University of Alberta

Fuzzy Clustering in Analysis of Multidimensional Data: A Study in Organization and Classification of Images

by

Tomasz Kaczmarzyk

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

Department of Electrical and Computer Engineering

Edmonton, Alberta
Spring 2007

Your file Votre référence
ISBN: 978-0-494-29971-5
Our file Notre référence
ISBN: 978-0-494-29971-5

Abstract

The objective of this study is to develop and examine the performance of an image classification system using Fuzzy C-Means (FCM) on a large set of images represented by MPEG-7 low-level descriptors. This experimental data set consists of five different categories of images. In a series of experiments we considered 5 different categories of the MPEG-7 descriptors related to colors and textures of images. Prior to any clustering the original space was reduced using the standard Principal Component Analysis (PCA). Additionally, in our experiments we are concerned with classifier fusion algorithms. A series of carefully organized experiments has led to a number of interesting findings as to the suitability of fuzzy sets in the framework in image organization and description, insights into the structure of various categories and their interrelationship.

Table of contents

List of Tables

List of Figures

1. Introduction

The rapid increase of multimedia data has created a need for new suited software systems to enable users to control and to retrieve efficiently this new sort of information. Today digital cameras are very widely used resulting in large personal and/or business-related digital photo collections. However, the current means of automatically sorting these images is very crude.

In recent years, the use of digital collections (say, art galleries, photograph archives, personalized digital albums, etc.) has become very common on the world wide web, as well as, in the preparation of both electronic and paper publications. Considering the existence of various archives there is a growing demand for advanced query mechanisms that are capable of addressing perceptual aspects of visual information. In order to address the needs of users, a number of content-based image retrieval (CBIR) techniques have been developed [2,4,6,7].

The main research objective in the management of multimedia data, is to allow users to manipulate (representation and retrieval) multimedia information as easily as traditional data (numbers, strings) and as intelligently as textual information. The biggest challenge in the imitation of human-like classification processes is the finding of the appropriate ways of abstracting complex objects and looking for a subset of features, which characterizes an object in the best way. For example, image classification can be realized manually by associating a collection of descriptors (words) to every single image in the database. Unfortunately, this is not feasible in practice. First, databases usually include thousands of images. Secondly, the content of an image cannot be fully annotated by a list of words [6]. Bearing this in mind, it is quite evident that a direct extraction of visual information from images is required. This gives rise to a series of low-level features (descriptors), which intend to capture the mapping between images (or their descriptors, to be more precise) and their categories. In general, we can say that the subsets of features have to be carefully selected and quantitatively measured to give us an opportunity to compare features of different objects.

Researcher's efforts lead to significant achievements and progress in this subject area. Their work showed many techniques that could be efficiently used in content-based image retrieval. More research will however be needed to effectively organize and classify large digital image sets.

1.1 Fuzzy sets in Image Classification

Fuzzy logic has been used in a wide range of problem domains [1,5]. Even though fuzzy logic is a relatively young theory, the range of applications is very wide. These include applications in the field of management and decision making, process control, operations research, economies and pattern recognition and classification. The idea behind the concept of fuzzy logic is that a binary answer (for example if something is true or false), in some cases, is not satisfactory. Prof. Zadeh introduced a new way of solving this problem by using a degree of membership. Further, a fuzzy set is a set whose elements have degrees of membership. More precisely, an element of a fuzzy set can be a full member (100% membership) or a partial member (between 0% and 100% membership). In other words, the membership value assigned to an element is no longer limited to only two values, but can be any value between zero and one. Mathematical function which defines the degree of an element's membership in a fuzzy set is called a membership function. In addition, a description of a problem in linguistic terms, rather than in terms of relationships between numerical values, is the key benefit of this theory.

This work will examine fuzzy clustering as an image classification and organization tool. Clustering is a method of unsupervised learning. The ultimate goal of clustering is to partition a given dataset into a number of clusters. It is worth stressing that no training data and no priori knowledge are used to influence the clustering process. In this algorithm, all patterns from a dataset are assigned to k groups of similar patterns (homogeneity within a cluster), where k is smaller than the number of samples in a given dataset. Objects which are not similar are located in different clusters (heterogeneity between clusters) [3].

Deeper analysis clearly shows that usually some clusters will overlap or that the smaller groups could be nested inside larger ones. Fuzzy sets are very useful in such scenarios. There are several reasons why fuzzy sets are so useful in classification and organization of multimedia information. At first, the fuzzy sets theory establishes the interface between higher level concepts represented as features and the computer computations driven by the quantitative measurements. Going further, the concept of fuzzy sets is very well suited for the real world data. Real data often do not have crisp boundaries and have overlapping clusters but in the case of fuzzy sets, a pattern with a certain degree of membership may belong to more than one cluster. Finally, the membership functions of the fuzzy sets model unsure patterns and are able to identify not well-defined patterns in the dataset. All these advantages of fuzzy sets are beneficial for our work. The Fuzzy C-Means algorithm will be of primary interest when clustering.

It is not easy to define a uniform "similarity" measurement in image classification systems. Similarity is very subjective and has different meanings in different application areas. When a user thinks that two images are quite similar, it is most likely that two images a re close in terms of their semantic meanings. The images might be similar in individual features with different weights or some combination of their features. This is an important conclusion which shows that fuzzy logic may be a powerful tool in an image classification and organization systems.

The purpose of this study is motivated by the challenges that exist in image classification and organization, as presented in the following chapters.

1.2 References

1. Bezdek J.C., "Pattern recognition with fuzzy objective function algorithms", Plenum Press, New York, 1981.
2. Bezdek J.C., Keller J., Krishnapuram R., Pal N.R., "Fuzzy Models and Algorithms for Pattern Recognition and Image Processing" , Kluwer Academic Publishers, Norwell, 1999
3. Cios K.J, Pedrycz W., Swiniarski R.W., "Data Mining: Methods for Knowledge Discovery", Kluwer Academic Publishers, Boston, 1998.
4. Hoppner F., Klawonn F., Kruse R., Runkler T., "Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition" , John Wiley and Sons, Ltd. New York 1999.
5. Pedrycz W., "Fuzzy sets in pattern recognition: methodology and methods", Pattern Recognition, Vol. 23, No. ½, 1990, pp. 121-146.
6. Smeulders A.W.M., Worring M., Santini S., Gupta A., Jani R., "Content-nased image retrieval at the end of the early years" ,IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 1349-1380
7. Yixin C., Wang J.Z., Krovetz R., "CLUE: Cluster-based Retrieval of Images by Unsupervised Learning", IEEE Transactions on Image Processing, vol. 14, no. 8, pp. 1187-1201, 2005.

2. Main objectives

In this chapter, we present motivations of our project and the main objectives with their justifications. We outline general ideas behind image organization and classification.

2.1 Motivation

The enormous growth of image archives requires the application of some new ideas in this field. A very essential problem in image management is the image retrieval. When images in a database are not well organized, their efficient retrieval is not a trivial task.

This project's objectives are as follows:

(i) Study in detail the MPEG-7 standard, algorithms responsible for a reduction of a feature space, as well as, image classification and organization systems.

(ii) Provide and implement a framework for image classification system based on fuzzy clustering algorithm (Fuzzy c-means) and study its performance on MPEG-7 feature space in terms of organization and classification of a large dataset of digital images.

(iii) Focus on the efficient manner of combining multiple classifiers and quantifying obtained results.

(iv) Explore the problem of overlapping image categories and investigate the correspondence between clusters and categories.

(v) Provide a practical guide for other image classification systems.

Most of the systems use combinations of image-features (low level features) such as texture or color to organize the images as a database of images and then to retrieve them from it. These image-features do not have any significance or are too complex to interpret and use by the user. It is worth stressing that the quality of a content-based image retrieval system strongly depends on the choice of the set of low-level visual features. For example, the indoor/outdoor classification [9] can be well performed using global color histograms and local color descriptors. On the other hand, edge histograms appear to be useful in the case of city/landscape classification [10], as city images usually contain horizontal and vertical edges. In our work we consider a large set of images being represented by MPEG-7 low-level descriptors [2,3,5] on which all classification activities will be performed. This experimental data set consists of five different categories of images. Our point of interest is to investigate the usefulness of particular MPEG-7 descriptors in terms of image classification. In a series of experiments we considered 5 different kinds of the MPEG-7 descriptors related to colors and textures of images. We compare obtained results and find out which descriptors provide best classification results.

Another important aspect of our work is related to the dimensionality of data. In general, the dimensionality of the problem (viz. the original feature space) is excessively high and this calls for the use of some reduction techniques. There are some important advantages behind space reduction. At first, clustering algorithms used in the reduced spaces are more effective. In this study we confine ourselves to the Principal Component Analysis (PCA) [5] regarded as a vehicle for dimensionality reduction. We compare results for different number of dimensions. Additionally, to compare the quality of classification results obtained for the reduced feature space, we complete classification task for the original feature space.

The advantages of fuzzy sets are beneficial to this project. Specifically, the work was confined to clustering methods based on fuzzy sets, which are able to deal very well with unclear patterns. The fuzzy clustering algorithms show obvious advantages for classification purposes. The objective of our work is to develop and examine the performance of an image classification system using Fuzzy C-Means (FCM) [4,9]. A series of carefully organized experiments have led to a number of interesting findings as to the suitability of fuzzy sets in the framework of image organization and description, insights into the structure of various categories, and their interrelationship. It is also the intention of this project to compare results of a large variety of different values of FCM parameters such as fuzzification parameter *m* or number of clusters. The role of these parameters will be examined and the observations quantified in this project.

It is worth stressing that some images can be assigned to more than a single category. Practically, in many cases image classification may strongly reflect individual preferences of the user. We may clearly see differences between the users in the assignment of the images to categories. In order to show that there is a substantial overlap between the categories and that some images can be assigned differently, we are going to thoroughly analyze the obtained results. More precisely, the intention of this project is to investigate how strong the relations between categories of images are. We investigated a correspondence between clusters and categories. Images could belong to several categories and we are going to capture this effect in the proposed system.

It is worth stressing that the fusion of several image descriptors is a crucial point for different retrieval systems [3,5,11]. A drawback of this fusion is that there is a risk of neglecting the good performances of a given descriptor because of the poor performances of another. In our work we are going to investigate if we can expect any improvements in terms of classification accuracy compared to classification when using only single descriptors. We perform a set of experiments using two different classifier fusion methods to assess their usefulness.

2.2 Conclusions

In this chapter we presented the main objectives of our project. We outlined the major problems and showed the novelty of our image organization and classification system. In the next chapters, we will provide more details about the framework of our system, the used algorithms, and the results obtained.

2.3 References

1. Bezdek J.C., Keller J., Krishnapuram R., Pal N.R., "Fuzzy Models and Algorithms for Pattern Recognition and Image Processing", Kluwer Academic Publishers, Norwell, 1999
2. Chang S.F., Sikora T., Puri A., "Overview of the mpeg-7 standard", IEEE trans. on Circuits and Systems for Video Technology 11 (2001) 688-695
3. Ho T.K, Hull J.J., Srihari S.N., "Decision Combination in Multiple Classifier Systems", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, pp. 66-75, Jan 1994.
4. Hoppner F., Klawonn F., Kruse R., Runkler T., "Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition", John Wiley and Sons, Ltd. New York 1999.
5. Jollife I., "Principal Component Analysis", Springer-Verlag, New York,1986
6. Lam L., Suen C.Y., "A Theoretical Analysis of the Application of Majority Voting to Pattern Recognition", Proc. 12th Int'l Conf. Pattern Recognition and Computer Vision, pp. 418-420, Jerusalem, 1994.
7. Manjunath B.S., Salembier, Sikora T., "Introduction to MPEG-7, Multimedia Content Description Interface" , J. Wiley, New York, 2002
8. Overview of the MPEG-7 Standard http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm
9. Szummer M., Picard R.: "Indoor-outdoor image classification" , IEEE international workshop on content-based access of images and video databases, 1998, Bombay, India

10. Vailaya A., Jani A., Zhang H.J.: "On image classification: City images vs landscapes" , Pattern Recognition vol. 31, pp. 1921-1936, 1998
11. Xu L., Krzyzak A., Suen C.Y., "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition", IEEE Trans. Systems, Man, and Cybernetics, vol. 22, no. 3, pp. 418-435, 1992.

3. Literature review

This chapter of our study presents the algorithms, methods, and concepts from fields such as fuzzy clustering, image representation, as well as, image classification and organization. We will carefully examine the existing trends and challenges in image classification.

3.1 *Image representation*

The contents of images can be very complicated and widely variant. Considerable effort has been devoted by image retrieval researchers to the search for compact image representation (choosing the "right" features). A digital image is a two-dimensional (2D) discrete signal. Mathematically, these signals can be represented as functions of two independent variables (for example a brightness function of two spatial variables).

There have been numerous numbers of approaches to the representation of images in the image database. There are two main types of this representation. The first of them consists of assigning a set of keywords to a given image [1,30,35]. In the second case an image is represented by a set of low-level features usually related to the shape, color or texture of an image [13,16]. As mentioned before, the quality of the content-based image retrieval system strongly depends on the choice of the features. Some image categories may be better recognized by some specific kinds of features.

It is important to mention here that keyword-based image search is a very powerful method. The problem is that it works only when all images are annotated with textual information. Annotating a large number of images is, however, a tough and time-consuming task. It is not an exaggeration to say that a process of image labeling requires a lot of effort. Usually, users write a query that contains a set of keywords related to the subject they want to search. If all images are categorized manually by the user or all images are labelled manually, it is possible to retrieve a set of images that more or less meets the user's requirements. But there is always possibility of a gap between requirements of a specific user and labels that are created by somebody else. An experimental study on manual indexing for information retrieval systems has shown that the degree of overlap in the keywords selected by two similarly trained people to represent the same document is not higher than 30%. We can assume that we can expect similar problem with labelling of images. The attachment of text labels to images is inadequate, since identical images can be described in different ways.

All drawbacks of keyword-based image representation give us a conclusion that a direct extraction of visual information from images is required. This gives rise to a series of low-level features (descriptors) whose use we intend to capture the mapping between images (or their descriptors, to be more precise) and the categories of images. A color digital image is typically represented by a triplet of values related to individual color channels. The main example of this kind of coding is the frequently used RGB color scheme. The individual color values are almost universally 8-bit and as the result one pixel is represented by 24 bits in total. This yields a threefold increase in the storage requirements for color versus monochrome images, and what is even more interesting in our case, the dimensionality of the problem is very large. We can assume that one pixel can be treated as one dimension. More precisely, the number of dimensions of a given image is equal to a number of its pixels. In this case the dimensionality of the problem is excessively high and representations of image data, where most information is packed into a small number of samples, are needed. Usually, these representations are obtained by non-redundant and invertible transforms. Nowadays, the most common techniques for this purpose are the wavelet transform [8,19,33] and the discrete cosine transform [23]. Moving Picture Experts Group (MPEG) developed MPEG-7 standard, which allows us to present images in different granularity in its descriptions by providing us a set of low-level descriptors [4,15,17,20]. These descriptors come in a compact form and their dimensionality is much lower as compared to a traditional representation of images.

Textual attributes or visual features alone, may be insufficient to correctly describe an image. Textual attributes should be used to describe the secret semantic of an image. On the contrary, visual features should describe the image content such as color and shape that may be extracted automatically or semi-automatically. In such a way, the textual descriptions are not redundant with regard to visual descriptions, but are highly complementary.

3.2 Image classification and organization systems

Content-based multimedia information retrieval has become one of the most active research areas in the past few years. Significant research has focused on determining efficient methodologies for retrieving images from the large image database. The need of tools to manage this rapidly-increasing quantity of visual data is greater than ever. Many visual representations (color, shape, texture, etc.) have been extensively explored and investigated. The current means of automatically sorting of images is very simple, usually based on filtering by considering the time they were created or their manually-defined location (for example a particular trip or event).

Relevant prior work is related to the problem of the dimensionality reduction of visual feature descriptors. One of the most popular feature space reduction method is the well known Principal Component Analysis [13,16].

Pratt [21] divided feature space reduction methods into two categories: the figure of merit and the prototype performance approach. In the first case, we use some criteria such as the Bhattacharyya distance or Mahalanobis distance in the comparison of objects' separability. In the second case, the algorithm consists in classifying data using different combinations of data and to keep the sets of data that gave us the best results.

Another space reduction method [9] maximizes the between-class variance and minimizes the average within-class variance. Other important methods are the canonical correspondence analysis [31] and the discriminant analysis [29].

Related prior work on image organization includes image clustering based on visual features and/or annotations [22], as well as, the thesaurus-based approaches [25,30].

In systems based on the Query-by-Example paradigm, the user provides to the system a piece of multimedia data that represents what the user wants to retrieve from the image database. This piece of data serves as the pattern for performing similarity search. The main examples of such systems are QBIC [10], Photobook [22] and Visualseek [28].

A number of work maps images onto one- and two-dimensional spaces [7] and [25] respectively, based on feature descriptors extracted from the images. The interest of these approaches is limited because they only display visual relationships and do not provide a structure to organize images.

Carson *et al.* proposed a blob-based image representation which calculates image similarities based on the visual similarities of image blobs [3]. The system developed by Torralba and Oliva uses discriminant structural templates to represent the global visual properties of natural scene images [25].

A system called MediaNet [2] incorporates lexical characterization, instance–based representations and the feature description of the multimedia entities. It makes use of complicated logical structures like ontologies [13] in order to infer knowledge from a given dataset.

Several works suggest organization of images based on existing thesauri [30,35]. In this case, relevant concepts in thesauri are found for images based on image annotations [30] or a given user feedback [35].

In another approach, image clustering algorithms group images hierarchically using feature descriptors extracted from images [5] or by modelling the distribution of visual descriptors and words [1].

In general, in most cases the set of retrieved images fits the user's needs only partly - no matter how suitable for the task at hand, the features and the similarity metrics were. In other words, there is a need for mechanisms that can adopt the CBIR system response based on some feedback from the user [25]. Usually, relevance feedback is used in CBIR systems to optimize some parametric similarity metric [11,27]. Relevance feedback can reduce a gap between the low-level features extracted from images like color or texture and the high-level semantic features that a user can use in describing an image. This idea is taken from relevance feedback of documents and adopted to content-based image retrieval.

There is a big number of techniques focused on exploiting such relevance feedback that have been proposed in the literature [12,27,34]. In most cases, they are based on the fact that the user does not know the actual distribution of images in the feature space, the feature space itself, nor the similarity metric employed [11]. In particular, the nature of the feedback provided by a user could be a selection of only positive or relevant examples of images. Some systems that are more complex permit both positive and negative feedback examples and even further descriptions such as degrees of relevance or irrelevance. In this case, information which is taken form the user is the most accurate. Obviously, as additional information can be captured from the user and properly used, the gap in retrieval semantic can be effectively reduced. In other words, such additional higher-level interactions of the user with the content-based image retrieval systems provide valuable information for efficient image retrieval.

