TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY
Julkaisu 556 • Publication 556

Iivari Kunttu

Shape and Gray Level Descriptors for Surface Defect Image Retrieval and Classification

Tampere 2005

Tampereen teknillinen yliopisto. Julkaisu 556
Tampere University of Technology. Publication 556

Iivari Kunttu

Shape and Gray Level Descriptors for Surface Defect Image Retrieval and Classification

Thesis for the degree of Doctor of Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 25th of November 2005, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2005

ISBN 952-15-1462-0 (printed)
ISBN 952-15-1497-3 (PDF)
ISSN 1459-2045

Preliminary assessors

Professor Björn Kruse
Linköping University
Sweden

Professor Juha Röning
University of Oulu
Finland

Opponents

Professor Josef Bigun
Halmstad University and Chalmers University of Technology
Sweden

Professor Juha Röning
University of Oulu
Finland

Custos

Professor Ari Visa
Tampere University of Technology
Finland

Abstract

Nowadays, the problem of image retrieval and classification plays an important role in the fields of image analysis and pattern recognition. With an increasing amount of real-world image data to be processed and stored, the development of powerful retrieval tools has also become a central problem in various industrial machine vision applications. The goal of finding similar objects from large and often distributed image collections is shared by the developers and users of machine vision systems. The focus of this thesis is on the field of surface defect imaging that has been applied to paper and metal manufacturing. Current surface inspection systems are capable of detecting various defects and producing gray level images of them. The defect images are collected into large image databases. Effective retrieval and classification methods are necessary to analyze the defects stored in the database.

The goal of this thesis is to present visual descriptors that characterize the defect shape and gray level distribution. The majority of the methods presented consider the shape using Fourier description of the boundary line of the defect. For this kind of shape description, novel Fourier-based approaches are presented. These approaches add a multiresolution property to conventional Fourier shape descriptors. This is achieved by combining discrete wavelet transform with discrete Fourier transform in shape description. Another approach to multiresolution shape description uses boundary smoothing combined with Fourier shape description. In addition, a method to combine defect boundary with defect's gray level information in Fourier description is presented. The gray level distribution of a defect image is described using binary co-occurrence matrix that outperforms commonly used second order statistical measures in defect image retrieval.

The proposed shape descriptors provide a significant improvement over the conventional Fourier shape description of the defects. The experimental results reveal that retrieval accuracy can be easily improved by using the proposed multiscale Fourier descriptors. The descriptors which use a combination of defect shape and gray level provide a novel method that is capable of improving retrieval performance without increasing descriptor dimensionality

Preface

The research presented in this thesis was carried out at the Institute of Signal Processing, Tampere University of Technology, Finland, during the period 2001-2005. The work formed part of the DIGGER project, which was jointly funded by industry and the Technology Development Centre of Finland.

I would like to express my thanks to my supervisor, Professor Ari Visa for his guidance and advice throughout this study. I am also grateful to the Institute of Signal Processing for the congenial working environment and its well-equipped facilities. In particular, I would like to extend my gratitude to Professor Moncef Gabbouj and Docent Alpo Värri for their generous help in the completion of this work. I would also like to thank Professor Josef Bigun for his valuable guidance and for the opportunity of working in his research group in Halmstad, Sweden in fall 2004.

The constructive comments of the reviewers, Professor Björn Kruse of Linköping University, Sweden, and Professor Juha Röning of the University of Oulu, Finland, are most gratefully appreciated. I am also indebted to Mr. Juhani Rauhamaa of ABB Oy, who has provided much valuable advice on issues concerning defect images and his assistance with the writing process of several publications. The language revision of the thesis was carried out by Mr. Alan Thompson, whom I would also like to thank.

My thanks are also due to the Technology Development Centre of Finland and ABB Oy, for their funding of this study. The financial support of the Foundation of the Promotion of Technology as well as the Jenny and Antti Wihuri Foundation is also gratefully acknowledged.

Finally, I wish to thank my beloved Leena for her unfailing support and encouragement throughout the duration of this research.

Tampere, October 2005

Iivari Kunttu

List of Abbreviations

AR	Autoregressive model
CBIR	Content-based image retrieval
CCD	Charge coupled device
CCV	Color coherence vector
CSS	Curvature scale space
DFT	Discrete Fourier transform
FD	Fourier descriptor
GLCM	Gray level co-occurrence matrix
k-NN	*k*-nearest neighbor classifier
WD	Wavelet descriptor
WWW	World Wide Web
ZMD	Zernike moment descriptor

Table of Contents

List of Publications

I. Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2003. Multiscale Fourier Descriptor for Shape Classification. In Proceedings of 12th International Conference on Image Analysis and Processing, Mantova, Italy, pp. 536-541.

II. Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2004. Multiscale Fourier Descriptor for Shape-Based Image Retrieval. In Proceedings of 17th International Conference on Pattern Recognition, Cambridge, UK, Vol. 2, pp. 765-768.

III. Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2005. Multiscale Fourier Descriptors for Defect Image Retrieval. Pattern Recognition Letters, to appear.

IV. Kunttu, I., Lepistö, L., Visa, A., 2005. An Efficient Fourier Shape Descriptor for Industrial Defect Images Using Wavelets. Optical Engineering 44(8), 080503.

V. Kunttu, I., Lepistö, L., Visa, A., 2005. Enhanced Fourier Shape Descriptor Using Zero-Padding. In Proceedings of 14th Scandinavian Conference on Image Analysis, Joensuu, Finland, LNCS 3540, pp. 892-900.

VI. Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2003. Binary Co-occurrence Matrix in Image Database Indexing, In Proceedings of 13th Scandinavian Conference on Image Analysis, Göteborg, Sweden, LNCS 2749, pp. 1090-1097.

VII. Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2005. Color Fourier Descriptor for Defect Image Retrieval. In Proceedings of 13th International Conference on Image Analysis and Processing, Cagliari, Italy, LNCS 3617, pp. 415-422.

1 Introduction

The rapid development of digital and information technologies has brought about the modern multimedia world. The amount of digital multimedia information available to ordinary people has increased enormously during the last decade. This development has been the result of the exponential growth of the Internet. In addition, different types of digital libraries and archives have made it possible to collect vast amounts of information in digital form. In addition to textual databases, multimedia data such as images, videos, or audio are nowadays often collected into digital archives. However, in order to make effective use of the available multimedia content, one needs to be able to find and locate the desired information from the huge variety of available information. The wealth of available digital information has given rise to a problem that is also shared by users of World Wide Web (WWW) as well as those seeking information in digital libraries. In consequence, the organization and search of multimedia content have become important issues.

Digital libraries and archives consist of one or several databases. The databases and the information stored in them have a great significance in several areas of current information technology. Different types of databases contain a wide variety of information that is part of everyone's daily life, such as personal, medical, financial, or consumer information. On the other hand, it is also common that people create their own databases containing multimedia content in digital form like photographs, music or videos. Another rapidly growing area is in industrial databases that contain information, such as measurement data, from different manufacturing processes.

Information retrieval means the storage, organization, retrieval and presentation of information (Baeza-Yates and Ribeiro-Neto, 1999). The basic requirement for a successful query is that the user defines the desired information in such a manner that the retrieval system is capable of understanding it. Hence, an essential problem with information retrieval is to define the information requirement and reply to it (Baeza-Yates and Ribeiro-Neto, 1999). A traditional approach is to use keywords that describe the desired information or document to be retrieved. This approach is still valid in the case of text documents such as books or journal articles, whose central content can be described using a few keywords. On the other hand, it is obvious that the whole content cannot be described using a limited number of keywords, which means that this information is omitted in the document search. Consequently, all available information cannot be retrieved from the database. This example illustrates one of the most essential problems for a retrieval system: how to find all the interesting information from the database. Another and equally important problem is how to limit the recalled information so that only the information on the user's interest is recalled.

1.1 Content-based retrieval in image databases

Images can be used in human communication to illustrate almost anything. Some typical areas can be, for example, enrichment of textual information by illustrations and images in writing, diagnosis and monitoring in health-care as well as different kinds of documentation, visual inspection, and recognition tasks. In recent years, the development of digital imaging tools has caused a rapid increase in image archives throughout the world and in several areas of modern life. These areas include education, entertainment, Internet, person identification, and industrial imaging solutions, to name just a few. Digital imaging has also meant that imaging costs have decreased significantly and this has caused strong growth in the sizes of the image databases. It is not unusual for current image archives to contain hundreds of thousands or even millions of images.

With the increase of digital image archives, the problem of image retrieval has become more pronounced. As a result, during the last decade, much research effort has been directed to this area. Early image retrieval applications used textual information i.e. keywords that were used to describe image content. The keywords had to be defined manually and this involves an enormous amount of work if the database is large. Another drawback with the keyword-based image retrieval is that the content description is always subjective. This means that the selection of keywords describing a particular image is dependent on the personal opinion of the person selecting the keywords. Moreover, it is often difficult to describe the visual content of an image using words.

Due to the serious drawbacks of keyword-based image retrieval systems, a new approach has been adopted to describe image content in retrieval applications. This approach employs visual features that are extracted from the image to describe its content. This is called image database indexing. In addition to images, this kind of indexing has been utilized in all kinds of multimedia, such as in videos and audio. Hence, the use of visual features makes it possible to describe effectively image content without the need for keywords. The most common visual features are the shapes, colors and textures occurring in the images. It is essential in using such a method that indexing can be performed automatically to make it suitable for large image databases. Furthermore, the use of automated indexing eliminates the problems caused by subjectivity in the indexing. The methods that use this approach are generally referred to as content-based image retrieval (CBIR) methods. The development of CBIR methods including visual features, image database indexing, similarity measures, semantic analysis, and system design have received much attention in the research community as well as in commercial applications. From the beginning of the 1990s a significant number of conference papers and journal articles as well as several books have been published on the topic, see (Gudivada and Raghavan, 1995; Gupta and Jain, 1997; Rui et al., 1997, 1999; Smeulders et al., 2000; Del Bimbo, 2001), for example.

In a CBIR system, the goal is to recall from the image databases those images that bear closest similarity to the query image provided by the user. As with information retrieval in general, the essential problem in image retrieval is to obtain a response to the user's query as accurately as possible. To fulfill this requirement, the recalled images should correspond to the query image in terms of their content. This is possible only if the selected features correspond to the visual content of the images. Selection of the features is, therefore, the essential problem in image retrieval. The retrieval is usually an on-line operation, and hence the features should be computationally inexpensive. This is particularly important in the case of large image databases. Successful indexing of an image database is a challenging task that is dependent on the nature of the database images. The research areas that are closely related to content-based image indexing are image analysis and pattern recognition.

Feature selection in image retrieval tasks is quite similar to that in image classification. This is due to fact that in content-based image classification the same features can often be utilized as in the case of retrieval. This means that the performance of particular visual features can be indicated and compared by means of simple image classification experiments.

1.2 Objectives of the thesis

Current image retrieval methods have been developed to make queries in different kinds of image archives. These archives may, for example, contain consumer photographs, medical images, satellite images, or newspaper images. This thesis presents the problem of image retrieval from an industrial viewpoint. Digital imaging tools are nowadays widely used in the process industry to monitor quality and production. Imaging tools can be used to acquire images of defects occurring in the production. Due to the speed of industrial processes such as those of a paper machine, the volume of acquired images is huge, and therefore effective retrieval methods are necessary for finding defects of a certain type from the database.

In defect image retrieval, as with image retrieval in general, shape, color and texture can be used to describe the image content. This thesis focuses on the defect image indexing and retrieval on the basis of shape and gray level features. The goal of this study is to develop effective visual features that are capable of describing the defect images in retrieval. The study shows that the defect image description can be significantly improved using the developed description methods.

This thesis concentrates on contour-based shape description and particularly on Fourier descriptors (FD). FD's are widely used and effective methods for shape description and they are well suited to on-line image retrieval because of their computational efficiency and compactness. Methods to enhance the retrieval performance of FD's are also presented. Another topic area of the thesis is defect image description using gray level distribution. For this purpose, a novel statistical measure, binary co-occurrence matrix, is presented. In addition to retrieval, the image classification aspect of the particular features is also considered. The image description methods developed in this research are such that they can be directly applied to defect image retrieval applications.

1.3 Organization of the thesis

The application area of the present study is defect image analysis. The defect images that are collected from industrial processes are stored in image databases, in which retrieval and classification operations are carried out. In Chapter 2 the area of defect imaging and the problems related to defect image recognition and classification are discussed. This discussion deals with paper and metal surface images obtained from industrial processes.

The principles of image retrieval and classification are covered in Chapter 3. The Chapter begins with a short historical review of image database indexing and content-based image retrieval methods. The principles of image retrieval using visual features are discussed in Section 3.1. The special characters of image retrieval and classification, as well as their differences are also discussed in this section. Section 3.2 considers the performance measurement of image retrieval and classification, including performance measures and validation methods.

Chapter 4 covers the area of shape description. The purpose of the Chapter is to give an overview of common shape description methods. The main focus is on the shape descriptors that are based on the boundary line of an object. The organization of the Chapter is based on the common taxonomy of shape description techniques.

In Chapter 5, the principle of Fourier descriptors is discussed. Section 5.1 presents the well-known discrete Fourier transform (DFT) and Section 5.2 discusses its application to boundary-based shape description.

Chapter 6 presents the subject of statistical gray level features. Beginning with general first and second order statistics, general statistical color description methods used in image retrieval are discussed.

Chapter 7 presents the contributions of this thesis in terms of publications dealing with the methods of defect image retrieval. The content of the publications is summarized and the author’s contributions are presented. Chapter 8 contains the conclusions to this thesis.

2 Industrial defect images

The use of machine vision applications is common in several areas of industry being utilized, for example, in quality and process control. Visual inspection systems are widely used in the process industry for inspecting goods and materials. Typical examples are the analysis of defects occurring in the paper manufacturing process (Rauhamaa and Reinius, 2002; Rauhamaa and Järvinen, 2002; Rauhamaa, 2004), the surface inspection of metals (Stolzenberg et al., 2003; Chen, 2004), wood inspection, (Niskanen, et al., 2001), and quality and defect control in the textile industry (Kumar, 2003), to name just a few. In this thesis, the focus is on the visual inspection of web materials, which include paper, metals, textiles, and plastics. In many cases the image material collected from a manufacturing process is stored in an image database. In the process industry these image databases are usually very large because of the huge number of images produced by inspection systems.

2.1 Surface inspection of web materials

This thesis examines two industrial visual inspection tasks, paper and metal defect analysis. Both defect types are related to the problem of surface inspection of web materials, which is a somewhat challenging problem. Iivarinen et al., (2004a) divide this problem to three parts, image acquisition, defect detection, and defect classification. In this thesis, the classification of the surface defects is considered, but the other elements of the problem are also briefly discussed here.

Paper defect imaging

There are various methods of quality control in use in paper mills. Laboratory tests have been the traditional way of precisely measuring the different chemical and physical properties of paper. These tests, however, are off-line operations and therefore cannot provide real-time feedback on paper properties. This has prompted the rapid growth in on-line scanning measurement systems that are capable of providing feedback on certain paper properties in real-time. Holes and other surface defects have been found to be the essential quality and productivity problems in the paper industry (Landry (2000)). This has led to the need for visual inspection systems

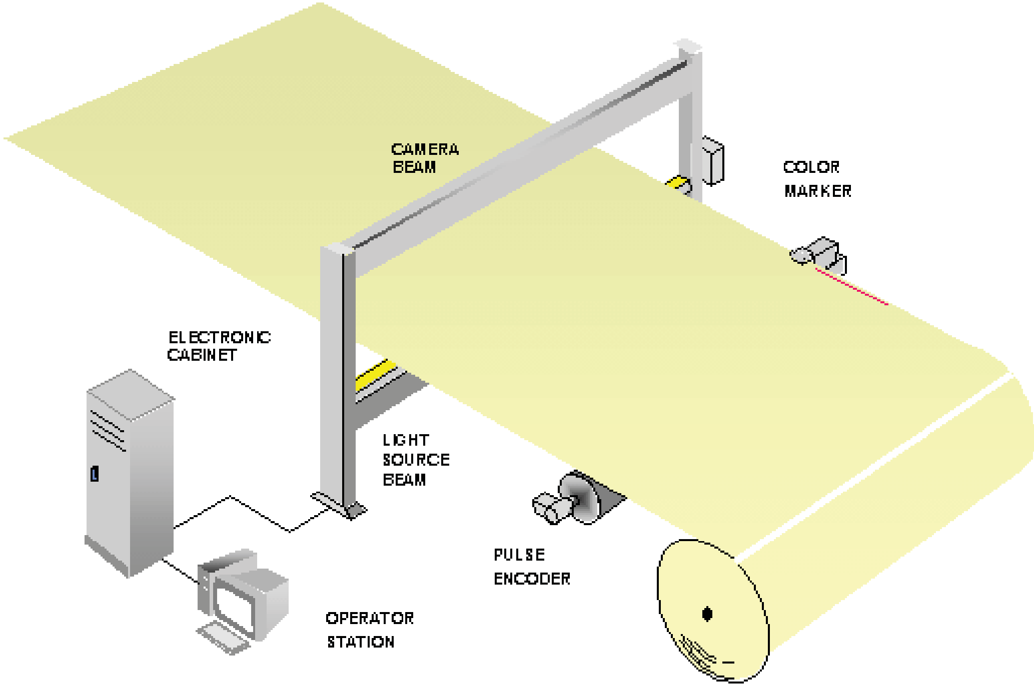

Figure 2.1. Paper web inspection system.

on the production line. In a typical paper inspection system the whole area of the paper web surface is inspected in real-time (Rauhamaa and Reinius, 2002). This is a challenging task due to the speed of the paper web, which may be up to 40m/s. The task of a paper surface inspection system is the detection of surface anomalies that deteriorate the quality of the final products or which are critical to the runnability and condition of production equipment (Iivarinen et al., 2004a). Holes, wrinkles or different kinds of thin or dirt spots are typical examples of paper surface defects.

Figure 2.1 shows the main parts of a modern paper surface inspection system. The light source beam provides strong illumination to the web area under inspection. The light is transmitted through the paper and the cameras above the web detect variations in its brightness. The transmission-based illumination (as in Figure 2.1) is suitable for paper grades of low opacity. With the paper grades of high opacity, the illumination is reflected so that both the camera beam and light source beams are on the same side of the web. The camera beam contains an array of CCD (charge-coupled device) cameras. The signal provided by CCD sensors can nowadays be converted to digital form in real-time. Therefore, the web imaging system is capable of producing high-quality gray level images of the defects. Additional devices include pulse encoder for speed measurement and color markers that are used to mark the defect locations.

The defect images obtained from a modern inspection system are like photographs with 256 gray levels. In early defect imaging methods, the defects were detected as simple holes or spots, with no precise information about their shape or structure (Rauhamaa and Reinius, 2002). Nowadays, due to advanced defect imaging, it is possible to use the shape and the internal structure of the defect as descriptors in defect recognition and classification.

The defect images are utilized in paper manufacturing process in several ways. They give immediate information to the process operators about the severity of the

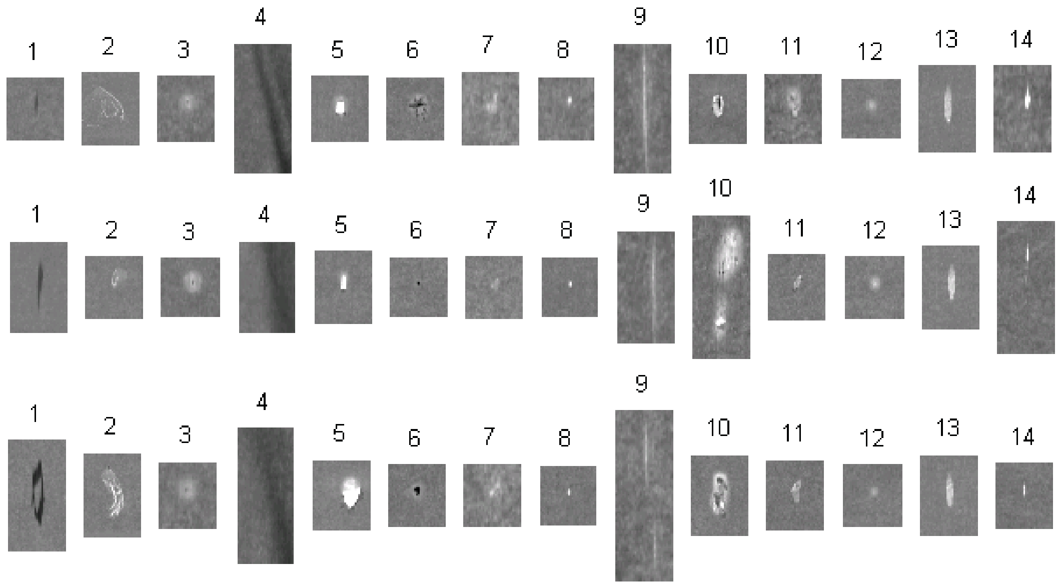

Figure 2.2. Examples of paper defect images. The number above the images indicates the class of the defect.

defects and, in many cases, their origin. On the basis of this information, the operators are able to make decisions about process control or possible maintenance shutdown. The images are also collected into databases, from which they can be later retrieved. This is particularly beneficial, because the occurrence of certain defect types can be analyzed afterwards by making queries in the database. It should be noted that a web inspection system may produce hundreds of even thousands of defect images in a day. Even though most defects do not necessitate immediate action, they are still stored into the database for future analysis tasks, making them very large sources of data.

There is a wide variety of different defects occurring in the paper web. Many defect types are common to all paper grades whereas some defects are specific to certain paper grades only. This is because different raw materials and equipment are used to manufacture different papers. For example, while making coated magazine paper, a coating layer is applied to the surface of the base paper. It is obvious that this process produces a different kind of defect from those arising in the manufacture of base paper. In addition, coating may even remove certain minor defects in the base paper by covering them over.

The defect images collected into the database vary greatly in size. The size of a paper defect may range between fractions of a millimeter to tens of centimeters. Figure 2.2 presents examples of typical paper surface defects. On the basis of manual inspection, these defects have been divided into 14 classes. The first two defect types are both caused by loose paper scrap. However, they are very dissimilar. Class 1 represents those cases where loose paper pieces are flying under a camera. In class 2 the paper scrap is run through a machine calender. Class 3 consists of dirt spots, but in this case, they are encircled by a wet collar. Sometimes most of a dirt spot may fall off, leaving a hole in the web, and the result will be a defect type represented by class 5, dirt holes. Class 4 contains images of paper fluttering, which is typically caused by a severe defect at the paper edge. Various effects in the wet end of a paper machine may cause disturbances in the web formation, and the results can be detected as thin areas or as a bad formation represented by class 7. The most probable cause of a clean hole (class 8) is most probable cause of a clean hole (class 8) is pitch or most probable

Figure 2.3. A surface inspection system monitoring metal strip quality.

cause of a clean hole (class 8) is pitch or some other material which plugs the wire in the forming area. The appearance of this defect type is quite clear. Wrinkles (class 9) are similar to web fluttering, but the folds are much narrower and sharper. Classes 10 and 11 are both the result of the same cause, i.e. by slime or also by slippery like mass of paper raw material that replaces fibers in the wet end. The result is a fairly weak area in the sheet. These two classes are almost the same in appearance; the only difference is that the defects in class 10 also contain at least one hole. It is often difficult to identify the causes of the small light spots in class 12 because there are many possible explanations for them, for example printing due to material sticking to a calender roll, or any random disturbance on the web, felt or roll. Class 13 contains marks caused by falling drops of water caused by condensation of water vapor in the hood of a drying section, for example. Defects are elliptical in shape and sometimes become so weak that holes can develop at the center. Oil spots (class 14) are similar to water spots, but smaller and more translucent.

Metal defect imaging

Another type of web material inspection considered here is related to the metal manufacturing process. Producers of metal materials use various methods to control the quality of their products. In the case of metal production, quality control is directed towards surface inspection (Chen, 2004; Henkemeyer, 2003, Stolzenberg et al., 2003). Traditionally, this has been carried out by human inspectors. However, the increasing speed of production lines has made it difficult to control such surface areas manually. Nowadays, machine vision systems are used to carry out surface inspection automatically. Current camera-based vision systems can detect anomalies occurring on metal surfaces, from small pimples to large area discoloration (Iivarinen et al., 2004b). Its speed, consistency, and reliability make the machine vision system a powerful inspection tool for metal strip manufacturing plants.

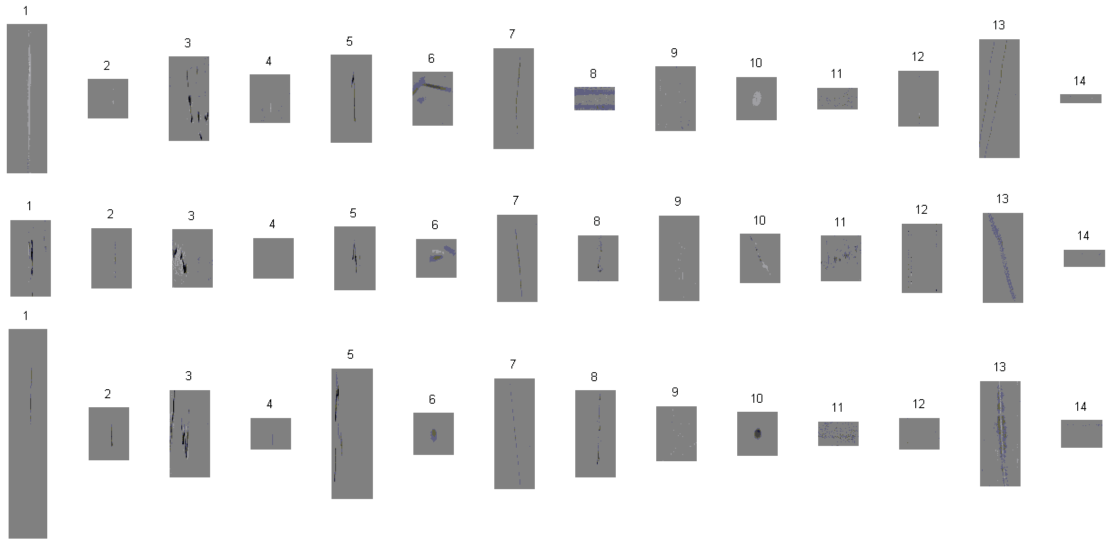

Figure 2.4. Examples of metal defect images. The number above the images indicates the class of the defect.

Figure 2.3 presents a metal surface inspection system installed on a strip handling line. There are two camera beams and a light source beam in the figure. The camera beam consists of several line-scan cameras utilizing CCD technology. The cameras are able to continuously monitor the whole width of the metal strip. Because certain defects are very small, the resolution of the imaging system needs to be of the order of fractions of a millimeter, both in the machine direction and the cross direction. As in the case of paper surface inspection, metal inspection systems also report defect information on the screen of the operator station. All detected defects are also stored into databases for future analysis.

There are dozens of defect types recognized in metal surface analysis. Common defect classes are shells, slag inclusions, scratches, roll imprints, pimples, indentations, flecks and holes (Iivarinen et al., 2004b). Figure 2.4 shows three examples of metal surface defects which are manually divided into 14 classes on the basis of defect cause.

2.2 Defect image analysis

Image segmentation

The defect images stored in the image databases need to be described using visual features to make possible their classification and retrieval. Certain texture and gray level features can be extracted from the defect images in a straightforward way, but in order to obtain shape representation, image segmentation phase is necessary. The defective area in the image needs to be separated from its background by a boundary line. Defect image segmentation is a challenging task because it is often difficult to distinguish the border outline of the defect from its background. There are several approaches to defect image segmentation (Rauhamaa and Reinius, 2002). The easiest approach is to draw a rectangle around the defect in a thresholded image. By means of this approach, the simplest measurements such as length, width, aspect ratio, and area of the defect can be defined. Such an approach, however, cannot be used for shape

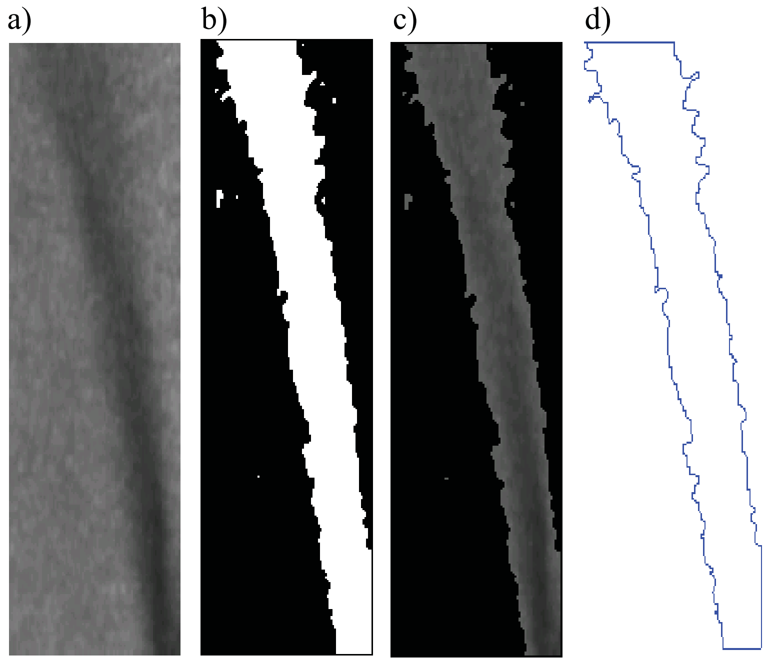

Figure 2.5. a) Paper defect image, b) segmentation mask, c) defect extracted from the background, d) boundary line.

description. Therefore, the segmentation line should correspond to the actual defect boundary. There have been a number of image segmentation methods presented in the literature, most of which are computing intensive and therefore unsuitable for real-time processing. Iivarinen et al. (1996) proposed a defect segmentation approach that was based on texture analysis. The defect contours used in the experiments in this thesis have been extracted from the images using ABB Oy's proprietary defect image segmentation algorithm.

Irrespective of the selected approach, the segmentation results in a binary mask. On the basis of this mask, the defect area can be distinguished from the other regions in the image. The defect boundary line can also be easily defined on the basis of the mask using some common algorithm (Gonzalez and Woods, 1993). Figure 2.5 presents a defect image, its segmentation mask, extracted defect region, and the boundary line of the defect.

Analysis methods

Active academic research work in the field of defect image analysis started at the beginning of 1990s. Since then, a variety of descriptors have been introduced to describe the visual content of defect images. The visual features used in defect image analysis and classification include texture and gray level descriptors. This section gives an overview of these three feature types in defect image analysis.

In defect shape description, simple descriptors (see section 4.2) have been used (Iivarinen, 1998). These descriptors, compactness, convexity, principle axis ratio, circular variance and elliptic variance were used in defect classification in (Iivarinen and Visa, 1998). Chain code histogram (CCH) was also tested in surface defect classification in (Iivarinen and Visa, 1996), in which CCH was proved to be fast but not a rotation invariant method for defect shape description. Recently, Rautkorpi and Iivarinen (2004) have obtained promising defect classification results with edge co-occurrence matrix (ECM). The ECM method indicates the joint probability of edge directions at a certain displacement in the image.

Various texture descriptors have been applied to surface defect images. Iivarinen and Visa (1998) used features extracted from a gray level co-occurrence histogram (Haralick, 1973) in defect classification. MPEG-7 standard (Manjunath et al., 2002) provides visual descriptors for defect characterization. Pakkanen et al. (2003) used MPEG-7 color and texture descriptors in defect image classification with good results. These descriptors have been applied to content-based image retrieval with PicSOM system (Laaksonen et al., 2000) in several papers, such as in (Iivarinen et al., 2004a, 2004b).

The gray level distribution of the surface defect image is also a significant feature in describing the defect. In this area, image histograms are used in the defect description (Iivarinen and Visa, 1998). Image histogram, however, is a first order statistical measure which ignores the spatial relationships between the gray levels in the image. Therefore, second order measures, like gray level correlograms are used in image characterization (Kunttu et al., 2003b). Gray level distribution has also been utilized in binary form, in which the effect of the defect image background can be minimized. The binary approaches have been applied to first order statistics (Kunttu et al., 2003a) as well as to second order co-occurrence matrix (Kunttu et al., 2002). In Publication VI, the binary co-occurrence matrix outperformed the correlograms in the retrieval of paper surface images. A simple unsupervised defect image classification based on gray level and shape is presented in (Kunttu et al., 2003c).

3 Principles of image retrieval and classification

In content-based image database indexing, the goal is to automatically describe the images in such a manner that it is possible to compare the visual content of the images and define the similarity between two images. It is possible to divide the whole image database into categories based on their visual features. However, it is more practical to retrieve only a limited number of images with the desired content from the database. The former approach is known as classification whereas the latter refers to image retrieval. The purpose of the current Chapter is to present the common principles of image retrieval and classification.

3.1 Image database indexing and organization

The indexing of the image databases has been a subject of research since the late 1970s (Rui et al.,1997). In early solutions, the database was indexed textually using keywords, which were used in retrieval. Chang and Fu (1980) used keywords to describe satellite images. In the late 1980s an iconic indexing method was presented by Chang et al. (1987, 1988) in which image content was described using two-dimensional strings. The keyword-based indexing methods have been compared in two articles. The article of Tamura and Yokoya (1984) compared early techniques whereas Chang and Hsu (1992) made a survey of more recent textual image indexing methods. The keyword-based methods have severe drawbacks including the labor intensive nature of the manual indexing of large image databases and the subjective judgment of a person who decides the keywords to be used in the indexing. Therefore, in the early 1990s the increase in computing resources as well as the strong growth of digital image archives caused the manual indexing methods to be replaced by automated ones. Since then, intensive research efforts have directed towards the area of content-based image indexing and retrieval. In the mid-1990's, the first commercial solutions were introduced. Among the first ones were *QBIC* (Flickner, 1995) and *Photobook* (Pentland, 1996). Other well-known content-based image indexing and retrieval systems include *Pictoseek* (Gevers and Smeulders, 2000), *VisualSEEk* (Smith and Chang, 1996a), *Virage* (Gupta and Jain, 1997), *Netra* (Ma and Manjunath, 1996), *MARS* (Mehrotra et al., 1997), PicSOM (Laaksonen et al., 2000), and MUVIS

(Cheikh, 2004). Several surveys and comparisons of the CBIR systems have been published (Rui et al., 1997, 1999; Smeulders et al., 2000).

In content-based image database indexing, the manual text-based content description is replaced by visual features that are automatically extracted from the images. The purpose of image database indexing is to present the content of the database image **I** using a feature vector **f**. Visual features p have been extracted from the images. In database indexing, the features extracted from the ith database image $\mathbf{I}_i$ are collected into a feature vector $\mathbf{f}_i$:

$$\mathbf{f}_i^{\mathbf{I}} = \{p_1, p_2, \ldots, p_n\} \tag{3.1}$$

in which n is the number of the features. This n-dimensional feature vector can be regarded as a point in an n-dimensional feature space.

Similarity measurement

In order to perform retrieval and classification operations in an image database, it is necessary to define the similarity or dissimilarity between two images. This can be done by evaluating the distance between their feature vectors. The distance measure should usually correspond to human perception which means that perceptually similar images should have a small distance between each other whereas perceptually dissimilar images have a larger distance between them. In addition, the computational efficiency of the selected distance metrics plays a role, especially in the case of on-line retrieval. The selection of the distance metrics is dependent on the feature vector type. A wide variety of metrics for different visual descriptors have been developed (Santini and Jain, 1999). For the shape descriptors, similarity measures have been reviewed in (Veltkamp, 2001).

