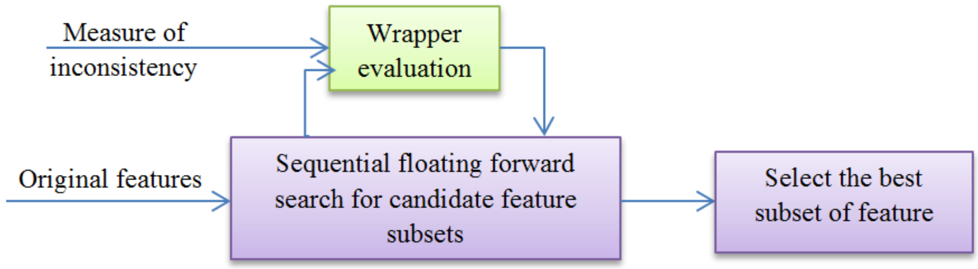

Measure of inconsistency
Wrapper evaluation
Original features
Sequential floating forward search for candidate feature subsets
Select the best subset of feature

Figure 2.13: The feature selection diagram with user feedback

A method to select texture features of Solitary Pulmonary Nodules (SPNs) which are detected by Computed Tomography (CT) is presented in (Zhu et al., 2010). Genetic algorithm-based feature selection technique is used to recreate multiple groups of feature subsets with different numbers of features. The reason for this work is because the performance of each classifier built by different groups of feature subsets is evaluated. The same number of features is used for each generation, and the fitness function is defined as the misclassification rate of a tenfold cross-validation procedure. In this method, the samples were randomly divided into ten groups; while one group was used as the test data, the rest of the samples were used to fit a multivariate normal density function. According to likelihood ratios, the test data were classified. The fitness function was calculated as the misclassification rate after each group had acted as a test group exactly once.

(Sérgio Francisco da Silva et al., 2011) utilises a genetic algorithm with an evaluation function based on the ranking concept, so as to perform feature selection for CBIR systems in the medical domain. This feature selection process employs supervised and wrapper strategy that searches for the best-reduced feature set. From a ranking evaluation function, three new fitness functions were proposed and evaluated. The problem of this method is the requirement of the user to state how many features are to be chosen.

(Rashedi et al., 2013) presented a simultaneous adaptive feature extraction and feature selection for the CBIR system. To extract the texture feature of images, two different analysis filters were applied to the columns and rows of the images. The parameters of these two wavelets were optimised by a Mixed Gravitational Search Algorithm (MGSA), also known as a swarm intelligence-based search technique.

Furthermore, feature selection is done with feature extraction in the chorus. That means, the parameters of feature extraction techniques are optimised to extract better features; meanwhile, the optimised group of features is also selected (see Figure 2.2). From Figure 2.14, it can be observed that both feature extraction and feature selection blocks are optimised simultaneously by a heuristic algorithm.

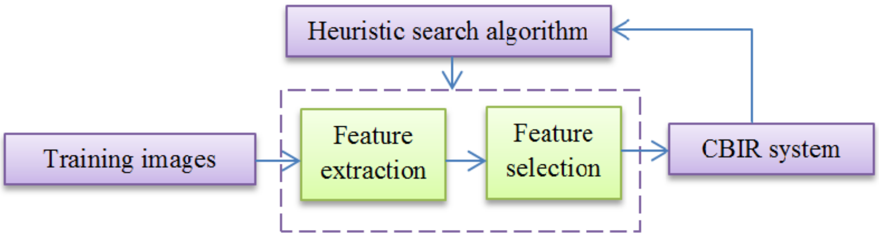


Figure 2.14: Simultaneous adaptive feature extraction and feature selection in the CBIR systems application

2.3.2.3 Embedded Model

Few CBIR systems use the embedded-based feature selection; the reason is that the search for an optimal subset of features is built into the classifier construction. Thus, most systems require the complicated simulation (Subrahmanyam, Maheshwari, & Balasubramanian, 2012b).

Relevance feature selection and classification are applied to the relevance feedback process (Prasanna et al., 2003). The stages of this method are as follows: a query image is provided to the CBIR system by the user. The set of *p* images that are classified as relevant return to the user (the number of images shown to the user is p). The user marks the images as relevant or irrelevant; these results are used for training a sparse classifier which classifies the images in the database. The images that are recognised as relevant by the sparse classifier are sorted in descending order of their distance to the separating hyper-plane. The most relevant images are farthest from the hyper-plane. The images which are shown to the user as feedback are p top images from sorted list have not been seen by user in previous iteration. The feedback obtained from this iteration and the previous iteration is combined to construct a new sparse classifier. The process of constructing a classifier, classifying images in the database, displaying the relevant images to the user and obtaining feedback is repeated until the p relevant images are farthest from the hyper-plane. The drawback of this method is that the algorithm termination needs to be determined by a threshold applied by the user.

Online feature selection in relevance feedback learning for region-based image retrieval systems is discussed in (Wei Jiang et al., 2006). Two criteria are used for the online feature selection. Firstly, Unified Feature Matching (UFM) measurement is used to calculate the similarity between the relevant and irrelevant image sets. Secondly, Fuzzy Feature Contrast Model (FFCM) is used for calculating the asymmetric similarity between images. The shorter version of this discussion can be found in (J. Wei, Guihua, Qionghai, Lian, & Yao, 2005).

The feature selection technique based on linear support vector machines is used in (Xin, Xin, & Hong, 2008) to select a low-dimensional feature subset from the original feature set. In addition, the relevance feedback technique used feature reweighting method to set suitable weights for each component of the selected feature subset. Finally, this feature subset with different weights is used to retrieve relevant images in the database, which are similar to query images submitted by the users. The problem is that when there are several highly correlated features, the linear support vector machine tends to pick only a few of them and removes the rest.

A hierarchical boosting algorithm based on feature selection for Synthetic Aperture Radar (SAR) image retrieval is used in (Mengling, Chu, Chao, & Hong, 2008). The statistics-based selecting method is used as the feature selection scheme in this paper. The principle of feature selection is making the inter-class dispersion large and the inner-class dispersion small.

In (Marakakis, Galatsanos, Likas, & Stafylopatis, 2009), a Relevance Feedback (RF) approach for CBIR is proposed. This is based on Support Vector Machines (SVMs) and uses a feature selection technique to reduce the dimensionality of the image feature space. The feature selection methodology called SVM Recursive Feature Elimination (SVM-RFE), is based on a recursive elimination of the less important features, based on the results of the classification of the training patterns using SVM classifiers. The problem is that it requires the user to state how many features are to be chosen.

The CBIR system used in (ElAlami, 2011) includes three steps: feature extraction from images database, feature discrimination and feature selection. The Sequential Forward Selection (SFS) technique is used in this paper. The first stage eliminated the features with dominant values. These features have less information and a slight effect on the classification. In addition, they can be ignored according to a certain threshold level. The most relevant features from the original features set are determined in the second stage via the calculation of certain evaluation function.

This function (E_S) depends on the calculation of the correlation and conditional probability between feature-to-feature and feature-to-target classes.

(D. Wang, Yuchun, & Binbin, 2011)) used a criterion function to measure the coherence between the metrics of the machine in the low-level feature and the subjective user and used it as the target function in the feature selection. After that, four feature selection methods are constructed. The Minimal-Redundancy-Maximal-Relevance criterion (MRMR) based on mutual information is a filter feature selection method. Three wrapper feature selection methods are Best Individual (BI), Sequential Forward Selection (SFS) and Plus-l-Minus-r (l-r).

In (Vavilin & Jo, 2013), the CBIR system is based on both appearance and contextual feature analyses. Thus, the training phase includes selecting the optimal appearance features and analysing the semantic relations between objects of different classes, so as to fix the context dependences. The classes are also analysed to find possible subclasses. The appearance and feature selection process is divided into three steps. The classes are analysed separately in the first step. Also, in this step, the properties which characterise each of the classes and subclasses are selected. The problem of feature subset selection is solved by Genetic Algorithm (GA), while the subclass detection is made using Fuzzy C-mean clustering. In the second step, the feature sets obtained are checked for their abilities to separate different classes. In order to select the best features for class separation, the classes are analysed pair-wise. Finally, the second step is repeated for the subclasses detected in the first step, in order to check the subclass hypothesis and improve the performance of class separation.

2.4 Limitation of Current Methods

The use of user-supplied information is essential to many existing algorithms for feature selection in the literature. This however, is a significant drawback. Some feature selectors require noise levels to be specified by the user beforehand, while others simply rank features, leaving the user to choose their own subset (Alattab & Kareem, 2013; Ji et al., 2008). Also, there are those that require the user to state how many features are to be chosen (Abdolhossein Sarrafzadeh et al., 2012; Sérgio Francisco da Silva et al., 2011; S. F. da Silva et al., 2009; Marakakis et al., 2009) or a threshold that determines when the algorithm should terminate (Prasanna et al., 2003; Tianzhong et al., 2008; Yi et al., 2012). All of these require the user to make a decision based on one's (possibly faulty) judgement.

In fact, for problems such as image retrieval where: a) there are usually large numbers of visual classes to be processed, b) the data is non-Gaussian and non-homogeneous (and the assumption of any parametric model is, therefore, highly restrictive) and c) there is a need to process very large training sets as traditional feature selection strategies either 1) simply fail to achieve meaningful results (Guldogan & Gabbouj, 2008; Vendrig et al., 1999) when the dataset is small (Kien-Ping et al., 2005; Manikandan & Rajamani, 2008) or when the number of interacting features are small (Valliammal & Geethalakshmi, 2012; Zhu et al., 2010) or 2) take an unrealistic (and practically infeasible) time to compute (Buijs & Lew, 1999; Guldogan & Gabbouj, 2008; Tsun-Wei et al., 2009; Vendrig et al., 1999). Some of these limitations, such as the dependence on particular probabilistic models (Yu & Bhanu, 2010), have been eliminated by recent advances in machine learning and hence, have already demonstrated successful applications (Rashedi et al., 2013;

Maryam Shahabi Lotfabadi et al., 2012b; Vavilin & Jo, 2013). On the other hand, these solutions also exacerbate some of the limitations enumerated above, namely the unavailability of practical extensions to problems with more than two classes and an immense training complexity. Due to this inherent lack of scalability, most existing feature selection methods are not applicable to large-scale problems, such as retrieval or recognition (Vasconcelos & Vasconcelos, 2004).

Many old (before the year 2000) CBIR systems used PCA and Karhunen-Loeve Transform dimensional reduction in their methodologies (Carlton W. Niblack et al., 1993; Michael Lew Nies & Lew, 1996; Sclaroff et al., 1997; Swets & Weng, 1995). However, there are also some limitations in their approach. Specifically, the size of the covariance matrix is proportional to the dimensionality of the data points. As a result, the computation of the eigenvectors might not be feasible for very high dimensional data. In addition, the problem of some feature selection methods, as mentioned in (Mengling et al., 2008), is the sensitivity to the dimension of the feature vector.

Some limitations in (Xu & Zhang, 2007) are as follows. Firstly, the discriminative features between the image categories are not well-leveraged. Secondly, the EM algorithm may lead to a local maximum.

The weakness of the Isomap in (Haiyu Song et al., 2010) is that it is topological unstable. The Isomap may construct erroneous connections in the neighbourhood graph. Such short-circuiting can severely impair the performance of the Isomap. Holes in the manifold are another problem of the Isomap. A third weakness is that the Isomap can fail if the manifold is non-convex.

As mentioned in (Dy et al., 2003), a drawback of EM clustering is the sensitivity of noise in the dataset. The noise makes it difficult for the algorithm to cluster an image to its suitable cluster. This will affect the results of the algorithm. In addition, an overfit with the presence of noise is a drawback of AdaBoost (Xun et al., 2007). Also, as stated in (Valliammal & Geethalakshmi, 2012) and (Zhu et al., 2010), it cannot deal with even small amounts of noise.

There are some drawbacks in (Xin et al., 2008). First, the results showed that the application of the proposed approach is limited. Second, when there are several highly correlated features, the linear SVM tends to pick only a few of them and removes the rest.

Some limitations of the Self-Organizing Map (SOM) which is used as feature selection in (G. Chen & Wilson, 2008) are as follows. It is difficult to find the necessary parameters for tuning the clustering algorithm to a specific application. Next, the distance measurement is task-dependent (although this drawback is not unique to the SOM). Last but not least, the SOM cannot handle random errors, which is an inherent drawback of neural networks.

In (Turcot & Lowe, 2009), researchers used just one method to select a restricted number of large-scale features for the treatment of unmatched images. However, the researchers of the paper recommended an exploration and use of other methods to improve the selection of features from unmatched images.

2.5 Summary

The objective of feature selection in CBIR systems is to reduce complexity, improve accuracy and retrieval results, as well as particularly, to improve semantic performance.

The properties used in the CBIR systems reviewed in this chapter are listed in Table 2.1. The properties are classified into the feature selection methods that each system used in their methodology, the datasets used in the reviewed systems and the year of the publications respectively. The summary table provides a quick glance of the feature selection methods and datasets that are popular in the CBIR research over the years listed.

Table 2.1: Summary of the 38 Content Based Image Retrieval Systems which use feature selection

Reference	Feature selection method	Type of data sets	Year
Carlton W. Niblack, Ron Barber et al. (Carlton W. Niblack et al., 1993)	Karhunen-Loeve Transform	Photo clip art subject matter (building, people, landscape, animal and etc.)	1993
Michael Lew Nies and Lew. (Michael Lew Nies & Lew, 1996)	Rotation, mirroring, intensity, band pass filter, Karhunen-Loeve Transform	Database of 16505 images of photographs taken from 1860 till 1914	1996
Sclaroff, Taycher et al. (Sclaroff et al., 1997)	Principle Component Analysis (PCA)	World wide web	1997
Buijs and Lew (Buijs & Lew, 1999)	Band pass filter	World wide web	1999
Vendrig, Worring et al. (Vendrig et al., 1999)	Statistical analysis	World wide web	1999
Dy, Brodley et al. (Dy et al., 2003)	Feature Subset Selection using	Medical images (lung images)	2003

	Expectation-Maximization clustering (FSSEM)		
Prasanna, Ramakrishnan et al. (Prasanna et al., 2003)	Sparse classifier	Corel database	2003
Kien-Ping, Chun-Che et al. (Kien-Ping et al., 2005)	Statistical discriminant analysis	Corel database	2005
Wei Jiang, Guihua Er et al. (J. Wei et al., 2005; Wei Jiang et al., 2006)	Online feature selection algorithm using FFCM and UFM	Corel database	2005 - 2006
Fei, Qionghai et al. (Fei et al., 2006)	Online feature selection	Corel dataset	2006
Wang and Khan. (L. Wang & Khan, 2006)	Weighted feature selection mechanism	http://corel.digitalriver.com	2006
Xun, Xian-Sheng et al. (Xun et al., 2007)	AdaBoost	Corel dataset + Musk dataset	2007
Xu and Zhang (Xu & Zhang, 2007)	Three step feature selection (Local salient feature detector + visual keyword construction + noise exclusion)	Corel dataset	2007
Chen and Wilson (G. Chen & Wilson, 2008)	Self-Organizing Map	Did not mention	2008
Guldogan and Gabbouj (Guldogan & Gabbouj, 2008)	Mutual information, inter-cluster and inner-cluster relations, a majority vote	Corel database	2008
Ji, Yao et al. (Ji et al., 2008)	FeatureBoost	Corel dataset	2008
Manikandan and Rajamani (Manikandan & Rajamani, 2008)	T-test	Fifty ultrasound kidney image	2008
Mengling, Chu et al. (Mengling et al., 2008)	Statistical based selecting method	KTH_TIPS image database PiSAR images of Japan area.	2008
Tianzhong, Jianjiang et al. (Tianzhong et al., 2008)	Real and binary Genetic Algorithm + KNN (4 different schemes)	Corel dataset	2008
Xin S, Xin et al. (Xin et al., 2008)	Linear support vector machine	SAR database	2008
da Silva, Traina et al. (S. F. da Silva et al., 2009)	Fitness function based on ranking evaluation	1. Breast Imaging Reporting and Data	2009

	function in genetic algorithm	System of the Department of Radiology of University of Vienna 2. Digital Database for Screening Mammography of the University of South Carolina 3. Mammograms images collected in the Clinical Hospital of University of Sao Paulo at Ribeiro Preto.	
Marakakis, Galatsanos et al. (Marakakis et al., 2009)	Support Vector Machine Recursive Feature Elimination	No mention of the source.	2009
Tsun-Wei, Yo-Ping et al. (Tsun-Wei et al., 2009)	Entropy-based	Bird species in Taiwan	2009
Turcot and Lowe. (Turcot & Lowe, 2009)	Unsupervised pre-processing step	Oxford building dataset + Flickr dataset	2009
Haiyu Song, Xiongfei Li et al. (Haiyu Song et al., 2010)	Isomap	Natural images from the Internet	2010
Yu and Bhanu. (Yu & Bhanu, 2010)	Measure of inconsistency + Bayesian classifier	Butterfly image database (http://janzen.sas.upenn.edu/) + Google images	2010
Zhu, Tan et al. (Zhu et al., 2010)	Genetic algorithm	Computed tomography (CT) lung images	2010
da Silva, Ribeiro et al. (Sérgio Francisco da Silva et al., 2011)	Genetic algorithm	1. Digital Database for Screening Mammography of the University of South Carolina 2. Mammogram images collected in the Clinical Hospital of University of Sao Paulo at Ribeiro Preto.	2011

		3. Computed Tomography (CT) of lunge exams.	
ElAlami (ElAlami, 2011)	Two-step feature selection (preliminary and deeply reduction)	Two nature datasets	2011
Han-ping and Zu-qiao (Han-ping & Zu-qiao, 2011)	Integrated visual attention analysis	ALOI image dataset+ Caltech 256 photo gallery	2011
Wang, Yuchun et al. (D. Wang et al., 2011)	Criterion function	FERET database	2011
Sarrafzadeh, Agh Atabay et al.(Abdolhossein Sarrafzadeh et al., 2012)	ReliefF algorithm	Coil-20	2012
Valliammal and Geethalakshmi(Valliammal & Geethalakshmi, 2012)	Genetic algorithm	Planet leaf image dataset	2012
Yi, Yihua et al. (Yi et al., 2012)	Genetic Algorithm	Digital Database for Screening mammography	2012
Alattab and Kareem (Alattab & Kareem, 2013)	Verbal description	Olivetti Research Lab (ORL) database + 1500 local Facial images of 150 participants from the University of Malaya (UM) in KualaLumpur, Malaysia.	2013
Rashedi, Nezamabadi-Pour et al. (Rashedi et al., 2013)	Binary gravitational search algorithm from Heuristic search algorithm	Corel dataset - Nature	2013
Vavilin and Jo (Vavilin & Jo, 2013)	Genetic algorithm + Fuzzy C-mean	Outdoor scenes (urban and nature)	2013

From Table 2.1, it is obvious that recent CBIR systems used soft computing methods as a feature selection in their methodology. Among these soft computing methods, Genetic Algorithm is used more like a feature selection as compared to the others (7 systems) in the CBIR systems. Following that, the Support Vector Machine is observed to be used in 2 CBIR systems.

Hybrid feature selection (Maryam Shahabi Lotfabadi et al., 2012b; Wei Jiang et al., 2006; Yu & Bhanu, 2010) and feature selection methods which have more than one step are used more often in recent years (ElAlami, 2011; Xu & Zhang, 2007). The reasons are when two or three methods are combined, the drawback of one method can be handled by the other method(s). This will, in turn, provide better and more accurate results.

From the analysis of the literature review regarding the datasets used, it was discovered that in the early years, before 2000, researchers usually work on images collected from the web. This could be due to the reason that there are not many commonly available data, so they can only rely on web images. From Table 2.1, it is observed that most of the CBIR systems used the Corel dataset for their experimental results (around 12 systems). As such, the Corel dataset is a more popular and perfect dataset for experimental results in CBIR systems because it is readily available and contains a wide range of subject matter. From these results, using the Corel dataset in the CBIR system has these benefits. Firstly, finding the other CBIR systems which use the Corel dataset for comparing purposes is easy. In addition, the dataset plays an important role in the retrieval process. Moreover, in order for a fair comparison, it is essential to find the same dataset for comparing the CBIR systems with each other so as to build the same situation.

Various medical datasets are popular too. Around 6 CBIR systems used mammography, kidney and lungs datasets for their experimental results. Approximately 8 CBIR systems used datasets containing nature, animals and buildings images. In addition, two systems used facial datasets, while two other systems used satellite images. Moreover, one system used greyscale images (Coil-20) for their experimental results.

The Corel dataset contains different semantic groups from nature, animals and buildings; in addition, 8 other CBIR systems used animal, nature and building images, while some systems which used the World Wide Web contained nature images. In general, around 24 out of 38 CBIR systems used outdoor images for their experimental results. Consequently, from Table 2.1, it can be observed that outdoor datasets and images are more popular for CBIR systems, which use feature selection in their methodology.

The advantages of using feature selection methods are summarised as follows:

- Reduction of data to fewer dimensions. Data visualisation is facilitated through the reduction of data. This allows the trends within the data to be more easily identified; this can be important when a few features may have an influence on data outcomes.
- Reduction in measurement and storage requirements. In domains where features correspond to particular measurements, fewer features are highly desirable, due to the expense and time-cost of taking such measurements.
- Reduction of training and utilisation times. With smaller dimensions, the run times of the learning algorithms improve significantly for both training and classification phases.

- Improvements in prediction performance. The classifier accuracy can be increased as a result of feature selection, through the removal of noisy or misleading features.

Additionally, for those methods which extract knowledge from data (e.g. rule induction technique), feature selection allows the readability of the model through knowledge discovery. When induction algorithms are applied to reduce data, the resulting rules are more compact. A good feature selection method will remove unnecessary attributes which affect both rule interpretation and prediction performance.

2.6 Conclusion

The main aim of feature selection is to determine a minimal feature subset from an image, which can be used to represent the original image features. In many real world problems like Content Based Image Retrieval Systems, feature selection is an important method that helps to remove noisy, irrelevant or misleading features. For example, by removing these features, learning techniques can improve their accuracy. This chapter provides a review of the different feature selection methods used in CBIR systems.

In this chapter, CBIR systems which use dimensionality reduction in their methodology are reviewed. The choice of a dimensionality reduction method plays a critical part in the success of a CBIR system. Dimensionality reduction includes features transform and feature selection methods. Feature selection basically narrows the semantic gap by selecting the feature subset that best represents the query and discards the redundant features. CBIR systems use the selected feature subset to search the image database, such that the retrieved images will be closer to a given query.

Sometimes, high-dimensional complex phenomena can be governed by significantly fewer and simple variables. The process of feature selection here will act as a tool for modelling these phenomena, thus improving their clarity. There is often a significant amount of redundant or misleading information present; this will need to be removed before any further processing can be carried out. In a CBIR system, the problem of deriving classification rules from large datasets is often managed by a feature selection pre-processing step. Not only does this reduce the time required to perform induction, but it also makes the resulting rules more comprehensible; hence, increasing the resulting classification accuracy.

From this chapter, it is observed that researchers used soft computing methods more than other methods (statistical) such as feature selection for CBIR systems. In addition, the Corel database is the most popular dataset used for research in generating experimental results.

Chapter 3: Fuzzy Rough Set Feature Selection in Content Based Image Retrieval Systems

3.1 Introduction

Feature selection refers to the problem of selecting input features that are more predictive of a given outcome (Foithong, Pinngern, & Attachoo, 2012). Feature selection is used in many areas such as image processing, machine learning, pattern recognition and signal processing (Y.-S. Chen, 2012). Unlike other dimensionality reduction methods, feature selectors try to preserve the original meaning of the features after reduction (Zhu et al., 2010). Besides applying to large datasets, feature selection methods have also been applied to small and medium-sized datasets, to locate more informative features for later use (J. Wang, Hedar, Wang, & Ma, 2012). Feature selection is an important step in processing the images especially for applications such as Content BasedImage Retrieval (CBIR) (Guldogan & Gabbouj, 2008). In large multimedia databases, it may not be practical to search through the entire database to retrieve similar images from a query. Good data structures for similarity search, and indexing is needed, as the existing data structures do not scale well for the high dimensional multimedia descriptors (Foithong et al., 2012). Thus, feature selection is an important step.

The main objective of this chapter is to develop and apply a different pre-processing stage to overcome problems as mentioned in Section 1.3 and improve the CBIR system with confidence. The Fuzzy Rough Set is applied as a feature selection to the pre-processing stage. It is specifically chosen for use in the pre-processing stage because of its particular characteristics. It can select important features from a massive number of image feature vectors and omit features which are not important. After all, redundant features could usually influence a further analysis in the wrong direction. Consequently, from these significant features, semantic rules that can classify the images more accurately are then extracted, so as to show more relevant images to the user; hence, improving the retrieval performance.

The use of the Fuzzy Rough Set theory in feature selection is one approach that has been explored in the last decade (Derrac, Cornelis, García, & Herrera, 2012; Ganivada et al., 2013). Fuzzy Rough feature selection can provide promising results mainly due to the following (Derrac et al., 2012; Hu, Zhang, An, Zhang, & Yu, 2012): (1) only the facts hidden in the data are analysed, (2) no additional information about the data such as thresholds or expert knowledge on a particular domain is required, (3) it finds a minimal knowledge representation, (4) Fuzzy Rough Set feature selection can deal with continuous dataset, (5) it has good results in noisy datasets, (6) the Fuzzy Rough Set with its lower and upper approximations can work better in vague and uncertain areas in the image feature vector and (7) rules extracted from the Fuzzy Rough Set are semantic. The aim of this thesis is to use these properties to overcome CBIR systems problem as mentioned in Section 1.3.

In order to provide an insight to the use of Fuzzy Rough feature selection in image retrieval, this thesis provides a comparison study with other feature selection methods. In this thesis, eight feature selection methods are compared with the Fuzzy Rough Set method. These eight feature selection methods are Gain Ratio, Genetic Algorithm, Information Gain, Isomap, Kernel PCA, OneR, Principal Component

Analysis (PCA) and Relief-F. The feature selection methods selected in this research are from Entropy, Statistical, Soft Computing, Decision Tree, Euclidean Distance, Kernel, Geodesic Distance and Nearest Neighbourhood based feature selection methods. Information Gain and Gain Ratio are entropy-based feature selection methods (Baranidharan & Ghosh, 2012; Chinpanthana, 2011). Isomap, PCA, Kernel PCA and OneR are based on Geodesic Distance, Euclidean Distance, Kernel (Yossi Rubner, Jan Puzicha, Carlo Tomasi, & Buhmann, 2001) and Decision Tree (Chinpanthana, 2011) respectively. In addition, Relief-F is based on its nearest neighbour (Abdolhossein Sarrafzadeh et al., 2012). Genetic Algorithm is an example of one of the soft computing methods, which is the same as the Fuzzy Rough Set. From Chapter 2 (the literature review chapter), it is observed that the Genetic Algorithm is used for feature selection in CBIR systems in many recent types of research. Consequently, it is essential to compare our proposed method with Genetic Algorithm. In addition, the other seven methods are some of the famous feature selection methods which are used in different CBIR systems and which have achieved reasonable results. As such, comparing our method with these seven feature selection methods shows robustness and effectiveness of our proposed feature selection phase.

This chapter is organised into five sections. Section 3.2 presents different pre-processing phases for the CBIR system based on Fuzzy Rough Set feature selection. The characteristics of the experimental images are demonstrated in Section 3.3. Section 3.4 presents the experiments and results of the improved CBIR system, compared against the other eight feature selection methods. Finally, the summary is described in the last section.

3.2 Pre-Processing Phase Based on Fuzzy Rough Set

A pre-processing phase based on the Fuzzy Rough Set is presented in this section. However, before that, Rough Set and Fuzzy Rough Set are described in detail. In addition, a brief description of the other eight feature selection methods will be included at the end of this section.

3.2.1 Rough Set

The Rough Set theory was proposed by Zdzislaw Pawlak in the early 1980s (Pawlak, 1982). It is a relatively new soft computing tool used to analyse vague descriptions of an object. The Rough Set theory has become a popular mathematical framework for pattern recognition (W. Wei, Liang, & Qian, 2012), image processing, feature selection (Foithong et al., 2012), neuro computing (J. Zhang, Li, Ruan, Gao, & Zhao, 2012), conflict analysis (Bello & Verdegay, 2012), decision support (Y.-S. Chen & Cheng, 2012), as well as data mining and knowledge discovery (J. Zhang et al., 2012) from large datasets.

Rough Set-based data analysis starts from a data table called the information table. The information table contains data about the objects of interest characterised by a finite set of attributes. Using the information table, some dependency relationships (or patterns) can be discovered (Huang, 2012). An information table, where condition attributes and decision attributes are distinguished, is also called a Decision Table. From a decision table, one can induce some patterns in the form of ‘‘if . . . then . . .’’ decision rules (C.-S. Son, Kim, Kim, Park, & Kim, 2012). More specifically, the decision rules state that if the condition attributes have given values, then the decision attributes have other given values (J. Liang, Li, & Qian, 2012).

The example below shows how the Rough Set theory can decrease image features and extract rules from the data table for classification. Table 3.1 is an information system that includes 4 conditional attributes (a, b, c, d), 1 decision attribute and 8 samples. The decision system is expressed as $I = (U, A \cup \{d\})$, where $d \notin A$; U is a set of samples or objects, with A being the set of conditional attributes, while d is the decision attribute.

The indiscernibility relation (Y.-L. Li, Tang, Chin, Luo, & Han, 2012; Z. Li, Xie, & Li, 2012) for each $P \subseteq A$, with an equivalent relation $IND(P)$, is described as:

$$IND(P) = \{(x, y) \in U^2 | \forall a \in P, a(x) = a(y)\}$$

If $(x, y) \in IND(P)$, x and y are not distinguishable by the P attributes. For example, if $p = \{b, c\}$, $\frac{U}{IND}(P)$ is computed as follows; here, each object which has the same value as the P attribute, is located in one group. As shown in Table 3.1, because row 0 and row 4 have the same value of b and c attributes, they have been located in one group.

$$\begin{aligned} U/IND(P) &= U/IND(b) \otimes U/IND(c) \\ &= \{\{0,2,4\}, \{1,3,6,7\}, \{5\}\} \otimes \{\{2,3,5\}, \{1,6,7\}, \{0,4\}\} \\ &= \{\{2\}, \{0,4\}, \{3\}, \{1,6,7\}, \{5\}\} \end{aligned}$$

Table 3.2: Information System

$x \in U$	a	b	c	d	e
0	1	0	2	2	0
1	0	1	1	1	2
2	2	0	0	1	1
3	1	1	0	2	2
4	1	0	2	0	1
5	2	2	0	1	1

6	2	1	1	1	2
7	0	1	1	0	1

For the set approximation (Cui, 2012; J. Zhang et al., 2012), we have I as an information system. First, let $P \subseteq A$ and $X \subseteq U$. Then, the X set can be approximated by the use of available information in P. This approximation is possible via “lower approximation” and “upper approximation” of X. Considering P, the lower approximation of X includes members that can be placed in X as discrete members, by paying attention to the P attributes. Considering P, the upper approximation of X includes members that can be placed in X as probable members, by paying attention to the P attributes.

$$\underline{P}X = \{x | [x]_P \subseteq X\}; \ \overline{P}X = \{x | [x]_P \cap X \neq \emptyset\}$$

Imagine P and Q have an equivalent relation with U. Therefore, the positive, negative and boundary regions are described as follows:

$$POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}\,X$$

$$NEG_P(Q) = U - \bigcup_{X \in U/Q} \overline{P}\,X$$

$$BND_P(Q) = \bigcup_{X \in U/Q} \overline{P}X - \bigcup_{X \in U/Q} \underline{P}X$$

The positive area includes all the objectives from U that can be sorted in U/Q class, by the use of P attribute information. For example, assume $P = \{b, c\}$ and $Q = \{e\}$. Therefore:

$$POS_P(Q) = \cup \left\{\emptyset, \{2,5\}, \{3\}\right\} = \{2,3,5\}$$

$$NEG_P(Q) = U - \cup \left\{\{0,4\}, \{2,0,4,1,6,7,5\}, \{3,1,6,7\}\right\} = \emptyset$$

$$BND_P(Q) = U - \{2,3,5\} = \{0,1,4,6,7\}$$

In Rough Set theory, dependency is described as follows (Diker & Altay UÄŸur, 2012; X. Li & Liu, 2012):

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|}$$

If $k = 1$, then Q is totally dependent on P. If $0 < k < 1$, then Q is dependent on P partially. If $k = 0$, then Q is not dependent on P.

According to the example, all possible subsets are computed as follows:

$\gamma_{\{a,b,c,d\}}(\{e\}) = 8/8$; $\gamma_{\{b,c\}}(\{e\}) = 3/8$

$\gamma_{\{a,b,c\}}(\{e\}) = 4/8$; $\gamma_{\{b,d\}}(\{e\}) = 8/8$

$\gamma_{\{a,b,d\}}(\{e\}) = 8/8$; $\gamma_{\{c,d\}}(\{e\}) = 8/8$

$\gamma_{\{a,c,d\}}(\{e\}) = 8/8$; $\gamma_{\{a\}}(\{e\}) = 0/8$

$\gamma_{\{b,c,d\}}(\{e\}) = 8/8$; $\gamma_{\{b\}}(\{e\}) = 1/8$

$\gamma_{\{a,b\}}(\{e\}) = 4/8$; $\gamma_{\{c\}}(\{e\}) = 0/8$

$\gamma_{\{a,c\}}(\{e\}) = 4/8$; $\gamma_{\{d\}}(\{e\}) = 2/8$

$\gamma_{\{a,d\}}(\{e\}) = 3/8$.