3.3 Conclusions

The main purpose of this chapter was to present main trends in areas such as image classification and representation. As we can see there are many approaches to this topics. We demonstrated differences between keyword-based representation and representation based on low-level features of images. Also, this chapter gave us a comprehensive review of many different classification systems and the ideas behind them. Additionally, we outlined the main idea of relevance feedback systems.

3.4 References

1. Barnard K., Duygulu P., Forsyth D., de Freitas N., Blei D., Jordan M.I., "Matching Words and Pictures", submitted to JMLR.
2. Benitez A.B., Smith J.R., Chang S.F., "MediaNet: A Multimedia Information Network for Knowledge Representation", Proceedings of SPIE, 4210, October 2000.
3. Carson C., Belongie S., Greenspan H., Malik J., "Region-based image querying", In CAIVL, Washington, DC, USA, 1997, IEEE Computer Society.
4. Chang S.F., Sikora T., Puri A., "Overview of the mpeg-7 standard", IEEE trans. on Circuits and Systems for Video Technology 11 (2001) 688-695
5. Chen J., Bouman C.A., Dalton J.C., "Hierarchical browsing and search of large image databases", IEEE Trans. On Image Processing, Vol. 9, No. 3, 2000.
6. Cox I.J., Miller M.L., Omohundro S.M., Yianilos P.N., Pichunter, "Bayesian relevance feedback for image retrieval system", in: International Conference on Pattern Recognition, August 1996, pp. 361-369
7. Craver S., Yeo B.-L., Yeung M., "Multi-Linearization Data Structure for Image Browsing", IS&T/SPIE-1999, Vol. 3656, pp. 155-166, Jan. 1999.
8. Donoho D. L., Vetterli M., DeVore R.A., Daubechies I., "Data compression and harmonic analysis", IEEE Trans. Inform. Theory, vol. 44, pp. 2435-2476, Oct. 1998.
9. Duda R.O., Hart P.E., "Pattern Classification and Scene Analysis", John Wiley & Sons, New York, 1973.
10. Fickner M., Sawhney H., Niblack W., Ashley J., Huang Q., Dom B., Gorkani M., Hafner J., Lee D., Petkovic D., Steele D., Yanker P., "Query by Image and Video Content: the QBIC System", Computer, 28(9), pp. 23-32, Sept. 1995
11. Giacinto G., Roli F., "Bayesian relevance feedback for content-based image retrieval", in: Pattern recognition 37 (2004), pp. 1499-1508

12. Koenen R., Pereira F., "MPEG-7: a standard description of audiovisual content", Signal Processing: Image Communication, 16, 2000, 5-13
13. Landgrebe D.A, "Signal Theory Methods in Multispectral Remote Sensing", Wiley-Intersciense, 2003.
14. Lijuan D., Wen G., Wei Z., Debin Z., "Adaptive relevance feedback based on Bayesian inference for image retrieval", in: Signal Processing 85 (2005), pp.395-399
15. Mallat S., "A Wavelet Tour of Signal Processing", 2nd ed. New York: Academic, 1999.
16. Manjunath B., Ohm J.R., Vasudevan V.V., Yamada A.: "Color and texture descriptors", IEEE trans. on Circuits and Systems for Video Technology 11 (2001) 703-71
17. Manjunath B.S., Salembier, Sikora T., "Introduction to MPEG-7, Multimedia Content Description Interface", J. Wiley, New York, 2002
18. Mezaris V., Kompatsiaris I., Strintzis M. G., "An Ontology Approach to Object-based Image Retrieval", In Proc. IEEE International Conference on Image Processing (ICIP), volume 2, pp. 511-514, Barcelona, Spain, September 2003.
19. MPEG-7 Description Definition Language http://archive.dstc.edu.au/mpeg-7-dll
20. Overview of the MPEG-7 Standard http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm
21. Pratt W.K., "Digital Image Processing", John Wiley & Sons, New York, 1991.
22. Pentland A., Picard R., Sclaroff S., "Photobook: Content-based manipulation of Image Databases", Intl. J. Computer Vision, 18(3), pp. 233-254, 1996.
23. Rao K. R., Yip P., "Discrete Cosine Transform: Algorithms, Advantages, Applications", New York: Academic, 1990.

24. Ronald R. Yager, Frederick E., Petry, "A framework for linguistic relevance feedback in content-based image retrieval using fuzzy logic", in: Information Sciences 173 (2005) 337-352

25. Rubner Y., Guibas L.J., Tomasi C., "The Earth Mover's Distance, Multi-Dimensional Scaling, and Color-Based Image Retrieval", DARPA Image Understanding Workshop 1997.

26. Rui R., et al., "A Relevance Feedback Architecture in Content-Based Multimedia Information Systems", Proc. of IEEE Workshop on CAIVL, 1997.

27. Smeulders A.W.M., Worring M., Santini S., Gupta A., Jain R. "Content-based image retrieval at the end of the early years", IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 1349–1380.

28. Smith J.R., Chang S.F., "Visualseek: A Fully Automated Content-based Image Query System", In ACM Multimedia, pp. 87-98, 1996.

29. Tabachnick B.G., Fidell L.S., "Using Multivariate Statistics", Ally and Bacon, 4th edition, Needham Heights, MA, 2001.

30. Tansley R., "The Multimedia Thesaurus: Adding A Semantic Layer to Multimedia Information", Ph. D. Thesis, Computer Science, University of Southampton, Southampton, UK, 2000.

31. Ter Braak C.J.F, Prentice I.C., "A Theory of Gradient Analysis", Advances in Ecological Research 18, pp. 272-317, 1998.

32. Torralba A.B., Oliva A., "Semantic organization of scenes using discriminant structural templates", In ICCV, volume 2, 1999.

33. Vetterli M., Kovacevic J., "Wavelets and Subband Coding", Englewood Cliffs, NJ: Prentice-Hall, 1995.

34. Wood M.E.J, Thomas B.T., Campbell N.W., "Iterative refinement by relevance feedback in content-based digital image retrieval", in: Proceeding of 5th ACM International Multimedia Conference, September 1998, pp. 13-20

35. Yang, J., Wenyin L., Zahng H., Zhuang Y., "Thesaurus-Aided Approach for Image Browsing and Retrieval", ICME- 2001, Tokyo, Aug. 2001.

4. Dataset description

This chapter presents our image dataset, image categories and describes MPEG-7 standard and descriptors. Additionally, we provide examples of MPEG-7 descriptors .

In our experiments, we consider 10,000 color images of different size coming from the Corel database (http://www.corel.com). The Corel database is a set of images which includes thousands of images from a very big variety of categories. Some of these images are labeled by professional annotators and described by a set of standardized keywords. In our case, a feature space is based on the contents of images rather that keywords related to them. Because of that fact, all photos are described using the standard MPEG-7 features (that will be presented later) which gives us an opportunity to present images in the numerical form.

4.1 Image categories

Our database of images was analyzed and we manually classified images into five categories (classes of images) [2]. Each category consists of 2,000 images. We consider that a certain image falls into a given class of pictures if the camera is primarily focused in the object that is adequate to the name of the category

Here is a brief characteristic of each category of the images used in the experiments:

animal – a picture is assigned to this category if there is an animal of a visible size. Figure 4-1 includes some examples of images falling under this category. As we can see on examples below, there are many different species represented on images. We note that animals could appear in its entirety or only a part of the animal could be present. Additionally, there are many different backgrounds, usually animals are surrounded by some vegetations but also there can be a plain unnatural background. It may be a very big problem for the classification system to interpret correctly this kind of images.

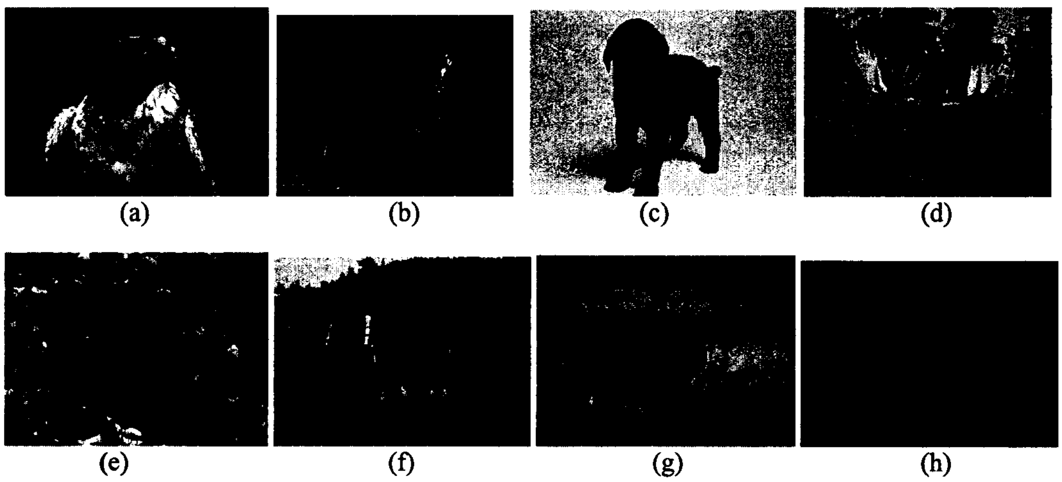

Figure 4-1. Examples of images belonging to the category of animals.

building – an image falls into this category when a building of a visible size appears in the scene. As we can see on examples presented below (Figure 4-2), there is a great deal of diversity in this category of images. We can assume that it may quite substantially contribute to possible challenges in their classification. In noticeable that we have to deal with different types of buildings (for example residential houses, churches, castles and many more). Additionally, possible difficulties could be expected by different distances used in the close-ups. It is worth stressing, that we may expect that descriptors related to the texture of the image will provide us the best classification accuracy. Images of buildings include characteristic edges and shapes, as well as, characteristic textures with regular shapes. However, we can expect some problems related to the background of images. Usually buildings are surrounded by environment which may be interpreted as landscape.

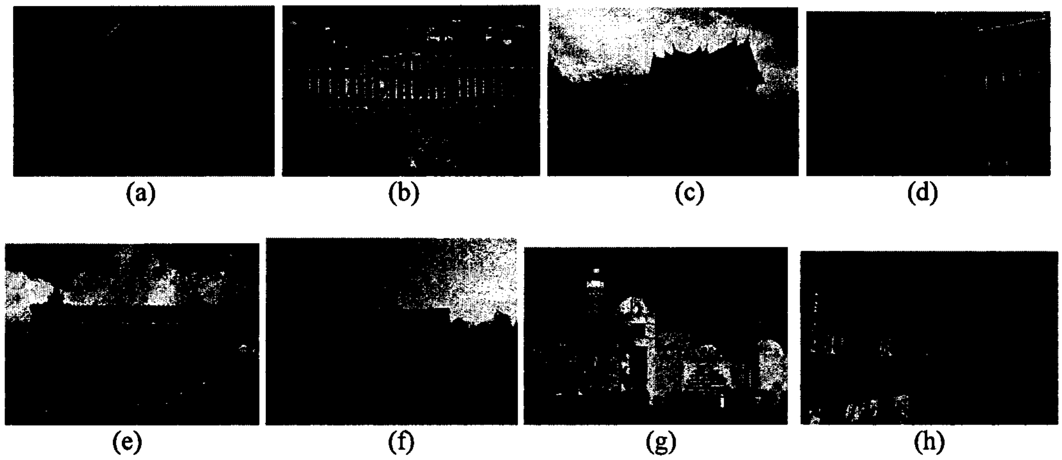

Figure 4-2. Examples of images belonging to the category of buildings

people – a picture is assigned to this category when people appear in the picture. As we can see on examples shown below (Figure 4-3) it can be not only a human face but also a whole human figure. It is noticeable that appeared people have different color of skin, sometimes they appear in groups, or they are on the huge variety of backgrounds. We may expect that there will be many overlaps with vegetation or animal category because these categories have many commonalities.

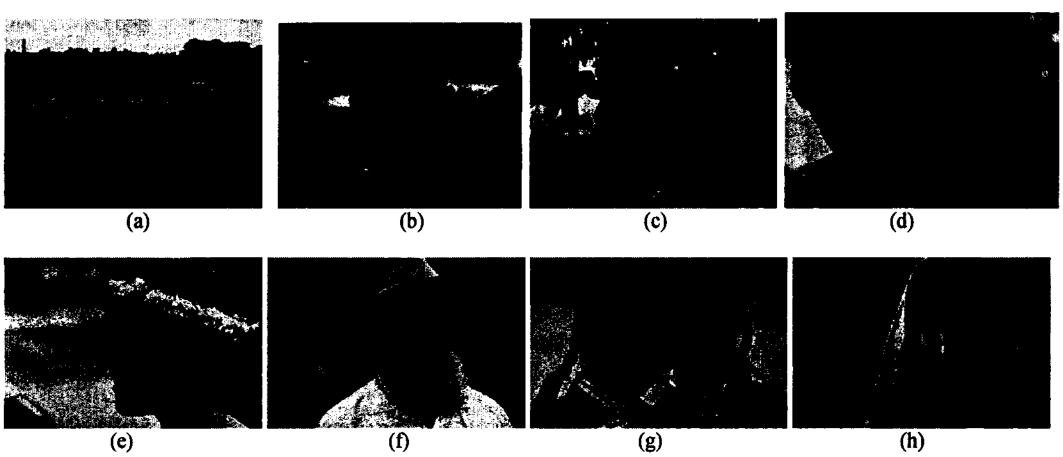

Figure 4-3. Examples of images belonging to the category of people.

landscape – an image belongs to this category when the image shows scenery with natural elements. (mountains, seashores, forest, etc.). Additionally, which may be a problem in classification, there can be some man-made objects in the pictures but they show up to be very small (for example buildings,) It is worth stressing that these images show significant similarity to the building category so we may expect some overlaps between these categories (Figure 4-4)

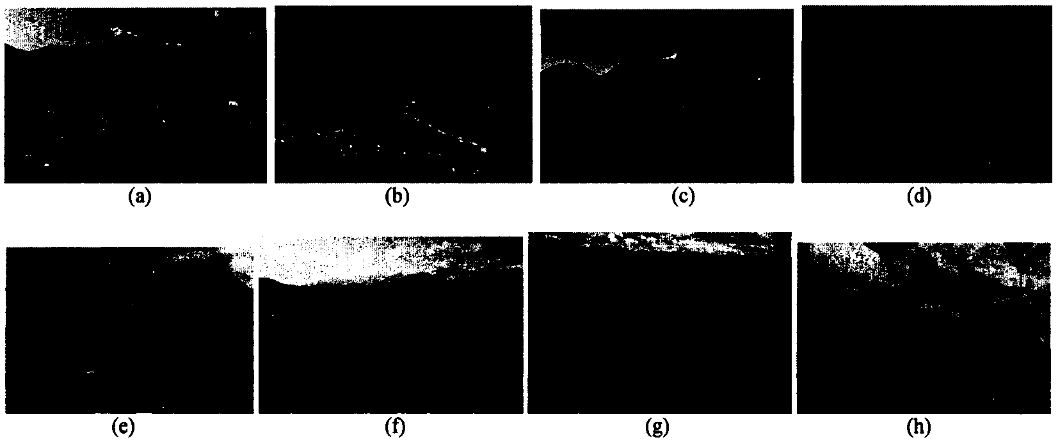

Figure 4-4. Examples of images belonging to the category of landscapes.

vegetation – a picture belongs to this category when nature appears in the picture. The main examples of natural elements may be plants, trees, flowers, etc. As we can see on pictures, it can be a single object (for example one flower) or a group of objects (i.e. a bunch of flowers). There is a huge variety of different colors and shapes. It is easy to see that elements of the vegetation category frequently appear in other categories. Figure 4-5 shows examples of these pictures.

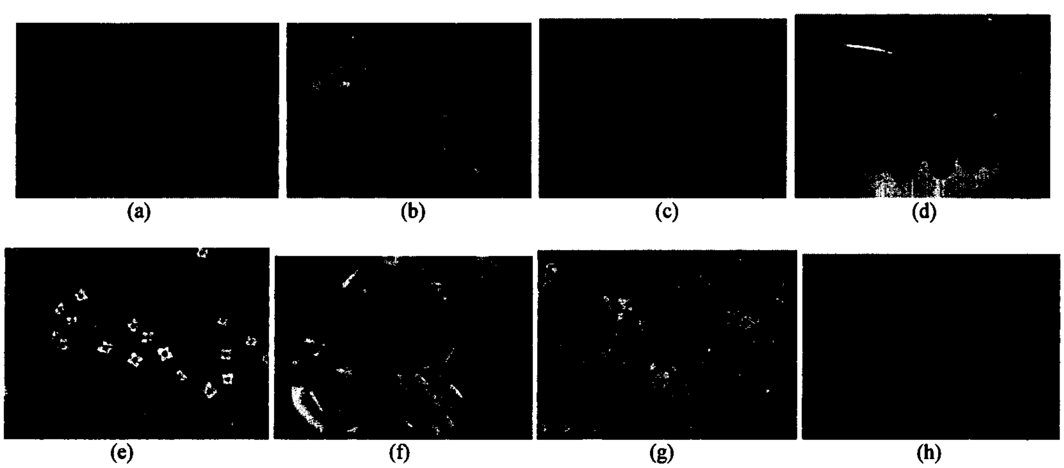

Figure 4-5. Examples of images belonging to the category of vegetation

The experimental data used in our research includes a large variety of images. It is worth stressing that the five categories that we formed are very general. Because of this fact, as shown and mentioned in above examples of images, an overlap between images from different categories is quite substantial and will have a big influence on classification results.

So, now we can summarize our expectations of overlaps that may appear in our dataset. In the case of the category of animals, animals usually appeared surrounded by vegetation. On the other hand, base on our visual assumptions we can say that they could be recognized as an object which is similar to the person, as well as, the person may be recognized as some kind of animal. Secondly, city views sometimes may include trees and mountains and we know that they are common to landscapes. Additionally as it was show in examples, many photos of buildings include the background (for example sky), that is very common in images coming from the landscape category. Based on these observations, we can assume that there will be many images that will be assigned to two or more different categories because of their several high-level and low-level commonalities.

In the study, we consider the five-class problem (because we divided our dataset into five categories). As we stressed above, by noting that some categories overlap quite substantially, additionally we investigate the use of three categories. More precisely, the categories of animals and people are put together forming a single combined category. On the other hand, the building and landscape categories will form the second combined category.