All the distance metrics obey certain rules (Duda et al., 2001). The distance metric is defined as a function that defines a generalized scalar distance between two patterns. Hence the distance between two vectors, say **a** and **b** can be defined as:

$$D_{\mathbf{ab}} = f_d(\mathbf{a}, \mathbf{b}) \tag{3.2}$$

where f_d is a distance function. Duda et al. (2001) define four general properties:

$$D_{\mathbf{ab}} \geq 0 \tag{3.3}$$

$$D_{\mathbf{ab}} = 0, \text{if and only if } \mathbf{a} = \mathbf{b} \tag{3.4}$$

$$D_{\mathbf{ab}} = D_{\mathbf{ba}} \tag{3.5}$$

$$D_{\mathbf{ab}} + D_{\mathbf{b}c} \geq D_{\mathbf{a}c} \tag{3.6}$$

A commonly used distance metric between two n-dimensional feature vectors is Minkowski distance:

$$L_k(\mathbf{a}, \mathbf{b}) = \left(\sum_{i=1}^{n} |\mathbf{a}(i) - \mathbf{b}(i)|^k \right)^{1/k} \tag{3.7}$$

When k is selected to be one, L_1-norm, known as Manhattan- or city block distance, is obtained. L_1-norm defines the distance between two vectors along the coordinate axes.

L_1-norm has been used in image retrieval in (Swain and Ballard, 1991; Stricker and Orengo, 1995) L_2-norm is known as Euclidean distance:

$$L_2(\mathbf{a},\mathbf{b}) = \sqrt{\sum_{i=1}^{d} (\mathbf{a}(i) - \mathbf{b}(i))^2} \tag{3.8}$$

Euclidean distance is the best-known distance metrics and it is commonly used in the similarity measurement of Fourier shape descriptors (Mehtre et al., 1997; Zhang and Lu, 2003a, 2003b).

Image retrieval

Visual information retrieval (Del Bimbo, 2001) is an extension of traditional information retrieval (Baeza-Yates and Ribeiro-Neto, 1999). A query image is selected by the user. The retrieval process is carried out by comparing the database images to the query image. The similarity between the images is based on the feature vectors that have been selected to describe the image content. The database images which are most similar to the query image are recalled and presented to the user.

The benefit of the retrieval approach is that it is not necessary to divide the whole image database into categories because only a subset of the database images is searched. This is particularly important in the case of large image databases.

Image classification

Classification operations can be divided into two categories: supervised and unsupervised. The supervised classification divides the database images into predefined categories whereas the unsupervised one makes the division without prior knowledge of the categories or even the number of them. In this thesis, the classification experiments are used to make comparisons between different visual descriptors in the classification of predefined defect types; hence, the use of the supervised approach.

In the supervised image classification task, the goal is to divide the database images into a fixed number of classes. The user has also defined the classes beforehand such as by using some selected example images to represent the content of each class. The set of example images is called a training set. A wide variety of different algorithms have been developed to train classifiers (Duda et al., 2001). In image classification, the division into classes is based on the visual features extracted from the images using selected distance metrics. Hence the database can be classified on the basis of the same feature vectors as in the case of retrieval. Therefore, the database indexing is not dependent on the selection between retrieval and classification.

Retrieval vs. classification

As noted in the two previous sections, image retrieval and classification obey different principles in the image database organization. However, retrieval and classification share certain common features. When the classification is carried out according to the visual content of the image, it is based on the visual indexing of the database. As in the case of image database retrieval, there are classification algorithms that seek the images most similar to the query image. An example of this type of classification technique is *k*-nearest neighbor classification (*k*-NN) principle (Duda et al., 2001) in which an unknown sample is assigned into the class, which is represented by the majority of its k nearest known samples.

Because there is a clear relation between image retrieval and k-NN classification, the image retrieval problem can be regarded as a classification problem as well. This is beneficial when different visual features are compared. A simple comparison can be carried out by making classification experiments.

3.2 Performance analysis

To make an evaluation between different features in image retrieval and classification, it is necessary to make a performance comparison between them. This performance evaluation, known as validation, is based on retrieval and classification experiments in the indexed image database.

Retrieval performance

When the retrieval performance of the content-based image retrieval techniques is measured, the validation methods of general database retrieval can be used. Perhaps the most widely used method for performance measurement of a particular retrieval task is to present precision as a function of recall (Baeza-Yates and Ribeiro-Neto, 1999). The information retrieval process can be modeled in the following manner. Let us consider a database (figure 3.1) that includes a set of objects of the user's interest R. Let the number of the objects in R be $|R|$. When the user performs a retrieval operation, the resulting recall set A contains $|A|$ objects. The size of the recall set $|A|$ can be defined by the user. Thus the intersection of the sets A and R can be defined as Ra, in which the number of objects is $|Ra|$. The set Ra consists of the recalled objects that are of the user's interest. On the basis of these sets it is possible to define the commonly used performance measures for information retrieval:

$$\text{Recall} = \frac{|Ra|}{|R|} \tag{3.9}$$

$$\text{Precision} = \frac{|Ra|}{|A|} \tag{3.10}$$

It is common that the objects in the recalled set are ranked according to their similarity to the query image. Hence the precision value varies with the size of the recall set. As a result, precision is usually presented as a function of recall. Consequently, the retrieval performance can be expressed by means of precision/recall curve. In practical retrieval experiments, several consequent queries are carried out. An average precision with fixed recall levels can be defined as:

$$\overline{P}(r) = \sum_{i=1}^{N_q} \frac{P_i(r)}{N_q} \tag{3.11}$$

in which $\overline{P}(r)$ corresponds to average precision at recall level r, N_q is the number of the queries and $P_i(r)$ means the precision of the i:th query at recall level r. The average precision of equation (3.11) is a common way of measuring the retrieval performance

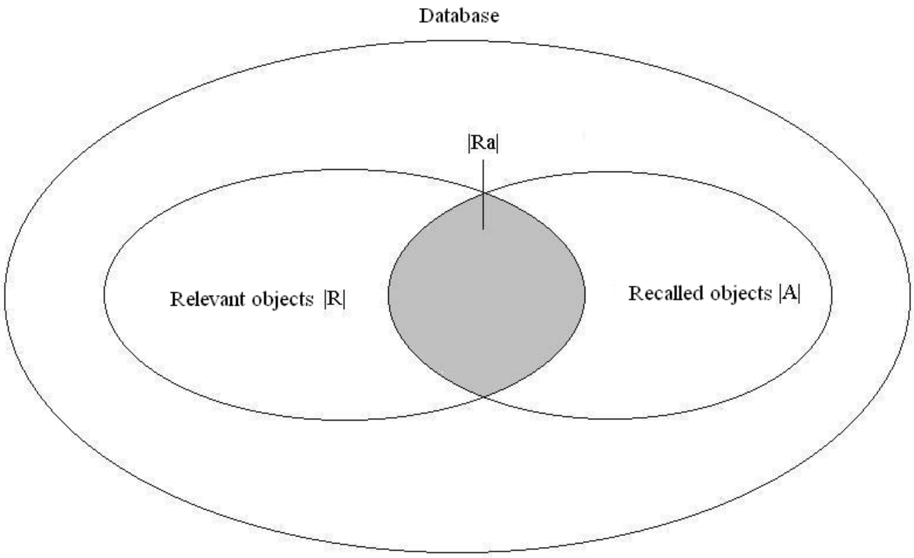

Figure 3.1. The sets *A*, *R*, and *Ra* in the database (Baeza-Yates and Ribeiro-Neto, 1999).

(Baeza-Yates and Ribeiro-Neto, 1999). The precision and recall values can be expressed as percentages. The selection of recall area is dependent on the application and the number of the images in the database. With small databases the whole recall area between 0 and 100% is usually used. In the case of large image databases it is unnecessary and often impractical to use the whole recall area because in real retrieval tasks the amount of the recalled images is usually limited to a small number of best matching images.

Classification performance

The measurement of classification performance is a straightforward operation based on the classification results. The result of the classification is compared to the real classes of the sample images. In this manner the classification rate can be defined. The classification rate can be expressed as an average value for the whole database or as an average classification rate of each image class. In addition, in the case of classification, the average classification rate is usually expressed in terms of percentages.

In practice, the classification experiments are carried out by dividing the available image data into training and testing data. It is usual that roughly 2/3 of the data is for training and the rest is used as testing data. When the classifier has been trained using training data, its classification ability is tested using testing data. The classification rate is then defined on the basis of these tests.

Cross validation

In validation, the retrieval and classification methods are experimentally tested. The problem of training and testing datasets is shared by both types of method. The division of the image database into these two datasets is sometimes impractical for two reasons. In the first place, the division into two sets should be such that both sets are equally representative. This is not always the case, especially with small or non-homogenous databases. By selecting certain type of data to be used in training or testing, the classification results can be corrupted and the testing is not repeatable. Secondly, in the case of retrieval the whole database is interesting; if it has been divided into testing database and a distinct set of query images, the results can be easily manipulated on the basis of the selection of the query images.

A simple and effective solution for the validation is the use of the leave-one-out validation principle (Hand et al., 2001). In this method each database image is employed as a query image in turn, whereas the rest of the database serves as a testing database. This principle can be applied to retrieval and classification. When all the images have been used as query images in retrieval or they have been classified, it is possible to calculate an average result from all the experiments.

4 Shape description

The goal in this thesis is to find effective shape descriptors for shape retrieval. The descriptors are used in image database indexing. This Chapter discusses the common methods of shape description. The focus of this study is on contour-based shape description methods, especially those which are based on transforms. Therefore, these methods also receive most attention in this Chapter, although other methods, including region-based approaches are also touched on briefly. There are several reasons to prefer contour-based shape descriptors in shape-based image retrieval (Zhang and Lu, 2005). Firstly, the contour-based descriptors are computationally less expensive than region-based ones and they are also easy to derive. Secondly, most real world objects have clear contours, which are readily available. Moreover, human beings are also able to discriminate shapes on the basis of their outlines. For reasons such as these, the contour-based methods are also popular also in the literature on shape description.

4.1 Shape description in image retrieval

In addition to texture and color, shape is one of the most essential visual features used in image retrieval. For humans and animals, shape is a dominant characteristic for the identification of similar objects (Belongie et al., 2002). The definition of perceptual shape features is a difficult task. It is even more difficult to perceptually measure the similarity between shapes. Furthermore, real-world shapes are often more or less noisy and distorted, which makes the shape recognition even more challenging. In recent decades, numerous approaches and solutions have been presented to characterize different shapes. The primary purpose of the early shape descriptors was based on object recognition and classification based on shape, whereas during recent years the use of shape description in image retrieval has received increasing interest among the research community.

In current CBIR systems, one of the problems is to answer the question: “Which of the database images contain the most similar shapes to the query image?” This type of image retrieval is called shape similarity-based retrieval (Mehtre et al., 1997). In this kind of retrieval, the aim is to find similar shapes from the database as accurately as possible. There are several principles that characterize a good shape descriptor (Carlin, 2001; Li and Edwards, 2004):

1. *Invariance*: If two boundaries have the same shape, irrespective of translation, scaling, rotation, and the boundary starting point, they will have the same descriptor.
2. *Uniqueness*: If two boundaries do not have the same shape, they will not have the same descriptor.
3. *Stability*: Reflection of shape differences between boundaries should be appropriate in their representations. For example, if two boundaries have small perceptual shape differences, the difference in descriptors will be small.
4. *Comprehensible*: Descriptors must fit the descriptive model of the shape, and hence must be aligned with our shape terminology.
5. *Invertibility*: If a boundary is given, its shape description should be computable; if a shape description is given, the boundary representation should be computable.
6. *Efficiency*: The representation should be efficient to compute and store. This is especially important when a retrieval or recognition is to be performed in real-time.

Hence, due to the increasing number of online retrieval solutions, computational efficiency is nowadays considered to be equally important as effectiveness, which refers to accuracy. A recently introduced multimedia standard, MPEG-7 (Manjunath et al., 2002), has also set six principles for measuring a shape descriptor (Kim and Kim, 2000):

1. *Good retrieval accuracy*: Shape descriptor should be able to effectively find perceptually similar shapes from the database. Rotated, translated, affinely transformed and scaled versions of a shape are usually regarded as perceptually similar.
2. *Compact features*: Small memory should be required for storing the descriptor. Low dimensionality is desirable especially in on-line retrieval.
3. *General application*: The descriptor should be able to effectively describe shapes in general, not only a specific shape type.
4. *Low computational complexity*: In addition to descriptor efficiency in feature extraction and similarity calculation, low computational complexity also refers to minimizing the effect of any uncertain factor on the computation processes (Zhang and Lu, 2004).
5. *Robust retrieval performance*: The robustness of a shape descriptor means that noise affected and distorted shapes are also tolerated. In addition, no drastic performance degradation should take place when the size of the database is increased.
6. *Hierarchical coarse to fine representation*: The descriptor should be able to describe the shape with incremental accuracy. That is, coarse level description can be used to eliminate clearly irrelevant shapes from the query results, and finer level description can be added to refine the result.

These principles were used as criteria in the study of Zhang and Lu (2004), in which common shape description techniques were reviewed. Another review of the state of the art in shape description techniques is provided by Loncaric (1998). Thorough discussion on methods and algorithms of certain areas of shape description is presented in (Costa and Cesar, 2001).

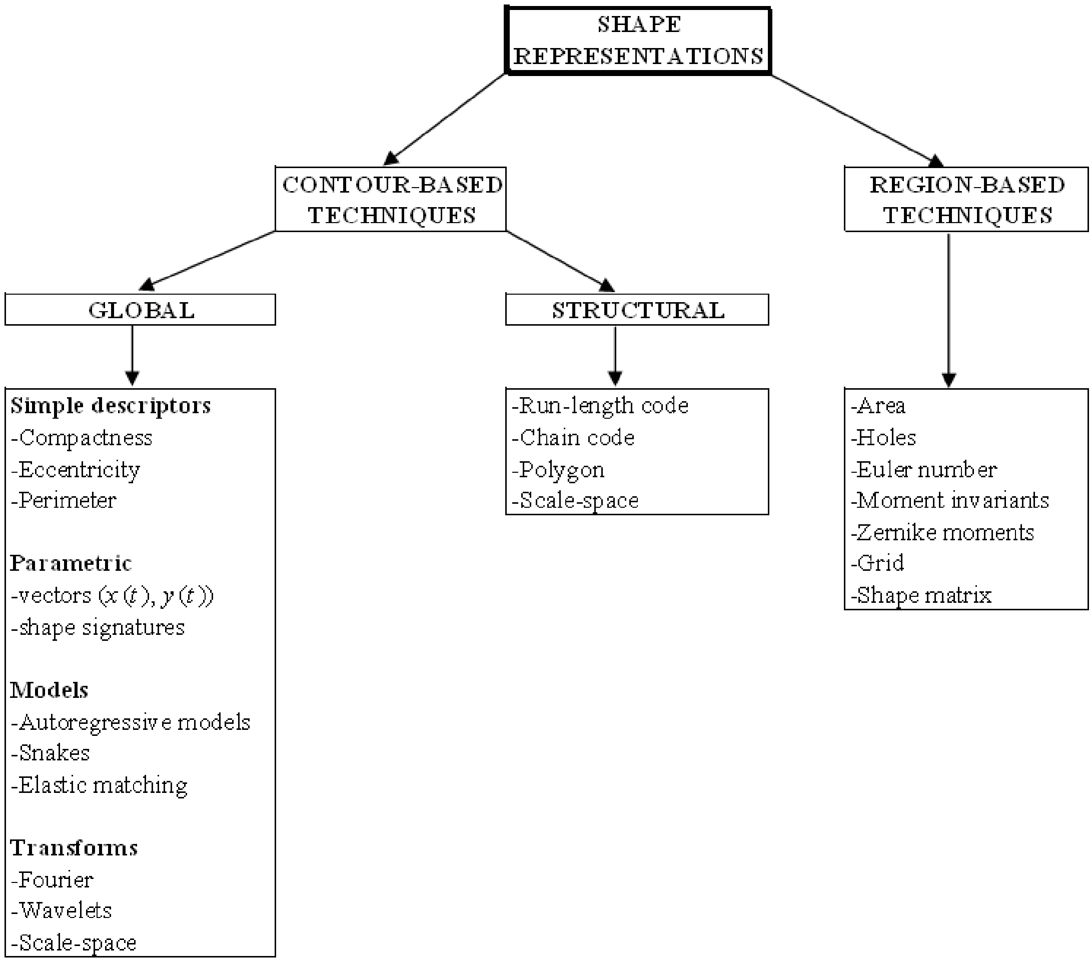

Figure 4.1. A taxonomy of shape representation techniques.

Shape description techniques

Several types of taxonomy have been presented for shape description techniques. The basic categorization into region-based and contour-based techniques is common to most of them. The region-based methods consider the entire area of the object whereas the contour-based methods use only the object boundary to characterize its shape. These are also known as boundary-based methods. In addition to these two main categories, Costa and Cesar (2001) consider transform-based shape descriptors as a separate category. In most of the other taxonomies, the transform-based methods are included in the two main categories. In the classification of Safar et al. (2000) the contour-based and region-based techniques have been divided into subcategories depending on their spatial or transform-based nature. This classification has also been adopted in the study of Cheikh (2004). Mehtre et al. (1997) also use this division for region-based techniques, but they further subdivide the spatial domain into geometric and structural methods. In the taxonomy introduced by Zhang and Lu (2004) region-based and contour-based approaches are divided into structural and global subcategories. Structural methods use segments as shape primitives whereas global methods consider the whole shape. This taxonomy, presented in figure 4.1, was selected for use in this thesis. It has been somewhat simplified by ignoring some less-known shape description methods and combining the region-based approaches into a single category as in (Zhang and Lu, 2003b).

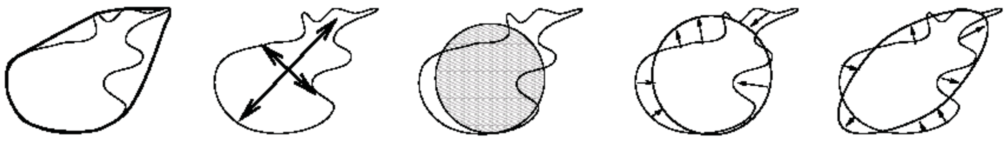

Figure 4.2. Simple shape descriptors (Iivarinen and Visa, 1998).

4.2 Overview of contour-based shape descriptors

The contour-based shape descriptors can are divided into global and structural domains, as presented in figure 4.1.

Global methods

The global domain methods can be further divided into four subclasses: simple descriptors, parametric contours, models, and transforms. Some simple shape descriptors can be considered to belong to the category of contour-based methods. Simple descriptors such as circularity (Costa and Cesar, 2001), eccentricity, convexity, principle axis ratio, circular variance and elliptic variance (Iivarinen and Visa, 1998) belong to this descriptor type as shown in Figure 4.2. These simple shape descriptors have been applied to defect image classification in (Iivarinen and Visa, 1998).

In the case of parametric contour methods, the shape outline is represented as parametric curves or signals ($x(t)$, $y(t)$). The parametric contour is first extracted using some contour following algorithm. A simple vector representation uses the x and y coordinates of the contour as vectors. To form a one dimensional vector, the coordinates can be expressed as a set of complex numbers $x(t)+jy(t)$ (Persoon and Fu, 1977). This kind of one-dimensional function which is used to describe two-dimensional shape is called *shape signature*. Other common shape signatures include tangent angle versus arc length function (Zahn and Roskies, 1972) and centroid distance function in which the boundary points are expressed in terms of their distance from the object centroid (Chang at al. 1991). Wang et al. (1994) used a sequence of line segment moments as a boundary function. The signature can be also expressed as a histogram, which is rotation invariant and therefore easy to match (Squire and Caelli, 2000). A general drawback of direct matching of shape signatures is their sensitivity to noise and distortions in the boundary line. Furthermore, the signature matching is computationally expensive. Therefore, it is common for the signatures to be applied by some transform or model-based method rather than used in shape matching directly. The previously presented parametric methods express the boundary line using some given order. On the other hand, it is also possible to present the boundary points without any given order, i.e. as a set. The use of set of contour points, however, often yields poorer shape representation than parametric methods (Costa and Cesar, 2001).

The third of the subclasses is the class of models. The object boundary can be modeled using *autoregressive models* (AR). These models are based on stochastic modeling of the one dimensional shape signature. An AR-model is a parametric equation that expresses each sample as the linear combination of a certain number of previous samples. The general form of an AR-model (Kashyap and Chellappa, 1981) of a closed boundary function f with m previous samples can be defined as:

$$f_t = \alpha + \sum_{j=1}^{m} \theta_j f_{t-j} + \sqrt{\beta}\omega_t \tag{4.1}$$

where θ_j correspond to the AR-model coefficients to be estimated and to be used as features. The model also produces values for two constants, α is proportional to the mean of the function values and $\sqrt{\beta}\omega_t$ is the error term. The algorithms for the estimation of the coefficients can be found in (Dubois and Glanz, 1986). The parameters $\{\alpha, \theta_1,\ldots,\theta_m, \beta\}$ can be estimated using *least square* (LS) criterion (Dubois and Glanz, 1986). The estimated coefficients θ_j as well as ratio $a/\sqrt{\beta}$ are translation, rotation, and scale invariant, and therefore they are used as features describing the object shape. The drawback of an AR-based shape description is that in the case of complex boundaries, a small number of parameters does not yield accurate shape description.

Snakes and *active contours* have received a certain amount of research interest in shape classification and retrieval. These methods belong to the shape transformation models (Widrow, 1973), which has been the basis of the snake model (Kass et al., 1988) and *elastic matching* (Del Bimbo and Pala, 1997) method. In the shape transformation models, the shape is regarded as a template which is deformed in order to match it with a target image.

In transform-based approaches, the parametric boundary function is transformed using some linear or non-linear transform. Spectral transforms form a subset of transform domain methods. These methods analyze the shape in spectral domain, which is beneficial because the problems with noise sensitivity and boundary variations can be minimized this way.

One of the most popular boundary-based shape representations is *Fourier descriptor* (FD). Instead of being a single shape representation method, FD refers to a class of methods that use Fourier transform to describe the shape (Costa and Cesar, 2001). The basic idea is to transform the selected shape signature using Fourier transform and use the obtained transform coefficients as shape descriptors. For a boundary function $z(k)$ of length N, the discrete Fourier transform (DFT) can be defined as:

$$F(n) = \frac{1}{N}\sum_{k=0}^{N-1} z(k)e^{-j2\pi nk/N} \tag{4.2}$$

for n=0,1,2,…,N-1 and $F(n)$ are the transform coefficients of $z(k)$. Typically the low frequency coefficients are used as shape descriptors because most of the boundary information is concentrated along them. The coefficients are usually normalized to achieve invariance to rotation and scaling. The descriptors can be made rotation invariant by ignoring the phase information and using only the magnitudes of the transform coefficients $|F(n)|$. The scale can be normalized by dividing the magnitudes of the transform coefficients by $|F(0)|$ or $|F(1)|$, depending on the application. The normalization process is discussed in more detail in Chapter 5.

FD's were originally introduced in the 1960s by Cosgriff (Zahn and Roskies (1972)). The first papers in this field were published by Granlund (1972) and Zahn and Roskies (1972) which were followed by the study of Persoon and Fu (1977). In addition to closed curves, Fourier description can be applied to partial shapes (Lin and Chellappa, 1987; Mitchell and Grogan, 1984). Affine shape description for three-dimensional shapes using Fourier descriptors was proposed by Arbter et al. (1989, 1990). *Short time Fourier descriptor* was introduced by Eichmann et al. (1990) but it has been found to be less accurate than conventional FD methods (Zhang and Lu, 2001). Different shape signatures used for Fourier shape description were compared in (Zhang and Lu, 2002, 2005). A good review and comparison of Fourier-based shape descriptor methods is provided by Zhang and Lu (2005).

The Fourier-based shape representations are easy to implement and they are computationally inexpensive methods, which is due to the use of efficient FFT algorithms (Gonzalez and Woods, 1993). The descriptors are also compact and low-dimensional, when a small number of low-frequency coefficients are considered. Furthermore, the FD's are easy to normalize and their matching is a very simple process. In addition, their sensitivity to noise is low when only low frequency Fourier coefficients are used as descriptors. A major disadvantage is that the frequency-based shape representation does not have a clear connection with human vision. The spatial relationships of the details in the boundary line are also ignored in shape description in frequency domain.

Another type of transform domain approach are *wavelet descriptors* (WD) (Ohm et al., 2000; Tieng and Boles, 1997; Yang et al., 1998). These descriptors are based on wavelet transform (Chui, 1992) that is a commonly used approach in signal processing and image analysis. The WD's have a multiresolution property that is of certain benefit in shape description applications. On the other hand, however, the wavelet transform coefficients of the boundary function are not scale or rotation invariant (Pfeiffer and Pandit, 1995). Normalization is more complicated than that of Fourier descriptors due to the spatial information included in the wavelet coefficients. Khalil and Bayoumi (2001) used affine invariant transform to normalize the coefficients whereas Kashi et al. (1996) proposed a normalization in frequency domain using Fourier transform and its inverse applied to boundary function before wavelet transform. Li and Edwards (2004) also presented a normalization method for wavelet coefficients that utilizes DFT. In this case, the rotation invariance is achieved by removing phase information from the wavelet coefficients. Chen and Bui (1999) presented the object shape using polar coordinates (r, θ). The rotation invariance was achieved by Fourier transforming the polar angle θ whereas the radius r was wavelet transformed. In addition to normalization, the matching scheme of the wavelet representations is more complicated and time consuming than that of Fourier descriptors. In addition, the dimensionality of WD's is usually higher than that of FD's.

In recent years, *scale space* shape description approaches first introduced by Witkin (1983) have received a certain amount of attention among the research community. In the scale space methods the problems caused by noise and variations in the boundary have been avoided by describing the boundary line on the basis of its extreme points. Asada and Brady (1986) first utilized the scale space method using tree structure which yields multiscale shape representation. In the *Curvature scale space* (CSS) method introduced by Mokhtarian and Mackworth (1986, 1992) the boundary line is iteratively smoothed using a low-pass Gaussian filter. In the CSS-method, the boundary is iteratively smoothed until the curvature function has no inflection points; i.e. the boundary is convex. The inflection points of the boundary line are sought using curvature function. When (x_k, y_k), where k=0,1,2,…,N-1, represent the object boundary coordinates, the curvature function of the boundary can be defined as:

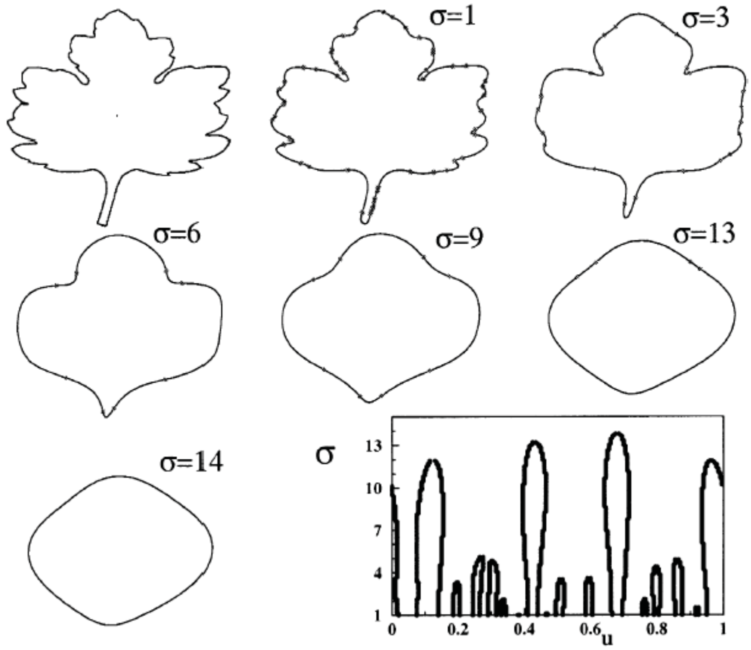

Figure 4.3. The smoothing process of a leaf contour. The CSS representation is formed on the basis of the zero-crossing points of the contours at each scale (Mokhtarian and Abbasi, 2004).

$$c(k) = \frac{\left(\dot{x}(k)\ddot{y}(k) - \ddot{x}(k)\dot{y}(k)\right)}{\left(\dot{x}^2(k) + \dot{y}^2(k)\right)^{3/2}} \tag{4.3}$$

where $\dot{x}(k)$, $\ddot{x}(k)$, $\dot{y}(k)$ and $\ddot{y}(k)$ are the first and second derivatives of the boundary coordinates, respectively. The zero-crossing points of the curvature function are regarded as inflection points. During the smoothing process, insignificant inflection points are eliminated whereas the significant ones remain. Figure 4.3 presents the smoothing process of a leaf contour as presented in (Mokhtarian and Abbasi, 2004). In this figure, the leaf is iteratively smoothed with one dimensional Gaussian kernel of width σ. Different scales of the shape are obtained as σ increases and the boundary becomes smoother. The shape descriptor obtained is a contour map that consists of the zero-crossing points that are presented at different degrees of boundary smoothness. The peaks in the contour map caused by the zero-crossing points are detected and they are used in shape matching. The matching is a complicated operation due to the varying number of contour peaks to be matched. Furthermore, the contour maps have to be scaled and shifted at each matching operation. The CSS method has been further developed for shape retrieval purposes (Abbasi et al., 1999, 2000). In these approaches the CSS matching has been made simpler by taking only the two highest peaks (maxima) of the CSS representation into account in the matching process. The scaling of the boundary line is also made beforehand, before applying the CSS method. The minima of the CSS contours are also included in the matching procedure in (Mokhtarian and Abbasi, 2004).

Structural methods

Common structural methods include *run-length codes* (Kim et al., 1988) and *chain codes* (Freeman, 1961; Freeman and Davis, 1977). Chain code representation is a well-known method for shape description. It represents the object boundary using a

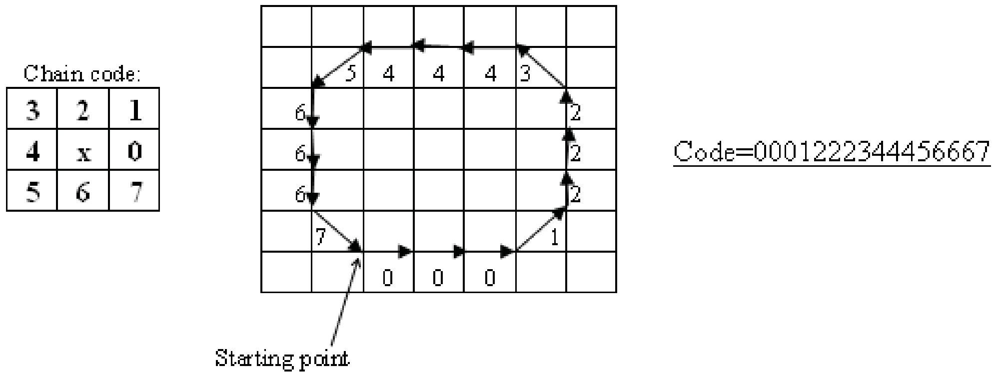

Figure 4.4. The principle of chain code.

sequence of small vectors of unit length. The set of possible directions is usually limited to 4 or 8. *N*-directional ($N>8$ and $N=2^k$) chain code is also possible (Freeman and Saghri, 1978). The object boundary is superimposed with a grid and the boundary points are approximated to the nearest grid points. From a selected starting point, a chain code can be generated by using code words, each corresponding to the direction of the vector drawn between the consecutive grid points. Figure 4.4 presents the principle of the chain code method. The matching of the chain codes is not straightforward because they are dependent on their starting points. Furthermore, the chain code method is sensitive to object scale and rotation and it is also a noise-sensitive approach. The problems with starting point dependence and rotation variance can be overcome by using *chain code histogram* (CCH) introduced by Iivarinen and Visa (1996). Translational and rotational invariant chain code approach is also *vertex chain code* as proposed by Bribiesca (1999).

The third of the spatial domain methods uses boundary approximation or interpolation. The inspiration behind these techniques is that rather than describing the whole boundary with a single function, it is more convenient to piecewise approximate each portion of the boundary (Costa and Cesar, 2001). The portions can be approximated using geometric primitives (Pavlidis, 1986). Polygonal Approximations (Ramer, 1972) uses straight segment lines as shape primitives. Such an approach yields a polygon-like shape approximation. The problem with the polygonal approaches is the selection of the boundary points which serve as endpoints of segment lines so that the polygonal approximation corresponds to the original boundary as closely as possible. A simple example of polygonal representation of a contour is presented in figure 4.5. A common approach is to draw the lines between high curvature points of the boundary line. This is motivated by human perception (Tsang et al., 1994). In the approach of Groskey et al. (1990, 1992), the polygon vertices are described with four parameters: internal angle, distance from the next vertex, and its x and y coordinates.

4.3 Overview of region-based shape descriptors

In region-based methods, the whole area of the object is considered. Hence, the pixels belonging to the object interior are of equal importance to those belonging to the boundary. The simplest area-based methods include the area and holes of the object. The Euler number (Gonzalez and Woods, 1993) is defined as a difference between the numbers of connected regions and of holes in an object.

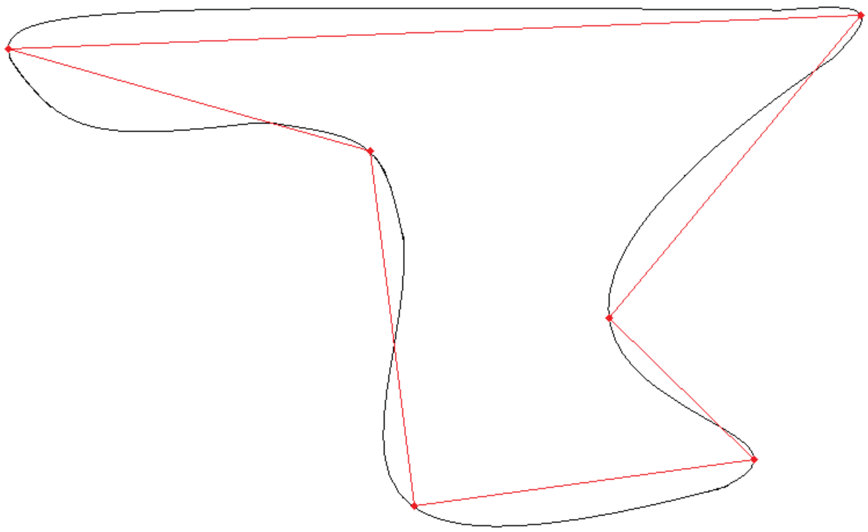

Figure 4.5. Polygonal contour representation.

Perhaps the most popular region-based shape description methods are *moment invariants*, first introduced by Hu (1962). The moment invariants are based on non-linear combinations of geometric moments:

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q \rho(x,y)dxdy, \quad p,q = 0,1,2,... \tag{4.4}$$

which are usually utilized using low orders. The obtained set of geometric moments is translation, scaling, and rotation invariant. The geometric moments have received a significant amount of interest (Teague, 1980; Liao and Pawlak, 1996). A drawback, however, is their increase in complexity with increasing order. The use of orthogonal moments, like Zernike moments (Teague, 1980) has been found to improve shape description (Teh and Chin, 1988). Zernike moments have also been selected for use as region-based shape descriptors of MPEG-7 standard (Zhang and Lu, 2003b). Shen and Ip (1999) used wavelet moment invariants in shape description. This approach outperformed those such as Zernike moments in shape classification.

Grid-based methods (Lu and Sajjanhar, 1999) use a binary grid structure that defines the image area covered by the shape. The grid needs to be normalized to be invariant for rotation, translation, and scaling. *Shape matrix* method (Goshtasby, 1985) uses normal raster sampling in a grid formed by concentric circles and radial lines. The binary value of the shape is sampled in the intersections of the circles and radial lines.