It is notable that the total dataset is $\gamma_{\{a,b,c,d\}}(\{e\}) = 1$. Therefore, we start from the smallest subsets and select the first and smallest subset, which equals to one as a decreasing subset. In this example, it is as below:

$$R_{min} = \{\{b,d\},\{c,d\}\}$$

If $\{b,d\}$, choose then the decreasing data set is in Table 3.2. Some rules extracted from Table 3.2 are:

Table 3.3: Decreasing dataset

$x \in U$	b	d	e
0	0	2	0
1	1	1	2
2	0	1	1
3	1	2	2
4	0	0	1
5	2	1	1
6	1	1	2
7	1	0	1

$$(b = 0, d = 2) \Rightarrow e = 0$$

$$(b = 1, d = 1) \Rightarrow e = 2$$

$$(b = 2) \Rightarrow e = 1$$

The main reasons why using the Rough Set theory is beneficial are because:

- It does not need any preliminary or additional information about data like probability in statistics, basic probability assignment in Dempster-Shafer theory, grade of membership or the value of possibility in the Fuzzy Set theory (Diker & Altay UÄŸur, 2012; Estaji, Hooshmandasl, & Davvaz, 2012).
- Rough Set theory is a powerful data analysis tool. It can handle and express incomplete data. It can also obtain minimum expressions of information, identify the dependencies between data and get a minimum regularity from the experienced data when the key meaning of the information is kept (Othman, Aris, Othman, & Osman, 2012; Xiang-wei & Yian-fang, 2012).
- Rules that are extracted using the Rough Set theory are helpful for improving the retrieval performance (Shi, Sun, & Xu, 2012; XiongWei, Qiuyan, & Jinlong, 2012).
- The Rough Set theory can be easily combined with other data analysis methods such as Fuzzy Theory, Neural Networks and other methods (D. Chen, Kwong, He, & Wang, 2012; Derrac et al., 2012; Huang, 2012; Yang, Li, Wang, & Wang, 2012).

3.2.2 Fuzzy Rough Set

As described in Section 3.2.1, the Rough Set theory is one of the efficient feature selection methods. The traditional Rough Set theory is restricted to crisp environments. However, in recent times, this has been extended to fuzzy environments, resulting in the development of the Fuzzy Rough Set (Feifei Xu, Duoqian Miao, & Wei, 2009). In addition, the Rough Set has been designed for processing discontinuous data. As such, it is necessary to quantify the data in the continuous area. This quantification causes the loss of some data and the increase of unreal data to the distance that can influence the final results. By using Fuzzy Rough Set, we can retrieve information from the continuous data without using the discontinuous methods (Ganivada et al., 2013; Hu et al., 2012). The Fuzzy Rough Set algorithm used in (Jensen & Shen, 2002) was selected and shown in Figure 3.1.

```
$C$, the set of all conditional features;
$D$, the set of decision features.
$R \leftarrow \{\}$; $\gamma'_{best} = 0$;  $\gamma'_{prev} = 0$
Do
    $T \leftarrow R$
    $\gamma'_{prev} = \gamma'_{best}$
    $\forall x \in (C - R)$
        $IF\ \gamma'_{R\cup\{x\}}(D) > \gamma'_{T}(D)$
            $T \leftarrow R \cup \{x\}$
            $\gamma'_{best} = \gamma'_{T}(D)$
    $R \leftarrow T$
$until\ \gamma'_{best} == \gamma'_{prev}$
$return\ R$
```

Figure 3.15: The Fuzzy Rough Feature Selection Algorithm (Jensen & Shen, 2002)

This algorithm employs the dependency function γ' to choose which features are added to the current reduced candidate. The dependency function is defined as follows:

$$\gamma'_P(Q) = \frac{\sum_{x\epsilon U} \mu_{POS_P(Q)}(x)}{|U|}$$

The function is determined by the fuzzy cardinality of $\mu_{POS_P(Q)}(x)$ divided by the total number of objects in the universe. The membership of an object $\in U$, belonging to the fuzzy positive region, can be defined as:

$$\mu_{POS_P(Q)}(x) = sup_{X\in U/Q}\, \mu_{P\_X(x)}$$

Object x does not belong to the positive region, only if the equivalence class it belongs to is not a constituent of the positive region.

Fuzzy lower and upper approximations are defined as (Hu et al., 2012):

$$\mu_{P\_X(x)} = sup_{F\epsilon U/P}\, min(\mu_F(x), inf_{y\epsilon U}\, max\{1 - \mu_F(y), \mu_X(y)\}$$

$$\mu_{P^-X(x)} = \mathrm{sup}_{F\epsilon U/P}\, \mathrm{min}(\mu_F(x), \mathrm{sup}_{\mathrm{y\epsilon U}}\, \mathrm{min}\{\mu_F(y), \mu_X(y)\}$$

During the implementation, not all $y\ \epsilon\ U$ need to be considered. Only those where $\mu_F(y)$ is non-zero, i.e. where object y is a fuzzy member of (fuzzy) equivalence class$F. < P_X, P^-X >$ is called a Fuzzy Rough Set (Yang et al., 2012).

The algorithm is terminated when the addition of any remaining feature does not increase the dependency.

3.2.3 Stages of the Proposed Content Based Image Retrieval System

Referring to Section 1.4, the objectives of this research is to develop an approach to reduce image features and preserve significant ones from huge amounts of features and applying it to the image retrieval system. In addition, this approach can reduce the semantic gap and improve the image retrieval performance. It can also work in vague and uncertain areas.

The framework of the CBIR system used in this thesis is shown in Figure 3.2. The framework consists of two phases, namely Train and Test. In this research, codes have been developed for the two phases. These two phases were simulated using the Matlab software with the available image processing toolbox, computer vision system toolbox, mapping toolbox and image acquisition toolbox.

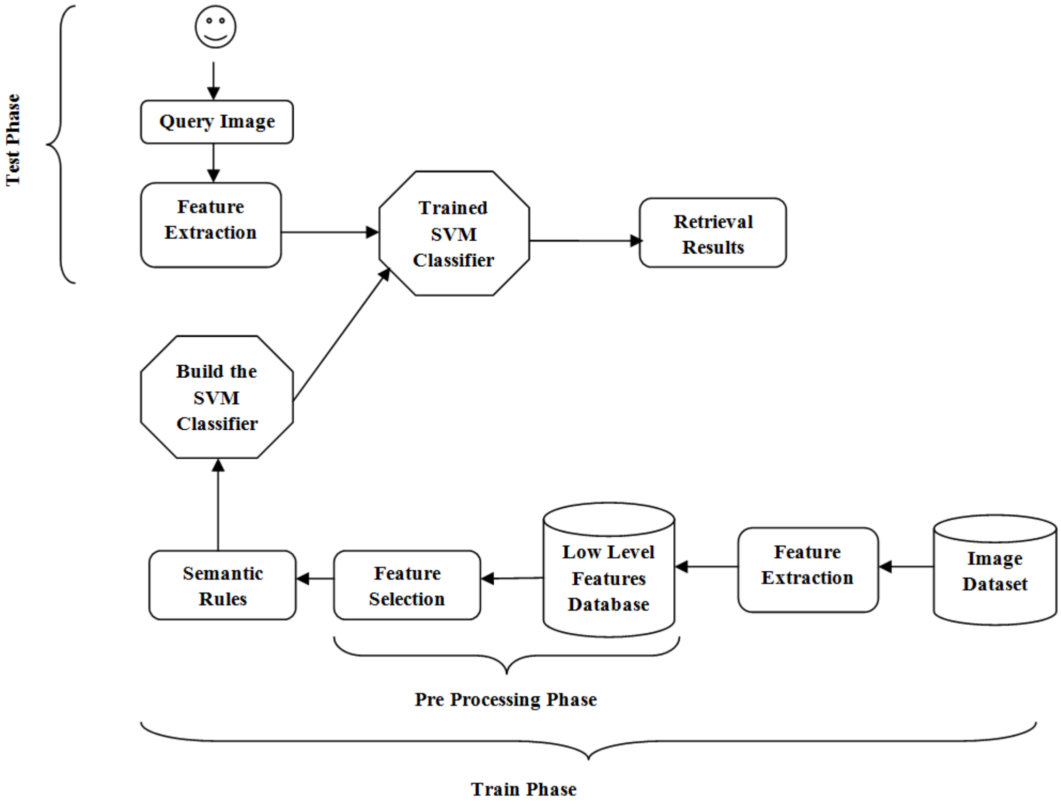


Figure 3.16: The framework of the Retrieval System in this thesis

In the Train phase, colour, shape and texture features are extracted from an image database. The colour features include Hue, Saturation and Intensity Means, as well as Hue, Saturation and Intensity Deviations; the shape features include Area, Circularity, Perimeter, Compactness, Number of Connected, Number of Holes, Convexity, Diameter, Anisomery, Bulkiness, Structure Factor, Euler Number, Inner Circle Radius etc., and the texture features include Gray Mean, Gray Deviation, Gray

Entropy, Gray Anisotropy, Fuzzy Perimeter, Fuzzy Entropy, Co-occurrence, Correlation, Energy, Inertia, Local Homogeneity etc. The most important features are then selected by using the Fuzzy Rough Set feature selection. Semantic rules are then generated with these features. After that, the Support Vector Machine (SVM) classifier is built using these semantic rules.

Still referring to Figure 3.2, the second part is the Test phase where the user feeds a query image into the retrieval system so that the queried image features could be extracted. The SVM classifier evaluates the queried image features with other image features in the database using semantic rules and shows the most relevant image to the user. The significant part of this research is that the relevant images are estimated by semantic rules extracted utilising the Fuzzy Rough Set theory. Therefore, the retrieval performance is improved and as a result, the semantic gap is reduced.

It is worth highlighting the following aspects of the proposed method.

- Image query removes the difficulty of describing the feature of an image into words when similar images are searched.
- The Fuzzy Rough Set method as a pre-processing stage makes the proposed approach more robust than conventional approaches.
- This proposed system can effectively and efficiently handle large image databases and can be smoothly embedded into different image retrieval systems.
- This retrieval system refines decision rules of image retrieval by the Fuzzy Rough Set.

- Rough Sets provide reasonable structures for the overlap boundary, given the domain knowledge.
- The proposed method can work efficiently in vague and uncertain areas.

The SVM is used as a classifier in the proposed method. In an image annotation and retrieval, the SVM is a widely used machine learning method (Chinpanthana, 2011). In addition, the SVM can generate a hyper plane to separate two data sets of features and provide good generalisation (S. Li, Wu, Wan, & Zhu, 2011). The SVM is also used as a learning tool to handle image retrieval problems (Z. Chen et al., 2012). Furthermore, the SVM is known to perform well with noisier data, as compared to other machine learning methods (Z. Chen et al., 2012). As for the SVM classifier, it is important to select the right kernel function.

The authors use the non-linear SVM, along with the Gaussian radial basis function kernel, in the system. This is because it achieves better results compared with linear and polynomial kernels (Z. Chen et al., 2012; Maryam Shahabi Lotfabadi & Mahmoudie, 2010).

3.2.4 Other Feature Selection Methods

Information Gain: The Information Gain (IG) is the expected reduction in entropy resulting from partitioning the dataset objects according to a particular feature (Baranidharan & Ghosh, 2012). The entropy of a labelled collection S of c objects is defined as:

$$Entropy\ (S) = \sum_{i=1}^{c} -p_i\, log_2 p_i$$

where p_i is the proportion of S belonging to class i. Based on this, the Information Gain metric is:

$$IG(S,A) = Entropy\ (S) - \sum_{v \epsilon values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

where $values(A)$ is the set of values for feature A, S is the set of training examples, while S_v is the set of training objects, where A has the value v. This metric is used in ID3 (decision tree) for selecting the best feature to partition the data.

Gain Ratio: One limitation of the IG measure is that it favours features with many values. The Gain Ratio (GR) seeks to avoid this bias by incorporating another term, “split information”, which is sensitive to how broadly and uniformly the attribute splits the considered data (Chinpanthana, 2011).

$$Split(S,A) = -\sum_{i=1}^{c} \frac{|S_i|}{|S|} log_2 \frac{|S_i|}{|S|}$$

where each S_i is a subset of objects generated by partitioning S with the c-valued attribute A. The Gain Ratio is then defined as follows:

$$GR(S,A) = \frac{IG(S,A)}{Split(S,A)}$$

Genetic Algorithm (GA): The presented method uses a genetic algorithm for feature selection. Genetic Algorithms (GAs), a form of inductive learning strategy, are adaptive search techniques initially introduced by Holland (Holland, 1975). Genetic algorithms derive their name from the fact that their operations are similar to the mechanics of genetic models of natural systems.

Genetic algorithms typically maintain a constant-sized population of individuals which represent samples of the space to be searched. Each individual is evaluated on the basis of its overall fitness with respect to the given domain application (Sérgio Francisco da Silva et al., 2011). New individuals (samples of the search space) are produced by selecting high performing individuals to produce "offsprings", which retain many of the features of their "parents". This eventually leads to a population that has improved fitness, with respect to the given goal. New individuals (offsprings) for the next generation are formed by using two main genetic operators - crossover and mutation (Tsai, Eberle, & Chu, 2013). Crossover operates by randomly selecting a point in the two selected parents' gene structures and exchanging the remaining segments of the parents to create new offspring. Therefore, crossover combines the features of two individuals to create two similar offspring. Mutation operates by randomly changing one or more components of a selected individual. It acts as a population perturbation operator and is a means for inserting new information into the population. This operator prevents any stagnation that might occur during the search process. The main issues in applying GAs to any problem are the selection of an appropriate representation and an adequate evaluation function (Huang, 2012).

Isomap: Classical scaling has proven to be successful in many applications, but it suffers from the fact that it mainly aims to retain pair wise Euclidean distances and does not take into account the distribution of the neighbouring data points (Balasubramanian & Schwartz, 2002). If the high-dimensional data lies on or near a curved manifold, classical scaling might consider two data points as near points, whereas their distance over the manifold is much larger than the typical interpoint distance. Isomap is a technique that resolves this problem by attempting to preserve pair wise geodesic (or curvilinear) distances between data points. Geodesic distance is the distance between two points measured over the manifold.

Kernel PCA (KPCA): This is the reformulation of traditional linear PCA in a high-dimensional space that is constructed using a kernel function (Hoffmann., 2007). KPCA computes the principal eigenvectors of the kernel matrix, rather than those of the covariance matrix. The reformulation of PCA in kernel space is straightforward since a kernel matrix is similar to the in-product of the data points in the high-dimensional space that is constructed using the kernel function. The application of PCA in the kernel space provides KPCA with the property of constructing non-linear mappings.

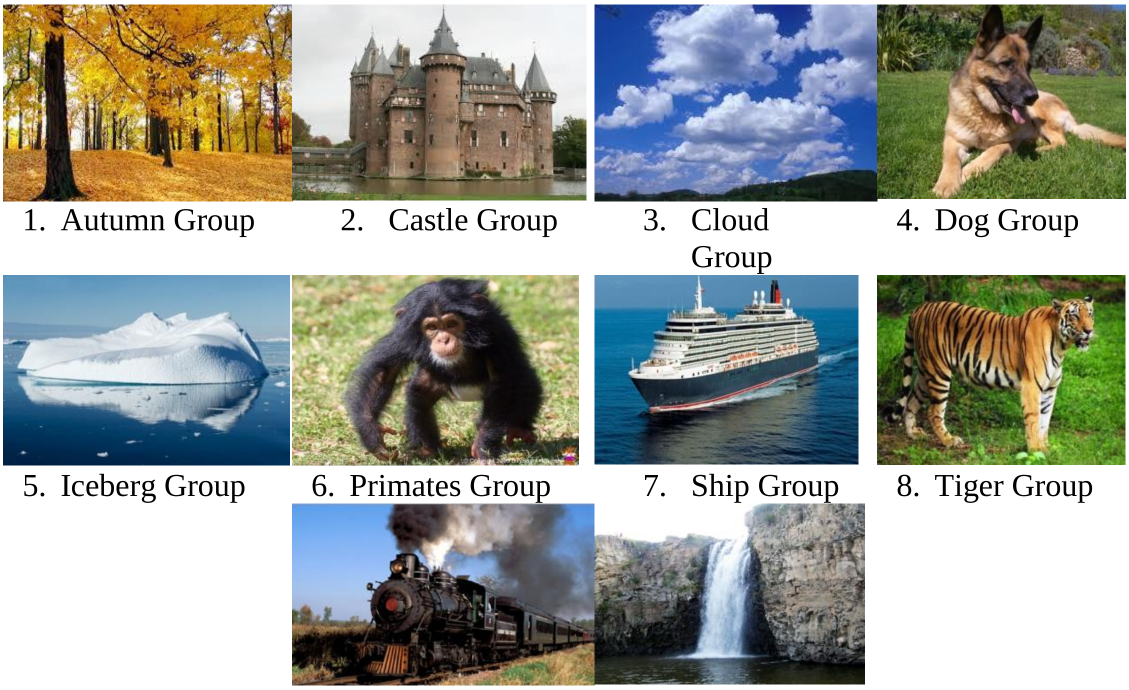
OneR: The OneR classifier learns a one-level decision tree (i.e. it generates a set of rules that test one particular attribute) (Chinpanthana, 2011). One branch is assigned for every value of a feature, and each branch is assigned to the most frequent class. The error rate is then defined as the proportion of instances that do not belong to the majority class of their corresponding branches. Features with higher classification rates are considered to be more significant than those resulting in lower values.

Principal Components Analysis (PCA): This is a linear method for dimensionality reduction that embeds the data into a linear subspace of lower dimensionality (Hotelling, 1933). PCA constructs a low-dimensional representation of the data that describes as much of the variance in the data as possible. This is done by finding a linear basis of reduced dimensionality for the data, in which the amount of variance in the data is maximal.

Relief-F: This is the Relief-F measure based on the original Relief measure (Abdolhossein Sarrafzadeh et al., 2012). Relief evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different classes (Kononenko, Šimec, & Robnik-Šikonja, 1997). Relief-F extends this idea to dealing with multi-class problems, as well as handling noisy and incomplete data.

3.3 Image Dataset Used in the Experiments

To investigate the function of the CBIR system based on the above-mentioned feature selection methods (Sections 3.2.2 and 3.2.4), the Corel image dataset is used. The Corel image dataset (http://corel.digitalriver.com/) has 68,040 images from various groups. For the experimental results, 2000 images were divided from the Corel image dataset into 10 semantic groups, e.g., autumn, castle, cloud, dog, iceberg, primates, ship, tiger, train and waterfall. Figure 3.3 shows an image example for each semantic group. We reorganised the Corel image dataset because 1) many images with similar concepts were not in the same group, 2) some images with different semantic contents were in the same group in the original dataset and 3) from the literature review chapter (Chapter 2), it is understandable that many CBIR systems used the Corel dataset to evaluate their algorithms, so comparing this improved CBIR system with other CBIR systems becomes easier. In the reorganised dataset, each group includes more than 150 images and the images in the group are category-homogeneous. These semantic groups were used in the evaluation of the results of our CBIR system based on the pre-processing phase. The images in the Corel database are 384*256 or 256*384 pixels in the JPEG format.



1. Autumn Group
2. Castle Group
3. Cloud Group
4. Dog Group
5. Iceberg Group
6. Primates Group
7. Ship Group
8. Tiger Group
9. Train Group
10. Waterfall Group

Figure 3.17: Example of Semantic Groups in Corel Image Dataset

3.4 Experimental Results with the Pre-Processing Phase

This section aims to study the performance of the proposed feature selection phase and compare the results with eight other feature selection methods. Different experiments are evaluated in the next four sub sections.

3.4.1 Experiment I: Precision-Recall Graph

The first experiment aims to study the eight feature selection methods with Fuzzy Rough feature selection to investigate which method can provide better results in terms of the Precision-Recall graph.

The Precision-Recall graph is the basic measure used in evaluating CBIR systems. The Precision equals the number of related retrieval images to the total number of retrieval images (Cerra & Datcu, 2012).

A= Number of relevant images retrieved, *B*= Number of irrelevant images retrieved.

$$\text{Precision}= \frac{A}{A+B}$$

The Recall equals to the number of the related retrieval images to the total number of the related images available in the image database (Z. Liang et al., 2013).

A= Number of relevant images retrieved, *C*= Number of relevant images not retrieved.

$$\text{Recall}= \frac{A}{A+C}$$

Figure 3.4 shows the Precision-Recall graph for the 10 semantic groups that are used for measuring the efficiency of the proposed CBIR system.

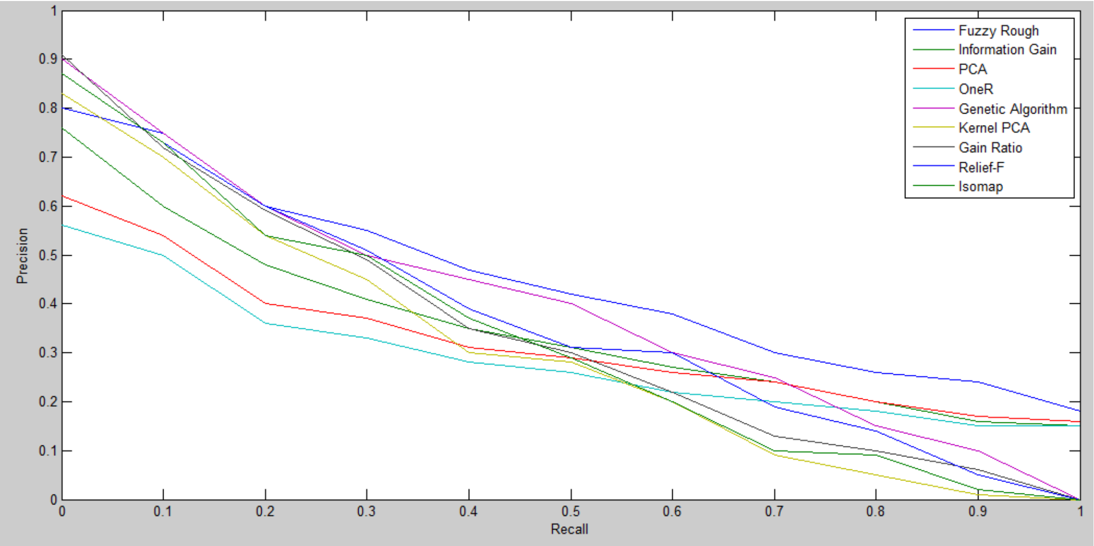


Figure 3.18: Precision-Recall Graph

From the graph, it is observed that the proposed CBIR system achieved better results than the other eight systems. The reason for this is due to the application of the proposed algorithm in the training phase to save and eliminate appropriate or useless image features respectively. With useful features, the system can train the SVM classifier with more accurate rules.

The Fuzzy Rough feature selection method, which performed efficiently in this experiment, needs to be investigated further into. This is done by comparing it with other feature selection methods for retrieval accuracy in the next experiment.

3.4.2 Experiment II: Retrieval Accuracy

This experiment aims to investigate the retrieval accuracy of the proposed CBIR system and compare the results with the other eight feature selection methods.

To investigate the total retrieval accuracy of the above mentioned CBIR systems, 100 images were fed into the system as queried images. The average of the retrieval accuracy is calculated for each class. Figure 3.5 shows the results. As expected, the results are better when using the proposed CBIR system. The average of the retrieval accuracy is 69.94%, 71.5%, 68.3%, 65.1%, 78.25%, 73.22%, and 74.28% for Information Gain, Isomap, PCA, OneR, Relief-F, Kernel PCA and Gain Ratio respectively and 86.34% for Genetic Algorithm. It is increased to 91.06% for Fuzzy Rough feature selection.

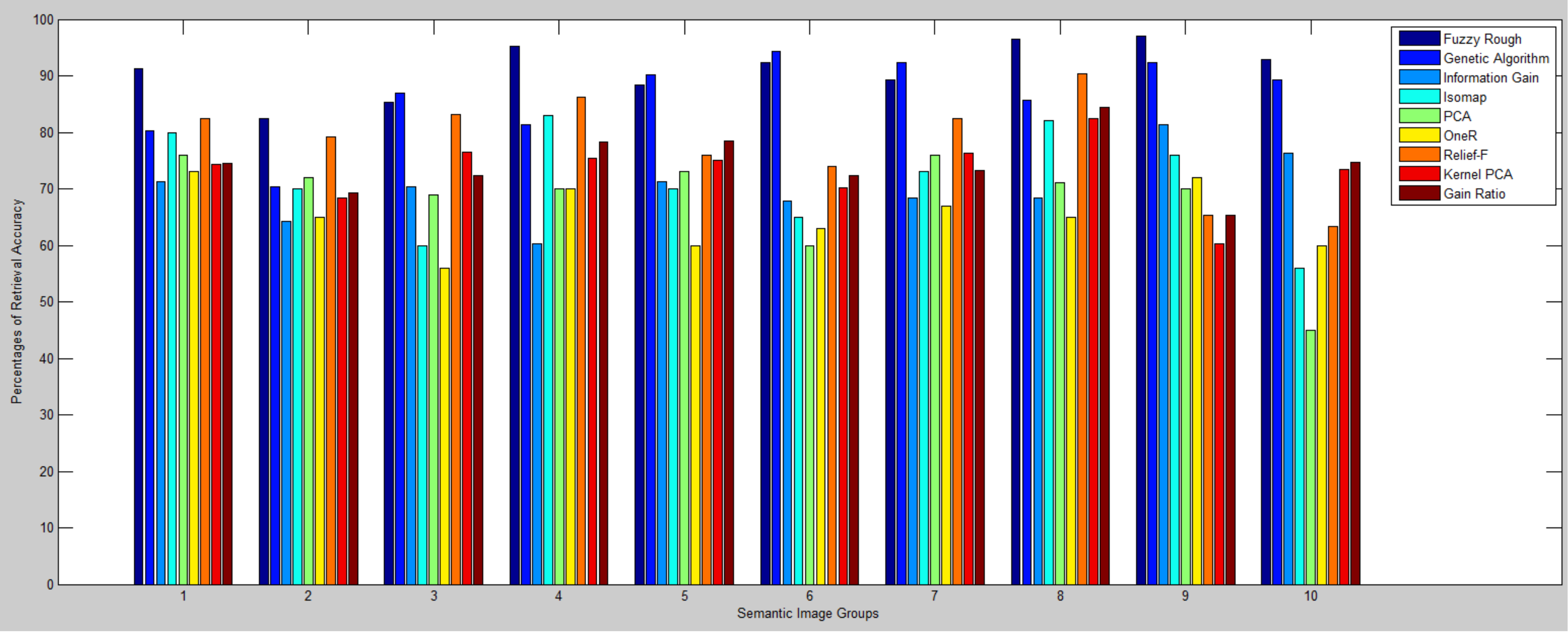


Figure 3.19: Retrieval Accuracy Graph

The reasons behind the superiority of the proposed CBIR system are:

- The Rough Set theory is a useful method for describing and modelling vagueness in ill-defined environments.
- The use of the membership function of a fuzzy set has many advantages in the definition, analysis and operation of fuzzy concepts.
- Combining Fuzzy Set and Rough Set (Fuzzy Rough Set) as a feature selection has increased retrieval accuracy because the rules which are generated from the Fuzzy Rough feature selection are semantic and helps the CBIR system achieve better results, compared to other CBIR systems which use the other eight feature selection methods.

3.4.3 Experiment III: Feature Selection Order

The first and second experiment already presented the increase of CBIR performance when the Fuzzy Rough feature selection is applied. However, the feature selection order did not investigate. Therefore, the third experiment aims to explore if the Fuzzy Rough feature selection can firstly recognise and select important features.

In this experiment, the image features that are important in the image retrieval application are defined and ranked. These features are collected and reviewed from literature (Acharya & Devi, 2012; Bird et al., 1996). However, to be consistent with the experiment used in this thesis, only literature using the Corel dataset with 1000 images in the 10 semantic groups will be examined. The features used in the analysis with their corresponding order of importance are shown in Table 3.3. From the literature, the most influential features are Mean Hue, Coarseness, Standard deviation, Wavelet Moment and Directionality. Contrast and Mean intensity are the next most influential features. Although an order is given to Edge and Roughness, it is difficult to differentiate between them. The features identified in Table 3.3 are only valid for the Corel dataset and may not apply to all image datasets.

Table 4.3: Important Features of Images

Feature number	Feature name	Defined ordering
1	Coarseness	1
2	Directionality	1
3	Mean hue	1
4	Standard deviation	1
5	Wavelet moments	1
6	Contrast	2
7	Mean intensity	2
8	Entropy	3
9	Euler number	3
10	Edge	4
11	Roughness	5
12	Bulkiness	6
13	Deviation intensity	6
14	Roundness	6
15	Structure factor	6
16	Convexity	7
17	Rectangularity	7
18	Sigma	7

The results of the comparison study using feature selection methods can be seen in Table 3.4. All methods rate features 3 (Mean hue) and 1 (Coarseness) highly. This is in agreement with the defined ranking as shown in Table 3.3. Only the Fuzzy Rough method correctly rates features 5 (Wavelet moments), 2 (Directionality) and 4 (Standard deviation) highly. After these features, Fuzzy Rough ranks Contrast and Mean intensity. In fact, Fuzzy Rough is the only method that can detect the importance of these two features. The results show that the Fuzzy Rough feature selection method is useful in producing results in line with the defined ranking. The reason is the Fuzzy Rough feature selection uses the dependency function to select the important features. This function uses a positive region that can deal with the vague areas and recognise more important features.

Feature number	**Defined ordering**	**FR**	**GA**	**Re**	**IG**	**KPCA**	**GR**	**1R**	**PCA**	**IM**
3	1	0.214	0.316	0.142	0.147	0.075	0.163	83.4	0.061	16.3
1	1	0.185	0.271	0.153	0.204	0.064	0.183	82.7	0.072	14.1
5	1	0.109	0.071	0.074	0.421	0.0	0.401	68.2	0.046	0.0
2	1	0.143	0.086	0.084	0.0	0.0	0.0	78.3	0.0	0.0
4	1	0.102	0.271	0.061	0.0	0.053	0.0	70.1	0.0	0.0
6	2	0.096	0.076	0.023	0.0	0.04	0.0	74.5	0.052	0.04
7	2	0.062	0.093	0.013	0.0	0.0	0.0	70.3	0.0	0.0
8	3	0.0	0.005	0.061	0.0	0.04	0.0	78.3	0.0	0.0
9	3	0.0	0.215	0.043	0.0	0.0	0.0	71.3	0.057	0.03
10	4	0.043	0.005	0.020	0.0	0.0	0.0	78.3	0.003	0.0
11	5	0.0	0.083	0.004	0.0	0.0	0.0	78.3	0.04	0.0
13	6	0.025	0.0	0.086	0.0	0.061	0.0	74.5	0.0	0.0
14	6	0.0	0.203	0.008	0.0	0.0	0.0	78.3	0.0	0.0
12	6	0.023	0.005	0.009	0.0	0.04	0.0	72.6	0.052	0.0
15	6	0.0	0.0	0.007	0.0	0.0	0.0	78.3	0.06	0.0
16	7	0.0	0.0	0.003	0.0	0.0	0.0	72.6	0.04	0.0
17	7	0.0	0.003	0.090	0.005	0.05	0.0	71.1	0.0	0.0
18	7	0.007	0.003	0.005	0.0	0.0	0.001	78.3	0.0	0.50

Table 3.5: Feature Ranker Results for the Corel Dataset

(FR=Fuzzy Rough, GA=Genetic Algorithm, Re=Relief-F, IG=Information Gain, KPCA=Kernel Principal Component Analysis, GR=Gain Ratio, 1R=OneR, PCA=Principal Component Analysis, IM=Isomap)

3.4.4 Experiment IV: The Image Comparison of the CBIR Systems

In the last test of this chapter, the retrieval results for all ten semantic groups are shown. The query images for each semantic group are based on Figure 3.3. The first, second and up to the ninth row in Figures 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, 3.14 and 3.15 are related to Fuzzy Rough, Genetic Algorithm, Information Gain, Isomap, PCA, OneR, Relief-F, Kernel PCA and Gain Ratio respectively. Referring to Figures 3.6 to 3.15, the retrieval system with the Fuzzy Rough Set method relates more output images to the user. The first left image in Figure 3.6 to 3.15 matches closely to the queried image.

