



This is to certify that the
thesis entitled
SHAPE-BASED IMAGE RETRIEVAL

presented by
Aditya Vailaya

has been accepted towards fulfillment
of the requirements for
M.S. degree in Computer Science

A handwritten signature in black ink, appearing to read "Paul Kummer Jr.", written over a horizontal line.

Major professor

Date April 23, 1996



PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

MSU is An Affirmative Action/Equal Opportunity Institution

c:\crl\datedue.pms-p.1

SHAPE-BASED IMAGE RETRIEVAL

By

Aditya Vailaya

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Computer Science

1996

ABSTRACT

SHAPE-BASED IMAGE RETRIEVAL

By

Aditya Vailaya

Retrieval efficiency and accuracy are two important issues in designing a content-based database retrieval system. We propose a new image database retrieval method based on shape information. This system achieves both the desired efficiency and accuracy using a two-stage hierarchy: in the first stage, simple and easily computable statistical shape features are used to quickly browse through the database to generate a moderate number of plausible retrievals; in the second stage, the outputs from the first stage are screened using a deformable template matching process to discard spurious matches. We have tested the algorithm using hand drawn queries on a trademark database containing 1,100 images. Each retrieval takes a reasonable amount of computation time (~ 45 seconds). The top most retrieved image by the system agrees with that obtained by human subjects, but there are significant differences between the top 10 retrieved images by our system and that provided by human subjects. This demonstrates the need for developing shape features that are better able to capture human perceptual similarity of shapes. An improved heuristic

has been suggested for more accurate retrievals. The system matches filled-in query images against filled-in images from the database thus using only the gross details in the image. Experiments have shown that matching on the filled-in database extracts more images that have similar content. The proposed scheme has also been tested on a database of master-marks of Norwegian silversmiths containing 171 images. The experiments show the effectiveness of the integration of shape features for matching faded, oriented, and scaled master-marks extracted from the bottom of the silver tankards. We believe that it is a challenge to develop an automatic retrieval algorithm which models the human perception well. However, considering the time and effort for humans to perform such a retrieval task, an automatic retrieval algorithm can provide a powerful tool to facilitate the retrieval task for human beings.

To My Family

ACKNOWLEDGMENTS

I would like to acknowledge all the people who have assisted me throughout my studies at Michigan State University. I am extremely grateful to my advisor, Professor Anil Jain, for his continuous guidance and encouragement, and the valuable time that he spent on this project with me. I owe my great interest in research to him.

I am extremely grateful to Dr. Ming-Yee Chiu and the technical staff at Siemens Corporate Research, Princeton for their guidance and encouragement during my stay there in 1995 summer and the use of their database for this research project.

I would also like to acknowledge Dr. Britt Kroepelien for providing a strong basis for the project on authentication of Norwegian silver tankards. It was with her patience and determination that we were able to start this project in such a short time. I would like to specially acknowledge Chitra Dorai for being my mentor in the initial period of my stay here and for patiently assisting me to adjust to the new environments.

I would like to dedicate this thesis to my family who have constantly encouraged me to achieve new heights. Without their love, patience, understanding, and support, I would not have finished this thesis.

I would like to thank Yu Zhong for her assistance in using deformable template matching. I would also like to acknowledge the help and support I have received from members of the PRIP lab and from my friends at Michigan State University.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
1 Introduction	1
1.1 Image Database	3
1.2 Information on Trademarks	6
1.2.1 Registration of Trademarks	7
1.2.2 Search for Conflicting Marks	7
1.3 Problem Definition	14
1.4 Motivation	16
1.4.1 Database	16
1.4.2 Similarity	17
1.5 Thesis Outline	18
2 Literature Review	22
2.1 Methods using a Single Cue	23
2.1.1 Color	23
2.1.2 Shape	24
2.1.3 Texture	26
2.2 Using Multiple Cues	27
2.2.1 QBIC	28
2.2.2 Photobook	30
2.2.3 STAR	32
2.2.4 Incorporating Learning	33
2.3 Discussion	34
3 An Integrated System for Efficient Retrieval	36
3.1 Image Attributes	37
3.2 Fast Pruning Stage	38
3.2.1 Edge Directions	38
3.2.2 Invariant Moments	42
3.2.3 Other Shape Features	43
3.2.4 Integration of Image Attributes	44
3.3 Deformable Template Matching	47
3.3.1 Representation of the Prototype Template	48
3.3.2 Deformation Transformation	48
3.3.3 Dissimilarity Measure	49

3.4	Summary	51
4	Experimental Results	52
4.1	Results on Trademark Images	53
4.1.1	Experiments on Accuracy	53
4.1.2	Experiments on the Proposed System	57
4.1.3	Retrieval by Humans	65
4.1.4	Retrieval using Additional Heuristics	67
4.2	A Case Study in Norwegian Silver	
	Authentication	91
4.2.1	Background	93
4.2.2	Database Search for Master-marks and City-marks	95
4.2.3	Preprocessing	97
4.2.4	Experimental Results	98
4.3	Discussion	99
4.3.1	Trademark Images	100
4.3.2	Norwegian Silver	101
5	Conclusions	109
5.1	Summary	109
5.2	Future Research	111
	APPENDICES	112
A	A Representative Subset of Trademark Images	112

LIST OF TABLES

4.1	Edge directions-based retrieval results for the 1,100 database images. . .	55
4.2	Invariant moments-based retrieval results for the 1,100 database images.	55
4.3	Integrated shape-based retrieval results for the 1,100 database images; n refers to the position of the correct retrieval, R : rotated query, S : scaled query, N : query with a noisy image; The last column indicates percentage of time the query image was not retrieved in the top 20 matches.	55
4.4	Dissimilarity values for the five query images when the deformable tem- plate matching is applied to the top 10 retrieved images from the fast pruning stage.	65
4.5	Recall and precision rates for three queries (shown in figures 4.14, 4.16, and 4.18) based on the image filling; n_1 refers to the number of images retrieved in the top 20 positions that are <i>similar</i> to the query, n_2 refers to the number of images in the database that are <i>similar</i> to the query. Recall rate is (n_1/n_2) , precision rate is $(n_1/20)$	86

LIST OF FIGURES

1.1	Model of a multimedia system.	2
1.2	Traditional image retrieval model.	6
1.3	Trademarks considered for opposition by the USPTO; registration was accepted by the PTO. The trademarks have been enlarged to show the details.	9
1.4	Trademarks chosen for opposition by the USPTO; registration was refused by the PTO. The trademarks have been enlarged to show the details.	10
1.5	Trademarks chosen for opposition by the USPTO; registration was accepted by the PTO. The trademarks have been enlarged to show the details.	11
1.6	Trademarks chosen for opposition by the USPTO; registration was refused by the PTO. The trademarks have been enlarged to show the details.	12
1.7	Trademarks chosen for opposition by the USPTO; registration was refused by the PTO. The trademarks have been enlarged to show the details.	12
1.8	Example of a trademark with multiple design codes.	14
1.9	Proposed image retrieval model.	15
1.10	Two perceptually similar images of a bull head.	19
1.11	Two perceptually similar images of a bear.	20
1.12	Another image of a bear.	20
1.13	Images of wheat and corn that vary in their appearance.	21
3.1	Example of shape representation using edge directions; (a)-(c) show 3 database images, (d)-(f) corresponding edge images, and (g)-(i) corresponding edge direction histograms; $D_s(g, h) = 0.065$, $D_s(g, i) = 0.69$, $D_s(h, i) = 0.70$	41
3.2	Example of shape representation using invariant moments; (a)-(c) show 3 database images; $D_m(a, b) = 0.033$, $D_m(a, c) = 0.85$, $D_m(b, c) = 0.85$	43
3.3	Constructing the query template for deformable template matching. (a) A hand drawn query trademark, and (b) the query template obtained by calculating the edge map of the image in (a)	48
4.1	Database images not retrieved correctly in the presence of rotation.	56
4.2	Examples of hand drawn query trademarks.	58
4.3	Database pruning results for the hand drawn bear sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.	59
4.4	Database pruning results for the hand drawn bull sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.	60

4.5	Database pruning results for the hand drawn boomerang sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.	61
4.6	Database pruning results for the hand drawn deer sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.	62
4.7	Database pruning results for the hand drawn kangaroo sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.	63
4.8	Deformable template matching; (a) initial position of the bull template overlaid on the edge map of a bull logo using the generalized Hough transform, (b) final match for the bull logo.	64
4.9	Deformable template matching; (a) initial position of the boomerang template overlaid on the edge map of a boomerang logo using the generalized Hough transform, (b) final match for the boomerang logo. . . .	65
4.10	Deformable template matching; final match for a bull template and a boomerang image.	66
4.11	Nine topmost retrieved trademark images by the human subjects for the bull template. The number below each retrieved image is the number of human respondents (out of 5) who placed the image in the top 10 retrievals.	68
4.12	Five images of a bull present in the database.	69
4.13	Filled-in images for the five bull images shown in Figure 4.12.	70
4.14	Database pruning results for a bull image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	72
4.15	Database pruning results for a bull image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity. Results are based on unfilled images.	73
4.16	Database pruning results for a bear image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	74
4.17	Database pruning results for a bear image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity. Results are based on unfilled images.	75
4.18	Database pruning results for a kangaroo image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	76
4.19	Database pruning results for a kangaroo image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity. Results are based on unfilled images.	77
4.20	Database pruning results for an image consisting a set of vertical bars (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	78

4.21 Database pruning results for a bee image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	79
4.22 Database pruning results for a Sun image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	80
4.23 Database pruning results for a word mark (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	81
4.24 Database pruning results for a lion image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	82
4.25 Two images containing a square border; information contained in the holes (white) is neglected when filling in the database.	83
4.26 Database pruning results for an image with a square border (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	84
4.27 Database pruning results for an image with a circular border (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.	85
4.28 Segmentation of image into object entities; (a) a trademark image; (b) segmented objects.	88
4.29 Segmentation of image into object entities; (a) a trademark image; (b) segmented objects.	89
4.30 Segmentation of image into object entities; (a) a trademark image; (b) segmented objects.	90
4.31 A block diagram of Master identification based on master and city-marks.	96
4.32 A representative set of master-marks in the database.	97
4.33 A representative set of real master-marks extracted from the tankards. .	98
4.34 Preprocessing an input master-mark; (a) Greyscale master-mark; (b) Thresholded (binary) master-mark; (c) Binary master-mark after noise cleaning.	98
4.35 Retrieval of master seals; (a) Input Master Mark; (b) Top 10 retrievals based on the integrated match.	99
4.36 Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.	103
4.37 Retrieval of master seals; (a) Input Master Mark; (b) Top 10 retrievals based on the integrated match.	104
4.38 Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.	105
4.39 Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.	106
4.40 Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.	107

4.41 Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.	108
A.1 Examples of trademarks present in the database.	112

Chapter 1

Introduction

Information is inherently multimodal. Humans can efficiently and effectively process information simultaneously in multiple dimensions. These multiple media, that aid effective communication, can be characterized into speech, audio, image, video, and textual data.

Advances in computing and networking are generating a significant amount of interest in multimedia services and applications [1, 2, 3, 4]. Powerful processors, high-speed networking, high-capacity storage devices, improvements in compression algorithms, and advances in processing of audio, speech, image, and video signals are making multimedia systems technically and economically feasible.

Multimedia systems suggest a wide variety of potential applications such as interactive entertainment, video news distribution, video rental services, and digital multimedia libraries. These services aim to provide the user with on-demand multimedia services. Any system providing these services will have to address the issue of representation, indexing, retrieval, and manipulation of the multimedia data. Fig-

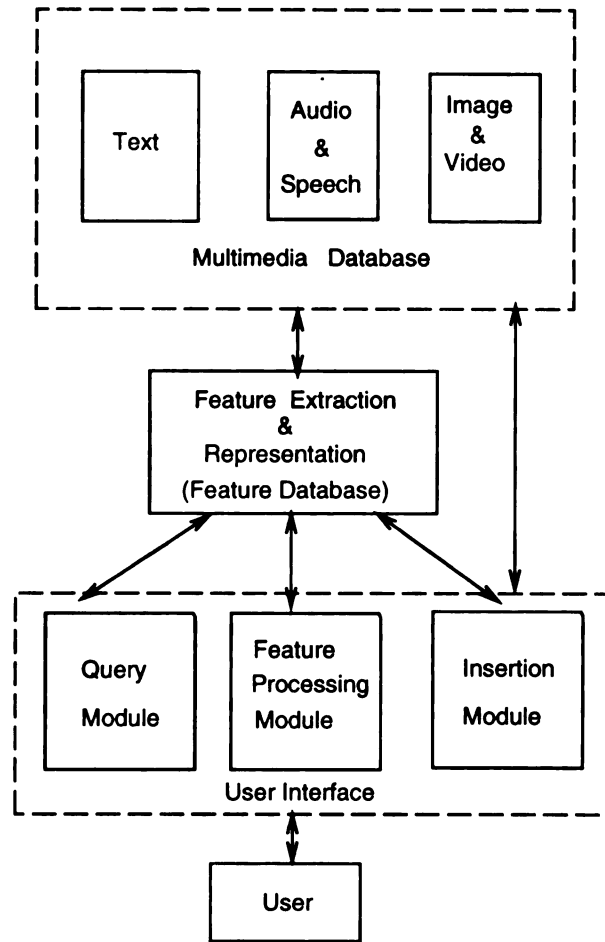


Figure 1.1: Model of a multimedia system.

Figure 1.1 represents a generic block diagram of a multimedia database. The multimedia objects are archived in a database. A feature extraction module extracts various features from the database, and represents the objects in terms of these features. The user interface provides the user with the capabilities to query, extract features, and insert multimedia objects.

Image and video are an integral part of multimedia data. There are a number of applications where images need to be automatically retrieved by their *content*. This necessitates the need for powerful image processing and understanding tools.

Various applications in digital libraries and image databases have been described in the literature. In the following section, we provide a brief introduction to image databases.

1.1 Image Database

Digital images are a convenient media for describing and storing spatial, temporal, spectral, and physical components of information contained in a variety of domains (e.g., aerial/satellite images in remote sensing, medical images in telemedicine, fingerprints in forensics, museum collections in art history, and registration of trademarks and logos). These databases typically consist of thousands of images, taking up gigabytes of memory space. While advances in image compression algorithms have alleviated the storage requirement to some extent, the large volume of these images makes it difficult for a user to browse through the entire database. Therefore, an efficient and automatic procedure is required for indexing and retrieving images from databases.

Traditionally, textual features, such as filenames, caption and keywords have been used to annotate and retrieve images. But, there are several problems with these methods. First of all, human intervention is required to describe and tag the contents of the images in terms of a selected set of captions and keywords. In most of the images there are several objects that could be referenced, each having its own set of attributes. Further, we need to express the spatial relationships among the various objects in an image to understand its content. As the size of the image databases

grow, the use of keywords becomes not only complex but also inadequate to represent the image content. The keywords are inherently subjective and not unique. Often, the preselected keywords in a given application are context dependent and do not allow for any unanticipated search. If the image database is to be shared globally then the linguistic barriers will make the use of keywords ineffective. Another problem with this approach is the inadequacy of uniform textual descriptions of such attributes as color, shape, texture, layout, and sketch. As an example, we can consider a picture of a group of people. Various objects that are associated with the picture which may later be used for querying, could be individual people in the picture, the clothes they are wearing, their gender, the color of their clothes, the background in the image, the activity of the people in the image (standing, sitting, laughing, etc.), the spatial relationships between the various people, and so on. A textual representation of all these attributes may become very complex. Later, while querying the database, a user can think of a number of other textual attributes to describe this picture, such as, "Show me the picture with X holding a glass of wine", which had not been tagged with the images in the database. There are thus many complex attributes that are typically needed to represent an image. In the case of humans, we surely do not store textual descriptions for various images, but have a general notion of what an image contains. The goal of research in content-based addressing is to make steps in the direction of extraction of features that aid in representing and retrieving pictorial data.

It is generally agreed that image retrieval based on image content is more desirable than text-based retrieval in a number of applications. As a result, there is a need to

automatically extract primitive visual features from the images and to retrieve images on the basis of these features. Humans use color, shape, and texture to understand and recall the contents of an image. Therefore, it is natural to use features based on these attributes for image retrieval. Most of the work in image database retrieval has concentrated on identifying appropriate models for image features such as color, shape, or texture. Figure 1.2 shows a block diagram of the traditional image retrieval model. The input images are preprocessed to extract the features which are then stored along with the images in the database. When a query image is presented, it is similarly preprocessed to extract its features which are then matched with the feature vectors present in the database. A ranked set of images with high matching scores are presented at the output. Retrieval based on a single image attribute might lack a sufficient discriminatory information and might not be able to accommodate large scale and orientation changes. Our goal is to build an image database retrieval system which is insensitive to large variations in image scale, rotation, and translation. In other words, even if the query image differs from its stored representation in the database in its orientation, position, or size, the image retrieval system should be able to correctly match the query image with its prototype in the database. Recent studies [5, 6, 7, 8] have combined various image features for an efficient and effective querying by image content.

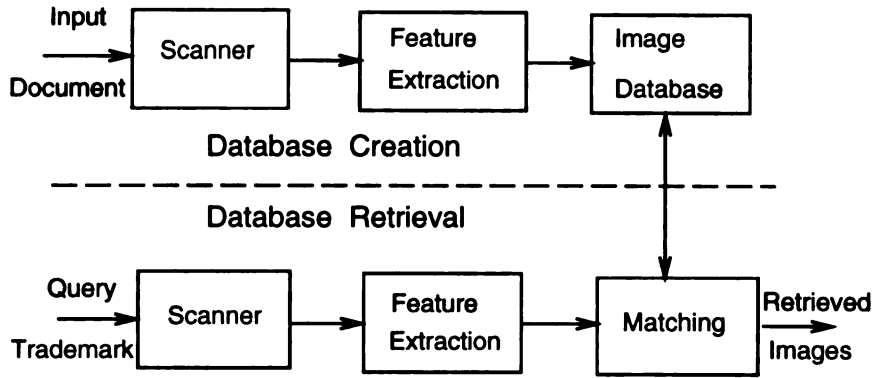


Figure 1.2: Traditional image retrieval model.

1.2 Information on Trademarks

Trademarks represent a gamut of pictorial data. There are over a million registered trademarks in the U.S. alone, and they represent a number of goods and products which are sold by different manufacturers and service and other organizations. Most of the trademarks are an abstract representation of a concept in the world, like an abstract drawing of an animal, or a natural object (Sun, Moon, etc.). It is extremely challenging and instructive to study and address the issue of image database retrieval on this huge source of pictorial data. *A trademark is either a word, phrase, symbol or design, or combination of words, phrases, symbols or designs, which identifies and distinguishes the source of goods or services of one party from those of others. A service mark is the same as a trademark except that it identifies and distinguishes the source of a service rather than a product [9].* Thus, while a trademark appears on the product or its packaging, the service mark appears in advertising for the services.

1.2.1 Registration of Trademarks

Trademarks in the U.S. are registered with the USPTO (U.S. Patent and Trademark Office). An applicant may apply for federal registration in three principle ways.

- An applicant can process a *use* application if the applicant has already commenced the use of the mark in products or services. This application is filed on the basis of the use of the mark.
- If the applicant has not yet used the mark but intends to use it in the future, the applicant can file an *intent-to-use* application on the basis of a bona fide intention to use the mark.
- The third option allows international applicants to apply for a mark that has already been applied or registered in another country. Thus, an applicant from outside the United States may file within the United States.

The application must be filed in the name of the owner of the mark. The owner is usually an individual, a corporation, or a partnership. This application may be submitted and executed by the owner for registration, or may be represented by an attorney. The nature of the goods and services provided by a mark are controlled by the owner of the mark, under whose name the application is filed.

1.2.2 Search for Conflicting Marks

Before a mark is registered with the USPTO, an examining attorney conducts a search for conflicting marks. Usually, it is not necessary for an applicant to conduct a

search for conflicting marks prior to filing an application. The application fee covers processing and search costs, and is not refunded in case a conflict is found and the mark cannot be registered.

To determine whether there is a conflict between two marks, the PTO determines whether there would be a likelihood of confusion, i.e., *whether relevant consumers would be likely to associate the goods or services of one party with those of the other party as a result of the use of the marks at issue by both parties* [10]. The principal factors to be considered in reaching this decision are the similarity of the marks and the commercial relationship between the goods and services identified by the marks [10]. In order for a conflict to occur, it is not necessary that the marks be very similar when subjected to side-by-side comparison, but the issue is whether the marks are sufficiently similar to produce a likelihood of confusion as to the source of the goods or services. Thus, marks representing similar meanings and used to sell similar goods or services may cause a conflict to the general public. While evaluating the similarities between the marks, emphasis must be placed on the recollection of the average purchaser who may normally retain a general rather than any specific details of the trademarks.

In case of design marks (trademarks consisting of symbols and designs), the issue of similarity is primarily based on *visual similarity*. Again, consideration must be placed on the fact that the purchaser's recollection of the mark is of a general and hazy nature. Figures 1.3, 1.4, 1.5, 1.6, and 1.7 show a few pictures of trademarks that were considered for opposition based on their visual similarity.