4.2 MPEG-7 Descriptors

MPEG-7 is a standard in multimedia description. The Moving Picture Experts Group (MPEG) developed this standard for description and search of audio and video content [1,3,4,5].

MPEG-7 descriptions do not depend on the way how a given content is stored or coded. For example, we can create a description of the picture that is printed on the paper as well as description of an analogue or digitalized movie. In general, MPEG-7 descriptions associated with audiovisual data content may include data related to visual information (for example still pictures, graphics, 3D models, videos, facial characteristic) and data related to audio data (for example speech). This standard also describes how all these elements are combined together in a multimedia presentation.

MPEG-7 standard gives us different points of view because allows us to present data in different granularity in its descriptions. It means that offers us the possibility to have different levels of discrimination.

More precisely, descriptions of content may include:

· the creation and production processes of the content , some general information about the multimedia presentation (director, title, short feature movie).

· the usage of the content (copyright pointers, usage history, broadcast schedule).

· storage features of the content (storage format, encoding method).

· structural information related to spatial, temporal or spatio-temporal components of the content (list of scene cuts, segmentation in regions, tracking of region motion).

· low level features of the content (for example colors, textures, sound timbres, melody description). These descriptions are used in our research.

· conceptual information (list of objects and events as well as interactions among objects).

· some additional information which helps us in browsing the content in an efficient way (for example summaries, variations, spatial and frequency subbands).

· collections of objects.

· interaction of the user with the content (for example user preferences, usage history).

All these descriptions are implemented in an efficient way for searching and filtering the content so they can be used widely in case of classification problems.

MPEG- 7 specifies standardized Descriptors and Description Schemes for audio and video as well as an integrated multimedia content. There are two main types of descriptors: low level and high level descriptors. Low-level descriptors are related to the visual construction of the scene This kind of descriptors captures information like color, shape, texture and motion. As it was pointed out, we are focused on some low-level descriptors. In this case, low-level descriptors come in the form of vectors involving a number of bins. So, we can say that MPEG-7 has some efficient color and texture descriptors for similarity matching. In our research we use five MPEG-7 low-level descriptors.

Basically, the low-level descriptors of image content are Color Layout, Color Structure, Dominant Color, Scalable Color, Edge Histogram and Homogeneous Texture.

4.2.1 Visual Color Descriptors

As we know, color is one of the most important features in image and video retrieval [1,3,4,5]. It is worth stressing that colors are independent of the image size and its orientation. We can expect that color features are relatively robust to changes in the background colors. Additionally, color descriptors can be useful for describing content in still images and as well as video. Because of this fact, MPEG-7 has some efficient color descriptors for similarity matching. We can assume that it may be hard to develop only one color descriptor that can be used for all foreseen applications. Taking into account this fact, a range of descriptors has been standardized. In other words each descriptor is suitable for achieving specific similarity-matching functionalities.

It is very important to have an interoperability between different color descriptors. Bearing in mind this objective, normative color spaces are constrained to hue-saturation value (HSV) and hue-min-max-diff (HMMD). HSV is well-known color space widely used in image applications. On the other hand, HMMD is a new color space defined by MPEG and is only used in the color structure descriptor (CSD).

There is a brief overview of each descriptor:

Scalable Color Descriptor (SCD)

We can say that this is the most basic description of color features. This descriptor is provided by describing color distribution in images. If this kind of distribution is measured over an entire image, global color features can be described. Scalable Color is a color histogram which is encoded by a Haar transform and uses the HSV color space uniformly quantized to 255 bins. Then, the histogram can be quantized in a wide range of values. The quantization step depends on requirements of a specific application.

There are five possible sizes of output vectors: 16, 32, 64, 128, 256. The system developed by us supports different quantization levels.

In the experiments reported there, we focused on the size of the output vector of 64 values. For this descriptor, only first 64 values of the output vector are meaningful. The rest of the values are very close to zero so we can assume that contribution given by these bins is not significant. Figure 4-6 shows some examples of this descriptor.

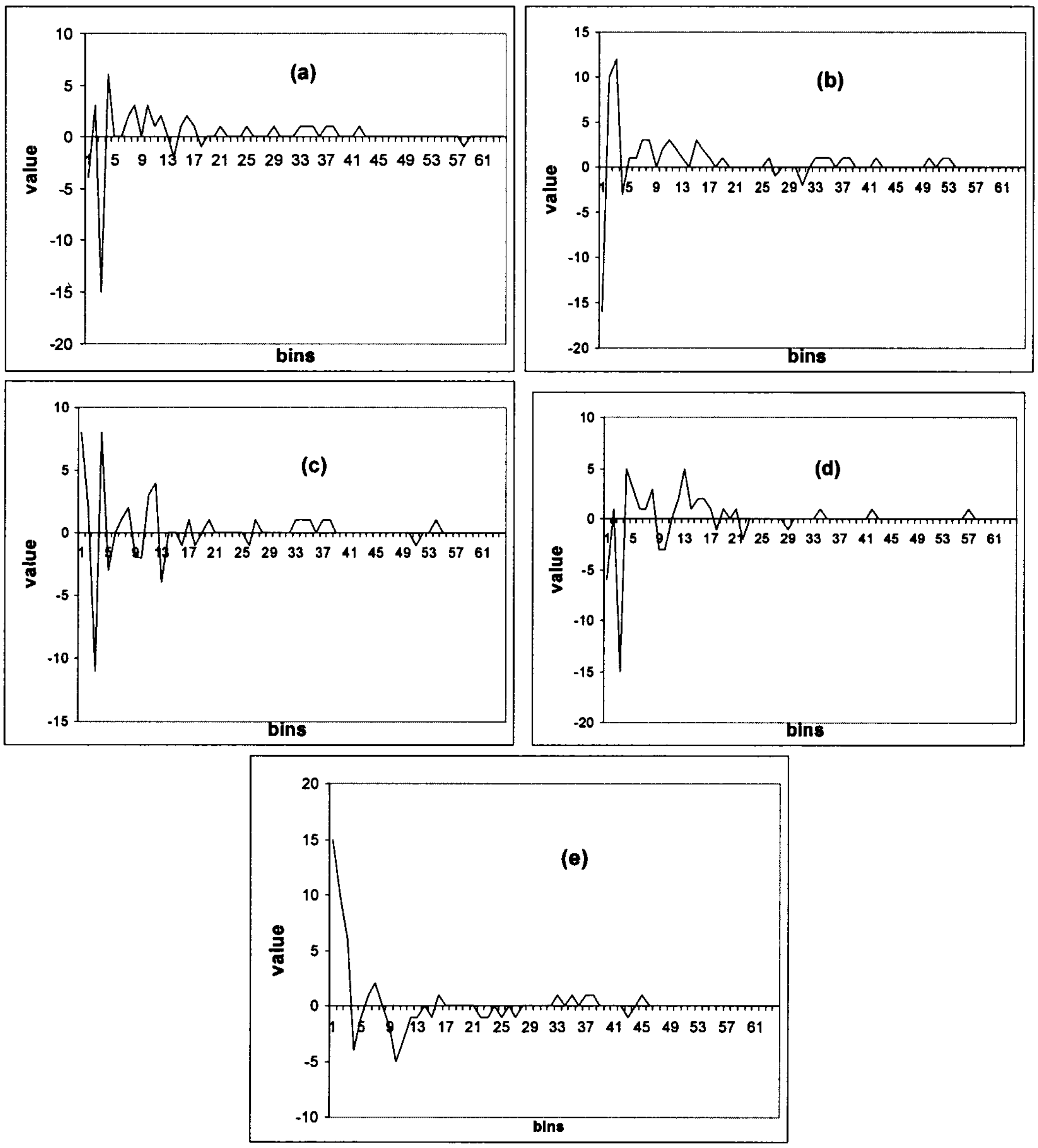

Figure 4-6. Scalable Color Descriptor individual bins: (a) an image from Animal category (Figure 4-1 (a)), (b) an image from Building category (Figure 4-2 (a)), (c) an image from Landscape category (Figure 4-3 (a)), (d) an image from People category (Figure 4-4 (a)), (e) an image from Vegetation category (Figure 4-5 (a))

Color Layout Descriptor (CLD)

This descriptor is designed to describe spatial distribution of color in an arbitrarily-shaped region. Color distribution in each region can be described using the Dominant Color Descriptor above. The spatial distribution of color is an effective description for sketch-based retrieval, content filtering using image indexing, and visualization. To obtain descriptor values the image is divided into 8x8 blocks. Next, for each block, a dominant color is selected and the resulting 8x8 image is transformed into a series of coefficients using dominant color descriptors transformation. Finally, these coefficients are quantized to fit an assigned number of bins.

The descriptor output is a vector with integer numbers, describing {Y,Cr,Cb} coefficients, where Y is the coefficient value for luminance, Cr, Cb coefficient values for chrominance.

There are five possible sizes of the output vector: 12 (6Y-3Cr-3Cb), 27 (15Y-6Cr-6Cb), 58 (28Y-15Cr-15Cb), 120 (64Y-28Cr-28Cb). In our experiments we focused on the vector size equal to 58.

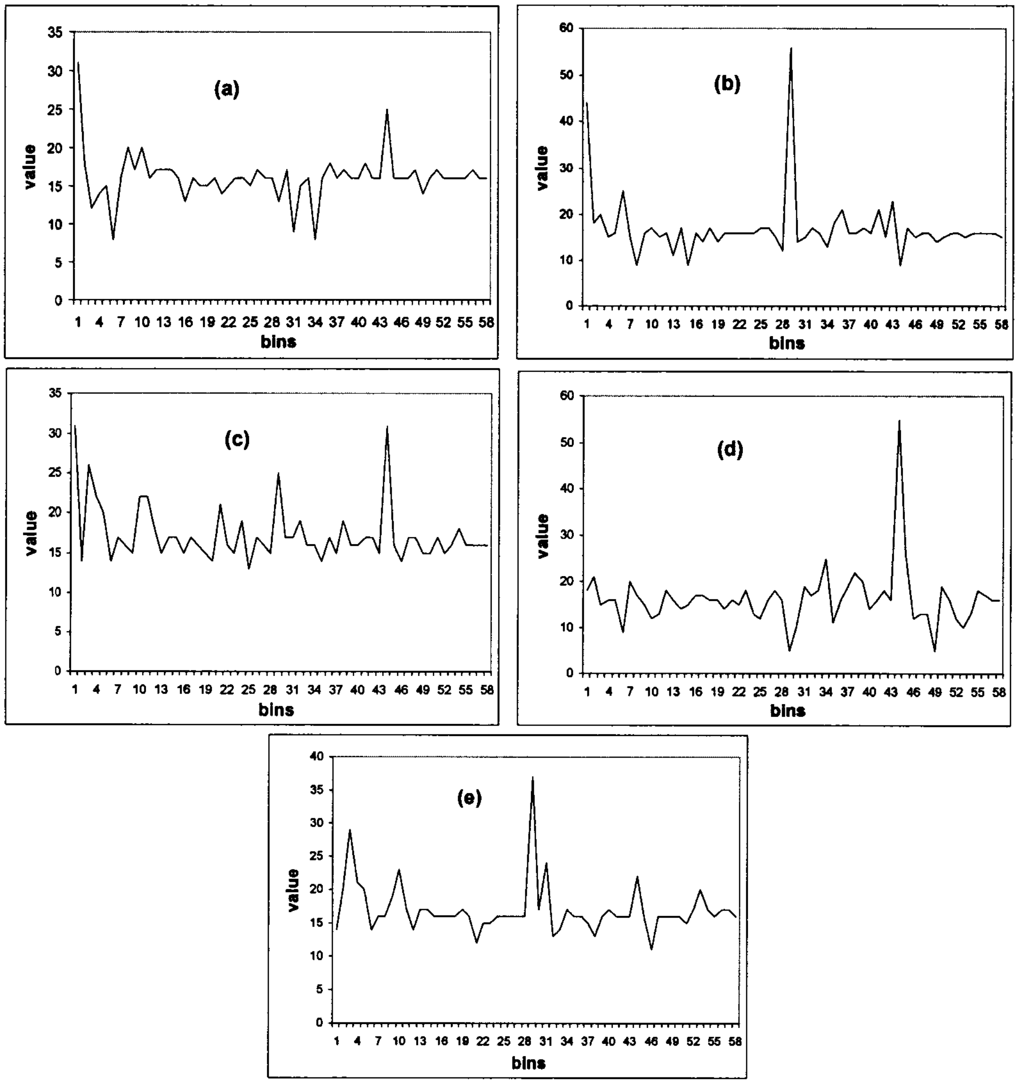

Figure 4-7. Color Layout Descriptor histograms: (a) an image from Animal category (Figure 4-1 (a)), (b) an image from Building category (Figure 4-2 (a)), (c) an image from Landscape category (Figure 4-3 (a)), (d) an image from People category (Figure 4-4 (a)), (e) an image from Vegetation category (Figure 4-5 (a))

Color Structure Descriptor (SCD)

This descriptor can be used if we want to express local color features in images. To obtain this, a 8x8 pel structuring block scans the image in a sliding window approach. After each shift of the structuring element, the number of times a given color is contained in the structure element is counted. Then, a color histogram is constructed. The structuring element always has dimensions 8x8, but the distance between the samples in the original image differs with the resolution.

The histogram is extracted in the HMMD color space and non-uniformly quantizing is performed over the histogram values.

The output of this descriptor is a vector with integer components, presented by a 256 bin histograms.

There are four quantization levels: 32, 64, 128, 256. 256 vector gives us the highest level of details.

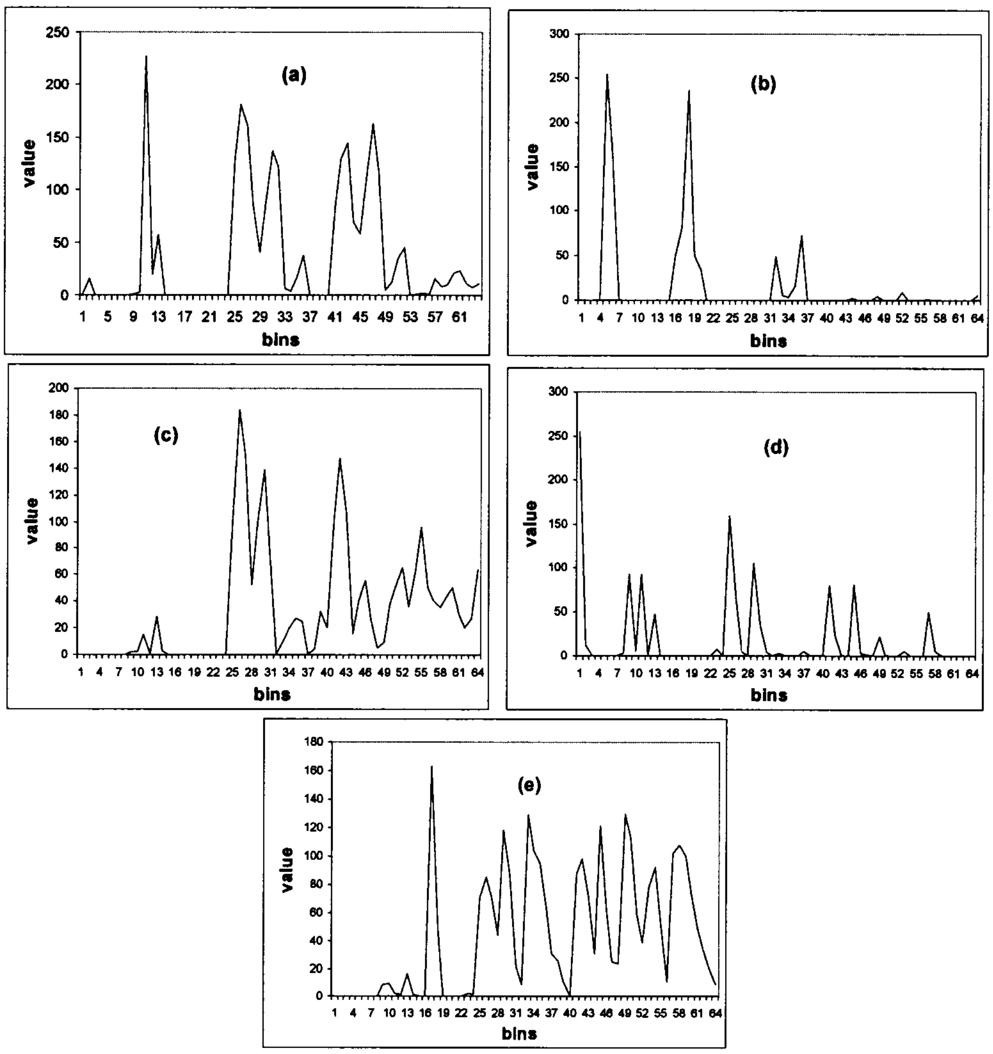

Figure 4-8. Color Structure Descriptor histograms: (a) an image from Animal category (Figure 4-1 (a)), (b) an image from Building category (Figure 4-2 (a)), (c) an image from Landscape category (Figure 4-3 (a)), (d) an image from People category (Figure 4-4 (a)), (e) an image from Vegetation category (Figure 4-5 (a))

4.2.2 Visual Texture Descriptors

The texture is a property of virtually any surface, including trees, walls, bricks, hair etc. [1,3,4,5]. The texture contains important structural information of surfaces as well as their relationship to the surrounding environment. MPEG-7 has defined appropriate texture descriptors that can be useful for a wide range of applications and tasks.

Homogenous Texture Descriptor (HT)

The Homogenous Texture Descriptor describes directionality, coarseness, and regularity of patterns in images. This descriptor can be very valuable for a quantitative characterization of texture which has homogenous properties. This descriptor is based on a filter bank approach employing scale and orientation sensitive filters. The descriptions are obtained in the frequency domain by computing mean and standard variation of frequency coefficients. A radon transform followed by Fourier transform can be employed to achieve adequate computational efficiency for low complexity applications. The frequency space is divided into 30 channels with equal division in the angular direction and octave division in radial direction. Feature channels are filtered using 2-D Gabor functions.

The output of the method is a vector of 63 values - the average value (an integer number in the interval [0,255]), standard deviation (an integer number in the interval [0,255], energy (30 integer numbers in the interval [0,255] and energy deviation (30 integer numbers in the interval [0,255]).