4.4 Comparisons

The area of shape description has been the subject of intensive research work for the past three decades. As a result, a wide variety of shape descriptors for various purposes have been presented and these have also been compared in several comprehensive comparisons. The retrieval performance of the most common boundary-based shape descriptors, Fourier descriptor and chain codes were compared with different moment invariants by Mehtre et al. (1997). In this study, the testing database consisted of 500 trademark images and the best results were obtained by using moments and Fourier descriptors whereas chain codes gave the lowest result. In the comparison made by Kauppinen et al. (1995) FD's were compared to autoregressive models in the boundary-based classification of planes and characters. In most cases, FD's outperformed autoregressive models. In the recent study of Zhang and Lu (2003a), FD's and CSS were compared in the retrieval of simple shapes. The conclusion of this study was that FD clearly outperforms CSS in terms of retrieval accuracy. The CSS is also computationally heavier than FD due to its complicated matching scheme. The retrieval performance of the MPEG-7 shape descriptors were compared to that of FD's in (Zhang and Lu, 2003b). In this comparison, FD proved to be a better boundary-based shape descriptor than CSS. In the case of region-based descriptors, Zernike moment descriptor (ZMD) outperformed geometric moments and grid descriptors. The ZMD also slightly outperformed FD in terms of retrieval effectiveness (Zhang and Lu, 2003b). However, the computational efficiency FD showed a clear advantage over ZMD.

The retrieval and classification performance of the shape descriptors is always dependent the image data used in testing. On the other hand, in the comparisons presented above, the performance of FD's was proved with several types of image data. It is obvious that boundary-based descriptors are not adequate shape descriptors for shapes with complicated interiors, such as trademarks. Therefore, a more suitable shape description for them is region-based representation. On the other hand, Mehtre et al. (1997) showed that Fourier description gives quite good results with trademarks, too. It is also important to note that region-based descriptors, such as moments, have significantly lower computational efficiency than FD's. This may limit their usability in on-line retrieval.

5 Fourier-based shape representation

The shape representation methods presented in this thesis are based on Fourier transform. This Chapter presents the general Fourier methods for boundary-based shape description.

5.1 Fourier transform

Fourier transform, originally introduced by the French mathematician Joseph Fourier at the beginning of 19th century, is a powerful tool for several types of signal analysis. The Fourier-based signal representation is based on the fact that any periodic[1] signal $f(t)$ of finite extension[2] can be expressed in terms of complex exponentials (Bigun, 1995):

$$\psi_m(t) = \frac{2\pi}{T} e^{i\omega_m t} = \exp(i\omega_m t) \tag{5.1}$$

where

$$\omega_m = m\frac{2\pi}{T}, \; m = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{5.2}$$

The series can also be expressed as:

$$\{\psi_m\}_m = \{\ldots, e^{-i2\omega_1 t}, e^{-i\omega_1 t}, 1, e^{i\omega_1 t}, e^{i\omega_1 t}, \ldots\}\omega_1 \tag{5.3}$$

in which $\omega_1 = 2\pi/T$. The basis of equation (5.3) is called the Fourier basis. The complex exponentials are orthogonal. On the basis of this, the well-known formulas

[1] A function is deemed periodic function if there is a positive constant T, called a period, such that $f(t)=f(t+nT)$ for all integers n.

[2] A function $f(t)$ is a finite extension if there exist finite real constants a and T such that $t \notin \left[a - \frac{T}{2}, a + \frac{T}{2}\right]$.

for Fourier transform (synthesis) and inverse Fourier transform (analysis) can be formed:

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \tag{5.4}$$

$$f(t) = \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \tag{5.5}$$

Function F represents the Fourier transform of the function f.

The discrete Fourier transform (DFT) for an N dimensional vector containing a function $f(t)$ of limited length is defined as:

$$F(n) = \frac{1}{N} \sum_{t=0}^{N-1} f(t) e^{-j2\pi tn/N} \tag{5.6}$$

and it has an inverse transform:

$$f(t) = \sum_{n=0}^{N-1} F(n) e^{j2\pi tn/N} \tag{5.7}$$

The DFT produces a sequence of complex numbers $F(n)$ of limited length (N). Hence, this sequence is also discrete-valued. To reduce the computation of DFT, several types of algorithms have been introduced to make *Fast Fourier Transform* (FFT). The most widely known of these is the algorithm proposed by Cooley and Tukey in 1961. Its performance is at maximum when the length of the signal N is 2^p in which p is a positive integer (Myers, 1990; Orfanidas, 1996).

5.2 Shape description

Discrete Fourier transform is a popular tool in the description of object boundary line (Zahn and Roskies 1972; Granlund, 1972; Persoon and Fu, 1977). Using DFT, the boundary can be expressed in frequency domain and in this way, the frequency content of the boundary line can be described.

Shape signatures for Fourier description

When the boundary line of an object is considered, the Fourier description is based on a selected shape signature $f(k)$ (i.e. a one dimensional function that describes a two-dimensional object boundary). A variety of complex-valued and real shape signatures for boundary representation have been introduced (see Section 4.2). In this discussion, three of those functions are presented.

Complex coordinate function is a simple and probably the best-known signature used in the Fourier-based shape description. Let (x_k, y_k), k=0,1,2,…,N-1 represent the object boundary coordinates, in which N is the length of the boundary. The complex coordinate function $z(k)$ expresses the boundary points in an object centered coordinate system in which (x_c, y_c) represents the centroid of the object:

$$z(k) = (x_k - x_c) + j(y_k - y_c) \tag{5.8}$$

Hence, the complex coordinate function includes two real-valued functions that are combined to a complex function. This way the resulting function is one-dimensional, and therefore easily applicable with DFT. Centroid distance function is a real-valued function that is defined as the distance between the boundary points and the object centroid:

$$r(k) = \sqrt{(x_k - x_c)^2 + (y_k - y_c)^2} \tag{5.9}$$

Another real-valued shape signature is area function (Zhang and Lu, 2002) that is defined as the area of the triangle formed by two boundary points and centroid in the object centered coordinate system:

$$a(k) = \frac{\left|(x_k - x_c)(y_{k+1} - y_c) - (x_{k+1} - x_c)(y_k - y_c)\right|}{2} \tag{5.10}$$

Hence, all three signatures represent the boundary irrespective of the location of the object in the image. Therefore, these signatures can be considered as translation invariant shape representations.

Discrete Fourier transform of an object shape

Irrespective of the selection of the signature, the boundary functions are closed curves, which mean that they are periodic. Since the functions are discrete signals, they can be easily transformed using DFT. For a selected real or complex-valued boundary function $f(k)$, DFT can be written as:

$$F(n) = \frac{1}{N}\sum_{k=0}^{N-1} f(k)e^{-j2\pi nk/N} \tag{5.11}$$

for n=0,1,2,…,N-1. The obtained coefficients $F(n)$ form a Fourier spectrum of the boundary function $f(k)$ and they are usually called Fourier descriptors (FD's). It is also possible to synthesize the original $f(k)$ from $F(n)$ via inverse DFT:

$$f(k) = \sum_{n=0}^{N-1} F(n)e^{j2\pi nk/N} \tag{5.12}$$

Figure 5.1 demonstrates the effect of DFT and its inverse in the shape description. Here, a defect boundary is transformed and inverse transformed. The inverse transform is performed using a reduced number of the coefficients $F(n)$. The coefficient numbers are reduced to 50%, 10%, 5%, and 2% of their original number so that high frequencies are omitted. Hence, the figure shows that the greater the number of coefficients used to reconstruct the original boundary function, the greater is the amount of fine detail preserved. On the other hand, the general shape can be reconstructed with a small set of low frequency coefficients.

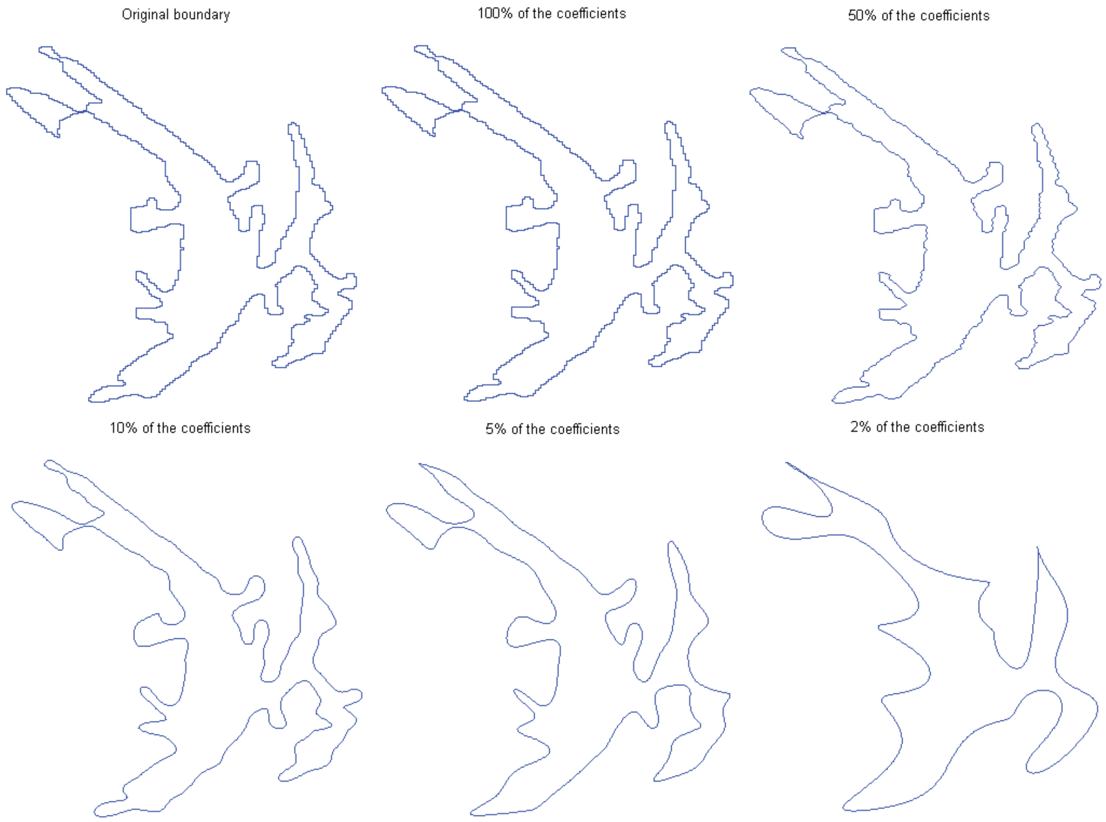

Figure 5.1. Reconstruction of a defect shape using limited number of Fourier transform coefficients.

Invariance of the Fourier descriptors

The Fourier transform coefficients $F(n)$ have several desirable properties for use in shape description (Bigun, 2005). Although the translation invariance of the obtained shape descriptors is due to the shape signature as discussed in section 5.2, the translational invariance is also one of the properties of Fourier coefficients $F(n)$. Let the boundary line be described by complex coordinate function of equation (5.8). In DFT, the object centroid is represented by $F(0)$ that is also known as DC component:

$$F(0) = \frac{1}{N}\sum_{k=0}^{N-1} z(k)e^{-j2\pi k*0/N} = \frac{1}{N}\sum_{k=0}^{N-1} z(k) \tag{5.13}$$

that is the mean of $z(k)$, i.e. the object centroid. On the other hand, all other Fourier coefficients than $F(0)$ are translation invariant. This can be proved by applying a translation $\Delta z = \Delta x + j\Delta y$ to the boundaries. Hence, new boundary coordinates $z'(k) = z(k) + \Delta z$ are obtained. The DFT is applied to this new boundary function:

$$\begin{aligned} F'(n) &= \frac{1}{N}\sum_{k=0}^{N-1} z'(k)e^{-j2\pi nk/N} = \frac{1}{N}\sum_{k=0}^{N-1}\left(z(k)+\Delta z\right)e^{-j2\pi nk/N} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} z(k)e^{-j2\pi nk/N} + \Delta z\cdot\frac{1}{N}\sum_{k=0}^{N-1} e^{-j2\pi nk/N} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} z(k)e^{-j2\pi nk/N} + \Delta z\cdot\delta(n) \\ &= F(n)+\Delta z\cdot\delta(n) \end{aligned} \tag{5.14}$$

in which $\delta(n)$ is Kronecker delta function. Hence, to obtain translation invariant Fourier boundary description, only $F(0)$ should be omitted.

In addition to translation invariance, the FD's should be made invariant to scale (Bigun, 2005). Let z' represent a scaled boundary so that $z'=\alpha z$, in which α is a positive real scalar. Then the DFT of z' is defined as:

$$\begin{aligned} F'(n) &= \frac{1}{N}\sum_{k=0}^{N-1} z'(k)e^{-j2\pi nk/N} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} \alpha z(k)e^{-j2\pi nk/N} = \alpha F(n) \end{aligned} \tag{5.15}$$

Hence, all the descriptors $F'(n)$ can be normalized by scaling them by one of them, e.g. $F'(1)$:

$$\frac{F'(n)}{|F'(1)|} = \frac{\alpha F(n)}{\alpha |F(1)|} \tag{5.16}$$

when $|F'(1)|>0$.

In shape description, invariance to rotation is usually also necessary. In other words, the shape description of an object should be the same irrespective of its rotation angle. This matter is related to the starting point of the boundary function. Hence, in the shape description, it is essential that the selection of the starting point has no influence on the description. It follows that the FD's have to be normalized in terms of rotation. When the boundary function is rotated at an angle θ so that $z' = e^{j\theta}z$ is a rotated version of the boundary function. Then the DFT of z' is:

$$\begin{aligned} F'(n) &= \frac{1}{N}\sum_{k=0}^{N-1} z'(k)e^{-j2\pi nk/N} \\ &= \frac{1}{N}\sum_{k=0}^{N-1} e^{j\theta}z(k)e^{-j2\pi nk/N} = e^{j\theta}F(n) \end{aligned} \tag{5.17}$$

The rotation invariance can be achieved in several ways in Fourier description. The simplest approach is to omit the phase information by using only the magnitudes of FD's, $|F(n)|$, which is based on the fact that $|F(n)|$ and $|F'(n)|$ are the same (Mitchell and Grogan, 1984). However, as presented in (Bigun, 1995), the use of the magnitudes is not necessary, because FD's can also be normalized by using $F^*(1)/|F(1)|$ as a normalizing factor for the coefficients $F(n)$.

Feature vectors for Fourier descriptors

As discussed in section 5.2, a translation, scaling and rotation invariant descriptor for coefficients $F(n)$ can be formed as:

$$\frac{|F(n)|}{|F(1)|} \quad \text{where } n = 2,3,\ldots,N-1 \tag{5.18}$$

provided that $|F(1)| >>0$. The general shape of the object is represented by the low frequency coefficients, whereas high frequency coefficients represent the fine details of the object shape. A common approach to shape representation is to use a subset of the low frequency coefficients as a shape descriptor. In this way the shape can be effectively presented using a relatively short feature vector.

For the coefficients obtained from complex valued signals, such as complex coordinate function, the coefficients at the positive and the negative frequency axis are essential. Therefore, the feature vector is formed as:

$$x = \left[\frac{|F_{-(L/2-1)}|}{|F_1|} \cdots \frac{|F_{-1}|}{|F_1|} \frac{|F_2|}{|F_1|} \cdots \frac{|F_{L/2}|}{|F_1|} \right]^T \tag{5.19}$$

in which L is a length of the feature vector. This method is known as *Contour Fourier* method (Kauppinen et al., 1995). It should be noted that the feature vector length can be decided in two ways, either by sampling the boundary function to L samples before DFT, or by selecting L samples from the transform coefficients $F(n)$, provided that $L \leq N$.

For real-valued boundary functions, the Fourier spectrum is symmetrical, i.e. the coefficients $F(n)$ are the same in both the positive and the negative frequency axes. An example of this kind of approach is the *Radius Fourier* method (Kauppinen et al., 1995) that uses the centroid distance function of equation (5.9) as a shape signature:

$$x = \left[\frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \cdots, \frac{|F_L|}{|F_0|} \right]^T \tag{5.20}$$

In the case of this feature vector type, Kauppinen et al. (1995) have used $F(0)$ in scale normalization. This is based on the fact that in the Fourier spectrum obtained from the centroid distance function, $F(0)$ represents the mean of the signal and therefore it can be used as a normalizing factor. The feature vector for area function of equation (5.10) is equivalent to the *Radius Fourier* method.

6 Statistical gray level features

In addition to texture and shape, the distribution of image colors (or gray levels) is an essential feature in content-based image retrieval. In color-based image retrieval, the goal is to find images whose colors are similar to those of a query image. In early retrieval systems, simple statistical measures such as average color and color histogram were employed; whereas recently second order statistics have received growing interest due to their ability to characterize the spatial relationships between the colors in the image, as well. The statistical measures used in image database indexing have been adopted from the areas of image analysis and pattern recognition. Some of the statistical methods are also closely related to texture analysis, which also uses statistical measures to describe textured areas in the image (Conners and Harlow, 1980; Ohanian and Dubes, 1991). On the other hand, in color description the statistical features are typically used to describe the color distribution of the whole image. The statistical measures can be divided into first, second and higher order measures.

6.1 First order statistical measures

First order measures consider the values of image pixels individually. The best-known first order feature is a histogram that characterizes the distribution of the colors or gray levels in the image[3]. Image histogram is a first order statistical measure that has been traditionally used in characterization of the global color distribution of the image. Let an image **I** consist of pixels $p(x,y)$, each pixel corresponding to level g. Thus a set of levels $[G]$ contains the levels $g_1 \ldots g_G$. Let $\mathbf{I}(p)$ correspond to a level of pixel p, and $\mathbf{I}_g$ refers to a pixel, for which $\mathbf{I}(p)=g$. Image histogram h is defined as:

$$h(g) = \frac{l_g}{l}, \;\; g \in [G] \tag{6.1}$$

[3] The statistical measures can be applied to gray levels and color components of the image. Henceforth the term *level* refers to colors and gray levels.

Figure 6.1. Two images with identical histograms.

where l_g and l are the numbers of pixels of level g and all the pixels in the image, respectively. Histogram can also be expressed as a probability of a certain level in the image:

$$h_{g_i}(\mathbf{I}) \equiv \Pr_{p \in \mathbf{I}} \left\lfloor p \in \mathbf{I}_{g_i} \right\rfloor \tag{6.2}$$

Image histogram is a simple and computationally effective statistical measure for the description of image level distribution. The dimensionality of the histogram is equal to the number of levels, G. In image retrieval, histograms are widely used (Flickner et al., 1995; Pentland et al., 1996; Ogle and Stonebraker, 1995). The similarity between histograms can be defined using histogram intersection introduced by Swain and Ballard (1991) or a weighted distance measure used in QBIC system (Hafner et al., 1995). Also L_1 and L_2 distances have been used in histogram matching (Del Bimbo, 2001)

The drawback of the histogram is the fact that it ignores the spatial relationships between the pixels. This is illustrated in figure 6.1, in which two images with totally different spatial organizations have identical histograms. Due to this drawback, the spatial relationships between the levels have been found to be essential in image description. Certain histogram-based techniques have been presented that also consider the spatial organization of the colors. Pass and Zabih (1996) introduced color coherence vector (CCV) that partitions the histogram bins by the spatial coherence of pixels. A pixel is considered to be coherent if it is a part of some similar colored region. Like histograms, CCV's are computationally fast methods, but they outperform histograms in color-based image classification. Rickman and Stonham (1996) used histograms that contain color tuples. Another approach to capture the spatial organization of the colors is the use of color sets (Smith and Chang, 1996b), which partition the image into regions of a certain color. Stricker and Dimai (1996) have used moments to describe the color regions in the image.

6.2 Second order statistical measures

The difference between second order and first order statistical measures is that second order measures utilize the spatial relationships between the pixel values. Hence, in addition to the level difference between two pixels, their spatial organization is also taken into account. Second order measures have traditionally been used in texture analysis, in which texture has been characterized in terms of correlation function or co-occurrence matrix, for example. In addition, correlograms and autocorrelograms have also been used in image retrieval. The use of second order measures in image analysis already started in the 1950s, when Kaizer (1955) applied correlation function to granularity measurement of aerial photographs.

Co-occurrence matrix

Since the early 1970s there has been wider use of second order measures when Haralick (1973) introduced gray level co-occurrence matrix (GLCM). The matrix indicates the joint probability of a gray level occurrence of two pixels at a certain displacement in the image. Let $\mathbf{d}=(d_x, d_y)$ be a displacement vector which determines the distance and direction of levels i and j in the image **I**. The (i,j)th element of the co-occurrence matrix is the number of times that levels i and j occur in the relative position **d** in the image **I**. The matrix is usually normalized, i.e. by the number of all occurrences in the image, #**I**. Haralick (1973) used the matrix to texture description by calculating certain features based on it. The most common of these 14 features are contrast, entropy, energy, correlation, and homogeneity. Shim and Choi (2003) have used the co-occurrence matrix in image retrieval to describe the relationships of color (hue) pairs in an image.

Correlogram

Correlation function has been utilized in image retrieval since the mid-1990s. In image retrieval, the correlation-based features are called correlograms. Huang et al. (1997) presented the first results of correlogram-based color description. In (Huang et al., 1998), they developed the method by introducing methods for querying objects and regions from the images. Ma and Zhang (1998) made a comparison between correlograms, histograms, color moments, and color coherence vectors in color-based image retrieval. The best results were obtained using correlograms. Ojala et al. (2001) proved that a correlogram also performs well in HSI color space. In their further research, HSI correlogram gave almost equal retrieval performance as MPEG-7 color descriptors (Ojala et al., 2002).

Image correlogram represents the correlations between the image pixel values. Hence, its principle is quite similar to co-occurrence matrix. The main difference between them is that instead of using a single displacement vector **d**, the correlogram uses a set of distances. The definition of image correlogram is the following (Huang et al., 1997; Ojala et al., 2001). Let $[D]$ denote a set of fixed distances $d_1 \ldots d_D$. Hence, the number of the distances in this set is D. The correlogram of the image **I** is defined for level pair (g_i, g_j) at a distance d:

$$\gamma^{(d)}_{g_i,g_j}(\mathbf{I}) \equiv \Pr_{p_1 \in I_{g_i}, p_2 \in I} \left\lfloor p_2 \in \mathbf{I}_{g_j} \middle\| p_1 - p_2 \right| = d \right\rfloor \tag{6.4}$$

which gives the probability that given any pixel p_1 of level g_i, a pixel p_2 at a distance d from the given pixel p_1 is of level g_i. In other words, the correlogram is a matrix that gives the probability that a certain level will occur at the distance d from each other. Correlogram is defined for several values of d defined in the set $[D]$. The size of the correlogram-based feature vector is G^2D. In image retrieval and classification, a commonly used distance measure between image correlograms is L_1-norm (Huang et al., 1997).

Autocorrelogram

Correlograms are high dimensional and therefore autocorrelograms have been proposed for use instead of correlograms (Huang et al., 1997; Ojala et al., 2001). Autocorrelogram is the subset of the correlogram. It captures only the spatial correlation of the identical levels. The autocorrelogram can be defined as:

$$\alpha_g^{(d)}(\mathbf{I}) = \gamma_{g,g}^{(d)}(\mathbf{I}) \tag{6.5}$$

and it gives the probability that a pixel p_2, d away from the given pixel p_1, is of level g. In the case of the autocorrelogram, the size of the feature vector is GD.

6.3 Comparisons

The image histogram is the best-known statistical measure with low computational cost. However, several studies (Huang et al., 1997; Ma and Zhang, 1998; Ojala et al., 2001) have shown that its ability to describe image content is inadequate for image retrieval purposes. Therefore, more effective second order statistics have been adopted for use in image retrieval. The correlogram is an accurate measure, but high dimensionality limits its usability in on-line retrieval. Though the autocorrelogram is computationally more efficient, it does not capture the whole probability distribution of the image, which decreases its performance. There is, however, a method for decreasing the dimensionality of the second-order statistics. Kunttu et al. (2003b) divided the images in the areas of similar level. This is near the principle of color sets (Smith and Chang, 1996b). The division was made by re-quantizing the color space of the images which yields a decreased number of image levels G. Such an approach makes it possible decrease the dimensionality of the correlogram significantly. The quantization can also be used to generalize the image content which yields better retrieval performance (Kunttu et al., 2003b).

7 Applications in defect image classification and retrieval

This thesis is a result of research work done in the DIGGER[4] research project in co-operation with industry. The author has been involved in the surface defect image analysis and retrieval part of the project since 2001. The goal was to develop effective and efficient visual descriptors for defect image classification and retrieval. The author's research area was defect image description in terms of shapes and gray level distributions. This thesis presents the main results of the research work undertaken by the author. In the present Chapter, the methods developed as well as their significance in defect image description are discussed.

7.1 The development of shape descriptors

The shape description was based on the boundary line of the defect extracted from its background. The shape description used Fourier-based methods, which were combined with a multiresolution property. This novel method can be achieved using two approaches: using wavelets or iterative boundary smoothing. The proposed multiresolution Fourier descriptors, *Multiscale Fourier* and *Boundary Scale Fourier*, consider the shapes in multiple scales, which make them more insensitive to fine details in the contour than conventional single-scale Fourier descriptors. This improves the shape classification and retrieval performance of the defect images. Previous work in the shape description of defect images has been confined to simple shape descriptors and chain code-based approaches (see section 2.2). The Fourier-based shape characterization is, therefore, a new approach in the field of defect description. By employing the multiresolution methods proposed, the Fourier-based shape description can be further improved. In addition, these methods provide a novel way for effectively describing any kind of complicated shape.

Multiscale Fourier descriptor uses wavelet transform to produce a multiresolution Fourier descriptor. However, a rotation, translation, and scale invariant wavelet-based

[4] Knowledge Mining and Managing in a Distributed Image Datawarehouse

shape descriptor can also be obtained at single scale by applying discrete Fourier transform to the wavelet coefficients of the boundary line. This kind of approach outperforms conventional Fourier description in defect shape retrieval without increasing descriptor dimensionality.

In Fourier-based defect shape description, small defects having short boundary function may cause problems. This is because the Fourier spectrum of short boundary functions does not have adequate frequency resolution. The performance of Fourier-based shape description can be improved by interpolating new values to the Fourier spectrum by using a zero-padding method, which provides a simple and fast way of obtaining a more accurate shape representation in frequency domain.

7.2 The developed gray level descriptors

The proposed description of the gray level distribution of the defect images was based on second order statistics, which have been found to be accurate descriptors in image retrieval. Compared to the former histogram-based defect image description, second order statistics represents a marked improvement in classification accuracy. The drawback with the correlation-based descriptors is the fact that they are computationally heavy methods for on-line image retrieval. Therefore, a new statistical descriptor for gray level characterization was developed. This descriptor, *Binary co-occurrence matrix*, represents the footprint distribution of the co-occurrence matrix. It is a computationally less expensive and more accurate description method than image correlogram. It is also essential that the binary co-occurrence matrix considers all the correlations occurring in the image equally, which means that image segmentation can be avoided in defect image description.

7.3 Combining defect shape and gray level information

In defect image description, it is also possible to combine the shape and gray level information into a single descriptor. In this way, matching and dimensionality problems caused by the use of separate shape and gray level descriptors can be overcome. This is facilitated by using Fourier-based techniques, which are powerful tools for the boundary-based object description. Using a novel technique, the boundary signature is combined with gray level information of the defect and the obtained signal is Fourier transformed. As a result, a new kind of object descriptor *Color Fourier descriptor* is obtained. This descriptor is as low dimensional and easy to match as any other shape-based Fourier descriptor. Such a descriptor provides an innovative approach to combining the shape and gray level information of the defect image.

7.4 Summary of publications

Publication I: A novel shape descriptor, *Multiscale Fourier* descriptor is presented. This descriptor is formed by applying Fourier transform to the coefficients of wavelet transform of the object boundary. In this way the Fourier descriptor can be presented in multiple resolutions. Classification experiments are carried out using paper and metal image databases. In addition, a set of general shapes is used. The classification results of *Multiscale Fourier* descriptor are compared to those of Fourier descriptors.

Publication II: The *Multiscale Fourier* descriptor is applied to the shape-based retrieval of paper and metal defect images. The retrieval results show that *Multiscale Fourier* is capable of outperforming conventional Fourier descriptors as well as CSS descriptors in defect image retrieval.

Publication III: In addition to *Multiscale Fourier* descriptor, another novel technique for Fourier-based multiresolution shape description for defect image retrieval is presented. This technique uses boundary smoothing (*Boundary Scale Fourier*) to produce the multiresolution property to the Fourier descriptor. The experimental results show that *Multiscale Fourier* and *Boundary Scale Fourier* descriptors outperform the most powerful single-scale Fourier descriptor, *Contour Fourier* in paper defect image retrieval. This can be done without increasing computational cost of retrieval, i.e. with feature vectors of equal dimensionality.

Publication IV: A wavelet-based shape descriptor for defect shapes at single scale is presented. In contrast to *Multiscale Fourier*, the proposed combination of wavelet and Fourier shape description uses single scale, which yields the same dimensionality as ordinary Fourier descriptors. However, the classification and retrieval experiments with paper defect shapes reveal that the proposed single scale descriptor clearly outperforms the most powerful single-scale Fourier descriptor, *Contour Fourier* method.

Publication V: A method for improving Fourier-based shape description of objects with short boundary line is presented. The frequency resolution of the Fourier spectrum calculated for the boundary function can be easily increased by using a zero-padding method. This method is used to improve the frequency resolution by adding zeros to the boundary function to be Fourier transformed. Consequently, new points are being interpolated to the spectrum. The experimental results show that by applying the zero-padding method, defect shape description can be easily improved in classification tasks.

Publication VI: A new statistical measure for the characterization of gray level distribution of the defect images is presented. This measure, *Binary co-occurrence matrix* is capable of accurate defect image description and it clearly outperforms image correlogram in the retrieval of paper and metal defect images.

Publication VII: The principle of *Color Fourier* descriptor is presented. In the proposed approach, the gray level distribution of the object is added to the Fourier-based contour description using area function. The Fourier descriptor using gray level and shape outperforms clearly shape based Fourier descriptors in defect image retrieval. However, the color information does not increase the dimensionality of the proposed descriptor.

7.5 Author's contributions to the publications

In all the above publications, the author has played the major role in developing the proposed methods. In all of them, the author developed the proposed defect image description methods and carried out the experiments with Ms. Leena Lepistö. In all but publications IV and V, Mr. Juhani Rauhamaa from ABB Oy has been a co-author responsible for the sections describing the surface defect images. Professor Ari Visa has acted as supervisor to verify the methods used in the publications.

8 Conclusions

Content-based image retrieval is nowadays an area of active research, which has generated a large amount of published research work and results. The main focus in the present study is the field of content-based image indexing. This has necessitated considerable research effort to find effective features for describing the content of the images to be retrieved. The majority of the existing content-based image retrieval systems have been developed for general image archives, such as newspaper images or consumer photographs. However, current machine vision systems are capable of producing vast amounts of images that are stored into databases. This has given rise to a growing need for image retrieval systems in industrial applications.

In an industrial process, different measurements play an essential role in the process and quality control. Visual inspection is one part of this measurement process. Nowadays, image-based automatic inspection has replaced manual inspection in several industrial areas. Industrial imaging systems are on-line applications, which set their own special requirements for retrieval system. In addition, the content of industrial images is often very specific and the images cannot be divided in a straightforward way into categories by human skill, like most of the photographs used in general image archives. For these reasons, the industrial defect image databases considered in this thesis constitute a challenging practical image retrieval task.

In practical defect inspection, the basic problem for the user is to find images of a certain defect type in the database. This is performed by providing an example image of the desired defect type. To satisfy this information requirement, the retrieval system should recall a set of defects that are as similar to the query image as possible. Therefore, effective visual descriptors that are able to describe the image content are necessary. Improved defect description provides a better understanding of the defect causes. This is because the information of the retrieved defect image can be combined with the knowledge of the defect location and other process measurements. It is, therefore, essential that the visual descriptors are capable of describing the special character of various defect types.

The main contribution of this thesis is the development of effective and computationally efficient visual descriptors for defect image classification and retrieval. The descriptors selected for study in this research consider shape and gray level distribution of the defect images. These features have not been studied earlier to the same extent as, for example, the texture features of the surface defects. The shape features developed in this thesis are based on well-known and effective boundary description that uses Fourier transform. However, a novel approach to the characterization of complicated shapes is to use multiresolution shape description based on Fourier description. The multiresolution property is capable of significantly increasing the classification accuracy of the defects. Moreover, the computational costs of the proposed multiresolution approaches are reasonable which makes them suitable for on-line retrieval. The gray level distribution of the defect images is described using a novel second order statistical measure, binary co-occurrence matrix. The experimental results show that its accuracy in defect image retrieval is higher than that of conventional correlation-based statistics. The developed methods are directly applicable in real defect image retrieval tasks.

Bibliography

Abbasi, S., Mokhtarian, F., Kittler, J., 1999. Curvature scale space image in shape similarity retrieval. Multimedia Systems 7(6), 467-476.

Abbasi, S., Mokhtarian, F., Kittler, J., 2000. Enhancing CSS-based shape retrieval for objects with shallow concavities. Image and Vision Computing 18(3), 199-211.

Arbter, K., 1989. Affine invariant Fourier descriptors. In: J.C. Simons (Ed.), From Pixels to Features, Elsevier Science Publishers, Amsterdam, 153-164.

Arbter, K., Snyder, W.E., Burkhardt, K., Hirzinger, G., 1990. Application of affine invariant Fourier descriptors to recognition of 3D-objects. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(7), 640-647.

Asada, H., Brady, Brady, M., 1986. The curvature primal sketch. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(1), 2-14.

Baeza-Yates, R., Ribeiro-Neto, B., 1999. Modern information retrieval. ACM Press, Addison-Wesley, New York.

Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(24), 509-522.

Bigun, J., 2005. Vision with direction. Springer.

Bribiesca, E., 1999. A new chain code. Pattern Recognition 32(2), 235-251.

Carlin, M., 2001. Measuring the performance of shape similarity retrieval methods. Computer Vision and Image Understanding 84(1), 44-61.

Chang, C.C., Hwang, S.M., Buehrer, D.J., 1991. A shape recognition scheme based on relative distances of feature points from the centroid. Pattern Recognition 24(11), 1053-1063.

Chang, N.-S., Fu, K.-S., 1980. Query-by-pictorial-example. IEEE Transactions on Software Engineering 6(6), 519-524.

Chang, S.K., Chi, Q.Y., Yan, C.W., 1987. Iconic indexing by 2D strings. IEEE Transactions on Pattern Analysis and Machine Intelligence 9(3), 413-428.

Chang, S.K., Yan, C.W., Dimitroff, D.C., Arndt, T., 1988. An intelligent image database system. IEEE Transactions on Software Engineering 14(5), 681-688.

Chang, S.K., Hsu, A., 1992. Image information systems: Where do we go from here? IEEE Transactions on Knowledge and Data Engineering 4(5), 431-442.

Cheikh, F.A., 2004. MUVIS: A system for content-based image retrieval. Ph.D. Thesis, Tampere University of Technology, Tampere, Finland.

Chen, G., Bui, D., 1999. Invariant Fourier-wavelet descriptor for pattern recognition. Pattern Recognition 32(7), 1083-1088.

Chen, X., 2004. Defect inspection of inhomogenous steel surfaces. Steel Grips 2(2), 109-112.

Chui, C.K., 1992. An introduction to wavelets. Academic Press, San Diego.

Conners, R., Harlow, C., 1980. A theoretical comparison of texture algorithms. IEEE Transactions on Pattern Analysis and Machine intelligence 2(3), 204-222.