Figure 1.3 shows the example where Red Carpet Corp. applied for a stylized house

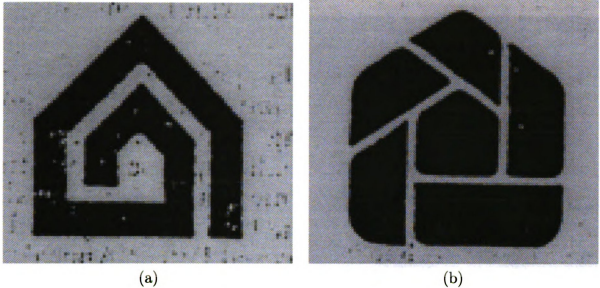


Figure 1.3: Trademarks considered for opposition by the USPTO; registration was accepted by the PTO. The trademarks have been enlarged to show the details.

design (a) for managing the real estate properties for others. Opposition was put up by Johnstown American Enterprises, Inc. with a registered trademark of a stylized house design (b) for real estate brokerage services. The PTO accepted the registration. Figure 1.4 presents the example where United Services Distributors, Inc. was refused registration of the design mark with a silhouette of two profiles facing right within a teardrop background (a) for services in the field of health and beauty aids against a registered mark of silhouette of two profiles facing left within an oval background (b) used for skin creams. Figure 1.5 presents the case between Ocean Spray Cranberries, Inc. and Ocean Garden Products, Inc. for their respective design marks. The PTO accepted the registration of the marks. Figure 1.6 presents the example where Steury Corp. was refused registration of its design mark (a) for boats and camper trailers as against a registered mark (b) for boats and campers. Figure 1.7 shows the case where Matsushita Electric Industrial Co. was refused registration of its design mark

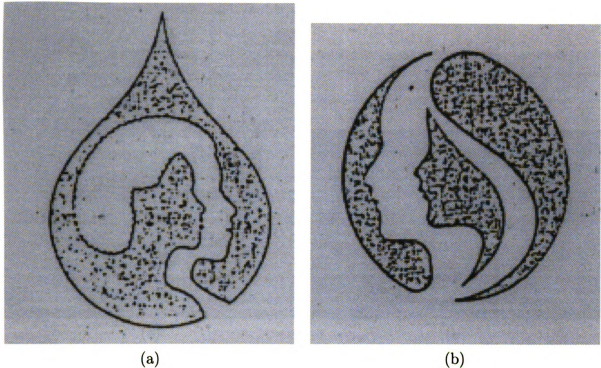


Figure 1.4: Trademarks chosen for opposition by the USPTO; registration was refused by the PTO. The trademarks have been enlarged to show the details.

of triangular arrow within a square border (a) for various items of electrical and electronic equipment as against a registered mark of Sanders Associates, Inc. with triangular arrow design (b) for various items of electrical and electronic components and equipment. For further details on these and other cases, please refer to the trademark manual for examining procedures [10].

In general, applicants want to conduct searches prior to submitting an application in order to save on the application fee ($\sim \$700$) if a conflicting mark does exist. Unless the PTO is acting on an application, it does not conduct searches for the public to determine if a conflicting mark is registered, or pending. The applicants are allowed to conduct searches of their own in a number of ways.

- The applicant can conduct a search in the PTO public search library. This is

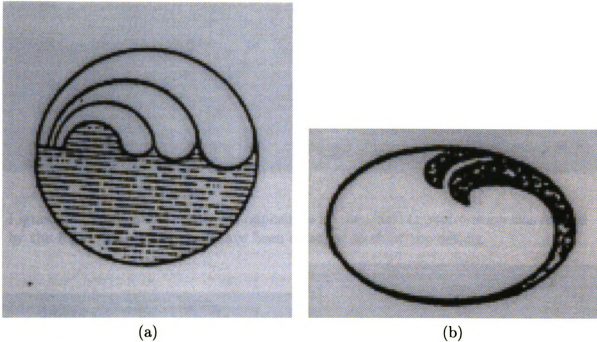


Figure 1.5: Trademarks chosen for opposition by the USPTO; registration was accepted by the PTO. The trademarks have been enlarged to show the details.

located on the second floor of the South Tower Building, 2900 Crystal Drive, Arlington, Virginia 22202.

- An alternate method is to visit a Patent and Depository Library located in every State. These libraries have CD-ROMS containing the trademark database of registered and pending marks.
- The search can also be conducted through a private trademark search company or an attorney who deals with trademark law.

The USPTO organizes the existing trademarks into categories for a more efficient retrieval of trademarks. The marks are initially grouped with respect to the goods or services they sell. For example, chemicals are assigned the code 001, vehicles are assigned the code 012, and services like advertising and business are assigned the

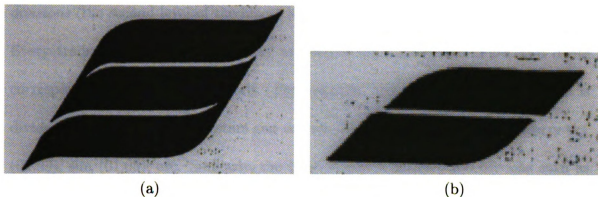


Figure 1.6: Trademarks chosen for opposition by the USPTO; registration was refused by the PTO. The trademarks have been enlarged to show the details.

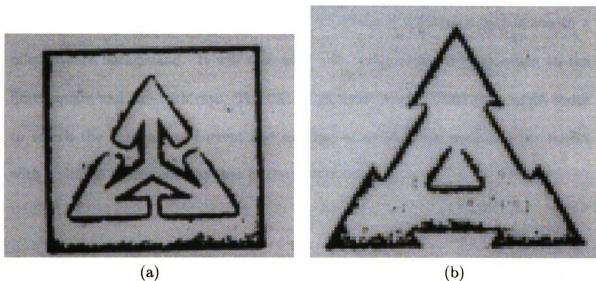


Figure 1.7: Trademarks chosen for opposition by the USPTO; registration was refused by the PTO. The trademarks have been enlarged to show the details.

code 035. The TRADEMARKSCAN Design Code Manual has been developed by Thomson and Thomson to assist an applicant in finding trademarks that have, or are, specific designs. A search mechanism based on design codes associated with the trademarks is incorporated in the CD-ROMS at depository libraries in each State. Trademarks are organized in a three-layered hierarchical structure. The highest level consists of 30 main *categories* (including celestial bodies, human beings, animals, foodstuffs, geometric figures and solids, etc). Each category is further divided into

divisions (the second level of the hierarchy), which are further divided into *sections*. Every trademark is assigned as many as possible six digit design codes, two digits corresponding to each of the levels. For example, the category celestial bodies is divided into divisions such as stars and comets (01.01), constellations and starry sky (01.03), sun (01.05), etc. Similarly, the division sun is divided into sections such as sun rising or setting (01.05.01), sun with rays or halo (01.05.02), sun representing a human face (01.05.03), etc. It can be seen that a trademark can be assigned more than one design code. For example, figure 1.8 shows a trademark with a sun in a filled square background. It will thus have two design codes corresponding to the filled square and sun with rays. The CD-ROM based search uses these design codes to search the database of current and pending trademarks to retrieve other marks with a similar design code as that of the query.

The current search strategy is based on manually assigning as many design codes as possible (referring to the design code manual) and then conducting the search based on the design codes at the Patent Depository Libraries. In general, tens of thousands of marks are retrieved for a single query due to the presence of multiple design codes for a trademark. An additional, content-based scheme is required to rank order these marks according to their visual similarity to the query trademark. This thesis addresses the problem of efficient shape-based retrieval from large databases.

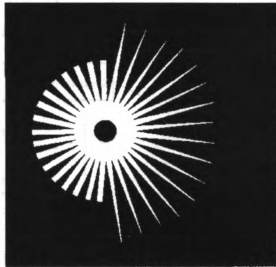


Figure 1.8: Example of a trademark with multiple design codes.

1.3 Problem Definition

In this thesis, we address the problem of an efficient and accurate retrieval from a large image database based on image content. Since we desire a system that has both high speed and high accuracy of retrievals, we propose a two-tiered hierarchical image retrieval system. Figure 1.9 shows a block diagram of our proposed image retrieval scheme. The first stage computes simple image features to prune the database to a reduced set of plausible matches. As simple shape features are used in screening the image database, this first stage can be performed very fast. The small set of plausible candidates generated by the fast screening stage are then presented to a detailed matcher in the second stage. This stage applies comprehensive feature matching techniques to eliminate false matches.

The proposed hierarchical content-based image retrieval algorithm is applied to a trademark image database with 1,100 images (See Appendix A for representative

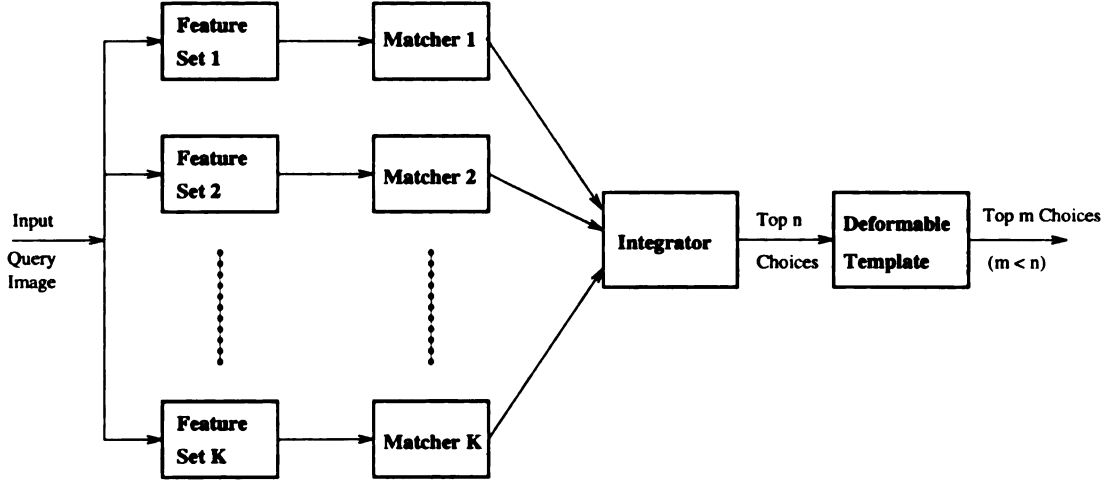


Figure 1.9: Proposed image retrieval model.

images belonging to this database). In the fast pruning stage, simple and easily calculated shape features, including edge direction histograms and invariant moments, are used to quickly browse the database for a small set of plausible matches. The retrieved images are then screened using a more elaborate, but costly deformable template matching process to remove false retrievals. The contributions of this work are as follows:

- We have explored how to sensibly combine some easily computable statistical shape features to give fast and accurate retrievals.
- The two-stage hierarchical retrieval process is able to improve the retrieval results presented to the user.
- The retrieval process is reasonably fast.
- Images resembling the queries are correctly retrieved even if they differ in their orientation, position, or size.

1.4 Motivation

Although content-based image retrieval is extremely desirable in many applications, it is a very difficult problem. The ease with which humans capture the image content has not been understood well enough to automate it. Problems arise in segmenting the images into regions corresponding to individual objects, extracting features from the images that capture the perceptual and semantic meanings from the images, and matching a database and a query image based on the extracted features. In this thesis, we conduct experiments on a large database of trademark images, addressing some of the issues in this application.

1.4.1 Database

The image database used in this study was created by scanning a large number of trademarks from several books [11, 12, 13]. About 400 images were scanned at MSU while 700 were scanned at Siemens Corporate Research, Princeton. The database consists of these 1,100 images. A design mark is registered in the binary form at the USPTO. Therefore, we have converted our scanned gray level images to binary images. The 400 trademarks gathered at MSU were scanned using a Hewlett Packard Scanjet IICx flatbed scanner at a resolution of roughly 75 dpi. The images are approximately 200×200 pixels in size. The images scanned at Siemens Corporate Research are 500×500 pixels in size. The trademarks in our database were selected so that many of them have similar perceptual meaning to make the retrieval problem more challenging. These trademarks encompass a wide variety of objects; some are

based on the alphabets of the English language, while others represent the Sun, Earth, humans, eyes, animals, etc. A representative subset of trademark images which reside in our database is shown in Appendix A.

1.4.2 Similarity

Whether two trademarks are similar enough to cause a (legal) infringement is decided by an attorney. The goal of the proposed retrieval system is to present to the user a subset of database images that are visually similar to the query. The aim is that all the registered trademarks that may cause a conflict to the query image be presented in the retrieved images. We can assume that a design code-based search has already been implemented to aid applicants to conduct searches for a conflict. A major drawback of the search based on the design codes is that it produces too many marks that have the same design code, regardless of the visual appearance of these marks. Thus, there is a need for some content-based (shape-based) retrieval to narrow down the search space.

A major drawback with content-based retrieval is that images that do seem to be perceptually similar need not be exactly similar in appearance. In fact, two images that might produce a conflict need not have a high pixel-based correlation value. As an example, figure 1.10 shows two images of a bull head. Although, these two images provide a similar perceptual meaning, they differ considerably in their correlation value. While one image is a filled head, the other is just an outline with some interior details. Figure 1.11 shows another pair of perceptually similar marks that differ in

their image correlation. While both the marks show a bear holding similar objects (a bottle in the first case, and a glass of wine and a bottle in the second), they are markedly different in their similarity to each other, for example, the direction in which the head is pointing, the difference in the bottles they are holding, and the manner in which the bottles are being held. We (humans) can identify these images as containing a bear but it is extremely difficult for a computer vision system to identify such objects automatically. As another example, figure 1.12 shows an image of a bear that is made up of a number of parts rather than a single component image. Though, humans can use Gestalt principles and perhaps top-down processing to identify the presence of a bear in this figure, it is not possible to do so with an automatic system, given the current state-of-the-art in computer vision. As a final example, figure 1.13 presents several trademark images containing corn and wheat, that are perceptually similar but are markedly different in their appearance. While some of these images are made up of a single component, others are made up of multiple components.

Since a content-based technique cannot by itself retrieve the perceptually similar images, we believe that it should be augmented by traditional text-based approaches to facilitate the search and retrieval tasks.

1.5 Thesis Outline

The outline of the thesis is as follows. Chapter 2 briefly reviews relevant literature on content-based image database retrieval methods. We describe the proposed hierarchical retrieval system in chapter 3. Experimental results on a digital trademark



(a)



(b)

Figure 1.10: Two perceptually similar images of a bull head.

library and a library of master marks of craftsmen of old Norwegian silverware are presented in chapter 4. Chapter 5 presents the conclusions and some ideas for future research.

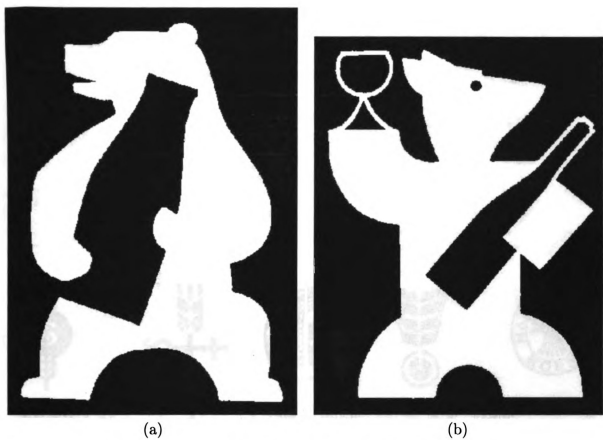


Figure 1.11: Two perceptually similar images of a bear.



Figure 1.12: Another image of a bear.



Figure 1.13: Images of wheat and corn that vary in their appearance.

Chapter 2

Literature Review

Much of the past research in image retrieval has concentrated on the feature extraction stage. For each database image, a feature vector which describes various visual cues, such as shape, texture, or color is computed. Given a query image, its feature vector is calculated and those images which are most similar to this query based on an appropriate distance measure in the feature space are retrieved. The traditional image matching methods based on image distance or correlations of pixel values are in general too expensive and not meaningful for such an application. The methods proposed in the literature can be broadly classified into two categories on the basis of the approach used for extracting the image attributes [14]. The *spatial information preserving* methods derive features that preserve the spatial information in the image. It is possible to reconstruct the image on the basis of these feature sets. Representative techniques include polygonal approximation of the object of interest, physics-based modeling, and principal component analysis (PCA). The *non-spatial information preserving* methods extract statistical features that are used to

discriminate among objects of interest. These include heuristics based on edge angle histograms, and other ad hoc shape measures. In the following sections, we briefly review the color-, shape-, and texture-based image retrieval methods.

2.1 Methods using a Single Cue

Traditional image retrieval systems use a single cue such as shape, texture, or color to represent the image and retrieval is based on the features that represent the chosen cue. Although color seems to be a highly reliable attribute for image retrieval, situations where color information is not present in the images require the use of shape and/or texture attributes for image retrieval. Retrieval based on a single image attribute might lack sufficient discriminatory information and might not be able to accomodate large scale and orientation changes. For example, color-based approaches cannot distinguish between a red apple and a red Ferrari. Additional shape information can very easily distinguish the two.

2.1.1 Color

We (humans) can effectively use color for object segmentation and recognition. Due to ease of availability of color information in digital images in the form of the three *RGB* color components, various systems have been built for color-based segmentation and retrieval.

Examples of recent work with color include color indexing using histogram intersection [15, 16], which considers intersection of color histograms; and retrieval using

distance and reference color methods [17], which employ metric distance-based comparison of color histograms. These schemes do not preserve the spatial information in the image, whereas the PCA on color features used in [6] maintains the spatial adjacencies.

Retrieval on the basis of color histograms has been shown to outperform retrieval on the basis of shape both in terms of efficiency and robustness. A 3-D histogram intersection technique [15] uses $16 \times 16 \times 16$ bins and the resulting matching is relatively fast. Reference color method [17] improves the performance further by using fewer reference color bins. Color histograms are generally invariant to translation and rotation of the images and normalizing the histograms leads to scale invariance. However, color histograms do not incorporate spatial adjacency of pixels in the image and may lead to inaccuracies in the retrieval.

2.1.2 Shape

Although color can be an effective means of querying, color alone as a retrieval cue cannot be effective for querying large image databases. Applications with grayscale or binary images have to use other cues such as shape and texture for retrieval. Although, humans can effectively use color to differentiate among natural objects, many artificial (man made) objects cannot be distinguished on the basis of color alone. Moreover, humans when presented with binary or grayscale images can easily distinguish among these.

Various schemes have been proposed in the literature for shape-based retrieval.

These include polygonal approximation of the shape [16]; shape matching using relaxation techniques [18], which uses relaxation methods to find acceptable combinations of matches between pairs of angles on two shapes; image representation on the basis of strings [19, 20], which represent the shape of objects as strings and consider string matching techniques for retrieval; comparing images using the Hausdorff distance [21], which measures the extent to which each point of the stored database image lies near some point of the query and vice versa; experiments in point matching techniques [22], which extract a set of distinctive local features from the query and the model, and then match the resulting point patterns; image registration by matching relational structures [23], which uses relational structures to represent images; shape matching based on chord distributions [24], which uses chord length distribution for image matching; image representation using Codons [25], which uses continuous curve segments in an image that are separated by concave cusps to represent the object shape; matching objects using Fourier descriptors [26]; and object matching using invariant moments [27].

The above schemes introduce various shape-based features for image retrieval. These techniques rely on a single concise feature to describe the shape. A major limitation of using a single shape model in image database retrieval is that it might not be possible to extract the corresponding features in a given application domain. Moreover, shape-based representation schemes are not generally invariant to large variations of image size, position, and orientation. In order to incorporate invariance to rigid motions (rotation and translation), these methods need to be applied for all possible rotations and translations thereby reducing the speed of the retrievals. When

we consider large image databases, this speed reduction can become considerable. We, therefore, need to identify shape features which are either invariant to rigid transformations or which can be efficiently computed over a number of possible rigid transformations.

2.1.3 Texture

Many machine vision and image processing algorithms neglect the presence of texture in an image. Variations of intensities that form certain repeated patterns are called visual texture [28]. These patterns can be the result of physical properties of the object surface (roughness, peakedness), or be the result of reflectance differences such as the color on a surface. Humans can very easily recognize a texture upon seeing it, yet it is very difficult to define it. Texture analysis is an important and useful area of study in computer vision. Most natural surfaces exhibit texture and it may be useful to extract texture features for querying.

Texture models developed in the literature can be divided into the following four classes [28]. *Statistical methods* define texture in terms of the spatial distribution of grey values. These include the use of co-occurrence matrices [29] and autocorrelation features (extracting repetitive nature of placement of texture elements). *Geometric methods* are characterized as being composed of “texture elements” or primitives. These include Voronoi tessellation features [30] and structural methods that extract the texture primitives [31]. *Model based methods* assume an underlying image model to describe and synthesize texture. These include the use of random field models [32]

and fractals [33]. *Signal processing methods* use frequency analysis of the image to classify texture. The schemes include the use of spatial domain [34] and Fourier domain [35] filtering, and the use of Gabor filters and wavelet models [36].

The above schemes define various simple texture analysis schemes. In the absence of color and shape information, texture can act as a vital cue for image classification. Scenes containing pictures of wood, grass, etc. can be easily classified based on the texture rather than shape or color. The numerous different texture definitions that have been attempted in the literature demonstrate the difficulty in defining texture. Although, humans can easily identify texture, a concise definition has evaded past research.

Retrieval based on a single image attribute lacks sufficient discriminatory information and might not be able to accomodate large scale and orientation changes. As we consider large databases, we need to extract multiple features for querying the database. Recently, a number of studies have been carried out [5, 6, 7, 8, 37] which combine the various features for efficient and effective querying by image content.

2.2 Using Multiple Cues

Recently, attempts have been made to develop general purpose image retrieval systems based on multiple features that describe the image content. These systems attempt to combine shape, color, and texture cues for a more accurate retrieval.

2.2.1 QBIC

The QBIC system has developed methods to query large on-line image databases using image content as the basis of the query. In other words, the system allows users to search through databases consisting of very large numbers of images using sketches, layout or structural descriptions, texture, color, and sample images. QBIC techniques serve as database filters and reduce the search complexity for the user. These techniques limit the content-based features to those parameters that can be easily extracted, such as color distribution, texture, global shape of an image, and layout. The system offers a user a virtually unlimited set of unanticipated queries thus allowing for general purpose applications rather than catering to a particular application. Color- and texture-based queries are allowed for both images and objects, whereas shape-based queries are allowed only for individual objects and layout-based queries are allowed only for the entire image. QBIC uses the following image features for content-based retrieval.