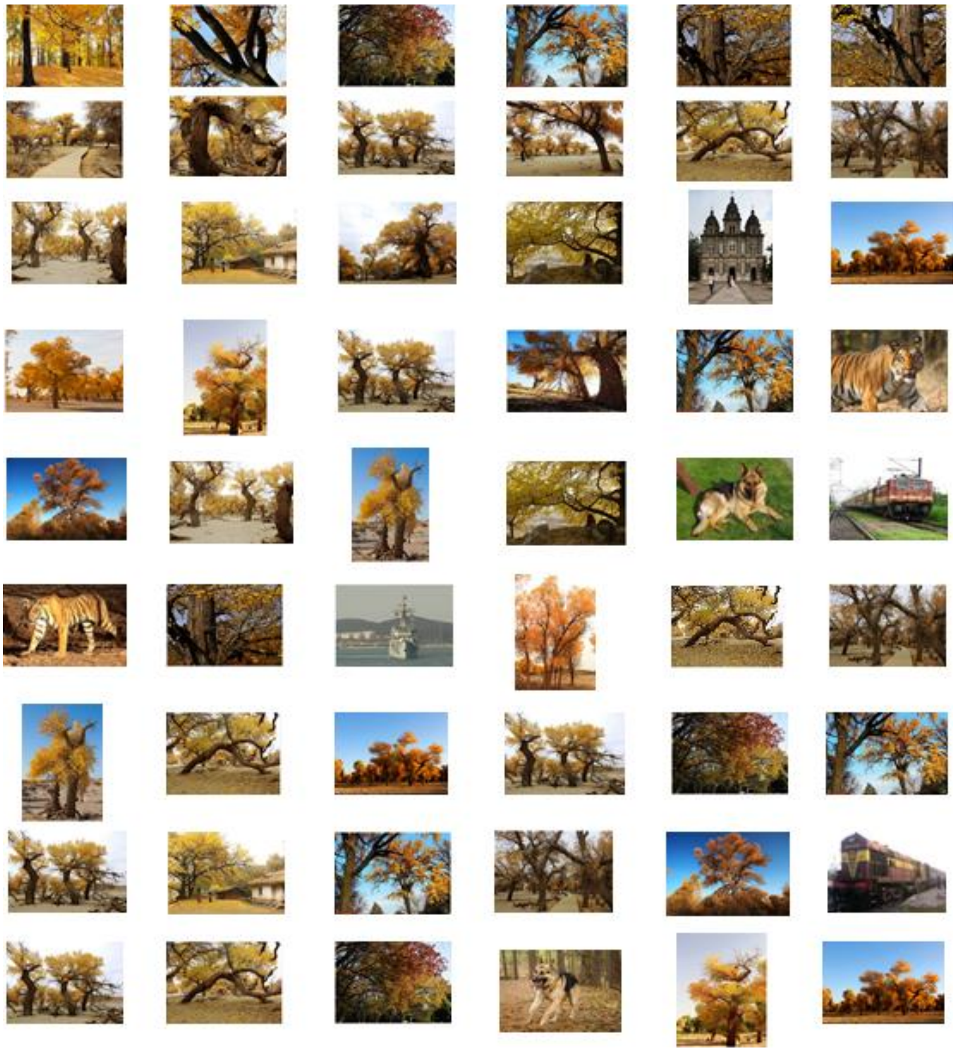


Figure 3.20: Retrieved Images for Autumn Query Image, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

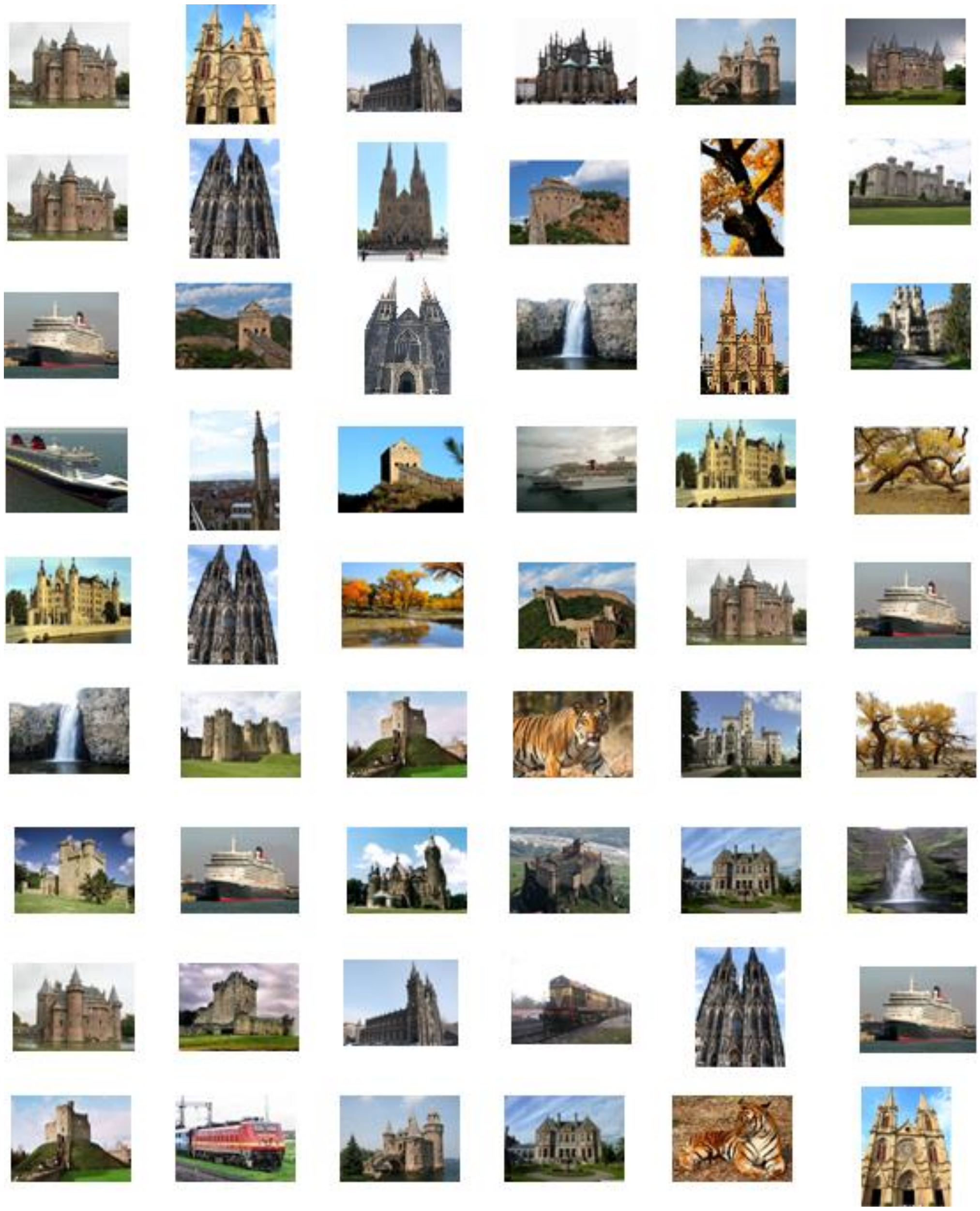


Figure 3.21: Retrieved Images for Castle Query Image, according to First Row-Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

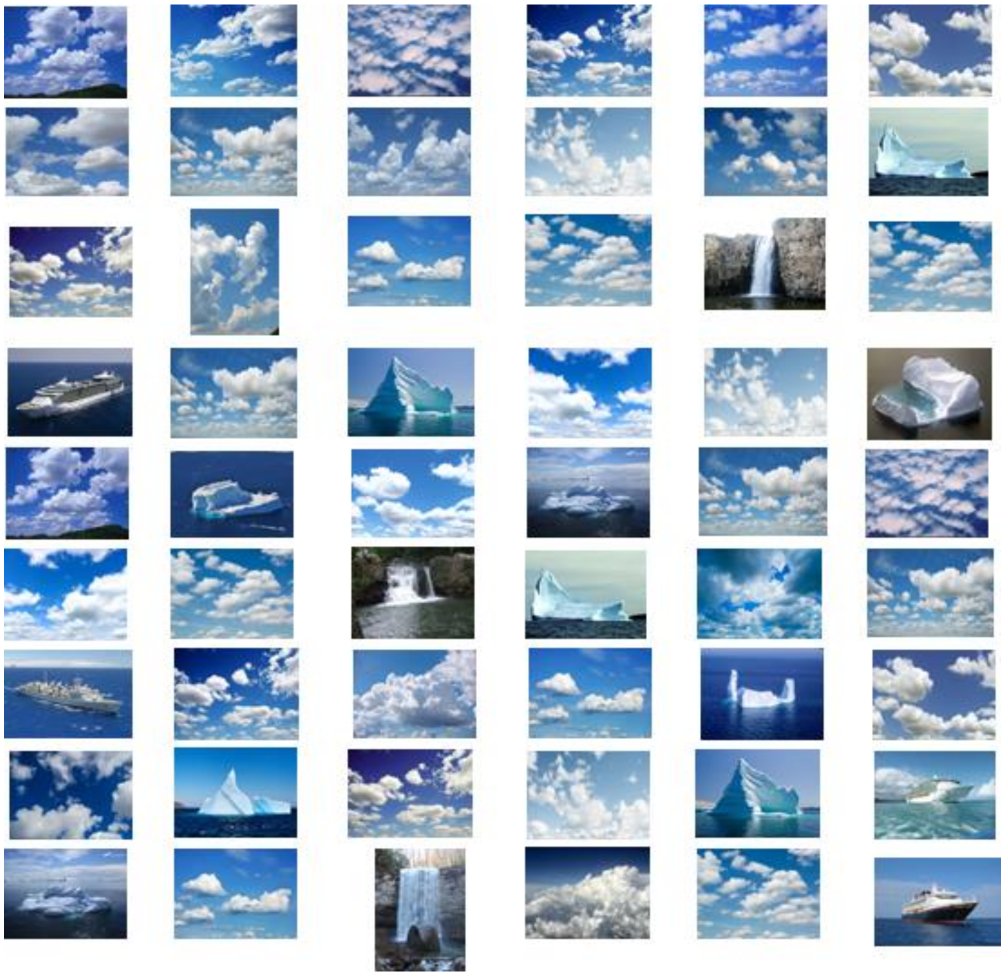


Figure 3.22: Retrieved Images for Cloud Query Image, according to First Row-Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

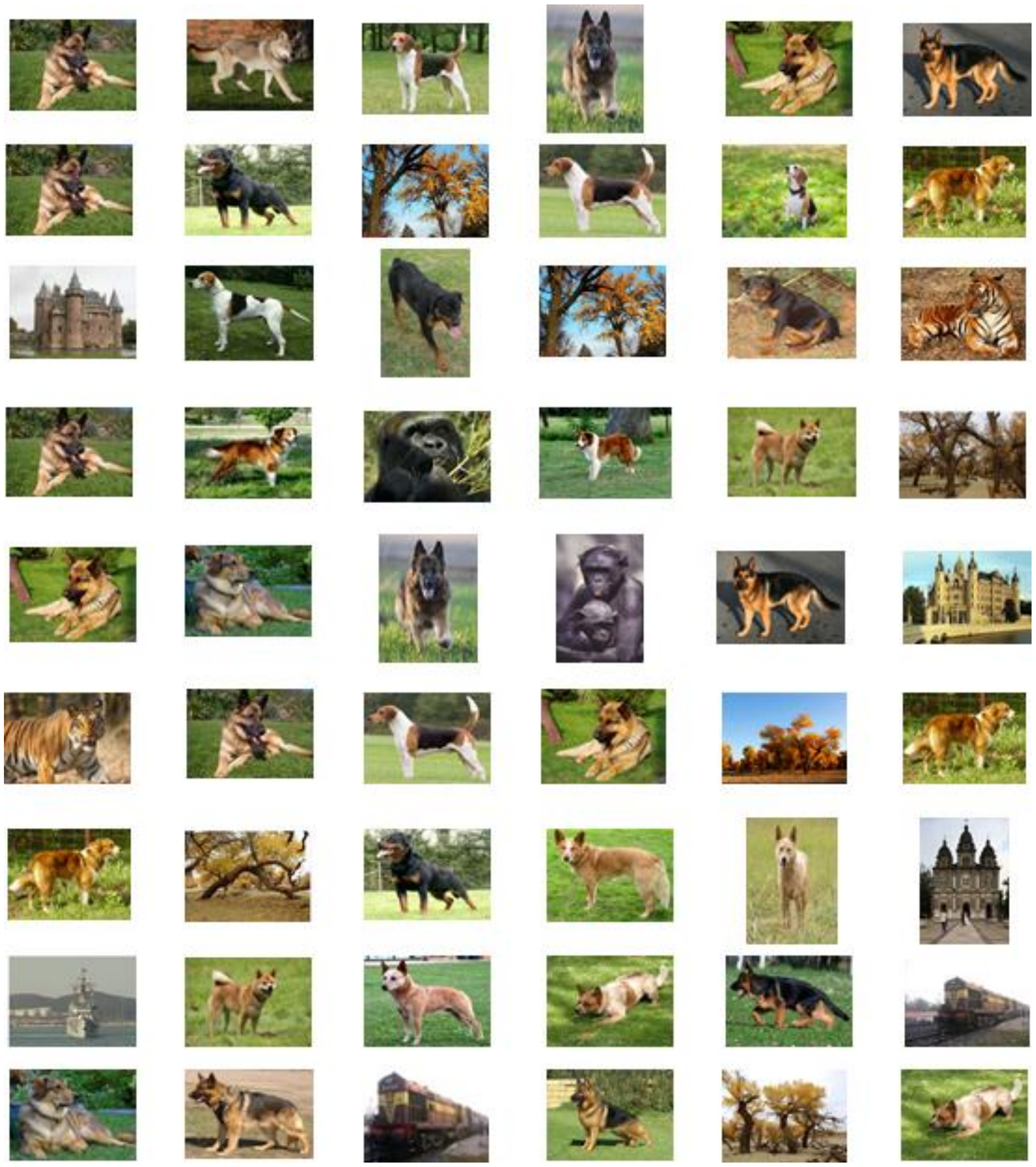


Figure 3.23: Retrieved Images for Dog Query Image, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

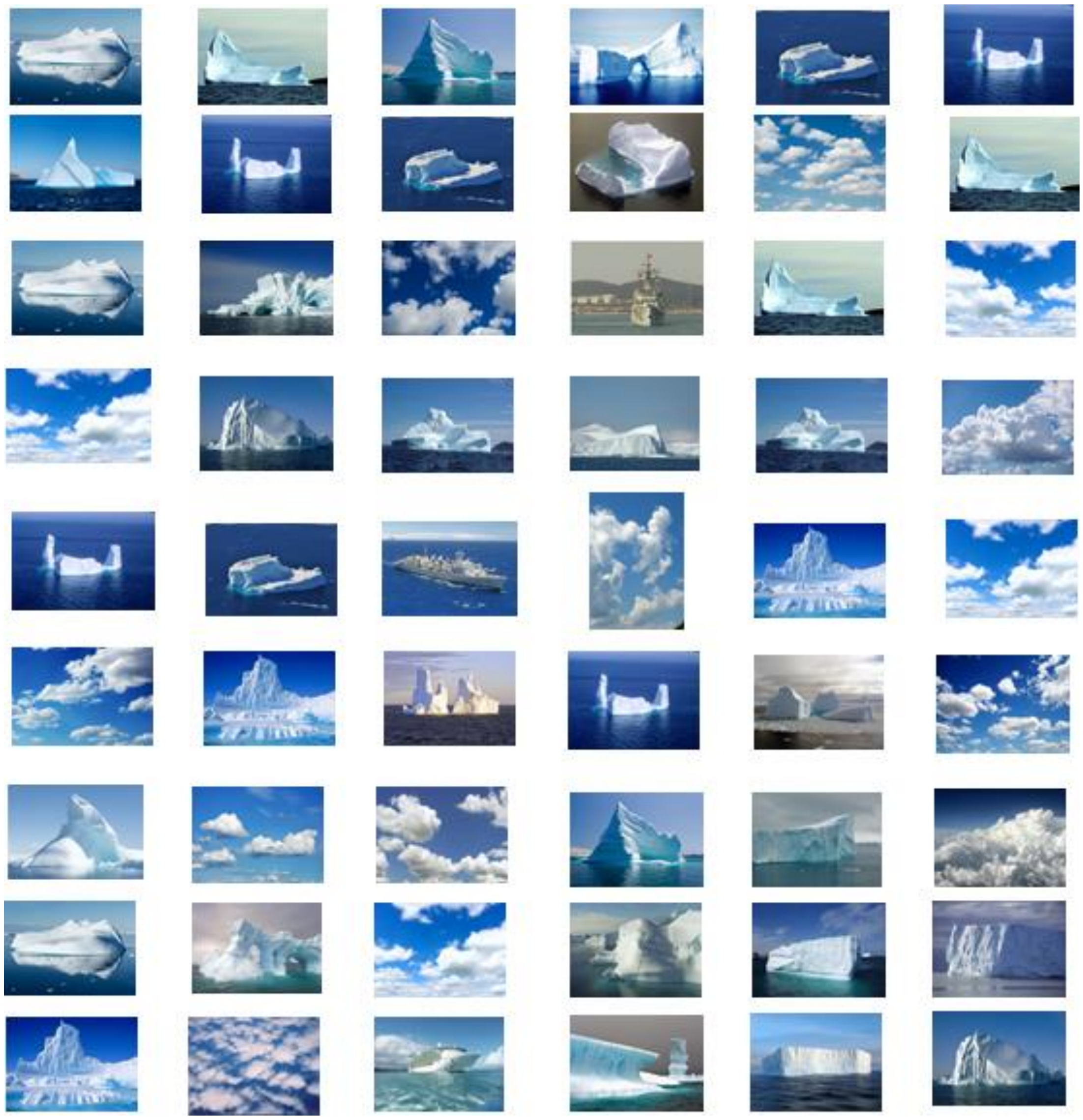


Figure 3.24: Retrieved Images for Iceberg Query Image, according to First Row-Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

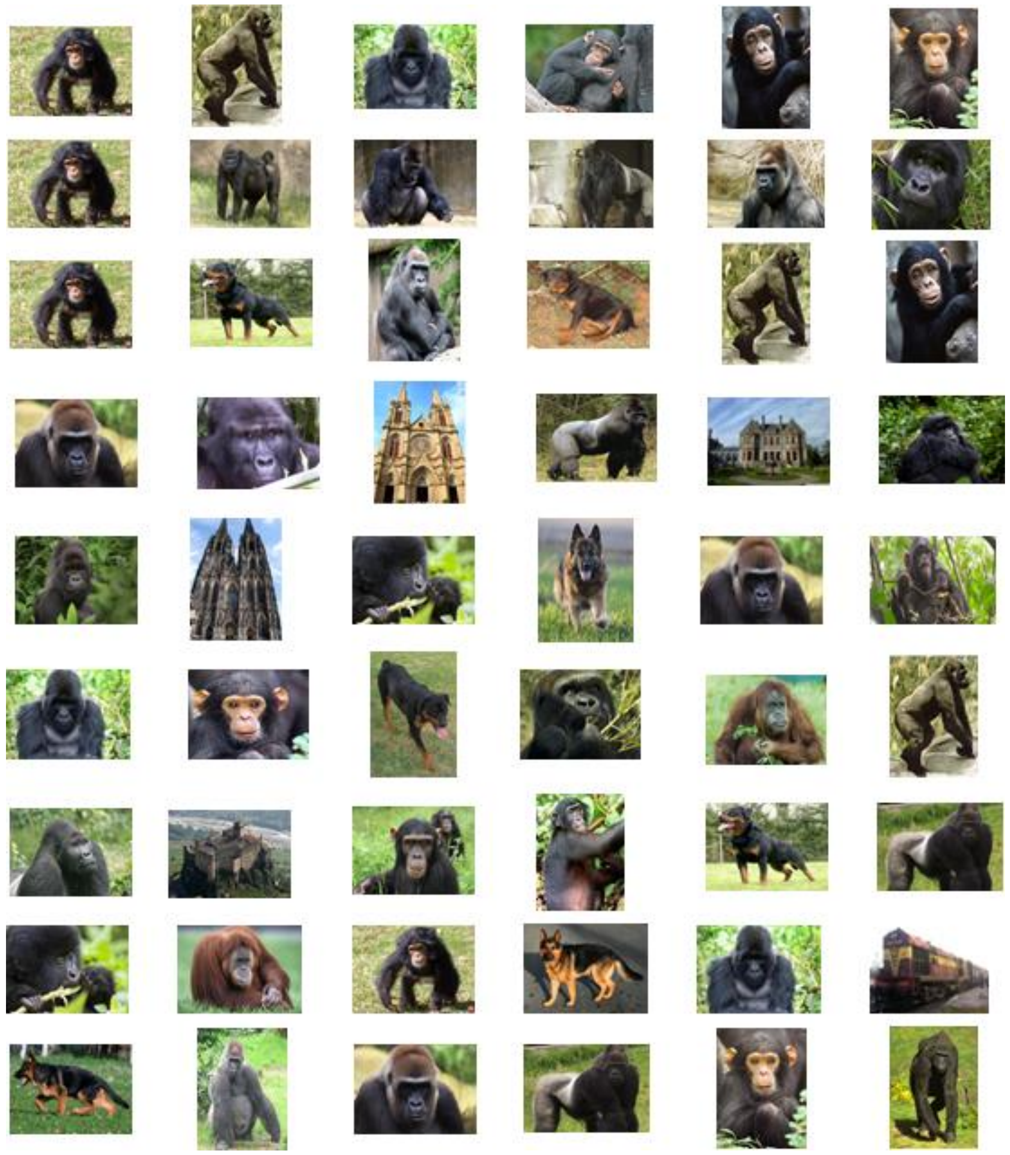


Figure 3.25: Retrieved Images for Primates Query Image, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

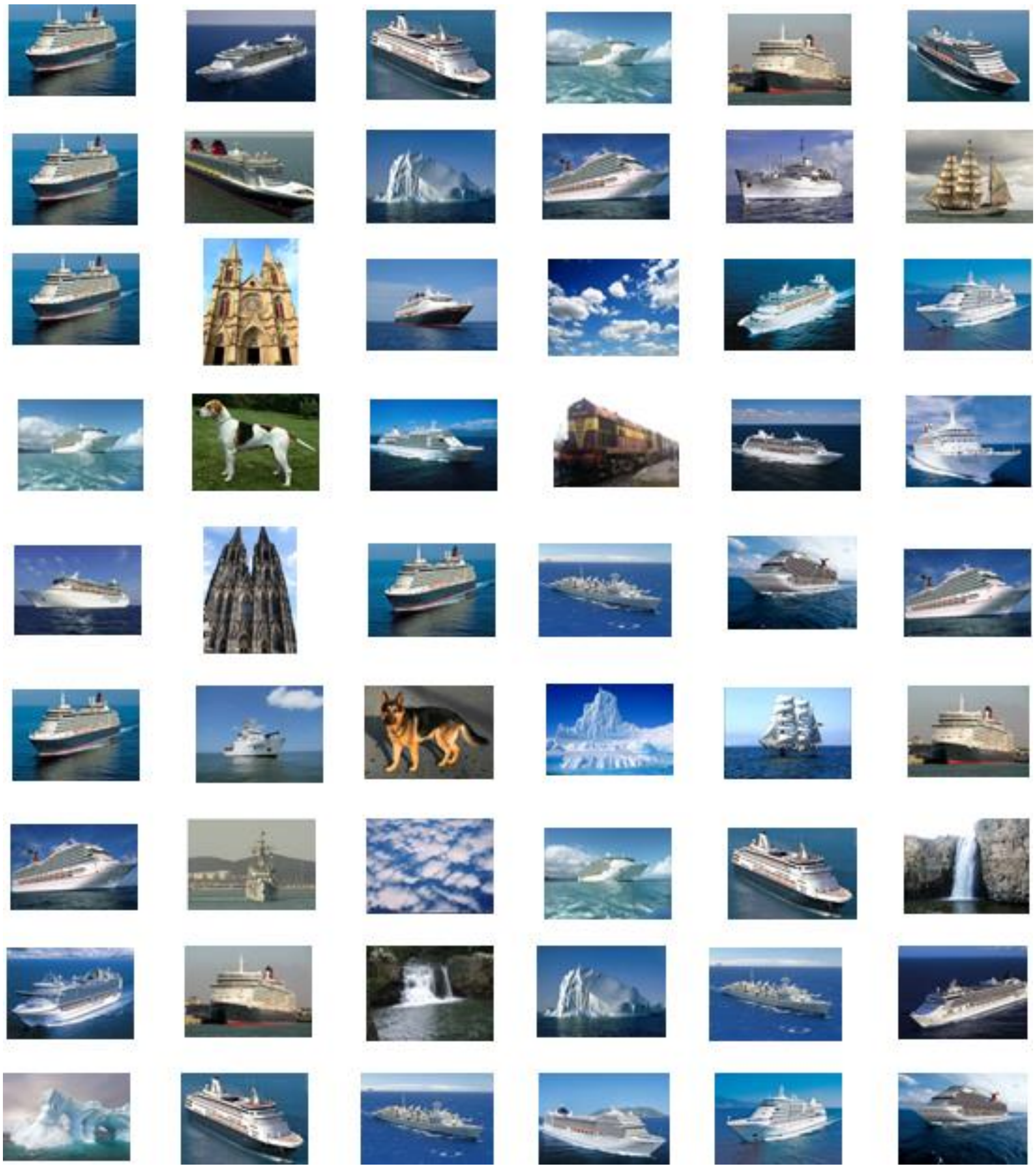


Figure 3.26: Retrieved Images for Ship Query Image, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

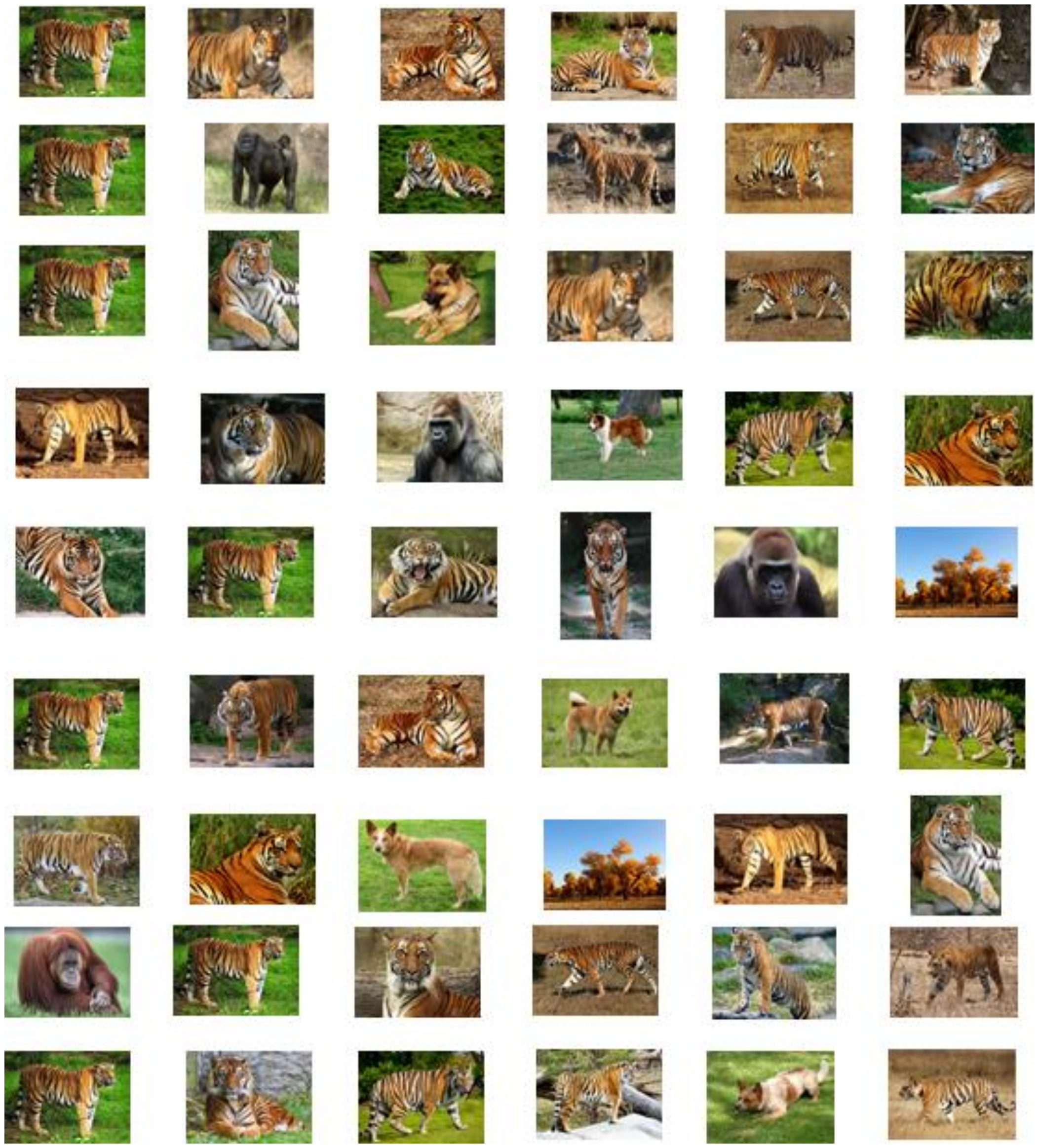


Figure 3.27: Retrieved Images for Tiger Query Image, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

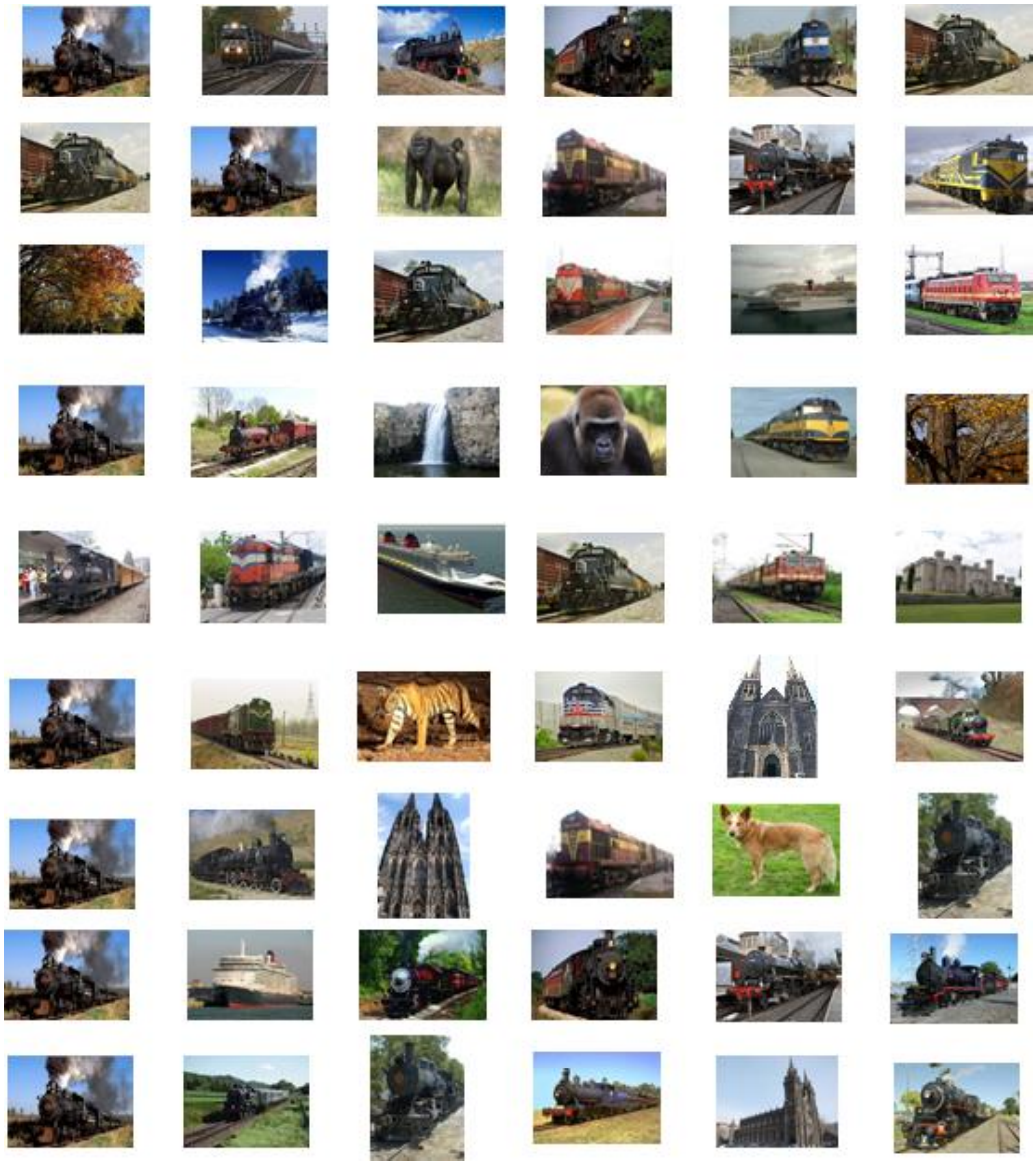


Figure 3.28: Retrieved Images for Train Query Image, according to First Row-Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

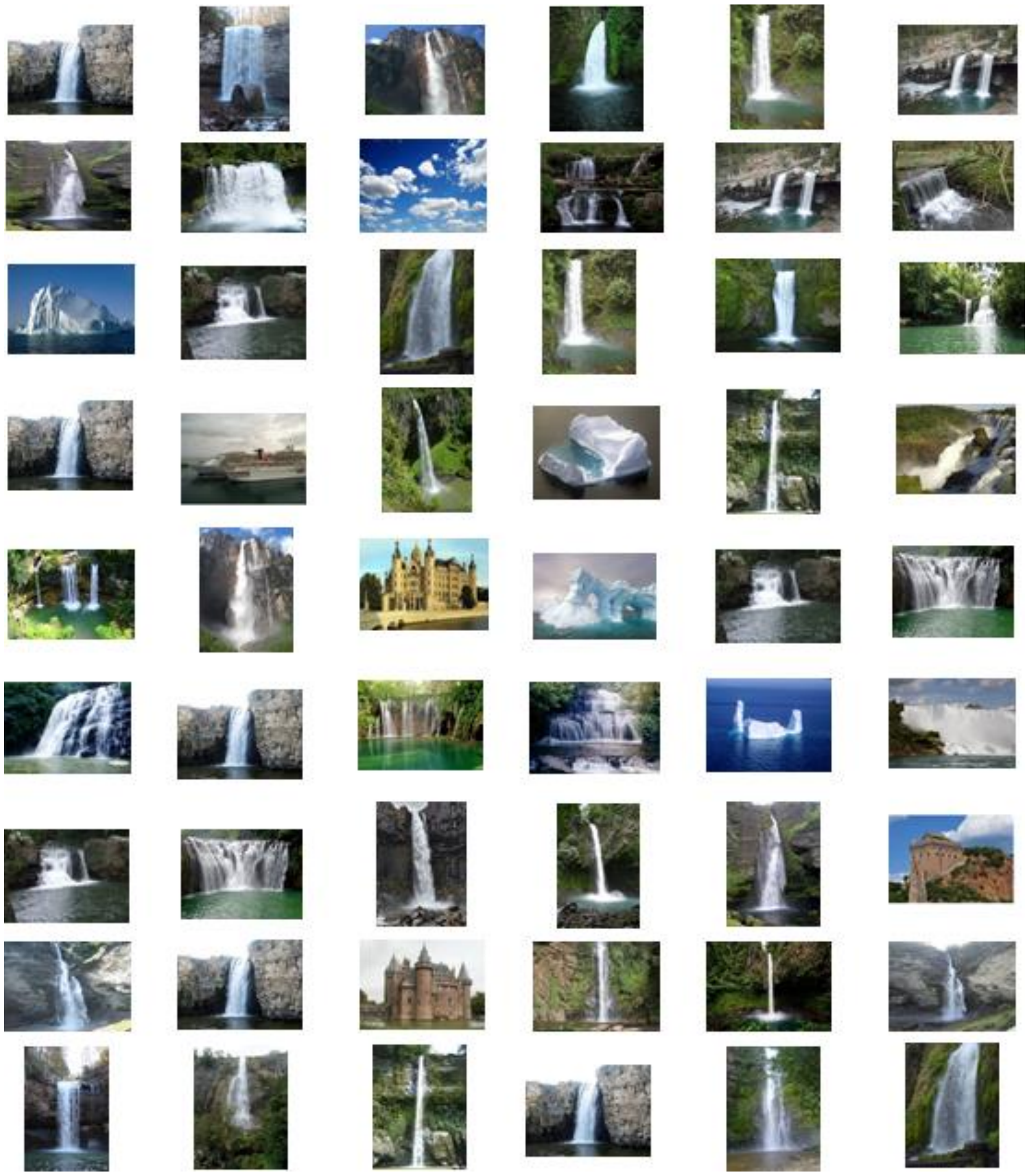


Figure 3.29: Retrieved Images for Train Query Image, according to First Row-Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- Isomap, Fifth Row- PCA, Sixth Row-OneR, Seventh Row-Relief-F, Eighth Row-Kernel PCA and Ninth Row-Gain Ratio

The reason why the proposed method has better results than those in other retrieval systems is that it does not require preliminary or additional parameters to describe the data; it works with missing values, it uses little time to generate rules; it can handle large amounts of quantitative and qualitative data; it yields easily understood decision rules supported by a set of real examples; it models highly non-linear or discontinuous functional relationships and is a powerful method for characterizing complex and multi-dimensional patterns; it discovers important facts hidden in data and expresses them in the natural language of decision rules; and the rules extracted from the features that are selected with the Fuzzy Rough Set are semantic and can be better used to train the SVM classifier. Consequently, the SVM classifier can show more relevant images to the user.