Costa, L.F., Cesar, R.M., 2001. Shape analysis and classification, Theory and Practice. CRC Press, Boca Raton, Florida.

Del Bimbo, A., Pala, P., 1997. Visual image retrieval by elastic matching of user sketches. IEEE Transactions on Pattern Analysis and Machine intelligence 19(2), 121-132.

Del Bimbo, A., 2001. Visual information retrieval. Morgan Kaufmann Publishers, San Fransisco.

Dubois, S.R., Glanz, F.H., 1986. An autoregressive model approach to two-dimensional shape classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(1), 55-66.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern classification. 2nd edition, John Wiley & Sons.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P., 1995. Query by image and video content: the QBIC system. IEEE Computer 28(9), 23-32.

Freeman, H., 1961. On the encoding of arbitrary geometric configurations. IRE Transactions on Electronic Computing EC-10, 260-268.

Freeman, H., Davis, L.S., 1977. A corner finding algorithm for chain coded curves. IEEE Transactions on Computers 26, 297-303.

Freeman, H., Saghri, A., 1978. Generalized chain codes for planar curves. In Proceedings of 4th International Joint Conference on Pattern Recognition, pp. 701-703.

Gevers, T., Smeulders, A.W.M., 2000. Pictoseek: Combining color and shape invariant features for image retrieval. IEEE Transactions on image Processing 9(1), 102-119.

Gonzalez, R.C., Woods, R.E., 1993. Digital image processing. Addison Wesley.

Goshtasby, A., 1985. Description and discrimination of planar shapes using shape matrices. IEEE Transactions on Pattern Analysis and Machine intelligence 7(6), 738-743.

Granlund, G.H., 1972. Fourier preprocessing for hand print character recognition. IEEE Transactions on Computers C-21(2), 195-201.

Groskey, W.I., Mehrotra, R., 1990. Index-based object recognition in pictorial data management. Computer Vision, Graphics and Image Processing 52(3), 416-436.

Groskey, W.I., Neo, P., Mehrotra, R., 1992. A Pictorial index mechanism for model-based matching. Data and Knowledge Engineering 8(4), 309-327.

Gudivada V.N., Raghavan, V.V., 1995. Content-based image retrieval systems. IEEE Computer 28(9), 18-22.

Gupta, A., Jain, R., 1997. Visual information retrieval. Communications of the ACM 40(5), 70-79.

Hafner, J., Sawhney, H.S., Equitz, W., Flickner, M., Niblack, W., 1995. Efficient color histogram indexing for quadratic form distance functions. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(7), 729-736.

Hand, D., Mannila, H., Smyth, P., 2001. Principles of data mining. MIT Press, Massachusetts.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-3(6) 610-621.

Henkemeyer, H., 2003. Long-term stability of surface inspection systems and resulting technical requirements. Steel Grips 1(1), 12-16.

Hu, M.K., 1962. Visual pattern recognition by moment invariants. IRE Transactions on Information Theory IT-8, 179-187.

Huang, J., Kumar, S.R., Mitra, M., Zhu, W.-J., Zabih, R., 1997. Image indexing using color correlograms. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, pp. 762-768.

Huang, J., Kumar, S.R., Mitra, M., Zhu, W.-J., 1998. Spatial color indexing and applications. In Proceedings of 6th International Conference on Computer Vision, Bombay, India, pp. 602-607.

Iivarinen, J., Visa, A., 1996. Shape recognition of irregular objects. In: D. P. Casasent (Ed.), Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling, Proceedings of SPIE, Vol. 2904, pp. 25-32.

Iivarinen, J., Rauhamaa, J., Visa, A., 1996. Unsupervised segmentation of surface defects. In Proceedings of 13th International Conference on Pattern Recognition, Wien, Austria, Vol. 4, pp. 356-360.

Iivarinen, J., 1998. Texture segmentation and shape classification with histogram techniques and self-organizing maps. Ph.D. Thesis, Helsinki University of Technology, Espoo, Finland.

Iivarinen, J., Visa, A., 1998. An adaptive texture and shape based defect classification. In Proceedings of 14th International Conference on Pattern Recognition, Brisbane, Australia, Vol. 1, pp. 117-122.

Iivarinen, J., Pakkanen, J., Rauhamaa. J., 2004a. SOM-based system for web surface inspection. In J. R. Price, F. Mériaudeau (Eds.), Machine Vision Applications in Industrial Inspection XII, Proceedings of SPIE, Vol. 5303, pp. 178-187.

Iivarinen, J., Rautkorpi, R., Pakkanen, J., Rauhamaa. J., 2004b. Content-based retrieval of surface defect images with PicSOM. International Journal of Fuzzy Systems 6(3), 160-167.

Kaizer, H., 1955. A quantification of textures on aerial photographs. Tech. Note No. 121, A.69484, Boston University Research Laboratories, Boston University.

Kashi R.S., Kavde, P., Nowakowski, R.S., Papathomas, T.V., 1996. 2-D shape representation and averaging using normalized wavelet descriptors. Simulation 66(3), 164-177.

Kashyap, R.L., Chellappa, R., 1981. Stochastic models for closed boundary analysis: representation and reconstruction. IEEE Transactions on Information Theory 27(5), 627-637.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: active contour models. International Journal of Computer Vision 3(6), 321-331.

Kauppinen, H., Seppänen, T., Pietikäinen, M., 1995. An experimental comparison of autoregressive and fourier-based descriptors in 2D shape classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(2), 201-207.

Khalil, M.I., Bayoumi, M.M., 2001. A dyadic wavelet affine invariant function for 2D shape recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(10), 1152-1163.

Kim, H., Kim, J., 2000. Region-based shape descriptor invariant to rotation, scale and translation. Signal Processing and Image Communication 16(1), 87-93.

Kim, S.D., Lee, J.H., Kim, J.K., 1988. A new chain-coding algorithm for binary images using run-length codes. Computer Vision, Graphics and Image Processing 41(1), 114-128.

Kumar, A., 2003. Neural network based detection of local textile defects. Pattern Recognition 6(7), 1645-1659.

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2002. Image retrieval without segmentation. In Proceedings of 10th Finnish AI Conference, Oulu, Finland, pp. 164-169.

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2003a. Binary histogram in image classification for retrieval purposes. Journal of WSCG 11(2).

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2003b. Image correlogram in image database indexing and retrieval. In Proceedings of 4th European Workshop on Image Analysis for Multimedia Active Services, London, UK, pp. 88-91.

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2003c. Classification method for defect images based on association and clustering. Data Mining and Knowledge Discovery: Theory, Tools, and Technology V, Proceedings of SPIE, Vol. 5098, Orlando, Florida, pp. 19-27.

Laaksonen, J.T., Koskela, J.M., Laakso, S.P., Oja, E., 2000. PicSOM – content-based image retrieval with self-organizing maps. Pattern Recognition Letters 21(13-14), 1199-1207.

Landry, H.J., 2000. Sheet hole and defect identification. TAPPI Journal 83(5), 64-67.

Li, Q., Edwards, J., 2004. An enhanced normalization technique for wavelet shape descriptors. In Proceedings of 4th Conference on Computer and Information Technology, pp. 722-729.

Liao, S.X., Pawlak, M., 1996. On image analysis by moments. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(3), 254-266.

Lin, C.C., Chellappa, R., 1987. Classification of partial 2D shapes using Fourier descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 9(5), 686-690.

Loncaric, S., 1998. A survey of shape analysis techniques. Pattern Recognition 31(8), 983-1001.

Lu, G., Sajjanhar, A., 1999. Region-based shape representation and similarity measure suitable for content-based image retrieval. Multimedia Systems 7(2), 165-174.

Ma, W.Y., Manjunath, B.S., 1996. Texture features and learning similarity. In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 425-430.

Ma, W.Y., Zhang, H.J., 1998. Benchmarking of image features for content-based retrieval. In Proceedings of 32nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, Vol. 1, pp. 253-257.

Manjunath, B.S., Salembier, P., Sikora, T., 2002. Introduction to MPEG-7 multimedia content description interface. John Wiley & Sons, UK.

Mehrotra, S., Rui, Y., Ortega, M., Huang, T.S., 1997. Supporting content-based queries over images in MARS. In Proceedings of IEEE International Conference on Multimedia Computing and Systems, Ottawa, Canada, pp. 632-633.

Mehtre, B.M., Kankanhalli, M.S., Lee, W.F., 1997. Shape measures for content-based image retrieval: a comparison. Information Processing Management 33(3), 319-337.

Mitchell, O.R., Grogan, T.A., 1984. Global and partial shape discrimination for computer vision. Optical Engineering 23(5), 484-491.

Mokhtarian, F., Mackworth, A.K., 1986. Scale-based description and recognition of planar curves and two-dimensional shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(1), 34-43.

Mokhtarian, F., Mackworth, A.K., 1992. A theory of multiscale, curvature-based shape representation for planar curves. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(8), 789-805.

Myers, D.G., 1990. Digital signal processing. Prentice-Hall, New York.

Niskanen, M., Silvén, O., Kauppinen, H., 2001. Color and texture based wood inspection with non-supervised clustering. In Proceedings of 12th Scandinavian Conference on Image Analysis, Bergen, Norway, pp. 336-342.

Ogle, V., Stonebraker, M., 1995. Chabot: retrieval from a relational database of images. IEEE Computer 28(9), 40-48.

Ohanian, P.P., Dubes, R.C., 1991. Performance evaluation for four classes of textural features. Pattern Recognition, 25(8), 819-833.

Ohm, J.R., Bunjamin, F.B., Liebsch, W., Makai, B., Muller, K., Somlic, A., Zier, D., 2000. A set of visual feature descriptors and their combination in a low-level description scheme. Signal Processing and Image Communication 16(1-2), 157-179.

Ojala, T., Rautiainen, M., Matinmikko, E., Aittola, M., 2001. Semantic image retrieval with HSV correlograms. In Proceedings of 12th Scandinavian Conference on Image Analysis. Bergen, Norway, pp. 621-637.

Ojala, T., Aittola, M., Matinmikko, E., 2002. Empirical evaluation of MPEG-7 XM color descriptors in content-based retrieval of semantic image categories. In Proceedings of 16th International Conference on Pattern Recognition, Quebec, Canada, Vol. 2, pp. 1021-1024.

Orfanidas, S.J., 1996. Introduction to signal processing, Prentice Hall

Pakkanen, J., Ilvesmäki, A., Iivarinen, J., 2003. Defect image classification and retrieval with MPEG-7 descriptors. In Proceedings of 13th Scandinavian Conference on Image Analysis, Göteborg, Sweden, LNCS 2479, pp. 349-355.

Pass, G., Zabih, R., 1996. Histogram refinement for content-based image retrieval. In Proceedings of IEEE Workshop on Applications of Computer Vision, pp. 96-102.

Pavlidis, T., 1986. Editoral. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(1).

Pentland, A., Picard, R.W., Schlaroff, S., 1996. Photobook: Content-based manipulation of image databases. International Journal on Computer Vision 18(3), 233-254.

Persoon, E., Fu, K., 1977. Shape discrimination using Fourier descriptors. IEEE Transactions on Systems, Man, and Cybernetics 7, 170-179.

Pfeiffer, M., Pandit, M., 1995. A parametrical description of plane curves using wavelet descriptors. In Proceedings of ICASSP Conference on Acoustics, Speech, and Signal Processing, Detroit, MI, Vol. 4, pp. 2531-2534.

Ramer, U., 1972. An iterative procedure for the polygonal approximation of plane curves. Computer Graphics and Image Processing 1, 244-256.

Rauhamaa, J., Reinius, R., 2002. Paper web imaging with advanced defect classification. In Proceedings of 2002 TAPPI Technology Summit, Atlanta, Georgia.

Rauhamaa, J., Järvinen, J., 2002. New challenges of paper inspection. In Proceedings of Asian Paper Conference, New Applied Technology Conference, Singapore.

Rauhamaa, J., 2004. Paper defect classification with neural networks. Machine Vision News 9, Published by Finnish Society of Automation.

Rautkorpi, R., Iivarinen, J., 2004. A novel shape feature for image classification and retrieval. In Proceedings of International Conference on Image Analysis and Recognition, Porto, Portugal, LNCS 3211, Part I, pp. 753-760.

Rickman, R., Stonham, J., 1996. Content-based image retrieval using color tuple histograms. In Storage and Retrieval for Still Image and Video Databases IV, Proceedings of SPIE, Vol. 2670, pp. 2-7.

Rui, Y., Chang, T.S., Chang, S.F., 1997. Image retrieval: past, present, and future. In Proceedings of International Symposium on Multimedia Information Processing, Taipei, Taiwan.

Rui, Y., Huang, T.S., Chang, S.F., 1999. Image retrieval: current techniques, promising directions, and open issues. Journal of Visual Communication and Image Representation 10(1), 39-62.

Safar, M., Shahabi, C., Sun, X., 2000. Image retrieval by shape: a comparative study. In Proceedings of IEEE International Conference on Multimedia and Expo, pp. 141-144.

Santini, S., Jain, R., 1999. Similarity measures. IEEE Transactions on Pattern Analysis and Machine intelligence 21(9), 871-883.

Shen, D., Ip, H.H.S., 1999. Discriminative wavelet shape descriptors for recognition of 2-D patterns. Pattern Recognition 32(2), 151-165.

Shim, S.-O., Choi, T.-S., 2003. Image indexing by modified color co-occurrence matrix. In Proceedings of International Conference on Image Processing, Barcelona, Spain, Vol. 3, pp. 493-496.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R., 2000. Content-based image retrieval at the end of the early years. IEEE Transactions Pattern Analysis and Machine Intelligence 22(12), 1349-1380.

Smith, J.R., Chang, S.-F., 1996a. Visualseek: a fully automated content-based image query system. In Proceedings of ACM Multimedia 96.

Smith, J.R., Chang, S.-F., 1996b. Tools and techniques for color image retrieval. Storage and Retrieval for Image and Video Databases IV, Proceedings of SPIE, Vol. 2670, pp. 1630-1639.

Squire, D.M., Caelli, T.M., 2000. Invariance signature: characterizing contours by their departures from invariance. Computer Vision and Image understanding 77(3), 284-316.

Stolzenberg, M., Gruber, C., Henkermeyer, H., Jonker, K., Geisler, S., 2003. Surface inspection for flat steel production. MPT International 26(1), 62-71.

Stricker, M., Orengo, M., 1995. Similarity of color images. Storage and Retrieval for Image and Video Databases III, Proceedings of SPIE, Vol. 2420, pp. 381-392.

Stricker, M., Dimai, A., 1996. Color indexing with weak spatial constraints. In Storage and Retrieval for Still Image and Video Databases IV, Proceedings of SPIE, Vol. 2670, pp. 29-40.

Swain, M.J., Ballard, D.H., 1991. Color indexing. International Journal of Computer Vision 7(1), 11-32.

Tamura, H., Yokoya, N., 1984. Image database systems: a survey. Pattern Recognition 17(1), 29-43.

Teague, M.R., 1980. Image analysis via the general theory of moments. Journal of Optical Society of America 70(8), 920-930.

Teh C.-H., Chin, R.T., 1988. On image analysis by the methods of moments. IEEE Transactions on Pattern Analysis and Machine Intelligence 10(4), 496-513.

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

Tsang W.M., Yuen, P.C., Lam, F.K., 1994. Detection of dominant points on an object boundary: a discontinuity approach. Image and Vision Computing 12(9), 547-557.

Veltkamp, R.C., 2001. Shape matching: similarity measures and algorithms. In Proceedings of International Conference on Shape Modeling and Applications, pp. 188-197.

Wang, S., Chen, P., Lin, W., 1994. Invariant pattern recognition by moment Fourier descriptor. Pattern Recognition 27(12), 1735-1742.

Widrow, B., 1973. The rubber mask technique-II – pattern storage and recognition. Pattern Recognition 5(3), 199-211.

Witkin, A.P., 1983. Scale space filtering. In proceedings of 8th international Joint Conference on Artificial Intelligence, pp. 1019-1022.

Yang, H.S., Lee, S.U., Lee, K.M., 1998. Recognition of 2D object contours using starting-point-independent wavelet coefficient matching. Journal of Visual Communication and Image Representation 9(2), 171-181.

Zahn, C.T., Roskies, R.Z., 1972. Fourier descriptors for plane closed curves, IEEE Transactions on Computers C-21, 269-281.

Zhang, D.S., Lu, G., 2001. A comparison of shape retrieval using Fourier descriptors and short-time Fourier descriptors. In Proceedings of 2nd IEEE Pacific-Rim Conference on Multimedia, pp. 855-860.

Zhang, D.S., Lu, G., 2002. A comparative study of Fourier descriptors for shape representation and retrieval. In Proceedings of the Asian Conference on Computer Vision, pp. 646-651.

Zhang, D.S., Lu, G., 2003a. A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval. Journal of Visual Communication and Image Representation 14(1), 41-60.

Zhang, D.S., Lu, G., 2003b. Evaluation of MPEG-7 shape descriptors against other shape descriptors. Multimedia Systems 9(1), 15-30.

Zhang, D.S., Lu, G., 2004. Review of shape representation and description techniques. Pattern Recognition 37(1), 1-19.

Zhang, D.S., Lu, G., 2005. Study and evaluation of different Fourier methods for image retrieval. Image and Vision Computing 23(1), 33-49.

Publications

Publication I

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2003. Multiscale Fourier Descriptor for Shape Classification.

Multiscale Fourier Descriptor for Shape Classification

Iivari Kunttu[1], Leena Lepistö[1], Juhani Rauhamaa[2], and Ari Visa[1]
[1]Tampere University of Technology
Institute of Signal Processing
P. O. Box 553, FIN-33101 Tampere, Finland
[2]ABB Oy
Paper, Printing, Metals & Minerals
P. O. Box 94, FIN-00381 Helsinki, Finland

Iivari.Kunttu@tut.fi

Abstract

The description of the object shape is an important characteristic of the image. In image processing and pattern recognition, several different shape descriptors are used. In human visual perception, the shapes are processed in multiple resolutions. Therefore multiscale shape representation is essential in the shape based image classification and retrieval. In the description of the object shape, the multiresolution representation provides also additional accuracy to the shape classification.

In this paper we introduce a new descriptor for shape classification. This descriptor is called multiscale Fourier descriptor, and it combines the benefits of Fourier descriptor and multiscale shape representation. This descriptor is formed by applying Fourier transform to the coefficients of wavelet transform of the object boundary. In this way the Fourier descriptor can be presented in multiple resolutions.

We make classification experiments using three image databases. The classification results of our method are compared to those of Fourier descriptors.

1. Introduction

The description of the object shape is an important task in image analysis and pattern recognition. The shapes occurring in the images have also a remarkable significance in image retrieval [4]. The basic problem in shape classification is to define the similarity between two shapes. In many cases, this similarity measurement should obey the human shape perception. Images can be classified based on their shape content using different types of shape descriptors [3]. In the field of psychophysics, it has been found that the human visual system processes and analyzes image information at different resolutions. Therefore multiscale shape representation is essential in the classification of the shapes occurring in the images.

Several methods for shape description have been developed. The shape description techniques can be divided into two types, boundary based and region based techniques [7]. The region based methods consider the whole area of the object whereas the boundary based methods concentrate merely on the object boundary line. In this paper we consider the boundary based methods. The most common boundary based shape descriptors are chain codes [6] and Fourier descriptors [15]. Also autoregressive (AR) [5],[12] models have been used to represent the boundaries of curves. During recent years, curvature scale-space (CSS) [14] shape representation has also been widely used. Kauppinen et al. [11] made a comparison between autoregressive models and Fourier-based descriptors in shape classification. In this comparison, Fourier-descriptors proved to be the best in the classification of different shapes. In the comparison made by Mehtre et al. [13], retrieval ability of chain codes, Fourier-descriptors, and different moments were compared in shape similarity-based retrieval. In this case, the best results were obtained using moments and Fourier-descriptors, whereas the lowest retrieval results were given by chain codes.

In this paper, we apply wavelet transform to the classification of different shapes. Wavelet transform has been widely used in multiscale image and signal analysis. It is used for example in image and signal compression and noise reduction. However, wavelet transform has only a few applications in the field of shape description. In [1] Chuang and Kuo used one-dimensional discrete periodized wavelet transform (DPWT) to describe planar curves. The same transform was used also in [9], in which the method was made rotation invariant.

In contrast to these studies, we use complex wavelet transform. In our approach, the boundary of the object is presented in complex form like in the case of Fourier descriptors. Using complex wavelet transform, the multiscale representation of the shape can be achieved. The multiscale Fourier descriptor is obtained by applying the Fourier transform to the coefficients of the multiscale wavelet transform. Using Fourier transform, the wavelet

Proceedings of the 12th International Conference on Image Analysis and Processing (ICIAP'03)
0-7695-1948-2/03 $17.00 © 2003 IEEE

coefficients can be presented in frequency domain, which makes the descriptor invariant for rotation and starting point of the boundary. The use of Fourier transform makes it also possible to present the descriptor in fixed length, independent on the length of the object boundary.

In section two, the general principle of Fourier descriptors and complex wavelet transform are presented. In the same section we show how the multiscale Fourier descriptor for object shape can be formed using complex wavelet transform. The classification ability of the descriptors is tested in section three using three sets of testing images. The results of the classification are discussed in section four.

2. Shape descriptors

The use of Fourier descriptors is common in pattern recognition and image analysis. The benefits of the Fourier descriptors are invariance to the starting point of the boundary and rotation [13]. However, the use of Fourier-based multiscale representation of shape is a new application in shape representation. In this section we present a multiscale shape descriptor based on the complex wavelet transform and Fourier transform. It is a simple descriptor that combines the benefits of Fourier representation of the object shape and the multiresolution nature of wavelet transform.

2.1. Representation of the object boundary

In this paper the shape description methods are based on the boundary of the object. Therefore, the boundary of the object has to be extracted from the image. In the presentation of the object boundary, we use the complex coordinate function [11]. This function is simply the coordinates of the boundary pixels in an object centered coordinate system, represented as complex numbers:

$$z(k) = (x_k - x_c) + j(y_k - y_c) \tag{1}$$

for k=0,1,2,…,N-1, in which N is the length of the boundary and (x_c, y_c) is the centroid of the object. Using complex coordinate function, the boundary can be represented independent on the location of the object in the image. In this way the translation invariance can be achieved.

2.2. Fourier descriptors

The shape descriptor based on the object boundary can be formed in several ways. Fourier transform [15] is the most common method for this purpose. Fourier transformation of a boundary function generates a set of complex numbers which are called Fourier descriptors. Fourier descriptors characterize the object shape in frequency domain. The Fourier descriptors can be formed for a complex boundary using discrete Fourier transform (DFT) [7]. Fourier transform of $z(k)$ is:

$$F(n) = \frac{1}{N}\sum_{k=0}^{N-1} z(k) e^{-j2\pi nk/N} \tag{2}$$

for n =0,1,2,…,N-1. The general shape of the object is represented by the lower frequency descriptors, whereas high frequency descriptors represent the small details of the object shape. A common approach to shape classification is to use only a subset of the descriptors. These subsets can be formed in several different ways. Kauppinen et al. [11] have compared *Curvature Fourier*, *Radius Fourier*, *Contour Fourier*, and *A-invariant* methods for Fourier-based shape representation. According to their experimental results, *Contour Fourier* and *A-invariant* methods were best approaches in shape classification. In this work, we selected *Contour Fourier* method for testing purposes.

The *Contour Fourier* method makes the Fourier transform directly for the complex coordinate function of the object boundary. In this method, the descriptors are taken both positive and negative frequency axis. The scaling of the descriptors is made by dividing the absolute values of the selected descriptors by the absolute value of the first non-zero component. The feature vector for this method is:

$$x = \left[\frac{|F_{-(L/2-1)}|}{|F_1|} \cdots \frac{|F_{-1}|}{|F_1|} \frac{|F_2|}{|F_1|} \cdots \frac{|F_{L/2}|}{|F_1|} \right]^T \tag{3}$$

in which L is a constant value that defines the number of the samples selected from the Fourier coefficients.

2.3. Complex wavelet transform

The multiscale representation of the object boundary can be achieved using wavelet transform. The boundary function is transformed using some wavelet Ψ. Complex wavelet transform is based on the continuous wavelet transform (CWT) [2]. In CWT, the wavelet coefficient of the boundary $z(k)$ at a scale a and position b is defined by:

$$C_a(b) = \frac{1}{\sqrt{|a|}} \int_R z(k) \psi\left(\frac{k-b}{a}\right) dk \tag{4}$$

As in the Fourier transform, also in case of CWT we obtain a set of complex coefficients $C_a(b)$ of scale a. The coefficients are defined for all the positions b=0,1,2,...,N-1.

2.4. Multiscale Fourier descriptor

The problem with the coefficients obtained from the complex wavelet transform is the fact that they are dependent on the starting point of the object boundary. Also the length of the feature vector depends on the length of the object boundary. Therefore, the coefficient vectors of different shapes cannot directly be matched in the image classification. The solution for this problem is to apply the Fourier transform to the coefficients obtained from the complex wavelet transform. In this way the multiscale shape representation can be transformed to the frequency domain. As a result, a multiscale Fourier descriptor is obtained. The descriptor is formed by applying the discrete Fourier transform of equation 2 to the set of complex coefficients $C_a(b)$:

$$F^a(n) = \frac{1}{N}\sum_{b=0}^{N-1} C_a(b) e^{-j2\pi nb/N} \qquad (5)$$

The multiscale descriptor x^a of each scale a is then formed from coefficients $F^a(n)$ using *Contour Fourier* method presented in equation 3:

$$x^a = \left[\frac{\left|F^a_{-(L/2-1)}\right|}{\left|F^a_1\right|} \cdots \frac{\left|F^a_{-1}\right|}{\left|F^a_1\right|} \frac{\left|F^a_2\right|}{\left|F^a_1\right|} \cdots \frac{\left|F^a_{L/2}\right|}{\left|F^a_1\right|} \right]^T \qquad (6)$$

The multiscale representation of the object shape can then be formed by defining the descriptor x^a using several different scales a, and combining the descriptors into a single feature vector, **FW** of length R. Let the set of scales be $A=\{a_1, a_2,..., a_r\}$. So the number of the scales in the descriptor is r.

3. Classification experiments

In this section, we make classification experiments using our method, multiscale Fourier descriptor. The classification results are compared to those of *Contour Fourier* approach.

3.1. Testing databases

For testing purposes, we used three image databases. Two of these databases were industrial defect image databases, which are quite difficult to classify. However, in these images, shape is one essential classifying feature and therefore these databases are used in the experimental part of this paper. In addition to these industrial databases, we had also a database of very simple shapes. Using these three databases, we can show that our method can be used in the shape-based classification of several different image databases.

Testing database I consisted of paper defect images. The images were taken from the paper manufacturing process using a paper inspection system [16]. The defects occurring in the paper can be for example holes, wrinkles or different kinds of dirt spots. The test set consisted of 1204 paper defects, which represented 14 defect classes so that each class consisted of 27-103 images. An example image of each class is presented in figure 1. Within the classes, there were differences in the size and orientation of the defects. This fact can be seen in figure 2, in which the variations of the defect class 1 are presented.

The second industrial image set, testing database II, contained 1943 metal defect images. Also this database contained 14 defect classes. In each class, there were 100-165 images. Figure 3 presents an example of each class and the variations in the defect class 1 are presented in figure 4.

Different defect types in both industrial databases can be distinguished using their shape or gray level. In this paper we concentrate on the shape information of the defects. The classification of the defect images is a demanding task, because in some classes the shapes are very similar. In the case of some defect classes, the shapes are also overlapping, which reduces the classification. The defects can be extracted from their background using an image segmentation method presented in [10].

The third test set, testing database III, consisted of 30 image classes selected from the MPEG-7 image database. Each class contained 20 images, so that the size of the whole testing database was 600 images. The images were silhouettes of some simple objects. An example of each image class in testing database III is presented in figure 5. In each class, the images were variations of the same object. In these images, shape, size, and orientation are varying. An example of the variations within the class "deer" is presented in figure 6. In all images, the object is deer, but the size, shape, and orientation of the deer varies significantly.

3.2. Classification

The database images were indexed by calculating the feature vector x^a for them. The selected wavelet ψ was complex gaussian wavelet of order two. The multiscale presentation was achieved using a set of three scales. The scale sets A were selected to be [10,15,20], [10,20,30], and [50,80,110] for testing databases I, II, and III, respectively. For comparison, also the feature vector x of *Contour Fourier* method was calculated for each test set image.

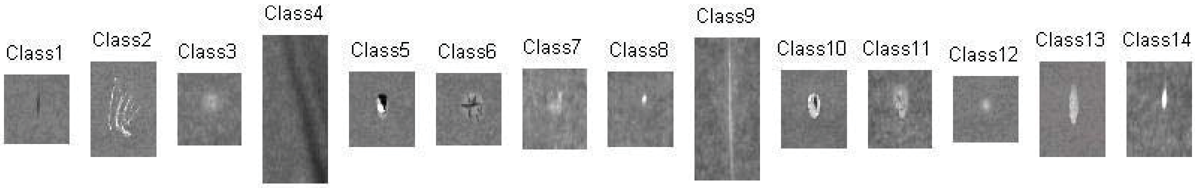

Figure 1. The example images of each paper defect image class of testing database I.

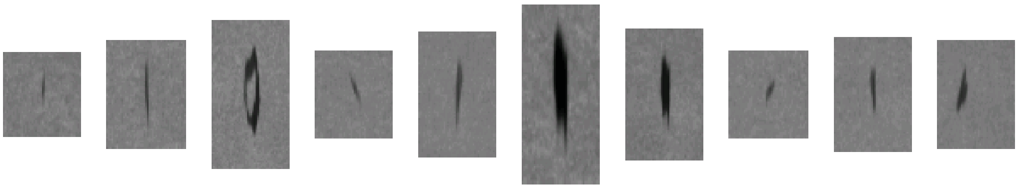

Figure 2. 10 examples of class 1 paper defect images in the testing database I.

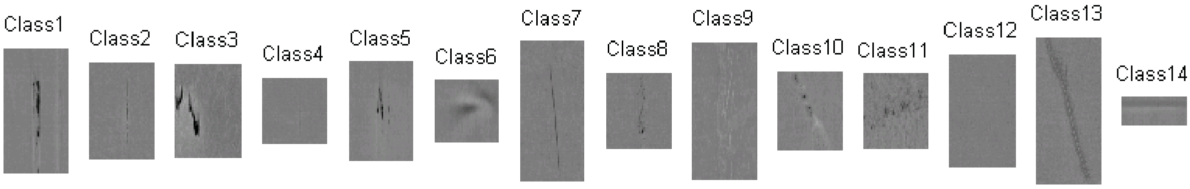

Figure 3. The example images of each metal defect image class of testing database II.

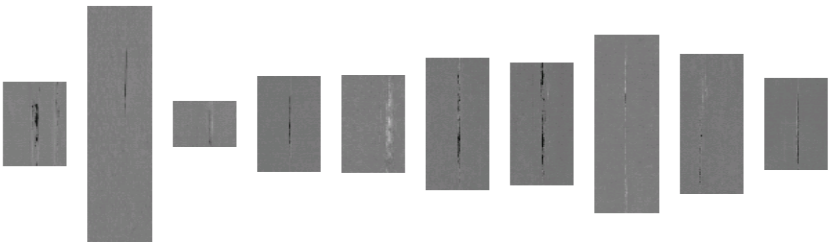

Figure 4. 10 examples of class 1 metal defect images in the testing database II.

The classification was made using nearest neighbor algorithm. The distance measure between the feature vectors was selected to be Euclidean distance. This distance can be calculated between the feature vectors (**FW**) of a query image Q and a database image D in the following way:

$$D_E(Q,D) = \sqrt{\sum_{i=1}^{R} \left(\mathbf{FW}^Q(i) - \mathbf{FW}^D(i)\right)^2} \tag{7}$$

The validation of the shape-based classification was made using *leaving one out* method [8]. In this method each image in turn is left out from the test set and used as a query image, whereas the other images in the test set form a testing database. The average classification rate was measured for both testing databases using three values for L. The results are presented in tables 1, 2, and 3.

The computational characteristics of the classification in both databases are presented in table 4. The computation was made using Matlab on a PC with 804 MHz Pentium III CPU and 256 MB primary memory.

Table 1. The average classification rate of test set I.

L	*Contour Fourier*	Multiscale Fourier
16	37.1 %	43.7 %
32	39.0 %	45.1 %
64	40.8 %	43.9 %

Table 2. The average classification rate of test set II.

L	*Contour Fourier*	Multiscale Fourier
16	26.6 %	31.6 %
32	27.8 %	30.4 %
64	29.2 %	30.5 %

Proceedings of the 12th International Conference on Image Analysis and Processing (ICIAP'03)
0-7695-1948-2/03 $17.00 © 2003 **IEEE**

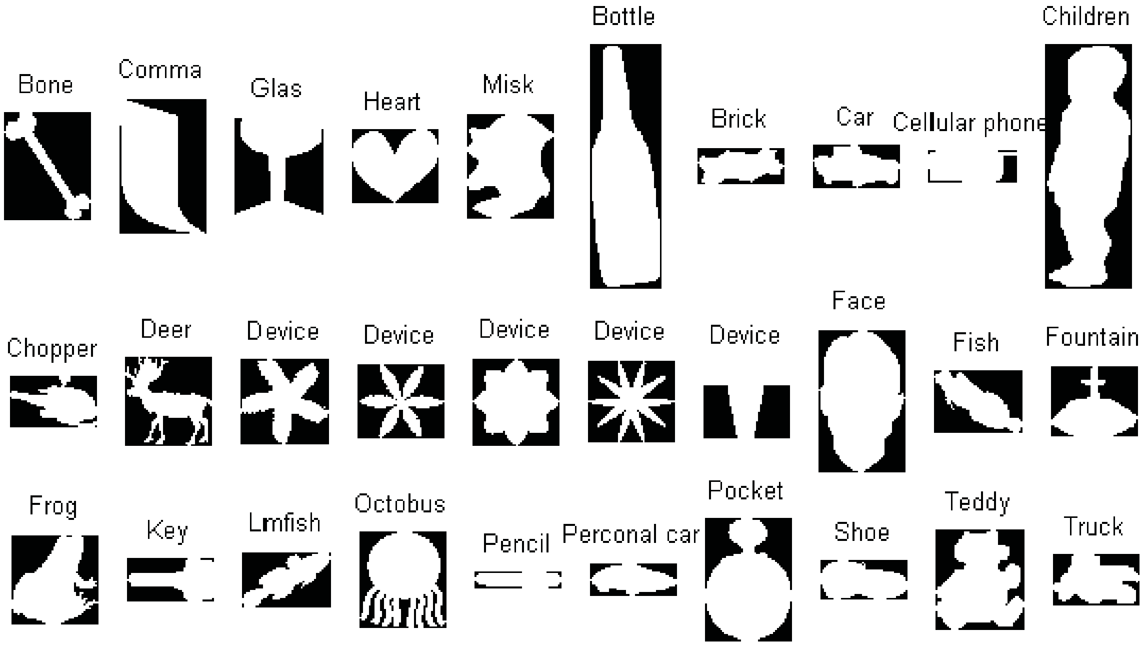

Figure 5. The example images of each defect image class of testing database II.