- **Color:** QBIC uses a k -element color histogram for matching based on color. The entire color space is quantized into k levels, such that these k cells correspond to super-cells in the color space. The similarity between a pair of color histograms, \mathbf{X} and \mathbf{Y} can be computed in a number of ways. The following method proposed in [38] is used for matching the color histograms.

$$d_{hist}^2(\mathbf{X}, \mathbf{Y}) = (\mathbf{X} - \mathbf{Y})^t A (\mathbf{X} - \mathbf{Y}) = \sum_i^k \sum_j^k a_{ij} (x_i - y_i)(x_j - y_j),$$

where the superscript t denotes vector transposition, and the matrix A has entries a_{ij} which describe the similarity between color i and color j . Euclidean distance is a special case of this distance measure when the matrix A is the identity matrix.

- Texture: QBIC uses modified versions of three texture features, namely, coarseness, contrast, and directionality. Coarseness measures the texture scale, contrast measures the gray level variance, and directionality describes whether an image has a preferred direction. The texture distance between two images is computed as a weighted Euclidean distance in the three-dimensional texture space, where the weights are the inverse of variance along the three dimensions.

Distance between image i and image j is thus defined as,

$$d_{ij} = \frac{(O_i - O_j)^2}{\sigma_O^2} + \frac{(C_i - C_j)^2}{\sigma_C^2} + \frac{(D_i - D_j)^2}{\sigma_D^2},$$

where O , C , and D are the coarseness, contrast, and directionality measures, respectively.

- Shape: QBIC primarily uses statistical features to represent the shape of an object. These include a combination of area, circularity, eccentricity, major axis orientation, and algebraic moment invariants. These features are computed from binary images representing individual objects. The distance between two images is computed similar to the distance in the case of texture using a weighted Euclidean distance with the inverse of feature variances used for normalization.

QBIC allows the user to use any subset of these shape features in their query.

- **Sketch:** A reduced resolution edge map is used to represent the sketch of an object. The similarity measure is computed based on the spatial correlation of the subimages.

Although the QBIC system allows the use of multiple cues such as color, shape, texture, and sketch for querying, users have to be trained well enough before they can use the system effectively. Due to the presence of a large variety of images (natural scenes, man made objects, etc), the user needs to use effective features that are most expressive in identifying the class of the query object and retrieve stored images from that class. For example, while querying for scenes of buildings, it is more meaningful to query on the basis of texture than on the basis of color, but while trying to identify images of plants and grass, it may be more meaningful to query on the basis of green color and texture. Since the user has to specify a combination of features to use for querying, the retrieval results critically depend on the ability of the user to identify expressive features for the query.

2.2.2 Photobook

Photobook is a set of interactive tools for browsing and searching image sequences. The features used for querying can be based on both text annotations and image content. The key idea behind the system is *semantics-preserving image compression*, which reduces images to a small set of perceptually significant coefficients. These features describe the shape and texture of the images in the database. Photobook

uses multiple image features for querying general purpose image databases. The user is given the choice to select features based on the appearance, shape, and texture to browse through large databases. These features can be used in any combination with textual features to improve the efficiency and accuracy of the retrievals. The following schemes are used to extract the semantics-preserving features.

- **Appearance Specific:** Appearance Photobook measures the effective similarity in visual appearance within an object class. Similarity measure is based on image correlation. Since, images are typically of large sizes (typically 512×512), a pixel by pixel correlation is very expensive. Moreover, this requires storing large size images. An efficient way to extract semantics-preserving features is by applying principal component analysis to gray scale images. The images are represented in terms of a small number of dominant eigenvectors which preserve a large fraction of the gray level variance present in the input image. The projection of the images onto these orthogonal eigenvectors represents the image features. Usually, a low dimensional (as compared to the original number of pixels) feature vector (about 20 features) is very effective in capturing the image content. Euclidean distance based on the low-dimensional image features is used as an approximation to the image correlation. Such a scheme makes no assumptions about the image content and uses the training samples (many objects belonging to a class) to extract the principal components.
- **Shape Specific:** Shape Photobook compares the shape of the two objects by measuring the deformations that relate them. Physics-based modeling is used

to define the transformations of real world objects. Due to changes in viewing direction and distance of the object from the sensor, and presence of physical deformities (bent, stretched objects), appearance-based matches produce inefficient results. Under these circumstances, the transformations between these objects can be modeled to extract a similarity value between the objects. The shape model uses a virtual material to fill up the space between various shape-based image features. These features include, edges, corners, and high-curvature points. A finite element method is used to compute this interconnectedness of the shape. Eigenvectors of the corresponding stiffness matrix are used as the semantics-preserving features.

- **Texture Specific:** Texture Photobook is applied to databases of texture patches, where both appearance- and shape-based features prove ineffective. The semantics preserving texture features are represented based on Wold decomposition for regular stationary stochastic processes in 2-D images [39]. Photobook also provides many other models for texture such as color histogram differences, color histogram energy and entropy, color histogram invariant features, eigenvectors of *RGB* covariance, multiscale simultaneous auto-regressive model, and tree-structured wavelet transforms [40].

2.2.3 STAR

STAR (System for Trademark Archival and Retrieval) [41, 42] uses a combination of color and shape features for retrieval purposes. The color of an image is represented

in terms of the R , G , and B color components, whereas the shape is represented in terms of combination of outline-based features (sketch of the images) and region-based features (objects in an image). The features used to describe both the color and shape in an image are non-information preserving or ambiguous in nature. Though these features cannot be used to reconstruct the image, they are useful as approximate indicators of shape.

- Color: Color in an image is represented as a histogram of reference colors. These reference colors are selected manually and the color space is divided to represent bins corresponding to each reference color. A histogram intersection based on 27 reference colors is used to query on the basis of color.
- Shape: Multiple features are used to query the database on the basis of shape. These include the use of invariant moments and Fourier descriptors for shape-based matches. In order to describe the Fourier descriptors, the system assumes a segmentation of objects of interest. An integrated retrieval based on the two features is used to query on the basis of shape.

2.2.4 Incorporating Learning

Although, the image retrieval systems based on multiple cues outperform systems based on a single cue, they have various limitations. The features used for image retrieval are hard-coded into the system and are usually application specific. For different applications, these features and the associated weights have to be specifically tweaked for efficient performance. A typical user of these systems does not have the

basic knowledge of feature extraction and thus, is unable to use the system effectively. A general purpose image database system should thus be able to automatically decide on the image features for retrieval purposes [43]. A recent project [44] describes an interactive learning system using a society of models. Instead of requiring universal similarity measures or manual selection of relevant features, this approach provides a learning algorithm for selecting and combining groupings of the data, where these groupings are generated by highly specialized and context-dependent features. The selection process is guided by a rich user interaction where the user generates both positive and negative retrieval examples. A greedy strategy is used to select a combination of existing groupings from the set of all possible groupings. These modified groupings are generated based on the user interactions and over a period of time, these replace the initial groupings that have a very low weightage. Thus, the system improves on its performance over a period of time through user interaction and feedback.

2.3 Discussion

While multiple cue-based schemes prove more effective than single cue-based retrieval systems, content-based retrieval still faces many problems and challenges. The multiple cue-based schemes generally employ various cues such as color, shape, and texture for the retrievals. In the case of trademark images, color does not play a useful role in distinguishing between various marks. The design marks are registered in binary image form with the USPTO. When searching for a conflict, the USPTO bases its

decision on the shape information present in the binary images. Thus, cues like color and texture are not applicable for query purposes. We therefore, need to investigate multiple shape models for the retrieval of trademark images. In the next chapter, we describe our proposed hierarchical retrieval scheme in detail.

Chapter 3

An Integrated System for Efficient Retrieval

Retrieval *speed* and *accuracy* are two main issues in designing image databases. System accuracy can be defined in terms of *precision* and *recall* rates. A *precision rate* can be defined as the percent of retrieved images similar to the query among the total number of retrieved images. A *recall rate* is defined as the percent of retrieved images which are similar to the query among the total number of images similar to the query in the database. It can be easily seen that both precision and recall rates are a function of the total number of retrieved images. In order to have a high accuracy, the system needs to have both a high precision and a high recall rate. Although, simple image features can be easily extracted, they lack sufficient expressiveness and discriminatory information to determine if two images have a similar content. Thus, there exists a trade-off between speed and accuracy.

In order to build a system with both high speed and accuracy, we use a hierarchical

two-level feature extraction and matching structure for image retrieval (Fig. 1.9). Our system uses multiple shape features for the initial pruning stage. Retrievals based on these features are integrated [8] for better accuracy and higher system recall rate. The second stage uses deformable template matching to eliminate the false retrievals present at the output of the first stage, thereby improving the precision rate of the system.

3.1 Image Attributes

In order to retrieve images, we must be able to efficiently compare two images to determine if they have a similar content. An efficient matching scheme further depends upon the discriminatory information contained in the extracted features.

Let $\{\mathbf{F}(x, y); x, y = 1, 2, \dots, N\}$ be a two-dimensional image pixel array. For color images, $\mathbf{F}(x, y)$ denotes the color value at pixel (x, y) . Assuming that the color information is represented in terms of the three primary colors (Red, Green, and Blue), the image function can be written as $\mathbf{F}(x, y) = \{F_R(x, y), F_G(x, y), F_B(x, y)\}$. For black and white images, $\mathbf{F}(x, y)$ denotes the gray scale intensity value at pixel (x, y) . Let f represent a mapping from the image space onto the n -dimensional feature space, $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, i.e.,

$$f : \mathbf{F} \rightarrow \mathbf{x},$$

where n is the number of features used to represent the image. The difference between two images, \mathbf{F}_1 and \mathbf{F}_2 , can be expressed as the distance, D , between the respective feature vectors, \mathbf{x}_1 and \mathbf{x}_2 . The problem of retrieval can then be posed as follows:

Given a query image \mathbf{P} , retrieve an image \mathbf{M} from the image database, \mathcal{S} , such that

$$D(f(\mathbf{P}), f(\mathbf{M})) \leq D(f(\mathbf{P}), f(\mathbf{F})), \quad \forall \mathbf{F} \in \mathcal{S}, \mathbf{F} \neq \mathbf{M}.$$

3.2 Fast Pruning Stage

It is desirable to have an image retrieval system which is insensitive to large variations in image scale, rotation, and translation. Hence, the pruning stage has to be not only fast but should also extract invariant features for matching. We use the following features to represent the shape of an image.

- **Edge Angles:** A histogram of the edge directions [8] is used to describe global shape information.
- **Invariant Moments:** The global image shape is also described in terms of seven invariant moments [27].

3.2.1 Edge Directions

A histogram of the edge directions is used to represent the shape attribute. The edge information contained in the database images is generated in the preprocessing stage using the Canny edge operator [45]. Euclidean distance is used to compute the dissimilarity value between two histograms.

There are many advantages and limitations in representing an image with its edge directions.

- Use of edge directions captures the general shape information.
- Histogram of the edge directions is invariant to translations in the image. Thus, the positions of the objects in the image have no effect on the edge directions. This may also turn out to be a limitation, as two totally different images may yield similar edge direction histograms.
- The use of edge directions is inherently not scale invariant. Two images identical in every respect except their size will yield different numbers of edge points and hence different histograms. In order to have invariance to scale, we normalize the histograms with respect to the number of edge points in the image. A drawback of normalized histograms is its inability to match parts of images. If an image, Q is a part of another image, I , then the histogram of Q is contained within the histogram of I . Normalized histograms do not satisfy this property.
- A histogram of the edge directions is also not invariant to rotation. A shift of the histogram bins during matching partially takes into account image rotations. A rotation of the image shifts each of the edge directions by the amount of rotation. But, due to quantization effects in binning the edge directions, the effect of rotation is more than a simple shift in the bins. Rotation also affects the bin memberships. For example, if a quantization of 45° (bin size = 45°) is used, then two edge points with directions of 30° and 40° will fall into the same bin. But, if the same image is rotated by 10° then these two points will fall into adjacent bins. We note that for a rotation of 45° , there is no such effect. To reduce this effect of rotation, we smooth the histograms. A histogram can be

treated as a 1-D discrete signal. Smoothing can then be defined as

$$I_s[i] = \frac{\sum_{j=i-k}^{i+k} I[j]}{2k+1}, \quad (3.1)$$

where I is the original histogram, I_s is the smoothed histogram, and the parameter k determines the degree of smoothing. In our experiments we have used $k = 1$.

- The matching results depend on the bin size. By choosing a larger bin size, the matching speed is increased. But, this also reduces the accuracy in case of an arbitrary rotation of the image. A rotation not only causes a shift in the histogram bins but also affects the bin membership. Use of a very small bin size reduces the matching speed (since the number of bins increases). Use of a small bin size also requires that the edge directions be found to a very high degree of accuracy.

Fig. 3.1 shows three database images ((a) and (b) are similar to each other but (c) is different from (a) and (b)) and their respective edge angle histograms. The value of D_s denotes the shape-based dissimilarity value between two images. Note that $D_s(g, h) < D_s(g, i)$ and $D_s(g, h) < D_s(h, i)$, where Figs. 3.1(g)-(i) represent the shape histograms for the three database images.

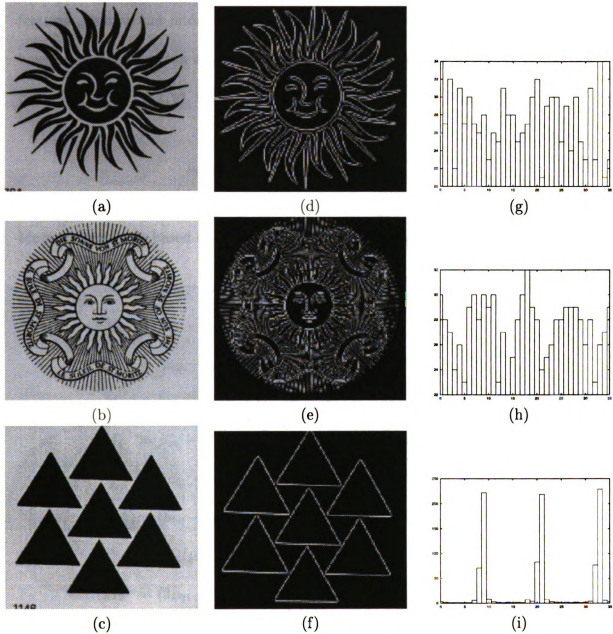


Figure 3.1: Example of shape representation using edge directions; (a)-(c) show 3 database images, (d)-(f) corresponding edge images, and (g)-(i) corresponding edge direction histograms; $D_s(g, h) = 0.065$, $D_s(g, i) = 0.69$, $D_s(h, i) = 0.70$.

3.2.2 Invariant Moments

We also represent the shape of an image in terms of 7 invariant moments [27]. These features are invariant under rotation, scale, translation, and reflection of images and have been widely used in a number of applications due to their invariance properties. For a 2-D image, $f(x, y)$, the central moment of order $(p + q)$ is given by

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y). \quad (3.2)$$

Moment invariants based on the 2nd- and 3rd-order moments are given as follows:

$$\begin{aligned} M_1 &= (\mu_{20} + \mu_{02}), \\ M_2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2, \\ M_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2, \\ M_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2, \\ M_5 &= (\mu_{30} + \mu_{12})(\mu_{30} - 3\mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ &\quad + (3\mu_{21} - \mu_{03})(\mu_{21} + 3\mu_{03})[3(\mu_{03} + \mu_{21})^2 - (\mu_{21} - \mu_{03})^2], \\ M_6 &= (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\ &\quad + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}), \\ M_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ &\quad - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{03} + \mu_{21})^2 - (\mu_{21} - \mu_{03})^2]. \end{aligned}$$

M_1 through M_6 are invariant under rotation and reflection. M_7 is invariant only in its absolute magnitude under a reflection. Scale invariance is achieved through the following transform.

$$M'_1 = M_1/n, \quad M'_2 = M_2/r^4, \quad M'_3 = M_3/r^6, \quad M'_4 = M_4/r^6,$$

$$M'_5 = M_5/r^{12}, \quad M'_6 = M_6/r^8, \quad M'_7 = M_7/r^{12},$$

where n is the number of object points and r is the radius of gyration of the object:

$$r = (\mu_{20} + \mu_{02})^{1/2}.$$

Fig. 3.2 shows 3 database images ((a) and (b) are similar to each other but (c) is different from (a) and (b)) and their respective distances based on the invariant moments. D_m represents the moments-based dissimilarity value of the images. Note that $D_m(a, b) < D_m(a, c)$ and $D_m(a, b) < D_m(b, c)$.

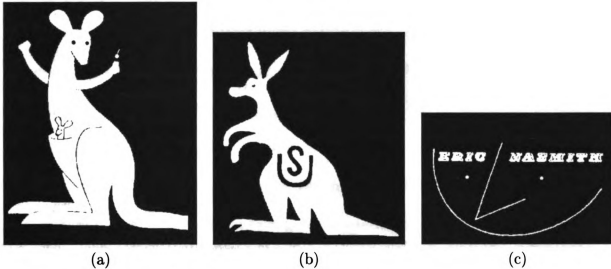


Figure 3.2: Example of shape representation using invariant moments; (a)-(c) show 3 database images; $D_m(a, b) = 0.033$, $D_m(a, c) = 0.85$, $D_m(b, c) = 0.85$.

3.2.3 Other Shape Features

We have also investigated several other simple shape features reported in the literature. We attempted to use the features based on the turning edge angles [46] on the

object boundary as the shape feature, based on the assumption that the object has only one closed contour. The object boundary is uniformly sampled and approximated by a polygon. The turning angle at each node represents the shape feature. When two shapes are compared, the correlation between the two turning angle signatures is calculated as the matching score. This measure is translation, scale and rotation invariant. However, it requires the object to have one closed boundary and can not be applied to complex images which are present in our database. Design marks are usually stylistic and artistic and turning angle of subsampled image points are not sufficient to capture the fine details in the image. We also tried to use the Principal Component Analysis (PCA) method to extract the shape features which has been used to organize and index a large image database [47]. However, this approach also does not work for our trademark image database. The PCA method merely computes the image distance in a *reduced* space (compared to the N^2 -dimensional feature space, where the image is of size $N \times N$), and it does not effectively capture the shape characteristics of our binary images. We feel that image correlation is not an efficient way to capture the perceptual similarity between images. Perceptually similar images as shown in Figures 1.10, 1.11, and 1.13 cannot be classified as belonging to the same class with a simple image correlation method.

3.2.4 Integration of Image Attributes

Use of a single image attribute for retrieval may lack sufficient discriminatory information and might not be able to support large variations in image orientation and

scale. In order to increase the accuracy of the retrievals, we need to integrate the results obtained from the query based on individual shape features.

The edge direction-based matching takes into consideration the boundary of the objects. It is thus a boundary based matching scheme. Invariant moments are defined over the object region. Thus, it is a region-based matching scheme. By integrating the two schemes, we try to retrieve those images that resemble the query image in either the boundary or its entirety. The output of a query on the basis of either the edge directions or the invariant moments is a ranked set (based on the dissimilarity value) of database images. An integrated rank of a retrieved image can be computed from either the ranks of the retrieved image in the individual queries (query on the basis of edge directions, or query on the basis of invariant moments) or the actual dissimilarity value of the retrieved image in the individual queries. We have integrated the results of the shape-based queries based on the rank of the retrieved images. This integrated scheme performed worse than each of the individual feature-based queries. Using the individual ranks may not be very effective since it does not use the dissimilarity value between the query and the retrieved image. The rank of the retrieved images may give a false sense of similarity when actually the dissimilarity value may be very high. Hence, we need to look into an integration scheme that takes into account the actual dissimilarity values between the images.

We have integrated the results of the shape-based retrievals by combining the associated dissimilarity values. Let Q be a query image and I be a database image. Let S_e be the dissimilarity index between Q and I on the basis of edge directions and S_m be the dissimilarity index between Q and I on the basis of invariant moments.

We define an integrated dissimilarity index S_t between Q and I as,

$$S_t = \frac{w_e * S_e + w_m * S_m}{w_e + w_m}, \quad (3.3)$$

where w_e and w_m are the weights assigned to the edge direction-based similarity and the invariant moment-based similarity, respectively. A set of top ten images on the basis of the total dissimilarity index S_t is presented as a result of the query. We have used equal weights ($w_e = w_m = 1$) in our experiment. Another way of determining the weights is on the basis of the accuracy of the individual feature-based queries. In general, we found that the edge direction-based queries are more accurate than moment-based queries and this fact can be used to assign a higher weight to the edge direction-based similarity values.

One of the difficulties involved in integrating the two shape-based retrievals is the difference in the range of values that are present for these retrievals. In order to have an efficient and robust integration scheme, we normalize the dissimilarity values based on the two shape-based schemes to be within the same range of $[0, 1]$. We normalize the retrieval of each query within the range of $[0, 1]$ using the following equation,

$$D'_s(i, j) = \frac{(D_s(i, j) - distmin)}{(distmax - distmin)}, \quad (3.4)$$

where D_s is the dissimilarity value between the i th query and the j th database image, and $distmin$ and $distmax$ are the distance of the most similar and dissimilar image in the database to the query image, respectively.

3.3 Deformable Template Matching

Both the edge direction histogram and the small number of invariant moments used in the previous section are necessary but not sufficient measures for shape matching. In other words, two dramatically different shapes can have very similar edge direction histograms and invariant moment features. It is observed that using the above features, similar shapes are likely to be among the top retrievals; however, the top retrievals also contain some trademarks that seem to be perceptually very different. To further refine the retrievals to guarantee that only visually similar shapes are reported, we use a more elaborate matching technique based on deformable templates [48, 49]. During the refined matching stage, the edge map of the query trademark is compared to the edge maps of the top N retrieved trademark images (referred to as test trademarks). The edge map of the query trademark is used as a prototype template; this template is deformed towards the edge map of the test trademark. The goodness of the matching is given by an energy function which describes the amount of deformation and how the deformed template fits the test edge map.

In the matching scheme defined in Jain et al. [48], the deformation model consists of (i) a prototype template which describes the representative shape of a class of objects, and (ii) a set of parametric transformations which deforms the template. The query image is used as the prototype, which is deformed to match the test trademarks. A fitness measure is then calculated to quantify the quality of the match.