3.5 Summary

This chapter presents how the Fuzzy Rough feature selection can be applied successfully in the CBIR system. The Fuzzy Rough feature selection can select important features accurately and thus provide better retrieval accuracy. Eight different feature selection methods, namely Gain Ratio, Genetic Algorithm, Information Gain, Isomap, Kernel PCA, OneR, Principal Component Analysis (PCA) and Relief-F, were compared with the Fuzzy Rough Set. Based on the results, it is shown that the Fuzzy Rough Set performs the best by using different tests and measurements.

By utilising the Fuzzy Rough Set, the proposed system has the advantage and deals efficiently in image feature environments that are vague and uncertain. In addition, the rules extracted from selecting features with Fuzzy Rough are semantic and can train the classifier presciently. An important advantage of this work is training the SVM with the semantic rules that can separate the relevant images from the irrelevant ones more accurately.

Furthermore, the Rough Set theory is a useful tool for describing and modelling vagueness in ill-defined environments. Besides, the use, of the membership function of the Fuzzy Rough Set has many advantages in the definition, analysis and operation of fuzzy concepts. The hybrid scheme that combines the advantages of rough and fuzzy set has better performance in image retrieval application. Overall, when the two theories are combined, the disadvantage of one is covered by the other; hence, better results can be obtained.

Chapter 4: Fuzzy Rough Set Feature Selection in Content Based Image Retrieval Systems with Noisy Images

4.1 Introduction

Image noise consists of a random (not present in the object image) variation of brightness or colour information in them and is usually an aspect of electronic noise (L. Chen et al., 2014; Jaime, Kerre, Nachtegael, & Bustince, 2013). Noise represents unwanted information which can deteriorate image quality (Frosini & Landi, 2013a; Walek, Jan, Ourednicek, Skotakova, & Jira, 2012). It can be produced by the sensor and circuitry of a scanner or digital camera. Image noise can also originate in film grain and in the unavoidable shot noise of an ideal photon detector (Frosini & Landi, 2013b). Image noise is an undesirable by-product of image capture that adds spurious and extraneous information (L. Chen et al., 2014; Frosini & Landi, 2013a).

In Chapter 3, the Fuzzy Rough Set has produced good results when used as a feature selection in the Corel image database as compared to other feature selection methods. Some of the publications based on the success of the Fuzzy Rough Set in the Corel image dataset (Maryam Shahabi Lotfabadi, Shiratuddin, & Wong, 2012a; Maryam Shahabi Lotfabadi et al., 2012b; Maryam Shahabi Lotfabadi, Mohd Fairuz Shiratuddin, & Kok Wai Wong, 2013; Maryam Shahabi Lotfabadi, M. F. Shiratuddin, & Kok Wai Wong, 2013) and other databases (Maryam Shahabi Lotfabadi & Eftekhari Moghadam, 2010) are listed here.

In this chapter, the performance of the Fuzzy Rough Set feature selection for noisy images is evaluated. The main objective of this chapter is to investigate the performance of the semantic rules extracted from the Fuzzy Rough Set with noisy images. In addition, the ability of these rules to recognise noisy images and allocate them to their related semantic groups has been studied in this chapter. The other objective of this chapter is to evaluate if the rules, which are extracted from the Fuzzy Rough Set and used for training the Support Vector Machine (SVM), can deal with the noisy query images as well as the originally queried images.

In the experiments, 10 semantic groups form the Corel image dataset (the same as those in Chapter 3) are used. The same 10 semantic groups were selected so as to make the comparison of the images (with and without noise) easier. In Chapter 2, it can be observed that the Corel image benchmark dataset has been commonly used in the literature. This image dataset contains minimum noise. Noise was added to the queried image to compare the performance of the Fuzzy Rough Set feature selection with other feature selection methods in a noisy environment.

The main purpose of doing this is to evaluate the effect of noise on the feature selection techniques. The three types used include the Gaussian noise, Poisson noise and Salt and Pepper noise. Using the Matlab software from Mathworks, the three types of noise were added to the queried images.

Furthermore, the results of the other methods such as the Genetic Algorithm (Tsai et al., 2013), Information Gain (Guldogan & Gabbouj, 2008), OneR (Hopfgartner et al., 2010) and Principle Component Analysis (PCA) (X.J. Shen & Wang, 2006) are compared with the proposed method. In this chapter, one of the feature selection methods used to compare with the proposed method is Genetic Algorithm; Genetic Algorithm is one of the soft computing methods that has demonstrated effective feature selection capability (S. Li et al., 2011; Tsai et al., 2013). In addition, Information Gain, OneR and PCA are well-known feature selection methods. Many researchers have used these methods for their feature selection tasks. Therefore, it is essential to compare the proposed method with them.

This chapter is structured into five sections. Section 4.2 presents three different kinds of image noises which are used in this chapter. Section 4.3 describes the stages of proposed framework for noisy images applied to Content Based Image Retrieval (CBIR) systems. The experimental results compared with other methods are discussed in Section 4.4. Finally, the conclusion is presented in Section 4.5.

4.2 Image Noise

For a better understanding of the differences between these three kinds of noise, a brief discussion of each type of noise is provided below. Figure 4.1 shows the three types of noise added to the original image (a), with (b), (c) and (d) showing the different effects each type of noise can produce.

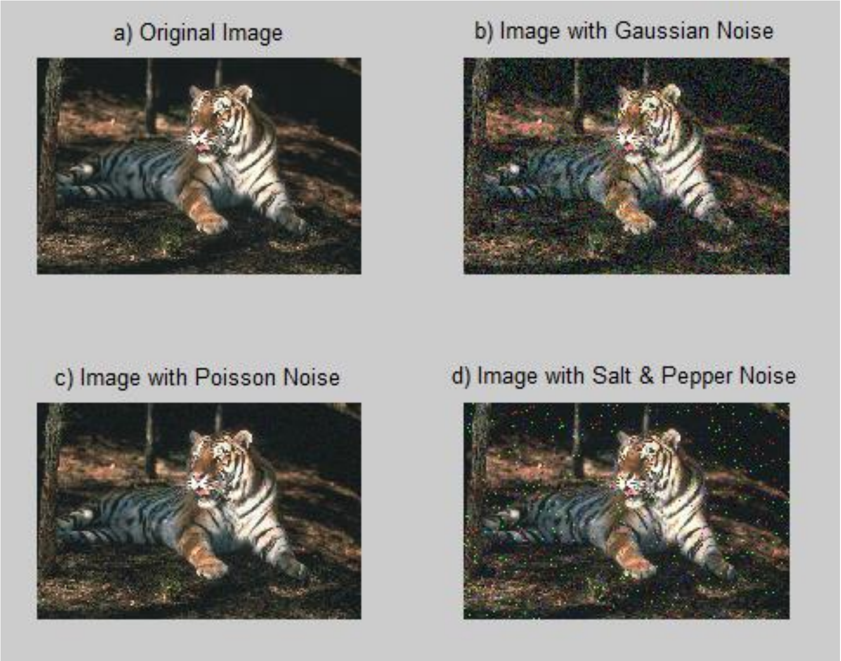

a) Original Image
b) Image with Gaussian Noise
c) Image with Poisson Noise
d) Image with Salt & Pepper Noise

Figure 4.30: Original image (a) and images with Gaussian, Poisson and Salt & Pepper noises added (b, c and d)

4.2.1 Gaussian Noise

Gaussian noise represents statistical noise having the Probability Density Function (PDF) equalling to that of the normal distribution, which is also known as the Gaussian distribution (Liu, Zeng, Shen, & Luan, 2013). In other words, the values that the noise can take on are Gaussian distributed. In this thesis, Gaussian white noise of mean *m* and variance *v* was added to the queried images. The mean noise and variance are 0 and 0.01 respectively for an image with Gaussian Noise (see Figure 4.1 in (b)).

4.2.2 Poisson Noise

Poisson noise, also known as Photon noise, is a basic form of uncertainty associated with the measurement of light, inherent to the quantized nature of light and the independence of photon detections (Setayesh, Zhang, & Johnston, 2013). Its expected magnitude is signal-dependent and constitutes the dominant source of image noise, except in low light conditions. Matlab syntax generates the Poisson noise from the data instead of adding artificial noise to the data. Figure 4.1 in (c) shows an image with Poisson noise.

4.2.3 Salt and Pepper Noise

Impulsive noise is sometimes called the Salt and Pepper noise or Spike noise (Yueyang Li, Sun, & Luo, 2014). An image containing Salt and Pepper noise will have dark pixels in its bright regions and bright pixels in its dark regions (Yueyang Li et al., 2014; C.-H. Son, Choo, & Park, 2013). Section (d) in Figure 4.1 represents an image with Salt and Pepper noise. The noise density of this image is 0.02.

4.3 Stages of the Content Based Image Retrieval System for Noisy Images

The Fuzzy Rough Set used in this section is the same as Section 3.2.2 in Chapter 3. Also, the diagram is similar to Figure 3.2 in Chapter 3, as in the testing phase, the user feeds the noisy queried image, instead of the normal query image, to the system. The system extracts the noisy queried image features and gives these features to the SVM classifier, which is then built into the training phase. This classifier will extract the relevant images based on the noisy queried image provided.

Some features selection methods (Yu & Bhanu, 2010) can only operate effectively with datasets containing discrete values and as such, have difficulty handling noisy data. As most datasets contain real-valued features, it is necessary to perform a discretization step beforehand. In the Fuzzy Rough feature selection method, this is typically implemented by standard fuzzification methods, enabling linguistic labels to be associated with the attributes values (Derrac et al., 2012). It also aids uncertainty modelling by allowing the possibility of the membership of a value to be assigned to more than one fuzzy label. However, membership degrees of feature values in the fuzzy sets are not exploited in the process of dimensionality reduction. By using Fuzzy Rough Sets, it is possible to use the membership information to better guide feature selection.

4.4 Experimental Results with the New Pre-Processing Phase for Noisy Images

This section presents three experiments. These three experiments are conducted to investigate the ability of the proposed feature selection method when dealing with the noisy images. Furthermore, the results that compare the four feature selection methods with the proposed retrieval system (by using the generated noisy images modified) are based on the three types of noise studied for each three experiments. To investigate the function of the image retrieval system based on the above-mentioned methods, the Corel image database, as shown in Section 3.3 (Chapter 3), is employed.

The reason Genetic algorithm, Information Gain, PCA and OneR are chosen for comparison with the Fuzzy Rough Set in a noisy environment is because these four methods had high (Genetic Algorithm), medium (Information Gain and PCA) and low (OneR) results when compared to the Fuzzy Rough Set in Chapter 3. Comparing the Fuzzy Rough Set in a noisy environment against a pure (without noise) environment (with high, medium and low methods) is really important for understanding the efficiency of the proposed method.

4.4.1 Experiment I: Precision-Recall Graph

Figure 4.2, 4.3 and 4.4 show the Precision-Recall graphs for ten semantic groups with Gaussian noise, Salt and Pepper noise and Poisson noise respectively. This is used for measuring the efficiency of the proposed retrieval system.

From the graphs, we observe that the proposed retrieval system achieved better results than the other four systems in all three kinds of noise. The reason for this is that a better feature extraction algorithm has been applied in the training phase to save appropriate image features, as well as eliminate the useless image features. With these useful features, the system can train the SVM classifier with more accuracy and semantic rules.

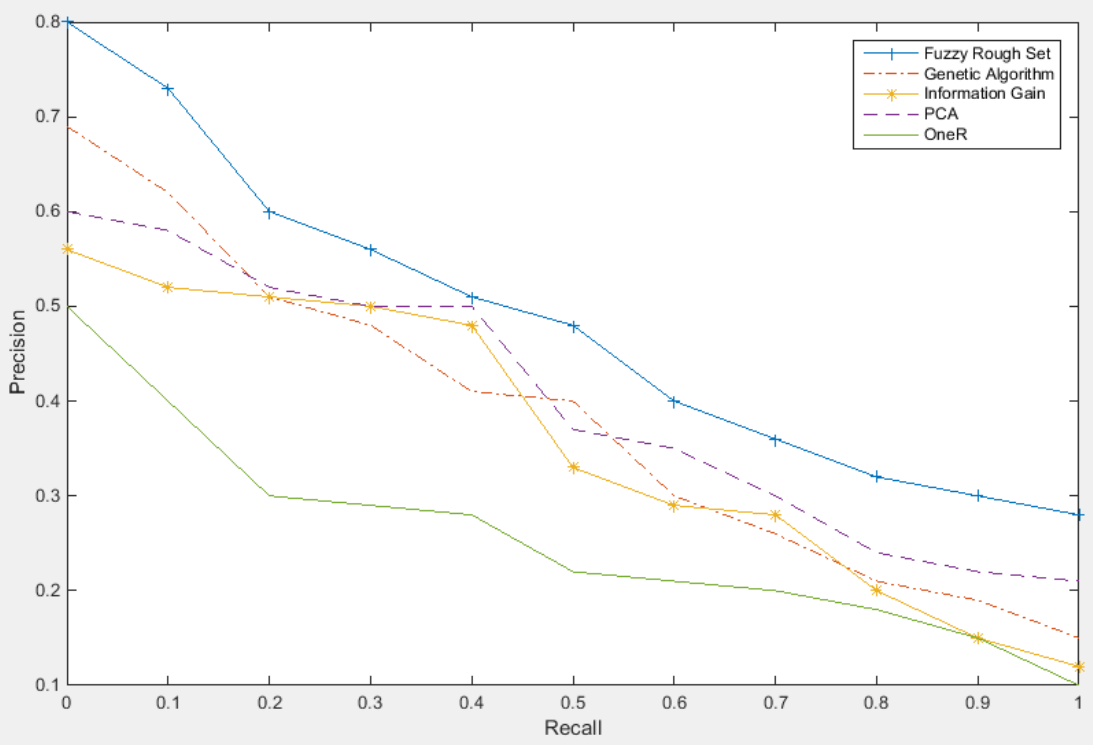


Figure 4.31: Precision-Recall Graph with Gaussian Noise

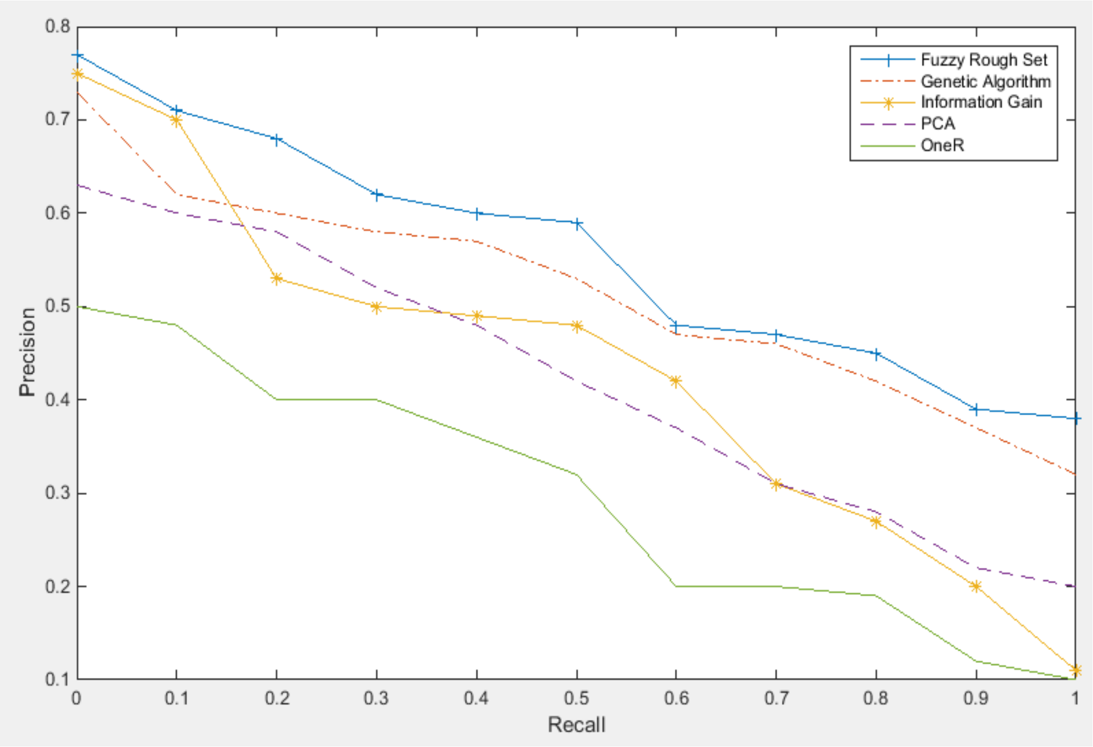


Figure 4.32: Precision-Recall graph with Salt and Pepper Noise

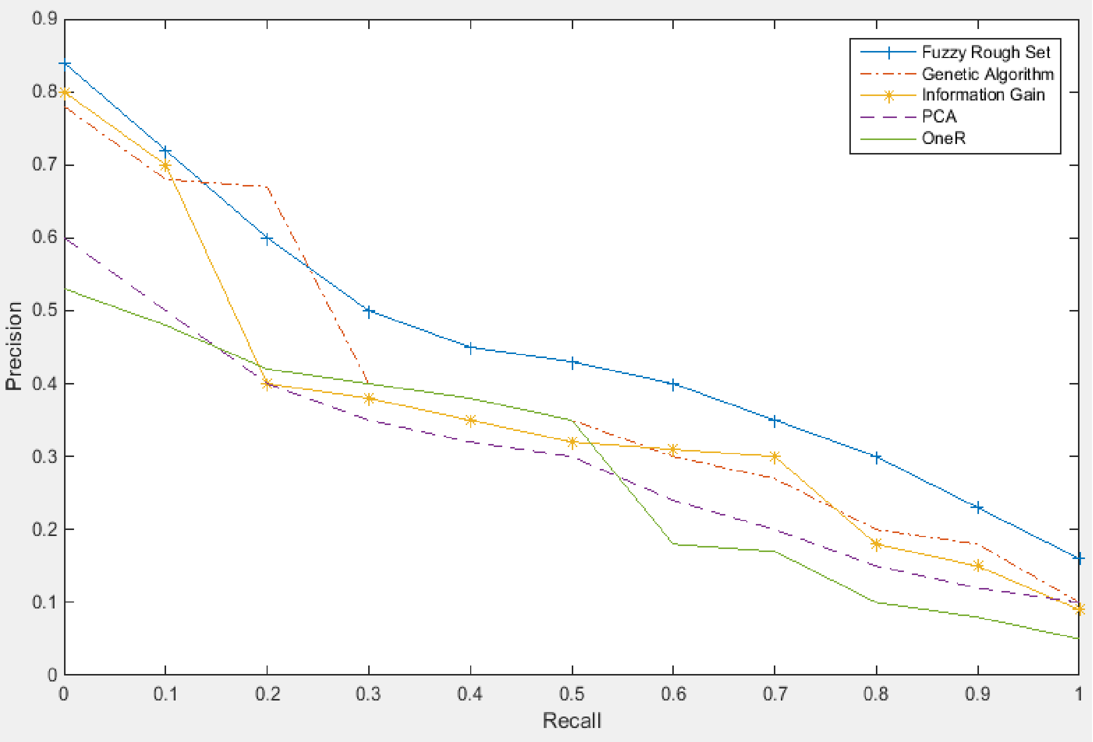


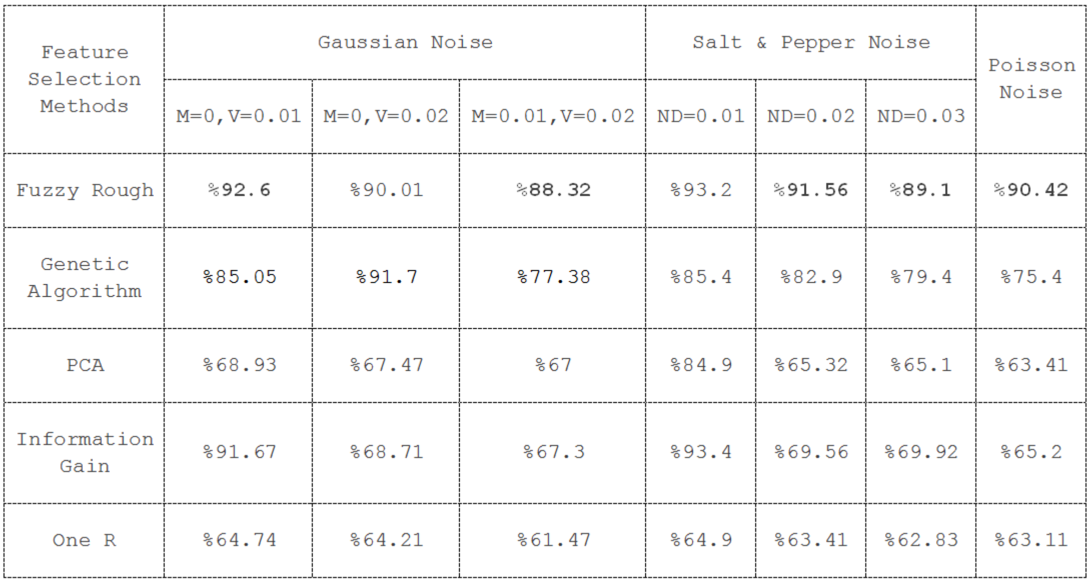
Figure 4.33: Precision-Recall graph with Poisson Noise

One of the features of the SVM is that it can perform well with noisier data. As such, although the SVM is used for all feature selections, the investigation performed in this section showed that the Fuzzy Rough Set has better results. The reason behind the better results of our proposed feature selection method is that the rules extracted from the Fuzzy Rough Set feature selection are semantic rules, which are used for training the SVM classifier. These semantic rules can train the SVM to get a more confident decision score for relevance measurement. In addition, these semantic rules can handle noisy positive images, while other methods cannot.

4.4.2 Experiment II: The Investigation of the Retrieval Accuracy

To investigate the total accuracy of the above-mentioned retrieval systems, 60 noisy images were fed into the system as queried images. That means 60 query images with Gaussian noise (Mean=0 and Variance= 0.01), 60 query images with Gaussian noise (Mean=0 and Variance= 0.02) etc. In addition, three different Noise Densities (ND) are used in the Salt and Pepper noise in the experimental results. The average of the retrieval accuracy is calculated for each system with the three types of noise. Table 4.1 shows the results. As expected, the results in most cases are better using our proposed feature selection method.

Table 4.6.Accuracy of Retrieval with Three Kinds of Noise



Feature Selection Methods	Gaussian Noise			Salt & Pepper Noise			Poisson Noise
	M=0,V=0.01	M=0,V=0.02	M=0.01,V=0.02	ND=0.01	ND=0.02	ND=0.03	
Fuzzy Rough	%92.6	%90.01	%88.32	%93.2	%91.56	%89.1	%90.42
Genetic Algorithm	%85.05	%91.7	%77.38	%85.4	%82.9	%79.4	%75.4
PCA	%68.93	%67.47	%67	%84.9	%65.32	%65.1	%63.41
Information Gain	%91.67	%68.71	%67.3	%93.4	%69.56	%69.92	%65.2
One R	%64.74	%64.21	%61.47	%64.9	%63.41	%62.83	%63.11

The results extracted from Table 4.1 are as follows:

- The image retrieval system which used the Fuzzy Rough Set for feature selection in their methodology had better results compared to the other retrieval systems which used other feature selection methods in their methodology.
- Overall, most of the feature selection methods had better results with the Salt and Pepper noise.
- When the mean and variance of the Gaussian noise were increased, the retrieval accuracy of all retrieval systems decreased because the mean and variance highly influenced the query image features. However, the Fuzzy Rough Set achieved better results compared to other methods in this situation.
- The Genetic Algorithm had the worst result with Poisson noise, compared to other types of noise.

Generally, the Fuzzy Rough Set feature selection provides satisfactory results as the best or second best method when compared against other methods in each type of noise. The reason being that the Fuzzy Rough feature selection removes only the high potential misclassification patterns, rather than eliminate all identified misclassification patterns from the training set. Thus, the proposed method can provide more confidence in the noise identification.

4.4.3 Experiment III: The Image Comparison of the CBIR Systems

In the last experiment of this chapter, the retrieval results for all 10 semantic groups with three kinds of noise are shown. The query images for each semantic group are based on Figure 3.3 (Chapter 3). Gaussian noise, Salt and Pepper noise and Poisson noise are three kinds of noise which is applied to each of the query images in Figure 3.3 (Chapter 3). For each semantic group with the Gaussian noise query image (mean=0 and variance=0.01), the first, second and up to the ninth row in Figures 4.5, 4.8, 4.11, 4.14, 4.17, 4.20, 4.23, 4.26, 4.29 and 4.32 are related to Fuzzy Rough, Genetic Algorithm, Information Gain, PCA and OneR respectively. Figures 4.6, 4.9, 4.12, 4.15, 4.18, 4.21, 4.21, 4.27, 4.30 and 4.33 are related to query images with Salt and Pepper noise (noise density is 0.02). Figures 4.7, 4.10, 4.13, 4.16, 4.19, 4.22, 4.25, 4.28, 4.31 and 4.34 are the result of all ten semantic groups with Poisson noise query image. Referring to Figures4.5 to 4.34, the retrieval system with the Fuzzy Rough Set method shows more related output images to the user. The first left image in Figures 4.5 to 4.34 matched closely to the queried image.

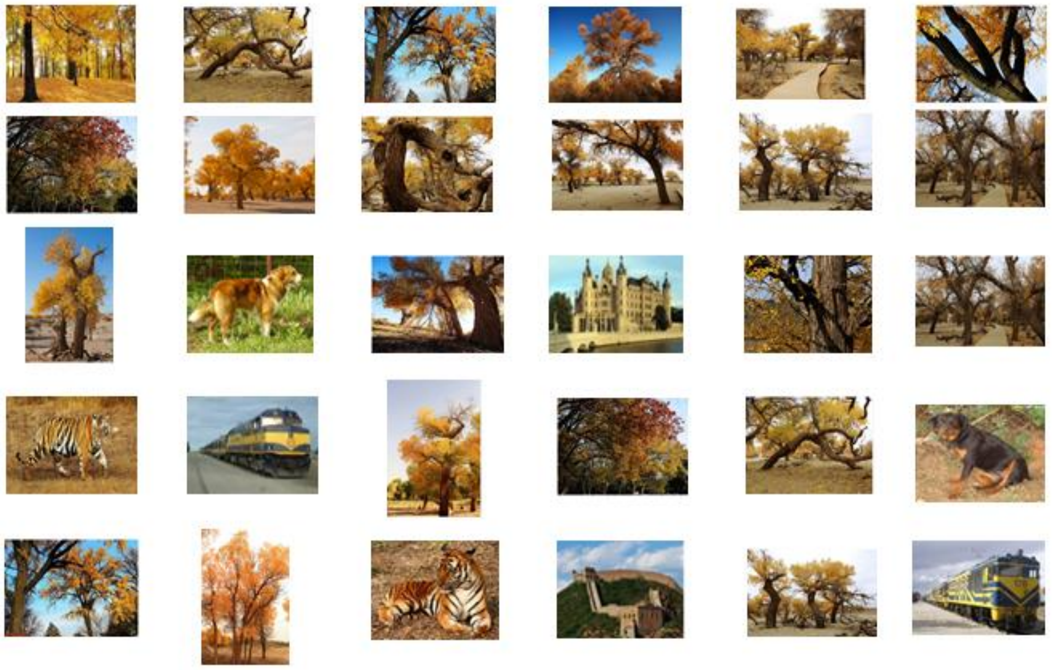


Figure 4.34: Retrieved Images for Autumn Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

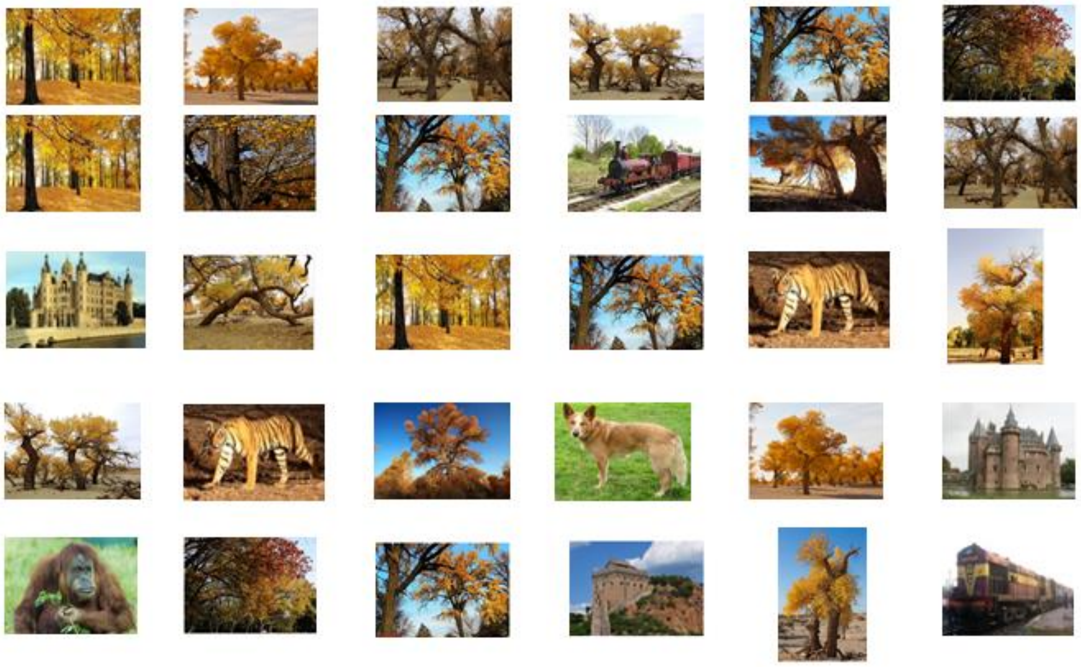


Figure 4.35: Retrieved Images for Autumn Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

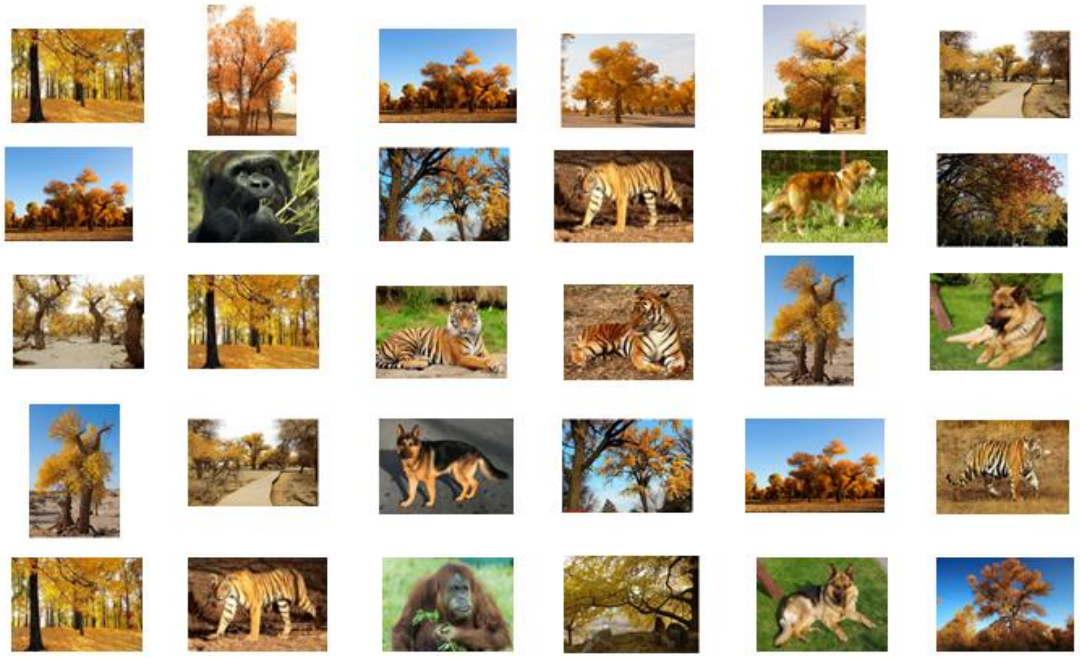


Figure 4.36: Retrieved Images for Autumn Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