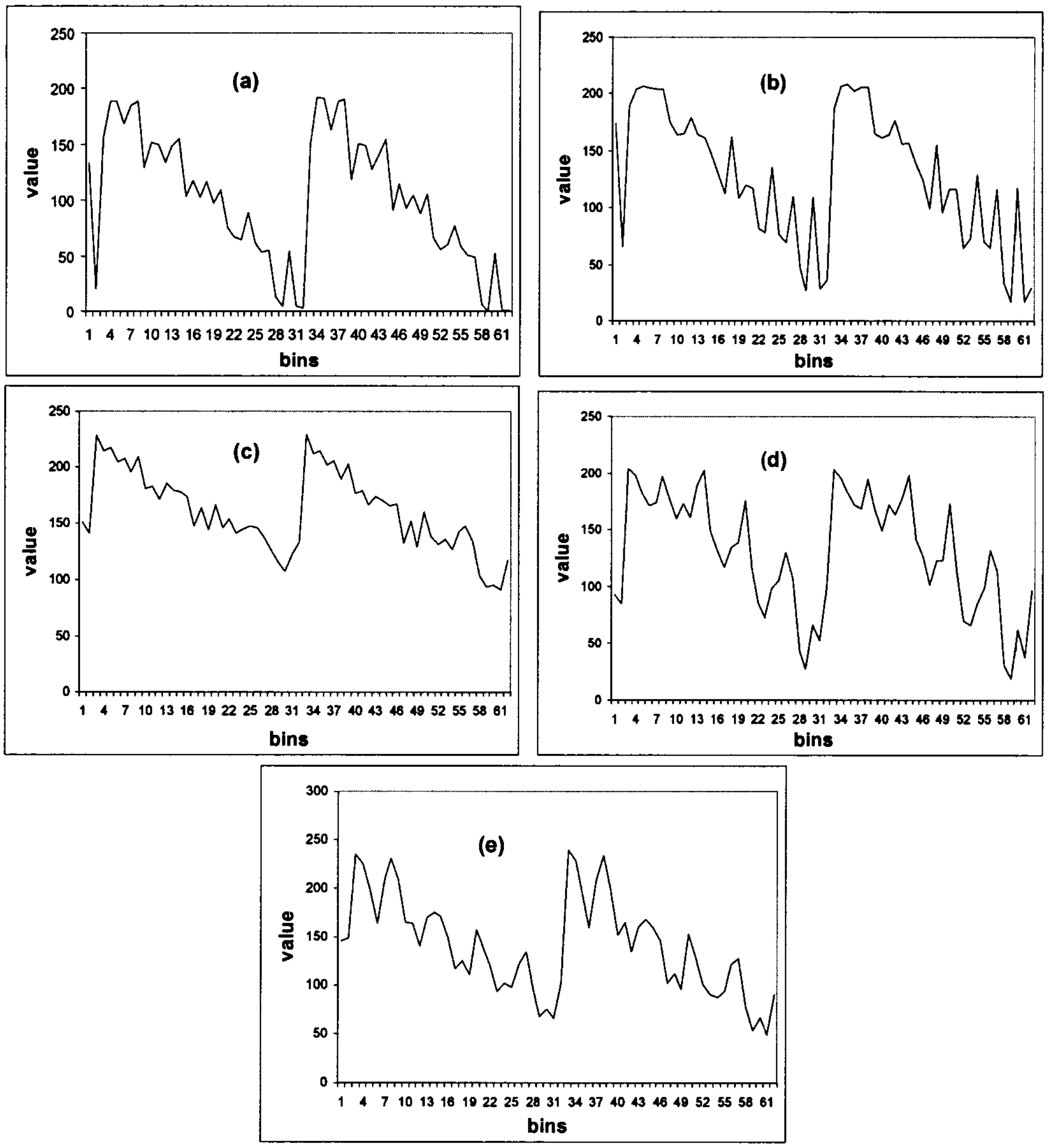

Figure 4-9. Homogenous Texture histograms: (a) an image from Animal category (Figure 4-1 (a)), (b) an image from Building category (Figure 4-2 (a)), (c) an image from Landscape category (Figure 4-3 (a)), (d) an image from People category (Figure 4-4 (a)), (e) an image from Vegetation category (Figure 4-5 (a))

Edge Histogram (EH)

Edge in the image is considered as an important feature to represent the content of the image. The spatial distribution of edges is captured by this descriptor. There are four directional edges and one nondirectional edge in three different levels of localization in an image. The localization levels are the global, the semi-global and the local level. An image is divided into 16 non-overlapping sub-images.

Additionally, it is divided into a preferred number of image-blocks. During the next step, for each image-block, a horizontal, a vertical, a 45 degree diagonal, a 135 degree diagonal and a nondirectional edge value is calculated using edge extraction filters applied on the average brightness values in four sub-blocks. The image block is considered to contain a corresponding edge if the maximum edge value is greater than a threshold value. A local edge histogram with a total of 80 bins (5 types of edges, for each of the 16 sub-images) is formed by the image-block edge composition in the sub-images. The global edge histogram is created by adding the corresponding local edge histogram bins into five global histogram bins one for each type of edge. The semi-global edge is created by accumulating the edge composition in the sub-image clusters.

The output is a vector of 80 integer numbers between [0,7].

There are some examples of MPEG-7 descriptors and their plots (Figure 4-10).

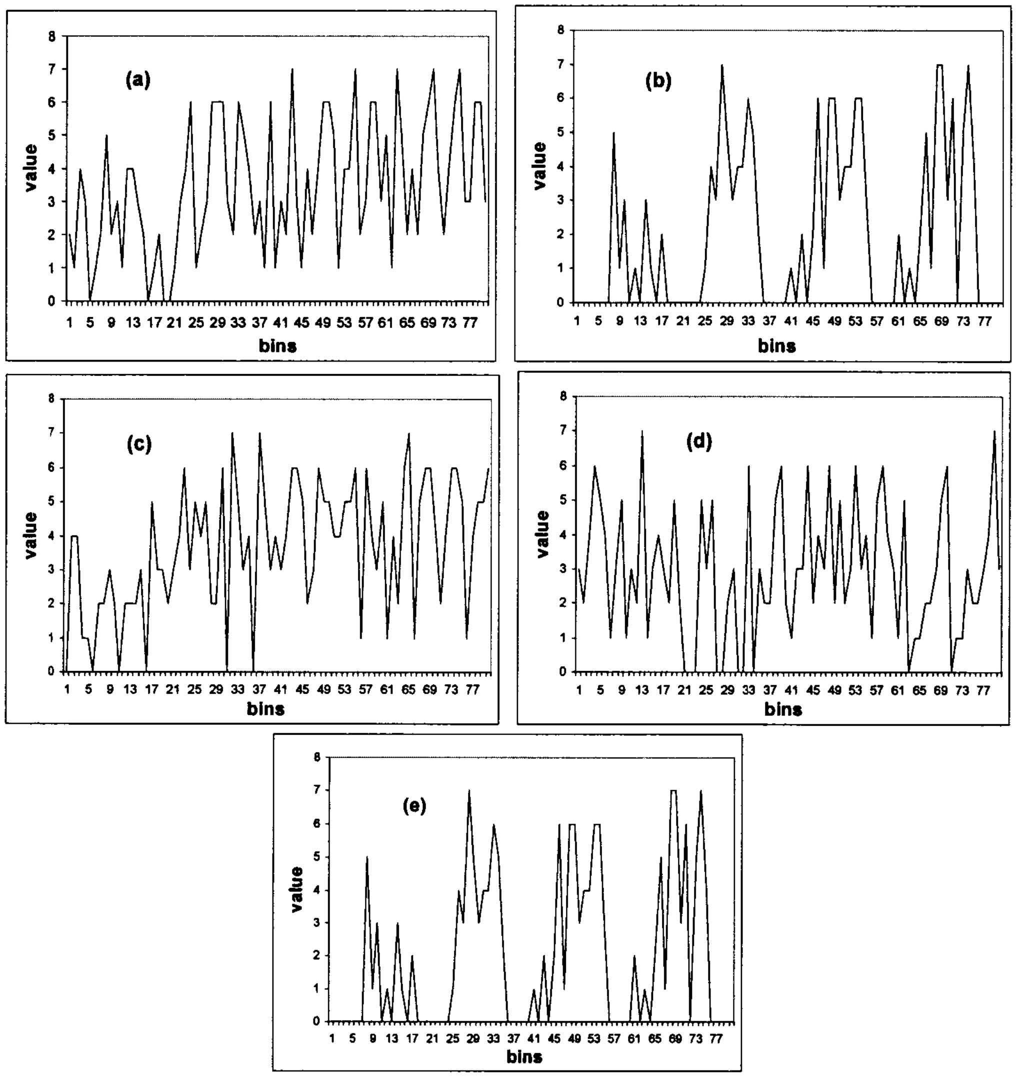

Figure 4-10. Edge Histogram histograms: (a) an image from Animal category (Figure 4-1 (a)), (b) an image from Building category (Figure 4-2 (a)), (c) an image from Landscape category (Figure 4-3 (a)), (d) an image from People category (Figure 4-4 (a)), (e) an image from Vegetation category (Figure 4-5 (a))

4.3 Conclusions

This chapter includes the description of image categories used in our experiments. We provided examples of images of every single category. Next, we presented details about MPEG-7 standard and its low-level descriptors. We provided examples of every single descriptor.

4.4 References

1. Chang S.F., Sikora T., Puri A., "Overview of the MPEG-7 standard", IEEE trans. on Circuits and Systems for Video Technology 11 (2001) 688-695
2. Kaczmarzyk T, Pedrycz W., "Content-based image retrieval: an application of MPEG-7 Standard and Fuzzy C-Means", 25th International Conference of the North American Fuzzy Information Processing Society (NAFIPS 2006) Proceedings, Montreal, June, 2006
3. Manjunath B.S., Salembier, Sikora T., "Introduction to MPEG-7, Multimedia Content Description Interface", J. Wiley, New York, 2002
4. MPEG-7 Description Definition Language *http://archive.dstc.edu.au/mpeg-7-dll*
5. Overview of the MPEG-7 Standard *http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm*

5. Algorithms

This chapter presents algorithms used in our experiments. At first, we describe Principal Component Analysis algorithm leading to the reduction of the original feature space. Next, we go to the description of the Fuzzy C-Means clustering and we finish with a presentation of classifier fusion algorithms.

5.1 Principal Component Analysis (PCA)

It is well known that the problem of reduction of a feature space is one of the most important aspect of pattern recognition. There is a big number of techniques which help us to create a reduced feature space such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA).

Principal Component Analysis is a statistical technique which shows its usefulness in many fields such as pattern recognition or image compression. PCA helps us to develop a reduced feature space. The reduction of a feature space may provide us better and more general results of image classification. PCA highlights differences and similarities between patterns. As mentioned, PCA reduces dimensionality of the problem without much of information.

To begin the transformation, the covariance matrix $\vec{C}$ of the original data is found. Using the covariance matrix, the eigenvalues $\vec{\lambda}$ are obtained from the equation:

$$\left|\vec{C} - \lambda_i \vec{I}\right| = 0 \quad (1)$$

Where $i \in [1,2,...,m]$ and m is equal to the dimensionality of the problem, and $\vec{I}$ is an identity matrix. The eigenvalues are equal to the variance of each corresponding principal component. The eigenvectors $\vec{e}_i$ define the axes of the components and are obtained from the equation:

$$\left(\vec{C} - \lambda_i \vec{I}\right)\vec{e}_i = 0 \qquad (2)$$

The principal components are then given as:

$$\vec{PC} = \vec{T} \cdot \vec{D} \qquad (3)$$

where $\vec{D}$ is the matrix of the original data, and $\vec{T}$ is the transformation matrix:

$$\vec{T} = \begin{bmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{m1} & \cdots & e_{mm} \end{bmatrix} \qquad (4)$$

If we wish to obtain the best n-dimensional representation of the m-dimensional problem, then we simply have to project the points onto the n-dimensional subspace defined by the first n principal components PC_1, PC_2, ..., PC_n with the highest values of eigenvalues.

Finally, we obtained our new feature space with lower dimensionality so we can expect better and more general results during classification process.

5.2 Knowledge discovery with fuzzy clustering

Clustering is an essential method of unsupervised learning. We can say that clustering can be defined as a search for a structure in data [1]. In general, the data can be any data taken from a physical process. Clustering process enables the computers to provide its findings to the researcher in usable and understandable forms. These forms on the methods, models, data used in experiments and the structure we expect to find. Precisely, the structure provides us information about relationships between variables in the process. The representation of the organized structure depends on the data, the model and the method of search. In summary, the data contains the information, the search process recognizes it and the structure represents it.

More specifically, the clustering provides a partitioning of a dataset $X=\{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N\}$ where N is a number of objects into $c\in\{2,...,N-1\}$ clusters. In general, *c* clusters were disjoint subsets of *X*. To illustrate the idea of clustering let us consider a following problem. The question is how to partition a set of three vehicles $X=\{\mathbf{x}_1=bicycle,\ \mathbf{x}_2=car,\ \mathbf{x}_3=motorcycle\}$ into 2 clusters. Let us consider A_1 as the subset containing a bicycle ($\mathbf{x}_1\in A_1$). Next, a car is completely different than a bicycle so we would create another subset A_1 ($\mathbf{x}_2\in A_2$). Both sets represent different features of the contained objects so we can say that $A_1\cap A_2=\varnothing$. Now, we consider the assignment of a motorcycle (a vehicle which has some features of a bike as well as of a car). Obviously, $\mathbf{x}_3\notin A_1\wedge\mathbf{x}_3\notin A_2$ so we have a constraint of this model resulting in impossibility of an assignment $\mathbf{x}_3$ (a motorbike) to any of the subsets. Precisely it means that, we cannot generate a 2-elements partition of the set *X* illustrating the features of objects from *X*. These limits can be successfully eliminated by using of fuzzy sets. In this case, we can imagine a function $u_1:X\to[0,1]$. Values of function $u_1(x)$ give us a grade of membership of *x* in the fuzzy set u_1. In practice, it helps us in a classification of a given object. A grade of membership of a motorbike can be expressed as $A_1(\mathbf{x}_3)=0.60$ and $A_2(\mathbf{x}_3)=0.40$. It tells us that a motorbike is partially similar to the bicycle as well as to the car.

We can define a fuzzy c-partition, a very suitable form of representation of a partitioning of *X*. *X* is a finite set,. The fuzzy c-partition of *X* is defined as:

$$\mathbf{M}_{fc}=\left\{\mathbf{U}\in\mathbf{V}_{cN}\mid u_{ik}\in[0,1]\,\forall i,k;\sum_{i=1}^{c}u_{ik}=1\,\forall k;0<\sum_{k=1}^{N}u_{ik}<N\,\forall i\right\} \tag{5}$$

Where:

V_{cN} is the set of real matrices,
c is an integer $2\le c<N$,
$u_{ik}=u_i(\mathbf{x}_k)$

We can represent an example solution of the problem of finding a 2-partition of *X*. The fuzzy c-partition may look as follows:

$$U = \begin{array}{c} \begin{array}{ccc} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{array} \\ \begin{bmatrix} 0.96 & 0.05 & 0.60 \\ 0.04 & 0.95 & 0.40 \end{bmatrix} \end{array} \qquad (6)$$

A membership function of $\mathbf{x}_1$ is equal to 0.96 in the first cluster. It means that $\mathbf{x}_1$ (bicycle) is more associated with the first cluster (first row of the matrix). Next, $\mathbf{x}_2$ belongs to the second cluster almost in 100%. On the other hand, $\mathbf{x}_3$ (motorcycle) combines features of both clusters, and this is represented in its membership functions – it belongs partially to both clusters but more strongly to the first cluster. For more detailed considerations the reader is referred to Bezdek [1].

As we presented above, partitioning of a given dataset is not a trivial process. We assume that the best fitted features were extracted and each pattern is represented by a feature vector. During clustering, it is necessary to distinguish between the different partitions and choose one which is the most suitable. In other words we have to quantify the quality of this process. The optimal solution of a given problem should minimize the error of clustering. It can be represented as the minimization of the distances between the patterns and the cluster centers to which these patterns belong. If the representation of the clustering process is the matrix $\mathbf{U} \in \mathbf{M}_{fc}$ as shown above (6), each distance can be weighted by the elements of the matrix $\mathbf{U}$. Finally, the expression to be minimized may look as follows:

$$J_m(\mathbf{U},\mathbf{v}) = \sum_{k=1}^{N} \sum_{i=1}^{c} (u_{ik})^m (d_{ik})^2 \qquad (7)$$

where:

$\mathbf{U} \in \mathbf{M}_{fc}$ is the fuzzy c-partition of X;

$u_{ik} \in [0,1]$ specifies the degree of membership of pattern $k = 1,2,\ldots,N$ in the cluster

$i = 1,\ldots,c$; $\mathbf{v}$ is the set of prototypes (centers of clusters)

$\mathbf{v} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c\}$ with $\mathbf{v}_i \in \mathbf{R}^p$

d_{ik} being the distance between each data vector $\mathbf{x}_k$ and a prototype $\mathbf{v}_i$:

$$d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\| \qquad (8)$$

Notation of $\|\bullet\|$ refers to a norm, in particular it is assumed as L_2 (Euclidean norm); $m \in (1,+\infty)$ is the fuzzification coefficient.

This is the objective function of FCM algorithm [1] and this function guides the clustering process. Each term of J_m is proportional to $(d_{ik})^2$, so J_m is a squared error-clustering criterion. In a consequence, we can say that the problem of generating partitions of X was reduced to a finding in iterative manner the optimal partition matrix $\mathbf{U}$. In other words, find the minimal squared error clustering criterion. The algorithm is presented below (Table 5-1).

Table 0-1. The overall scheme of the Fuzzy C-Means (FCM).

I. Initialization:

Set c, $(2 \le c < N)$

Set m, $(1 < m < \infty)$;

Initialize randomly the partition matrix $\mathbf{U} \in \mathbf{M}_{fc}$.

Proceed with $l = 0,1,\ldots$

II. Calculation of prototypes:

Calculate the c prototypes $\{\mathbf{v}^{(l)}{}_{\mathrm{i}}\}$ *with* $\mathbf{U}^{(l)}$ *according to equation:*

$$\mathbf{v}_{\mathrm{i}} = \sum_{k=1}^{N} (u_{ik})^m \mathbf{x}_{\mathrm{k}} \Big/ \sum_{k=1}^{N} (u_{ik})^m \; \forall i \tag{9}$$

III. Calculations of partition matrix:

Calculate $\mathbf{U}^{(l)}$ *using* $\{\mathbf{v}^{(l)}{}_{\mathrm{i}}\}$ *if* $d_{ik} > 0$:

$$u_{ik} = 1 \Bigg/ \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)} \right] \tag{10}$$

Otherwise:

Set $u_{ik} = 0$ *and impose* $\sum_{i=1}^{c} u_{ik} = 1$, $\forall i \; 1 \le i \le c$ *that produces* $d_{ik} = 0$.