3.3.1 Representation of the Prototype Template

The prototype template consists of a set of points on the object contour, which is not necessarily closed, and can consist of several connected components. In particular, the edge map of the query trademark image is used (Fig. 3.3 (b)). Since the images in the database are binary, the edge map is obtained as the set of foreground pixels which have at least one neighboring background pixel.

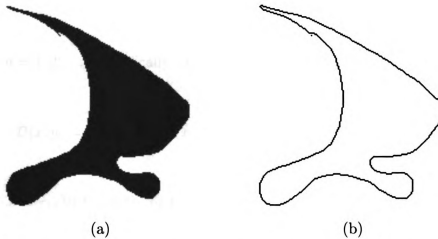


Figure 3.3: Constructing the query template for deformable template matching. (a) A hand drawn query trademark, and (b) the query template obtained by calculating the edge map of the image in (a).

3.3.2 Deformation Transformation

The prototype template describes only one of the possible (though most likely) instances of the shape of interest. Therefore, it has to be deformed to match similar trademarks. A deformation of the template is performed by introducing a displacement field in the 2D template image. Without a loss of generality, it is assumed that the template is drawn on a unit square $\mathcal{S} = [0, 1]^2$. The points in the square

are mapped by the function $(x, y) \mapsto (x, y) + (\mathcal{D}^x(x, y), \mathcal{D}^y(x, y))$, where the displacement functions $\mathcal{D}^x(x, y)$ and $\mathcal{D}^y(x, y)$ are continuous and satisfy the following boundary conditions: $\mathcal{D}^x(0, y) \equiv \mathcal{D}^x(1, y) \equiv \mathcal{D}^y(x, 0) \equiv \mathcal{D}^y(x, 1) \equiv 0$. The space of such displacement functions is spanned by the following orthogonal bases [50]:

$$\begin{aligned} \mathbf{e}_{mn}^x(x, y) &= (2 \sin(\pi n x) \cos(\pi m y), 0) \\ \mathbf{e}_{mn}^y(x, y) &= (0, 2 \cos(\pi m x) \sin(\pi n y)), \end{aligned} \quad (3.5)$$

where $m, n = 1, 2, \dots$. Specifically, the displacement function is chosen as follows:

$$\mathcal{D}(x, y) = (\mathcal{D}^x(x, y), \mathcal{D}^y(x, y)) = \sum_{m=1}^M \sum_{n=1}^N \frac{\xi_{mn}^x \cdot \mathbf{e}_{mn}^x + \xi_{mn}^y \cdot \mathbf{e}_{mn}^y}{\lambda_{mn}}, \quad (3.6)$$

where $\lambda_{mn} = \alpha \pi^2 (n^2 + m^2)$, $m, n = 1, 2, \dots$ are the normalizing constants. The parameters $\underline{\xi} = \{(\xi_{mn}^x, \xi_{mn}^y), m, n = 1, 2, \dots\}$, which are the projections of the displacement function on the orthogonal basis, uniquely define the displacement field, and hence the deformation.

3.3.3 Dissimilarity Measure

The discrepancy of a trademark edge map to the query template is described by a dissimilarity measure. This measure consists of two terms: the first term takes into account the amount of deformation: the larger the deformation, the more the deformed template deviates from the prototype; the second term considers the fidelity of the deformed prototype in terms of the discrepancy between the deformed template

and the edge map of the test trademark. These two factors account for the difference between the query and test trademarks.

Formally, the dissimilarity measure of a trademark Y and query template \mathcal{T} is defined as:

$$\mathcal{L}(\mathcal{T}, Y) = \min_{s, \Theta, \underline{\xi}, \underline{d}} \left\{ \gamma \sum_{m=1}^M \sum_{n=1}^N (\xi_{mn}^x)^2 + (\xi_{mn}^y)^2 + \mathcal{E}(\mathcal{T}_{s, \Theta, \underline{\xi}, \underline{d}}, Y) \right\}, \quad (3.7)$$

where $\mathcal{T}_{s, \Theta, \underline{\xi}, \underline{d}}$ denotes the deformed template \mathcal{T} with scale s , orientation Θ , deformation $\underline{\xi}$, and position \underline{d} ; the first term, which is the sum of squares of the deformation parameters, measures the deviation of the deformed template from the prototype template; the second term is the energy function that relates the deformed template $\mathcal{T}_{s, \Theta, \underline{\xi}, \underline{d}}$ to the edges in the test trademark image Y :

$$\mathcal{E}(\mathcal{T}_{s, \Theta, \underline{\xi}, \underline{d}}, Y) = \frac{1}{n_{\mathcal{T}}} \sum (1 + \Phi(x, y) |\cos(\beta(x, y))|), \quad (3.8)$$

where $\Phi(x, y) = -\exp\{-\rho(\delta_x^2 + \delta_y^2)^{1/2}\}$ is defined in terms of the displacements (δ_x, δ_y) of a pixel (x, y) to its nearest edge pixel, $\beta(x, y)$ is the angle between the tangent of the nearest edge and the tangent direction of the template at (x, y) , ρ is a smoothing factor which controls the degree of smoothness of the potential field, the summation is over all the pixels on the deformed template, $n_{\mathcal{T}}$ is the number of pixels on the template, and the constant 1 is added so that the potentials are positive and take values between 0 and 1. Intuitively, the first term in Eq. (3.7) favors small deformations, and the second term requires that the deformed template be in the proximity of and aligned

with the edge directions of the test image. The parameter γ provides a relative weighting of the two penalty measures; a larger value of γ implies a lower variance of the deformation parameters, and as a result, a more rigid template. This dissimilarity is always nonnegative and it is zero if and only if the query template matches the edge map of the test trademark image exactly.

The function \mathcal{L} defined in Eq. (3.7) involves minimizing an energy function *w.r.t.* the pose and deformation parameters $s, \Theta, \underline{\xi}$, and \underline{d} . The optimization is approximated by finding an initial guess for the pose parameters s, Θ and \underline{d} using the generalized Hough transform, and then performing a gradient descent search in the parameter space.

The proposed deformable template-based matching scheme has been applied to several objects in different scenes. The work presents a systematic approach to object localization and retrieval using deformable templates. The scheme has been tested on a number of cluttered images as well as images with objects present under different scales and orientations. For a detailed analysis of the results of this work, interested readers are requested to refer to [48].

3.4 Summary

We have presented our framework for a hierarchical system for image retrieval. The retrieval consists of a two-stage process, the fast pruning stage for efficient selection of a small subset of plausible similar images, and a comprehensive matching strategy based on deformable template analysis to discard false retrievals.

Chapter 4

Experimental Results

We have applied the hierarchical shape-based retrieval algorithm to a trademark image database and a database of master-marks on Norwegian silver tankards.

- **Trademark Database:** In the case of trademark images, our goal is to present the user with a subset of images that are most similar to the user. We conduct experiments with rotated, scaled, and noisy versions of database images, and with hand drawn sketches of trademarks present in the database. The system integrates the retrievals based on the edge directions and invariant moments to present to the second stage a small subset of images ranked in their similarity to the query image. The second stage applies deformable templates to the pruned set of images to further refine the output.
- **Norwegian Silver Authentication:** In the case of master-marks, the system uses the integrated shape-based match (edge directions and invariant moments) on a database of stored master-marks. The query images are master-marks cap-

tured from the bottom of the silver tankards. These marks are very noisy and sometimes the marks are faded due to the wear and tear of the tankard. The master-mark images are processed to improve their quality and the user is presented with a subset of database images with a high similarity value to the query image. For the retrieval results to be good, we expect that the actual master-mark be retrieved, given the noisy query. Once, the master has been identified, the input tankard can be matched in detail with the known tankards from the same master for authentication.

4.1 Results on Trademark Images

In this section, we present experimental results on the accuracy of the pruning stage and the need for integration of multiple shape features. We next present the retrieval results of the entire system. In the end, we discuss experiments with human subjects and possible heuristics to improve the retrieval results.

4.1.1 Experiments on Accuracy

In order to measure the accuracy of the retrievals of the pruning stage, we conducted the following three experiments.

- *Rotated (R)*: Every image in the database was rotated arbitrarily and then presented as the query image.
- *Scaled (S)*: Every image in the database was scaled and presented as the query image.

- *Noisy (N)*: A uniform i.i.d. additive noise model was used to change either 5% of the pixel values (regions for moments) or 5% of the edge orientations (addition of noise to edges in the case of edge directions).

Tables 4.1 and 4.2 present the results of the retrieval using shape features based on the edge direction histograms and the invariant moments. We notice that both the individual shape features are not very effective in retrieving rotated images. The performance of these shape-based features improves for scaled and noisy images. The results of the integrated query are presented in Table 4.3. We present the results with the integration weights $w_e = w_m = 1$. Another possibility is to choose the weights on the basis of the accuracy of the individual feature-based queries. For retrieving scaled and noisy images, the topmost retrieved image is the correct result for each of the presented query. In the case of rotated images, the query matches the correct image within the top 20 positions in all but 17 of the 1,100 cases. Figure 4.1 shows the 17 database images that were not retrieved in our rotation experiments. Most of these 17 images are line drawings and present very few image points for robust calculation of the invariant moments. Moreover, the edge directions also cannot be retrieved accurately for thin line drawings and small isolated blobs. We feel that both the edge directions and invariant moments are not very robust shape measures for line drawings. Integrating the two schemes reduces the number of false retrievals as it is highly unlikely that a pair of perceptually different images is assigned a high similarity value in both the schemes.

Query Nature	$n = 1$ (%)	$n \leq 2$ (%)	$n \leq 5$ (%)	$n \leq 20$ (%)	Not Retrieved (%)
R	73	79	86	94	6
S	100	100	100	100	0
N	92	95	96	100	0

Table 4.1: Edge directions-based retrieval results for the 1, 100 database images.

Query Nature	$n = 1$ (%)	$n \leq 2$ (%)	$n \leq 5$ (%)	$n \leq 20$ (%)	Not Retrieved (%)
R	44	55	67	88	12
S	100	100	100	100	0
N	88	95	97	100	0

Table 4.2: Invariant moments-based retrieval results for the 1, 100 database images.

Query Nature	$n = 1$ (%)	$n \leq 2$ (%)	$n \leq 5$ (%)	$n \leq 20$ (%)	Not Retrieved (%)
R	85	91	95	98.5	1.5
S	100	100	100	100	0
N	96	99	100	100	0

Table 4.3: Integrated shape-based retrieval results for the 1, 100 database images; n refers to the position of the correct retrieval, R : rotated query, S : scaled query, N : query with a noisy image; The last column indicates percentage of time the query image was not retrieved in the top 20 matches.

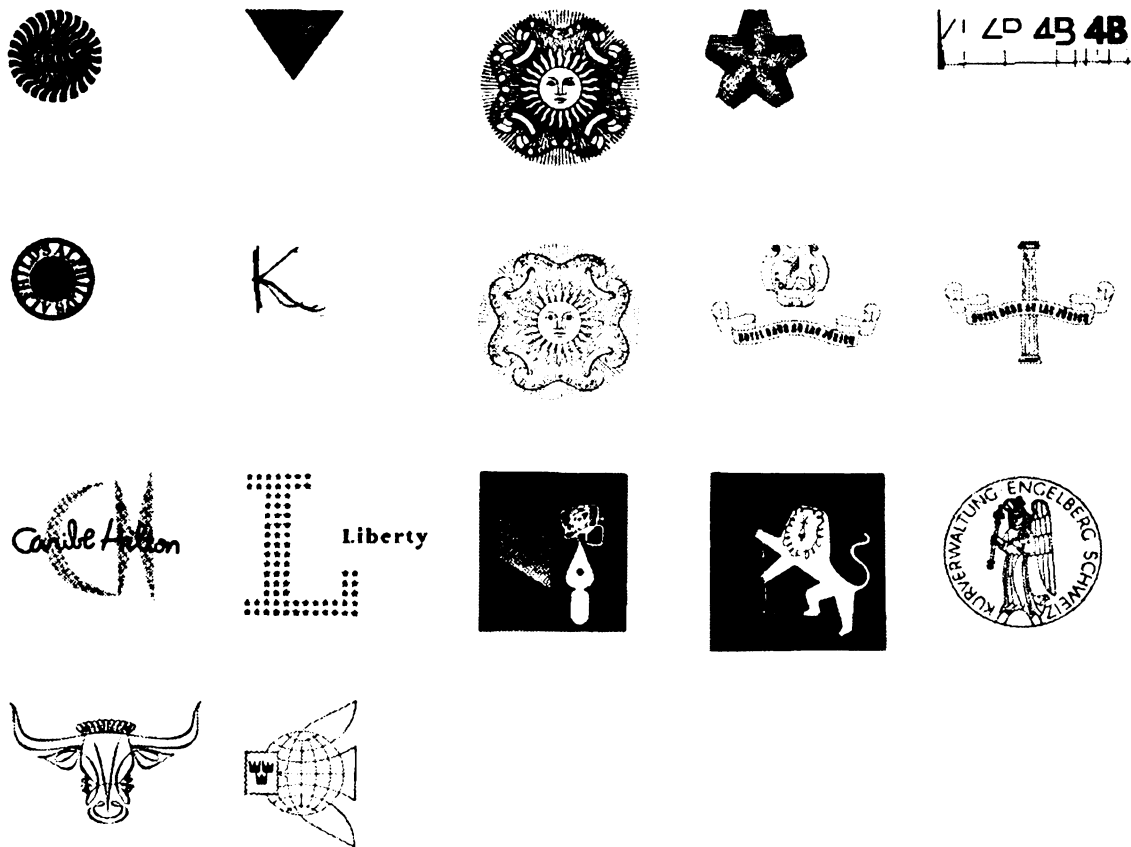


Figure 4.1: Database images not retrieved correctly in the presence of rotation.

4.1.2 Experiments on the Proposed System

The various stages involved in the image retrieval are mentioned below.

- *Preprocessing*: The edge direction and invariant moment features are pre-calculated and stored for all the images in the database. As a result, two feature vectors are associated with each database image.
- *Query image*: A query consists of a hand drawn image of a shape. It can be disconnected, or may contain holes. Fig. 4.2 shows five hand drawn query trademarks used in the experiments.
- *Fast pruning*: The query image is compared to the database images based on the edge direction histogram and invariant moment shape features using the integrated dissimilarity index D_t (Eq. (3.3)). Figures 4.3, 4.4, 4.5, 4.6, and 4.7 show the top 20 retrieved images in the order of increasing dissimilarity for each of the five hand drawn query images. Note that the correct database image that is retrieved has the smallest dissimilarity value for queries involving bear, bull, boomerang, and deer, and the second smallest dissimilarity value in the case of kangaroo query. Although the query image is linearly compared to all the images in the database, the pruning process is still reasonably fast. For each query, this stage takes about 45 seconds on a Sun Sparc 20 workstation (for a database containing 1,100 images).
- *Comprehensive matching*: Under the assumption that all plausible candidates for a query image are contained in the top 10 retrievals in the fast pruning stage,



Figure 4.2: Examples of hand drawn query trademarks.

we apply the deformable matching scheme on these candidates only to further refine the results. The initial pose parameters of the deformable template (position, scale, and orientation) are estimated using the generalized Hough transform. Figures 4.8 and 4.9 illustrate the initial and final configurations of the deformable template match for the bull and boomerang trademarks.

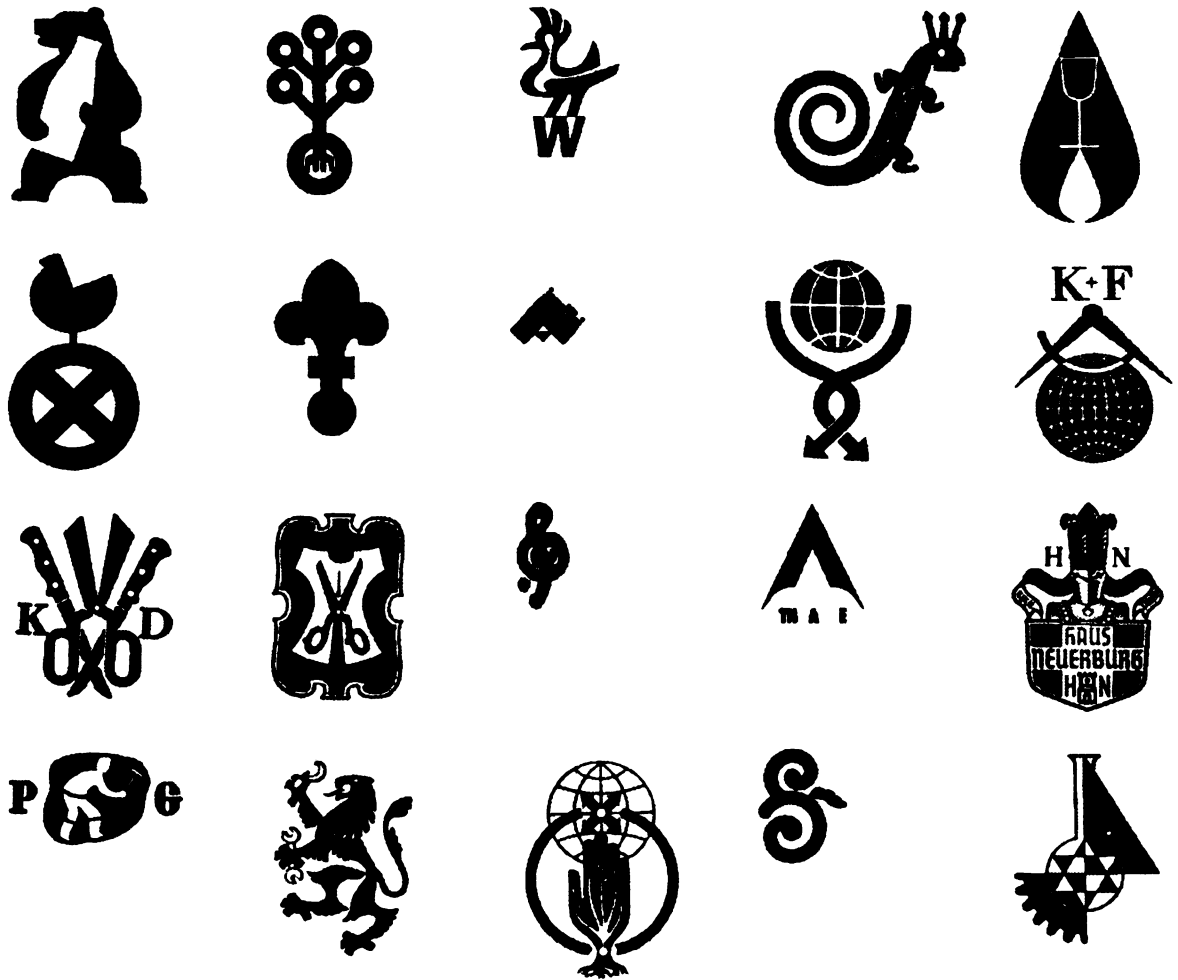


Figure 4.3: Database pruning results for the hand drawn bear sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.

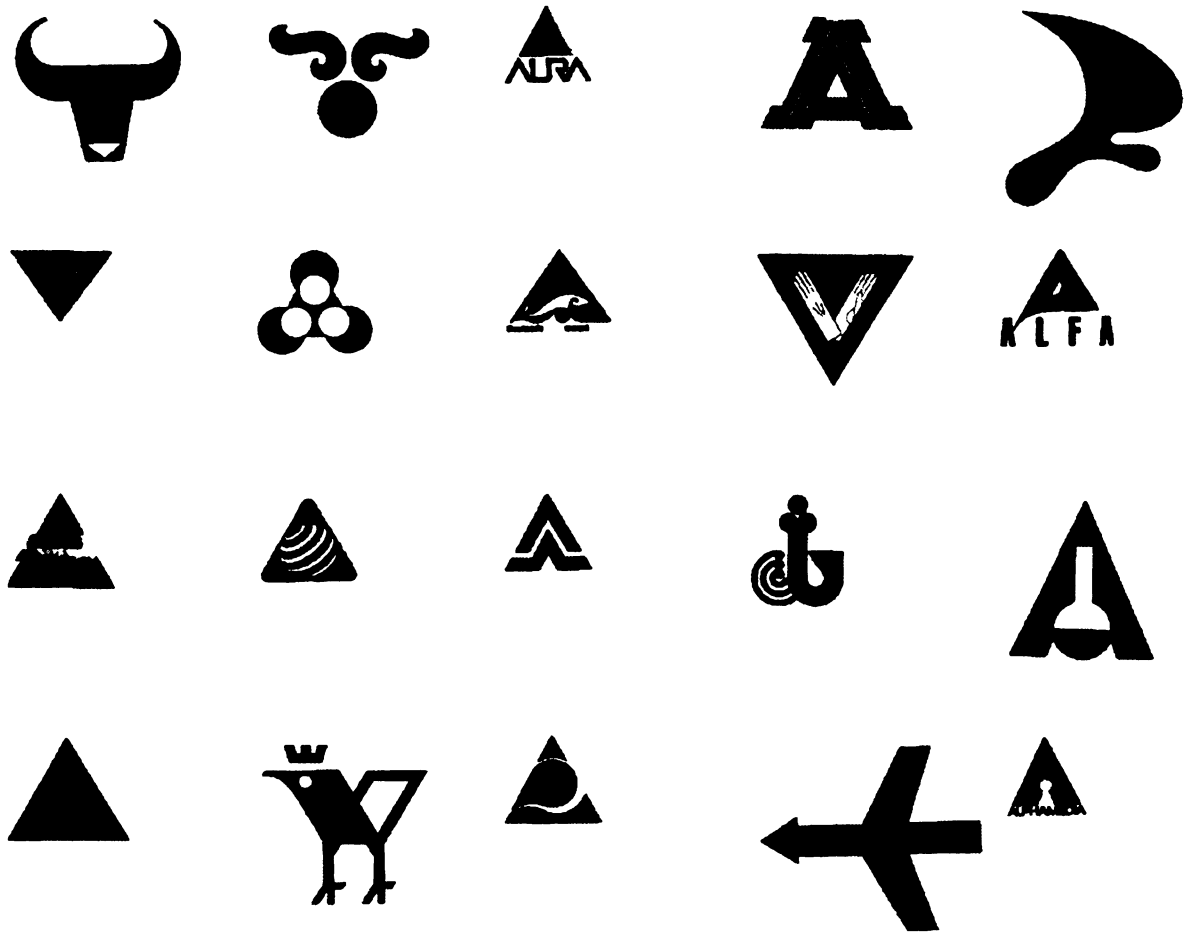


Figure 4.4: Database pruning results for the hand drawn bull sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.