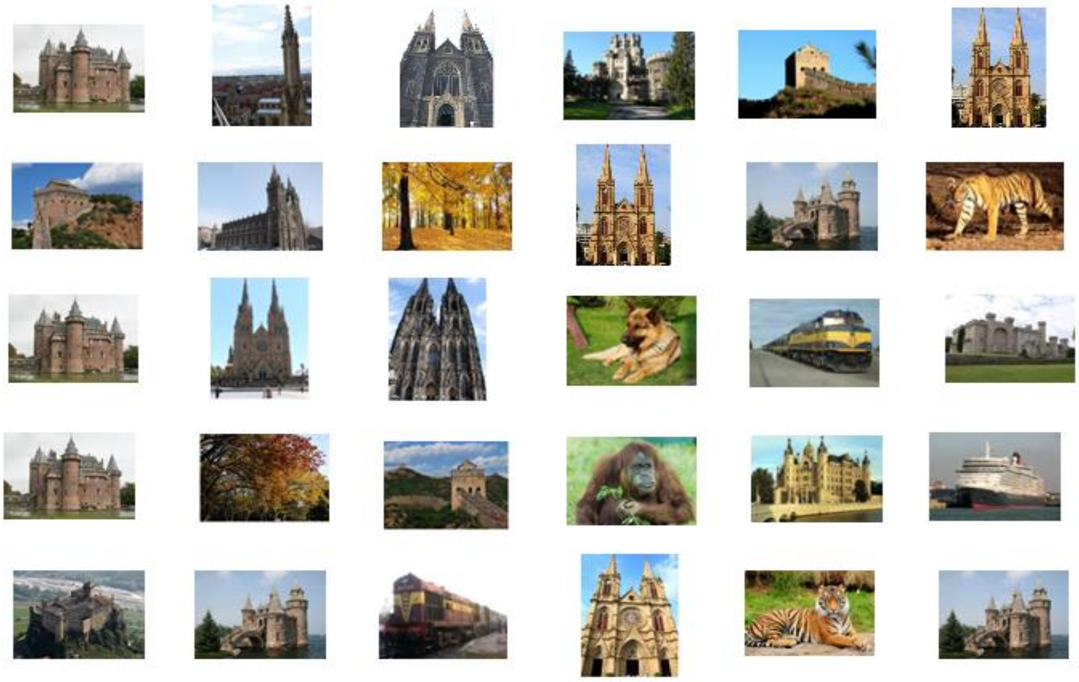


Figure 4.37: Retrieved Images for Castle Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

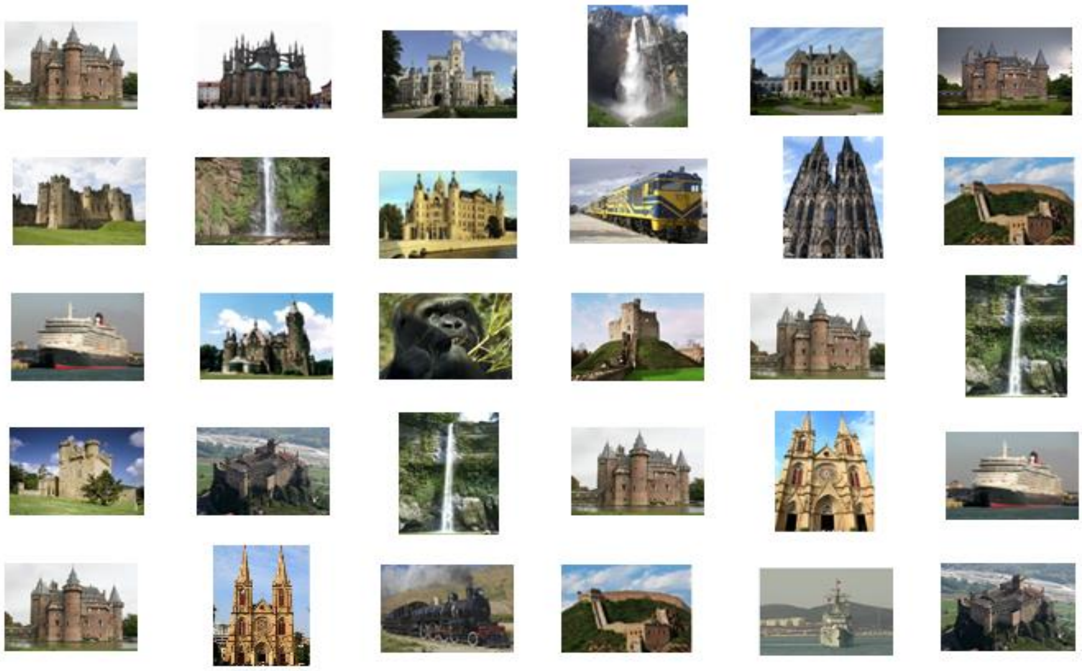


Figure 4.38: Retrieved Images for Castle Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

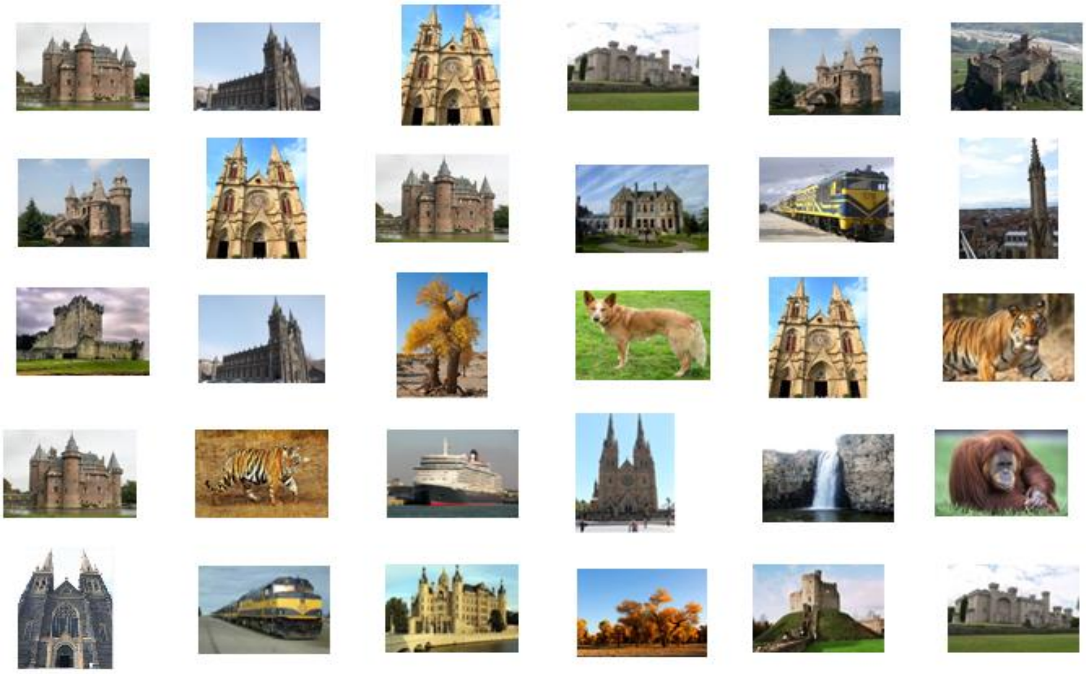


Figure 4.39: Retrieved Images for Castle Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

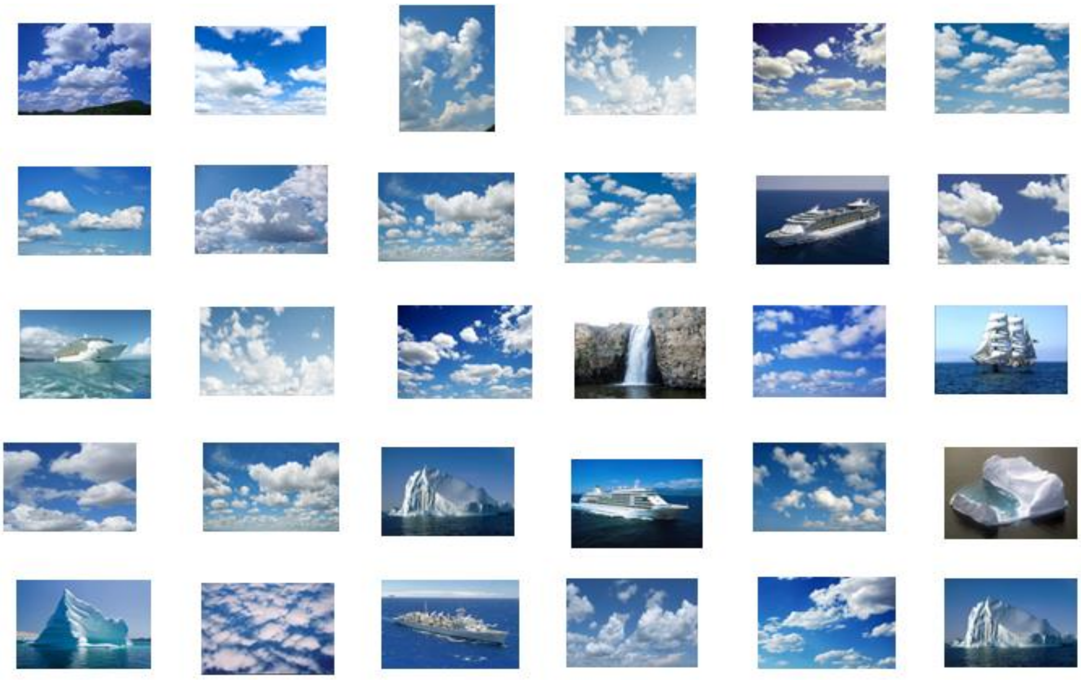


Figure 4.40: Retrieved Images for Cloud Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

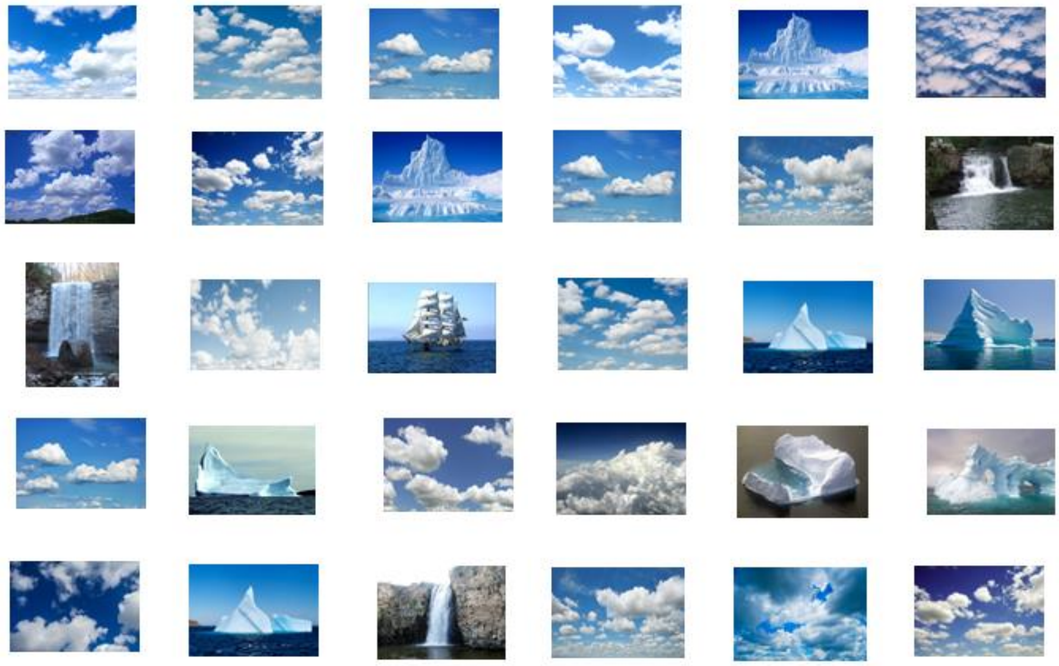


Figure 4.41: Retrieved Images for Cloud Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

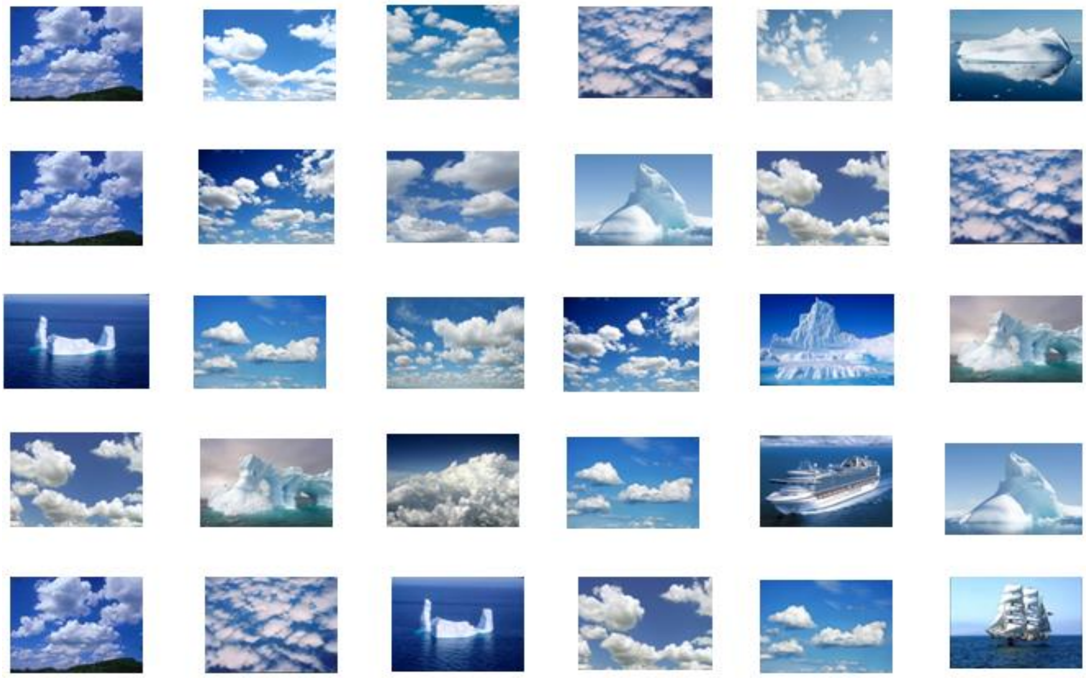


Figure 4.42: Retrieved Images for Cloud Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

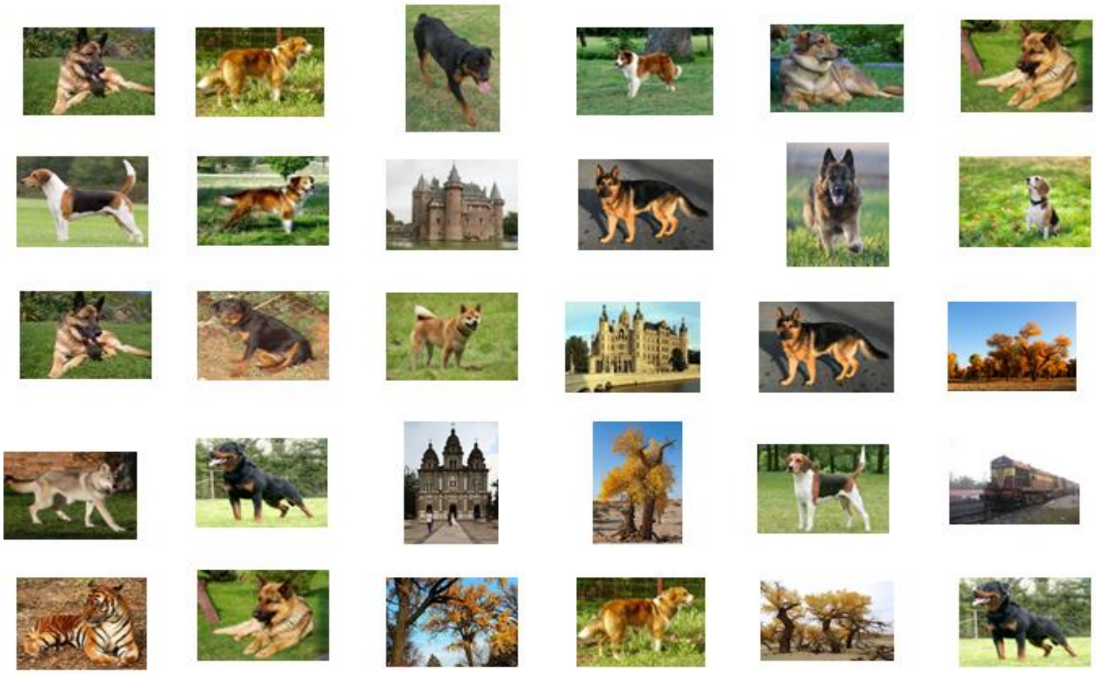


Figure 4.43: Retrieved Images for Dog Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

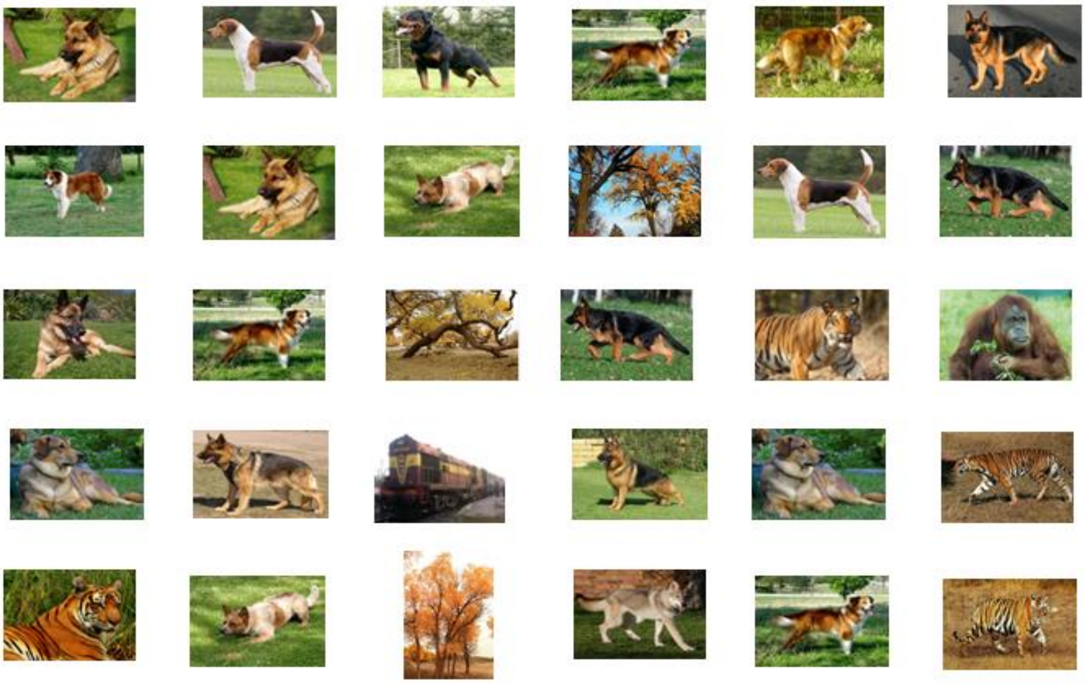


Figure 4.44: Retrieved Images for Dog Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

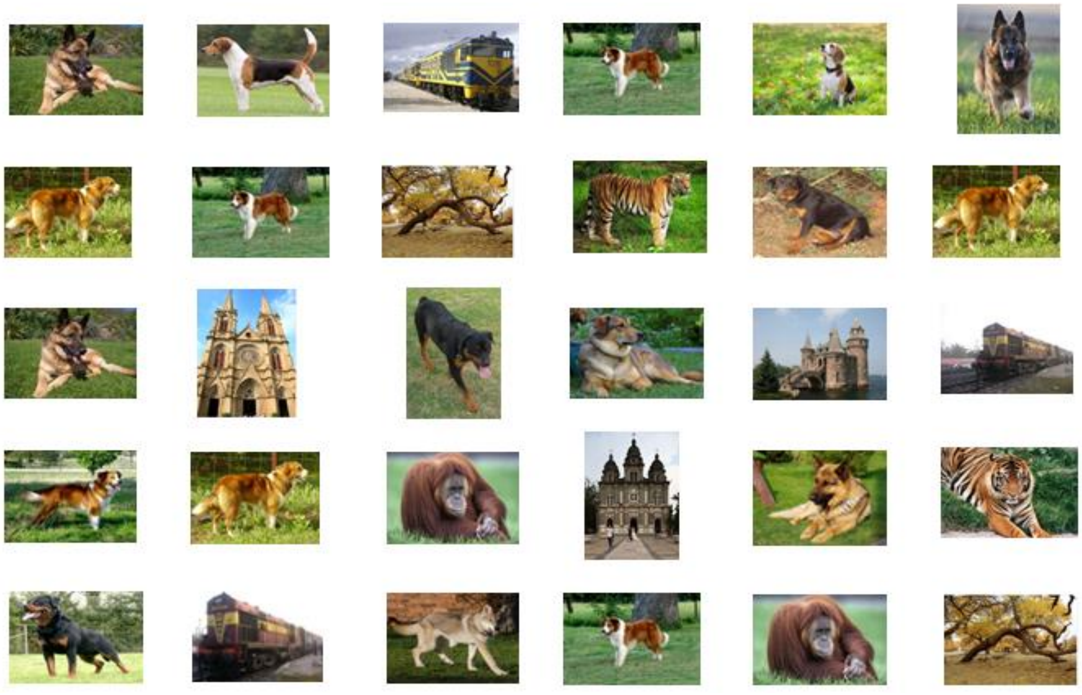


Figure 4.45: Retrieved Images for Dog Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

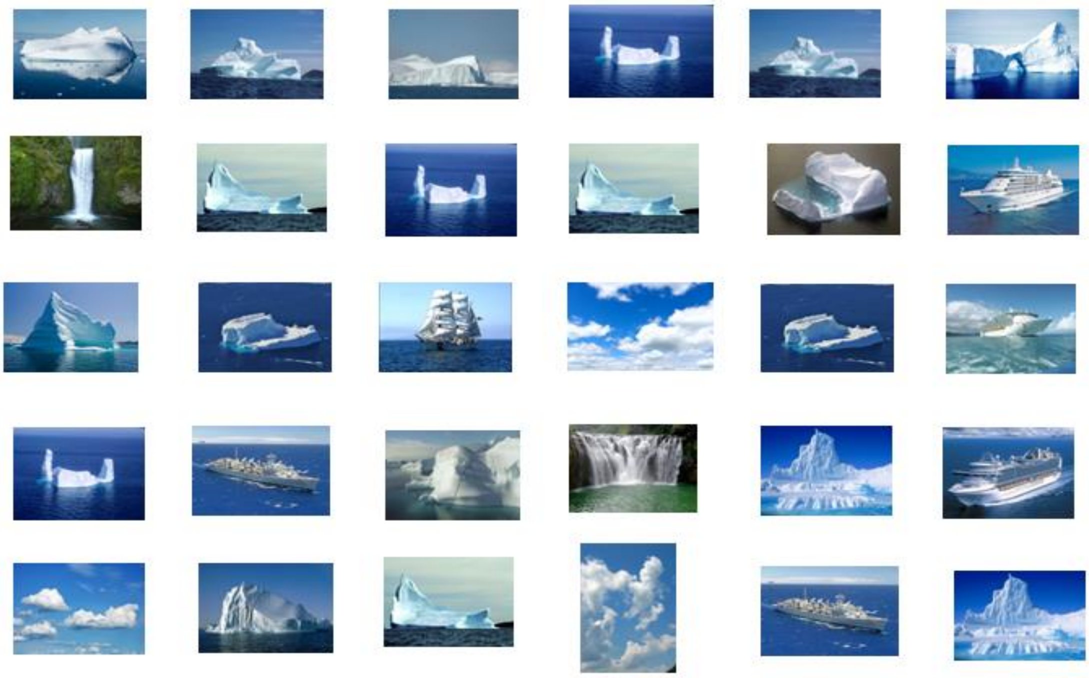


Figure 4.46: Retrieved Images for Iceberg Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

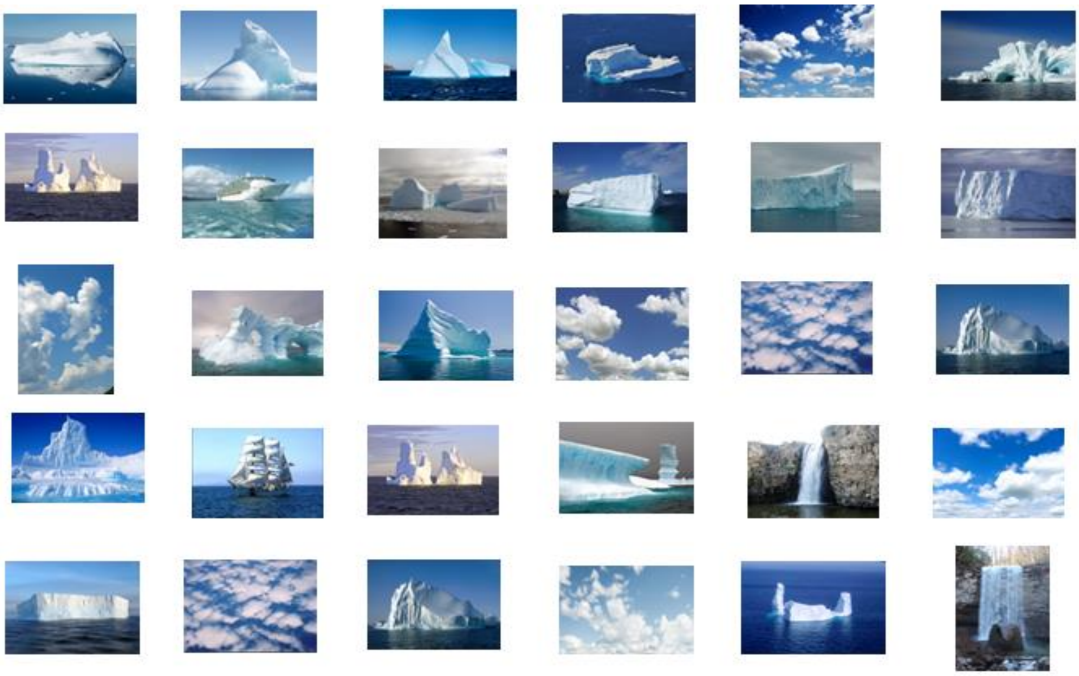


Figure 4.47: Retrieved Images for Iceberg Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

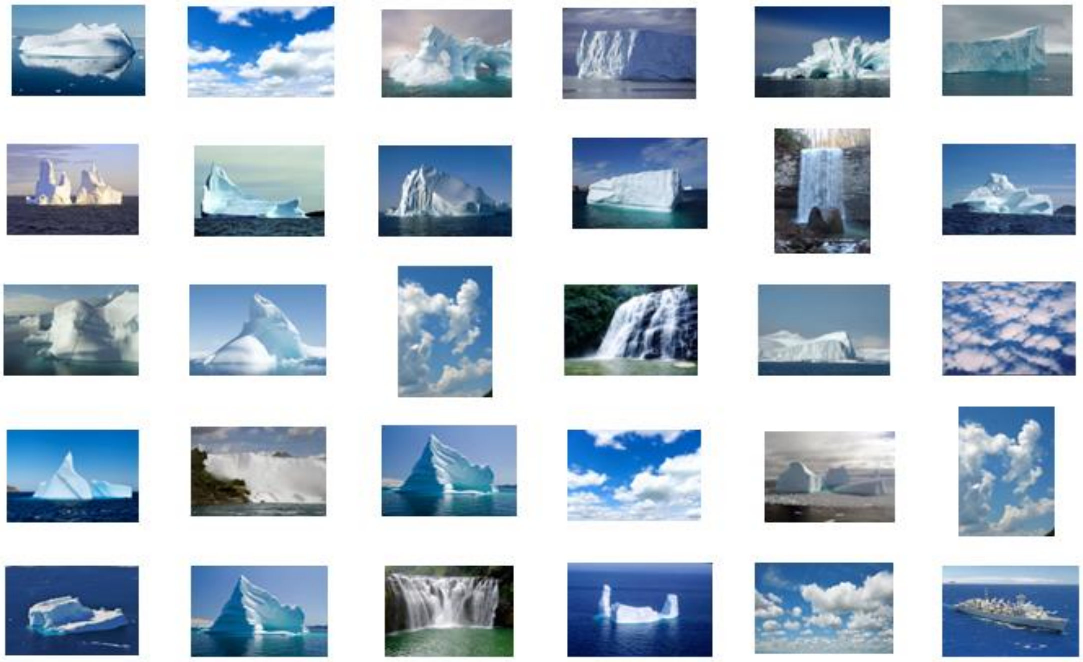


Figure 4.48: Retrieved Images for Iceberg Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

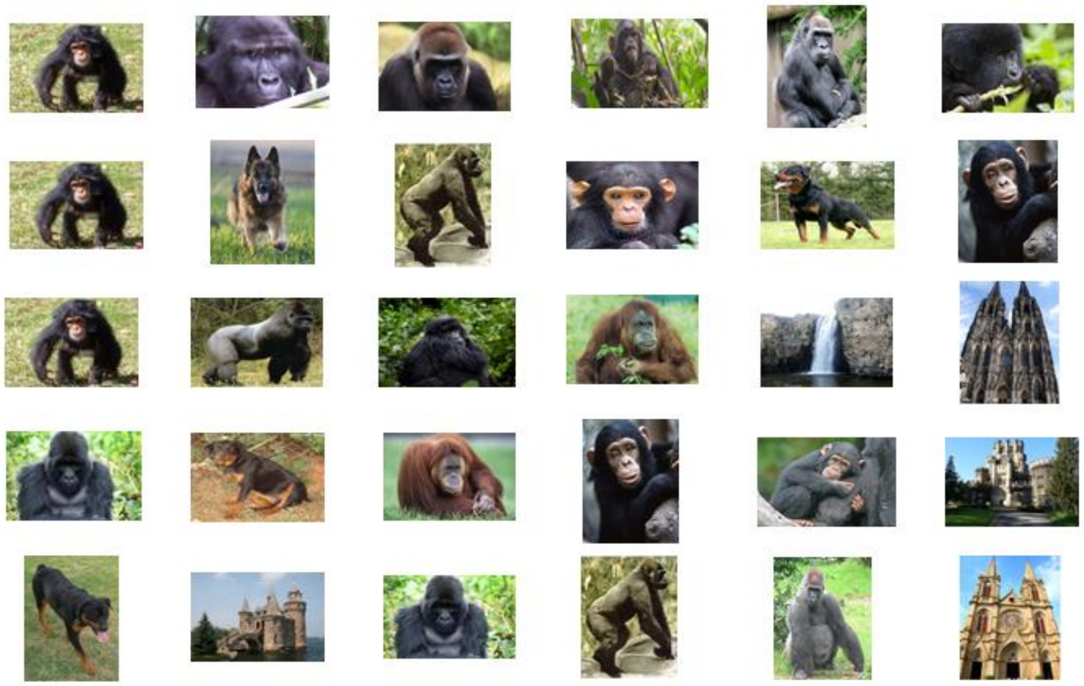


Figure 4.49: Retrieved Images for Primates Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

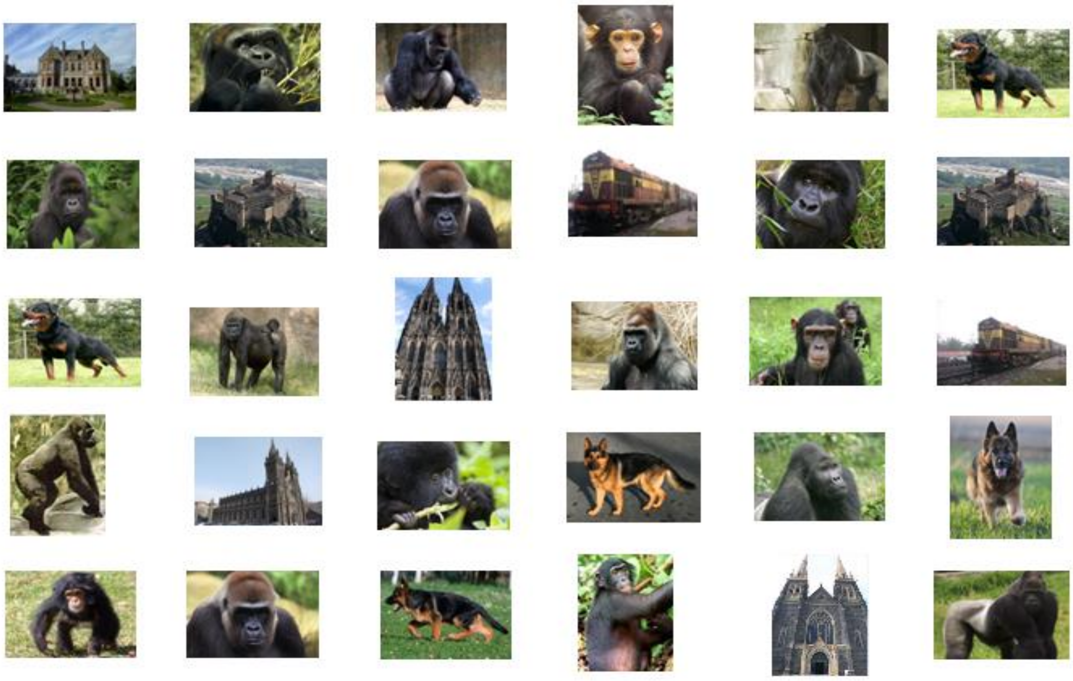


Figure 4.50: Retrieved Images for Primates Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