IV. Check the termination condition or if the maximal number of iterations is reached:

Compare in the convenient matrix norm if termination condition was reached $\left\|\mathbf{U}^{(l+1)} - \mathbf{U}^{(l)}\right\| \leq \varepsilon$ *where* ε *is a very small positive constant number*

Or

$l = \max\_iterations$ *holds true.*

FCM is a very good clustering algorithm for regular, distinguishable clusters of a similar size. It computes the cluster centre using all patterns from the cluster. This fact brings us the first disadvantage of this algorithm, FCM can be misguided in a noisy data environment. The second weak point of FCM is its sensitivity to the initial settings of the partition matrix. In other words, for different settings and the nature of the data, we can obtain different prototypes which means a different structure of clusters. FCM gives us an opportunity to control the "amount" of fuzziness in the partition matrix. This is obtained by the fuzzification parameter called m. If we increase the value of m the results will include larger amount of fuzziness in the partition matrix. On the other hand, for $m \to 1$ the membership values of a pattern are increased and favour a single cluster. The best value of m depends on the type of data but is frequently assumed by researchers to be $m = 2$.

5.3 Classifier fusion

There are numerous approaches of a combination of multiple classifiers methods in the literature. We can divide them into two main groups: classifier fusion and dynamic classifier selection. In case of classifier fusion, individual classifiers are applied in parallel and after that results of classification are combined in some manner to obtain a consensus within a classifier group. In case of dynamic classifier selection, we are trying to predict which classifier is most likely to be correct for a given sample and the final decision is based only on this classifier. Many different classifier fusion algorithms were developed. The main examples consist of: majority voting [16][13], unanimous consensus [13], thresholded voting [13], the Borda count [20], the average Bayes classifier [13], pooling methods which utilize heuristic decision rules [16][17], logistic regression to assign weights to the ranks produced by each classifier [20], methods of multiusage classification [11] and Dempster-Shafer theory to derive weights for each classifier's vote [13].

A method of partitioning the input samples is required for dynamic classifier selection. Partitions can be determined by a set of individual classifier decisions in which classifiers agree with each other. During this process, we use training or validation data to determine the "best" classifier for each partition [12]. During the classification process an unknown sample is assigned to a given partition and after that the output of the best fitted classifier for a given partition is used to make a decision.

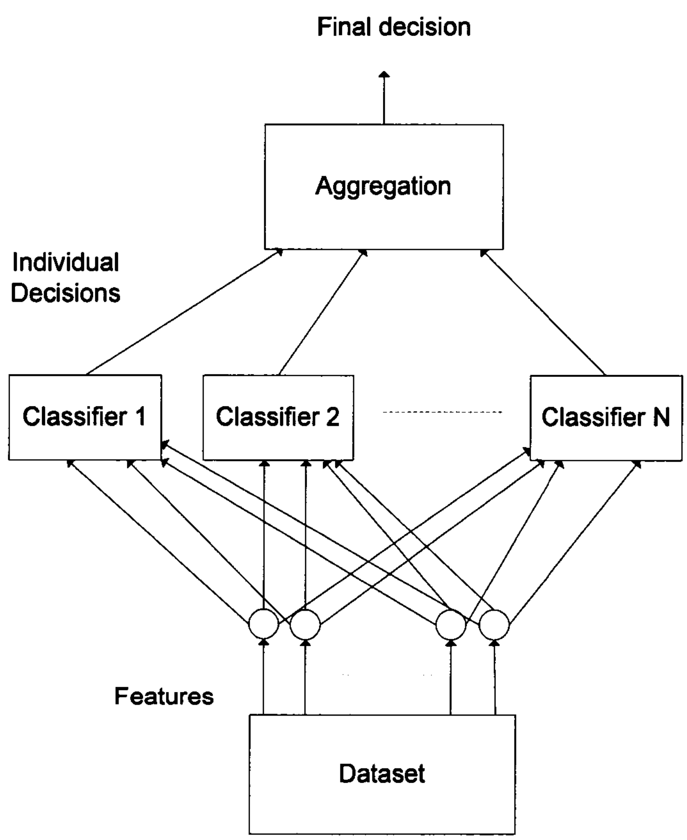

Figure 0-1. The idea of the classifier fusion

Majority voting is one of the most popular method of classifier fusion and one of the most easiest to implement. The idea of this algorithm is such that the individual classifiers "vote" for a given class, and the class with most votes is accepted as a final decision. If there is no winner, the combined classifiers make a decision based on the random assignment to the class. Let assume that we have $\boldsymbol{x}$ different classes. The output of D_i – classifier is denoted as $D_i(\boldsymbol{x}) = [d_{i,1}(\boldsymbol{x}), \ldots d_{i,c}(\boldsymbol{x})]$ where 'c' is a number of classes. The entry $d_{i,j}(x) \in [0,1]^c$ denotes the support that $\boldsymbol{x}$ may come from class ω_j. In order to find the classification decision we can create the maximum membership formula:

$$\text{Choose class } \omega_k \Leftrightarrow d_{i,k}(x) = \max_j \{d_{i,j}(x)\} \quad (11)$$

Based on this formula, we can create the hardened classification decision of each D_i as the binary vector D_i^h (h means "hardened") which contains 1 at position k and 0 elsewhere:

$$d_{i,j}^h(x) = \begin{cases} 1, & \text{if} \quad j = k \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

Finally, the majority vote aggregation can be defined as

$$F_{maj} = D(x) = [d_{1(x)}, \ldots, d_c(x)]^T, \ d_j(x) \in \{0,1\} \quad (13)$$

and

$$d_j(x) = \begin{cases} 1, \ \text{if} \ \sum_{i=1}^{L} d_{i,j}^h(x) = \max_k \sum_{i=1}^{L} d_{i,k}^h(x) \\ 0, \ \text{otherwise} \end{cases} \quad (14)$$

where L is a number of classifiers.

So, after aggregation of classifier results we have a binary vector with element 1 corresponding to the most supported class, and with 0 elsewhere.

In case of our experiments, we have a set of different MPEG-7 descriptors which are corresponding to different properties of the image. It means, that combining them using classifier fusion methods should provide interesting results in terms of image classification.

5.4 Conclusions

This chapter consists of detailed descriptions of algorithms used in our framework. At first, we described a creation of the reduced feature space. Next, we presented the idea of the fuzzy clustering and the clustering algorithm itself. Finally, we showed some classifier fusion methods which can be adopted in our system.

5.5 References

1. Bezdek J.C., "Pattern recognition with fuzzy objective function algorithms", Plenum Press, New York, 1981.
2. Bezdek, J.C., Hathaway, R.J., "A note on two clustering algorithms for relational network data", SPIE Vol. 1293 Applications of Artificial Intelligence VIII, 1990, pp. 268-277.
3. Bezdek, J.C., Hathaway, R.J., Windham, M.P., "Numerical comparison of the RFCM and AP algorithms for clustering relational data", Pattern Recognition, Vol. 24, No. 8, 1991, pp. 783-791.
4. Hathaway, R.J., Bezdek, J.C., "Clustering incomplete relational data using non-Euclidean relational fuzzy c-means algorithm", Pattern Recognition Letters 23, 2002, pp.151-160.
5. Hathaway, R.J., Bezdek, J.C., "Nerf c-means: Non-Euclidean relational fuzzy clustering", Pattern Recognition, Vol. 27, No. 3, 1994, pp. 429-437.
6. Hathaway R.J., Bezdek J.C., Davenport W.J, "On relational data versions of c-means algorithms", Pattern Recognition Letters, Vol. 17, 1996, pp. 607-612.
7. Ho T.K, Hull J.J., Srihari S.N., "Decision Combination in Multiple Classifier Systems", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, pp. 66-75, Jan 1994.

8. Huang Y.S, Suen C.Y., " A Method of Combining Multiple Classifiers – A Neural Network Approach", Proc. 12th Int'l Conf. Pattern Recognition and Computer Vision, pp. 473-475, Jerusalem 1994.
9. Huang Y.S, Suen C.Y., "A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, no. 1, pp. 90-94, 1995.
10. Kersten, P. R., "Fuzzy Order Statistics and Their Application to Fuzzy Clustering", IEEE Transactions on Fuzzy Systems, Vol. 7, No. 6, 1999.
11. Kersten, P. R., "Implementation issues in the fuzzy c-medians clustering algorithm", Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, Vol. 2, 1997.
12. Kersten, P. R., The Fuzzy Median and the Fuzzy MAD, Proceedings of ISUMA-NAFIPS, 1995.
13. Kimura F .,Shridar M., "Handwritten Numerical Recognition Based on Multiple Algorithms", Pattern Recognition, vol. 24, no. 10, pp. 969-983, 1991.
14. Lam L., Suen C.Y., "A Theoretical Analysis of the Application of Majority Voting to Pattern Recognition", Proc. 12th Int'l Conf. Pattern Recognition and Computer Vision, pp. 418-420, Jerusalem, 1994.
15. Loia V., Pedrycz W., Senatore S., "P-FCM: a proximity-based fuzzy clustering for user-centered web applications", International Journal of Approximate Reasoning, Vol. 34, Issues 2-3, 2003, pp. 121-144.
16. Mandler E., Schurmann J., "Combining the Classification Results of Independent Classifiers based on the Dempster-Shafer Theory of Evidence", Pattern Recognition and Artificial Intelligence , pp. 381-393, North Holland. Elsevier Science Publishers B.V., 1988.
17. Nadal C., Legault R., Suen C.Y., "Complementary Algorithms for the Recognition of Totally Unconstrained Handwritten Numerals", Proc. 10th Int'l Conf. Pattern Recognition, pp. 443 – 449, Atlantic City, N.J, 1990.

18. Pedrycz W., Loia V., Senatore S., "P-FCM: proximity—based fuzzy clustering", Fuzzy Sets and Systems, Vol. 148, Issue 1, 16, 2004, pp. 21-41.
19. Pedrycz W., Waletzky J., "Fuzzy clustering with partial supervision", IEEE Trans. on Systems, Man, and Cybernetics 5, 1997, pp. 787-795.
20. Xu L., Krzyzak A., Suen C.Y., "Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition", IEEE Trans. Systems, Man, and Cybernetics, vol. 22, no. 3, pp. 418-435, 1992.

6. Image classification system development

This chapter presents the framework of our classification system, all steps of the classification process and all functionalities provided by the system.

6.1 General framework of the system

Figure 6-1 outlines the general framework of our system and shows all steps in the image classification process. [1]

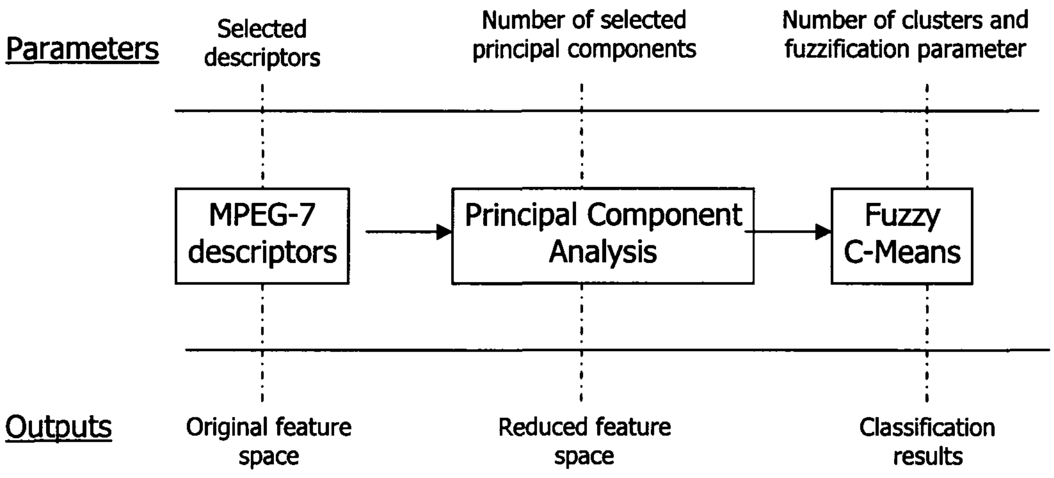

Figure 6-1. General framework

At the very beginning of our experiments, we extract MPEG-7 descriptors (we decided to use five of them) of all images from our image database. The content of these descriptors is related to colors of images, as well as, to their textures. After this process, the extracted descriptors create five separate feature spaces (Figure 6-2) [1]. Next, all other classification activities take place in these feature spaces.

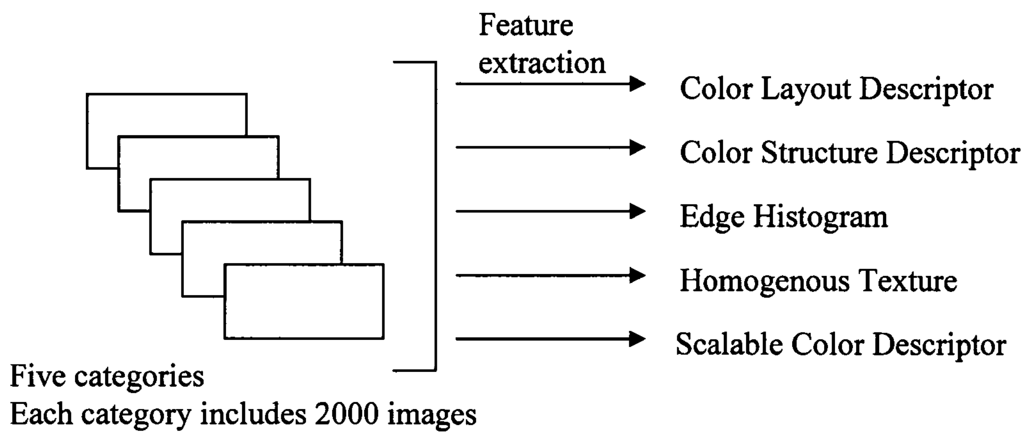

Figure 6-2. Creating of an original feature space

Our system enables us to choose between different sizes of descriptors. We can choose between all values mentioned in chapter 4.2.

Typically, as it was underlined previously, the dimensionality of such feature spaces is usually high and so their reduction is always strongly recommended. There are many techniques but in our experiments the reduction of the feature space is performed by the Principal Component Analysis (PCA). During the creating of the reduced feature space we have to decide how many dimensions it should have.

Further, image classification is performed with the use of the Fuzzy C-Means (FCM). This kind of clustering is carried out for each descriptor's feature space separately. During this process we can choose between different values of fuzzification parameter 'm', as well as, different number of created clusters.

6.2 Proposed classification system

Our proposed system is divided into several parts responsible for all steps of the classification process (Figure 6-1). All values of descriptors, all results of the PCA algorithm and all classification results are stored in the database. Because of that fact, we do not need to create hundreds of files. Everything we need is stored at one place and that fact gives us a simple access to required data. The user interface is created by using the Visual Studio .Net environment.

The part of the system responsible for extracting of MPEG-7 descriptors is based on MPEG-7 eXperimentation model (XM). MPEG-7 XM includes a set of classes which provide functions for extracting of descriptors. Figure 6-3 shows the main view of this part of the system. Here we can create a set of images which will be used in our experiments. The below is a list of main functionalities of this part:

- to add or remove individual images or folder of images to/from our database
- to set different extraction parameters as mentioned in chapter 4.2
- to browse included images and their previously extracted descriptors
- to decide if all added images or only a random number of them will be included in the experiment
- to start the process of descriptor extraction and after that add all extracted descriptors into the database

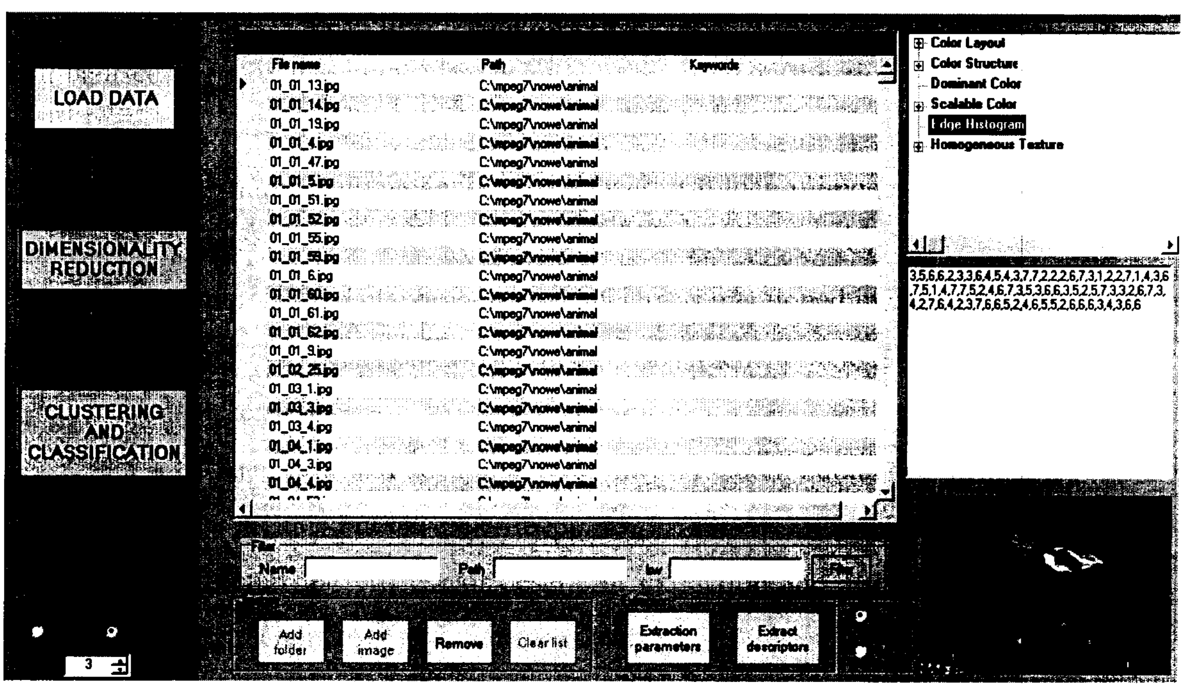

Figure 6-3. System functionalities – extraction of MPEG-7 descriptors

At this stage of the classification process, we have a database with all extracted descriptors.

The next part of our system is related to the creation of the reduced feature space. We can go to this section if a set of images was created and all chosen descriptors were extracted. The main view of this part is presented in Figure 6-4.