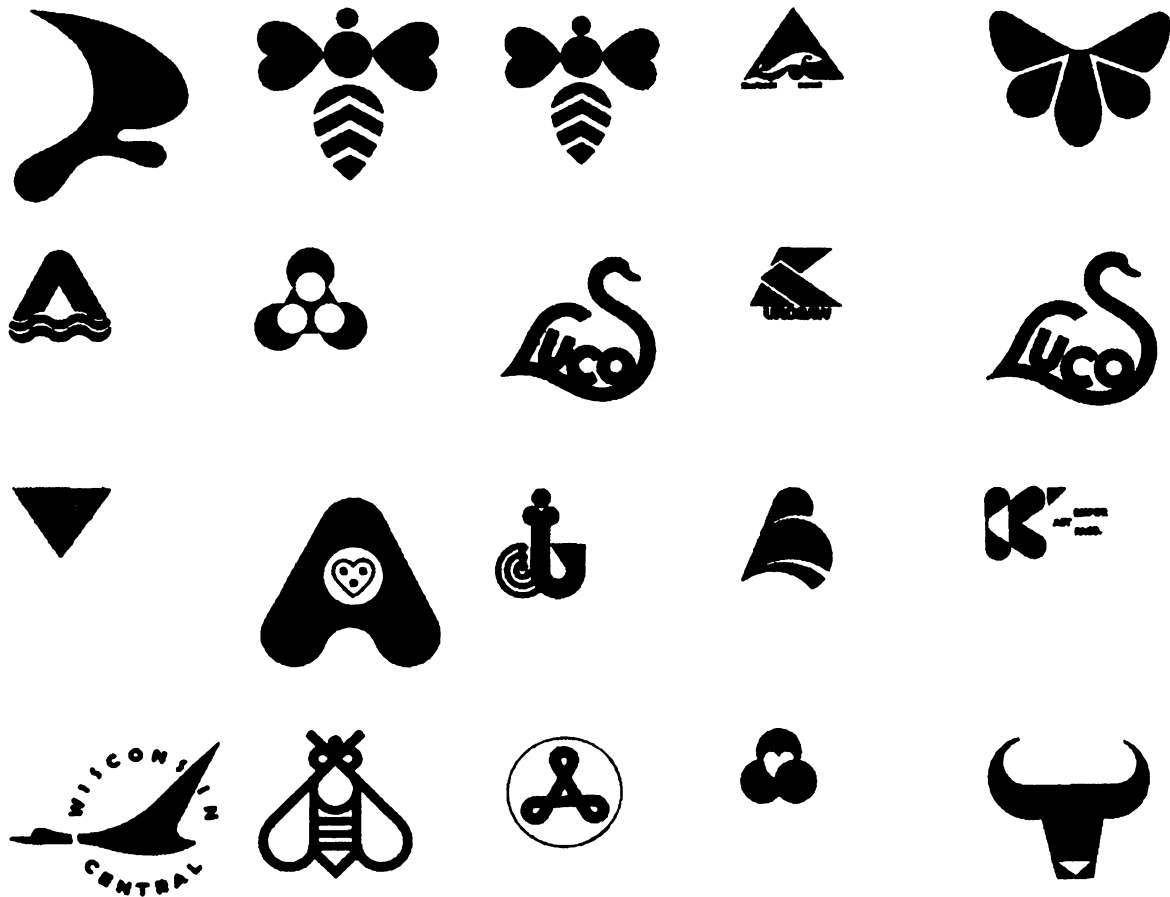


Figure 4.5: Database pruning results for the hand drawn boomerang sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.

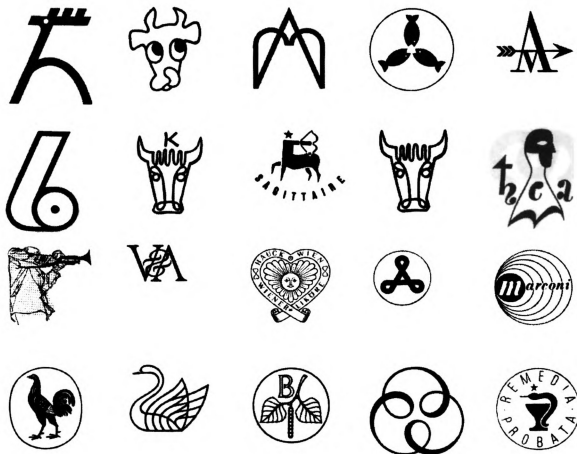


Figure 4.6: Database pruning results for the hand drawn deer sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.

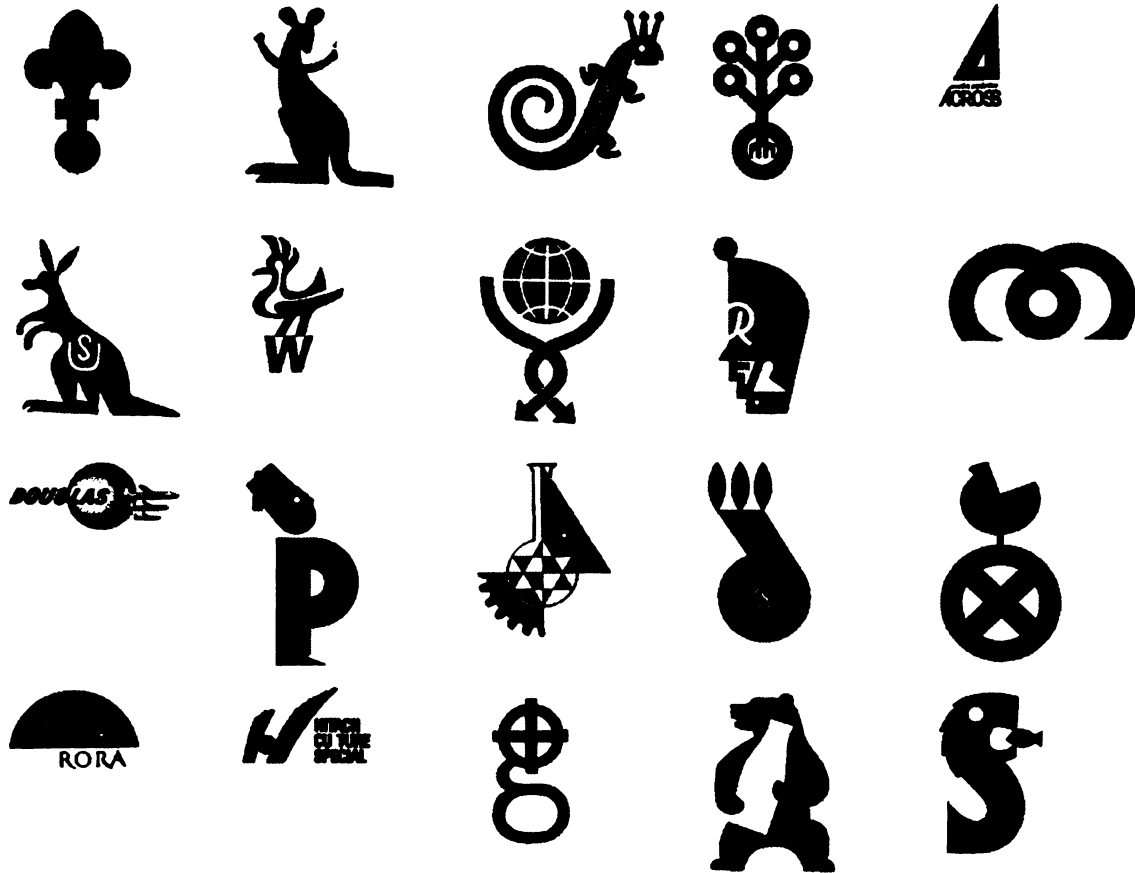


Figure 4.7: Database pruning results for the hand drawn kangaroo sketch shown in Fig. 4.2. The top 20 retrievals are given in the increasing order of dissimilarity.

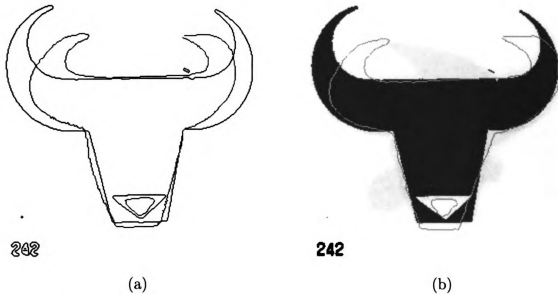


Figure 4.8: Deformable template matching; (a) initial position of the bull template overlaid on the edge map of a bull logo using the generalized Hough transform, (b) final match for the bull logo.

Table 4.4 presents the dissimilarity measures of the five hand drawn logos (Fig. 4.2) to the top 10 retrieved images by the pruning stage. In four out of the five queries, the simple integrated shape dissimilarity index ranks the correct logo in the first place, and in one case, the correct logo is ranked in the second place. The dissimilarity score using the deformable matching ranks the desired images (underlined) in the first place for all the five queries. An incorrect match should result in a large dissimilarity measure (Fig. 4.10) so that the matching hypothesis is rejected. It typically takes 5–8 seconds to calculate the initial configuration using the generalized Hough transform. The iterative deformable matching process takes about 6 seconds on a Sun Sparc 20 workstation.

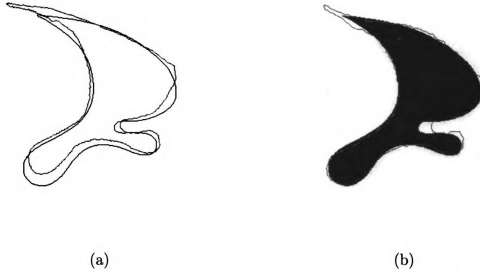


Figure 4.9: Deformable template matching; (a) initial position of the boomerang template overlaid on the edge map of a boomerang logo using the generalized Hough transform, (b) final match for the boomerang logo.

<i>template</i>	<i>Top 10 retrievals from the fast pruning stage</i>									
	1	2	3	4	5	6	7	8	9	10
bull	<u>.149</u>	.670	.959	.856	.847	.862	.803	.784	.820	.913
boomerang	<u>.137</u>	.596	.731	.820	.628	.785	.794	.857	.771	.804
bear	<u>.425</u>	.639	.504	.509	.705	.688	.640	.669	.574	.609
kangaroo	.751	<u>.422</u>	.521	.630	.877	.725	.639	.628	.645	.559
deer	<u>.392</u>	.457	.662	.857	.677	.665	.488	.787	.686	.425

Table 4.4: Dissimilarity values for the five query images when the deformable template matching is applied to the top 10 retrieved images from the fast pruning stage.

4.1.3 Retrieval by Humans

It is instructive to compare the automated retrieval results with those obtained by human subjects. For this purpose, we asked five human subjects to retrieve images from the same database using the five queries. Fig. 4.11 shows the top nine retrievals for the bull query by the human subjects. To find the top ten retrievals for the five query images, it took each subject between 1 to 2 hours. In comparing Figs. 4.4



Figure 4.10: Deformable template matching; final match for a bull template and a boomerang image.

and 4.11, we make the following observations. The top most retrieval for the bull query and the human respondents is the same. In fact, the top most retrievals for all the five queries are consistent for both the human respondents and the proposed algorithm. This is expected since the hand drawn queries closely resemble one of the database images. As there is a substantial amount of diversity in the image database, only two to three (out of the ten) top retrievals are in common between the outputs of the algorithm and human respondents for each query. We note that for all the queries, the retrievals obtained by the five respondents are somewhat consistent for the following reasons: (i) human subjects can easily decide the foreground greyscale of the object, no matter whether it is 0 or 255, and (ii) human subjects tend to abstract the query image for some conceptual information. As an example, for the bull query, most human respondents retrieved all the trademark images in the database which contain a bull head, even though the shape of the bull in the

retrieved images is quite different from the query shape. Although, we (humans) can extract other images with similar semantic content (like bull head in this example), we lack the ability to match objects at different orientations. The proposed system retrieves images invariant to orientation, and hence, picks up images that resemble a triangle though at arbitrary orientations. These observations explain the difference between the retrievals by human subject and the proposed algorithm. We believe it is a challenge to develop an automatic retrieval algorithm which models the human perception well. However, considering the time and effort for humans to perform such a retrieval task, an automatic retrieval algorithm can provide a powerful tool to facilitate the retrieval task for human beings.

4.1.4 Retrieval using Additional Heuristics

The performance of the proposed system does not approach that of humans. We can see that simple shape features do not capture the semantic content of an image. A concise theory on how to automatically extract the semantics of an image from its content has not been developed so far. At best, with the current state-of-the-art knowledge, we can propose certain heuristics that improve the retrieval accuracy of the system.

Image Filling

Matching of trademarks is based more on the general global shape of the marks rather than on fine details. Referring to Figs. 4.4 and 4.11, we see that while querying for the filled bull image, the system tends to retrieve images that were filled, rather than the



5



4



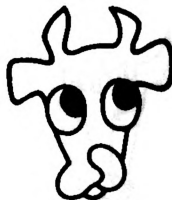
4



4



3



3



3



3



2

Figure 4.11: Nine topmost retrieved trademark images by the human subjects for the bull template. The number below each retrieved image is the number of human respondents (out of 5) who placed the image in the top 10 retrievals.

bull images made up of line drawings. This suggests that the object outline itself can be used for extracting the semantic content in the images. It may thus seem reasonable to extract the outline of objects within the images and extract edge direction- and moment-based features on the outlines. We tried this heuristic but found that the moments and the edge directions did not yield good results when applied to the few outline points only. Thus, instead of using just the outline, we filled in the objects and based our features on the filled-in images. Figures 4.12 and 4.13 present five bull images in the database, and their corresponding filled-in images.

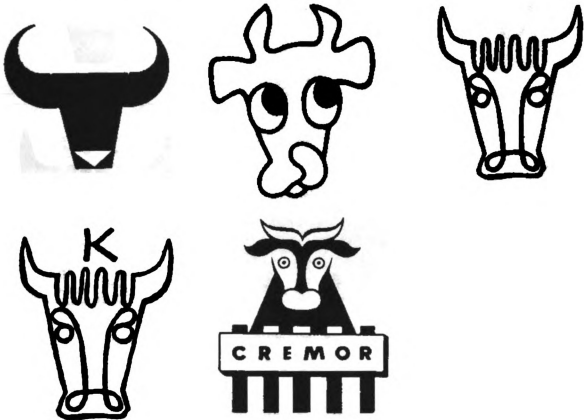


Figure 4.12: Five images of a bull present in the database.

We have conducted experiments to match the filled-in query images against the filled-in images from the database. The aim of these experiments was to study the

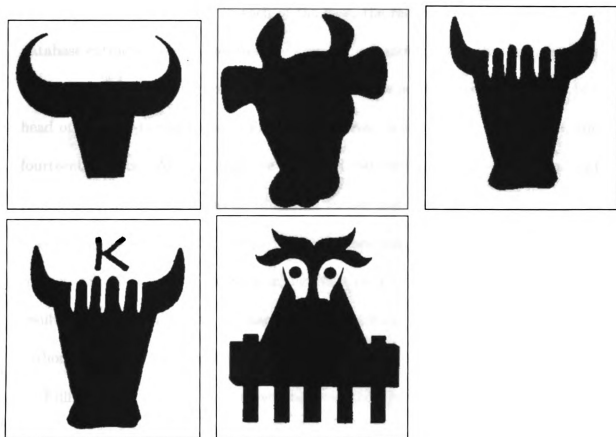


Figure 4.13: Filled-in images for the five bull images shown in Figure 4.12.

effect of using only gross details in the images (filling in leads to a more general representation of the images in the database) on the efficiency and accuracy of the fast pruning stage. Figures 4.14, 4.16, 4.18, 4.20, 4.21, 4.22, 4.23, 4.24 show retrieval results for eight query images (a bull image, a bear image, a kangaroo image, a set of bars, a bee image, a Sun image, a mark consisting of text (word mark), and a lion image). It can be seen that in each of the case, the retrieval based on the filled-in database extracts more images that have similar semantic content. In the case of the query on a bull image (figure 4.14), the system extracts four other images of a bull head other than the query one. These are retrieved in the second, third, sixth, and fourteenth places. We note that the bull head retrieved in the third place was not retrieved by the human subjects. Similarly, we see that with the bear and kangaroo queries (figures 4.16 and 4.18), we retrieve two more images of a bear and a kangaroo, respectively. As a basis of comparison, figures 4.15, 4.17, and 4.19 show the retrieval results when the bull, bear and kangaroo images from the database were retrieved without filling the holes in the images.

Filling in the holes in the image improves the efficiency of a match making it more robust, since the finer details are neglected (in the case of edge directions, holes contribute to the edge direction histograms and in the case of invariant moments the moments are evaluated over larger number of points improving their robustness). With the proposed filling in, we can now match images which are either just outlines or are regions. A major drawback of the scheme though, is the disregard for the information contained in the holes. Many a times, certain marks contain useful information in their holes. Figure 4.25 shows two database images where the holes

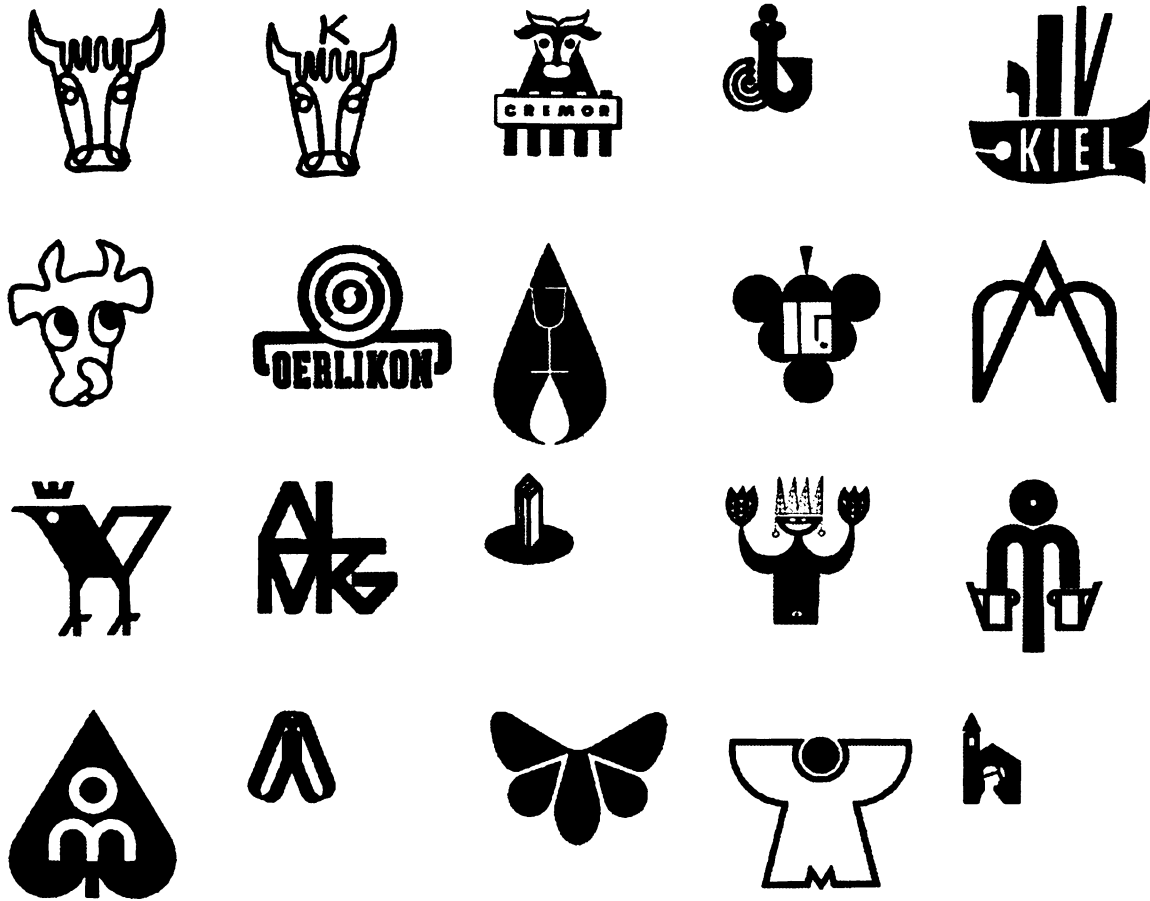


Figure 4.14: Database pruning results for a bull image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.

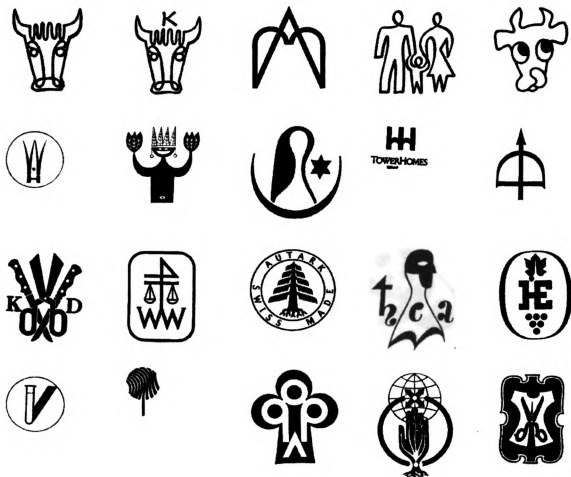


Figure 4.15: Database pruning results for a bull image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity. Results are based on unfilled images.