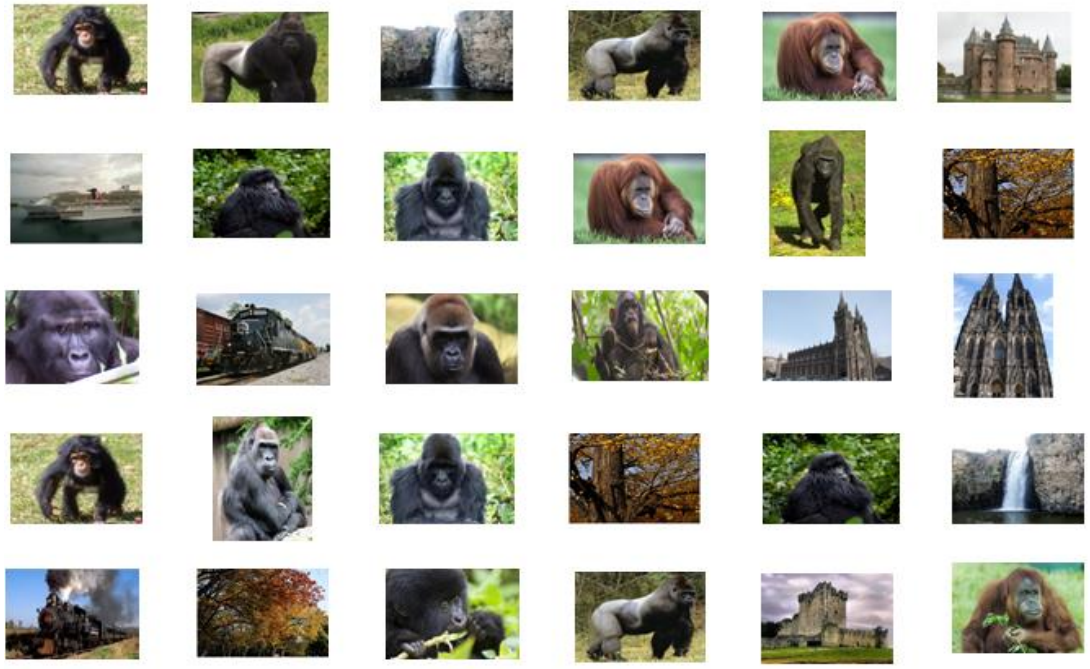


Figure 4.51: Retrieved Images for Primates Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

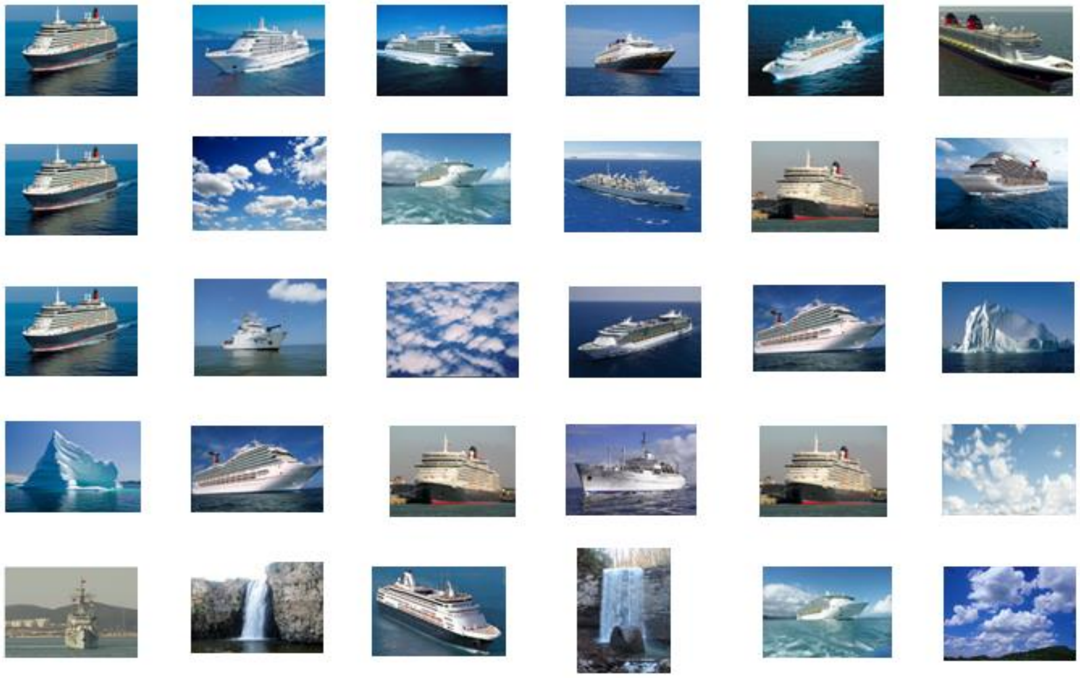


Figure 4.52: Retrieved Images for Ship Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

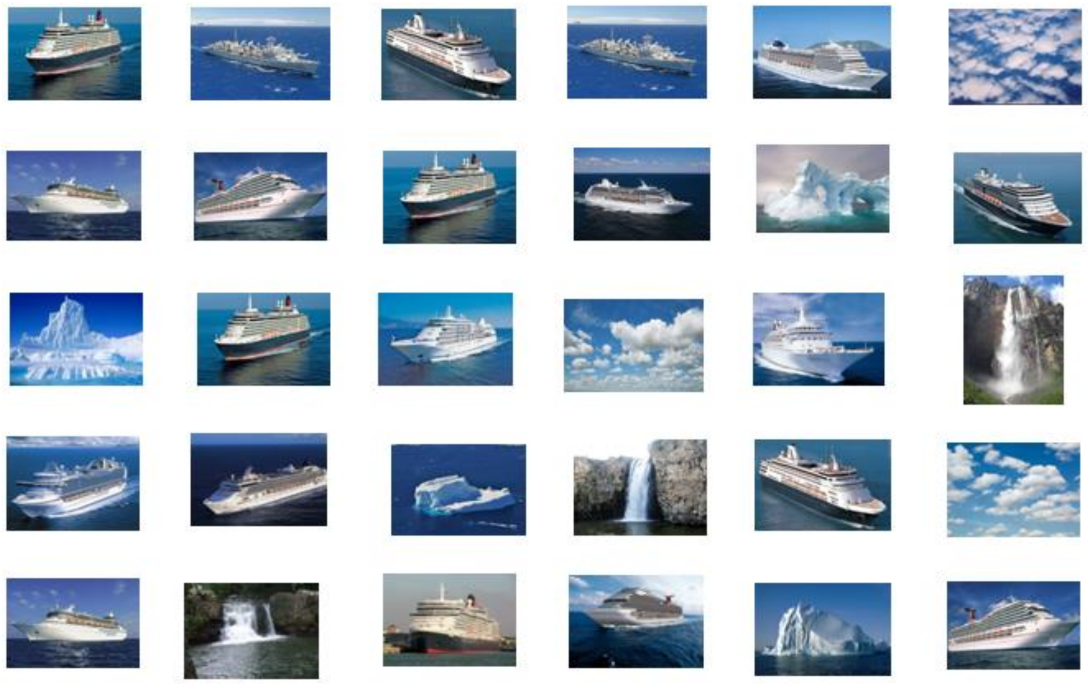


Figure 4.53: Retrieved Images for Ship Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

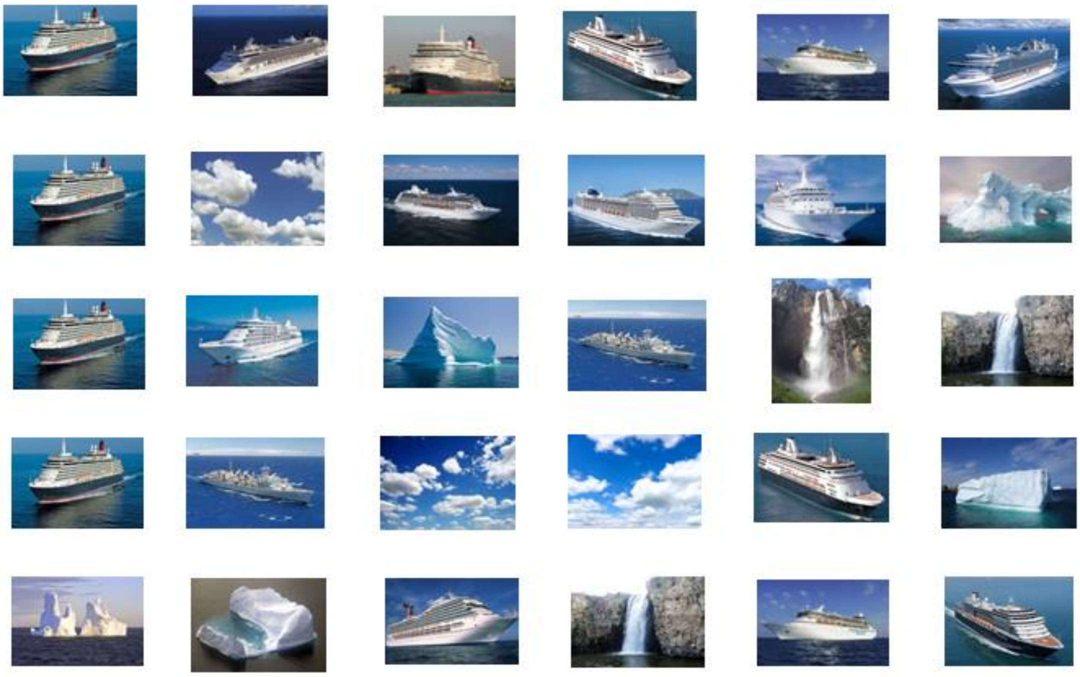


Figure 4.54: Retrieved Images for Ship Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

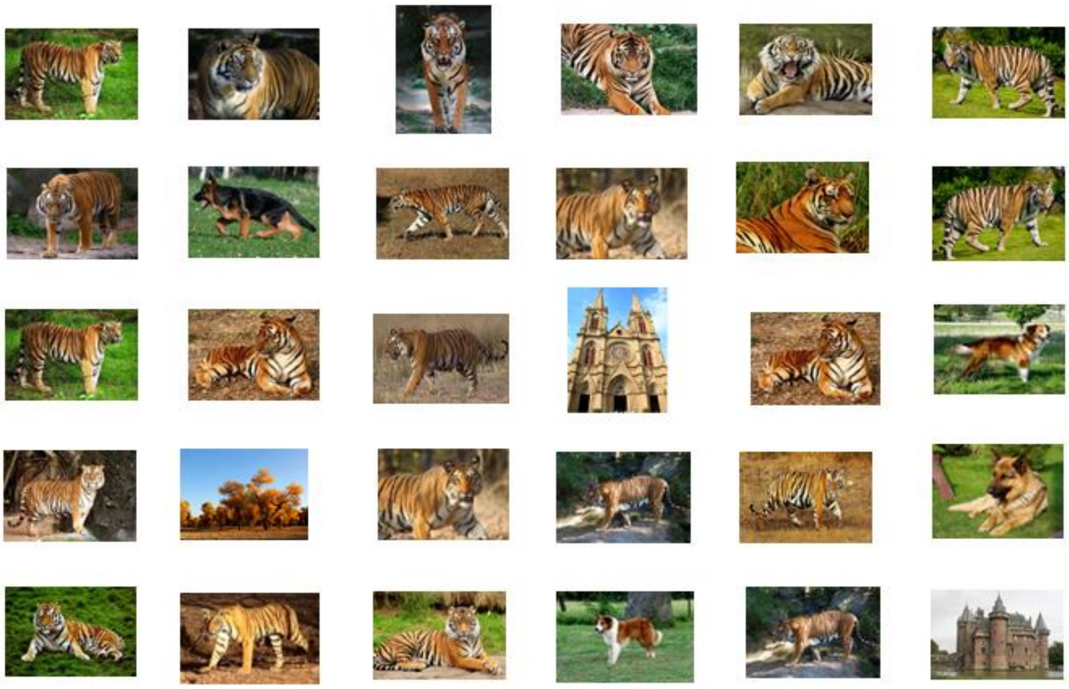


Figure 4.55: Retrieved Images for Tiger Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

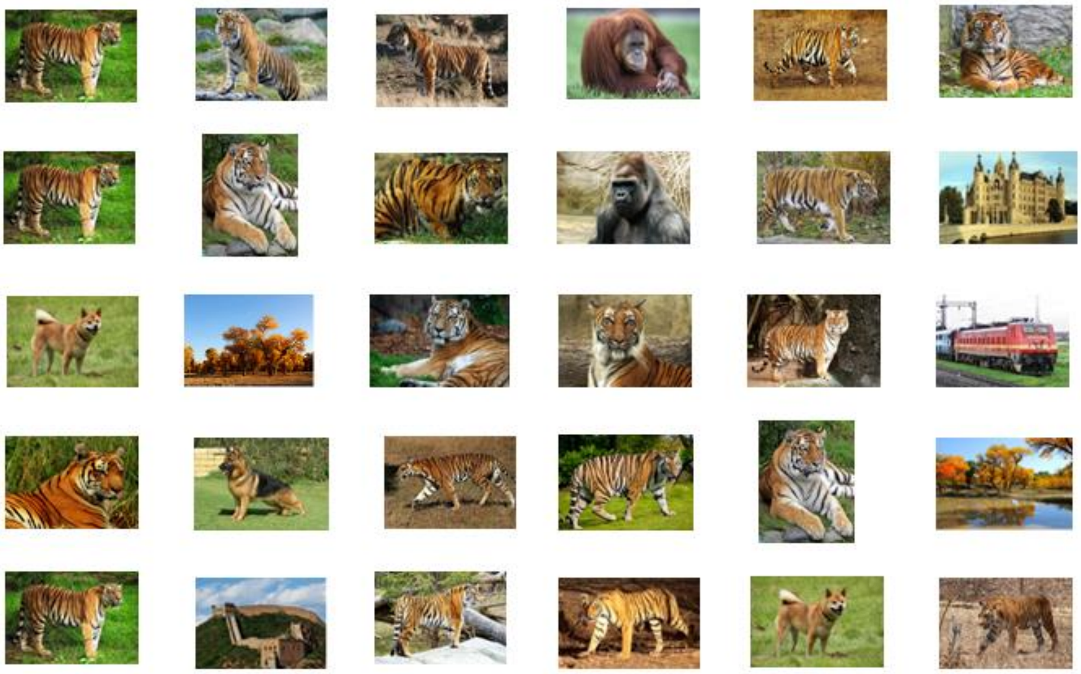


Figure 4.56: Retrieved Images for Tiger Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

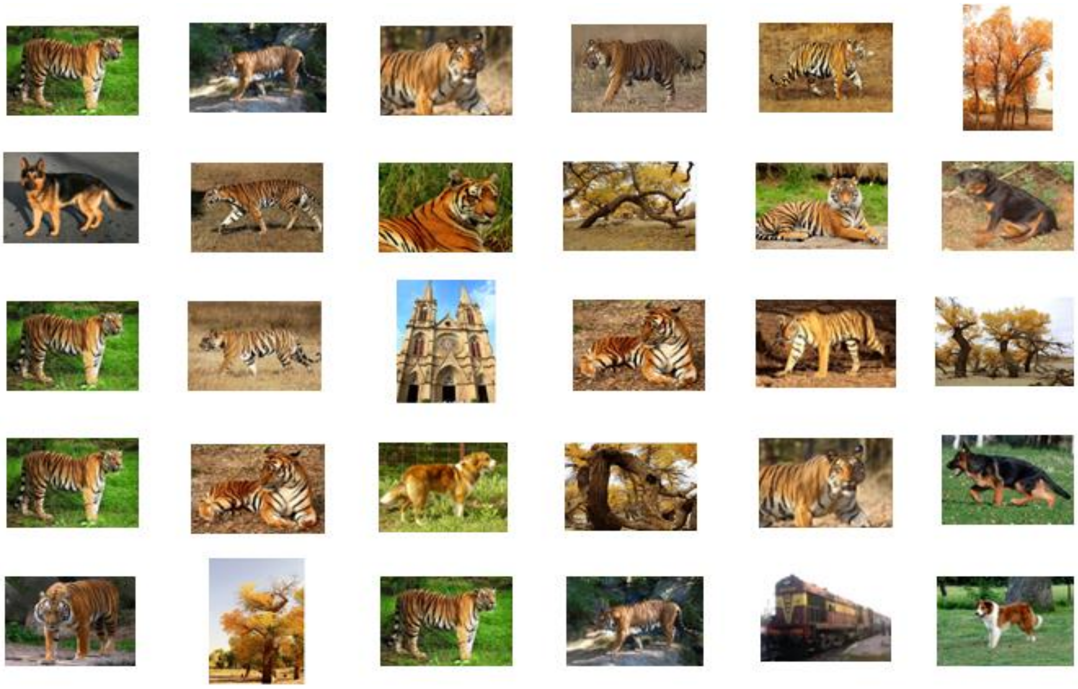


Figure 4.57: Retrieved Images for Tiger Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

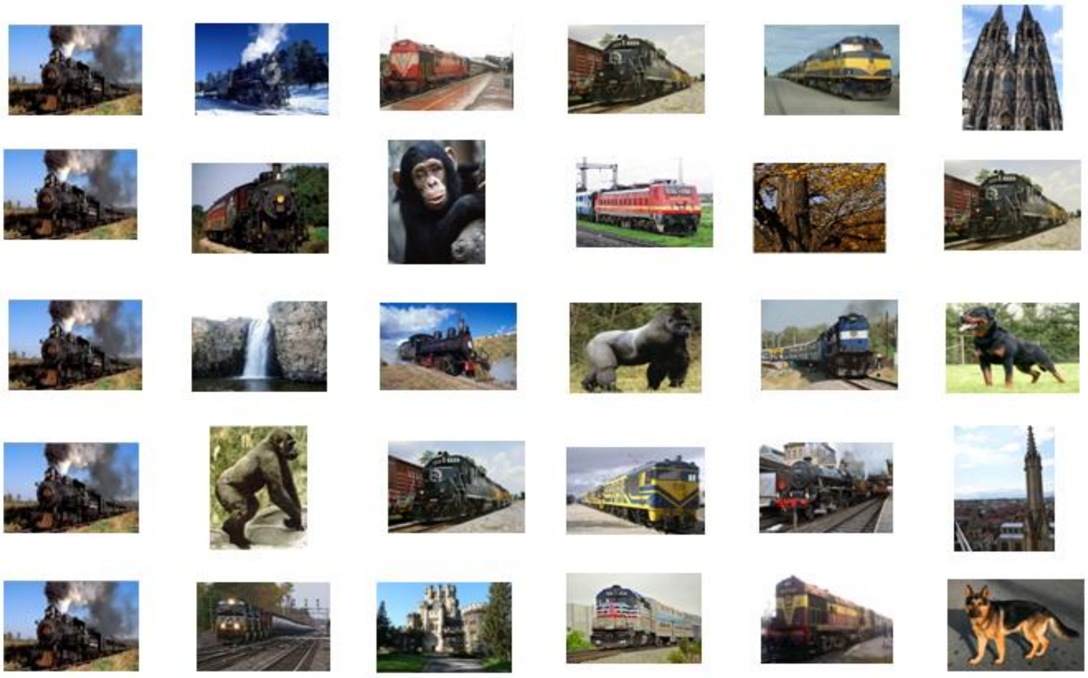


Figure 4.58: Retrieved Images for Train Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

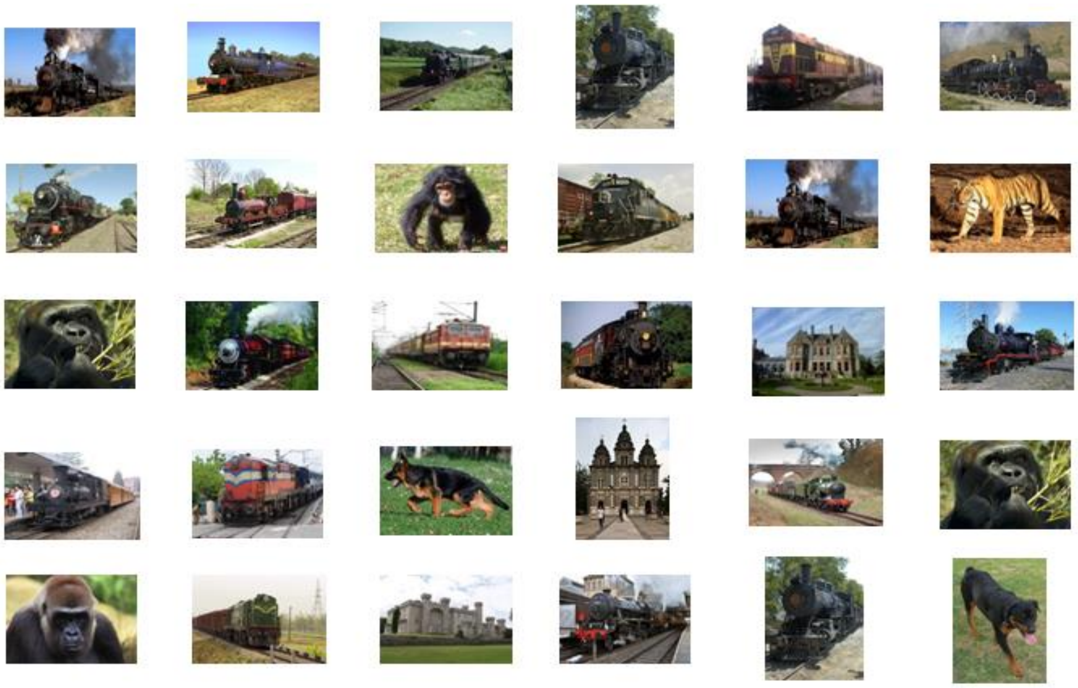


Figure 4.59: Retrieved Images for Train Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

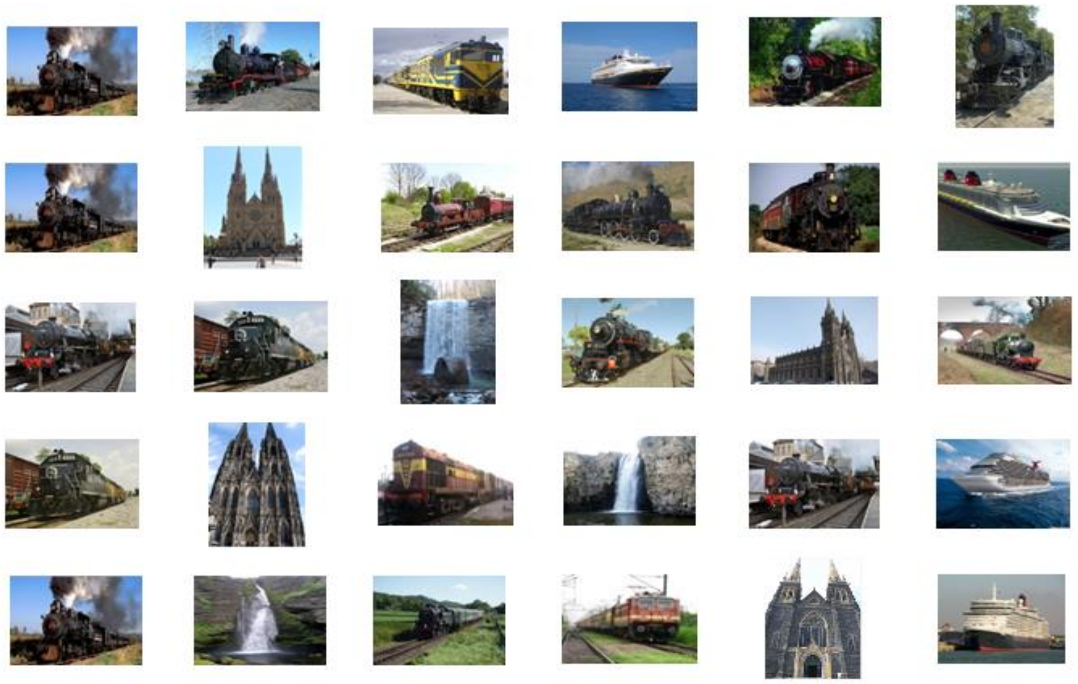


Figure 4.60: Retrieved Images for Train Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

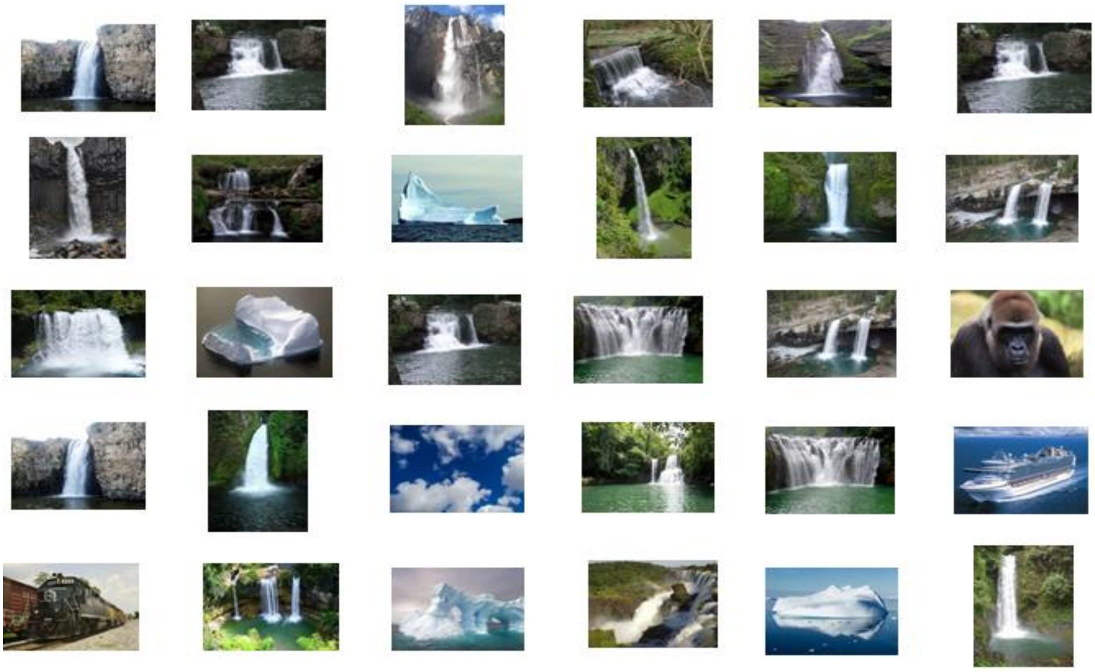


Figure 4.61: Retrieved Images for Waterfall Query Image with Gaussian Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

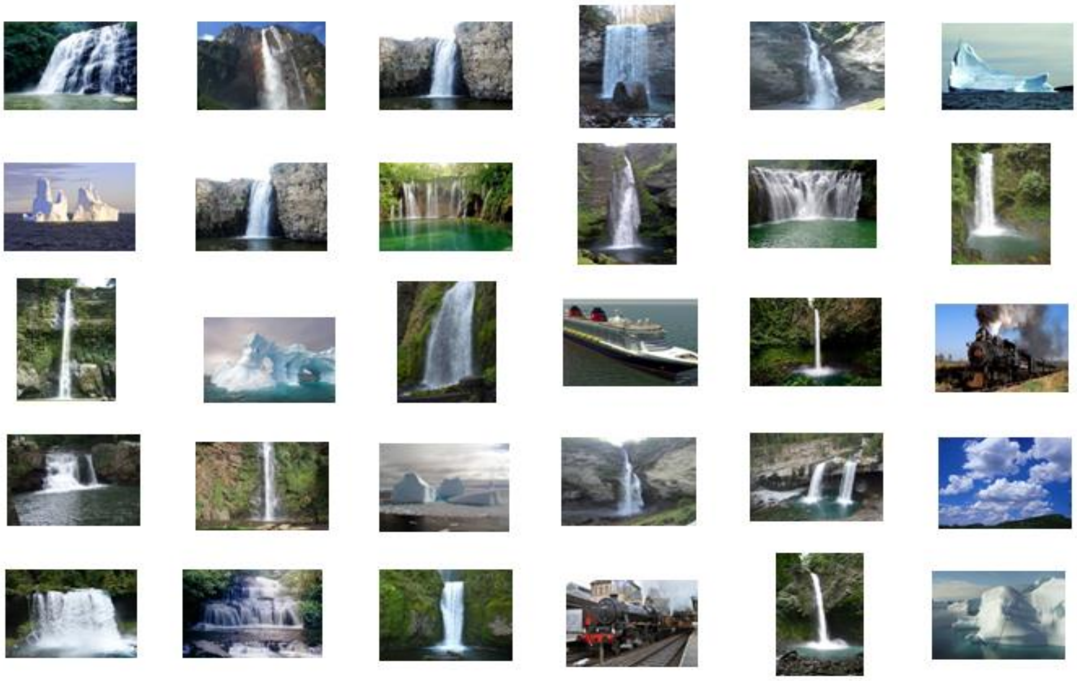


Figure 4.62: Retrieved Images for Waterfall Query Image with Salt and Pepper Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

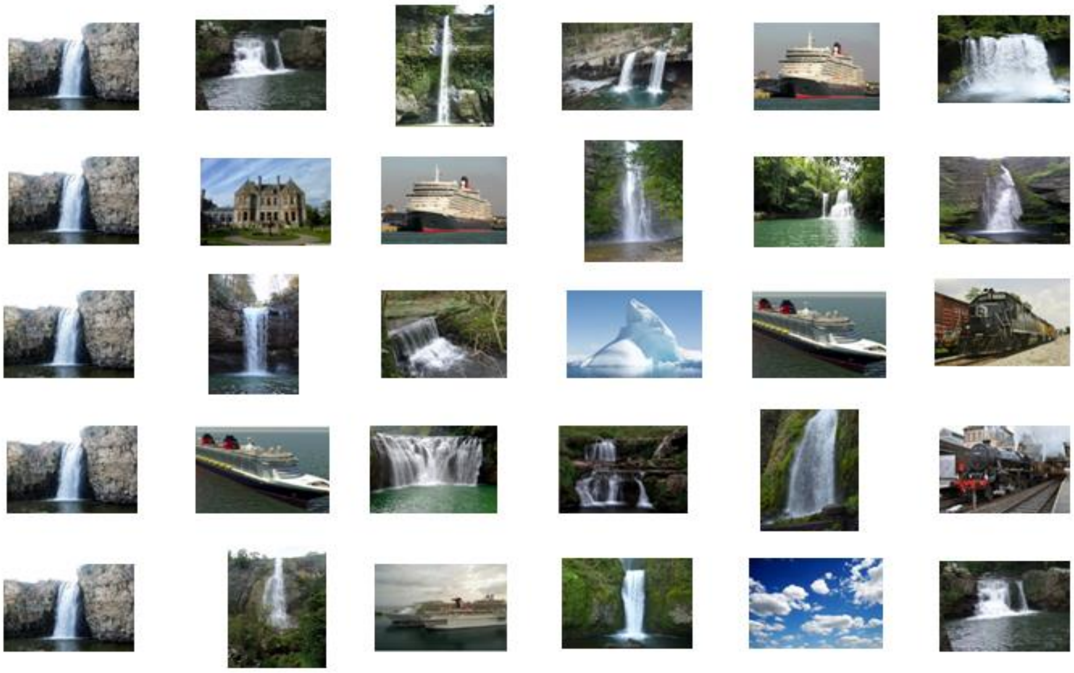


Figure 4.63: Retrieved Images for Waterfall Query Image with Poisson Noise, according to First Row- Fuzzy Rough, Second Row- Genetic Algorithm, Third Row- Information Gain, Fourth Row- PCA and Fifth Row- OneR

The Fuzzy Rough Set consistently shows performance results above the average in most semantic groups among different feature selection methods. This is because when the training features are applied during the pre-processing phase, it can create both human understanding and semantic rules. Human understanding rules are important to ensure that the semantic rules are readily interpretable by the user and that the inference performed is explainable to the user. This can be especially beneficial in a situation where no human experts are available. Subsequently, when the image features are applied by the Fuzzy Rough Set as feature selection, the training features can eliminate most of the possible misclassification features detected by the SVM. Moreover, when a number of features removed from the training sets are compared, it is found that misclassification features eliminated by the Fuzzy Rough Set are greater than other methods.

4.5 Summary

This chapter presents the proposed pre-processing phase of the noisy environment to solve the CBIR system problems. The Fuzzy Rough Set is applied to detect misclassification features in the noisy environment. Gaussian noise, Poisson noise, as well as Salt and Pepper noise, were used to estimate the Fuzzy Rough Set feature selection accuracy in a CBIR system.

To evaluate the Fuzzy Rough Set feature selection, 10 semantic groups from the Corel image dataset including autumn, castle, cloud, dog, iceberg, primates, ship, tiger, train, and waterfall are employed with the experimental results. Some defects were purposely placed in these images via the addition of Gaussian, Poisson and Salt and Pepper noises of different magnitudes using the Matlab software. The Corel image dataset is one of the most important image datasets and is widely used in many papers and researches.