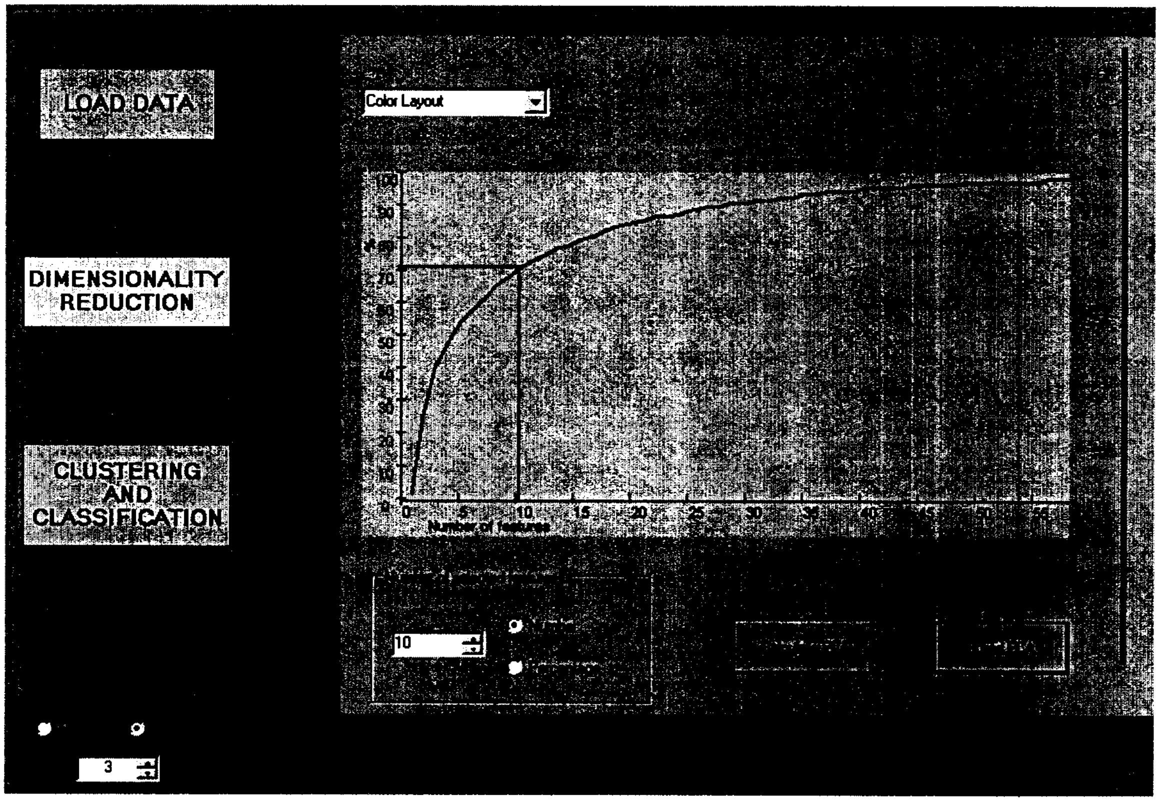

Figure 6-4. System functionalities – performing of PCA and feature selection

The main functionalities are:

- to load previously calculated eigenvectors and eigenvalues and visualize loaded results,
- to calculate PCA for every single feature space created by descriptors and insert results into the database,
- to set the number of eigenvectors that have to be included in the newly created feature space,
- to visualize results of PCA (importance of eigenvalues)

The final part of our system is related to the clustering and the classification process. Here we can perform all activities related to these processes as well as investigate obtained results. This section of our system is shown in Figure 6-5.

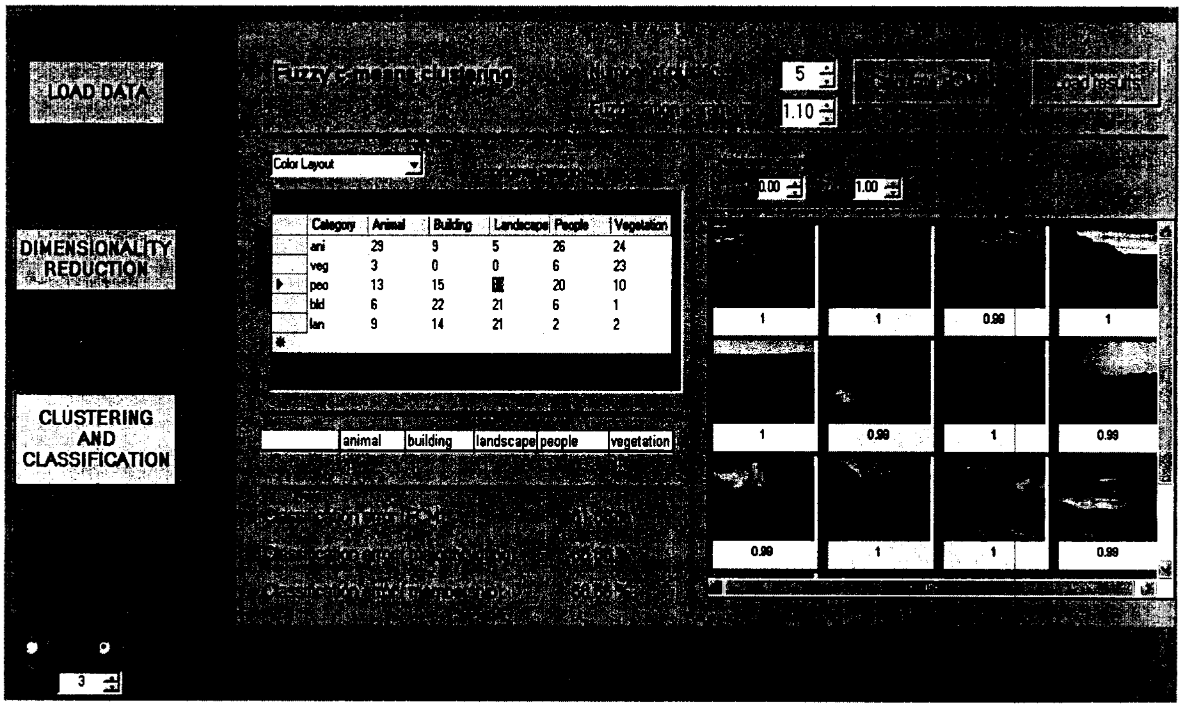

Figure 6-5. System functionalities – clustering and classification of images

In particular we can:

- to load previously made experiments,
- to set a number of clusters 'c' and set a fuzzification parameter 'm',
- to perform clustering and classification of images,
- to obtain the overall classification error for FCM, as well as, the classification error in case of classifier fusion
- to browse all images assigned to created clusters
- to browse memberships assigned to images,
- to see how many clusters are assigned to a given category

As we can see, our system provides many functionalities, which helps in automation of the classification task. We can choose between many options, browse our image database, and visualize results.

6.3 Conclusions

In this chapter our image classification and organization system was presented. At the very beginning, we outlined all steps of the classification process as well as all input and output parameters of every single step. Next, we demonstrated an implemented system, its interface with the user and all functionalities related to previously presented framework.

6.4 References

1. Kaczmarzyk T, Pedrycz W., "Content-based image retrieval: an application of MPEG-7 Standard and Fuzzy C-Means", 25th International Conference of the North American Fuzzy Information Processing Society (NAFIPS 2006) Proceedings, Montreal, June, 2006

7. Corel dataset experimentation

At the beginning of this chapter we present the experimental setup. Further, we demonstrate results of image classification. We deeply investigate the quality of the proposed classification system and give some examples of correctly and incorrectly classified images [1].

7.1 Experimental setup

It was pointed out that our system provides us possibility of choosing between different sizes of descriptors. In our experiments we use typical sizes of them. In particular, for Color Layout Descriptor – 58 bins, Color Structure Descriptor – 64 bins, Edge Histogram - 80 bins, Homogenous Texture - 62 bins, Scalable Color Descriptor – 64 bins.

Extracted descriptors are used to create five separate original feature spaces. Further, we have to decide how many dimensions a reduced feature space should have. In order to achieve optimal classification errors, we decide experimentally how many principal components should be taken to new feature spaces.

During the next step, FCM is carried out for each descriptor's feature space separately. Given a number of categories (classes) in the problem, the minimal number of clusters is equal to the number of classes yet a higher number of groups is also investigated. It arises a question if clusters related to all classes of images appear. To investigate this problem, we perform experiments also for a higher number of clusters. In other words, we find out which categories are the most complicated to classify and which of them dominates the rest.

We compare results of classification for the original and the reduced feature space using FCM and compare results of FCM to results achieved by the standard K-Means clustering. It gives us idea about accuracy of classification and tells us which method is better [1].

Additionally, we complete experiments for different values of the fuzzification parameter 'm' used in the objective function of the FCM algorithm (starting from m = 1.1). We investigate which values give us the best results in terms of classification.

We perform experiments separately for each descriptor. Our point of interest is to investigate which descriptors are the best choices in terms of image classification using Fuzzy c-means clustering algorithm.

7.2 The development of the reduced feature space

As mentioned earlier, the original feature space formed by the MPEG-7 descriptors is excessively high (between 58 and 80 dimensions) and comes with a significant level of redundancy [1]. Lower dimensionality usually gives us more general results and lower classification results. The PCA can be considered as a generic vehicle to complete its reduction. In the experiments, we performed the reduction of dimensionality separately for each descriptor feature space (Color Layout, Color Structure, Edge Histogram, Homogenous Texture and Scalable Color Descriptor). The plot of the level of retained variability versus the number of the most dominant eigenvalues for all used descriptors is shown in Figure 7-1.

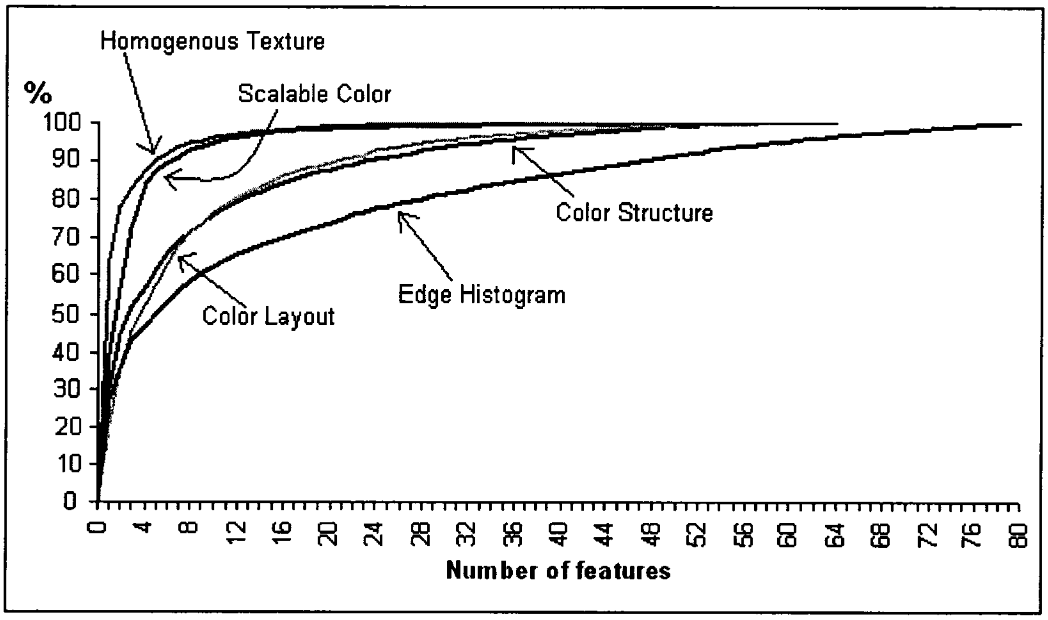

Figure 7-1. Variability (%) explained by the most significant eigenvalues

We note that Scalable Color and Homogenous Texture have a few eigenvalues whose importance is much higher that the importance of the rest of eigenvalues. As a result, we may anticipate that the reduction of the feature space in these cases could be the highest because there are some features which dominate over the rest of them. In case of Edge Histogram we can see that there are many important features so we may anticipate that the reduction will not be very efficient. Results for Color Layout and Color Structure are quite similar. In their case we can expect that the reduction of the feature space will be not as good as in case of Homogenous Texture and Scalable Color but should be significantly better than in case of Edge Histogram.

To find out an optimal number of principal components we performed a set of experiments [1]. Precisely, for the fixed number of clusters equal to the number of categories (so we have 5 clusters – each cluster may consist images of a given category) and two different fuzzification parameters (m = 1.1 and m = 2.0, to see how the algorithm works in less and more fuzzy case) we completed the FCM algorithm for a different number of principal components and investigate obtained results. The range of number of selected principal components is set from 2 to 20 and for every single experiment the overall classification error is calculated (see equation (1)). For m = 1.1 we can assume that the result of FCM is very close to those produced by the standard K-means (where membership function is limited to values 1 and 0). Next, we calculate the overall classification error for the originally formed five categories. Additionally, we calculate error for the three-class problem. Based on obtained results we choose the number of principal components, which corresponds to the lowest value of the overall error. The classification error is determined by calculating a difference of the total number of images in the dataset and images that are classified correctly, divided by the total number of images in the dataset:

$$Error = \frac{n_{all} - n_{correct}}{n_{all}} \quad (1)$$

The resulting classification error for the experiment for m = 1.1 is shown in Figure 7-2.

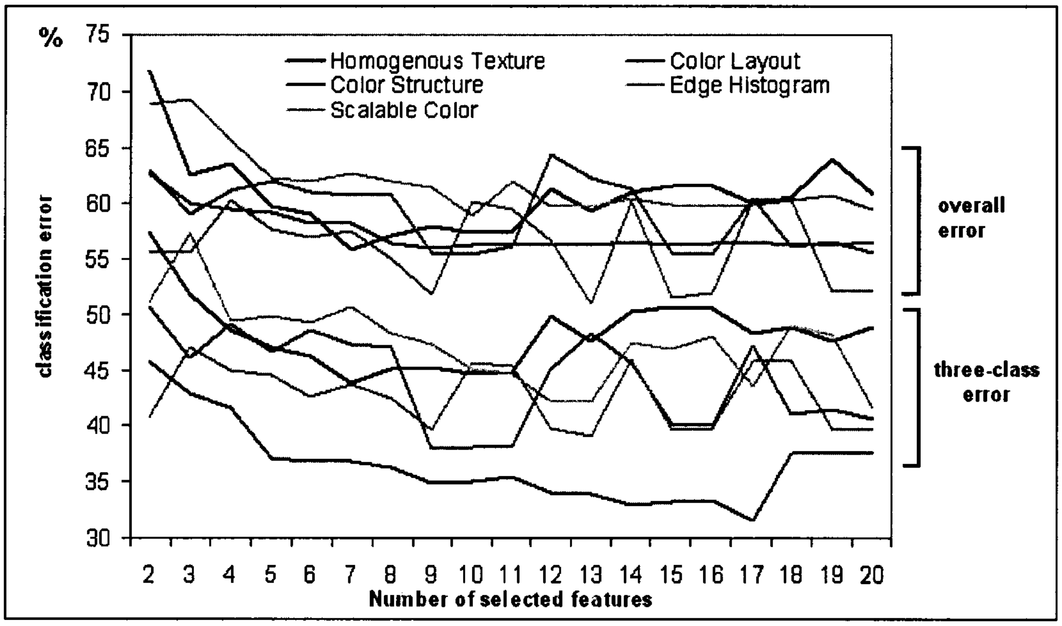

Figure 7-2. Classification errors for different numbers of selected features (five-class and three-class problem), five clusters and m = 1.1

The results of this experiment for m = 2.0 are shown on Figure 7-3.

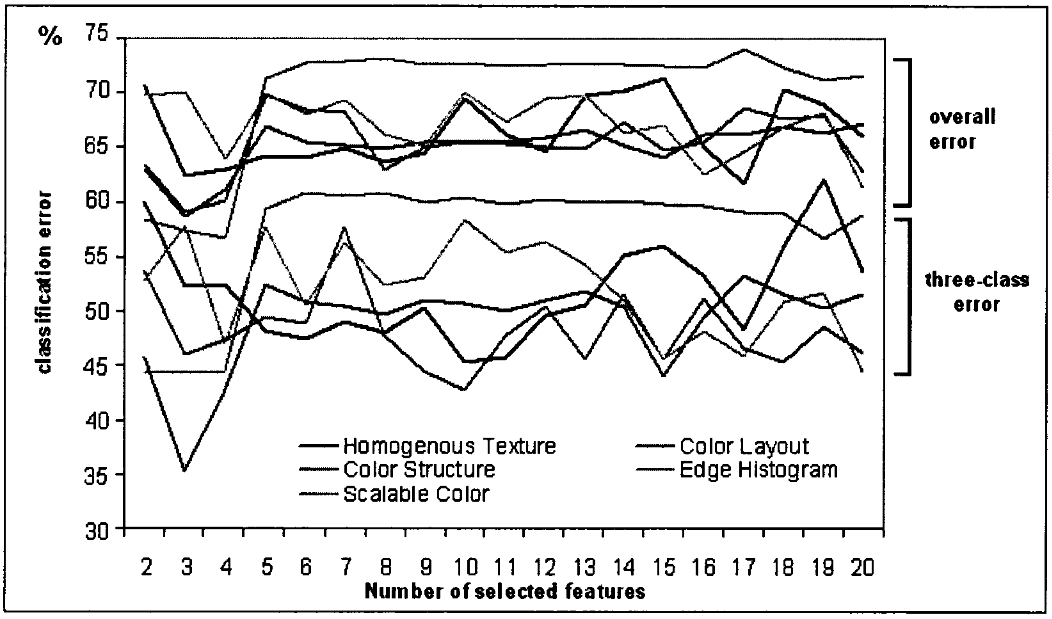

Figure 7-3. Classification errors for different numbers of selected features (five-class and three-class problem), five clusters and m = 2.0

We can easily note that far better results are achieved for m = 1.1 [1]. It means that with increasing of fuzziness, obtained classification results are getting worse, classification error is going up. This indicates the choice of the value of the fuzzification coefficient is crucial and FCM works much better when algorithm is set to be close to a standard K-means. In general, value of m has to be investigated when building a classifier.

Taking into account the above result of experiments we can find out best possible numbers of selected principal components. Our intent is to select a number of components which gives rise to the lowest classification error. On the other hand our demand is to create a feature space which is as small as possible. The optimal numbers of the selected features are shown in Table 7-1.

Table 7-1. The reduction of the original feature space – the optimal numbers of selected features.