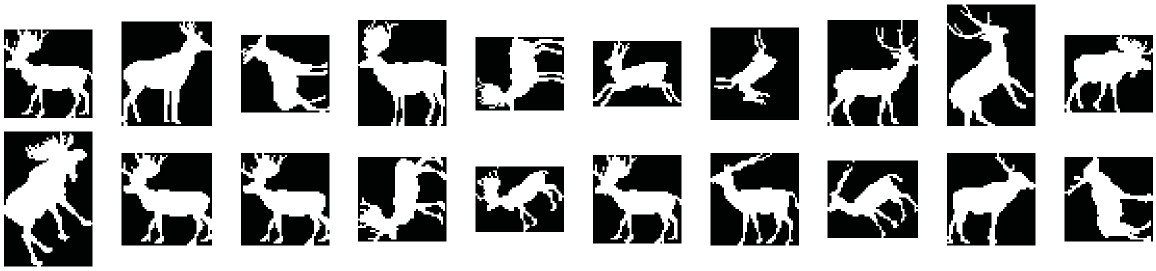

Figure 6. 20 examples of deer images in the testing database II.

Table 3. The average classification rate of test set III.

L	*Contour Fourier*	Multiscale Fourier
16	93.5 %	96.3 %
32	93.5 %	94.7 %
64	94.2 %	94.2 %

4. Results and discussion

In this paper we presented a new shape representation method, multiscale Fourier descriptor, for shape-based image classification. This descriptor combines wavelet transform and Fourier transform. In this way, the benefits of both transforms can be utilized. Therefore, when the wavelet transform is applied to the object boundary, the shape description is obtained in multiple resolutions. This is remarkable because human vision system uses multiresolution representation of shape. This representation improves also the classification of the shapes occurring in the images.

Table 4. The Computational characteristics of the methods. The computing times are presented for the classification of the whole databases.

FEATURE	Vector length	Classification time		
		DB I	DB II	DB III
Multiscale Fourier	$L*r$			
L=16	48	56 sec	67 sec	28 sec
L=32	96	81 sec	97 sec	36 sec
L=64	192	128 sec	159 sec	49 sec
Contour Fourier	L			
L=16	16	44 sec	46 sec	10 sec
L=32	32	47 sec	56 sec	25 sec
L=64	64	64 sec	83 sec	29 sec

In our approach, the obtained multiscale shape representation is transformed into frequency domain using Fourier transform. In this way, our shape description approach is invariant for rotation and the starting point of the boundary line.

According to the results presented in tables 1, 2, and 3, our method, multiscale Fourier, gives better classification results than *Contour Fourier* method in all testing databases. The classification accuracy was very high in case of the database III, in which the object shapes were quite simple and easy to distinguish from each other. On the other hand, the classification rate was relatively low in the industrial image databases I and II. This is because the shape classes of these databases are much harder to distinguish from each other. In fact, the classification of the defect images is a demanding task even to an expert. However, the results show that our method is applicable in several types of image databases.

The computational cost of multiscale Fourier is also reasonable. Compared to the *Contour Fourier*, multiscale approach demands more computation time due to the increased feature vector length. On the other hand, the difference between the classification times is not remarkable, and in the case of small values of L, the whole databases can be classified in the less than 100 sec.

In conclusion, the multiscale Fourier descriptor proved to be an effective tool for classifying different types of shapes. The classification results show that when multiscale representation is combined to the commonly used Fourier-based shape description, the classification results can be easily improved.

5. Acknowledgment

The authors wish to thank the Technology Development Centre of Finland (TEKES's grant 40397/01) for financial support.

6. References

[1] G.C.-H. Chuang and C.-C.J. Kuo, "Wavelet Descriptor of Planar Curves: Theory and Applications," *IEEE Transactions on Image Processing,* Vol. 5, No. 1, Jan. 1996, pp. 56-70.

[2] C.K. Chui, *An Introduction to Wavelets,* Academic Press, San Diego, 1992.

[3] L.F. Costa and R.M. Cesar, *Shape Analysis and Classification, Theory and Practice*, CRC Press, Boca Raton, Florida, 2001.

[4] A. Del Bimbo, *Visual Information Retrieval,* Morgan Kaufmann Publishers, San Fransisco, California, 2001.

[5] S.R. Dubois and F.H. Glanz, "An Autoregressive Model Approach to Two-Dimensional Shape Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 8, 1986 pp. 55-66.

[6] H. Freeman and L.S. Davis, "A Corner Finding Algorithm for Chain Coded Curves," *IEEE Transactions on Computers,* 26, 1977, pp. 297-303.

[7] R.C. Gonzalez and R.E. Woods, *Digital Image Processing,* Addison Wesley, 1993.

[8] D. Hand, H. Mannila, and P. Smyth: *Principles of Data Mining*, MIT Press, Massachusetts, 2001.

[9] K.-C. Hung, "The Generalized Uniqueness Wavelet Descriptor for Planar Closed Curves," *IEEE Transactions on Image Processing,* Vol. 9, No. 5, May 2000, pp. 834-845.

[10] J. Iivarinen, J. Rauhamaa and A. Visa, "Unsupervised segmentation of surface defects," *Proceedings of 13th International Conference on Pattern Recognition*, Vol. 4, Wien, Austria, Aug. 25–30, 1996, pp. 356–360.

[11] H. Kauppinen, T. Seppänen, and M. Pietikäinen, "An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 17, No. 2, Feb. 1995, pp. 201-207.

[12] R.L. Kayshap and Chellappa, "Stochastic Models for Closed Boundary Analysis: Representation and Reconstruction", *IEEE Transaction on Information Theory,* Vol. 27, No. 5, 1981, pp. 627-637.

[13] B.M. Mehtre, M.S. Kankanhalli, and W.F. Lee, "Shape Measures for Content Based Image Retrieval: A Comparison," *Information Processing & Management*, Vol. 33, No 3, 1997, pp. 319-337.

[14] F. Mokhtarian and A.K. Mackworth, "A Theory of Multiscale, Curvature-Based Shape Representation of Planar Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 14, No. 8, Aug. 1992, pp. 789-805

[15] E. Persoon and K. Fu, "Shape Discrimination Using Fourier Descriptors," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 7, 1977, pp. 170-179.

[16] J. Rauhamaa and R. Reinius, "Paper Web Imaging with Advanced Defect Classification," *Proceedings of the 2002 TAPPI Technology Summit*, Atlanta, Georgia, March 3-7, 2002.

Publication II

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2004. Multiscale Fourier Descriptor for Shape-Based Image Retrieval.

Multiscale Fourier Descriptor for Shape-Based Image Retrieval

Iivari Kunttu[1], Leena Lepistö[1], Juhani Rauhamaa[2], and Ari Visa[1]

[1]Tampere University of Technology
Institute of Signal Processing
P.O. Box 553, FI-33101 Tampere, Finland

[2]ABB Oy
Process Industry
P.O. Box 94, FI-00381 Helsinki, Finland

Abstract

The shapes occurring in the images are important in the content-based image retrieval. In this paper we introduce a new Fourier-based descriptor for the characterization of the shapes for retrieval purposes. This descriptor combines the benefits of the wavelet transform and Fourier transform. This way the Fourier descriptors can be presented in multiple scales, which improves the shape retrieval accuracy of the commonly used Fourier-descriptors. The multiscale Fourier descriptor is formed by applying the complex wavelet transform to the boundary function of an object extracted from an image. After that, the Fourier transform is applied to the wavelet coefficients in multiple scales. This way the multiscale shape representation can be expressed in a rotation invariant form. The retrieval efficiency of this multiscale Fourier descriptor is compared to an ordinary Fourier descriptor and CSS-shape representation.

1. Introduction

In addition to color and texture, shape is one of the most important features in characterization of image content in content-based image retrieval systems [6]. In these systems, the problem is to answer the question: "Which of the database images contain the most similar shapes to the query image?" This type of image retrieval is called shape similarity-based retrieval [13]. In this kind of retrieval, it has been found that shapes can be effectively characterized using a description that uses multiple resolutions [5],[14]. Therefore, multiscale shape representation [5] is essential in the shape-based classification and retrieval.

The shape description techniques can be divided into two types, region-based and boundary-based techniques [5]. The region-based methods consider the whole area of the object. In this paper we concentrate on the boundary-based shape descriptors that use only the object boundary in the description of the object shape. The most common boundary-based shape descriptors are chain codes [8] and Fourier descriptors [12]. Recently, growing research interest has been focused on *Curvature Scale Space* (CSS) shape representation [14] that has been selected to be used in the contour-based shape description of MPEG-7 standard [2]. Kauppinen et al. [11] made a comparison between autoregressive models [7] and Fourier-based descriptors in shape classification. In this comparison, Fourier descriptors proved to be best in the classification of different shapes. In the comparison made by Mehtre et al. [13], the retrieval ability of chain codes, Fourier descriptors, and different moments were compared in shape similarity-based retrieval. In this case, the best results were obtained using moments and Fourier descriptors. In the recent study of [17] Zhang and Lu, Fourier descriptors gave better experimental results in image retrieval than CSS-representation. These experimental results show that Fourier descriptor is an effective tool in the shape classification and retrieval.

In this paper, we apply wavelet transform to the retrieval of different shapes. Despite the fact that wavelet transform has been widely used in multiscale image analysis, it has only a few applications in the shape description. In [3] Chuang and Kuo used one-dimensional discrete periodized wavelet transform (DPWT) to describe planar curves. The same transform was used also in [10], in which the method was made rotation invariant. In contrast to these studies, we use complex wavelet transform. In our approach, the object boundary is presented in complex form like in the case of Fourier descriptors. Using complex wavelet transform, the multiscale representation of the shape can be achieved. Our method, *Multiscale Fourier* descriptor is obtained by applying the Fourier transform to the coefficients of the multiscale wavelet transform. This approach has given promising results in the shape classification in [12].

2. Shape description

The use of Fourier descriptors is common in pattern recognition and image analysis. The benefits of the Fourier descriptors are invariance to the starting point of the boundary and rotation [5]. However, the use of Fourier-based multiscale representation of shape is a new application in shape representation. In this section, we present a multiscale shape descriptor based on the complex wavelet transform and Fourier transform

In this paper, the shape description methods are based on the boundary of the object. In the presentation of the object boundary, we use the complex coordinate function [11]. This function represents simply the coordinates of

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

the boundary pixels in an object centered coordinate system, presented as complex numbers:

$$z(k) = (x_k - x_c) + j(y_k - y_c) \tag{1}$$

for k=0,1,2,…,N-1, in which N is the length of the boundary and (x_c, y_c) is the centroid of the object. Using the complex coordinate function, the boundary can be represented independent of the location of the object in the image. In this way the translation invariance can be achieved.

2.1. Fourier descriptors

The shape descriptor based on the object boundary can be formed in several ways. Fourier transform [15] is a commonly used method for this purpose. Fourier transformation of a boundary function generates a set of complex numbers which are called Fourier descriptors. Fourier descriptors characterize the object shape in a frequency domain. The Fourier descriptors can be formed for a complex boundary using discrete Fourier transform (DFT) [9]. Fourier transform of $z(k)$ is:

$$F(n) = \frac{1}{N}\sum_{k=0}^{N-1} z(k)e^{-j2\pi nk/N} \tag{2}$$

for n=0,1,2,…,N-1. The general shape of the object is represented by the lower frequency descriptors, whereas high frequency descriptors represent the small details of the object shape [11]. A common approach to shape classification is to use only a subset of the descriptors. These subsets can be formed in several different ways. Kauppinen et al. [11] have compared *Curvature Fourier*, *Radius Fourier*, *Contour Fourier*, and *A-invariant* methods for Fourier-based shape representation. According to their experimental results, *Contour Fourier* and *A-invariant* methods were the best approaches in the shape classification. In this work, we selected *Contour Fourier* method for testing purposes.

The *Contour Fourier* method [11] makes the Fourier transform directly for the complex coordinate function of the object boundary. In this method the descriptors are taken both positive and negative frequency axis. The scaling of the descriptors is made by dividing the absolute values of the selected descriptors by the absolute value of the first non-zero component. The feature vector for this method is:

$$x = \left[\frac{|F_{-(L/2-1)}|}{|F_1|} \cdots \frac{|F_{-1}|}{|F_1|}\,\frac{|F_2|}{|F_1|} \cdots \frac{|F_{L/2}|}{|F_1|}\right]^T \tag{3}$$

In which L is a constant value that defines the length of the feature vector.

2.2. Multiscale Fourier descriptor

The multiscale representation of the object boundary can be achieved using wavelet transform. The boundary function is transformed using some wavelet Ψ. Complex wavelet transform is based on the continuous wavelet transform (CWT) [4]. In CWT, the wavelet coefficient of the boundary $z(k)$ at a scale a and position b is defined as:

$$C_a(b) = \frac{1}{\sqrt{|a|}}\int_R z(k)\psi\left(\frac{k-b}{a}\right)dk \tag{4}$$

As in the Fourier transform, also in case of CWT we obtain a set of complex coefficients $C_a(b)$ of scale a. The coefficients are defined for all positions b=0,1,2,...,N-1.

The problem with the coefficients obtained from the complex wavelet transform is the fact that they are dependent on the starting point of the object boundary. Also the length of the feature vector depends on the length of the object boundary. Therefore, the coefficient vectors of different shapes cannot directly be matched in the image retrieval. The solution for this problem is to apply the Fourier transform to the coefficients obtained from the complex wavelet transform. In this way, the multiscale shape representation can be transformed to the frequency domain. Hence the benefits of multiscale representation and Fourier shape representation can be combined. The *Multiscale Fourier* descriptor is formed by applying the discrete Fourier transform of equation 2 to the set of complex coefficients $C_a(b)$:

$$F^a(n) = \frac{1}{N}\sum_{b=0}^{N-1} C_a(b)e^{-j2\pi nb/N} \tag{5}$$

The multiscale descriptor x^a of each scale a is then formed from coefficients $F^a(n)$ using *Contour Fourier* method presented in equation 3:

$$x^a = \left[\frac{|F^a_{-(L/2-1)}|}{|F^a_1|} \cdots \frac{|F^a_{-1}|}{|F^a_1|}\,\frac{|F^a_2|}{|F^a_1|} \cdots \frac{|F^a_{L/2}|}{|F^a_1|}\right]^T \tag{6}$$

The multiscale representation of the object shape can then be formed by defining the descriptor x^a using several different scales a, and combining the descriptors into a single feature vector, **FW** of length R. Let the set of scales be $A=\{a_1, a_2,..., a_r\}$. So the number of the scales in the descriptor is r.

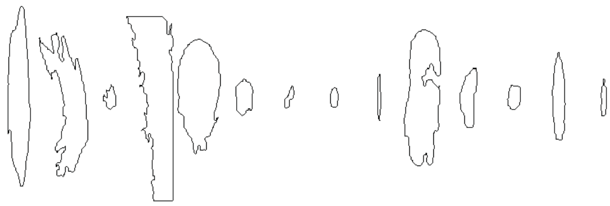

Figure 1. Examples of the boundaries extracted from each class of database I paper defects.

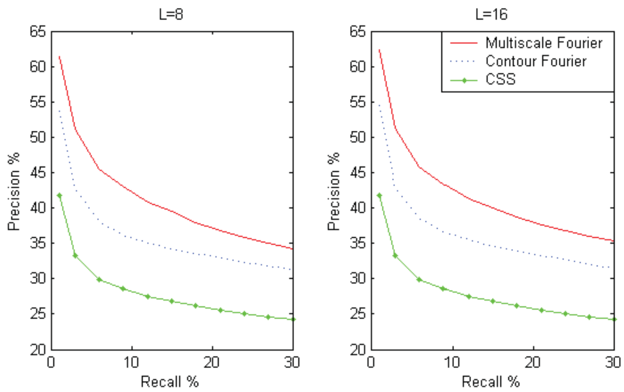

Figure 2. Average precision/recall curves of the database I paper defect images.

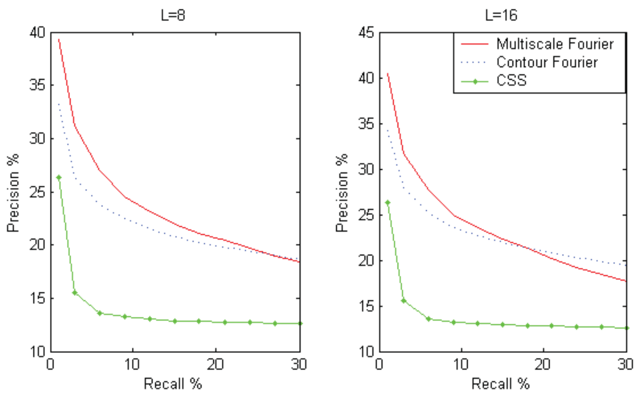

Figure 3. Average precision/recall curves of the database II metal defect images.

3. Retrieval experiments

For testing purposes, we used paper and metal defect images that were collected from industrial processes. The reason for the collection of the defect image databases in the process industry is the practical need of controlling the quality and production [16]. In the industrial imaging solutions, there is a need to retrieve the defect images from the databases. In these images, the defect shape is one essential feature that describes the defect type. Therefore, effective methods for the shape representation are needed in the retrieval and classification of the defect images.

Testing database I consisted of paper defect images. The images were taken from the paper manufacturing process using a paper inspection system [16]. The defects occurring in the paper can be for example holes, wrinkles or different kinds of dirt spots. The test set consisted of 1204 paper defects, which represented 14 defect classes so that each class consisted of 27-103 images. An example contour of each defect class is presented in figure 1. Testing database II consisted of 1943 metal defect images. Also this database contained 14 defect classes. In each class, there were 100-165 images. Within the classes of both defect databases, there were differences in the size and orientation of the defects.

The indexing of the testing databases was carried out by calculating the feature vectors x^a of *Multiscale Fourier* method for each image. The selected wavelet ψ was complex gaussian wavelet of order two. The multiscale presentation was achieved using a set of three scales. In this experiment, the scale sets A was selected to be $\{10,15,20\}$ and $\{10,20,30\}$ for the database I and II, respectively. For comparison, also the feature vector x of *Contour Fourier* method was calculated for each test set image. We compared our method also to CSS-presentation [14] that has recently received lot of attention in the research of shape description. The method for measuring the similarity between the CSS-presentations was presented in [14]

In the retrieval experiments, the distance measure between the feature vectors was selected to be Euclidean distance. The retrieval experiments were made using *leaving one out* method. In this method, each image in turn is left out from the test set and used as a query image, whereas the other images in the test set form a testing database. The performance of the retrieval was measured by calculating a *precision versus recall curve* [1] for each query. Figures 2 and 3 present the average precision-recall curves for the images in both databases using two different values for L. In both databases, the best retrieval performance is achieved using *Multiscale Fourier*. In the figures 2 and 3, the precision values are not significantly high for two reasons: 1) shape is only one of the features (in addition to gray level distribution and texture) that describe the defect and 2) the defect classes are often fuzzy and overlapping, which makes it difficult to distinguish them from each other. However, the results of figures 2 and 3 present the practical comparison between the methods in the shape-based retrieval. Also the retrieval accuracy of the metal defect shapes (database II) is lower than in the case of paper defect shapes of database I. This is due to the nature of metal defects, which are more difficult to classify based on their shapes. However, figure 3 shows that *Multiscale Fourier* is the most effective method also in the shape-based retrieval of metal defect images.

When we compare the computational cost of *Multiscale Fourier* to the *Contour Fourier* method, the computational cost of the *Multiscale Fourier* descriptor is larger due to the dimensionality of the feature vectors of multiscale representation. The dimensionality is dependent on the number of the scales, *r*. Hence, the length of the feature vector is $r*L$, However, using three scales the cost is still reasonable, which makes the *Multiscale Fourier* suitable for indexing of large image databases.

4. Discussion

In this paper, we presented *Multiscale Fourier* method for shape description in image retrieval. In this method, the object boundary is transformed using complex wavelet transform. The obtained multiscale shape representation is then transformed into the frequency domain using Fourier transform. Therefore, the resulting descriptor is independent on rotation as well as the starting point of the boundary. *Multiscale Fourier* method combines the benefits of the wavelet transform and Fourier transform. When the wavelet transform is applied to the object boundary, we obtain the shape description in multiple resolutions. This is important, because multiscale shape representation improves the shape retrieval accuracy.

We made also a comparison between CSS-representation that has been selected to be the basis of contour-based shape representation of MPEG-7 standard. In our comparison, CSS-representation was outperformed by both Multiscale and Contour Fourier descriptors. This result supports the conclusion of Zhang and Lu [17], which is that Fourier descriptors give more accurate retrieval results than CSS-representation. However, the experimental results of this paper show that even better results can be achieved using *Multiscale Fourier* method. The computational cost of *Multiscale Fourier* remained still reasonable.

5. Acknowledgment

The authors wish to thank Technology Development Centre of Finland (TEKES's grant 40120/03) for financial support.

6. References

[1] R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, ACM Press, Addison-Wesley, New York, 1999.

[2] M. Bober, "MPEG-7 Visual Shape Descriptors", *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 11, No. 6, 2001, pp. 716-719.

[3] G.C.-H. Chuang and C.-C.J. Kuo, "Wavelet Descriptor of Planar Curves: Theory and Applications," *IEEE Transactions on Image Processing,* Vol. 5, No. 1, 1996, pp. 56-70.

[4] C.K. Chui, *An Introduction to Wavelets,* Academic Press, San Diego, 1992.

[5] L.F. Costa and R.M. Cesar, *Shape Analysis and Classification, Theory and Practice*, CRC Press, Boca Raton, Florida, 2001.

[6] A. Del Bimbo, *Visual Information Retrieval,* Morgan Kaufmann Publishers, San Fransisco, 2001.

[7] S.R. Dubois and F.H. Glanz, "An Autoregressive Model Approach to Two-Dimensional Shape Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 8, 1986, pp. 55-66.

[8] H. Freeman and L.S. Davis, "A Corner Finding Algorithm for Chain Coded Curves," *IEEE Transactions on Computers,* Vol. 26, 1977, pp. 297-303.

[9] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Addison Wesley, 1993.

[10] K.-C. Hung, "The Generalized Uniqueness Wavelet Descriptor for Planar Closed Curves," *IEEE Transactions on Image Processing,* Vol. 9, No. 5, 2000, pp. 834-845.

[11] H. Kauppinen, T. Seppänen, and M. Pietikäinen, "An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 17, No. 2, 1995, pp. 201-207.

[12] I. Kunttu, L. Lepistö, J. Rauhamaa, A. Visa, "Multiscale Fourier Descriptor for Shape Classification", *Proceedings of 12th International Conference on Image Analysis and Processing*, 2003, pp. 536-541.

[13] B.M. Mehtre, M.S. Kankanhalli, and W.F. Lee, "Shape Measures for Content Based Image Retrieval: A Comparison," *Information Processing Management*, Vol. 33, No 3, 1997, pp. 319-337.

[14] F. Mokhtarian "Silhouette-Based Isolated Object Recognition through Curvature Scale Space" *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 17, No. 5, May. 1995, pp. 539-544.

[15] E. Persoon and K. Fu, "Shape Discrimination Using Fourier Descriptors," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 7, 1977, pp. 170-179.

[16] J. Rauhamaa and R. Reinius, "Paper Web Imaging with Advanced Defect Classification," *Proceedings of the 2002 TAPPI Technology Summit*, Atlanta, Georgia, March 3-7, 2002.

[17] D. Zhang and G. Lu, "A Comparative Study of Curvature Scale Space and Fourier Descriptors for Shape-Based image Retrieval", *Journal of Visual Communication and Image Representation,* Vol. 14, No. 1, 2003, pp. 41-60.

Publication III

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2005. Multiscale Fourier Descriptors for Defect Image Retrieval.

ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Pattern Recognition Letters xxx (2005) xxx–xxx

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Multiscale Fourier descriptors for defect image retrieval

Iivari Kunttu [a,*], Leena Lepistö [a], Juhani Rauhamaa [b], Ari Visa [a]

[a] Tampere University of Technology, Institute of Signal Processing, P.O. Box 553, FIN-33101 Tampere, Finland
[b] ABB Oy, Process industry, P.O. Box 94, FIN-00381 Helsinki, Finland

Received 12 November 2004; received in revised form 4 August 2005

Communicated by R. Manmatha

Abstract

Shape is an essential visual feature of an image and it is widely used to describe image content in image classification and retrieval. In this paper, two new Fourier-based approaches for contour-based shape description are presented. These approaches present Fourier descriptors in multiple scales, which improves the shape classification and retrieval accuracy. The proposed methods outperform ordinary Fourier descriptors in the retrieval of complicated shapes without increasing computational cost.

Keywords: Shape retrieval; Shape classification; Fourier descriptor; Multiresolution shape description

1. Introduction

The description of the shapes in image is an essential task in the field of pattern recognition. During recent decades, a number of approaches and solutions have been presented to characterize different shapes. The primary purpose of the early shape descriptors was shape classification, whereas during recent years the use of shape description in image retrieval has received increasing interest (Del Bimbo, 2001). For example, in multimedia applications, content-based image retrieval (CBIR) plays a significant role.

In CBIR systems, one of the problems is to answer the question: "Which of the database images contain the most similar shapes to the query image"? This type of image retrieval is called shape similarity-based retrieval (Mehtre et al., 1997). In this kind of retrieval, the aim is to find similar shapes from the database as accurately as possible. On the other hand, the classification accuracy (effectiveness) of a certain descriptor is not an adequate measure for its usefulness in the retrieval. Due to the increasing number of online retrieval solutions, computational lightness (efficiency) is nowadays considered equally important as effectiveness (Zhang and Lu, 2004). A recently introduced multimedia standard, MPEG-7 (Manjunath et al., 2002), has set several principles for measuring a shape descriptor. The principles are good retrieval accuracy, compact features, general application, low computational complexity, robust retrieval performance, and hierarchical coarse to fine representation (Kim and Kim, 2000). These principles were used as criteria in the study of Zhang and Lu (2004), in which common shape description techniques were reviewed. Another review of the state of the art in shape description techniques is provided by Loncaric (1998).

Shape description techniques can be divided into two types: region- and boundary-based techniques (Costa and Cesar, 2001). Region-based methods consider the whole area of an object. Different moments (Hu, 1962), including for example Zernike moments (Teague, 1980) are popular descriptors. Boundary-based shape descriptors use only the object boundary in the description of the object shape. The most common boundary-based shape descriptors are chain codes (Freeman and Davis, 1977),

* Corresponding author. Tel.: +358 3 3115 4393; fax: +358 3 3115 4989.
E-mail address: Iivari.Kunttu@tut.fi (I. Kunttu).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2005.08.022

Fourier descriptors (Persoon and Fu, 1977) and simple descriptors such as circularity (Costa and Cesar, 2001), eccentricity, convexity, principle axis ratio, circular variance and elliptic variance (Iivarinen and Visa, 1998). Recently, growing research interest has been focused on curvature scale space (CSS) shape representation (Mokhtarian and Mackworth, 1986) that has been selected to be used as the boundary-based shape descriptor of MPEG-7 standard (Bober, 2001). However, despite the fact that the Fourier descriptor method is over 30 years old (Granlund, 1972; Persoon and Fu, 1977), it is still found to be a valid shape description tool. In fact, Fourier descriptor has proved to outperform most other boundary-based methods in terms of retrieval accuracy and efficiency. This has been verified in several comparisons. Kauppinen et al. (1995) made a comparison between autoregressive models (Dubois and Glanz, 1986) and Fourier-based descriptors in shape classification. In most cases, Fourier descriptors proved to perform better than autoregressive models. In a comparison made by Mehtre et al. (1997), the retrieval ability of chain codes, Fourier descriptors, and different moments were compared in shape similarity-based retrieval. In this case, the best results were obtained by using moments and Fourier descriptors. In the recent studies of Zhang and Lu (2003a,b), Fourier descriptors and Zernike moments outperformed CSS-representation in terms of retrieval accuracy and efficiency. Similar results were also obtained by Kunttu et al. (2004).

In addition to good retrieval and classification accuracy, there are also other reasons which make Fourier descriptors probably the most popular of the boundary-based shape representations. The main advantages of the Fourier-based shape representations are that they are compact and computationally light. Furthermore, they are easy to normalize and their matching is a very simple process. Also their sensitivity to noise is low when only low frequency Fourier coefficients are used as descriptors.

It has been found that complicated shapes can be effectively characterized by using a description with multiple resolutions (Costa and Cesar, 2001; Mokhtarian, 1995). CSS-representation uses multiple resolutions that are achieved by smoothing the boundary. However, the drawbacks of CSS are relatively low shape classification accuracy and efficiency compared to Fourier descriptors (Zhang and Lu, 2003a,b). Also the matching procedure of the CSS-features is not as straightforward as that of Fourier descriptors. Wavelet transform (Chui, 1992) has been widely used in multiscale image analysis. However, it has only a few applications in the shape description (Yang et al., 1998; Tieng and Boles, 1997). The obtained descriptors are not rotation invariant. Furthermore, the matching scheme of these wavelet representations is more complicated and time consuming than that of Fourier descriptors. This reduces their usability in on-line retrieval solutions.

In this paper, two multiresolution approaches to shape description are presented. The first one, called here *Multiscale Fourier*, utilizes a combination of wavelet and Fourier transforms. *Multiscale Fourier* descriptor is obtained by applying the Fourier transform to the coefficients of the multiscale wavelet transform. Consequently, the Fourier descriptor is formed from multiresolution representation of the shape. The *Multiscale Fourier* approach has given promising results in the classification of general shapes as well as shapes of industrial defects in (Kunttu et al., 2003). In (Kunttu et al., 2004), this method also outperformed CSS-representation and *Contour Fourier* descriptor (Kauppinen et al., 1995) in the shape-based retrieval of different kinds of defect images. The second multiresolution approach presented in this paper is called *Boundary Scale Fourier* descriptor. This descriptor is obtained by using different scales of the boundary line. The scales are achieved by smoothing the boundary line iteratively and applying the Fourier transform to the boundary of different degrees of smoothness. In both of these descriptors, the matching is as simple as in the case of Fourier descriptors, which is a benefit in the on-line retrieval.

The outline of this paper is the following. Section 2 presents the methodology of shape description using Fourier-based methods. In addition to ordinary single-scale Fourier descriptors, both of the proposed multiresolution methods are presented in that section. Section 3 is the experimental part of this paper. In the experiments, the proposed methods are evaluated and compared to ordinary Fourier descriptors in retrieval. For retrieval experiments, a database of real industrial defect shapes is used. In Section 4, the results and performance of the methods are discussed. Section 5 concludes this study.

2. Shape description

In this paper, the shape description methods are based on the object boundary. In shape description, a boundary is usually presented using some shape signature i.e. a function derived from the boundary coordinates. Complex coordinate function (Kauppinen et al., 1995) is a simple and probably the best-known signature used in the Fourier-based shape description. Let (x_k, y_k), $k = 0, 1, 2, \ldots, N-1$ represent the object boundary coordinates, in which N is the length of the boundary. The complex coordinate function $z(k)$ expresses the boundary points in an object centered coordinate system in which (x_c, y_c) represents the centroid of the object:

$$z(k) = (x_k - x_c) + j(y_k - y_c) \tag{1}$$

Hence, using this function, the boundary is represented independent of the location of the object in the image. In this way the translation invariance can be achieved.

2.1. Fourier descriptors

Fourier descriptors characterize the object shape in a frequency domain. The descriptors can be formed for the complex-valued boundary function using the discrete Fourier transform (DFT) (Gonzalez and Woods, 1993). The Fourier transform of $z(k)$ is

$$F(n) = \frac{1}{N}\sum_{k=0}^{N-1} z(k)\mathrm{e}^{-j2\pi nk/N} \tag{2}$$

for $n = 0, 1, 2, \ldots, N-1$ and $F(n)$ are the transform coefficients of $z(k)$. The translational invariance is based on the shape signature. Furthermore, the coefficients have also to be normalized to achieve invariance to rotation and scaling. The descriptors can be made rotation invariant by ignoring the phase information and using only the magnitudes of the transform coefficients $|F(n)|$. In the case of complex-valued boundary function, the scale can be normalized by dividing the magnitudes of the transform coefficients by $|F(1)|$ (Kauppinen et al., 1995).

The general shape of the object is represented by the low frequency coefficients, whereas high frequency coefficients represent the fine details of the object shape. A common approach to shape representation is to use a subset of the low frequency coefficients as a shape descriptor. This way the shape can be effectively presented using a relatively short feature vector. In our experiments, the feature vector is formed using *Contour Fourier* method (Kauppinen et al., 1995), which applies the complex coordinate function. In this method the descriptors are taken from positive and negative frequency axis. The feature vector for this method is

$$x = \left[\frac{|F_{-(L/2-1)}|}{|F_1|} \cdots \frac{|F_{-1}|}{|F_1|}\,\frac{|F_2|}{|F_1|} \cdots \frac{|F_{L/2}|}{|F_1|}\right]^{\mathrm{T}} \tag{3}$$

In which L is a constant value that defines the dimensionality of the feature vector. Fig. 1a presents the outline of the *Contour Fourier* method.

2.2. Multiscale Fourier descriptor using wavelet transform

The multiscale representation of the object boundary can be achieved using wavelet transform (Kunttu et al., 2003, 2004). The boundary function is transformed using some wavelet Ψ (Chui, 1992). Complex wavelet transform is based on the continuous wavelet transform (CWT) (Teolis, 1998). In continuous wavelet transform (CWT), the wavelet coefficient of the boundary $z(k)$ at a scale a and position b is defined as

$$C_a(b) = \frac{1}{\sqrt{|a|}}\int_R z(k)\,\psi\left(\frac{k-b}{a}\right)\mathrm{d}k \tag{4}$$

As in the Fourier transform, in the case of CWT we obtain a set of coefficients $C_a(b)$ of scale a. The coefficients are defined for all positions $b = 0, 1, 2, \ldots, N-1$. The families of complex wavelets include e.g. complex Gaussian, complex Morlet, and complex Shannon wavelets (Teolis, 1998). In the experiments of this paper we have used complex Gaussian wavelets (Misiti et al., 2001).

The problem with the coefficients obtained from the wavelet transform is the fact that they are dependent on the starting point of the object boundary. Hence the obtained descriptor is not rotation invariant. Also the dimensionality of the feature vector depends on the boundary length. Therefore, the coefficient vectors of different shapes cannot be directly matched in the image retrieval. The proposed solution for this problem is to apply the Fourier transform to the coefficients obtained from the wavelet transform. In this way, the multiscale shape representation can be transformed to the frequency domain, in which normalization and matching are straightforward operations. Hence the benefits of multiscale representation and Fourier shape representation can be combined. The *Multiscale Fourier* descriptor is formed by applying the discrete Fourier transform of Eq. (2) to the set of complex coefficients $C_a(b)$:

$$F^a(n) = \frac{1}{N}\sum_{b=0}^{N-1} C_a(b)\mathrm{e}^{-j2\pi nb/N} \tag{5}$$

The multiscale descriptor x^a of each scale a is then formed from coefficients $F^a(n)$ using *Contour Fourier* method presented in Eq. (3):

$$x^a = \left[\frac{|F^a_{-(L/2-1)}|}{|F^a_1|} \cdots \frac{|F^a_{-1}|}{|F^a_1|}\,\frac{|F^a_2|}{|F^a_1|} \cdots \frac{|F^a_{L/2}|}{|F^a_1|}\right]^{\mathrm{T}} \tag{6}$$

The multiscale representation of the object shape can then be formed by defining the descriptor x^a using several differ-

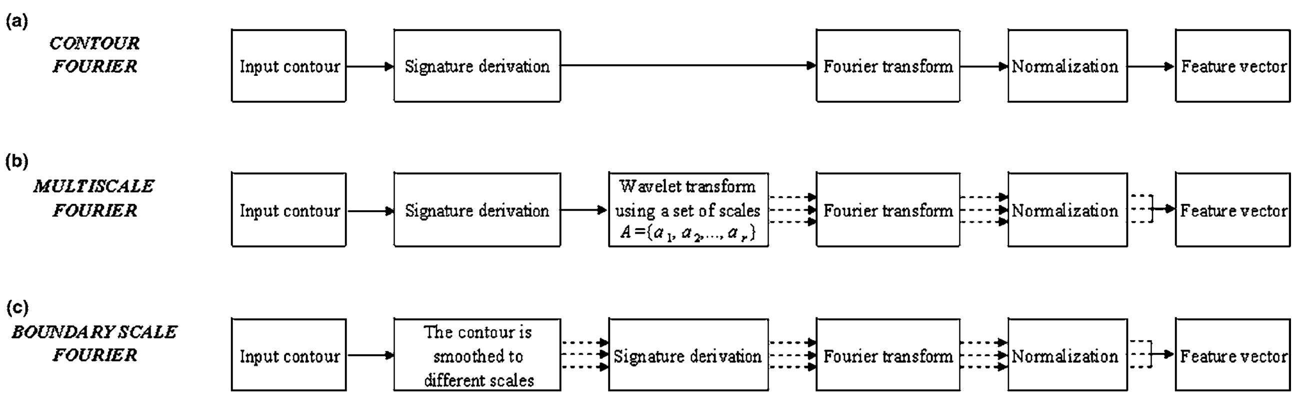

Fig. 1. The outlines of different Fourier-based shape description methods used in the comparison.

ent scales a, and combining the descriptors into a single feature vector. If the set of scales is defined as $A = \{a_1, a_2, \ldots, a_r\}$, the number of the scales in the descriptor is r. The dimensionality of the obtained feature vector is r^*L. The outline of the *Multiscale Fourier* method is presented in Fig. 1b.