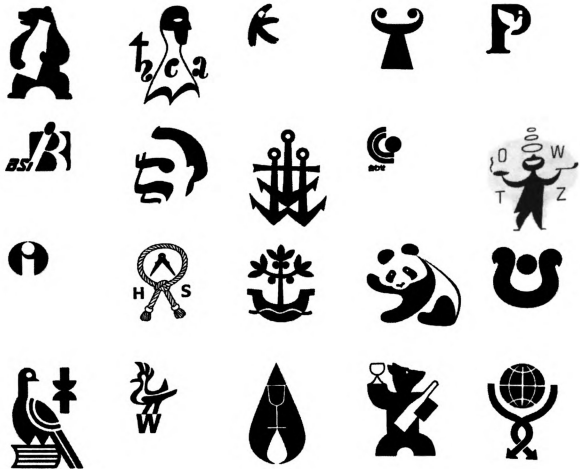


Figure 4.16: Database pruning results for a bear image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.

Figure 4.17: Database pruning results for a bear image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity. Results are based on unfilled images.

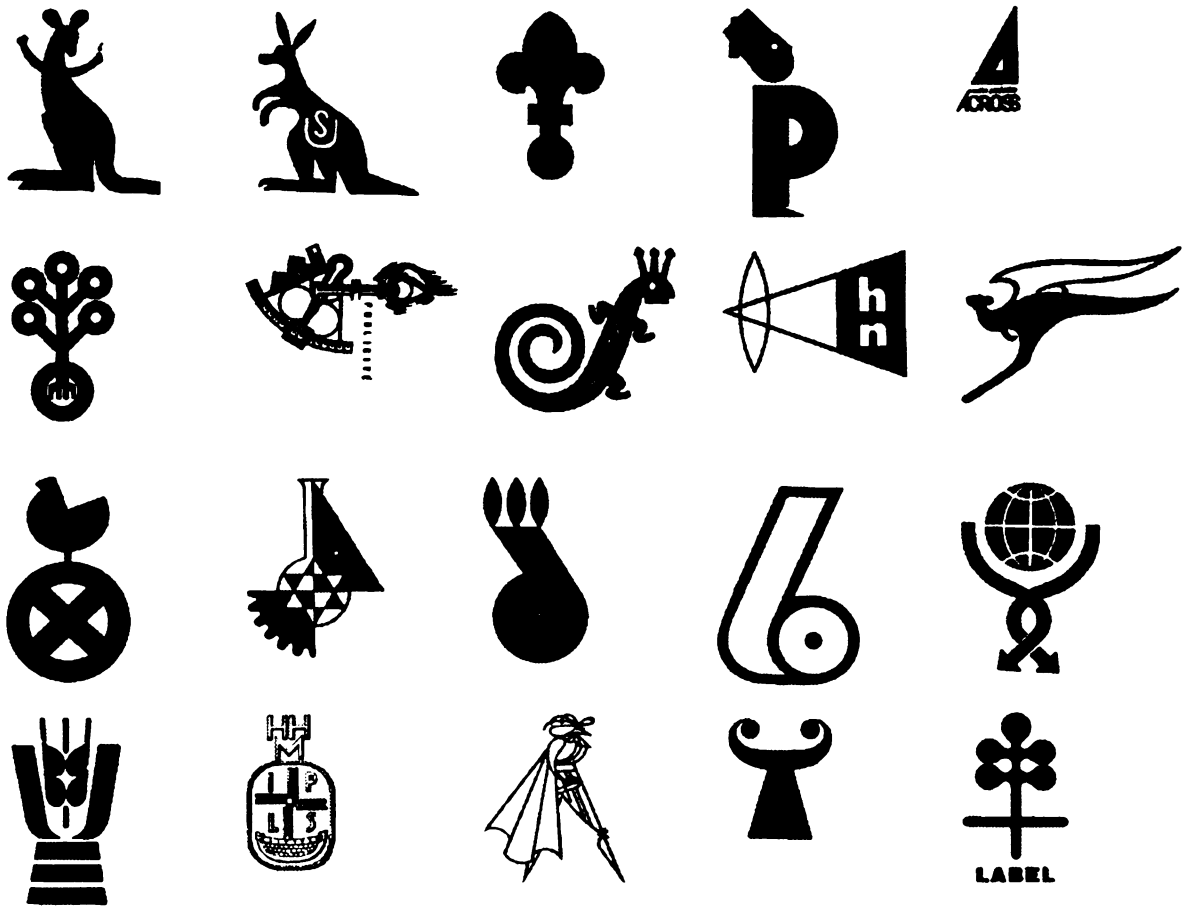


Figure 4.18: Database pruning results for a kangaroo image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.

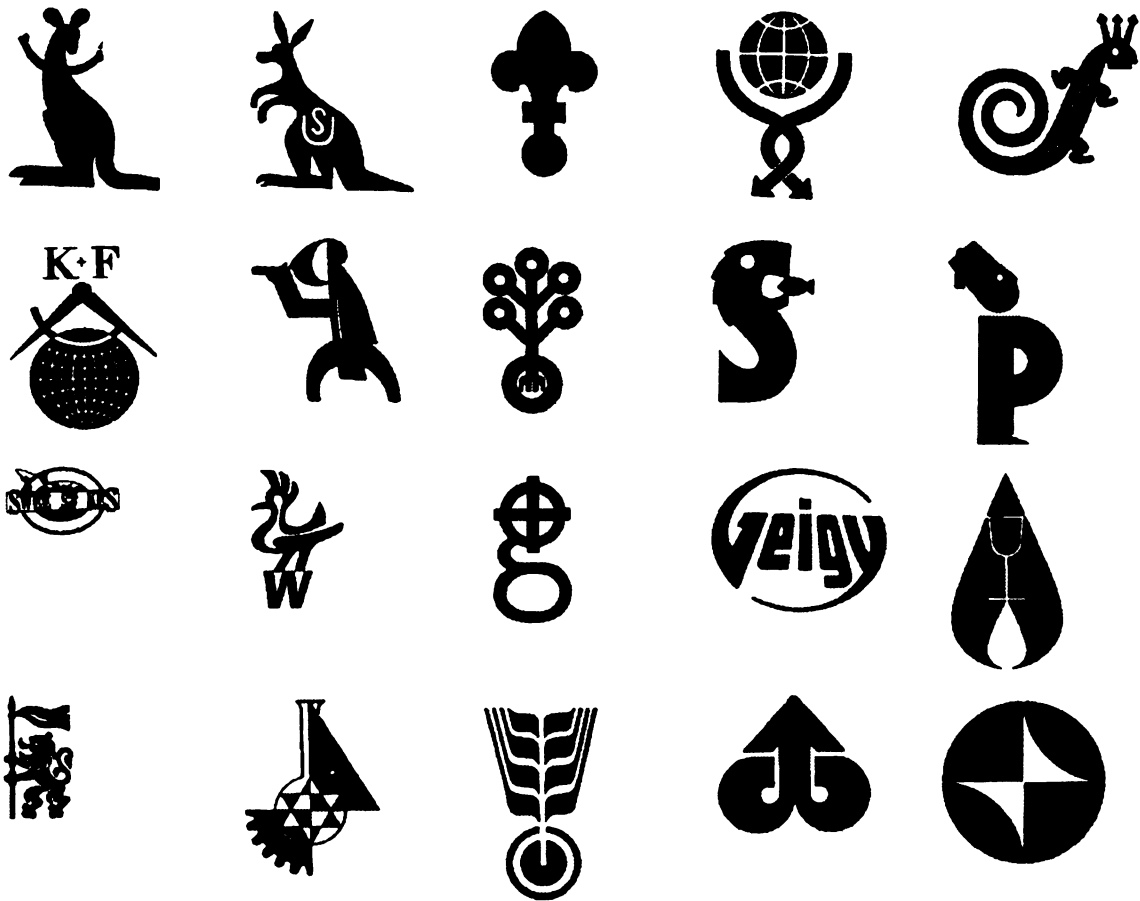


Figure 4.19: Database pruning results for a kangaroo image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity. Results are based on unfilled images.

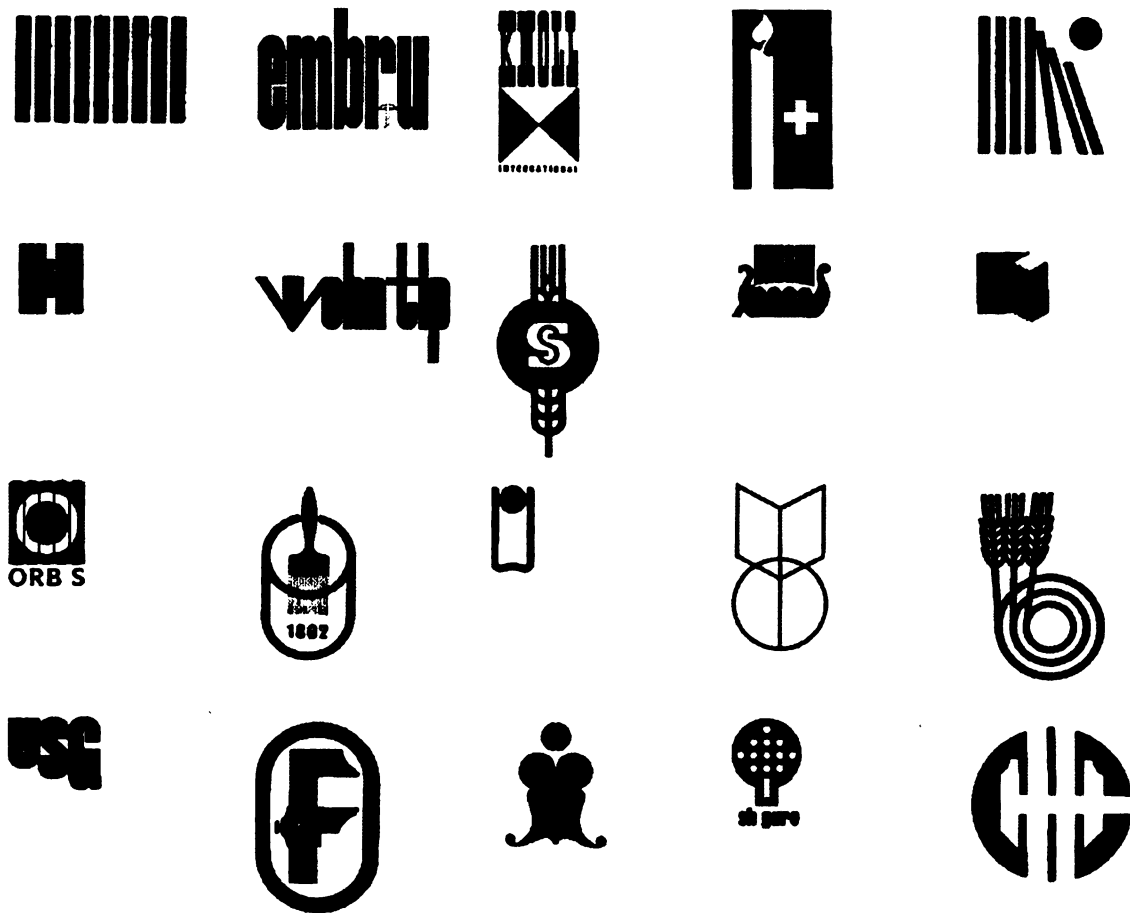


Figure 4.20: Database pruning results for an image consisting a set of vertical bars (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.



Figure 4.21: Database pruning results for a bee image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.

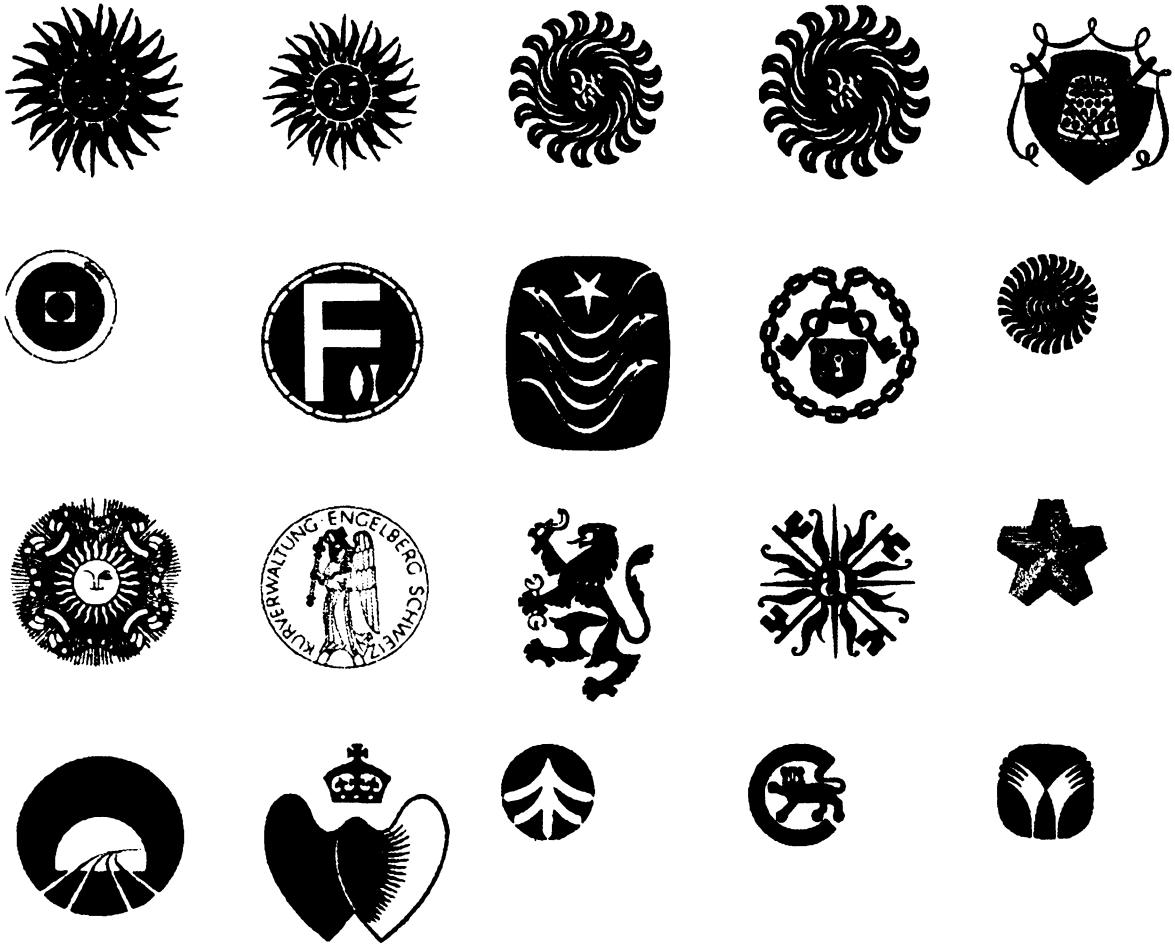


Figure 4.22: Database pruning results for a Sun image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.

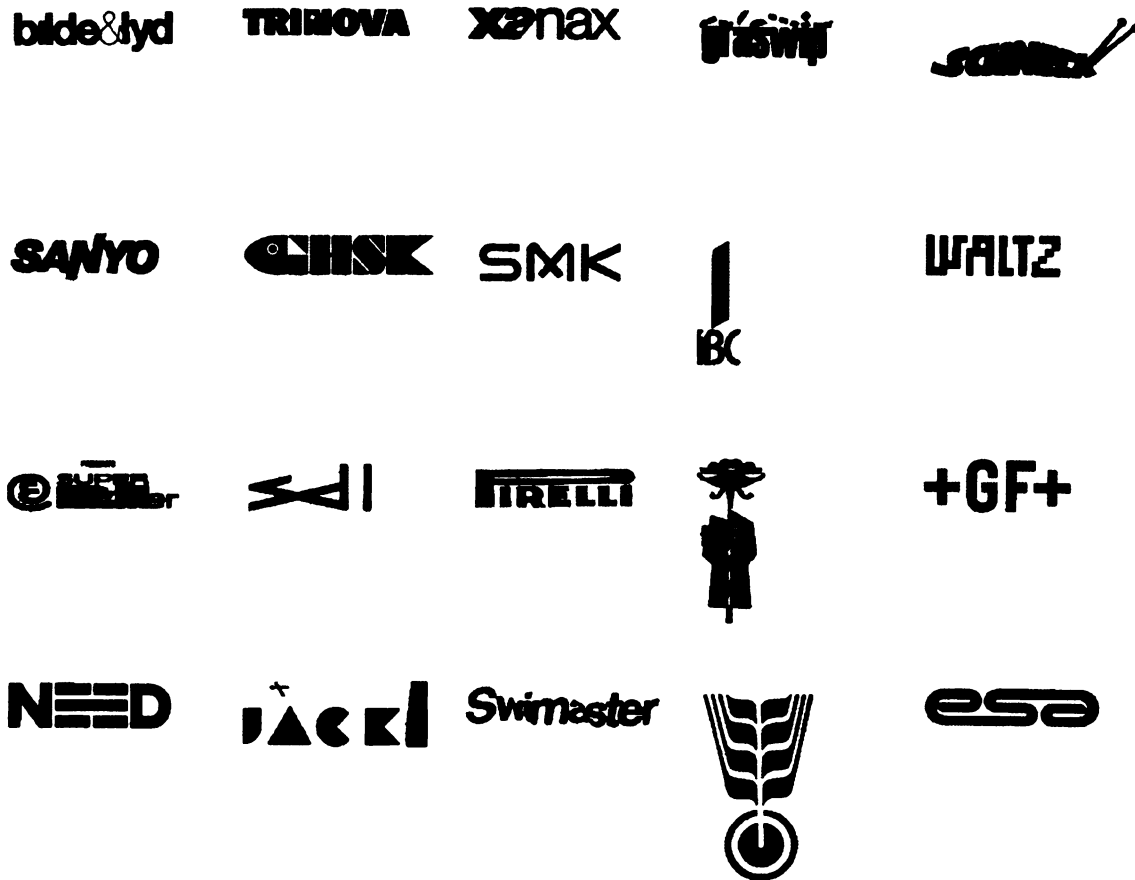


Figure 4.23: Database pruning results for a word mark (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.

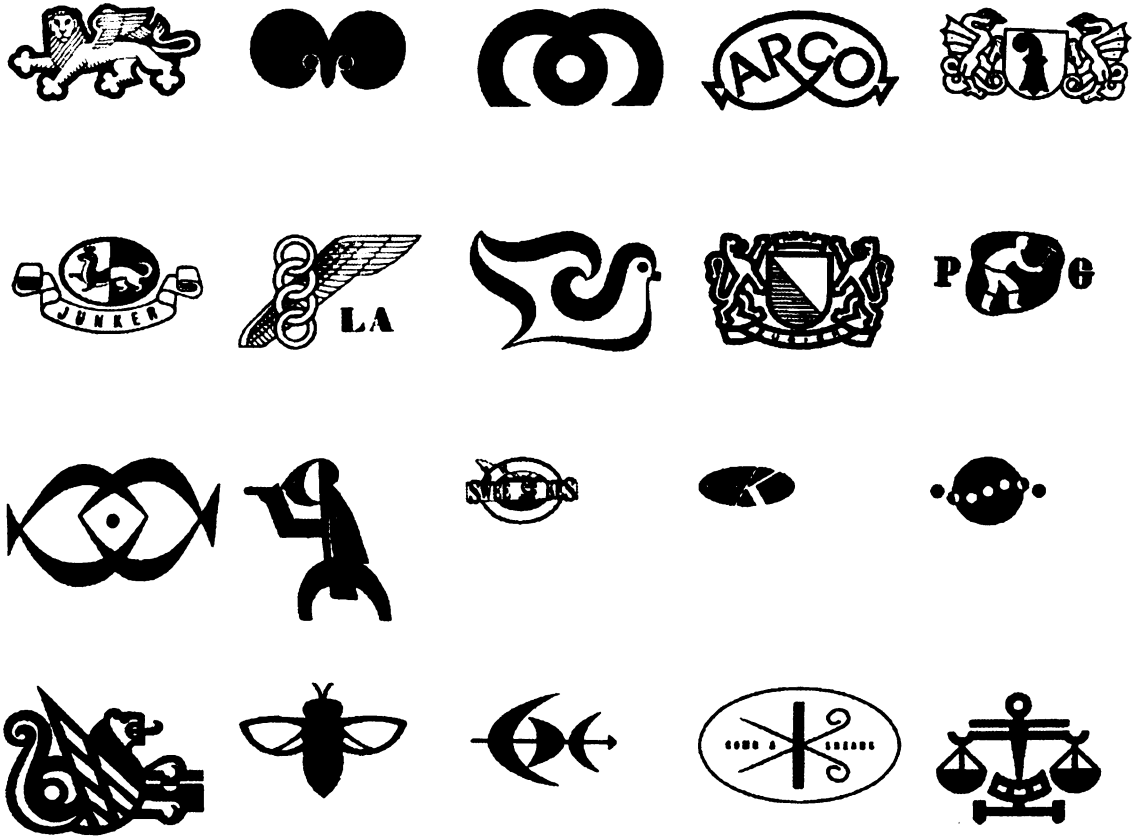


Figure 4.24: Database pruning results for a lion image (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.

contain significant information. Note that, for the sake of clarity, the holes are marked in white where as the foreground for these trademarks is black. When filling in the database images, these holes are neglected, leading to false matches (here matching is based on squares). A human can very easily identify the foreground and background in an image and extract various objects from the image. Our proposed system is not able to do so automatically. We believe that if the system is presented with segmented objects from images, then it can perform matching based on these objects. At present, the system extracts images similar to the entire query image, rather than on the basis of the objects present there in. Figures 4.26 and 4.27 present the retrieval results based on queries containing images with significant information in the holes. In the case of figure 4.26 all images with a square border are retrieved, whereas in figure 4.27 all images that have a circular border are retrieved.

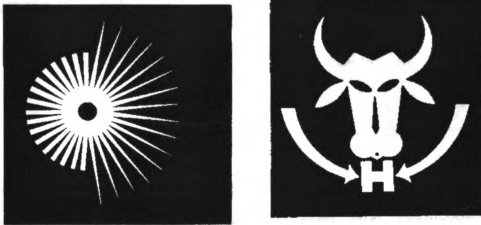


Figure 4.25: Two images containing a square border; information contained in the holes (white) is neglected when filling in the database.