In the experimental results, the Fuzzy Rough Set feature selection was compared with four other feature selection methods. These four feature selection methods are Genetic Algorithm, Information Gain, OneR and Principle Component Analysis. From the experimental results with a noisy queried image, it can be observed that the CBIR system using the Fuzzy Rough Set feature selection has a better retrieval accuracy and Precision-Recall graph when compared to the other four retrieval systems. This is because it can identify and eliminate noise with confidence. Furthermore, the Fuzzy Rough Set feature selection removes only the highly possible misclassification features, rather than eliminating all possible misclassification features.

The drawbacks of these four feature selection methods described in this paper are as follows: (1) In PCA, the computation of the eigenvectors might not be feasible for very high dimensional data, (2) The OneR algorithm is topologically unstable, (3) The Genetic Algorithm cannot find the best features and is stuck in a local maximum; hence, the best features are not guaranteed. Furthermore, it increases the computational complexity and lastly, (4) The Information Gain has a problem when it is applied to features that can take on a large number of distinct values. Based on these drawbacks, the four retrieval systems cannot achieve better results than our proposed feature selection method.

By utilising the Fuzzy Rough Set feature selection method, the proposed system has the advantage that it deals efficiently with an image feature environment that is noisy, vague and uncertain. In addition, the rules extracted from the selecting features of the Fuzzy Rough Set feature selection are semantic and can train the classifier properly. An important advantage of this work is training the SVM with semantic rules that can separate the relevant images from irrelevant ones more accurately. It can be concluded that the Fuzzy Rough Set method is an effective method that can be recommended to handle the CBIR system problems in noisy environments.

Chapter 5: Improved Classifier using Support Vector Machine and Rough Set for the Content Based Image Retrieval System

5.1 Introduction

Image classification is one of the most important aspects of Content Based Image Retrieval (CBIR) systems (Datta et al., 2008; Dharani & Aroquiaraj, 2013; Maryam Shahabi lotfabadi, Shiratuddin, & Wong, 2014). Therefore, using the appropriate classifier for CBIR systems is critical. Many past research used different classifiers for their CBIR systems (Remco C. Veltkamp & Tanase, 2002; Xun et al., 2007; Yildizer, Balci, Hassan, et al., 2012). However, some problems remain. Some of the drawbacks of current classifiers is the lengthy training time (Alattab & Kareem, 2013; Lu, Burkhardt, & Boehmer, 2006), high storage requirements (Dharani & Aroquiaraj, 2013; Xiaohong Yu & Liu, 2009), inability to achieve the required semantic results (Mukhopadhyay, Dash, & Das Gupta, 2013; Xiaohong Yu & Liu, 2009), as well as the inability to deal with incomplete and uncertain data and features (Jyothi & Eswaran, 2010; Z. Liang et al., 2013).

A combination of the Rough Set and two types of Support Vector Machine (SVM), which includes 1-v-1 (one-versus-one) SVM and 1-v-r (one-versus-all) SVM as the classifiers, are proposed by Pawan Lingras and CoryJ Butz in (Lingras & Butz, 2005). In (Lingras & Butz, 2005), these two classifiers have many advantages which could address some of the problems that exist in the CBIR systems. However, to our knowledge, these techniques have not been used in CBIR systems before. Thus, the purpose of this chapter is to examine the suitability of these two classifiers for a CBIR system. A faster and more accurate CBIR system is required for real-time application. This can be achieved by employing a classifier such as the SVM. However, the SVM has some problems as mentioned below. These problems can be reduced using the Rough Set, so it is worth examining the Rough Set with 1-v-1 SVM and 1-v-r SVM in CBIR systems. In the 1-v-1 SVM, one SVM is constructed for each pair of the classes (Xiaoyuan Zhang, Zhou, Guo, Zou, & Huang, 2012). However, the largest class or group is classified first. In the 1-v-r SVM, the positive region of this group is eliminated from further classification so that the positive region will definitely have features or objects related to this group (Lingras & Butz, 2007). Further classification will be done on the negative and boundary regions. This process continues until all of the groups are classified.

The reason why the Rough Set with 1-v-1 SVM and 1-v-r SVM as a classifier have better results compared to the conventional SVM classifier is because the conventional 1-v-1 SVM has high storage requirements and lack the semantic interpretation of the classification process (Lingras & Butz, 2007; Xiaoyuan Zhang et al., 2012). The Rough Set can reduce the storage requirements by using upper and lower approximations. This means conventional 1-v-1 SVM needs $N \times (N-1)/2$ rules. However, this amount is reduced to $2 \times N$ for the classifier and it includes the Rough Set and 1-v-1 SVM (Lingras & Butz, 2007). In addition, the combination of the Rough Set and 1-v-1 SVM can provide a better semantic interpretation of the classification process using properties of the Rough Set boundary region (Lingras & Butz, 2005).

Furthermore, conventional 1-v-r SVM also has issues with respect to long training time, low training performance and the fact that it cannot deal with noisy data (Lingras & Butz, 2007; Xiaoyuan Zhang et al., 2012). However, the Rough Set can reduce the training time and improve the training performance using lower approximations (Derrac et al., 2012), by omitting positive region from further classifications (Lingras & Butz, 2007). The combination of Rough Set and 1-v-r SVM can provide better results in noisy areas using the properties of the Rough Set boundary region (J. Chen & Li, 2012; Xianyong Zhang, Mo, Xiong, & Cheng, 2012).

The main focus of this chapter is to investigate the Rough Set with 1-v-1 SVM classifier and Rough Set with 1-v-r SVM classifier in a CBIR system and evaluate the performance of these two classifiers with the image features. These two Content Based Image Retrieval systems are compared with other image retrieval systems that use Decision Tree (C5.0), K-nearest neighbour, Neural Network, and Support Vector Machine as the classifier in their methodology. These four classifiers are some of the more popular classifiers (Xiang-Yang Wang, Zhang, & Yang, 2013; Zhu et al., 2010), which are used in different CBIR systems and provide reasonable results. Therefore, comparing the Rough Set with 1-v-1 SVM and the Rough Set with 1-v-r SVM classifiers to these four classifiers (Decision Tree (C5.0), K-nearest neighbour, neural network, and Support Vector Machine) show robustness and effectiveness of both the Rough Sets. Also, the experiments are carried out using similar Corel image datasets used in other chapters in order to test the accuracy and robustness of the classifiers.

This chapter is organised as follows. Section 5.2 presents Rough Set to 1-v-1 SVM and 1-v-r SVM classifiers. In Section 5.3, the experiment setup is described and in Section 5.4, the experimental results are presented. Finally, the summary is presented in the last section.

5.2 Rough Set Method to Support Vector Machine 1-v-1 and 1-v-r Multi-Classifiers

This section describes the Rough Set method to SVM 1-v-1 and 1-v-r multi-classifiers as proposed in (Lingras & Butz, 2005). First, a non-linear separable feature space is transformed to a linear separable feature space using a Radial Basis Function Kernel (RBF Kernel). The reason for choosing this kernel is the RBF Kernel has better results in CBIR systems (Maryam Shahabi Lotfabadi & Mahmoudie, 2010). The perfect situation is that the SVM can find the hyper-plane by maximising the margin between the two classes, and no example is found in the margin, i.e. after transforming the non-linear feature space into the linear feature space (Chapelle, Vapnik, Bousquet, & Mukherjee, 2002; Cortes & Vapnik, 1995) (see Figure 5.1).

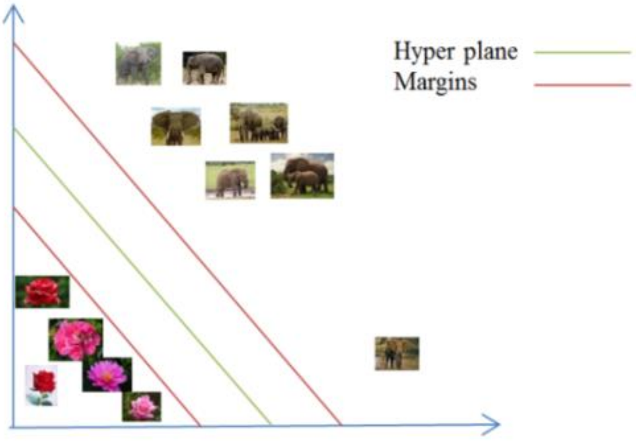


Figure 5.64: Maximising the margin between the two classes

However, when there are some examples between the margin, applying a method like the Rough Set, which can deal with vague and uncertain spaces, is essential. The margin can be used as the Rough Set boundary region. Using the formulas shown below, the Rough Set is applied to the SVM and b_1 and b_2 correspond to the boundaries of the margin in Figure 5.3- 5.7 (red lines).

b_1 is defined as follows: $y \times [< x, w > +b_1] \geq 0$, for all $(x, y) \in S$, and there exists at least one training example $(x, y) \in S$ such that $y = 1$ and $y \times [< x, w > +b_1] = 0$.

b_2 is defined as follows: $y \times [< x, w > +b_2] \geq 0$, for all $(x, y) \in S$, and there exist at least one training example $(x, y) \in S$ such that $y = -1$and $y \times [< x, w > +b_2] = 0$.

The above variables are defined as follows: Assume x is an input vector in the input space X and y is the output in $Y = \{-1, +1\}$. The training set used for supervised classification is $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots\}$. $< x, w >= \sum_j x_j \times w_j$ is the inner product and x_j and w_j are components of two vector x and w.

According to R1, R2 and R3 rules, a Rough Set based SVM binary classifier can be defined when:

[R1] If $< x, w > +b_1 \geq 0$, classification of x is positive (+1).

[R2] If $< x, w > +b_2 \leq 0$, classification of x is negative (-1).

[R3] Otherwise, classification of x is uncertain.

In the SVM 1-v-1 multi-classifier, one binary SVM is constructed for each pair of classes (i, j). According to the rules R1, R2 and R3, three equivalence classes can be defined for each pair. $P_{ij}(POS_i)$, $P_{ij}(POS_j)$ and $P_{ij}(BND)$ are the set of x (or region) that follows the rules R1, R2 and R3 respectively. The lower ($\underline{A}(class_i)$) and upper approximations ($\overline{A}(class_i)$), as well as the boundary regions for class i and j are summarised in Table 5.1.

Table 5.7: Lower and upper approximations and boundary region for class **i** and **j**

	Lower approximation	Upper approximation
Class i	$P_{ij}(POS_i)$	$P_{ij}(POS_i) \cup P_{ij}(BND)$
Class j	$P_{ij}(POS_j)$	$P_{ij}(POS_j) \cup P_{ij}(BND)$
Overall lower approximation for class i (N is number of classes)	$\underline{A}(class_i) = \bigcap_{\substack{j=1 \\ j\neq i}}^{N} P_{ij}(POS_i)$	
Overall Boundary region for class i (N is number of classes)	$\overline{A}(class_i) - \underline{A}(class_i)$ $= \bigcup_{\substack{j=1 \\ j\neq i}}^{N} P_{ij}(BND) - \bigcup_{j=1}^{N} \underline{A}(class_j)$	

Overall upper approximation for class i (N is number of classes)	$\overline{A}(class_i) = \bigcup_{\substack{j=1 \\ j\neq i}}^{N} P_{ij}(BND) - \bigcup_{j=1}^{N} \underline{A}(class_j) + \underline{A}(class_i)$
Some rules extracted from the above formula	$P_{ij}(POS_i) \cap P_{ij}(POS_j) = \emptyset;$ $\underline{A}(class_i) \subseteq P_{ij}(POS_i)$; $\underline{A}(class_j) \subseteq P_{ij}(POS_j)$ $\underline{A}(class_i) \cap \underline{A}(class_j) = \emptyset$

A classification problem with the three classes, i.e. Flower, Elephant and African people, is shown in Figure 5.2. Figures 5.3, 5.4 and 5.5 show the Rough Set method to SVM 1-v-1 classification for the classes, Flower and Elephant, Flower and African people, as well as Elephant and African people respectively.

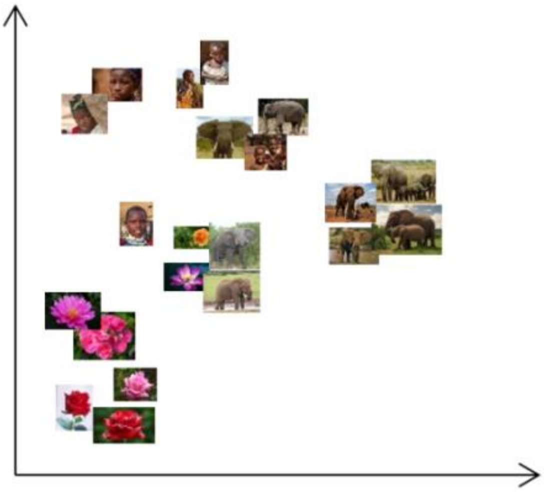


Figure 5.65: A classification problem including the three classes - Flower, Elephant and African people

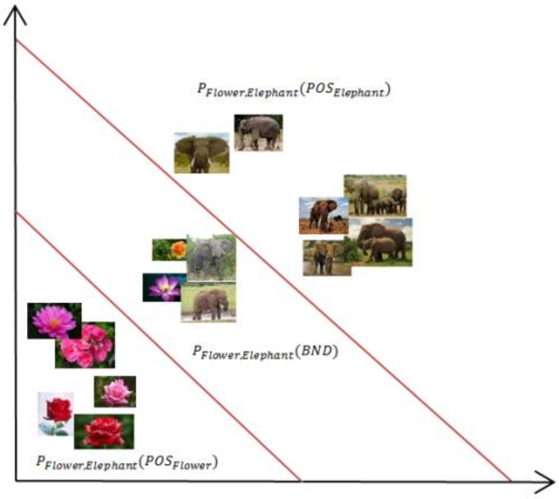

$P_{Flower,Elephant}(POS_{Elephant})$
$P_{Flower,Elephant}(BND)$
$P_{Flower,Elephant}(POS_{Flower})$

Figure 5.66: A Rough Set method to SVM 1-v-1 classification for the classes, Flower and Elephant

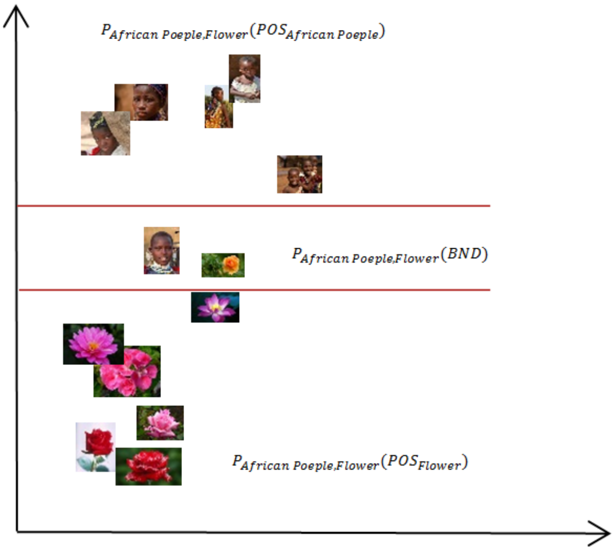

$P_{African\ Poeple,Flower}(POS_{African\ Poeple})$
$P_{African\ Poeple,Flower}(BND)$
$P_{African\ Poeple,Flower}(POS_{Flower})$

Figure 5.67: A Rough Set method to SVM 1-v-1 classification for the classes, Flower and African people

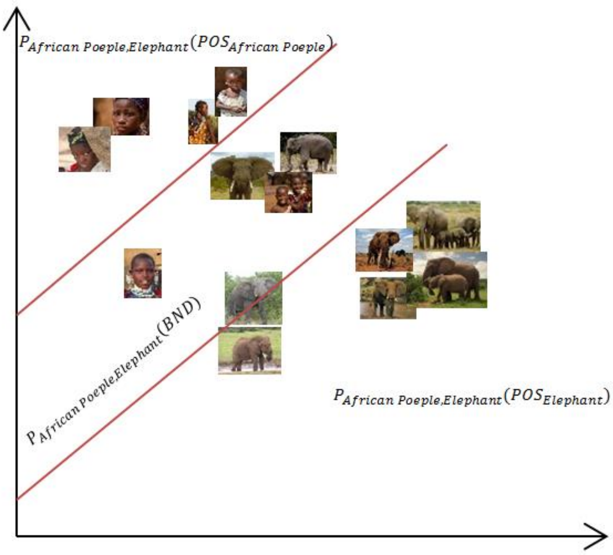

$P_{African\ Poeple,Elephant}(POS_{African\ Poeple})$
$P_{African\ Poeple,Elephant}(BND)$
$P_{African\ Poeple,Elephant}(POS_{Elephant})$

Figure 68.5: A Rough Set method to SVM 1-v-1 classification for the classes, Elephant and African people

According to rules R1, R2 and R3, the training sample, 1-v-r strategy and three equivalence classes, $Q_{Flower}(POS), Q_{Flower}(BND), Q_{Flower}(NEG)$, for the class Flower are created.

$Q(POS)$ is the set of regions that follows the rule [R1], $Q(NEG)$ is the set of regions that follows rule [R2] and $Q(BND)$ is the set of regions that follows rule [R3]. By creating $Q_i(POS)$, $Q_i(BND)$, $Q_i(NEG)$ for each subsequent class i, $1 < i < N$ (class $i-1$ has more object than class i), $Q_{i-1}(BND) \cup Q_{i-1}(NEG)$ is refined. $Q_N(POS) = Q_{N-1}(NEG)$ and $Q_N(BND) = Q_{N-1}(BND)$ are defined for the last class. Lower and upper approximations for all N classes and boundary regions are represented in Table 5.2.

Table 5.8: Lower, Upper and Boundary regions

Upper approximation for all N classes	$\underline{A}(class_i) = Q_i(POS)$
Lower approximation for all N classes	$\overline{A}(class_i) = Q_i(POS) \cup Q_i(BND)$
Boundary region	$Q_i(BND) = Q_i(BND) - \bigcup_{j=i}^{N} Q_j(POS)$

The classification problem with the three classes - Flower, Elephant, and African people is shown in Figure 5.2. Using R1, R2 and R3 rules, $Q_{Flower}(POS)$, $Q_{Flower}(BND)$, $Q_{Flower}(NEG)$ are calculated for the class Flower in Figure 5.6. Images in the region $Q_{Flower}(POS)$ definitely belong to class Flower. $Q_{Flower}(NEG)$ corresponds to images that do not belong to class Flower, while $Q_{Flower}(BND)$ may or may not belong to the class Flower. There is no need to further classify images in $Q_{Flower}(POS)$ because it only contains images belonging to class Flower (black and white flower images in Figure 5.7). However, $Q_{Flower}(BND) \cup Q_{Flower}(NEG)$ should be further refined.

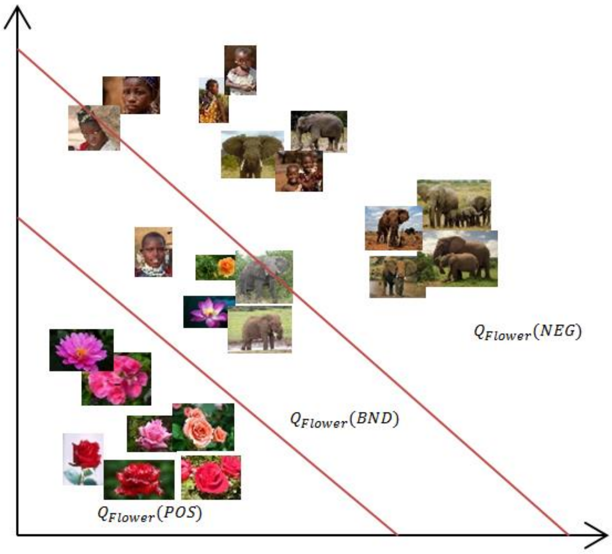


Figure 5.69: A Rough Set method to 1-v-r classification for class Flower.

The classification results are shown in Figure 5.7 for the next class (class Elephant). 1-v-r classification can identify the images that definitely belong to class Elephant. In 1-v-r support vector machine, until the number of classes is reduced to two, the process will be further repeated. According to the algorithm, these two equations are extracted:

$$Q_{Elephant}\ (NEG) = Q_{African\ People}(POS)$$

$$Q_{African\ People}(BND) = Q_{Elephant}(BND)$$

Figure 5.7 shows the final classification using the Rough Set method to SVM 1-v-r classification.

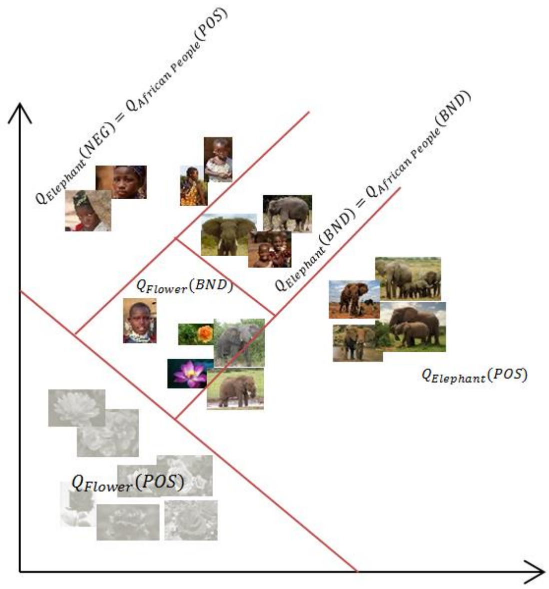

$Q_{Elephant}(NEG) = Q_{African\ People}(POS)$
$Q_{Elephant}(BND) = Q_{African\ People}(BND)$
$Q_{Flower}(BND)$
$Q_{Elephant}(POS)$
$Q_{Flower}(POS)$

Figure 5.70: A Rough Set method to 1-v-r classification for classes, Elephant and African People

5.3 Experimental Results with Rough Set 1-v-1 SVM and Rough Set 1-v-r SVM Classifiers

In this section, the results that compare the four retrieval systems with the Rough Set 1-v-1 SVM and the Rough Set 1-v-r SVM retrieval systems are presented. These four retrieval systems used Decision Tree (C5.0), Support Vector Machine (SVM) (Xiaohong Yu & Liu, 2009), Neural Network (NN) (Jyothi & Eswaran, 2010) and K-nearest Neighbour (KNN) (Lu et al., 2006) as the classifiers in their methodology.

To investigate the function of the image retrieval system based on the above-mentioned methods, we used the COREL image database that is the same as the last chapters. In the following two sub-sections, a comparison of the Precision-Recall Graph and retrieval accuracy is calculated for all the classifiers.

5.3.1 Experiment I: Precision-Recall Graph

Recall equals to the number of the related retrieval images to the number of the related images available in the image database. Precision equals to the number of the related retrieval images to all of the retrieval images (Dharani & Aroquiaraj, 2013). Figure 5.8 shows the Precision-Recall graph for 10 semantic groups that are used for measuring the efficiency of the proposed retrieval system. In Figure 5.8, the Rough Set 1-v-1 SVM and Rough Set 1-v-r SVM are labelled as RS 1-v-1 SVM and RS 1-v-r SVM respectively. From the graph, we observed that the Rough Set 1-v-1 SVM and Rough Set 1-v-r SVM retrieval systems achieved better results than the other four systems.

In order to explain why this combined method can generate a satisfactory outcome (which means the present images are more relevant to the user), the characteristics of the combined method need to be discussed. On one hand, the classifier using the Rough Set can enhance the quality of the training data by removing the most identified misclassification pattern from the majority class. On the other hand, the SVM gains the benefits of avoiding the over-fitting problems of the minority class by interpolating new minority class instances, rather than duplicating the existing instances.

Furthermore, a better algorithm has been applied as a classifier in the image classification part. Also, the Rough Set 1-v-1 SVM and Rough Set 1-v-r SVM methods have the ability to handle the uncertain boundaries better, enabling it to classify those images in the region more accurately. Uncertain boundaries relate to those boundaries that cannot be clearly defined and are not distinct from others. For example, the exact border of a cloud in the grey area. That means the CBIR system cannot understand if this area is of the object or the background. However, when using Rough Set 1-v-1 SVM and Rough Set 1-v-r SVM, the CBIR system can recognise these kinds of boundaries easily.

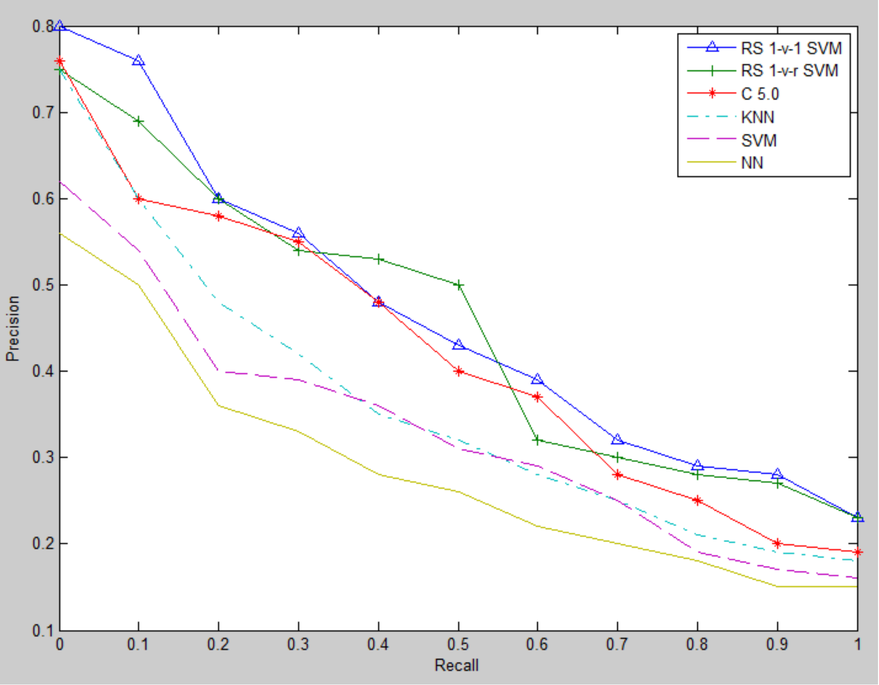


Figure 5.71: Precision-Recall Graph

5.3.2 Experiment II: The Investigation of the Retrieval Accuracy

To investigate the total accuracy of the above-mentioned retrieval systems, 1000 images were fed into the system as the queried images. The average of the retrieval accuracy is calculated for each class. Figure 5.9 shows the results using the different classifiers. As anticipated, the results are better using the proposed system. The averages of the retrieval accuracy are 85.4% for NN, 87.3% for SVM, 89.4% for

KNN and 89.9% for C5.0 respectively. The results also show an increase to 91.4% for Rough Set 1-v-r SVM and 92% for Rough Set 1-v-1 SVM, when these two classifiers are used.

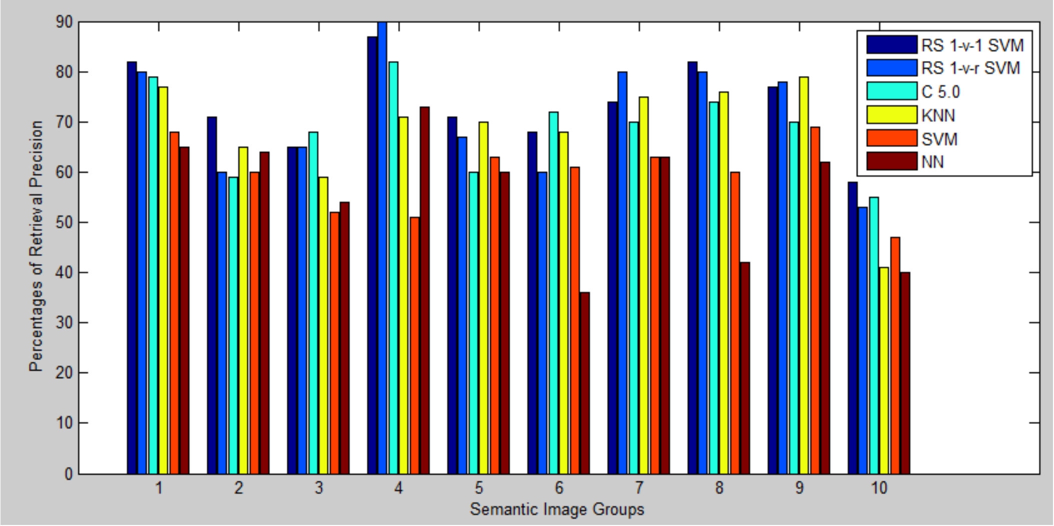


Figure 5.72: Retrieval Accuracy Graph

The reasons behind the results when Rough Set 1-v-1 SVM and Rough Set 1-v-r SVM are used include:

1) The overlapped region in the classification problem can be described more accurately using the boundary region in the Rough Set.
2) The optimal separation of the hyper-plane by maximising the margin is effectively constructed using the SVM.
3) The perfect generation ability is one of the SVM’s properties. However, it cannot deal with imprecise or incomplete data.

4) The most important property of the Rough Set is that it can deal with vague and incomplete data efficiently (Maryam Shahabi lotfabadi, Mohd Fairuz Shiratuddin, & Kok Wai Wong, 2013).

In addition, the Rough Set 1-v-1 SVM and Rough Set 1-v-r SVM classifiers have some advantages compared to the conventional SVM. One of the advantages is that the Rough Set 1-v-1 SVM reduces the storage requirements. Rough Set 1-v-1 SVM is required to store just $2 \times \mathrm{N}$ rules for each class, i.e. one rule for the lower approximation and another for the upper approximation, as compared to the conventional SVM that is required to store $\mathrm{N} \times (\mathrm{N} - 1)/2$ rules (Xiaoyuan Zhang et al., 2012). Another advantage of the Rough Set 1-v-1 SVM is that it has a better sematic interpretation of the classification process, compared to the conventional SVM (Lingras & Butz, 2007).

On the other hand, one of the advantages of the Rough Set 1-v-r SVM is that it can deal with noisy data better than conventional SVM (Lingras & Butz, 2007). Another advantage is that it can reduce the training time and increase the training performance. As described in the second section of this chapter, the positive region (lower approximation) of the largest group is eliminated from further classification, followed by the next largest group and so on. By reducing the size of the training set, this elimination process increases the training performance over the conventional 1-v-r SVM (Lingras & Butz, 2007).

5.4 Summary

Due to the vital and useful advantages that the 1-v-1 Support Vector Machine and 1-v-r Support Vector Machine produce when combined with the Rough Set, we examined these two classifiers for use in a Content Based Image Retrieval system in this chapter. These two image retrieval systems are compared with other image retrieval systems which utilise other classifiers such as Decision Tree (C5.0), Neural Network, K-nearest neighbour and Support Vector Machine. The main focus of this chapter is to examine the Rough Set with 1-v-1 SVM classifier and the Rough Set with 1-v-r SVM classifier in a CBIR system and evaluate the performance of these two classifiers with image features.

The experiment is conducted using the Corel image dataset. Based on the experiment results using the Precision-Recall graph and retrieval accuracy, it can be concluded that the CBIR system with Rough Set and the two 1-v-1 SVM and 1-v-r SVM classifiers produce better results when compared to the conventional SVM. Both the Rough Set with 1-v-1 SVM classifier and the Rough Set with 1-v-r SVM classifier increased the retrieval accuracy and retrieved more relevant images. The priorities of the Support Vector Machine 1-v-1 compared to the conventional Support Vector Machine are better semantic interpretations of the classification process. In addition, the former classifier reduced the storage requirements because it only requires storing a $2 \times N$ rule.

Chapter 6: Conclusion and Future Research

6.1 Introduction

This research concentrates mainly on noisy and vague image features and feature selection problems that are found in Content Based Image Retrieval (CBIR) systems. These problems normally affect the CBIR system's performance. From the literature review, various approaches have been proposed to overcome these problems. However, each approach still has some disadvantages as discussed in Chapter 2.

The motivation of this research originates from the shortcomings of available approaches to deal with noisy and vague image features and feature selection problems. The challenge of this research study is to create techniques on how to handle noise and vague image features, as well as feature selection problems effectively, without losing the important features, so as to obtain better image retrieval results. This research study has investigated and explored applicable techniques to handle noisy and vague image features, as well as feature selection problems using the Rough Set method. As a result, two algorithms have been proposed to handle vague and noisy image features and feature selection problems. One algorithm, which works with vague and noisy image features, is presented in Chapters 3 and 4. The other algorithm, which handles feature selection and classification problems in CBIR systems, is shown in Chapter 5. The two algorithms were implemented and tested on 10 semantic groups of the Corel image dataset. Finally, the research study has deduced that the effective approaches are those developed by using the Fuzzy Rough Set to handle noisy image features and feature selection problems. These approaches are able to provide satisfactory outcomes in terms of retrieval accuracy, Precision-Recall (PR) graph and image comparison.

The concluding chapter is organised into four sections. In Section 6.2, the summary of contributions is presented. Each method which can handle noisy image features and feature selection problems effectively is summarised. Section 6.3 presents the limitations of this research study, while the last section features suggested future research.

6.2 Summary of Contributions

According to the research objective and aims as stated in Chapter 1 - Section 1.4, the major contribution and significance of the research can be summarised by the achievement of the pre-processing stage in using the Fuzzy Rough Set for handling the vague and noisy image features, reducing the semantic gap and improving the image retrieval performance in the CBIR system. In order to achieve these contributions, two methods are presented in this thesis. These methods are Fuzzy Rough Set feature selection for normal and noisy images, as well as a combination of two kinds of Support Vector Machine with the Rough Set as two new classifiers. The following section presents the summary.