2.3. Boundary scale Fourier descriptor

Another approach to multiresolution Fourier-based shape representation is related to the scaling of the object boundary. Hence, when the boundary signature is presented in multiple scales and the descriptors are formed for each scale, the resulting Fourier representation has a multiresolution property. The multiresolution boundary representation can be achieved in several ways. A simple method for this is the CSS-representation (Mokhtarian and Mackworth, 1986), which uses smoothing of the boundary. In the CSS-method, the object boundary is iteratively smoothed until the curvature function has no zero-crossing points, i.e. the boundary is convex. Hence the multiresolution representation is achieved based on boundary lines of different curvatures. In the proposed approach, the Fourier descriptors are defined for the boundaries of different smoothnesses. This way a more robust shape description can be obtained, because fine details and distortions in a shape are likely to be removed during the smoothing process. When (x_k, y_k), where $k = 0, 1, 2, \ldots, N-1$, represent the object boundary coordinates, the curvature function of the boundary can be defined as

$$c(k) = \frac{(\dot{x}(k)\ddot{y}(k) - \ddot{x}(k)\dot{y}(k))}{(\dot{x}^2(k) + \dot{y}^2(k))^{3/2}} \tag{7}$$

where $\dot{x}(k), \ddot{x}(k), \dot{y}(k)$ and $\ddot{y}(k)$ are the first and second derivatives of the boundary coordinates, respectively (Mokhtarian and Mackworth, 1986). The boundary curvature is iteratively smoothed and the zero-crossing points are defined for each scale. When the contour becomes smoother, the number of curvature zero-crossing points is decreased. The boundary at each degree of smoothing represents a different scale. Therefore, in the *Boundary Scale Fourier* descriptor, the shape descriptors are defined for the boundaries at the desired scales. The smoothness of the boundary can be indicated based on the number of the zero-crossing points. If p is the original number of the zero-crossing points and p_f is the number of these points in a smoothed boundary, the degree of smoothness or scale can be expressed as the ratio $s = p_f/p$. Unlike in the CSS-representation, in this approach it is not necessary to smooth the boundary until it is convex, but the smoothing is stopped when the desired scale is achieved. This saves the computation time. An example of smoothing process of a boundary extracted from a defect image using different values of s is presented in Fig. 2. Thus the Fourier descriptor for shape signature at a selected scale s is defined by Eqs. (1)–(3). The final feature vector is a combination of the Fourier descriptors obtained from the signatures at different scales. If the set of scales is defined as $S = \{s_1, s_2, \ldots, s_r\}$, the number of the scales in the descriptor is r. Hence, the dimensionality of the feature vector is r^*L. The outline of the *Boundary Scale Fourier* descriptor is presented in Fig. 1c.

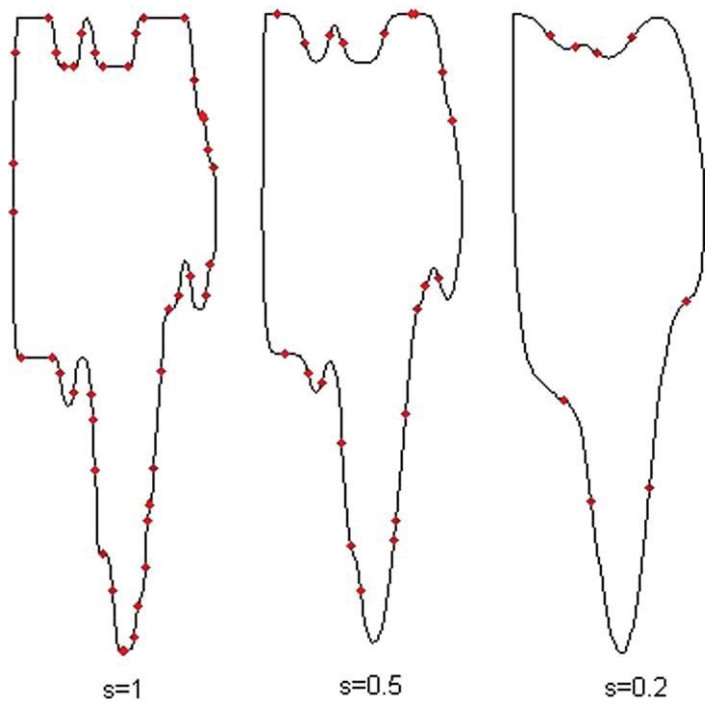

Fig. 2. Smoothing process of a defect boundary using three values of s. The zero-crossing points are marked on the boundaries.

3. Experiments with industrial defect shapes

In the experimental part of this paper the accuracy and efficiency of the proposed methods are tested and compared to those of the commonly used Fourier descriptors using a real defect shape database. The proposed descriptors are compared to the *Contour Fourier* method, because it has been found to be accurate in shape classification (Kauppinen et al., 1995). Furthermore, the preliminary experiments carried out in the testing database of this study showed that *Contour Fourier* method outperforms the other Fourier-based shape representations. Being simple and computationally light method, *Contour Fourier* is suitable for on-line shape retrieval tasks. The CSS-representation is not included in the comparison, because previous results indicate that it performs with lower retrieval accuracy than any of the Fourier-based methods. This conclusion was drawn from experiments with databases of general shapes (Zhang and Lu, 2003a,b) and with defect shapes (Kunttu et al., 2004).

The validation presented in this section is twofold. Simple classification experiments are first carried out to show the influence of scale selection on the shape description. Furthermore, the classification results are used to show that the proposed shape descriptors outperform the *Contour Fourier* descriptors in shape classification. The second part of the validation presents the shape retrieval experiments, in which the retrieval accuracy of the proposed methods is compared to that of *Contour Fourier* method.

3.1. Testing database

For testing purposes, we used defect images that were collected from a real industrial process using a paper inspection system (Rauhamaa and Reinius, 2002). A reason for collecting defect image databases in process industry is a practical need for controlling the quality of production (Rauhamaa and Reinius, 2002). When retrieving images from a database, the defect shape is one essential property describing the defect class. Therefore, effective methods for the shape representation are necessary. The defects occurring in paper can be for example holes, wrinkles or different kinds of thin or dirt spots. The test set consisted of 1204 paper defects which represented 14 defect classes with each class consisting of 27–103 images. Example contours of each defect class are presented in Fig. 3. Within each class, there are defects of different size and orientation. Furthermore, in some classes the boundaries are very varying and sometimes distorted (classes 2, 4 and 10, for example).

3.2. Classification and retrieval

The feature extraction in the testing database was carried out by calculating all descriptors for the images in the database. In the case of *Multiscale Fourier* method, the selected wavelet ψ was Gaussian wavelet of order two (Misiti et al., 2001). To select of the scale sets, we made preliminary k-nearest neighbor (k-NN) classification

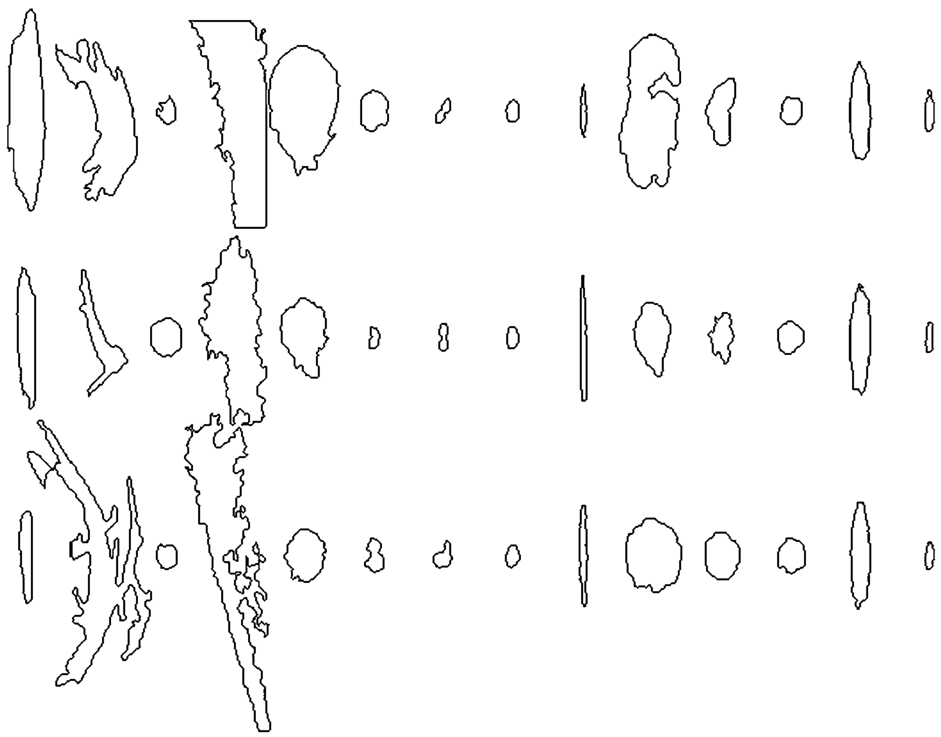

Fig. 3. Three example contours of each 14 paper defect classes in the testing database.

experiments with different scales. In the classification experiments the proposed methods were applied using each scale separately. In other words, the methods used single-scale wavelet transform and boundary smoothing combined with Fourier transform as described in Section 2. Fig. 4 presents the average classification rates of *Multiscale Fourier* and *Boundary Scale Fourier* methods at different scales using 5-NN classifier. In this figure, also the classification rate of the *Contour Fourier* descriptor (41.87%) is presented.

The scales that produce the highest classification rates were selected to be used in the retrieval experiments. On the other hand, the classification errors of the selected scales were wanted to be as decorrelating as possible. Therefore, the selected scales were not close to each other. To reduce the dimensionality of the features, the number of scales was selected to be two in the both multiresolution approaches. The approach of using three scales was tested in (Kunttu et al., 2003, 2004). However, in the retrieval experiments of this paper, the scale set A was selected to be $\{6, 14\}$. In the case of the *Boundary Scale Fourier* descriptor, the scale set S for the defect shapes was $\{1, 0.05\}$. Thus the scale 1 corresponds to the boundary without smoothing i.e. it is equal to *Contour Fourier*. In the *Boundary Scale Fourier* method, the boundary smoothing was carried out by application of a low-pass filtering with the kernel $[0.25, 0.5, 0.25]$ to coordinates (x_k, y_k) of the boundary line (Manjunath et al., 2002). The filtering was applied repetitively until a desired scale was achieved. To demonstrate the classification performance of the selected scale combinations, their classification rates (48.42% and 43.85% for *Multiscale Fourier* and *Boundary Scale Fourier*, respectively) are marked in Fig. 4.

In the retrieval experiments, the distance measure between the feature vectors was selected to be Euclidean distance, which is the most common distance metrics for Fourier descriptors. The retrieval and classification experiments were made using *leaving one out* method. In this method, each shape in turn is left out from the test set and used as a query shape; whereas the other shapes in the test set form a testing database. The performance of the retrieval was measured by calculating a *precision versus recall curve* (Baeza-Yates and Ribeiro-Neto, 1999) for each query. Let $|A|$ and $|R|$ be the number of all retrieved shapes and the number of query class shapes in the whole testing database, respectively. If $|Ra|$ is the number of retrieved query class shapes, precision and recall are defined as $|Ra|/|A|$ and $|Ra|/|R|$, respectively. The retrieval performance of each feature can be presented by calculating the average precision-recall curve for each query.

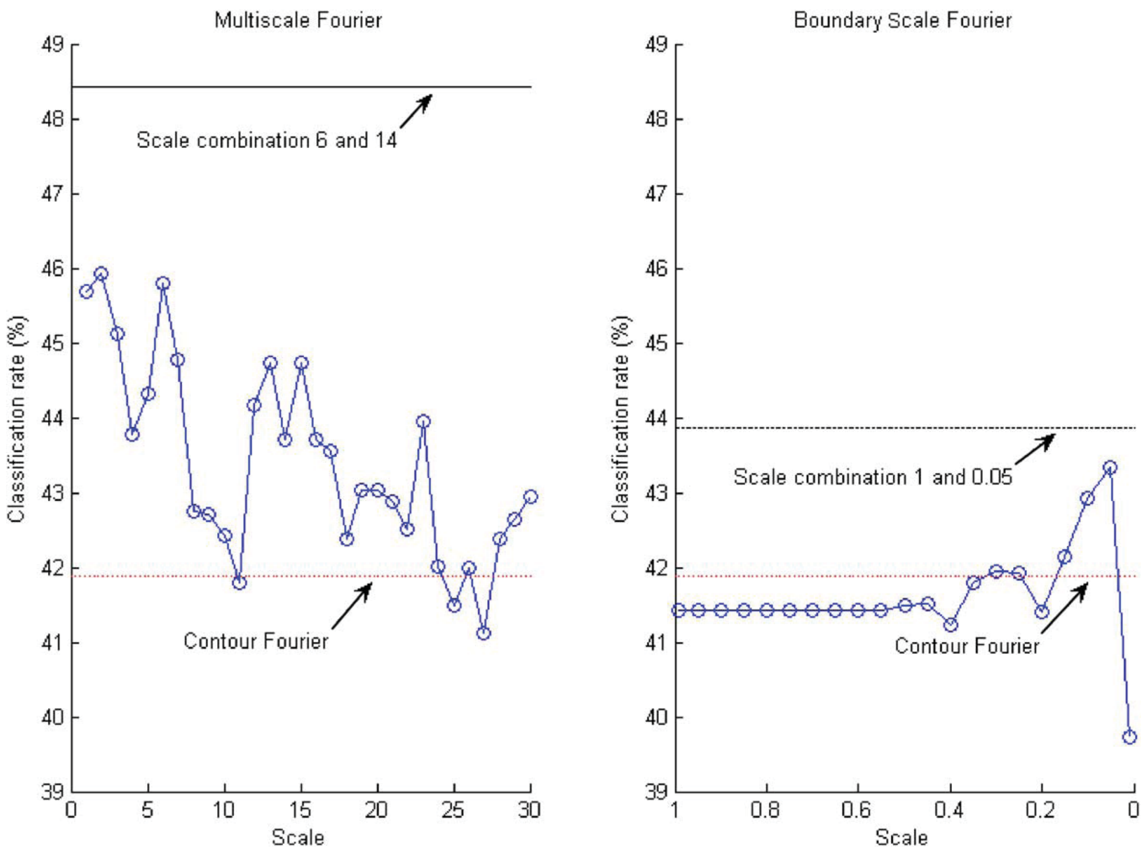

Fig. 4. The average 5-NN classification rates of the proposed methods using different scales. The classification rates of *Contour Fourier* and the selected scale combinations are marked into the figure.

3.3. Results

The results of the classification are presented in Fig. 4, which shows that the scale selection has essential impact on the classification results. In the *Multiscale Fourier* descriptor, small scales give the best classification rates and they clearly outperform *Contour Fourier* in classification. In the case of *Boundary Scale Fourier* descriptor, the large scales are preferable. Furthermore, the classification rates of selected multiscale combinations outperformed clearly the descriptors at any single scale in classification. In addition, both of the proposed descriptors are capable of outperforming *Contour Fourier* in shape classification also at single scales.

Fig. 5 presents the average precision-recall curves for the images in the testing database using different values of Fourier coefficients selected to be used in the shape description (L). The best retrieval performance is achieved using *Multiscale Fourier* at the whole recall scale. The results obtained from the *Boundary Scale Fourier* descriptor outperform *Contour Fourier* with low recall values. It is important to note that in real defect image retrieval only the most similar images to the query image are usually recalled from the database. Therefore, it is not necessary to retrieve all the images in a particular class. For this reason, with a database of about 1000 images or more, the recall area up to 30% is the most interesting from the user's viewpoint. This comparison is performed using fixed values of L, which means that the dimensionalities of the multiresolution approaches are higher than those of *Contour Fourier*. However, it is essential to note that in the multiresolution approaches, the number of Fourier coefficients is not necessary to be as large as in the ordinary Fourier descriptors. This fact is proved in Fig. 6, in which different descriptors

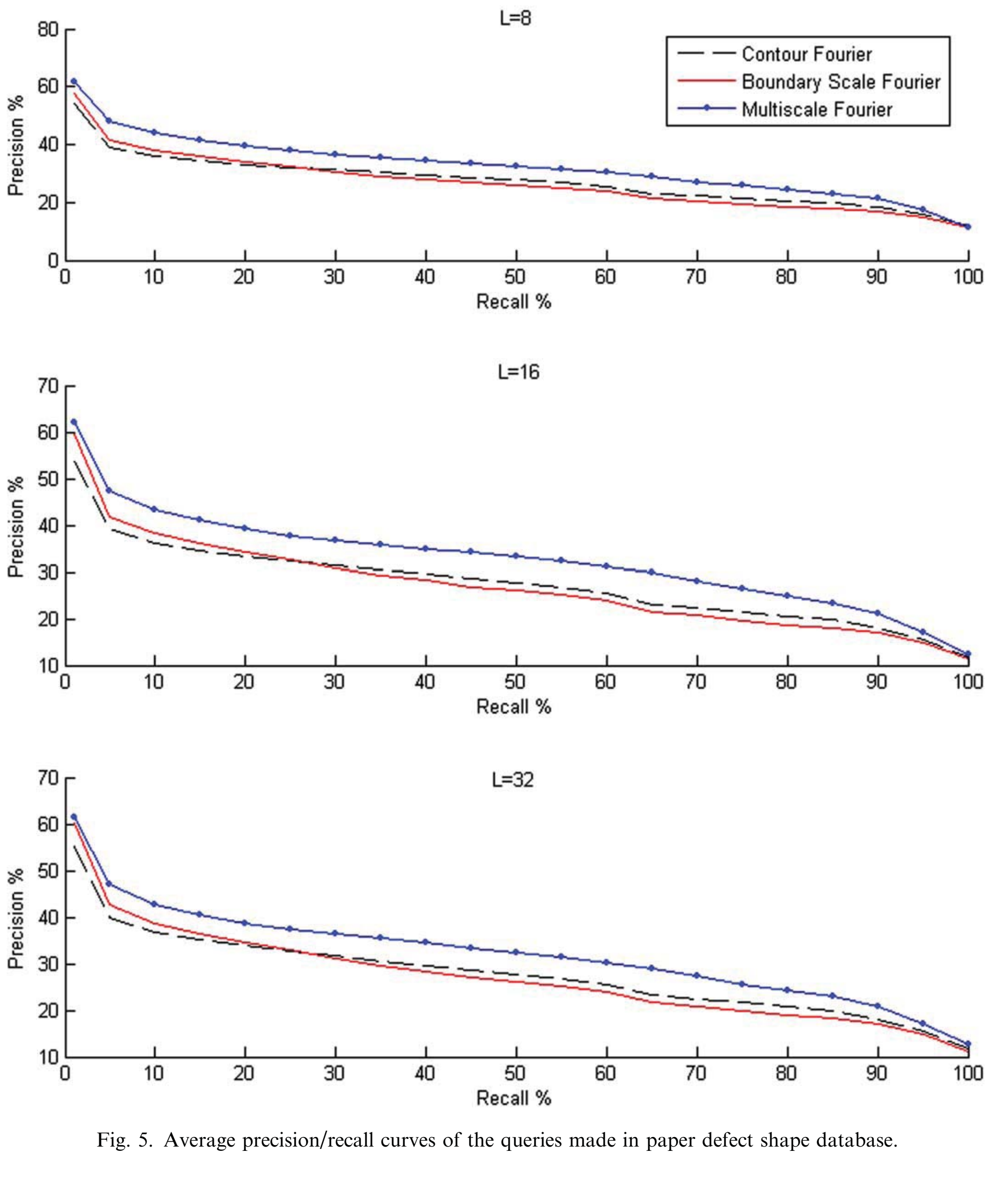

Fig. 5. Average precision/recall curves of the queries made in paper defect shape database.

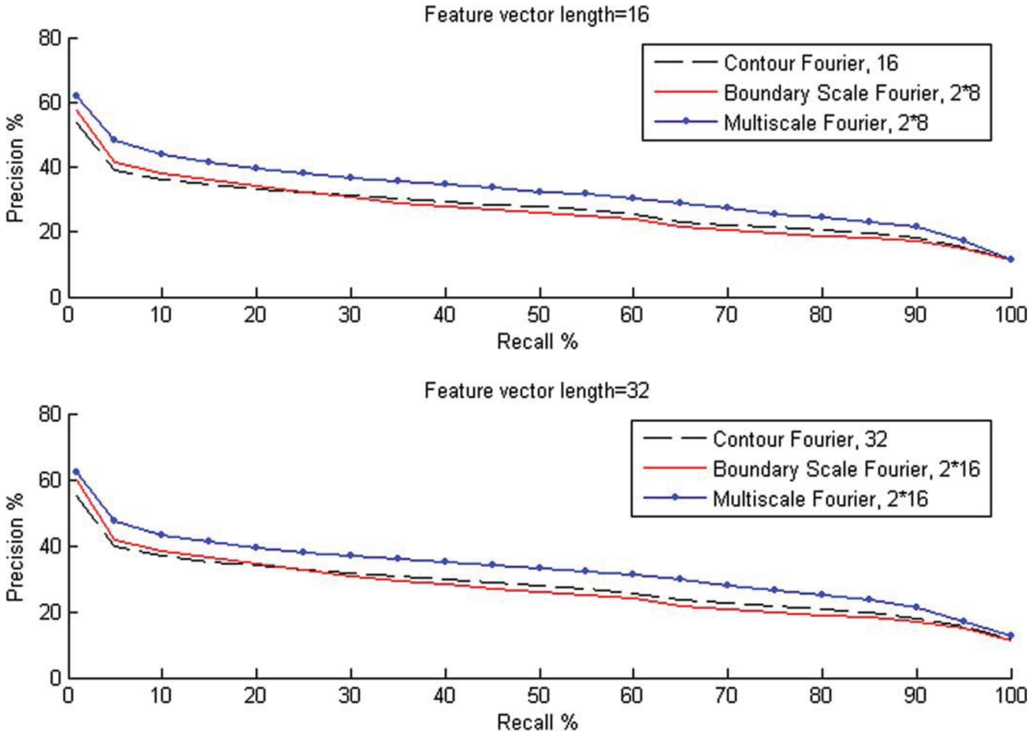

Fig. 6. Comparison of average precision/recall of the descriptors of similar lengths.

of the same dimensionality are compared. Hence, the multiresolution descriptors, *Multiscale Fourier* and *Boundary Scale Fourier* of two scales are compared to *Contour Fourier* descriptor, whose length is similar. The comparison is made with two feature vector lengths. The results presented in Fig. 6 show that *Boundary Scale Fourier* is better than *Contour Fourier* of the same length with low recall values. In the case of *Multiscale Fourier*, the result outperforms the *Contour Fourier* at the whole recall scale. This comparison shows that increased retrieval accuracy can be achieved using multiresolution property of Fourier descriptors without increasing dimensionality.

3.4. Computational efficiency and implementation

The computational cost of the image retrieval methods can be divided into the cost of retrieval process and the cost of feature extraction. When the retrieval process is in all the methods identical (i.e. the same distance metrics is used), the retrieval time is completely dependent on the dimensionality of the feature vectors if the feature values are calculated before the actual retrieval.

In the multiresolution approaches, the dimensionality is dependent on the number of the scales, r. Hence, the dimensionality of the feature vector is r^*L, However, using small number of scales the cost is still reasonable, which makes the *Multiscale Fourier* and *Boundary Scale Fourier* suitable for indexing of large image databases. Furthermore, as discussed in the previous section, the dimensionality of the multiscale approaches does not need to be higher than in the case of the ordinary Fourier descriptors to obtain better retrieval accuracy. Therefore, the use of multiresolution approaches does not necessarily increase the computational cost of retrieval.

Another part of the discussion about the computational cost is related to the complexity of feature extraction. The computational time required by the wavelet transform is relatively low. Being a standard signal processing operation, the computationally efficient implementation of wavelet transform is available in the most common computing tools (Misiti et al., 2001). On the other hand, boundary smoothing that uses conventional filtering tools is also a fast and straightforward operation that is easy to implement.

4. Discussion

An essential matter of the usability of the proposed multiresolution shape descriptors is their capability of being used in on-line retrieval systems. As mentioned in the introduction, six principles (Kim and Kim, 2000) have been set to measure the shape descriptors. The properties of the descriptors presented in this paper are discussed in the terms of these principles.

Several different studies, (among others Zhang and Lu, 2003a,b; Mehtre et al., 1997; Kauppinen et al., 1995) have shown that Fourier descriptor has a very good shape retrieval and classification accuracy also in the case of rotated, translated and scaled shapes. The retrieval results obtained in this study and in (Kunttu et al., 2004) show that the multiresolution property still increases the accuracy of Fourier descriptor. The proposed descriptors provide good retrieval accuracy when only a small number of scales is used. Therefore, they are relatively compact features. As the ordinary Fourier descriptors also the multiresolution Fourier descriptors can be regarded as general tools, because they are capable of classifying several types of complex shapes. In addition to defect shapes, also a common shape database has been used in testing of the multiresolution approach in (Kunttu et al., 2003).

As discussed in Section 3.4, the computational complexity of the methods is reasonable. The computational cost of feature extraction with the multiresolution approaches is somewhat heavier than in the case of ordinary Fourier descriptors. This is particularly a drawback of *Boundary Scale Fourier* descriptor. However, the dimensionality of the descriptors is more essential than the feature extraction time, because in retrieval applications the feature extraction is usually an off-line operation. As discussed in Section 3.3, the dimensionality of the proposed descriptors is not a problem. A significant benefit of the proposed multiresolution descriptors is their straightforward matching. Like the ordinary Fourier descriptors, also the methods presented in this paper can be matched using Euclidean distance. Compared to the complex matching of other multiresolution shape descriptors, like CSS and wavelet descriptors, the use of Euclidean distance is computationally very light operation, especially when the dimensionality is low.

The robustness of a shape descriptor means that the descriptor is able to find also noise-affected, defective and distorted shapes (Zhang and Lu, 2004). In general, when Fourier descriptor is formed using low frequency coefficients, the noise occurring in the high frequencies can be effectively removed. This makes Fourier descriptors insensitive to noise. The proposed multiresolution Fourier descriptors consider the shapes in multiple scales, which make them more insensitive to fine details in the shape boundary than the single-scale Fourier descriptors. In addition, good accuracy in retrieval of quite noisy and in some cases distorted paper defect shapes can perhaps be regarded as an adequate proof of robustness. The hierarchical coarse to fine representation is related to shape matching efficiency (Zhang and Lu, 2004). In the matching process, clearly dissimilar shapes can be eliminated at coarse level, whereas detailed matching is performed at finer levels. This kind of representation can be achieved by selection of the number of the Fourier coefficients and scales. Coarse representation can be obtained using only a few low frequency coefficients and scales. When their number is increased, the representation is finer and the shape is characterized in more details in the descriptor.

5. Conclusions

In this paper, two types of approaches for Fourier-based multiresolution shape descriptors were presented. The methods combine the commonly used and effective Fourier shape description with multiple resolutions, which improves the shape classification and retrieval performance. The multiresolution approaches use wavelet transform and boundary smoothing to produce the multiresolution property to the Fourier descriptor. The proposed descriptors, *Multiscale Fourier* and *Boundary Scale Fourier* produced better retrieval accuracy in experiments than the most powerful single-scale Fourier descriptor, *Contour Fourier*.

In the experiments, a database of complex shapes was used. The shapes in the database represent defects that occur in an industrial process. Using this real-world shape retrieval problem, the proposed methods were validated based on the common criteria that are required in shape similarity-based retrieval. The experiments showed that applying the multiresolution property to Fourier descriptor, additional retrieval accuracy can be achieved without increasing computational cost of retrieval. Therefore, the multiresolution Fourier descriptors presented in this paper are effective and efficient ways of describing complicated shapes in image retrieval.

Acknowledgement

The financial support of Technology Development Centre of Finland (TEKES's grant 40120/03) is gratefully acknowledged.

References

Baeza-Yates, R., Ribeiro-Neto, B., 1999. Modern Information Retrieval. ACM Press, Addison-Wesley, New York.

Bober, M., 2001. MPEG-7 Visual Shape Descriptors. IEEE Trans. Circuits Systems Video Technol. 11 (6), 716–719.

Chui, C.K., 1992. An Introduction to Wavelets. Academic Press, San Diego.

Costa, L.F., Cesar, R.M., 2001. Shape Analysis and Classification, Theory and Practice. CRC Press, Boca Raton, Florida.

Del Bimbo, A., 2001. Visual Information Retrieval. Morgan Kaufmann Publishers, San Fransisco.

Dubois, S.R., Glanz, F.H., 1986. An autoregressive model approach to two-dimensional shape classification. IEEE Trans. Pattern Anal. Machine Intell. 8, 55–66.

Freeman, H., Davis, L.S., 1977. A corner finding algorithm for chain coded curves. IEEE Trans. Comput. 26, 297–303.

Gonzalez, R.C., Woods, R.E., 1993. Digital Image Processing. Addison Wesley.

Granlund, G.H., 1972. Fourier preprocessing for hand print character recognition. IEEE Trans. Comput. C-21 (2), 195–201.

Hu, M.K., 1962. Visual pattern recognition by moment invariants. IRE Trans. Inf. Theory IT-8, 179–187.

Iivarinen, J., Visa, A., 1998. An adaptive texture and shape based defect classification. Proc. 14th Internat. Conf. on Pattern Recognition, Vol. 1, pp. 117–122.

Kauppinen, H., Seppänen, T., Pietikäinen, M., 1995. An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification. IEEE Trans. Pattern Anal. Machine Intell. 17 (2), 201–207.

Kim, H., Kim, J., 2000. Region-based shape descriptor invariant to rotation, scale and translation. Signal Process. Image Comm. 16 (1), 87–93.

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2003. Multiscale Fourier descriptor for shape classification. Proc. of 12th Internat. Conf. on Image Analysis and Processing, pp. 536–541.

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2004. Multiscale Fourier descriptor for shape-based image retrieval. Proc. of 17th Internat. Conf. of Pattern Recognition, Vol. 2, pp. 765–768.

Loncaric, S., 1998. A survey of shape analysis techniques. Pattern Recognition 31 (8), 983–1001.

Manjunath, B.S., Salembier, P., Sikora, T., 2002. Introduction to MPEG-7 Multimedia Content Description Interface. John Wiley and Sons, UK.

Mehtre, B.M., Kankanhalli, M.S., Lee, W.F., 1997. Shape measures for content based image retrieval: a comparison. Inf. Process. Manage. 33 (3), 319–337.

Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.-M., 2001. Wavelet Toolbox for Use with Matlab. Mathworks Inc.

Mokhtarian, F., 1995. Silhouette-based isolated object recognition through curvature scale space. IEEE Trans. Pattern Anal. Machine Intell. 17 (5), 539–544.

Mokhtarian, F., Mackworth, A.K., 1986. Scale-based description and recognition of planar curves and two-dimensional shapes. IEEE Trans. Pattern Anal. Machine Intell. 8 (1), 34–43.

Persoon, E., Fu, K., 1977. Shape discrimination using Fourier descriptors. IEEE Trans. Systems Man Cyber. 7, 170–179.

Rauhamaa, J., Reinius, R., 2002. Paper web imaging with advanced defect classification. Proc. 2002 TAPPI Technology Summit.

Teague, M.R., 1980. Image analysis via the general theory of moments. J. Opt. Soc. Amer. 70 (8), 920–930.

Teolis, A., 1998. Computational Signal Processing with Wavelets. Birkhauser.

Tieng, Q.M., Boles, W.W., 1997. Recognition of 2D object contours using the wavelet transform zero-crossing representation. IEEE Trans. Pattern Anal. Machine Intell. 19 (8), 910–916.

Yang, H.S., Lee, S.U., Lee, K.M., 1998. Recognition of 2D object contours using starting-point-independent wavelet coefficient matching. J. Vis. Comm. Image Represent. 9 (2), 171–181.

Zhang, D.S., Lu, G., 2003a. A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval. J. Vis. Comm. Image Represent. 14 (1), 41–60.

Zhang, D.S., Lu, G., 2003b. Evaluation of MPEG-7 shape descriptors against other shape descriptors. Multimedia Syst. 9 (1), 15–30.

Zhang, D.S., Lu, G., 2004. Review of shape representation and description techniques. Pattern Recognition 37 (1), 1–19.

Publication IV

Kunttu, I., Lepistö, L., Visa, A., 2005. An Efficient Fourier Shape Descriptor for Industrial Defect Images Using Wavelets.

Efficient Fourier shape descriptor for industrial defect images using wavelets

Iivari Kunttu
Leena Lepistö
Ari Visa, MEMBER SPIE
Tampere University of Technology
Institute of Signal Processing
P.O. Box 553
FI-33101 Tampere, Finland
E-mail: Iivari.Kunttu@tut.fi

Abstract. The use of image retrieval and classification has several applications in industrial imaging systems, which typically use large image archives. In these applications, the matter of computational efficiency is essential and therefore compact visual descriptors are necessary to describe image content. A novel approach to contour-based shape description using wavelet transform combined with Fourier transform is presented. The proposed method outperforms ordinary Fourier descriptors in the retrieval of complicated industrial shapes without increasing descriptor dimensionality.
[DOI: 10.1117/1.1993687]

Subject terms: shape retrieval; Fourier descriptors; wavelet descriptors; defect images.

Paper 050289 LR received Apr. 22, 2005; revised manuscript received May 23, 2005; published online Aug. 10, 2005.

1 Introduction

The recognition and classification of objects based on their visual similarity has become a central task in current industrial imaging systems. With increasing amounts of real-world image data to be processed and stored, the development of powerful retrieval tools also has become necessary in machine vision applications. Along with texture and color, shape is an essential feature used to describe the objects in the images. Therefore, effective shape description is essential in retrieval systems.

Due to the increasing number of on-line solutions, computational lightness is nowadays considered equally important as classification accuracy. In retrieval, computational efficiency of a particular descriptor is generally dependent on two matters, descriptor dimensionality and matching procedure.

The Fourier descriptor (FD)[1] is probably the best-known boundary-based shape descriptor. It has been proven to outperform most other boundary-based methods in terms of retrieval accuracy and efficiency.[2] In addition to good retrieval and classification performance, the main advantages of FDs are that (1) they are compact and computationally light, (2) they are easy to implement, (3) their matching is straightforward, (4) they are very easy to normalize to be scale and rotation invariant, and (5) their sensitivity to noise is low.