It is extremely difficult to quantify the results of the system in terms of the actual recall or precision rates. The *precision rate* can be defined as the percent of retrieved



Figure 4.26: Database pruning results for an image with a square border (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.

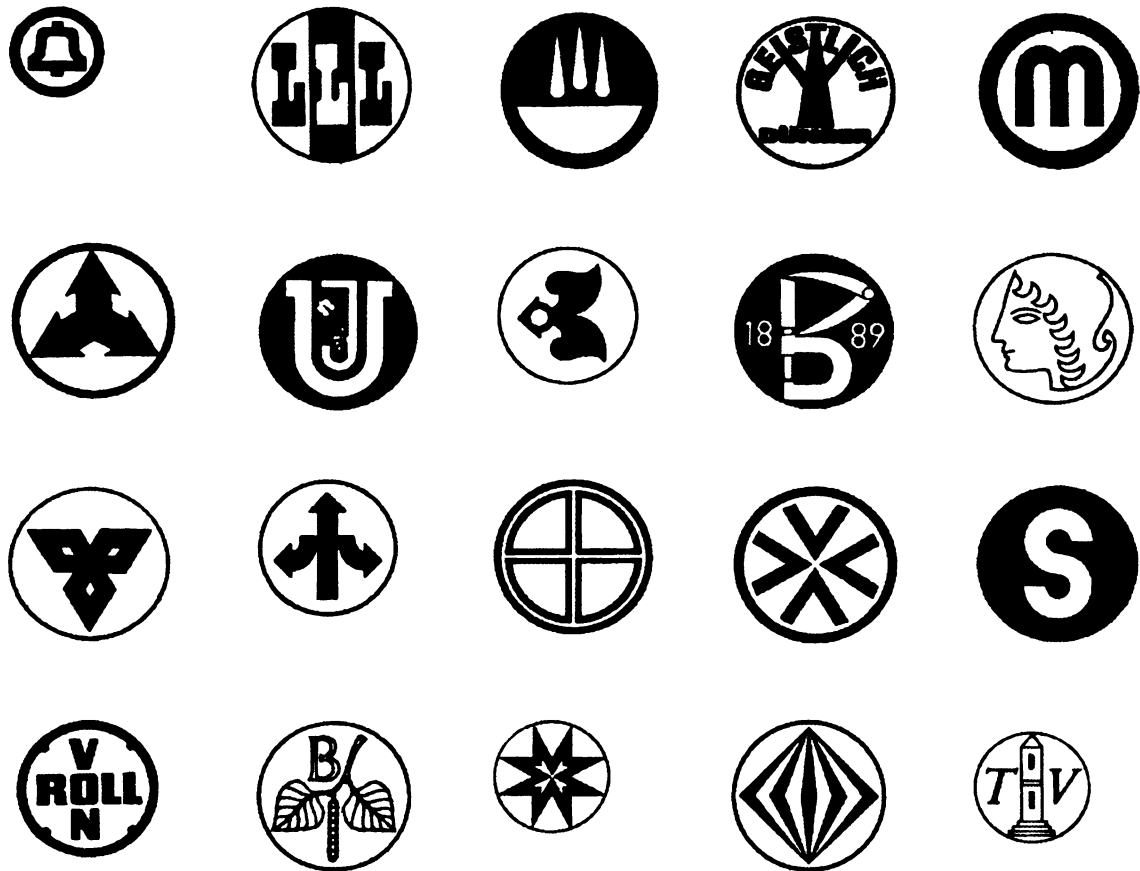


Figure 4.27: Database pruning results for an image with a circular border (query image corresponds to the top most retrieval). The top 20 retrievals are given in the increasing order of dissimilarity.

images *similar* to the query among the total number of retrieved images. A *recall rate* is defined as the percent of retrieved images which are *similar* to the query among the total number of images *similar* to the query in the database. It can be easily seen that both precision and recall rates are a function of the total number of retrieved images. In order to have a high accuracy, the system needs to have both a high precision and a high recall rate. A major drawback in quantifying the precision and recall rates is to identify images that are *similar* to the query image. We have found that it is extremely difficult and time consuming for human subjects to identify all images in the database that are *similar* to a query image. We have defined *similarity* to be perceptual similarity. Thus, we consider two images containing a bear as *similar*. We compute the precision and recall rates using this definition of similarity. Table 4.5 presents the recall and precision rates for three queries as shown in figures 4.14, 4.16, and 4.18. We have been able to identify 10 images of a bullhead, 4 images of a bear, and 3 images of a kangaroo in the database. The precision and recall rates are based on these values. A more detailed study is needed to have a better understanding of the system performance in terms of the precision and recall rates.

Query	n_1	n_2	Recall Rate	Precision Rate
<i>Bullhead</i>	4	10	40%	20%
<i>Bear</i>	3	4	75%	20%
<i>Kangaroo</i>	3	3	100%	15%

Table 4.5: Recall and precision rates for three queries (shown in figures 4.14, 4.16, and 4.18) based on the image filling; n_1 refers to the number of images retrieved in the top 20 positions that are *similar* to the query, n_2 refers to the number of images in the database that are *similar* to the query. Recall rate is (n_1/n_2) , precision rate is $(n_1/20)$.

Object Segmentation

Another major issue in image retrieval is the issue of retrieval based on multiple parts or components in the image. A trademark can consist of multiple objects and it may be necessary to query the objects rather than the images. Such a scheme would require a priori segmentation of objects from the database images. We have tried various heuristics to segment object entities from an image. An object entity may be defined as a single component or multiple components that can denote a single entity in the world. Extraction of object entities from binary images can be a very difficult problem. Quite often an object entity is made up of multiple components. Many researchers are currently working on the problem of perceptual grouping for object segmentation. We use a very simple heuristic to extract object entities from images. The heuristic extracts single components in an edge image and combines objects that are similar in size. These objects then represent an object entity in the image. Figures 4.28, 4.29, and 4.30 show six database images and the extracted object entities based on the heuristic described above. Note that the objects are filled to extract the general shape and neglect the details. Holes are also extracted as objects. Since the components are extracted from edge images.

The proposed heuristic extracts objects that humans perceive as a single entity, although it may consist of multiple components. But, this heuristic fails in case of objects having a similar size but constituting different entities (in Figure 4.30(i), the bird is mixed in the first object entity). Spatial proximity may be applied to overcome this problem. Another drawback of the scheme is that a large number of noisy

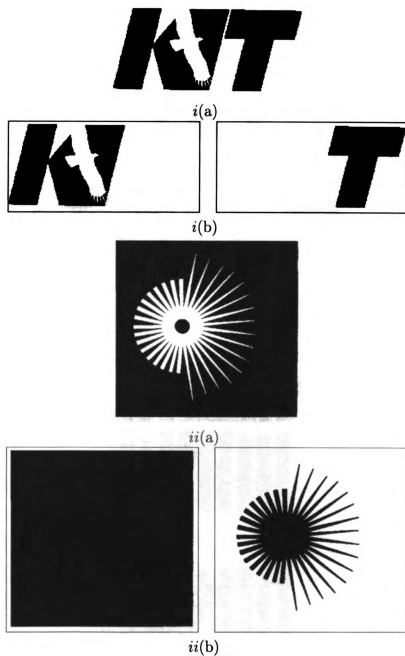


Figure 4.28: Segmentation of image into object entities; (a) a trademark image; (b) segmented objects.

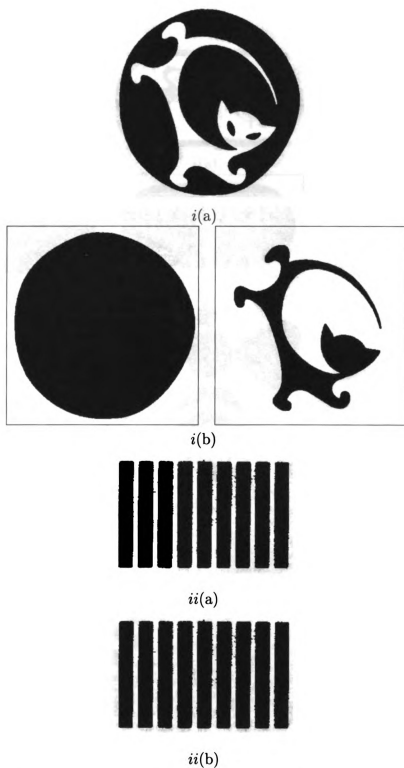


Figure 4.29: Segmentation of image into object entities; (a) a trademark image; (b) segmented objects.

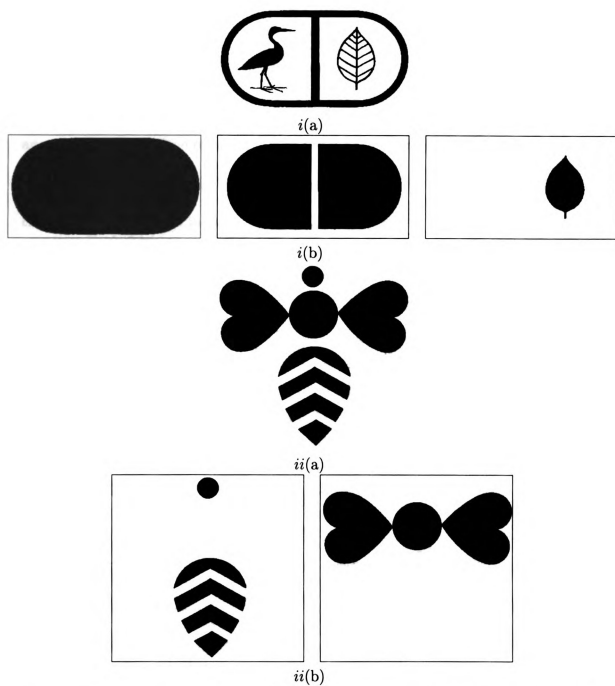


Figure 4.30: Segmentation of image into object entities; (a) a trademark image; (b) segmented objects.

objects are also extracted from the images. Figure 4.30(ii) shows how components representing a single entity can sometimes be divided into two objects. The threshold parameters used to merge components would need to be tweaked for every image. Thus, during archival of the database, this heuristic can be used to extract individual entities from the images with the user tweaking the thresholds for each of the images (this needs to be done once only during archival). During retrieval, the thresholds may be tweaked to select individual entities from the query image. The retrieval can then be based on these individual objects. As future work, we plan to develop this framework for matching on the basis of objects rather than images.

4.2 A Case Study in Norwegian Silver

Authentication

Norwegian silverware from the middle of 1500, represents an important part of Norway's cultural heritage. Silver objects especially from the year 1568 until the beginning of 1700 are highly prized because of their advanced craftsmanship. In this period the Norwegian silversmiths work had attained an artistic quality that was comparable with the best international piece of the time. These objects continue to fetch enormous prices at major auction houses around the world. Museums, private art collectors and antique dealers are the customers, and the price they are willing to pay usually depends upon when the silver object was produced and who made it. To *date* and *attribute* silver is, however, not a simple task. An expert on silver considers

a number of different criteria. The three most important are: (1) the marks which are usually stamped on the silver objects, (2) the form of the object, and (3) the style of decoration. Most silverware is stamped with marks, and the oldest types of marks are the *makers mark* (also called *master-mark*), and a mark representing the city (city-marks) of production. There are about 6,000 different known master-marks and 250 city-marks for Norwegian silver, so a manual identification of a mark takes a long time. However, the identification of a master is not always sufficient to do a proper dating of an object. A master-mark may have been used in more than one generation, e.g., by the master's son and grandson. In such situations, a stylistic evaluation of the decoration of the object is needed. An expert usually studies the form and decoration of the object and compares the style with the styles of similar silver objects. This process is also very time consuming and requires a careful study of the form and decoration of the object. This study is part of a project to develop a computer-based system that can assist in dating and attributing of Norwegian silverware [51, 52].

Our project is divided in two major parts. The first part is aimed at developing a system to recognize the different types of marks which are stamped on the tankards. The second part is the design of an interactive system which enables the user to accomplish a proper investigation of the style of decoration, based on a comparison with other objects guided by the expert. The major effort required here will be to develop a graphical user interface and to structure the expert's knowledge accordingly. While the first part will be mostly automatic, a high degree of user interaction will be needed in the second part.

Our system will work on images taken by a Kontron - ProgRes 3012 digital color

camera, at a resolution of 3400×2320 . The camera has a dynamic range of 12 bits.

4.2.1 Background

Stamping of silver guarantees the pureness of Sterling's worth. As early as in 1314, Norwegian king Haakon V appointed two persons to monitor that the silversmiths followed the existing rules. From the year 1384, the makers were obliged to stamp their own mark (master-mark) on the object, and in 1491 a city-mark with a crown was made compulsory.

The master-mark is extremely important for authenticating and dating an object. The same master can have many different marks, and some silversmiths applied these different marks at the same time, while others used them in a chronological order. The earliest of the master-marks are often trademarks, but usually they are composed of a silversmith's initials written, either side by side, or more or less intertwined (see Fig. 4.32). Some marks are easy to read, while others are almost unreadable. The shape of the frame enclosing master-marks can vary: some are rectangular, square, oval, heart shaped, a cartouch, etc., often depending on the fashion of the style at the time. If the silversmith's name is made up of two syllables, like for instance Hornemand (Horne - mand), he may well use the initial HM instead of H only. If his first name is Isach, he can let the first line in the letter H represent the letter I in Isach. There are numerous other possibilities. In addition to his initials, the silversmith sometimes added a date, that usually referred to the time he acquired citizenship, and not to the year when the object was produced.

The city-mark is usually composed of the city's initial with a crown. Two cities, like for instance Skien and Stavanger, have the same initials, which makes it difficult to differentiate between them. In 1540 a law was passed in Copenhagen - Norway was in a union with Denmark at that time - that the chairman of the guild should personally stamp this particular mark. At that time, there were no organized guilds in Norway, except for the ones that were a part of the Hanseatic League in Bergen. The first Norwegian guild was established in Bergen in 1568, and fifty years later a guild was established in Oslo. From that time, we find a certain date written on the city-marks which refers, most probably, to the time when a new chairman of the guild was appointed. However, it seems obvious that the silversmiths themselves and not the chairman stamped these marks, because they differ from silversmith to silversmith. Some of the silversmiths in Vestfold and Telemark used a mark with the initials CL, which refers to the Christiania Guild.

Besides the master-mark and the city-marks, several other marks were also applied from 1740 onwards: the assay-mark, year-mark, month-mark, and the standard-mark. Before 1719, when an official assayer was appointed, we rarely find the hallmark on the silverwork. But, from this time on the mark of the assayer can be recognized in known order. After 1781, the year-mark is separated from the city-marks, usually limited to the two first ciphers, although during the decade 1800-1810 the complete date is given. The month-mark is in the form of a fraction until 1765 and after 1821, but between these years the signs of the Zodiac are used. The standard-mark on silver applied after 1840 is $13\frac{1}{3}$ L and is often used instead of the assay-mark [53].

4.2.2 Database Search for Master-marks and City-marks

There are about 6,000 known master-marks and 250 city-marks. Given a query mark, a manual search through this database is demanding and takes a very long time. An image retrieval system can do this job much faster and more reliably [5]. Our goal is, therefore, to develop an efficient retrieval of images of master- and city-marks from large databases based on the shape content of the query image. We will build an automatic retrieval system to browse through the entire database. The first step is to extract primitive visual features from the images and to retrieve database images on the basis of these features. So far, we have developed two prototype image databases; one with master-marks and another with city-marks both obtained from a catalog. We have scanned about 300 marks using an HP Scanjet IICx flatbed scanner. The eventual goal is to build an image retrieval system which is insensitive to large variations in image scale, rotation, translation and occlusion. Even if the query image differs from its stored representation in the database in its orientation, position, or size, the image retrieval system should be able to correctly match the query image with its prototype in the database. Figure 4.31 shows a block diagram of the recognition part of the system.

The prototype databases are divided into a Renaissance subset and a Baroque subset. The first step is, therefore, to establish whether the form of the input image (the unknown tankard) belongs to one of these two periods. The next step is the preprocessing of the master- and city-mark input images. The images in the prototype databases have been preprocessed to extract shape features which have been stored

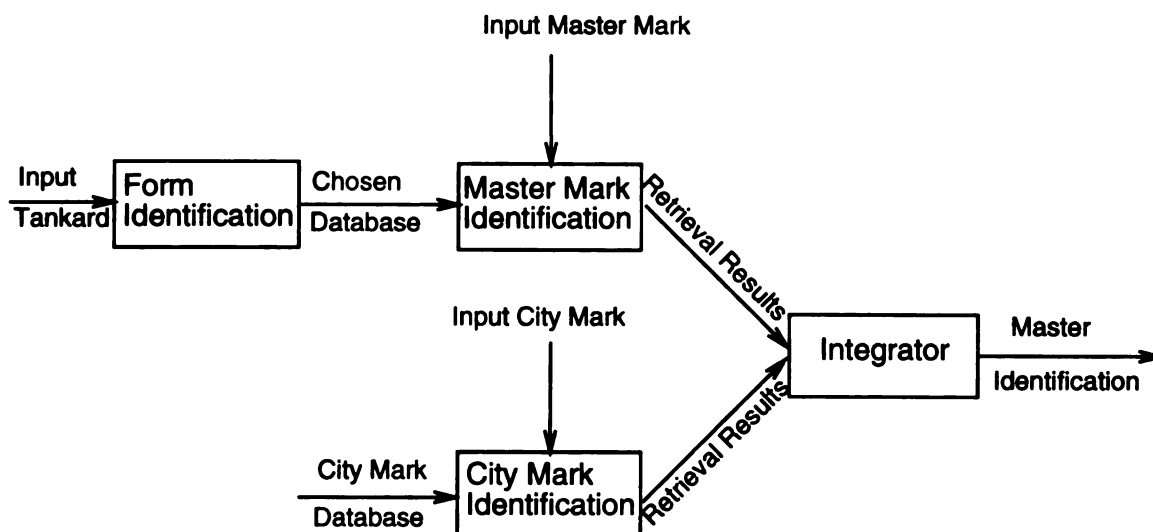


Figure 4.31: A block diagram of Master identification based on master and city-marks.

along with the images in the database. When a master-or a city-marks image of an unknown tankard is presented to the system, it has to be preprocessed in the same way to extract features similar to the features which are stored (as a feature vector) in the databases. The query image is then matched with the feature vectors present in the database. There is an integration module, where the system integrates the retrieved results based on the master- and city-marks. The output is a ranked set of retrieved images with high matching scores. If one of the retrieved images is similar to the master-mark on the unknown tankard, the user accepts the automated search and the identification of the goldsmith comes up on the screen. If none is acceptable, then a new search starts.

We use the integration of the edge direction histograms and the invariant moments as the basis of the retrievals. Please refer to Chapter 3 and Section 4.1. A mark can sometimes be worn down and almost invisible. Some image processing techniques are needed to enhance and restore such images.

4.2.3 Preprocessing

Experiments were conducted on a master-mark database of 171 marks. These marks were scanned from the books [54, 53] using a HP Scanjet IIcx flatbed scanner. Figure 4.32 shows 10 of the 171 master-marks stored in the database.

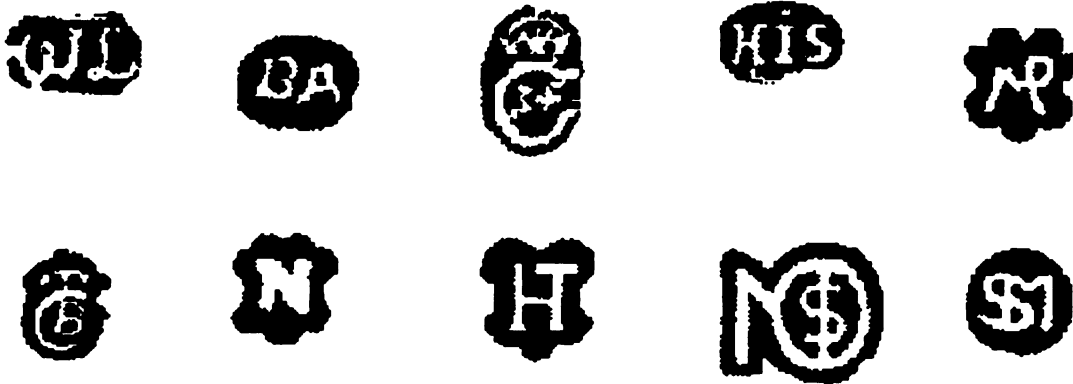


Figure 4.32: A representative set of master-marks in the database.

Real master-marks present at the bottom of several silver tankards were digitized using the Kontron-ProgRes3012 camera and presented as query images. Unlike the database images, the master-marks captured from the tankards are in color and rather noisy. In addition, due to the age of the tankards, many of these marks are faded. Figure 4.33 shows a set of real master-marks extracted from the bottom of various tankards.

We require a preprocessing stage to binarize and clean the noise in the master-marks shown in Figure 4.33. The color images are first converted into greyscale images. The greyscale images are converted to binary images using a global thresholding scheme. A second stage of preprocessing is then applied to the binary image for noise cleaning. Noise in the binary images is usually in the form of speckles. These

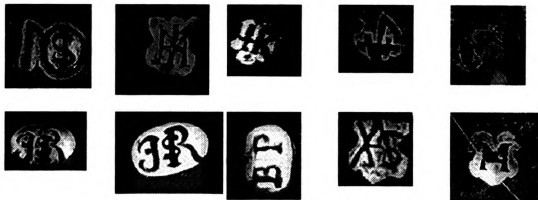


Figure 4.33: A representative set of real master-marks extracted from the tankards.

speckles are filtered out using a connected component analysis, where all components of a size less than a threshold are removed. The filtering allows for removal of both white as well as black components. Figure 4.34 shows an input greyscale master-mark with the corresponding thresholded binary and cleaned image.