	# of features in original feature space	# of features in reduced feature space
Color Layout	58	9
Color Structure	64	9
Edge Histogram	80	13
Homogenous Texture	62	7
Scalable Color	64	10

As we can see the reduction of the feature space is quite significant. The best result is obtained for Homogenous Texture descriptor. In this case we take only 7 eigenvectors into the reduced feature space compared to an original number of 62 dimensions We may expect this based on Figure 7-1. For Edge Histogram descriptor the reduced feature space has the highest dimensionality. We noted above that the original feature space for this descriptor has many equally important features and it is reflected in this result. In this case we reduced the original feature space (80 dimensions) into 13 dimensions.

7.3 Classification results

The next step of our experiments is related to the classification of images [1]. At this point we have a reduced features space and all experiments are performed on them. At the very beginning, we completed clustering and classification for the original feature space. Furthermore, we want to compare the quality of classification results obtained for the reduced feature space with classification results for the original feature space. Additionally, our point of interest is to compare these results with those obtained for the standard K-means. Table 7-2. summarizes the classification errors for the optimal numbers of features obtained in previously.

Table 7-2. Classification errors for all descriptors for original and reduced feature space (five-class and three-class problem)

	Overall classification error		Three-class classification error	
	Original feature space (FCM)	Reduced feature space (FCM)	Original feature space (FCM)	Reduced feature space (FCM)
Color Layout	63.5%	55.99%	41.73%	34.93%
Color Structure	60.1%	55.64%	45.33%	37.99%
Edge Histogram	57.51%	51.05%	42.76%	39.11%
Homogenous Texture	71.54%	55.85%	55%	43.77%
Scalable Color	60.88%	58.82%	43.43%	45.15%

Presented above results show that in most cases the classification is much better when dealing with the reduced feature space. In case of Color Layout, Color Structure, Edge Histogram and Homogenous Texture we obtained better results in overall classification error, as well as, in the three-class error. The exception is the Scalable Color Descriptor. For this descriptor the overall classification error is better but the three-class classification error is a little higher in the case of the reduced feature space. In general, as we expected the reduction of the feature space gives us more promising results.

Next, we compare these results with the results obtained with the use of the standard K-means (Table 7-3)

Table 7-3. Comparison of classification error using FCM (m = 1.1) and K-means clustering on the original feature space

	Overall classification error		Three-class classification error	
	FCM	K-means	FCM	K-means
Color Layout	63.5%	62.44%	41.73%	38.51%
Color Structure	60.1%	60.47%	45.33%	45.06%
Edge Histogram	57.51%	58.10%	42.76%	42.80%
Homogenous Texture	71.54%	71.42%	55%	54.98%
Scalable Color	60.88%	61.11%	43.43%	43.26%

In general, these results point out that the results obtained by these two methods are very close to each other. We may expect these results because for m = 1.1 FCM works almost like a standard K-means algorithm. It means that the membership values are usually very close to 0 and to 1.

In the next step of our research, we completed a series of experiments considering the reduced feature spaces and varying the values of the fuzzification parameter 'm'. The range of m is between 1.1 (close to a standard K-means) and 2 (much more fuzziness in memberships). The results are shown in Figure 7-4.

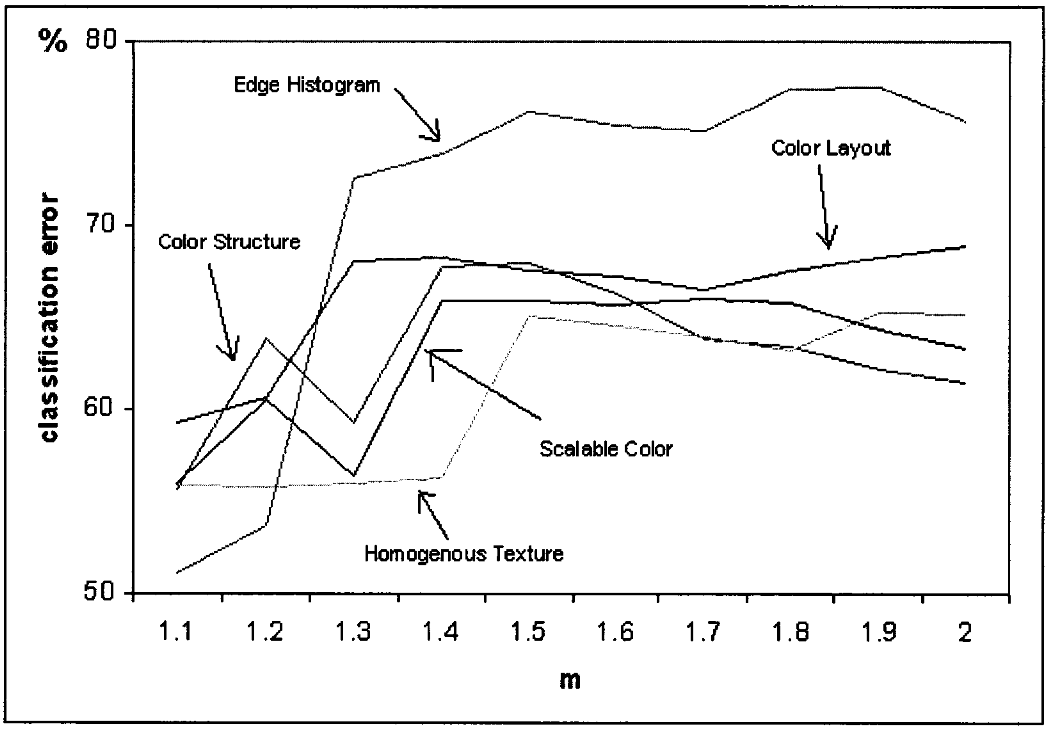

Figure 7-4. The overall classification error regarded as a function of the fuzzification parameter 'm'

It is significant, that the lowest classification errors are reported for lower values of "m". With increasing of the value of m, the classification error increases radically (for example from about 50% to about 80% in the worst case). The most sensitive to changes of 'm' is Edge Histogram descriptor. The rest of descriptors show similar behaviour in this case. We noted this fact previously but this experimental evidence leads us one more time to consider the fuzzification coefficient as an important design parameter. Furthermore, we have to stress that the values of m tend to be far lower than those we typically encounter in the literature. Usually researchers use value of m equal to 2.0 but as we can see in case of this dataset the best results are obtained for much lower value.

Additionally, we investigated how fuzzy are membership values taken from the membership matrices. The equation (2) is used to measure the fuzziness of membership values of a given sample 'k':

$$\psi = 1 - c^{c}\prod_{i=1}^{c} u_{ik} \qquad (2)$$

Basically, if the result of this equation is close to 1, it means that membership values are very close to 1 or 0. In other words, the result of classification may be very similar to the result obtained by the K-means algorithm. On the other hand, when the result of this equation is close to 0, it means that membership values of a given image are very close to 1/c where 'c' is a number of clusters.

The result of this experiment is shown in Figure 7-5.

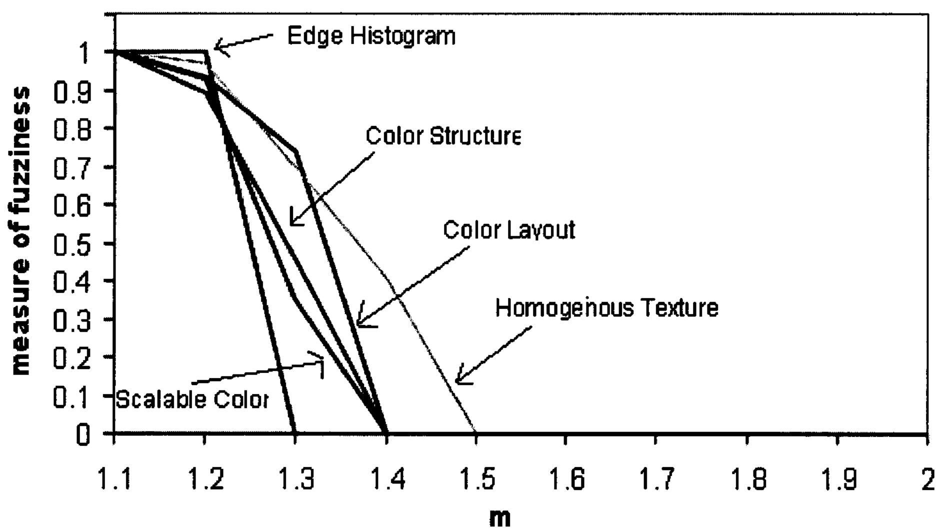

Figure 7-5. The overall classification error regarded as a function of the fuzzification parameter 'm'

Above results present clearly that some descriptors show more fuzziness in their membership matrices. For example, in case of Homogenous Texture we have to increase value of 'm' up to 1.5 to achieve fuzziness ratio which is close to 1/c. On the other hand, in case of Edge Histogram this ratio is obtained for m = 1.3. For Edge Histogram only a small increasing of 'm' (from 1.2 to 1.3) causes a big change in the membership matrix.

7.4 Categories and clusters

Different categories could exhibit different levels of complexity and typically there might not be a one-to-one correspondence between clusters and categories (classes). Some categories, especially those of higher diversity may require more clusters to represent them. By investigating the cluster-category correspondence, we can learn about the character of the categories. In our dataset, this leads to several interesting findings. We completed experiments for different number of clusters. The number of clusters is set from 5 to 20. Figures from 7-6 to 7-10 present the distribution of clusters assigned to categories for different overall number of clusters for all five descriptors.

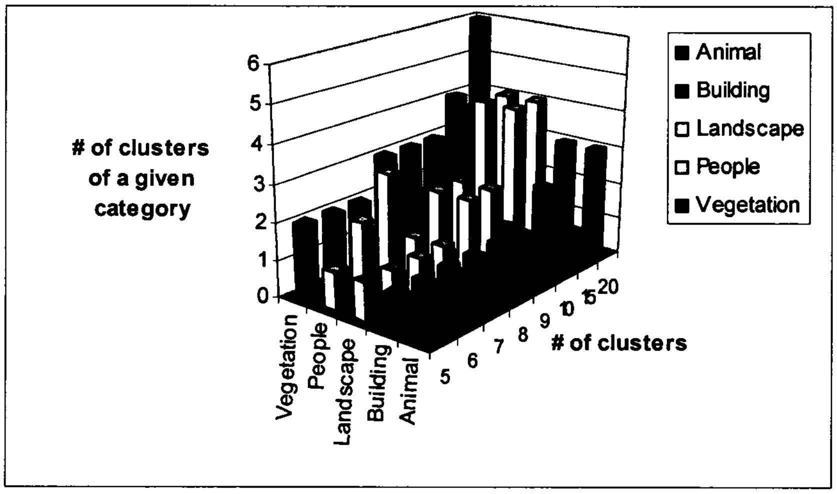

Figure 7-6. The relation between the overall number of clusters and the number of clusters assigned to categories for Color Layout Descriptor

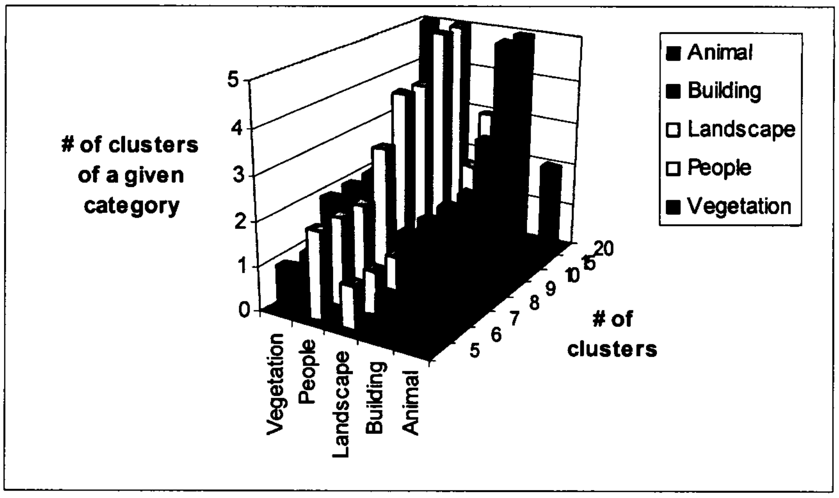

Figure 7-7. The relation between the overall number of clusters and the number of clusters assigned to categories for Color Structure Descriptor

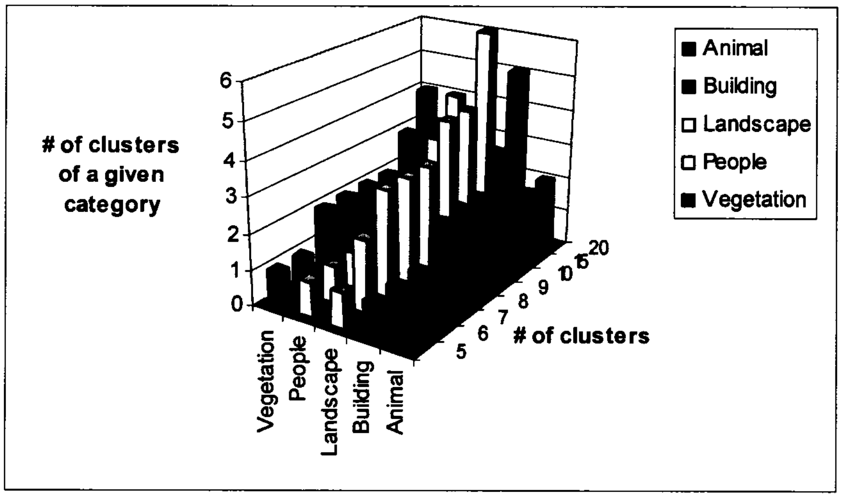

Figure 7-8. The relation between the overall number of clusters and the number of clusters assigned to categories for Edge Histogram

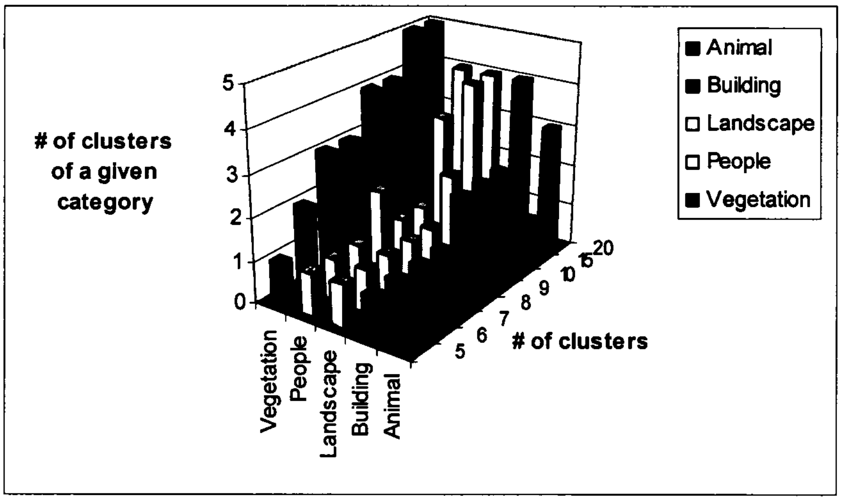

Figure 7-9. The relation between the overall number of clusters and the number of clusters assigned to categories for Homogenous Texture

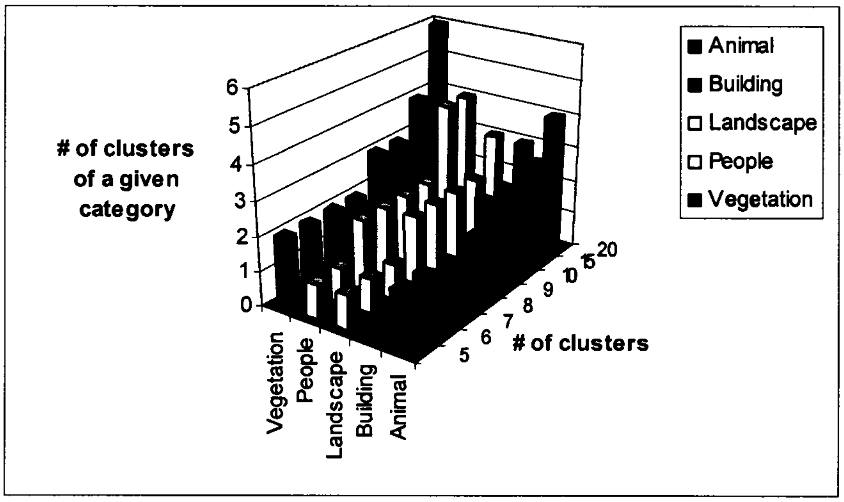

Figure 7-10. The relation between the overall number of clusters and the number of clusters assigned to categories for Scalable Color Descriptor

We clearly see that in some cases it is quite tough to create clusters reflective of some categories. For example, clusters representing the category of animals show only if the number of clusters gets higher. Otherwise, for a small number of clusters, animals do not create a separate cluster but become distributed among the remaining clusters. Usually, they are assigned to the clusters that are representative of people or vegetation. It means that this category is really hard to classify. Images of this category could be interpreted in many different ways. On the other hand, the clusters of vegetation images are very easy to build. We have also found that the number of correctly classified images from vegetation category is the highest.

Above results show that the distribution of clusters strongly depends on the choice of the descriptors. In general, when we look at results of classification of a given image for various descriptors, it is noticeable that a given image can be classified differently when dealing with various descriptors. MPEG-7 descriptors emphasize different aspects of the image content and this fact affects the distribution of clusters and classification results. In case of Edge Histogram and Homogenous Texture even if we set a number of clusters to 5 we get clusters of every single category. On the other hand for Color Layout Descriptor we have to set a number of clusters up to 20 to see that there are some clusters dominated by animal category.

7.5 Voting classification results

In this section we investigate the accuracy of two kinds of classifier fusion approaches. As it was noted, classifier fusion methods may improve the accuracy of classification. Taking into account that we have five different MPEG-7 descriptors, we can assume that their combination may give us significant improvements in terms of image classification. In general, we can treat the result of every single descriptor as a single classifier and then combine these results together. Additionally, we know that some descriptors may provide better results for some categories (for example descriptors related to texture usually work better in classification of images of buildings and landscapes). Bearing this fact in mind, we can assume that the combination of different descriptors may provide us some interesting results.

In our experiments we use two kinds of voting. The first approach is called majority voting. In this case every single descriptor vote for a given category. As a result we have a set of votes and based on these votes we decide which category is chosen. In the second case, we take into account membership values of every single image. More precisely, for a given image we create five sums related to five categories. Each sum consists of membership values related to clusters of a given category taken from all descriptors. Further, we choose the category with the highest sum. We performed experiments for a different number of clusters (from 5 to 15). For all descriptors the fuzzification parameter m is set to 1.2 to provide a significant amount of fuzziness in membership matrices (refer to Figure 7-5).