0091-3286/2005/$22.00 © 2005 SPIE

Wavelet transforms[3] have been widely used in multiscale image analysis and also have a few applications in shape description. In Ref. 4, the wavelet descriptors (WDs) are based on zero-crossing points of wavelet approximation of the shape and hence the similarity measurement is dependent on the shape complexity. In Ref. 5, moment invariants are employed in shape description using wavelets. It is also possible to combine wavelets with Fourier descriptors, which yields to rotation and scale invariance. This can be made based on polar coordinates of a shape[6] or by Fourier transforming the wavelet coefficients obtained from the complex-valued boundary function.[7] On the other hand, when WDs are formed using several scales, the resulting feature vector is typically high dimensional due to spatial information caused by multiple scales.

In this paper, we present an effective approach to wavelet-based shape representation at single scale. We show that it is possible to form rotation and translational invariant WDs, whose matching is as simple and fast as that of FDs. The proposed approach is applied to a practical industrial image retrieval and classification problem.

2 Shape Description

The contour-based shape description is based on one-dimensional boundary function (shape signature). Let (x_k, y_k), $k=0,1,2,\dots,N-1$ represent the object boundary coordinates, in which N is the boundary length. Complex coordinate function $z(k)$ (Ref. 2) expresses the boundary points in an object centered coordinate system:

$$z(k) = (x_k - x_c) + j(y_k - y_c) \tag{1}$$

in which (x_c, y_c) is the object centroid.

2.1 *Fourier Descriptors*

Fourier descriptors can be formed for the boundary function $z(k)$ using the discrete Fourier transform (DFT):

$$F_n = \frac{1}{N}\sum_{k=0}^{N-1} z(k)e^{-j2\pi nk/N} \tag{2}$$

for $n=0,1,2,\dots,N-1$ and F_n are the transform coefficients of $z(k)$. The descriptors can be made rotation invariant using the magnitudes of the transform coefficients, $|F_n|$. The scale can be normalized by dividing the magnitudes of the coefficients by $|F_1|$.

The general shape of the object is represented by the low-frequency coefficients, which are usually selected to be the descriptor. In the contour Fourier method,[2] the feature vector of length L is formed as:

$$x = \left[\frac{|F_{-(L/2-1)}|}{|F_1|},\dots,\frac{|F_{-1}|}{|F_1|},\frac{|F_2|}{|F_1|},\dots,\frac{|F_{L/2}|}{|F_1|}\right]^T. \tag{3}$$

2.2 *Wavelet Shape Descriptor Using Fourier Transform*

In the wavelet-based approach, the boundary function $z(k)$ is transformed using some wavelet Ψ.[3] The complex wavelet transform[8] is based on the continuous wavelet transform (CWT). The CWT of the boundary $z(k)$ is defined as:

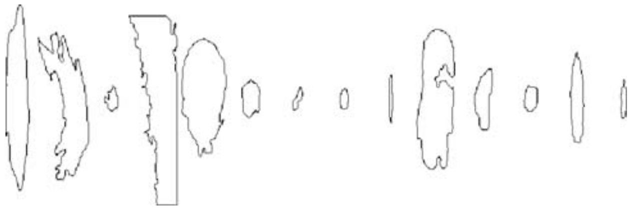

Fig. 1 Example contours of each 14-paper defect class in the testing database.

$$C_a(b) = \frac{1}{\sqrt{|a|}} \int_R z(k)\psi\left(\frac{k-b}{a}\right) dk. \quad (4)$$

In the case of CWT, a set of coefficients $C_a(b)$ of scale a are obtained. The coefficients are defined for all positions $b=0,1,2,\ldots,N-1$.

The problem with the CWT coefficients is that they are dependent on the starting point of the object boundary. Hence, the obtained descriptor is not rotation invariant. Also the dimensionality of the feature vector depends on the boundary length. Therefore, the coefficient vectors of different shapes cannot be directly matched. The proposed solution for this problem is to apply the Fourier transform to the whole set of wavelet coefficients. This way the normalization and matching are straightforward operations. The proposed descriptor is formed by applying the DFT to the coefficients $C_a(b)$:

$$F_n^a = \frac{1}{N} \sum_{b=0}^{N-1} C_a(b) e^{-j2\pi nb/N}. \quad (5)$$

In this paper, we use the wavelet shape descriptor at a single scale to keep the descriptor dimensionality as low as in the case of Fourier descriptors. The feature vector of this new descriptor is equal to that of the contour Fourier descriptor presented in Eq. (3).

3 Experiments with Industrial Defect Shapes

The validation presented in this section is twofold. Simple classification experiments are first carried out to show the influence of scale selection on the shape description. The second part of the validation, the retrieval accuracy of the proposed methods, is compared to that of an ordinary FD (contour Fourier). In all the experiments, Euclidean distance and the "*leave one out*" validation principle are used.

3.1 *Testing Database*

For testing purposes, we use defect images that are collected from an industrial process using a paper inspection system.[9] A reason for collecting defect image databases in process industry is a practical need for controlling the quality of production.[9] When retrieving images from a database, the defect shape is one essential property describing the defect class. Therefore, effective methods for the shape representation are necessary. The test set consisted of 1204 paper defect shapes, which represented 14 defect classes with each class consisting of 27–103 images (Fig. 1).

3.2 *Classification and Retrieval*

The feature extraction in the testing database was carried out by calculating the descriptors for the images in the database. The dimensionality (L) was 8 with all the descriptors [Eq. (3)]. In the case of the wavelet-based approach, the selected wavelets ψ were first and second order complex Gaussian wavelets that have been implemented in the Matlab wavelet toolbox.[8] To compare different scales, we made preliminary k-nearest neighbor (k-NN) classification experiments. Figure 2(a) presents the average classification rates of the proposed wavelet descriptors at different scales using a 5-NN classifier. In this figure, the classification rate

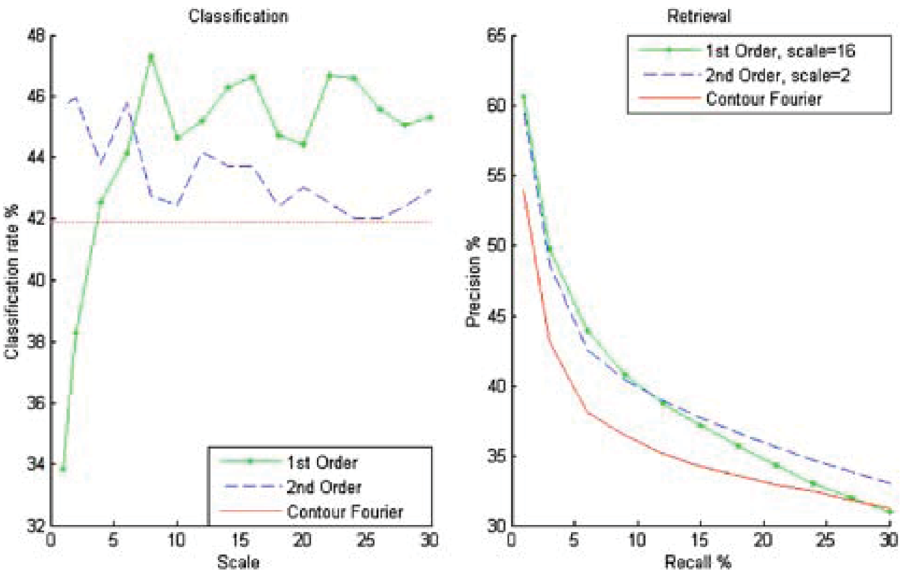

Fig. 2 (a) The average 5-NN classification rates of the proposed methods using different scales of the first and second order complex Gaussian wavelets. (b) Average precision/recall curves of the queries using proposed descriptors that employ the first and second order complex Gaussian wavelets the scales 16 and 2, respectively.

of the contour Fourier descriptor (41.87%) is also presented. The scales that produce the highest classification rates were compared to contour Fourier in the retrieval experiment by calculating average precision versus recall curves for the queries [Fig. 2(b)].

4 Discussion

In this paper, we showed that it is possible to overcome the difficulties with shape description using wavelet coefficients (rotational variance and complicated matching) by Fourier transforming the coefficients. The results of the classification and retrieval experiments reveal that the proposed wavelet-based shape description approach clearly outperforms ordinary FDs in defect shape description. It is also essential to note that the proposed descriptors have the same dimensionality and matching procedure as FDs. The computational cost of the feature extraction is somewhat higher than that of FDs due to the wavelet transform. However, the dimensionality of the descriptors is more essential than the feature extraction time, because in retrieval applications the feature extraction is usually an off-line operation. If the computational efficiency of feature extraction is critical, the cost of wavelet transform can be decreased using the algorithm presented in Ref. 10.

Acknowledgment

The authors wish to thank ABB Oy (Mr. Juhani Rauhamaa) for the paper defect image database used in the experiments.

References

1. C. T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Comput.* **C-21**(3), 269–281 (1972).
2. H. Kauppinen, T. Seppänen, and M. Pietikäinen, "An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(2), 201–207 (1995).
3. C. K. Chui, *An Introduction to Wavelets*, Academic Press, San Diego (1992).
4. Q. M. Tieng and W. W. Boles, "Recognition of 2D object contours using the wavelet transform zero-crossing representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(8), 910–916 (1997).
5. D. Shen and H. Ip, "Discriminative wavelet shape descriptors for recognition of 2-D patterns," *Pattern Recogn.* **32**, 151–165 (1999).
6. G. Chen and T. Bui, "Invariant Fourier-wavelet descriptors for pattern recognition," *Pattern Recogn.* **32**, 1083–1088 (1999).
7. I. Kunttu, L. Lepistö, J. Rauhamaa, and A. Visa, "Multiscale Fourier descriptor for shape-based image retrieval," *Proc. 17th Intl. Conf. Patt. Recog.*, Vol. **2**, pp. 765–768 (2004).
8. M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, *Wavelet Toolbox for Use with Matlab*, Mathworks Inc. (2001).
9. J. Rauhamaa and R. Reinius, "Paper web imaging with advanced defect classification," *Proc. 2002 TAPPI Technology Summit*, (2002).
10. F. Nicolier, O. Laligant, and F. Truchetet, "Discrete wavelet transform implementation in Fourier domain for multidimensional signal," *J. Electron. Imaging* **11**(3), 338–346 (2002).

Publication V

Kunttu, I., Lepistö, L., Visa, A., 2005. Enhanced Fourier Shape Descriptor Using Zero-Padding.

Enhanced Fourier Shape Descriptor Using Zero-Padding

Iivari Kunttu, Leena Lepistö, and Ari Visa

Tampere University of Technology,
Institute of Signal Processing,
P.O. Box 553, FI-33101 Tampere Finland
{Iivari.Kunttu, Leena.Lepisto, Ari.Visa}@tut.fi
http://www.tut.fi

Abstract. The shapes occurring in the images are essential features in image classification and retrieval. Due to their compactness and classification accuracy, Fourier-based shape descriptors are popular boundary-based methods for shape description. However, in the case of short boundary functions, the frequency resolution of the Fourier spectrum is low, which yields to inadequate shape description. Therefore, we have applied zero-padding method for the short boundary functions to improve their Fourier-based shape description. In this paper, we show that using this method the Fourier-based shape classification can be significantly improved.

1 Introduction

The description of the object shape is an important task in image analysis and pattern recognition. The shapes occurring in the images have also a remarkable significance in image classification and retrieval. The basic problem in shape classification is to define similarity between two shapes. Therefore, different visual features (descriptors) have been developed to characterize the shape content of the images. Common shape description techniques have been reviewed in a recent study of Zhang and Lu [17]. Another review of the state of the art in shape description techniques is provided by Loncaric [9].

The shape description techniques can be divided into two types, boundary based and region based techniques [1]. The region based methods consider the whole area of the object whereas the boundary based shape descriptors use only the object boundary in the shape description. The most popularly used region-based methods are different moments [5],[15]. The best-known boundary based shape descriptors include chain codes [3] and Fourier descriptors [13]. Also autoregressive (AR) [2] models have been used in shape description. Simple shape features such as circularity [1], eccentricity, convexity, principle axis ratio, circular variance and elliptic variance [6] include boundary-based descriptors. Recently, growing research interest has been focused on Curvature Scale Space (CSS) shape representation [11] that has been selected to be used as the boundary-based shape descriptor of MPEG-7 standard.

H. Kalviainen et al. (Eds.): SCIA 2005, LNCS 3540, pp. 892 – 900, 2005.

However, despite the fact that Fourier descriptor is over 30 years old method [4],[13], it is still found to be valid shape description tool. In fact, Fourier descriptor has proved to outperform most other boundary-based methods in terms of classification accuracy and efficiency. This has been verified in several comparisons. Kauppinen et al. [7] made a comparison between autoregressive models and Fourier descriptors in shape classification. In most cases, Fourier descriptors proved to be better in the classification of different shapes. In the comparison made by Mehtre et al. [10], the accuracy of chain codes, Fourier descriptors, and different moments was compared in the shape retrieval. In this case, best results were obtained using moments and Fourier descriptors. In a recent study of Zhang and Lu [16], Fourier descriptors and Zernike moments outperformed CSS representation in terms of retrieval accuracy and efficiency. Similar results were obtained also in [8], in which Fourier descriptors outperformed CSS in defect shape retrieval.

In addition to good classification and retrieval performance, there also other reasons which make Fourier descriptors probably the most popular of the boundary-based shape representations. The main advantages of the Fourier-based shape descriptors are that they are compact and computationally light methods with low dimensionality. Furthermore, they are easy to normalize and their matching is a very simple process. Also their sensitivity to noise is low when only low frequency Fourier coefficients are used as descriptors.

In this paper, the area of Fourier shape descriptors is revisited. We present a method for enhancing the performance of Fourier-based shape description by increasing frequency resolution of the Fourier spectrum calculated for the boundary function of an object shape. Using this method, a more accurate shape representation in frequency domain can be achieved. This is particularly beneficial in the case of objects with relatively short boundary function, in which cases the spectrum estimate has low resolution. The experiments presented in this paper prove that using this technique, the shape classification accuracy can be easily improved.

2 Shape Representation Using Fourier Descriptors

In this paper, the shape description methods are based on the boundary line of the object. The boundary can be presented using some shape signature i.e. function derived from the boundary coordinates of the object [1]. Complex coordinate function is a well-known shape signature [7]. It presents the boundary coordinates in an object centered complex coordinate system. Let (x_k, y_k), k=0,1,2,…,N-1 represent the boundary coordinates, in which N is the length of the boundary. The complex coordinate function $z(k)$ expresses the boundary points in an object centered coordinate system in which (x_c, y_c) represents the centroid of the object:

$$z(k) = (x_k - x_c) + j(y_k - y_c) \tag{1}$$

Hence, using this function, the boundary is represented independent of the location of the object in the image. In this way the translation invariance can be achieved.

2.1 Fourier Description of the Boundary Function

Fourier descriptors characterize the object shape in a frequency domain. The descriptors can be formed for the complex-valued boundary function using the discrete Fourier transform (DFT). The Fourier transform of a boundary function generates a set of complex numbers, which characterize the shape in frequency domain. Fourier transform of $z(k)$ is:

$$F(n) = \frac{1}{N}\sum_{k=0}^{N-1} z(k) e^{-j2\pi nk/N} \tag{2}$$

for n=0,1,2,…,N-1. The transform coefficients $F(n)$ form the Fourier spectrum of the boundary function. The translational invariance of this shape representation is based on the object centered shape signature. Furthermore, the coefficients have also to be normalized to achieve invariance to rotation and scaling. The descriptors can be made rotation invariant by ignoring the phase information and using only the magnitudes of the transform coefficients $|F(n)|$. In the case of complex-valued boundary function, the scale can be normalized by dividing the magnitudes of the transform coefficients by $|F(1)|$ [7].

2.2 Zero-Padding Method

Even if Fourier descriptor is a powerful tool of boundary-based shape description, its performance is somewhat dependent on the frequency resolution of the Fourier spectrum. When the boundary function $z(k)$ is Fourier transformed, the resulting Fourier spectrum is of the same length as boundary function. Therefore, in the case of short boundary functions the frequency resolution is also low. To obtain better resolution, the number of the datapoints in the boundary function should be increased. In practice, this is not always feasible, because the boundary lines of the objects are usually defined pixel-by-pixel, and the number of the boundary points depends on the image resolution. However, there is an alternative approach for this purpose. Zero-padding [12] is a commonly used method in signal processing. It can be used to increase the frequency resolution by adding zeros to the function to be Fourier transformed. Hence, a new function is defined as:

$$z_{zp}(k) = \begin{cases} z(k) & for\; 0 \le k \le N-1 \\ 0 & for\; N \le k \le N_{zp}-1 \end{cases} \tag{3}$$

in which N_{zp} is the length of a desired frequency spectrum. By using additional zeros in the input signal of DFT, new spectrum values are being interpolated among the original values in the spectrum. This way, the density of the frequency samples is increased in the spectrum. In practice, the desired spectrum length is selected such that $N_{zp}=2^p$ in which p is a positive integer. This is beneficial because DFT is usually implemented using FFT algorithm, in which input functions of length 2^p are preferred to decrease computing time.

2.3 Descriptors

The Fourier spectrum represents the frequency content of the boundary function. General shape of the object is represented by the low frequency coefficients of $F(n)$, whereas high frequency coefficients represent the fine details in the object shape. A common approach to shape representation is the use of a subset of low-frequency coefficients as a shape descriptor. Consequently the shape can be effectively represented using a relatively short feature vector. In our experiments, the feature vector is formed using *Contour Fourier* method [7], which applies the complex coordinate function. In the case of complex valued boundary functions, the coefficients are taken both positive and negative frequency axis. The feature vector is formed as:

$$x = \left[\frac{|F_{-(L/2-1)}|}{|F_1|} \cdots \frac{|F_{-1}|}{|F_1|} \, \frac{|F_2|}{|F_1|} \cdots \frac{|F_{L/2}|}{|F_1|} \right]^T \tag{4}$$

where L is a constant value that defines the feature vector length (dimensionality).

3 Experiments

In this paper, we make experiments to demonstrate the effect of the zero-padding on the shape classification performance. In this experimental part we use a database of industrial defect shapes as a test set.

3.1 Testing Database

For testing purposes, we used defect images that were collected from a real industrial process using a paper inspection system [14]. A reason for collecting defect image databases in process industry is a practical need for controlling the quality of production [14]. When retrieving images from a database, the defect shape is one essential property describing the defect class. Therefore, effective methods for the shape representation are necessary. The defects occurring in paper can be for example holes, wrinkles or different kinds of thin or dirt spots. The test set consisted of 1204 paper defects which represented 14 defect classes with each class consisting of 27-103 images. Three example contours of each defect class are presented in figure 1. Within each class, there are defects of different size and orientation. Furthermore, in some classes the boundaries are very varying and sometimes distorted (classes 2, 4 and 10, for example).

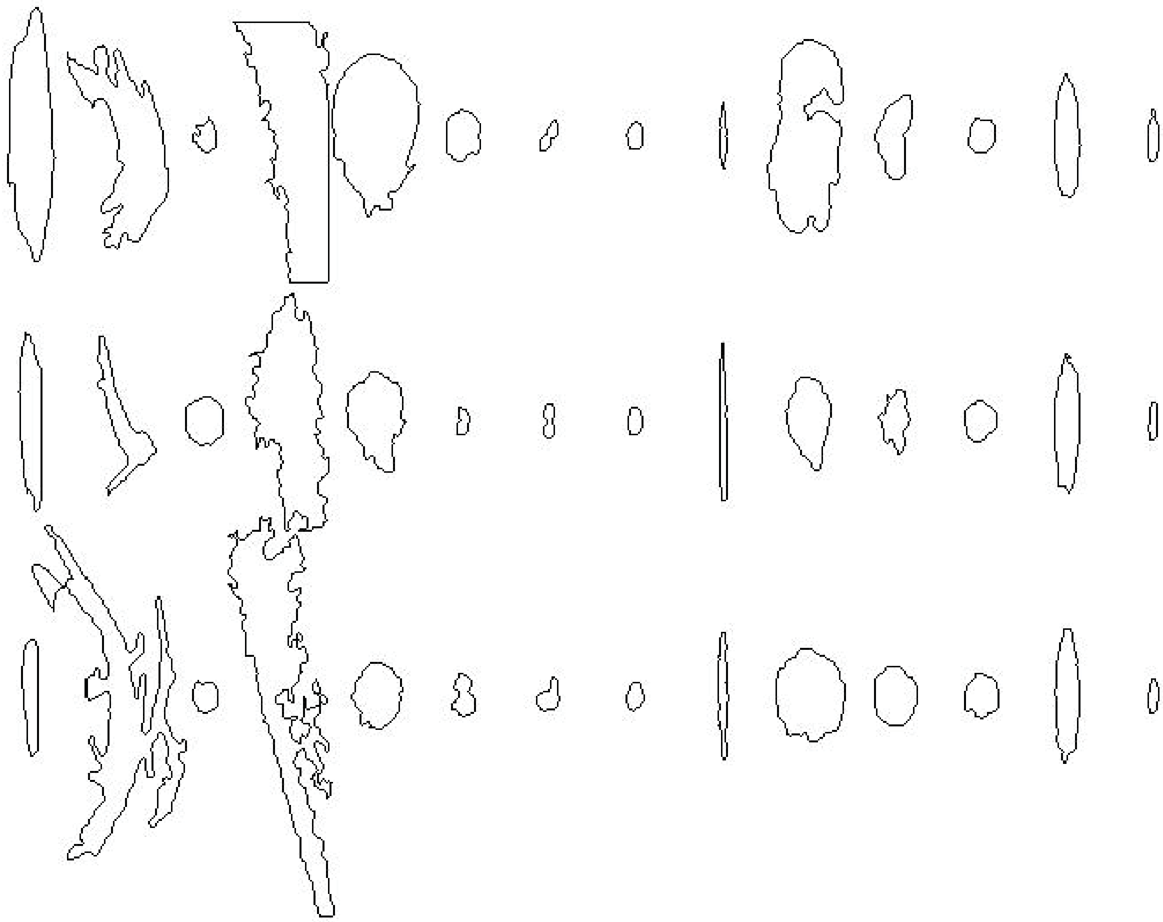

Fig. 1. Three example contours of each 14 paper defect class in the testing database

3.2 Classification

The feature vectors for the shapes in the database were calculated for ordinary *Contour Fourier* descriptors as well as for the same descriptors using zero-padding. The lengths of the boundary lines of the defect shapes were very varying, which can be seen in figure 2. In this figure, the lengths of each 1204 defect boundaries are presented. In our experiments, the boundaries of lower lengths than N_{zp} were inserted with zeros. Two values of N_{zp} were used, namely 512 and 1024. However, preliminary experiments have showed that in some cases zero-padding also decreases the Fourier-based shape distinction. This can be avoided by emphasizing the zero-padding only to the shortest boundaries in the test set. Therefore, we made an additional experiment, in which only very short boundaries whose length was less than 100 points, were used. These boundaries were zero-padded quite strongly, to 1024 points. The length of the Fourier descriptor (L) was selected to be 16 in all the experiments. It is important to note that the use of zero-padding has no influence on the descriptor dimensionality.

In classification, we used 5-nearest neighbor classifier and leave-one-out validation. The distance metrics was selected to be Euclidean distance, which is a standard

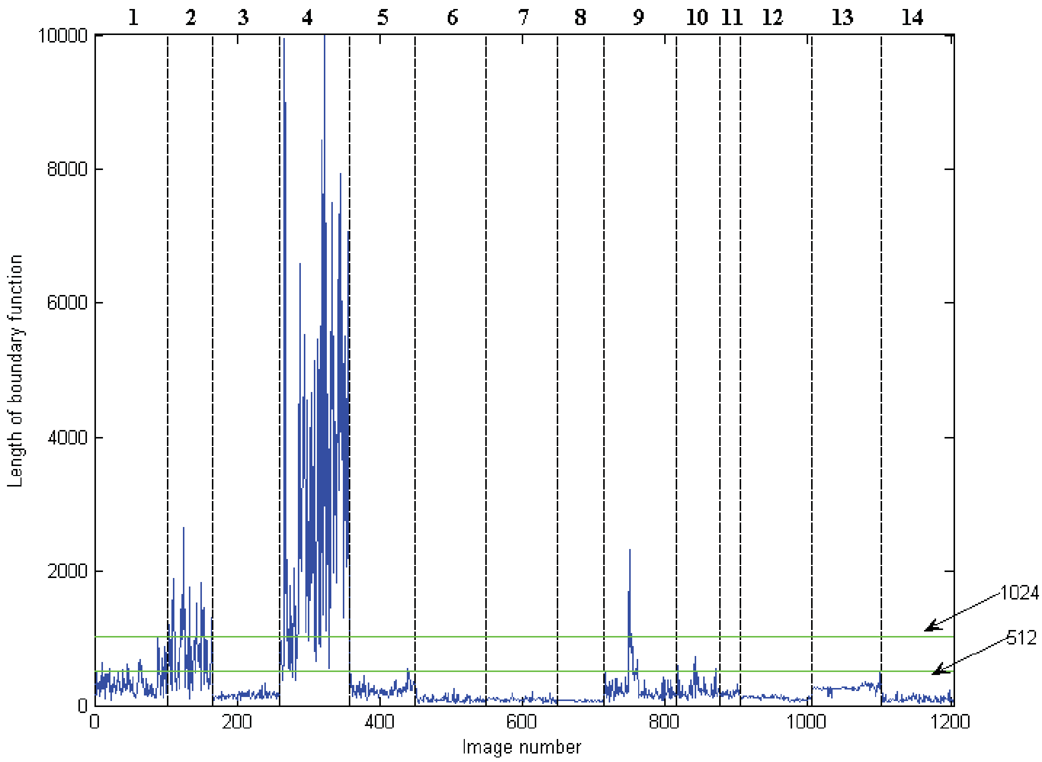

Fig. 2. The lengths of the boundary functions of 1204 paper defects in the testing database. The numbers above the plot correspond to the classes

approach with Fourier descriptors. We carried out the classification using four selected methods, which were ordinary *Contour Fourier*, *Contour Fourier* with zero-padding to 512 of 1024 points, and *Contour Fourier* with zero padding of very short boundaries to 1024 points. Average classification rates of the descriptors are presented for each 14 defect class in figure 3.

3.3 Results

The results presented in figure 3 show that the zero-padding method is able to improve the shape classification rates in most of the defect classes. Particularly, in the classes containing short boundaries the overall result was improved. Therefore, the application of the zero-padding was capable of increasing the average classification performance from 41.2 % to 52.2 %. The best average result was achieved using the adaptive zero-padding of very short boundaries to 1024 datapoints. However, all the zero-padding approaches were able to improve the classification results, especially in the classes with short boundaries. For example, in class 11 the classification rate was improved from zero to over 30 %.

According to the obtained results, it seems that the most recommendable approach is to use the zero-padding method only to the shortest boundaries, though the other presented approaches produce clear improvement in classification performance as well.

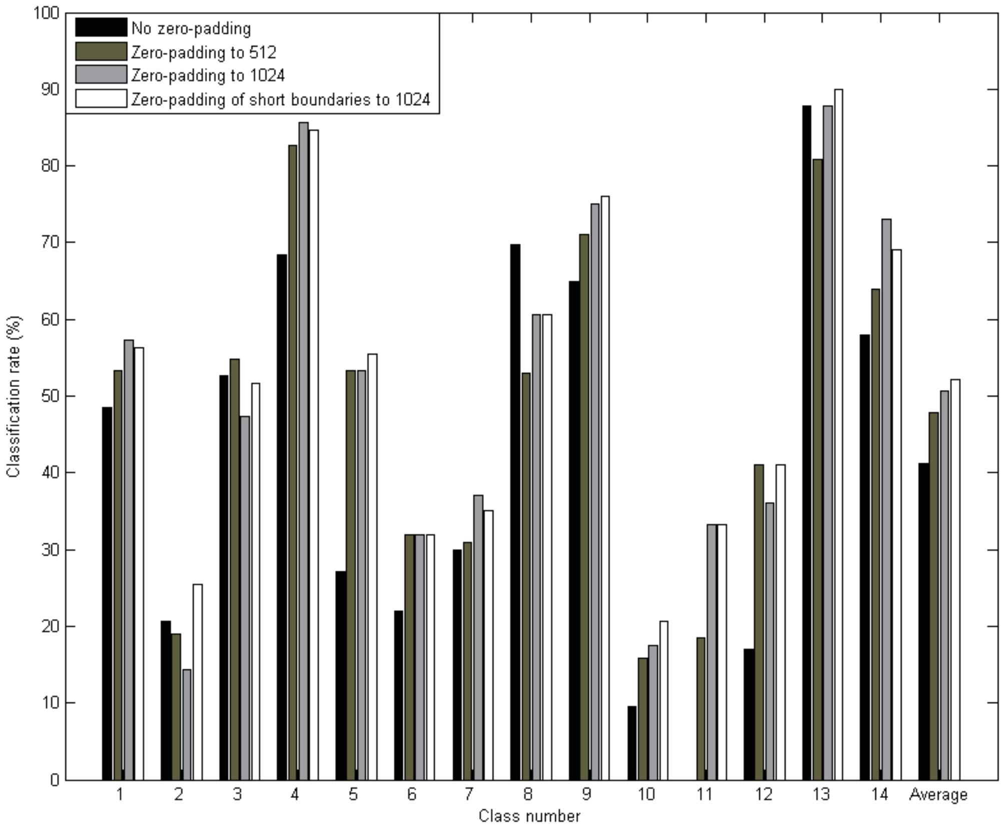

Fig. 3. The average results of 5-NN classification in each 14 defect class

4 Discussion

In this paper, a method for improving Fourier-based shape description was presented. The proposed method is a simple and fast tool for improving the shape distinction ability of Fourier descriptors. It is particularly beneficial in the case of short boundary functions, which do not produce adequate frequency resolution to their Fourier spectrum. The zero-padding method as itself is not a novel method, because it has been applied to different kinds of signal processing problems before. However, the new way in which it has been employed to improve shape description has a certain practical value in different shape classification problems.

For experimental purposes, we used industrial defect images, which are quite complicated shape classification task. This is due to the irregularities and variations in the defect shapes. Some of the classes are also somewhat overlapping. It is essential to note that in real defect image classification the shape is not the only classifying feature, because also texture and gray level distribution play a role in defect image description.

The experimental results obtained from this real-world shape classification problem show that using zero-padding the classification performance can be significantly improved. This, on the other hand, does not increase the computational cost, because the dimensionality of the feature vectors remains the same. Zero-padding method does not require additional computational capacity in feature extraction, either. This is due to the advanced FFT algorithms [12].

In conclusion, the zero-padding method has proved to an effective tool for enhancing Fourier-based shape description. It is especially effective with short boundary functions of complicated shapes.

Acknowledgment

The authors wish to thank ABB Oy (Mr. Juhani Rauhamaa) for the paper defect image database used in the experiments.

References

1. Costa, L.F., Cesar, R.M.: Shape Analysis and Classification, Theory and Practice, CRC Press, Boca Raton, Florida (2001).
2. Dubois, S.R., Glanz, F.H.: An Autoregressive Model Approach to Two-Dimensional Shape Classification, IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 8 (1986) 55-66
3. Freeman, H., Davis, L.S.: A Corner Finding Algorithm for Chain Coded Curves, IEEE Transactions on Computers Vol. 26 (1977) 297-303
4. Granlund, G.H.: Fourier Preprocessing for Hand Print Character Recognition, IEEE Transactions on Computers Vol. C-21, No. 2 (1972) 195-201
5. Hu, M.K.: Visual Pattern Recognition by Moment Invariants, IRE Transactions on Information Theory Vol. IT-8 (1962) 179-187
6. Iivarinen, J., Visa, A.: An adaptive texture and shape based defect classification, Proceedings of the 14th International Conference on Pattern Recognition, Vol. 1, (1998) 117-122
7. Kauppinen, H., Seppänen, T., Pietikäinen, M.: An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 2 (1995) 201-207
8. Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A.: Multiscale Fourier Descriptor for Shape-Based Image Retrieval, Proceedings of 17th International Conference on Pattern Recognition, Vol. 2 (2004) 765-768
9. Loncaric, S.: A survey of shape analysis techniques, Pattern Recognition Vol. 31, No. 8 (1998) 983-1001
10. Mehtre, B.M., Kankanhalli, M.S., Lee, W.F.: Shape Measures for Content Based Image Retrieval: A Comparison, Information Processing Management, Vol. 33, No 3, (1997) 319-337
11. Mokhtarian, F., Mackworth, A.K.: Scale-based description and recognition of planar curves and two-dimensional shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 1 (1986) 34-43
12. Orfanidas, S.J.: Introduction to Signal Processing, Prentice Hall (1996)
13. Persoon, E., Fu, K.: Shape Discrimination Using Fourier Descriptors, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 7 (1977) 170-179
14. Rauhamaa, J., Reinius, R.: Paper Web Imaging with Advanced Defect Classification, Proceedings of the 2002 TAPPI Technology Summit (2002)

15. Teague, M.R.: Image Analysis via the General Theory of Moments, Journal of Optical Society of America, Vol. 70, No. 8 (1980) 920-930
16. Zhang, D., Lu, G.: A Comparative Study of Curvature Scale Space and Fourier Descriptors for Shape-Based image Retrieval, Journal of Visual Communication and Image Representation, Vol. 14, No. 1 (2003) 41-60
17. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition, Vol. 37, No. 1, (2004) 1-19

Publication VI

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2003. Binary Co-occurrence Matrix in Image Database Indexing,

Binary Co-occurrence Matrix in Image Database Indexing

Iivari Kunttu[1], Leena Lepistö[1], Juhani Rauhamaa[2], and Ari Visa[1]

[1] Tampere University of Technology, Institute of Signal Processing
P.O. Box 553, FIN-33101 Tampere, Finland
{Iivari.Kunttu, Leena.Lepisto, Ari.Visa}@tut.fi
http://sigwww.cs.tut.fi/
[2] ABB Oy, Paper, Printing, Metals & Minerals, Automation,
P. O. Box 94, FIN-00381 Helsinki, Finland
Juhani.Rauhamaa@fi.abb.com
http://www.abb.com/

Abstract. The use of the second order statistical measures has became popular in the image database indexing and retrieval. Unlike the common approach, image histogram, second order statistics like image correlogram and autocorrelogram consider also the spatial organization of the image colors or gray levels. Recently, correlograms and autocorrelograms have been widely used in the image database indexing. In this paper we present binary co-occurrence matrix, a new statistical measure for image indexing. This measure represents the footprint distribution of the co-occurrence matrix. Compared to image correlogram, this approach provides better retrieval accuracy at lower computational cost. We make retrieval experiments using two industrial image databases. These databases contain images collected from paper and metal manufacturing processes. In the experiments, we compare the retrieval performance of our approach to that of correlograms and autocorrelograms.

1 Introduction

In addition to texture and shape, the distribution of image colors (or gray levels) is an essential feature in content-based image retrieval. Image histogram is a first order statistical measure that has been traditionally used in characterization global color distribution of the image. The benefit of the image histogram is its low computational cost. However, histogram describes only the global distribution of the colors ignoring their spatial organization. This drawback has a remarkable effect on the image retrieval accuracy.