6.2.1 Fuzzy Rough Set Feature Selection

The Fuzzy Rough Set is applied as a feature selection to the pre-processing stage to identify and eliminate vague, redundant image features from the image feature vector. These redundant features could influence further analysis in the wrong direction. Consequently, from the remaining significant features, semantic rules that can classify the images more accurately and show more relevance images to the user are then extracted; hence, improving the retrieval performance. Unlike the black box process of Artificial Neural Network in which the output is blindly trusted (although the knowledge is not comprehensible and easily justifiable), the rule discovering process of Fuzzy Rough Set is intuitively comprehensible and can be interpreted and analysed for intelligent decision-making support. The Fuzzy Rough Set generates semantic rules to identify important knowledge hidden in the original data. This rule-based method is suitable for knowledge discovery, especially in complex professional domains, such as an image.

In order to conduct the experiment, 10 semantic groups from the Corel image dataset were selected. They include autumn, castle, cloud, dog, iceberg, primates, ship, tiger, train and waterfall. The Fuzzy Rough Set feature selection shows that it can provide higher retrieval accuracy over the other feature selection methods, such as Gain Ratio, Genetic Algorithm, Information Gain, Isomap, Kernel PCA, OneR, Principal Component Analysis (PCA) and Relief-F. The reason is the semantic rules, which are generated from the Fuzzy Rough feature selection are semantic and help the CBIR system achieve better retrieval performance and accuracy. These semantic and human understanding rules can help those interested in the image features, understand which features are crucial, as well as make effective and accurate decisions to achieve specific objectives. Unlike statistical methods, the Fuzzy Rough Set can automatically extract semantic rules from an image dataset and construct different model representations that explain the image dataset.

Rather than just using the Precision-Recall graph and retrieval accuracy, another experiment has been done for the evaluation of the performance of the pre-processing stage. This experiment is used to explore if the Fuzzy Rough feature selection can first recognise and select important features. In this experiment, the image features that are important in the image retrieval application are defined and ranked. This information is collected from different parts of the literature. In particular, the most influential features are Mean Hue, Coarseness, Standard Deviation, Wavelet Moment and Directionality. The results show that the Fuzzy Rough feature selection method is useful in producing results in line with the defined ranking. The reason is that the Fuzzy Rough feature selection uses the dependency function to select the important features. This function uses the positive region that can deal with the vague area and recognise more important features.

6.2.2 Fuzzy Rough Set Feature Selection for Noisy Images

In order to evaluate the pre-processing stage and the Fuzzy Rough Set performance with noisy image features, three types of noise, namely the Gaussian noise, Poisson noise and Salt and Pepper noise, were added to the queried images.

In the testing phase, the user feeds the noisy queried image, instead of the normal query image, to the system. The system then extracts the noisy queried image features and gives these features to the SVM classifier, which is built into the training phase. This classifier will subsequently extract the relevant images based on the noisy queried image provided.

When all experimental results are compared, it can be concluded that the Fuzzy Rough Set feature selection performs better than eight other common feature selection methods (Gain Ratio, Genetic Algorithm, Information Gain, Isomap, Kernel PCA, OneR, Principal Component Analysis (PCA), and Relief-F). This is because the Fuzzy Rough Set feature selection removes only the probable noisy and misclassification features, rather than eliminate all identified misclassification features from the image features. Thus, this feature selection method can provide higher confidence in noise identification and elimination. When the quality of the image features is improved by removing the noise, CBIR systems tend to increase their image retrieval performances.

6.2.3 Combination of Two Kinds of Support Vector Machine with Rough Set

These methods are presented to handle the multi-class classification and reduce some of the SVM problems for CBIR systems. These methods aim to provide the solution to two questions namely: “Does the Rough Set improve the SVM classifier?” and “How can one increase the retrieval accuracy using the Rough Set and SVM?” The two classifiers are based on a combination of firstly, 1-v-1 (one-versus-one) Support Vector Machine and Rough Set and secondly, 1-v-r (one-versus-all) Support Vector Machine and Rough Set. In the experiment, 10 semantic groups of the Corel image dataset were used, and two new classifiers were compared with the Decision Tree (C5.0), K-nearest Neighbour, Neural Network and Support Vector Machine. It is shown that the Rough Set can enhance the overall performance in terms of retrieval accuracy, reduce storage requirements for 1-v-1 SVM, training time, as well as work better with noisy images for 1-v-r SVM.

6.3 Limitations

Although the research has reached its aims, there were some unavoidable limitations. First, the feature selecting time for the Fuzzy Rough Set is two times more than other feature selections. However, the higher selection time is not a critical issue for this research because the performance of the computer hardware these days can help to reduce this. The other limitation is that an excessive number of decision rules are generated, thus increasing the complexity of CBIR systems. However, having said that, these rules are semantic and aid human understanding. Therefore, they helped the system have better image retrieval results, as well as helped the user understand the system better.

6.4 Suggestions for Future Research

Many possible directions related to this research can be taken up for future investigations. Here are several interesting issues:

In this study, the experiments have focused only on the colour image dataset of nature or animals, rather than the black and white or x-ray image datasets. As such, it is a good idea to continue the research by examining black and white or x-ray image datasets during the new pre-processing stage as well.

For future research, it is valuable to go into the area of why the classical Fuzzy Rough Set method generates many rules. Also, it is a good idea to continue the research in this direction and present improved robust models with lesser rules based on the Fuzzy Rough Set.

This study has explored the use of Rough Set and Fuzzy Rough Set methods to overcome the semantic gap and retrieval accuracy in CBIR systems. It is believed that the experimental studies and results from this research have contributed to the improvement of image retrieval performances in the Corel image dataset. Although many other possible research directions are not included in this section, it is hoped that research studies of Rough Set and Fuzzy Rough Set methods will continue further to solve other complex problems such as edge detection, face recognition and image representing.

List of References

Abdolhossein Sarrafzadeh, Habibollah Agh Atabay, Mir Mosen Pedram, & Shanbehzadeh, J. (2012). ReliefF based feature selection In Content-based image retrieval. *International of multi conference of Engineers and computer Scientists, 1,* 107-110.

Acharya, S., & Devi, M. R. V. (2012). Image retrieval based on visual attention model. *Procedia Engineering, 30,* 542-545.

Alattab, A. A., & Kareem, S. A. (2013). *Semantic Features Selection and Representation for Facial Image Retrieval System.* Paper presented at the 2013 4th International Conference on Intelligent Systems Modelling & Simulation (ISMS).

Balasubramanian, M., & Schwartz, E. L. (2002). The Isomap algorithm and topological stability. *Science, 295*(5552), 7.

Baranidharan, T., & Ghosh, D. K. (2012). Medical Image Classification Using Genetic Optimized Elman Network. *American Journal of Applied Sciences, 9*(1), 123-126.

Bello, R., & Verdegay, J. L. (2012). Rough sets in the Soft Computing environment. *Information Sciences, 212,* 1-14.

Bird, C. L., Elliott, P. J., & Griffiths, E. (1996). User interfaces for content-based image retrieval. *Intelligent Image Databases, IEE Colloquium on*, 8/1 - 8/4

Buijs, J., & Lew, M. (1999). Visual Learning of Simple Semantics in ImageScape. In D. Huijsmans & A. M. Smeulders (Eds.), *Visual Information and Information Systems* (Vol. 1614, pp. 131-138): Springer Berlin Heidelberg.

Carlton W. Niblack, Ron Barber, Will Equitz, Myron D. Flickner, Eduardo H. Glasman, Dragutin Petkovic, . . . Taubin, G. (1993). QBIC project: querying images by content, using color, texture, and shape. *Storage and Retrieval for Image and Video Databases, 1908*(1), 173-187.

Cerra, D., & Datcu, M. (2012). A fast compression-based similarity measure with applications to content-based image retrieval. *Journal of Visual Communication and Image Representation, 23*(2), 293-302.

Chapelle, O., Vapnik, V., Bousquet, O., & Mukherjee, S. (2002). Choosing Multiple Parameters for Support Vector Machines. *Machine Learning, 46*(1-3), 131-159.

Chen, D., Kwong, S., He, Q., & Wang, H. (2012). Geometrical interpretation and applications of membership functions with fuzzy rough sets. *Fuzzy Sets and Systems, 193*, 122-135.

Chen, G., & Wilson, C. (2008). *Use of Self-Organizing Maps for texture feature selection in content-based image retrieval.* Paper presented at the 2008 IEEE International Joint Conference on Neural Networks (IJCNN).

Chen, J., & Li, J. (2012). An application of rough sets to graph theory. *Information Sciences, 201*, 114-127.

Chen, L., Huang, X., Tian, J., & Fu, X. (2014). Blind noisy image quality evaluation using a deformable ant colony algorithm. *Optics & Laser Technology, 57,* 265-270.

Chen, Y.-S. (2012). Classifying credit ratings for Asian banks using integrating feature selection and the CPDA-based rough sets approach. *Knowledge-Based Systems, 26*, 259-270.

Chen, Y.-S., & Cheng, C.-H. (2012). A soft-computing based rough sets classifier for classifying IPO returns in the financial markets. *Applied Soft Computing, 12*(1), 462-475.

Chen, Z., Hou, J., Zhang, D., & Qin, X. (2012). An Annotation Rule Extraction Algorithm for Image Retrieval. *Pattern Recognition Letters, 33*(10), 1257–1268.

Chinpanthana, N. (2011). Integrating Qualitative Features Selection for Semantic Image Classification with Support Vector Machine. *3rd International Conference on Information and Financial Engineering*, 117-122.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297. doi: 10.1007/BF00994018

Cui, M. (2012). Rough set model under a limited asymmetric similarity relation and an approach for incremental updating approximations. *Physics Procedia, 24, Part A,* 603-610.

da Silva, S. F., Ribeiro, M. X., Batista Neto, J. d. E. S., Traina-Jr, C., & Traina, A. J. M. (2011). Improving the ranking quality of medical image retrieval using a genetic feature selection method. *Decision Support Systems, 51*(4), 810-820.

da Silva, S. F., Traina, A. J. M., Ribeiro, M. X., do E.S.Batista Neto, J., & Traina, A. J. M. (2009, 2-5 Aug. 2009). *Ranking evaluation functions to improve genetic feature selection in content-based image retrieval of mammograms.* Paper presented at the 22nd IEEE International Symposium on Computer-Based Medical Systems, CBMS 2009. .

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys, 40*(2), 1-60.

Derrac, J., Cornelis, C., García, S., & Herrera, F. (2012). Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection. *Information Sciences, 186*(1), 73-92.

Dharani, T., & Aroquiaraj, I. L. (2013). *A survey on content based image retrieval.* Paper presented at the 2013 International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME).

Diker, M., & Altay UÄŸur, A. l. (2012). Textures and covering based rough sets. *Information Sciences, 184*(1), 44-63.

Dy, J. G., Brodley, C. E., Kak, A., Broderick, L. S., & Aisen, A. M. (2003). Unsupervised feature selection applied to content-based retrieval of lung images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(3), 373-378.

ElAlami, M. E. (2011). A novel image retrieval model based on the most relevant features. *Knowledge-Based Systems, 24*(1), 23-32. doi: http://dx.doi.org/10.1016/j.knosys.2010.06.001

Estaji, A. A., Hooshmandasl, M. R., & Davvaz, B. (2012). Rough set theory applied to lattice theory. *Information Sciences, 200*, 108-122.

Fanjie, M., Baolong, G., & Xianxiang, W. (2012). Localized Image Retrieval Based on Interest Points. *Procedia Engineering, 29*, 3371-3375.

Fei, L., Qionghai, D., & Wenli, X. (2006). *Improved Similarity-Based Online Feature Selection in Region-Based Image Retrieval.* Paper presented at the 2006 IEEE International Conference on Multimedia and Expo.

Feifei Xu, Duoqian Miao, & Wei, L. (2009). *Fuzzy-Rough Attribute Reduction Via Mutual Information With An Application To Cancer Classification*. *Computers And Mathematics With Applications, 57*, 1010_1017.

Feng, Y., Xiao, J., Zha, Z., Zhang, H., & Yang, Y. (2012). Active learning for social image retrieval using Locally Regressive Optimal Design. *Neurocomputing, 95*, 54-59.

Foithong, S., Pinngern, O., & Attachoo, B. (2012). Feature subset selection wrapper based on mutual information and rough sets. *Expert Systems with Applications, 39*(1), 574-584.

Frosini, P., & Landi, C. (2013a). Corrigendum to "Persistent Betti numbers for a noise tolerant shape-based approach to image retrieval". *Pattern Recognition Letters, 34*(11), 1320-1321.

Frosini, P., & Landi, C. (2013b). Persistent Betti numbers for a noise tolerant shape-based approach to image retrieval. *Pattern Recognition Letters, 34*(8), 863-872.

Ganivada, A., Ray, S. S., & Pal, S. K. (2013). Fuzzy rough sets, and a granular neural network for unsupervised feature selection. *Neural Networks, 48*, 91-108.

Gavves, E., Snoek, C. G. M., & Smeulders, A. W. M. (2012). Visual synonyms for landmark image retrieval. *Computer Vision and Image Understanding, 116*(2), 238-249.

Grana, M., & Veganzones, M. A. (2012). An endmember-based distance for content based hyperspectral image retrieval. *Pattern Recognition, 45*(9), 3472–3489.

Guldogan, E., & Gabbouj, M. (2008). Feature selection for content-based image retrieval. *Signal, Image and Video Processing, 2*(3), 241-250.

Haiyu Song, Xiongfei Li, & Wang, P. (2010). Adaptive Feature Selection and Extraction Approaches for Image Retrieval based on Region *Journal of Multimedia, 5*(1), 85-92.

Hammami, M., Ben Jemaa, S., & Ben-Abdallah, H. (2012). Selection of discriminative sub-regions for palmprint recognition. *Multimedia tools & Applications*, 1-28.

Han-ping, G., & Zu-qiao, Y. (2011). *Integrated Visual Saliency Based Local Feature Selection for Image Retrieval.* Paper presented at the 2011 2nd International

Symposium on Intelligence Information Processing and Trusted Computing (IPTC)

Hoffmann., H. (2007). Kernel PCA for novelty detection. *Pattern Recognition, 40*(3), 863-874.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*: University of Michigan Press.

Hopfgartner, F., Urruty, T., Lopez, P., Villa, R., & Jose, J. (2010). Simulated evaluation of faceted browsing based on feature selection. *Multimedia Tools & Applications, 47*(3), 631-662.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(6), 417-441.

Hu, Q., Zhang, L., An, S., Zhang, D., & Yu, D. (2012). On Robust Fuzzy Rough Set Models. *IEEE Transactions on Fuzzy Systems, 20*(4), 636-651.

Huang, K. Y. (2012). An enhanced classification method comprising a genetic algorithm, rough set theory and a modified PBMF-index function. *Applied Soft Computing, 12*(1), 46-63.

Iqbal, K., Odetayo, M. O., & James, A. (2012). Content-based image retrieval approach for biometric security using colour, texture and shape features controlled by fuzzy heuristics. *Journal of Computer and System Sciences, 78*(4), 1258-1277.

Jaime, L. G., Kerre, E. E., Nachtegael, M., & Bustince, H. (2013). Consensus image method for unknown noise removal. *Knowledge-Based Systems*. doi: http://dx.doi.org/10.1016/j.knosys.2013.10.023

Jensen, R., & Shen, Q. (2002). Fuzzy-rough sets for descriptive *dimensionality* reduction. *Proceedings of the 2001 IEEE International Conference Fuzzy Systems*, 29-34.

Ji, R., Yao, H., & Liang, D. (2008). DRM: dynamic region matching for image retrieval using probabilistic fuzzy matching and boosting feature selection. *Signal, Image and Video Processing, 2*(1), 59-71.

Jyothi, B. V., & Eswaran, K. (2010, 27-29 Jan. 2010). *Comparative Study of Neural Networks for Image Retrieval.* Paper presented at the 2010 International Conference on Intelligent Systems, Modelling and Simulation (ISMS).

Kien-Ping, C., Chun-Che, F., & Kok-Wai, W. (2005). *A Feature Selection Framework for Small Sampling Data in Content-based Image Retrieval System.* Paper presented at the 2005 Fifth International Conference on Information, Communications and Signal Processing.

Kiktova-Vozarikova, E., Juhar, J., & Cizmar, A. (2013). Feature selection for acoustic events detection. *Multimedia Tools & Applications*, 1-21.

Kononenko, I., Šimec, E., & Robnik-Šikonja, M. (1997). Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF. *Applied Intelligence, 7*(1), 39-55.

Krishnamoorthi, R., & Sathiya devi, S. (2012). A multiresolution approach for rotation invariant texture image retrieval with orthogonal polynomials model. *Journal of Visual Communication and Image Representation, 23*(1), 18-30.

Li, D.-x., Fan, J.-l., Wang, D.-w., & Liu, Y. (2012). Latent topic based multi-instance learning method for localized content-based image retrieval. *Computers &amp; Mathematics with Applications, 64*(4), 500-510.

Li, S., Wu, H., Wan, D., & Zhu, J. (2011). An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. *Knowledge-Based Systems, 24*(1), 40-48.

Li, X., & Liu, S. (2012). Matroidal approaches to rough sets via closure operators. *International Journal of Approximate Reasoning, 53*(4), 513-527.

Li, Y.-L., Tang, J.-F., Chin, K.-S., Luo, X.-G., & Han, Y. (2012). Rough set-based approach for modeling relationship measures in product planning. *Information Sciences, 193*, 199-217.

Li, Y., Geng, B., Yang, L., Xu, C., & Bian, W. (2012). Query difficulty estimation for image retrieval. *Neurocomputing, 95*, 48-53.

Li, Y., Sun, J., & Luo, H. (2014). A neuro-fuzzy network based impulse noise filtering for gray scale images. *Neurocomputing, 127*, 190-199.

Li, Z., Xie, T., & Li, Q. (2012). Topological structure of generalized rough sets. *Computers & Mathematics with Applications, 63*(6), 1066-1071.

Liang, J., Li, R., & Qian, Y. (2012). Distance: A more comprehensible perspective for measures in rough set theory. *Knowledge-Based Systems, 27*, 126-136.

Liang, Z., Zhuang, Y., Yang, Y., & Xiao, J. (2013). Retrieval-based cartoon gesture recognition and applications via semi-supervised heterogeneous classifiers learning. *Pattern Recognition, 46*(1), 412-423. doi: http://dx.doi.org/10.1016/j.patcog.2012.06.025

Lingras, P., & Butz, C. (2005). Reducing the Storage Requirements of 1-v-1 Support Vector Machine Multi-classifiers. In D. Ślęzak, J. Yao, J. Peters, W. Ziarko & X. Hu (Eds.), *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing* (Vol. 3642, pp. 166-173): Springer Berlin Heidelberg.

Lingras, P., & Butz, C. (2007). Rough set based 1-v-1 and 1-v-r approaches to support vector machine multi-classification. *Information Sciences, 177*(18), 3782-3798. doi: http://dx.doi.org/10.1016/j.ins.2007.03.028

Liu, L., Zeng, L., Shen, K., & Luan, X. (2013). Exploiting local intensity information in Chan–Vese model for noisy image segmentation. *Signal Processing, 93*(9), 2709-2721. doi: http://dx.doi.org/10.1016/j.sigpro.2013.03.035

Lu, Z. M., Burkhardt, H., & Boehmer, S. (2006). Fast Content-Based Image Retrieval Based on Equal-Average K-Nearest-Neighbor Search Schemes. In Y. Zhuang, S.-Q. Yang, Y. Rui & Q. He (Eds.), *Advances in Multimedia Information Processing - PCM 2006* (Vol. 4261, pp. 167-174): Springer Berlin Heidelberg.

Luo, R., Zhang, Y., Fan, Y., & Deng, X. (2001). *Research on content-based remote sensing image retrieval: the strategy for visual feature selection, extraction, description and similarity measurement.* Paper presented at the 2001 International Conferences on Info-tech and Info-net (ICII).

Manikandan, S., & Rajamani, V. (2008). A Mathematical Approach for Feature Selection & Image Retrieval of Ultra Sound Kidney Image Databases *European Journal of Scientific Research, 24*(2), 163-171.

Marakakis, A., Galatsanos, N., Likas, A., & Stafylopatis, A. (2009). Relevance Feedback for Content-Based Image Retrieval Using Support Vector Machines and Feature Selection. In C. Alippi, M. Polycarpou, C. Panayiotou & G. Ellinas (Eds.), *Artificial Neural Networks – ICANN 2009* (Vol. 5768, pp. 942-951): Springer Berlin Heidelberg.

Mengling, L., Chu, H., Chao, Q., & Hong, S. (2008, 26-29 Oct. 2008). *A hierarchical boosting algorithm based on feature selection for Synthetic Aperture Radar image retrieval.* Paper presented at the 9th International Conference on Signal Processing (ICSP).

Michael Lew Nies, & Lew, M. S. (1996). *Content Based Image Retrieval: KLT, Projections, or Templates*: Amsterdam University Press.

Mukhopadhyay, S., Dash, J. K., & Das Gupta, R. (2013). Content-based texture image retrieval using fuzzy class membership. *Pattern Recognition Letters, 34*(6), 646-654. doi: http://dx.doi.org/10.1016/j.patrec.2013.01.001

Najjar, M., Ambroise, C., & Cocquerez, J. P. (2003). *Feature selection for semisupervised learning applied to image retrieval.* Paper presented at the 2003 International Conference on Image Processing (ICIP).

Nikhil Naikal, Allen Y. Yang, & Sastry, S. (2011). Informative Feature Selection for Object Recognition via Sparse PCA*. *In Proceedings of ICCV 2011*, 818-825.

Othman, M. L., Aris, I., Othman, M. R., & Osman, H. (2012). Rough-Set-based timing characteristic analyses of distance protective relay. *Applied Soft Computing, 12*(8), 2053–2062.

Pawlak, Z. (1982). Rough sets. *International Journal of Parallel Programming, 11*(5), 341-356.

Penatti, O. A. B., Valle, E., & Torres, R. d. S. (2012). Comparative study of global color and texture descriptors for web image retrieval. *Journal of Visual Communication and Image Representation, 23*(2), 359-380.

Prasanna, R., Ramakrishnan, K. R., & Bhattacharyya, C. (2003). *Simultaneous feature selection and classification for relevance feedback in image retrieval.* Paper presented at the Conference on Convergent Technologies for the Asia-Pacific Region (TENCON 2003).

Rashedi, E., Nezamabadi-Pour, H., & Saryazdi, S. (2013). A simultaneous feature adaptation and feature selection method for content-based image retrieval systems. *Knowledge-based systems, 39*, 85-94.

Remco C. Veltkamp, & Tanase, M. (2002). Content-Based Image Retrieval Systems: A Survey *Technical Report* (pp. 1-62). UU-CS.

Sclaroff, S., Taycher, L., & La Cascia, M. (1997). *ImageRover: a content-based image browser for the World Wide Web.* Paper presented at the 1997 IEEE Workshop on Content-Based Access of Image and Video Libraries.

Setayesh, M., Zhang, M., & Johnston, M. (2013). A novel particle swarm optimisation approach to detecting continuous, thin and smooth edges in noisy images. *Information Sciences, 246*, 28-51.

Shahabi Lotfabadi, M., & Eftekhari Moghadam, A. M. (2010). *The Comparison of Different Feature Decreasing Methods Base on Rough Sets and Principal Component Analysis for Extraction of Valuable Features and Data Classifying Accuracy Increasing.* Paper presented at the First International Conference on Integrated Intelligent Computing (ICIIC 2010), Bangalore, India

Shahabi Lotfabadi, M., & Mahmoudie, R. (2010). *The Comparison of Different Classifiers for Precision Improvement in Image Retrieval.* Paper presented at the Sixth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS 2010).

Shahabi Lotfabadi, M., Shiratuddin, M. F., & Wong, K. W. (2012a). *Using Rough Set Theory to Improve Content Based Image Retrieval System*. Paper presented at the IEEE. Eleventh Postgraduate Electrical Engineering and Computing Symposium (PEECS2012).

Shahabi Lotfabadi, M., Shiratuddin, M. F., & Wong, K. W. (2012b). *Utilising Fuzzy Rough Set based on Mutual Information Decreasing Method for Feature Reduction in an Image Retrieval System.* Paper presented at the Proceedings of the International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE 2012).

Shahabi Lotfabadi, M., Shiratuddin, M. F., & Wong, K. W. (2013). *Feature Decreasing Methods Using Fuzzy Rough Set based on Mutual Information*. Paper presented at the Proceeding of the 8th IEEE Conference on Industrial Electronics and Applications (ICIEA 2013).

Shahabi lotfabadi, M., Shiratuddin, M. F., & Wong, K. W. (2013). *Using Fuzzy Rough Feature Selection for Image Retrieval System.* Paper presented at the Proceeding of the 2013 IEEE Symposium Series On Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP 2013), Singapore.

Shahabi Lotfabadi, M., Shiratuddin, M. F., & Wong, K. W. (2013, 16-19 April 2013). *Using fuzzy rough feature selection for image retrieval system.* Paper presented at the IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP 2013).

Shahabi lotfabadi, M., Shiratuddin, M. F., & Wong, K. W. (2014). *Evaluation of Fuzzy Rough Set Feature Selection for Content Based Image Retrieval System with Noisy Images.* Paper presented at the 22rd International Conference in Centeral Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2014), Plzen-Czech Republic.

Shi, F., Sun, S., & Xu, J. (2012). Employing rough sets and association rule mining in KANSEI knowledge extraction. *Information Sciences, 196*, 118-128.

Son, C.-H., Choo, H., & Park, H.-M. (2013). Image-pair-based deblurring with spatially varying norms and noisy image updating. *Journal of Visual Communication and Image Representation, 24*(8), 1303-1315. doi: http://dx.doi.org/10.1016/j.jvcir.2013.09.001

Son, C.-S., Kim, Y.-N., Kim, H.-S., Park, H.-S., & Kim, M.-S. (2012). Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. *Journal of Biomedical Informatics, 45*(5), 999-1008.

Subrahmanyam, M., Maheshwari, R. P., & Balasubramanian, R. (2012a). Expert system design using wavelet and color vocabulary trees for image retrieval. *Expert Systems with Applications, 39*(5), 5104-5114.

Subrahmanyam, M., Maheshwari, R. P., & Balasubramanian, R. (2012b). Local maximum edge binary patterns: A new descriptor for image retrieval and object tracking. *Signal Processing, 92*(6), 1467-1479.

Swets, D. L., & Weng, J. J. (1995, 21-23 Nov 1995). *Efficient content-based image retrieval using automatic feature selection.* Paper presented at the International Symposium on Computer Vision.

Tianzhong, Z., Jianjiang, L., Yafei, Z., & Qi, X. (2008). *Feature Selection Based on Genetic Algorithm for CBIR.* Paper presented at the 2008 Congress on Image and Signal Processing (CISP).

Tsai, C.-F., Eberle, W., & Chu, C.-Y. (2013). Genetic algorithms in feature and instance selection. *Knowledge-Based Systems, 39*, 240-247. doi: http://dx.doi.org/10.1016/j.knosys.2012.11.005

Tsun-Wei, C., Yo-Ping, H., & Sandnes, F. E. (2009). *Efficient entropy-based features selection for image retrieval.* Paper presented at the 2009 IEEE International Conference on Systems, Man and Cybernetics (SMC 2009).

Turcot, P., & Lowe, D. G. (2009). *Better matching with fewer features: The selection of useful features in large database recognition problems.* Paper presented at the 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops).

Valliammal, N., & Geethalakshmi, S. N. (Eds.). (2012). *Efficient feature fusion, selection and classification technique for Plant Leaf Image Retrieval system*: ACM.

Vasconcelos, N. (2003). *A family of information-theoretic algorithms for low-complexity discriminant feature selection in image retrieval.* Paper presented at the 2003 International Conference on Image Processing (ICIP).

Vasconcelos, N., & Vasconcelos, M. (2004, 27 June-2 July 2004). *Scalable discriminant feature selection for image retrieval and recognition.* Paper presented at the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).

Vavilin, A., & Jo, K.-H. (2013). Automatic context analysis for image classification and retrieval based on optimal feature subset selection. *Neurocomputing, 116*, 201-207.

Vendrig, J., Worring, M., & Smeulders, A. W. M. (1999). *Filter Image Browsing - Exploiting Interaction in Image Retrieval.* Paper presented at the Proceedings of the Third International Conference on Visual Information and Information Systems.

Walek, P., Jan, J., Ourednicek, P., Skotakova, J., & Jira, I. (2012). Preprocessing for Quantitative Statistical Noise Analysis of MDCT Brain Images Reconstructed Using Hybrid Iterative (iDose) Algorithm. *Journal of WSCG, 20*(1), 73-80.

Wang, D., Yuchun, F., & Binbin, H. (2011, 15-17 Oct. 2011). *Feature selection in interactive face retrieval.* Paper presented at the 4th International Congress on Image and Signal Processing (CISP).

Wang, J., Hedar, A.-R., Wang, S., & Ma, J. (2012). Rough set and scatter search metaheuristic based feature selection for credit scoring. *Expert Systems with Applications, 39*(6), 6123-6128.

Wang, L., & Khan, L. (2006). Automatic image annotation and retrieval using weighted feature selection. *Multimedia Tools & Applications, 29*(1), 55-71.

Wang, X.-y., Chen, Z.-f., & Yun, J.-j. (2012). An effective method for color image retrieval based on texture. *Computer Standards and Interfaces, 34*(1), 31-35.

Wang, X.-Y., Zhang, B.-B., & Yang, H.-Y. (2013). Active SVM-based relevance feedback using multiple classifiers ensemble and features reweighting. *Engineering Applications of Artificial Intelligence, 26*(1), 368-381. doi: http://dx.doi.org/10.1016/j.engappai.2012.05.008

Wei, J., Guihua, E., Qionghai, D., Lian, Z., & Yao, H. (2005, March 18-23, 2005). *Relevance Feedback Learning With Feature Selection In Region-Based Image Retrieval.* Paper presented at the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

Wei Jiang, Guihua Er, Qionghai Dai, & Jinwei Gu. (2006). Similarity-Based Online Feature Selection in Content-Based Image Retrieval. *IEEE Transactions on Image Processing, 15*(3), 702-712.

Wei, W., Liang, J., & Qian, Y. (2012). A comparative study of rough sets for hybrid data. *Information Sciences, 190*, 1-16.

X.J. Shen, & Wang, Z. F. (2006). Feature selection for image retrieval. *Electronics Letters 42*(6).

Xiang-wei, L., & Yian-fang, Q. (2012). A Data Preprocessing Algorithm for Classification Model Based On Rough Sets. *Physics Procedia, 25*, 2025-2029.

Xiaohong Yu, & Liu, H. (2009). Image Semantic Classification Using SVM In Image Retrieval *Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT '09)*, 458-461.

Xin, S., Xin, L., & Hong, S. (2008, 26-29 Oct. 2008). *Feature selection and re-weighting in content-based SAR image retrieval.* Paper presented at the 9th International Conference on Signal Processing (ICSP).

XiongWei, Qiuyan, S., & Jinlong, L. (2012). The group decision-making rules based on rough sets on large scale engineering emergency. *Systems Engineering Procedia, 4*, 331-337.

Xu, F., & Zhang, Y.-J. (2007). Integrated patch model: A generative model for image categorization based on feature selection. *Pattern Recognition Letters, 28*(12), 1581-1591.

Xun, Y., Xian-Sheng, H., Meng, W., Qi, G. J., & Xiu-Qing, W. (2007). *A Novel Multiple Instance Learning Approach for Image Retrieval Based on Adaboost Feature Selection.* Paper presented at the 2007 IEEE International Conference on Multimedia and Expo.

Yang, H.-L., Li, S.-G., Wang, S., & Wang, J. (2012). Bipolar fuzzy rough set model on two different universes and its application. *Knowledge-Based Systems, 35*, 94-101.

Yi, C., Yihua, L., & Haozheng, R. (2012). *A Feature Selection Method Base on GA for CBIR Mammography CAD.* Paper presented at the 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)

Yildizer, E., Balci, A. M., Hassan, M., & Alhajj, R. (2012). Efficient content-based image retrieval using Multiple Support Vector Machines Ensemble. *Expert Systems with Applications, 39*(3), 2385-2396.

Yildizer, E., Balci, A. M., Jarada, T. N., & Alhajj, R. (2012). Integrating wavelets with clustering and indexing for effective content-based image retrieval. *Knowledge-Based Systems, 31*(0), 55-66.

Yossi Rubner, Jan Puzicha, Carlo Tomasi, & Buhmann, J. M. (2001). Empirical Evaluation of Dissimilarity Measures for Color and Texture. *Computer Vision and Image Understanding, 84*, 25-43.

Yu, S., & Bhanu, B. (2010, 26-29 Sept. 2010). *Image retrieval with feature selection and relevance feedback.* Paper presented at the 2010 17th IEEE International Conference on Image Processing (ICIP).

Zhang, J., Li, T., Ruan, D., Gao, Z., & Zhao, C. (2012). A parallel method for computing rough set approximations. *Information Sciences, 194*, 209-223.

Zhang, X., Mo, Z., Xiong, F., & Cheng, W. (2012). Comparative study of variable precision rough set model and graded rough set model. *International Journal of Approximate Reasoning, 53*(1), 104-116.

Zhang, X., Zhou, J., Guo, J., Zou, Q., & Huang, Z. (2012). Vibrant fault diagnosis for hydroelectric generator units with a new combination of rough sets and support vector machine. *Expert Systems with Applications, 39*(3), 2621-2628.

Zhu, Y., Tan, Y., Hua, Y., Wang, M., Zhang, G., & Zhang, J. (2010). Feature Selection and Performance Evaluation of Support Vector Machine (SVM)-Based Classifier for Differentiating Benign and Malignant Pulmonary Nodules by Computed Tomography. *Journal of Digital Imaging, 23*(1), 51-65.