In Figure 7-11 we show obtained results and compare them with the classification error for a single descriptor.

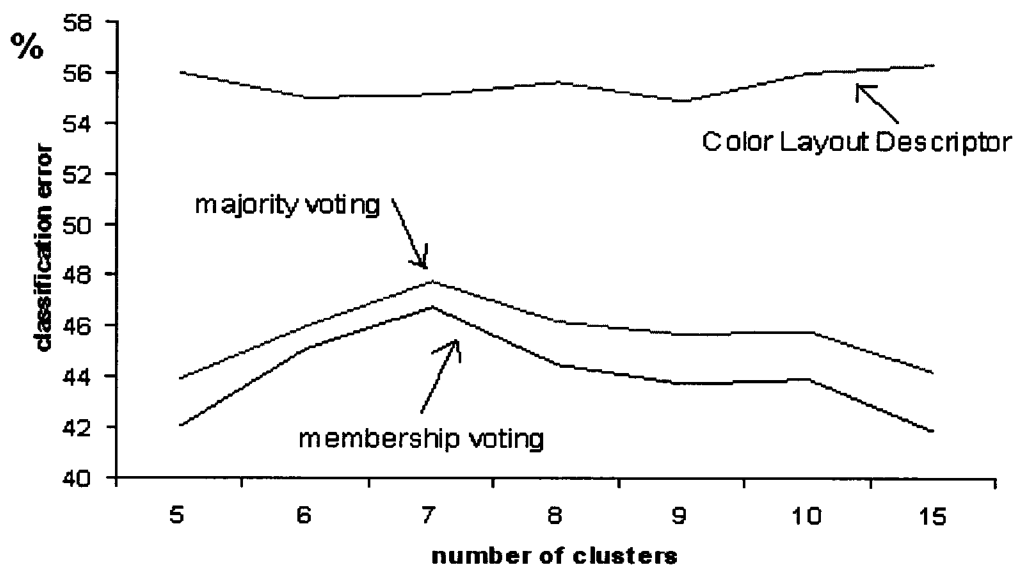

Figure 7-11. Classification errors for membership voting and majority voting

Above results show that the classification accuracy is much better in both cases of classifier fusion. Compared to the accuracy of a single descriptor (Color Layout Descriptor) the difference is more than 10%. Additionally, results obtained by the membership voting method are up to 2% better that results obtained by the majority voting algorithm. It means that the usage of membership values rather than votes is a better idea in this case. It is worth stressing, that with increasing the number of clusters, results do not show improvements.

7.6 *Examples of classification*

We know that image classification is a very complex process. It was noted before, that some images can be assigned to different categories by different people. It strongly depends on their preferences. It means that some images can be assigned to more than a single category. We may clearly see differences between the users in the assignment of the images to categories.

At the very beginning let us refer to some correctly classified images. Figures from 7-12 to 7-16 show examples of correctly classified images.

Figure 7-12. Correctly classified images from Animal category

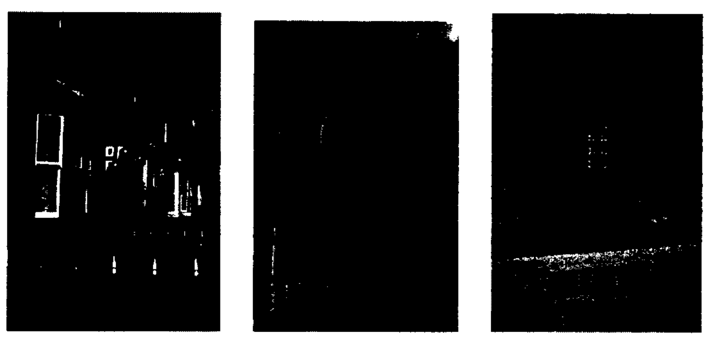

Figure 7-13. Correctly classified images from Building category

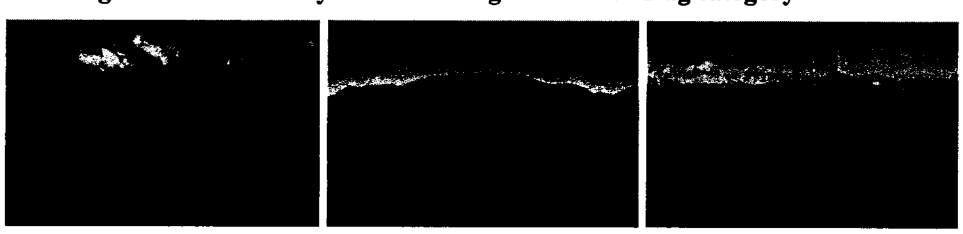

Figure 7-14. Correctly classified images from Landscape category

Figure 7-15. Correctly classified images from People category

Figure 7-16. Correctly classified images from Animal category

Above examples prove that our system can perform an image classification quite well. When we look at these pictures we can note that even pictures which fall into the same category show a large diversity of shapes and colors.

More interesting part of our results is related to overlaps between categories. In order to show that there is a substantial overlap between the classes and some images can be assigned differently, let us refer to some examples.

Figure 7-17 includes images which are assigned to the animal category but they should fall into another one.

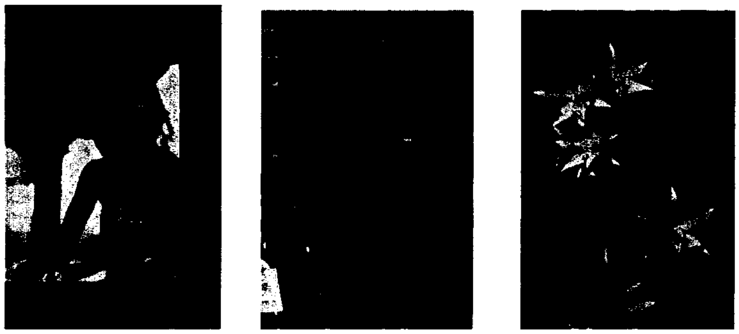

Figure 7-17. Images incorrectly classified as animals

These examples gives us idea about misclassified images which fall into the animal category. In general, most of these images come from the people and the vegetation category. We may anticipate that fact because the overlap between these categories is quite significant. Animals are usually surrounded by vegetation. On the other hand, some pictures of people are quite similar to the pictures of animals.

Figure 7-18 includes images which are assigned to the building category but they should fall into another one.

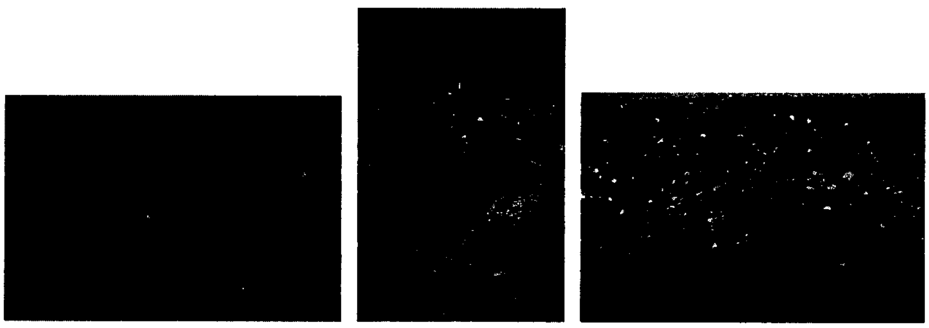

Figure 7-18. Images incorrectly classified as buildings

In this case we can easily see why these images were misclassified. They include some textures which can be recognized as some kind of building. We observe that usually misclassified images come from the landscape category. Landscape and building categories have many commonalities related to their textures. Additionally, as mentioned before, pictures of buildings frequently have backgrounds which can be interpreted as landscape. That is the main problem in this case. It is clearly visible in Figure 7-19.

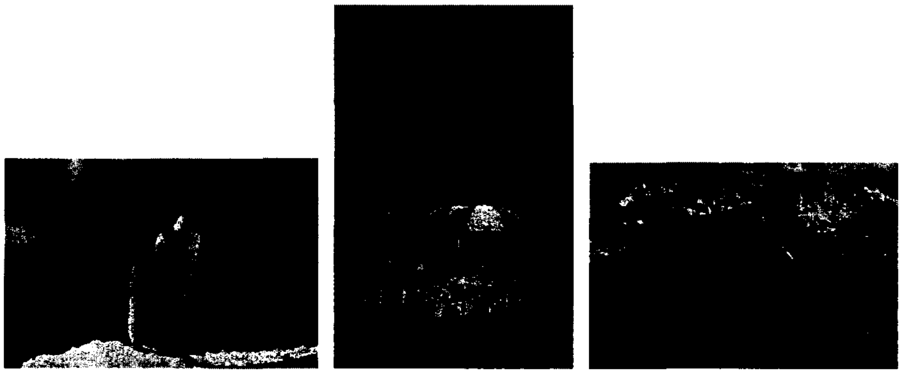

Figure 7-19. Images incorrectly classified as landscape

In case of the landscape category we deal with the same problem. The examples evidently show that even if the main part of the picture is the building, its background may be a reason to classify the image into the landscape category. As previously, most of misclassified images come from the building category.

Figure 7-20 includes images which are assigned to the people category but they should fall into another one.

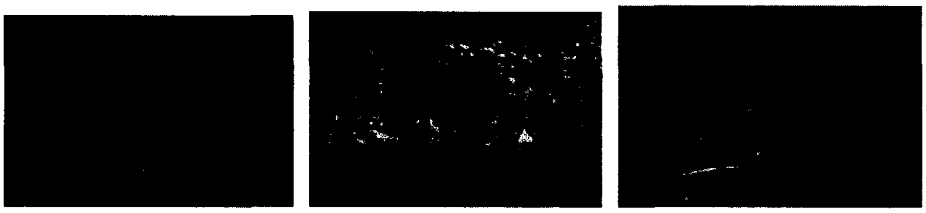

Figure 7-20. Images incorrectly classified as a person

Our results obviously show that usually misclassification in this case is caused by images of animals. When looking at above examples, we can say that they include many commonalities with pictures of people. On the other hand, pictures of people are frequently classified as animals. It means that these two categories are quite hard to distinguish and their classification is a very tough task to perform.

Finally, Figure 7-21 presents images which are misclassified as vegetation.

Figure 7-21. Images incorrectly classified as vegetation

These misclassified images include colors that can be interpreted by our classification system as some kind of vegetation. This may be the main reason why they are not classified properly. We can notice, that these images come from different categories but in general we observed that the misclassification comes from the animal and from the people category.

Let us refer to some images which are classified to different categories by different descriptors. Figure 7-22 presents three examples of them.

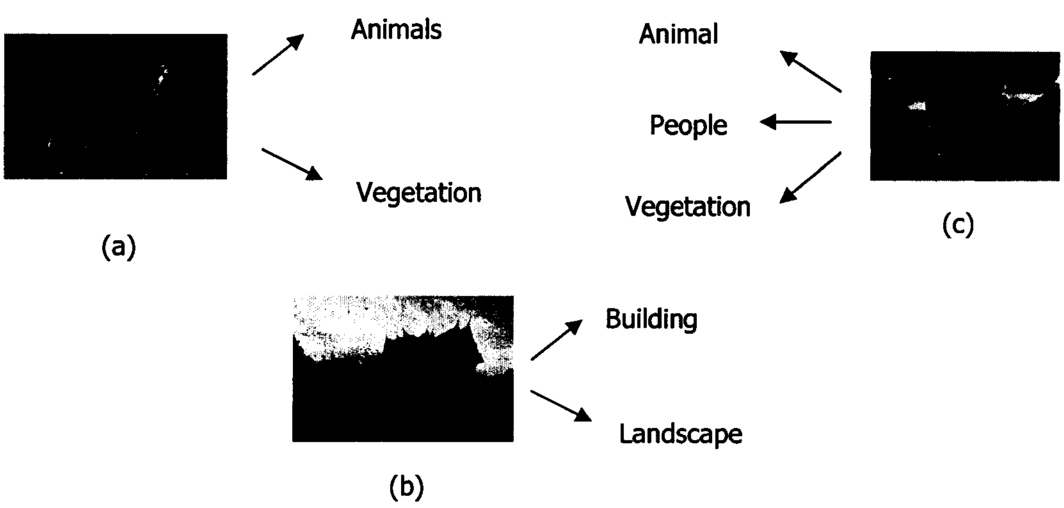

Figure 7-22. Examples of classification – overlapped categories.

When we are looking at the photo (a) in Figure 7-22 , we can say that the main topic of this image is the animal. Three descriptors classified this image as animal, but two of them classified it as belonging to the class of vegetation. The animal is surrounded by vegetation so the classification of this image to the vegetation category is quite reasonable. The photo (b) is assigned to the category of buildings however its background (sky) is characteristic for landscape images. So, we can anticipate that results will not be unequivocal. In this case, three descriptors assigned it to the landscape class while the two others pointed at it as a building.

To highlight other difficult case, consider image (c) from Figure 7-22. This photo was labeled to the three categories (animal, people, and vegetation).

In general, the results of our experiments show that some types images are classified appropriately only when we use a particular type of the descriptor. The correctness of classification for landscapes and buildings is much better when dealing with the texture descriptors (Edge Histogram and Homogenous Texture). The descriptors related to colors offer better classification accuracy for the classes of animals, people, and vegetation.

7.7 References

1. Kaczmarzyk T, Pedrycz W., “Content-based image retrieval: an application of MPEG-7 Standard and Fuzzy C-Means”, 25th International Conference of the North American Fuzzy Information Processing Society (NAFIPS 2006) Proceedings, Montreal, June, 2006

8. Conclusions

Our ultimate objective was to study the effectiveness and performance of fuzzy clustering (FCM) and MPEG-7 feature space in the problems of organization and classification of digital images. We performed a set of experiments and investigated their results.

A number of conclusions can be drawn from this study:

- The experimental study demonstrated that the reduction of an original MPEG-7 feature space significantly increases the accuracy of classification. As it follows from all the experiments, almost all classification errors are smaller compared to those obtained when using the original feature space. More precisely, in case of Color Layout, Color Structure, Edge Histogram and Homogenous Texture classification errors were reduced up to 16%. The only exception is the Scalable Color Descriptor. For this descriptor the three-class classification error was a little higher in the case of the reduced feature space. It shows that clustering algorithms used in the reduced feature spaces are more effective.
- We observed that the accuracy of classification strongly depends on the values of the fuzzification parameter "m". The best results are reported for values of "m" close to 1. All MPEG-7 descriptors used in our experiments are very sensitive to changes of 'm'. These experiments lead us to consider the fuzzification coefficient as an important design parameter. It is worth stressing that the values of "m" tend to be far lower than those we typically encounter in the literature (m=2.0).
- We contrasted the results of classification with those obtained for the standard K-means. This experiment points out that the results obtained by FCM (m = 1.1) and K-means are very close to each other.
- We pointed at several interesting properties of fuzzy clustering that are of interest in this application. In particular, the categories are not clearly

delineated. Fuzzy clustering offers a quantification of this effect. Images could belong to several categories and this effect is also captured by the proposed system. We presented several examples of correctly and incorrectly classified images and demonstrated that some images can be interpreted in many different ways.

- We investigated a correspondence between clusters and categories. The more heterogeneous (diversified) the category, the more clusters are needed to describe it. The study presented some quantitative findings regarding this: it was shown that the category of animals is quite difficult to capture and with the limited number of clusters, the images of this category are lumped into different categories. The distribution of clusters strongly depends on the choice of the descriptors. A given image can be classified differently when dealing with various descriptors. This strongly suggests to consider a fusion of several classifiers.
- The experiments showed that depending upon the specific feature space, the resulting clustering may produce different values of the classification error. We found that some categories of images are better classified by particular kinds of descriptors. Vegetation, people, and animal images are better recognized by color descriptors. On the other hand, texture descriptors are more suitable in case of building and landscape categories.
- We investigated the accuracy of classifier fusion. In both ways of classifier fusion (that is majority voting and membership voting), the classification errors were lower than the classification errors calculated for every descriptor separately. Our experiments show that usage of classifier fusion is appropriate but requires to be studied further.

The results presented here can serve as a practical guide for any other image classification system. Nevertheless, the concepts and approaches applied in this work might be beneficial and transferable for other methods used for classification of images.

We can envision several main directions this study could be easily expanded. The following are the areas worth exploring:

- Investigation of the performance of different categories of clustering algorithms on the collected image dataset.

  In our project we have used only the FCM clustering algorithm but there are many other methods of organizing objects into groups. For example, the Mercer kernel method could be easily introduced into the generic version of the FCM algorithm. Let us recall that kernels map implicitly the input data into the high-dimensional (or even infinite) feature space through some nonlinear transformation. An interesting aspect of the method lies in its abilities to handle outliers and noise immunity than FCM. A main drawback of these kernel clustering algorithms is that the clustering prototypes lie in the high dimensional feature space and hence there could be a lack of their clear and intuitive description.

  Additionally, relationships with neural networks could be also worth exploring.

- Investigation of the accuracy of classification for different quantization levels of image descriptors.

  The system developed in this study provides us with an interesting functional capability of choosing between different quantization levels of descriptors. While the experiments carried out so far explored a single level, it would be interesting to run experiments involving various quantization levels and assess the performance of classification as well as determine possible tradeoffs between accuracy and associated computing overhead.

- Thorough experimentation with a suite of schemes of classifier fusion.

  Our experiments have demonstrated that the usage of classifier fusion is legitimate and may significantly increase the accuracy of classification results. So far we investigated the performance of two selected fusion methods. As there far more alternatives, they are definitely worth investigating. . Some examples of the classifier fusion algorithms were presented in Chapter 5 and thorough studies of their performance could be a very interesting and useful component of future research.

An implementation of mechanisms that can adopt the system response based on some feedback from the user. Relevance feedback can reduce a gap between the typically low-level features extracted from images like color or texture and the high-level semantic features that a user can use in describing an image. There is a big number of techniques focused on exploiting such relevance feedback that have been proposed in the literature and we mentioned about some of them in chapter 3. Obviously as additional information can be captured from a user and properly used, the gap in retrieval semantic can be effectively reduced. In other words, such additional higher-level interactions of the user with the content-based image retrieval systems provide considerable valuable information for efficient image retrieval. During searching, the user should have a possibility of changing his expectations when he knows what he is looking for. Using techniques like Proximity Fuzzy C-Means can add some additional possibilities for the user during searching and categorizing process.

- Enlarge the number of image categories, as well as, the number of images. It may be interesting to increase the size of our image database and introduce new image categories. Our experiments showed that there are many overlaps between categories and this problem may be more deeply investigated. We presented that some categories are more complex than others and the next version of the classification system may use these observations in order to increase the accuracy of classification.