A simple improvement to the color-based image retrieval is the use of second order statistics. The second order statistical measures utilize the spatial organization between the pixel pairs occurring in the image. Correlation-based methods have been used in texture analysis since 1950's. Kaizer [6] was the first who used autocorrelation function to measure texture coarseness. Co-occurrence matrix introduced by Haralick [4] is a correlation-based tool for texture analysis. Correlation function has

J. Bigun and T. Gustavsson (Eds.): SCIA 2003, LNCS 2749, pp. 1090–1097, 2003.

been used also in the field of image retrieval. Huang et al. [5] introduced color correlogram, a measure that describes the spatial correlation of image colors as a function of their spatial distance *d*. In fact, the principle of correlogram is the equal to co-occurrence matrix. The difference between these measures is that whereas co-occurrence matrix uses a single distance *d*, correlogram is calculated for a set of different distances. Because of its computational lightness, Huang et al. preferred auto-correlogram to correlogram in image indexing. Autocorrelogram is a subset of correlogram. It defines the probability of finding identical colors at distance *d*. In [5] retrieval experiments showed that autocorrelogram gives significantly better retrieval results than image histogram. For computational reasons, also Ojala et al. [9] chose autocorrelogram instead of correlogram in image indexing. However, the information carried by correlogram covers the image color content significantly better than auto-correlogram. In [8], we showed retrieval results achieved using correlogram were remarkably better than in the case of autocorrelogram.

In this paper we apply binary co-occurrence matrix in image database indexing. The method was introduced in [7] as a tool for image retrieval without segmentation. The binary co-occurrence matrix is a simple and effective second order statistic for image database indexing. We compare the binary matrix to correlogram-based image retrieval tools. In section two, we present the principles of correlation-based statistical tools and the binary co-occurrence matrix. The retrieval ability of these methods is measured in section three.

The number of digital imaging and image databases in industry has strongly increased during recent years. For example, in process industry, digital imaging solutions are used to control the process and quality. In many cases, these solutions store the image data in image databases. These industrial image databases containing real image data are a challenging retrieval task. In this paper, we use two industrial image databases for testing purposes. The first of these databases is collected from the paper manufacturing process, and it contains 1308 paper defect images. The second testing database is from metal industry. In this database there are 1955 images of defects occurring in the metal surface.

2 Image Database Indexing Using Second Order Statistics

Statistical methods for the analysis of image gray levels or colors are commonly used tools for characterization of the image content. First order statistical methods, like histogram, consider image pixels separately ignoring their spatial relationships. Second- and higher order measures estimate the relationships between two or more pixel values occurring at specific locations relative to each other. In this section, we consider second order statistical measures for image retrieval. In addition to the commonly used methods, we present our approach, binary co-occurrence matrix.

2.1 Statistical Tools for Image Retrieval

Second order statistical measures have traditionally been used in texture analysis. In addition, correlograms and autocorrelograms have also been used in image retrieval. In this part, we present commonly used second order statistical measures.

Image correlogram represents the correlations between the image pixel values. The definition of image correlogram is the following [5] [9]. Let I be an XxY image which comprises of pixels $p(x,y)$. Each pixel has a certain color- or gray level (henceforth level). Let $[G]$ be a set of G levels g_1 ... g_G that can occur in the image. For a pixel p, let $I(p)$ denote its level g, and let I_g correspond to a pixel p, for which $I(p)=g$. Let $[D]$ denote a set of fixed distances d_1 ... d_D. Hence, the number of the distances in this set is D. The correlogram of the image I is defined for level pair (g_i, g_j) at a distance d:

$$\gamma_{g_i,g_j}^{(d)}(I) \equiv \Pr_{p_1 \in I_{g_i}, p_2 \in I} \left\lfloor p_2 \in I_{g_j} \middle\| p_1 - p_2 \right| = d \right\rfloor \tag{1}$$

which gives the probability that given any pixel p_1 of level g_i, a pixel p_2 at a distance d from the given pixel p_1 is of level g_i. In other words, the correlogram is a matrix that gives the probability of certain level to occur at the distance d from each other. Correlogram is defined for several values of d defined in the set $[D]$. The size of the correlogram-based feature vector is G^2D.

Autocorrelogram [5], [9] is the subset of the correlogram. It captures only the spatial correlation of the identical levels. The autocorrelogram can be defined as:

$$\alpha_g^{(d)}(I) = \gamma_{g,g}^{(d)}(I) \tag{2}$$

and it gives the probability that a pixel p_2, d away from the given pixel p_1, is of level g. In case of the autocorrelogram, the size of the feature vector is GD.

Co-occurrence matrix introduced by Haralick et al. [4] is the basis of the statistical texture analysis. It is a matrix that express the probability of two pixels to occur at certain distance from each other. In fact, co-occurrence matrix is the same as image correlogram defined for a single distance d.

2.2 Binary Co-Occurrence Matrix

A new second order statistical feature to be used in the image retrieval is binary co-occurrence matrix [7]. It is formed by means of the co-occurrence matrix (or a correlogram calculated for a single distance d). In the binary form of the matrix, all the occurrences between the image pixel levels are considered equally. This is done by quantizing the matrix into two levels, “zero” and “non-zero” values. In this way, a binary matrix containing only zeros and ones is formed. The size of the binary co-occurrence matrix is hence G^2.

Table 1. The computational cost based on the length of the feature vectors

FEATURE	FEATURE VECTOR LENGTH
BINARY CO-OCCURENCE MATRIX	G^2
32 gray levels	1024
16 gray levels	256
CORRELOGRAM	G^2D
32 gray levels	4096
16 gray levels	1024
AUTOCORRELOGRAM	GD
32 gray levels	128
16 gray levels	64

Binary co-occurrence matrix has two benefits that make it effective in the image retrieval. First, it is computationally light method compared to the image correlogram (that in our experiments in [8] proved to be the more powerful statistic in image retrieval than image autocorrelogram and histogram). However, the retrieval results of binary co-occurrence matrix are at the same level or better than correlogram. The second reason for the use of binary co-occurrence matrix is the fact that it considers all the correlations occurring in the image equally, which means that in many cases image segmentation can be avoided [7].

2.3 Statistical Measures in Image Database Indexing

Computational cost is an essential property of the indexing methods used in image retrieval. The computational cost of each method is proportional to the length of the feature vector, and therefore short vectors are preferred. The lengths of the feature vectors used in this paper are presented in table 1. In this table, the number of distances (D) is selected to be 4, as in [5] and [9].

Computational cost has been a reason in [5] and [9] for the use of the image autocorrelograms instead of correlograms in the image database indexing. However, in the description of image content, tools based on the whole probability distribution of the image (like correlogram and co-occurrence matrix) are clearly better. In [8] we solved the problem of the computational cost by dividing the database images in the areas of similar color (or gray level). This method is near the principle of color sets presented in [11]. In our approach, this division was made by re-quantizing the color space of the images. This way the number of the image levels G was decreased. Also the quantization of the image generalizes the image content and yield to better retrieval results [8]. Therefore, this quantization is used also in the experiments presented in this paper.

In image retrieval the similarity between the query image **Q** and the database image **I** is measured by distance metrics. In [5] and [9], the distance measure between the autocorrelograms is L_1 norm [2]:

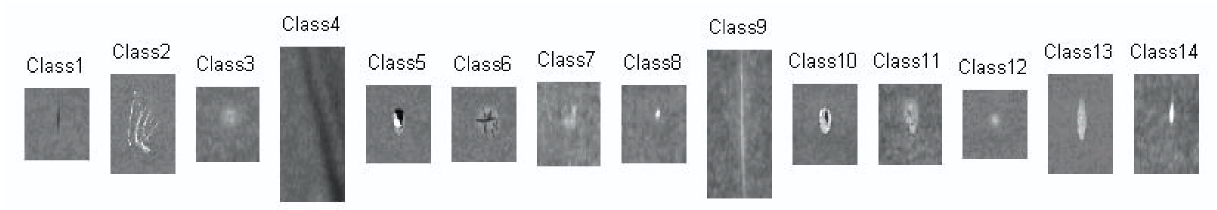

Fig. 1. An example of each paper defect image class in testing database I

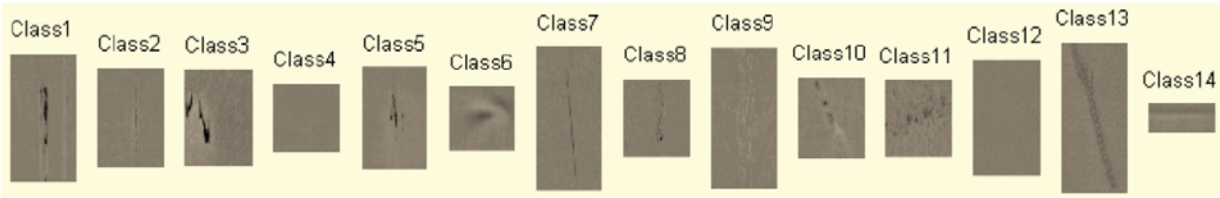

Fig. 2. An example of each metal defect image class in testing database II

$$L_1 = \sum_{i=1}^{N} \left| \alpha_{g\mathbf{Q}}^{(d)}(i) - \alpha_{g\mathbf{I}}^{(d)}(i) \right| \tag{3}$$

We use the same distance measure also in case of the correlograms. In case of binary co-occurrence matrices, binary distances are required. When comparing two binary matrices, $\mathbf{B_1}$ and $\mathbf{B_2}$, let $n_{1,1}$ denote the number of the elements, whose value is 1 in both matrices. In a similar way, $n_{1,0}$, $n_{0,1}$ and $n_{0,0}$ denote numbers of matrix elements, which have values 1 and 0, 0 and 1, 0 and 0, respectively. Jaccard coefficient [3] is a popular similarity measure for binary data. This coefficient is defined as:

$$S_J = \frac{n_{1,1}}{n_{1,1} + n_{1,0} + n_{0,1}} \tag{4}$$

3 Experiments

In this section we test our approach to the correlogram-based image retrieval using real industrial image databases. For testing purposes we had two sets of defect images. In testing database I, the defects were collected from the paper web using a paper inspection system [10]. The objects in the images were typical paper surface defects. The test set consisted of 1308 paper defects, which represented 14 defect classes so that each class consisted of 32-100 defect images. An example image of each paper defect image class is presented in figure 1. The second test set, testing database II, there were 1955 metal surface defect images. Also in this case, there were 14 defect classes (figure 2). Each class contained 100-150 metal defect images. In both databases, the images were intensity images containing 256 gray levels. The image size had also strong variations (dimension of the image varied from 100 to 2000 pixels).

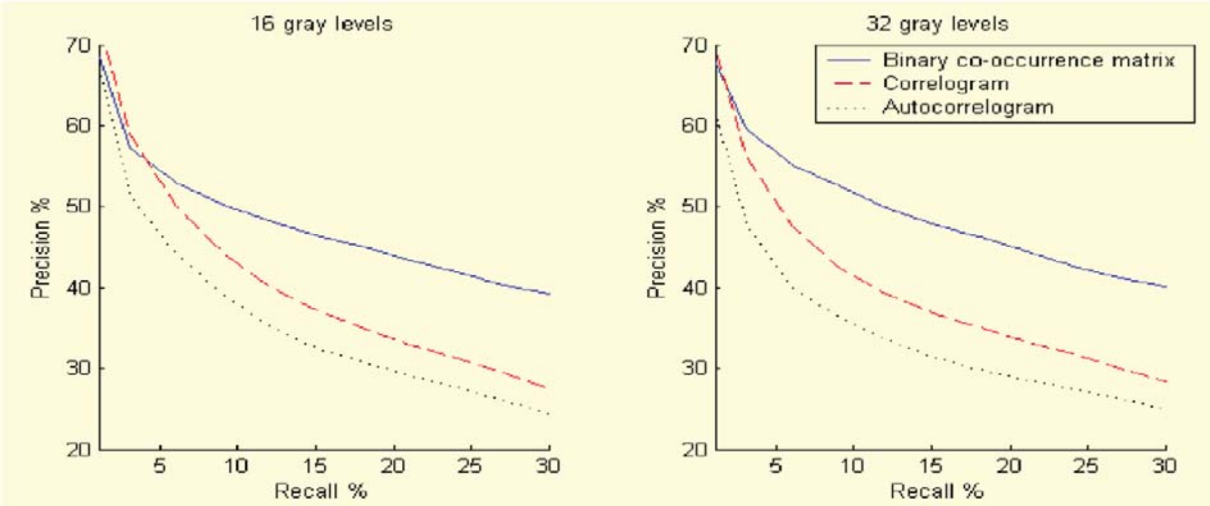

Fig. 3. Average retrieval performance of the features in case of paper defect images in testing database I

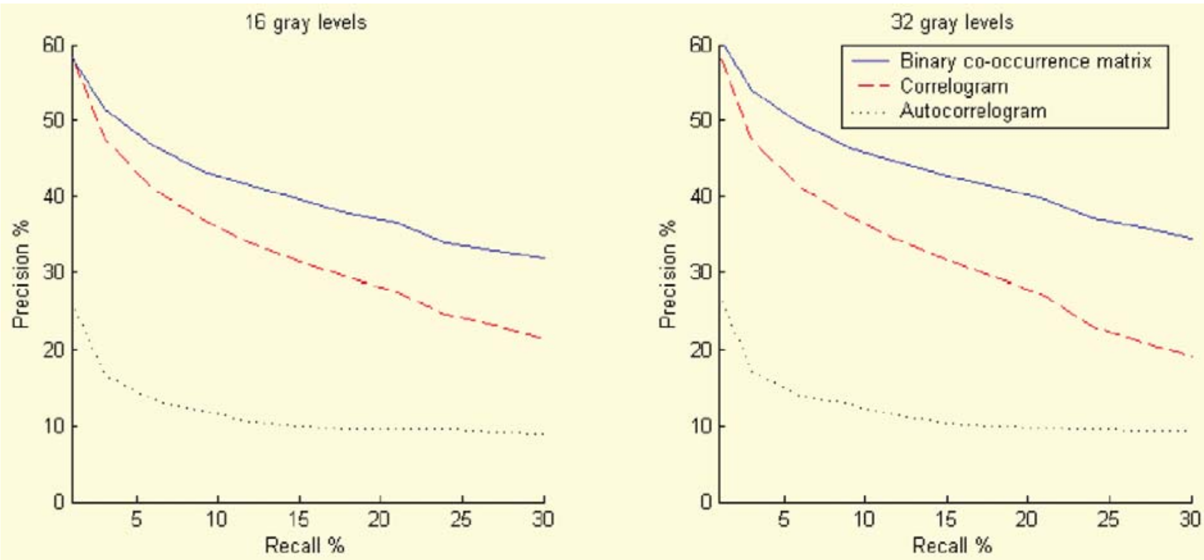

Fig. 4. Average retrieval performance of the features in case of metal defect images in testing database II

We calculated the correlograms, autocorrelograms as well as the binary co-occurrence matrices for the database images quantized to 32 and 16 gray levels. In the calculation of the autocorrelogram and the correlogram we used the set of distances [D]={1,3,5,7}, which is the same as in [5] and [9]. Both databases were indexed using these features. The purpose of the retrieval experiments was to test the retrieval ability of each feature. The retrieval experiments were made using *leaving one out* method [3]. In this method, each image in turn is left out from the test set and used as a query image, whereas the other images in the test set form a testing database. In the queries, the nearest images to the query image are retrieved based on their feature vectors. The performance of the retrieval was measured by calculating a *precision* *versus recall curve* [1] for each query. If $|A|$ is the number of all retrieved images, $|R|$ is the number of query class images in the whole testing database and $|Ra|$ is the number of retrieved query class images, precision and recall can be defined in the following way [1]:

$$\text{Precision} = \frac{|Ra|}{|A|} \tag{5}$$

$$\text{Recall} = \frac{|Ra|}{|R|} \tag{6}$$

The retrieval performance of each feature can be presented by calculating the average precision-recall curve for each query.

In the retrieval experiments, the similarity measure for the binary co-occurrence matrices was the Jaccard coefficient. For the correlograms and the autocorrelograms, L_1 norm was used. Figures 3 and 4 present the average precision-recall curves for the images in both testing databases.

4 Discussion

In this paper we presented a new approach to the statistical image retrieval. Our method, binary co-occurrence matrix is a simple tool for the characterization of the gray level distributions of the database images. The experimental results presented in figures 3 and 4 show that the binary co-occurrence matrix is an effective method in image retrieval. Compared to the correlogram and autocorrelogram, our method gives clearly better results in retrieval accuracy. Binary co-occurrence matrix is also computationally lighter method than image correlogram.

For testing purposes, we had two industrial image databases. The testing databases contained 1308 paper defect images and 1955 images of metal defects. These kinds of industrial databases provide a good opportunity to test the retrieval methods with real image data. On the other hand, retrieval of the defect images is a quite demanding task. This is because some defect classes are very similar to each other and also overlapping. In these classes, the retrieval results could be improved by using some shape descriptors together with the statistical tools.

In this work, the testing databases contained only gray level images. Our method, binary co-occurrence matrix could be applied also to the classification and retrieval of color images. In that case the color quantisation can be made using tools presented in [11]. This could be a subject of further studies in the field of statistical image retrieval methods.

Acknowledgment

The authors wish to thank the Technology Development Centre of Finland (TEKES's grant 40397/01) for financial support.

References

1. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press, Addison Wesley, New York (1999)
2. Duda, R.O., Hart, P.E., Stork, .G.: Pattern Classification. 2nd ed. John Wiley (2001)
3. Hand, D., Mannila, H., Smyth, P.: Principles of Data Mining. MIT Press, Massachusetts, USA (2001)
4. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-3, 6 (1973)
5. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.-J., Zabih, R.: Image indexing Using Color Correlograms. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Juan, Puerto Rico (1997) 762-768
6. Kaizer, H.: A Quantification of Textures on Aerial Photographs. Tech. Note No. 121, A.69484. Boston University Research Laboratories, Boston University (1955)
7. Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A.: Image Retrieval without Segmentation. Proceedings of 10th Finnish AI Conference. Oulu, Finland (2002) 164-169
8. Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A.: Image Correlogram in Image Database Indexing and Retrieval. Proceedings of 4th European Workshop on Image Analysis for Multimedia Active Services, London, UK, (2003) 88-91
9. Ojala, T., Rautiainen, M., Matinmikko, E., Aittola, M.: Semantic Image Retrieval with HSV Correlograms. Proceedings of 12th Scandinavian Conference on Image Analysis. Bergen, Norway (2001) 621-637
10. Rauhamaa, J., Reinius, R.: Paper Web Imaging with Advanced Defect Classification. TAPPI Technology Summit, Atlanta, Georgia (2002)
11. Smith, J.R., Chang, S.F.: Tools and Techniques for Color Image Retrieval. Storage and Retrieval for Image and Video Databases IV, SPIE Proceedings, Vol. 2670 (1996) 1630-1639

Publication VII

Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A., 2005. Color Fourier Descriptor for Defect Image Retrieval.

Color Fourier Descriptor for Defect Image Retrieval

Iivari Kunttu[1], Leena Lepistö[1], Juhani Rauhamaa[2], and Ari Visa[1]

[1] Tampere University of Technology, Institute of Signal Processing
P.O. Box 553, FI-33101 Tampere, Finland
{Iivari.Kunttu, Leena.Lepisto,Ari.Visa}@tut.fi
http://www.tut.fi
[2] ABB Oy, Process Industry P.O. Box 94, FI-00381 Helsinki, Finland
Juhani.Rauhamaa@fi.abb.com
http://www.abb.com

Abstract. The shapes of the objects in the images are important in the content-based image retrieval systems. In the contour-based shape description, Fourier descriptors have been proved to be effective and efficient methods. However, in addition to contour shape, Fourier description can be used to characterize also the color of the object. In this paper, we introduce new Color Fourier descriptors. In these descriptors, the boundary information is combined with the color of the object. The results obtained from the retrieval experiments show that by combining the color information with the boundary shape of the object, the retrieval accuracy can be clearly improved. This can be done without increasing the dimensionality of the descriptor.

1 Introduction

Nowadays, the problem of image retrieval plays a remarkable role in the fields of image analysis and pattern recognition. With increasing amount of real-world image data to be processed and stored, the development of powerful retrieval tools has become an essential subject of research work. The description of the objects occurring in the images is based on visual features extracted from them. In addition to color and texture, shape is one of the most important features used to characterize the objects occurring in the images as accurately as possible. These features are widely used in content-based image retrieval systems [1],[8].

On the other hand, classification accuracy (effectiveness) of a certain descriptor is not an adequate measure for its usefulness in the retrieval. Due to the increasing number of online retrieval solutions, computational efficiency is nowadays considered equally important as effectiveness [10]. In retrieval applications, the matter of computational complexity is twofold, namely the cost of image database indexing and retrieval. In the indexing, the features (descriptors) are extracted from the database images. Although this part is not always online operation, the feature extraction should not be a computationally heavy. More importantly, retrieval is always performed in real time. Therefore, the descriptors used in retrieval are required to be compact. The compactness of a de-

F. Roli and S. Vitulano (Eds.): ICIAP 2005, LNCS 3617, pp. 415 – 422, 2005.

scriptor depends on its dimensionality, because the retrieval time is proportional to the descriptor dimensions. Consequently, low-dimensional descriptors are preferred.

In this paper, we concentrate on object description that is based on Fourier transform. Fourier-based methods are widely used in shape description [6]. Fourier descriptors have been found to be accurate in shape classification in several comparisons, [2],[3],[4],[9]. In addition to good retrieval and classification accuracy, there are also other reasons which make Fourier descriptors popular among the contour-based shape representations. The main advantages of the Fourier-based shape representations are that they are compact and computationally light methods. Furthermore, they are easy to normalize and their matching is very simple. Also the sensitivity to noise is low, when only low frequency Fourier coefficients are used as descriptors.

In addition to object shape, its color is often equally important feature. In retrieval systems, colors are usually characterized using relatively high-dimensional descriptors, like histograms [1] or other statistical measures. On the other hand, the number of descriptors that efficiently combine color and shape is very small. In the work of Mehtre et al., [5] color and shape of the object were combined. This approach, however, was based on quite complicated clustering method. Furthermore, the approach used moment-based shape features that are computationally more expensive than for example Fourier descriptors.

In this paper, we present a new approach to the use of Fourier descriptors in the characterization of image content. Hence, we show that the Fourier descriptor is capable of describing also other features of the object than its contour. In our approach, we add the object color to the Fourier-based contour description. In this way, the obtained descriptor is able to more accurate object description in the retrieval process. However, the color information does not increase the dimensionality of the obtained descriptor.

The organization of this paper is the following. In section two, the principles of Fourier descriptors and our approach, *Color Area Fourier*, are presented. Section three reports the retrieval experiments carried out using real industrial defect images. The proposed method is discussed in section four.

2 Object Description

In this section, the common methods for shape description using Fourier-based methods are presented. In addition to that, our approach to combine object color information with its contour in Fourier description is presented.

2.1 Fourier Descriptors

Shape signatures. Shape signature is a 1D boundary function $f(k)$ that represents the boundary of a 2D object. The functions are either real-valued or complex. Complex coordinate function [2] is the simplest and best-known boundary presentation. It presents the coordinates of the boundary (x_k, y_k) in an object centered complex coordinate system:

$$z(k) = (x_k - x_c) + j(y_k - y_c) \tag{1}$$

for k=0,1,2,…,N-1, in which N is the length of the boundary and (x_c, y_c) is the centroid of the object. Area function [11] is an example of real-valued shape signatures and it is defined as the area of the triangle formed by two boundary points and centroid in the object centered coordinate system:

$$a(k) = \frac{\left|(x_k - x_c)(y_{k+1} - y_c) - (x_{k+1} - x_c)(y_k - y_c)\right|}{2} \tag{2}$$

Hence, both signatures represent the boundary independent of the location of the object in the image.

Signatures for color and shape. The object signatures of the proposed descriptors use the same basic approach as the complex-valued shape signatures of equation (1). Hence, by combining two real-valued 1D signals it is possible to form a 1D complex signal. In this paper, we combine the object color to its boundary information. This is made by combining the color of the object region defined by the shape signature with the signature itself. The color value C_k can be e.g. the mean of the selected color component at each object region k. In the case of area function, the region corresponds to the image pixels covered by the area of the triangle defined by equation (2). The signature for color and shape of an object is expressed as complex numbers:

$$c_a(k) = a_k + jC_k \tag{3}$$

Hence the signature of equation (3) combines the real-valued boundary information with object color distribution.

Fourier description. The descriptor based on a signature function can be formed in several ways. Fourier transform is a commonly used method for this purpose. Fourier transformation of a boundary function generates a set of complex numbers which are called Fourier descriptors. Fourier descriptors characterize the object shape in a frequency domain. The discrete Fourier transform for a boundary function $f(k)$ is:

$$F_n = \frac{1}{N}\sum_{k=0}^{N-1} f(k) e^{-j2\pi nk/N} \tag{4}$$

for n=0,1,2,…,N-1. The general shape of the object is represented by the lower frequency descriptors, whereas high frequency descriptors represent the fine details of the object shape. The descriptors have to be normalized to achieve invariance to translation, rotation, and scaling. Translation invariance of is based on the object centered shape signatures. The descriptors can be made rotation invariant by ignoring the phase information and using only the magnitudes of the transform coefficients $|F_n|$. The scale can be normalized by dividing the magnitudes by $|F_0|$ or $|F_1|$, depending on the shape representation method [2].

Feature vectors. A common approach to object description is to use only a subset of low frequency coefficients that carry the most of the object information. This way the shape can be effectively presented using a relatively short feature vector. For complex-valued shape signatures, the coefficients are taken from positive and negative frequency axis:

$$x = \left[\frac{|F_{-(L/2-1)}|}{|F_1|}, \dots, \frac{|F_{-1}|}{|F_1|}, \frac{|F_2|}{|F_1|}, \dots, \frac{|F_{L/2}|}{|F_1|} \right]^T \tag{5}$$

in which L is a constant value that defines the dimensionality of the feature vector. When this description is formed for the transform coefficients obtained from complex coordinate function of equation (1), it is called *Contour Fourier* method [2]. In this paper, this kind of feature vector is applied also to *Color Area Fourier* descriptor that uses complex-valued signature of equation (3). However, in the case of *Color Area Fourier* descriptor, the normalization is carried out using $|F_0|$ instead of $|F_1|$:

$$x = \left[\frac{|F_{-(L/2)}|}{|F_0|}, \dots, \frac{|F_{-1}|}{|F_0|}, \frac{|F_1|}{|F_0|}, \dots, \frac{|F_{L/2}|}{|F_0|} \right]^T \tag{6}$$

The difference between the feature vectors can be explained by differences between the signatures. In the case of *Contour Fourier* method of equation (5), the signature uses merely boundary information that is represented in location-independent manner. Therefore, scale is normalized using the first non-zero coefficient, $|F_1|$. On the other hand, *Color Area Fourier* descriptor uses complex-valued signature of equation (3) in which contour shape is represented by centroid distance. Therefore, the mean value of the signature function differs from zero. This causes the normalization by $|F_0|$, which is the transform coefficient representing the mean value of the signal.

The real-valued shape representation, *Area Fourier* [2] uses the area function as shape signature. Because this signature is real, only half of the transform coefficients are needed to characterize the shape [2]:

$$x = \left[\frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \dots, \frac{|F_L|}{|F_0|} \right]^T \tag{7}$$

Also with this descriptor type, the normalization is carried out using the mean component, $|F_0|$, to remove the effect of the mean of the area function.

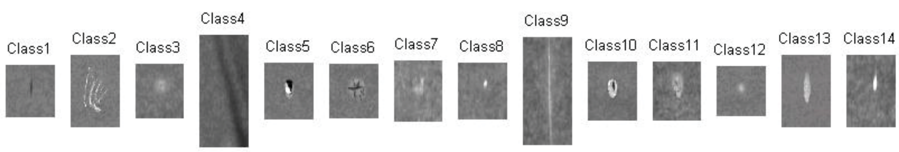

Fig. 1. An example image of each paper defect class in the testing database

3 Retrieval Experiments

In this section, we present the retrieval experiments carried out using a real defect image database. We compare the retrieval performance of the proposed *Color Area Fourier* approach to the ordinary Fourier shape description methods, which describe only the boundary line of the object.

3.1 Defect Image Database

For testing purposes, we used paper defect images that were collected from an industrial process. The images were taken from the paper manufacturing process using a paper inspection system [7] that produces gray level images of the defects. The reason for the collection of the defect image databases in the process industry is the practical need of controlling the quality and production [7]. In industrial imaging solutions, there is a need to retrieve the defect images from the databases. In these images, the defect shape and gray level are the most essential features that describe the defect type. Therefore, effective methods for the shape and gray level representation are needed in the retrieval and classification of the defect images. The defects occurring in the paper can be for example holes, wrinkles or different kinds of dirt spots. The test set consisted of 1204 paper defects, which represented 14 defect classes so that each class consisted of 27-103 images. An example image of each paper defect class is presented in figure 2. Within the classes of the defect database, there were differences in the gray levels, size, and orientation of the defects.

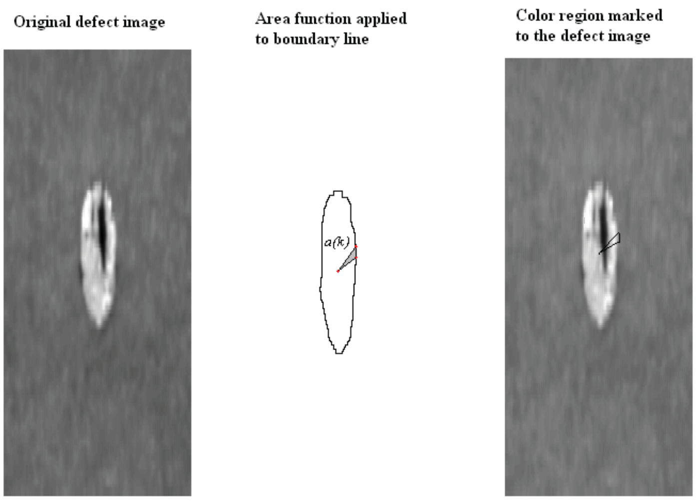

Fig. 2. The principle of *Color Area Fourier* descriptor applied to a paper defect image

3.2 Retrieval

In this paper, we compare our *Color Area Fourier* approach to *Area Fourier* method and *Contour Fourier* method that have been proved to be effective Fourier-based shape descriptors in retrieval of defect shapes [3].

The Fourier-based shape descriptors (*Area Fourier* and *Contour Fourier*) were calculated based on the defect contours. The descriptors used triangular color regions that were defined from the defect images so that the triangle was drawn between object centroid and two consequent boundary points (figure 2). The feature vectors of the descriptors were formed using equations (7) and (5), respectively. In the case of *Color Area Fourier* descriptor, the color information was added to the area-based shape description. The defects are presented as gray level images, which means that only the intensity component was used to represent the color. The color value C_k was selected to be the mean gray level of the triangular region. The descriptors were formed using feature vector of equation (6).

In the comparison, low-dimensional descriptors were preferred, and hence we used two lengths of the vectors (L), namely 8 and 16. In the retrieval experiments, the distance measure between the feature vectors was selected to be the Euclidean distance. The retrieval experiments were made using *leaving one out* method. In this method, each image in turn is left out from the test set and used as a query image; whereas the other images in the test set form a testing database. The performance of the retrieval was measured by calculating a *precision versus recall curve* for each query.

The average precision/recall curves for the database are presented in figure 3. The results show that the Fourier descriptors combined with object color (*Color Area Fourier*) outperform clearly the *Area Fourier* descriptor. On the other hand, the

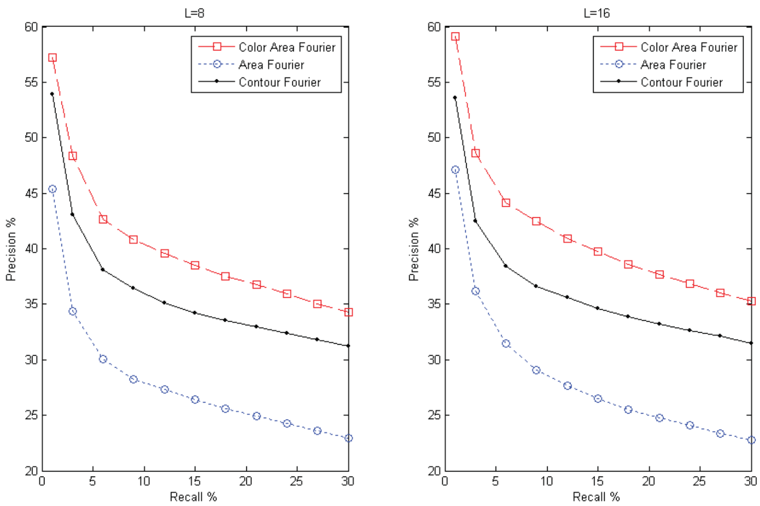

Fig. 3. The average precision/recall curves of the retrieval experiments calculated for each descriptor type

proposed descriptors outperform also *Contour Fourier* method that is the most accurate shape-based Fourier descriptor in defect image retrieval [3]. It is essential to note that this improvement does not increase dimensionality of the feature vectors, when *Color Area Fourier* and *Contour Fourier* descriptors are compared. Furthermore, when the same distance metrics is applied, the computational cost of retrieval is equal with the proposed descriptors and the conventional Fourier shape descriptors.

4 Discussion

In this paper, a new approach to Fourier-based object presentation was introduced. Our approach, *Color Area Fourier* descriptor, combines the color and shape information of an object into a single feature vector. The obtained vector is as low dimensional and easy to match as any other shape-based Fourier descriptor. However, our experiments showed that *Color Area Fourier* descriptor outperforms the other Fourier-based shape descriptors in terms of retrieval accuracy.

In the experiments, a database of complex shapes was used. The shapes in the database represent defects that occur in an industrial process. Self-evidently, the accuracy of the descriptors is the most important criterion also in this retrieval problem. On the other hand, the matter of computational efficiency is essential in this case, like in the most of the real-world image retrieval tasks. Therefore, compact features are required. The experiments showed that combining the color information to a Fourier descriptor, additional retrieval accuracy can be achieved without increasing computational cost of retrieval. Therefore, the *Color Area Fourier* descriptor presented in this paper is an effective and efficient tool for describing complex objects in image retrieval and classification.

References

1. Del Bimbo, A.: Visual Information Retrieval, Morgan Kaufmann Publishers, San Fransisco, (2001)
2. Kauppinen, H., Seppänen, T., Pietikäinen, M.: An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 2 (1995) 201-207
3. Kunttu, I., Lepistö, L., Rauhamaa, J., Visa, A.: Multiscale Fourier Descriptor for Shape-Based Image Retrieval, Proceedings of 17th International Conference on Pattern Recognition, Cambridge, UK, Vol. 2 (2004) 765-768
4. Mehtre, B.M., Kankanhalli, M.S., Lee, W.F.: Shape Measures for Content Based Image Retrieval: A Comparison, Information Processing Management, Vol. 33, No 3, (1997) 319-337
5. Mehtre, B.M., Kankanhalli, M.S., Lee, W.F.: Content-Based Image Retrieval Using a Composite Color-Shape Approach. Information Processing & Management Vol. 34, No. 1, (1998) 109-120.
6. Persoon, E., Fu, K.: Shape Discrimination Using Fourier Descriptors, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 7 (1977) 170-179

7. Rauhamaa, J., Reinius, R.: Paper Web Imaging with Advanced Defect Classification, Proceedings of the 2002 TAPPI Technology Summit, Atlanta, Georgia (2002)
8. Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-Based image Retrieval at the End of the Early Years. IEEE Transactions Pattern Analysis and Machine Intelligence Vol. 22, No. 12 (2000) 1349-1380.
9. Zhang, D., Lu, G.: A Comparative Study of Curvature Scale Space and Fourier Descriptors for Shape-Based image Retrieval, Journal of Visual Communication and Image Representation, Vol. 14, No. 1 (2003) 41-60
10. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition, Vol. 37, No. 1, (2004) 1-19
11. Zhang, D.S., Lu, G.: Study and evaluation of different Fourier methods for image retrieval. Image and Vision Computing Vol. 23, (2005) 33-49

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O. Box 527
FIN-33101 Tampere, Finland