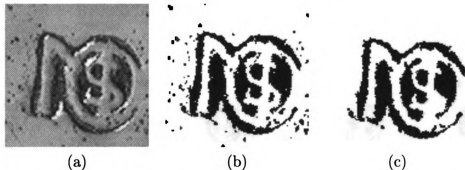


Figure 4.34: Preprocessing an input master-mark; (a) Greyscale master-mark; (b) Thresholded (binary) master-mark; (c) Binary master-mark after noise cleaning.

4.2.4 Experimental Results

The cleaned and binarized master-marks were presented as input queries to the retrieval system. The system extracts the shape features for the query images and matches these with those stored in the database. A set of top 10 matches

based on the integrated dissimilarity index are presented as the output. Figures 4.35, 4.36, 4.37, 4.38, 4.39, 4.40, and 4.41 present the retrieval results for seven master-marks. The correct mark is retrieved at the 1st, 2nd, 8th, 8th, 4th, 15th, and 1st places, respectively, for the seven query marks. These results demonstrate the robustness of the matching scheme employed in the presence of noise, loss of data, and variations in scale and orientation.

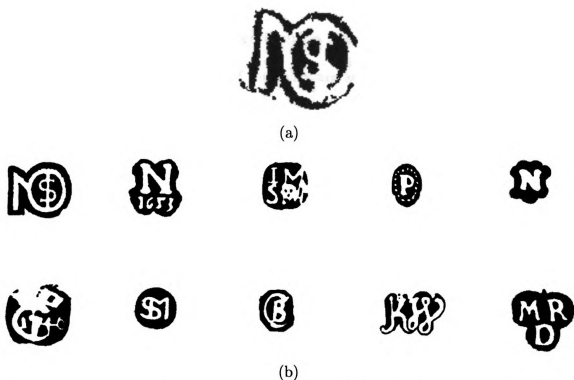


Figure 4.35: Retrieval of master seals; (a) Input Master Mark; (b) Top 10 retrievals based on the integrated match.

4.3 Discussion

Our initial experiments demonstrate the effectiveness of integration of simple shape features for pruning large databases to a smaller set. The proposed pruning strategy

is robust under noisy (hand drawn trademark images and faded master-marks), arbitrarily oriented (master-mark images) and scaled images (multiple trademarks in database with different scales as well as scale difference in master-marks in database with respect to real marks extracted from the bottom of the tankards).

4.3.1 Trademark Images

An efficient shape-based retrieval algorithm has been developed to retrieve trademark images. Efficiency and accuracy of retrievals are achieved by designing a two-stage hierarchical retrieval system: *(i)* a simple statistical feature-based process quickly browses through a database for a moderate number of plausible retrievals; and *(ii)* a deformable matching process screens the candidate set for the best matches. Preliminary results on a trademark image database show that this is a promising technique for content-based image database retrieval. The technique is robust under rotated, scaled and noisy versions of the database images.

We note that image retrieval based on user provided information such as hand drawn sketches is a challenging problem in multimedia applications. A human's perception of shape can be rather subjective and as a result, the representation of the desired object tends to vary a lot. Our goal is to extract images which have a similar semantic content to a query image. Figure 4.11 shows that semantically similar images may actually be visually very different from each other. In order to retrieve these images in the fast pruning stage, we need to somehow extract semantic meaning from the images. We present a heuristic that extracts the outline of the

objects in an image and fills up the interior as a basis to generalize the shape of the objects in the image. This scheme improves the retrieval results for the trademark images. Another way to extract the semantic content is through a semi-automatic (manual intervention needed in preprocessing) scheme using textual description of the trademark images. Text-based search can then be incorporated prior to the pruning stage. A more automatic and objective approach to preserving the semantic content is to extract image components that represent different entities in the trademark images, and to query on the basis of these component entities. We have used a simple heuristic that can extract individual object entities from an image. At present the segmentation scheme has not been completely implemented and in its current state, many noisy components are also represented as objects. The system does not have the capability of matching occluded objects. We feel that trademark images represent a very wide gamut of images, and incorporating partial matches may make the system highly unstable.

4.3.2 Norwegian Silver

We have been encouraged by the results of our preliminary experiments to recognize and match master-marks. However, our ultimate goal is to build a system which dates and identifies the craftsman even if the master-mark is not present. In addition, we plan to develop a system which is usable by non-experts by incorporating the expertise of the connoisseur into the system. In this way, we can preserve the expertise for future generations. Since our system stores digital images, it can also be used to match the

input query with objects which reside in distant museums, provided that the scanned images of these objects are available. The level of object detail captured with high-resolution, digital cameras also allows a detailed examination of the object without handling the object. Furthermore, the application of our system is not limited to Norwegian silver. The comparison of artifacts by shape (and color) features is a common element in the identification and authentication of artistic objects. Our system will be suitable for the identification of other art objects such as ancient pottery, woodwork, or china which have significant identifiable features.



(a)



(b)

Figure 4.36: Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.

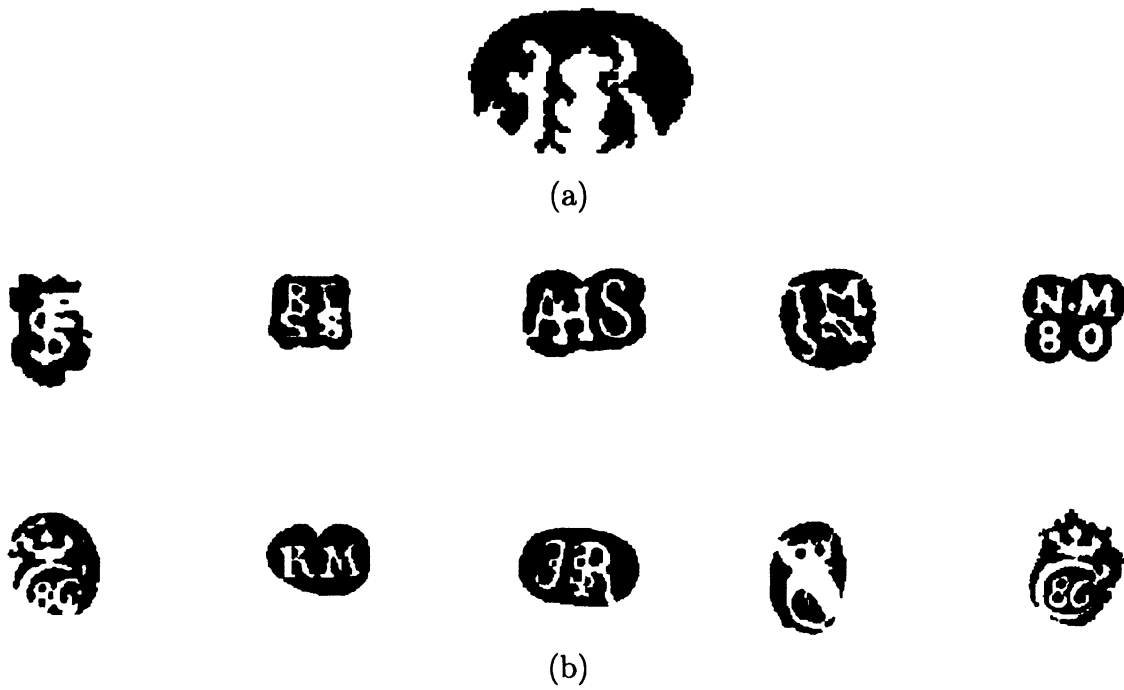


Figure 4.37: Retrieval of master seals; (a) Input Master Mark; (b) Top 10 retrievals based on the integrated match.

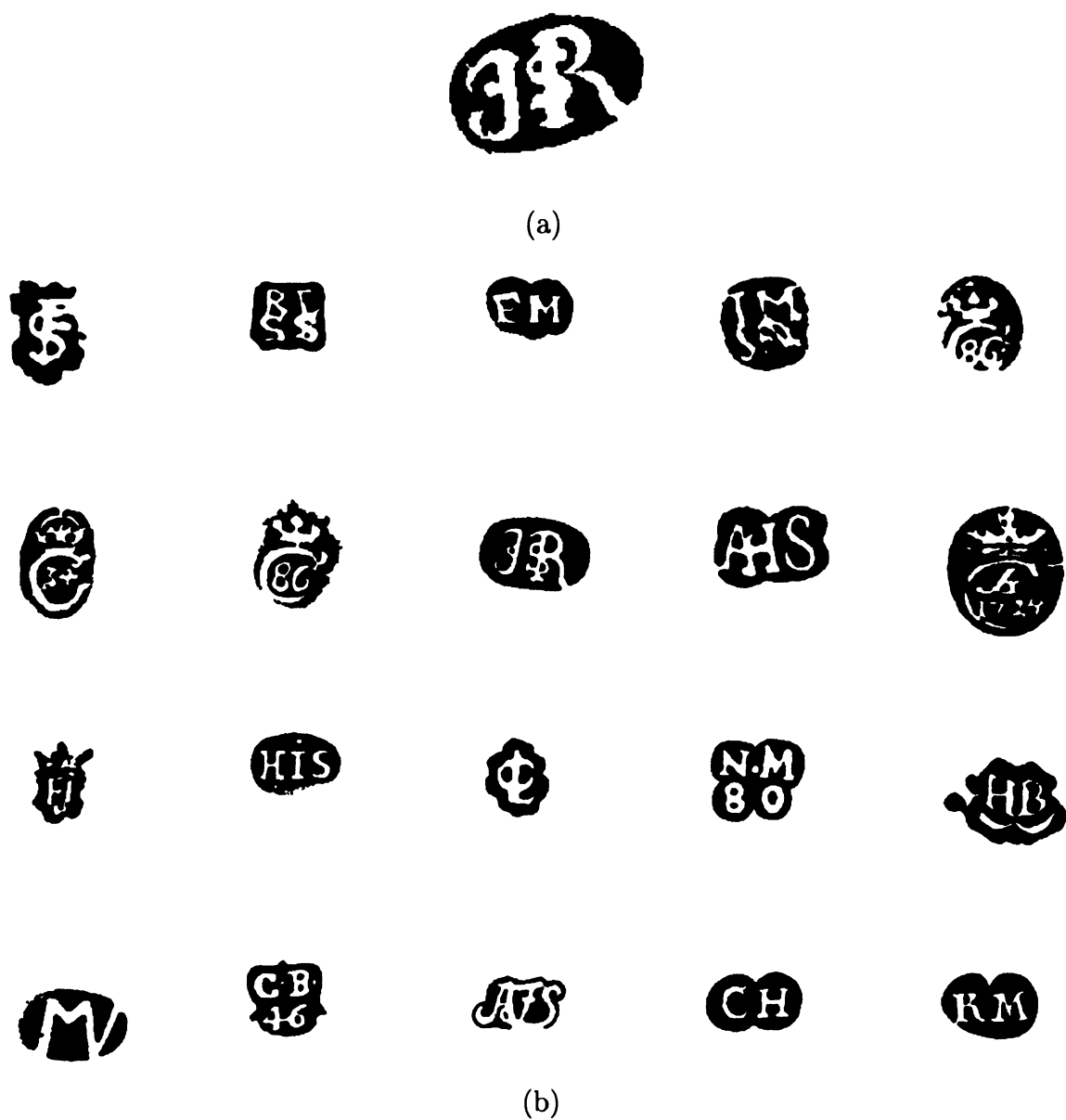


Figure 4.38: Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.

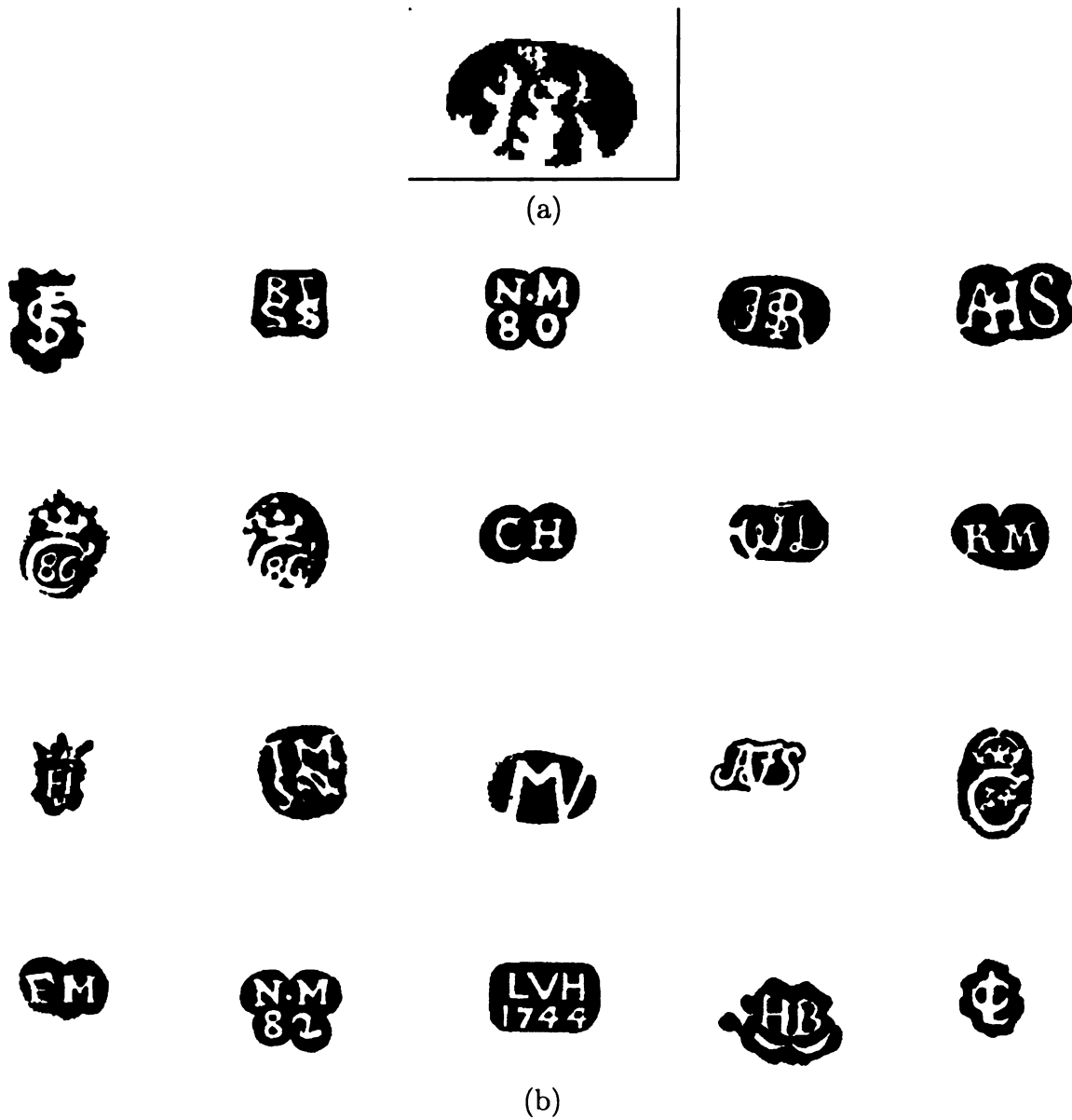


Figure 4.39: Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.

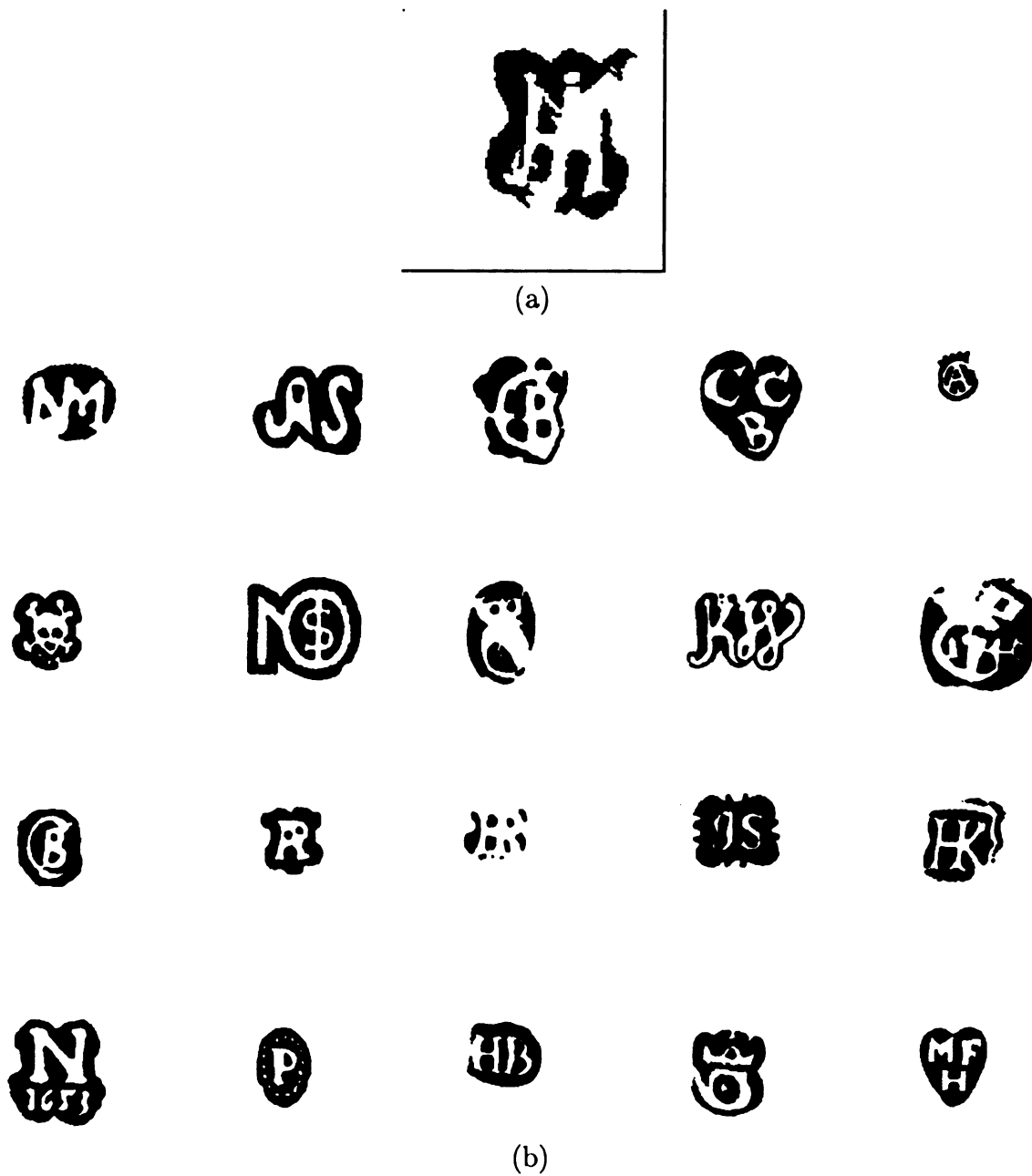


Figure 4.40: Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.



(a)



(b)

Figure 4.41: Retrieval of master seals; (a) Input Master Mark; (b) Top 20 retrievals based on the integrated match.

Chapter 5

Conclusions

We have addressed the problem of content-based retrieval from large image databases. Traditional text-based image database retrieval systems have many limitations and content-based search may be required to augment the text-based search. Any image database retrieval system must have a high speed and high accuracy of retrievals. Since image databases may consist of thousands of images, matching between images should be based on simple features that can be extracted efficiently and that can capture the contents of an image. This chapter summarizes the results of the research reported in this thesis. Several directions for future research are also outlined.

5.1 Summary

We have developed a hierarchical two-level feature extraction and matching methodology for image database retrieval for a trademarks database. The system takes as input a query mark, that is either a scanned drawing, or a hand drawn sketch. Using

a two-level hierarchy, the system uses simple and easily computable shape features to prune the large dataset to a small subset of plausible matches. The query is then matched with the pruned set of database images to eliminate false retrievals. A final subset of similar matches is then presented to the user.

For an efficient and accurate retrieval, we require that the pruning stage be not only fast, but also should not eliminate any images that are similar to the query. In the case of trademark images, similarity is based on semantic information, i.e., two images that are perceptually similar must be included in the pruning stage. These requirements mean that the pruning stage be general in nature rather than be specific in terms of its capability to handle images in different orientations, of different sizes, and at different positions. The proposed pruning stage integrates the results of edge direction histograms and invariant moments to select a plausible set of similar images. The moment features are invariant to rigid transformations, where as the edge direction histograms can be normalized and shifted to make them invariant to rigid transformations. The retrieved images from the fast pruning stage are then screened using a more elaborate, but costly deformable template matching process to remove false retrievals.

We have tested the performance of the entire matching system on a database of 1,100 trademark images, and 171 master-marks. We have explored how to sensibly combine some easily computable statistical shape features to yield fast and accurate retrievals. We have demonstrated how a two-stage hierarchical retrieval system can improve the results presented to the user. We have shown how images can be queried on the basis of hand drawn sketches and demonstrated that the system is robust

under noisy, oriented, and scaled data (identification of faded, noisy, and possibly scaled and oriented master-marks). We have developed a heuristic to extract the outline of objects in an image and query on the basis of filled images to improve the recall and precision rates at the pruning stage. We have also proposed a strategy to query on the basis of individual objects rather than the entire image. We have implemented a simple segmentation scheme for extracting object entities from an image.

5.2 Future Research

There are a number of research issues which need to be addressed in the future. The system can be extended to match on the basis of segmented objects rather than the entire image. A more robust and automatic segmentation algorithm would be required to match on the basis of objects rather than the entire image. A further step could be in the direction for allowing partial matching of objects based on local features. Currently, features that we use in the fast pruning stage are global features. The edge direction histograms and invariant moments are global statistical features and do not maintain local information. Local information can be very useful in eliminating many false matches. The system currently has no learning capabilities. It cannot improve its performance with time. We feel that for any system to perform robustly under various conditions, one needs to incorporate learning capabilities into the system. A future extension may be in the direction of allowing the system to learn through positive and negative examples of every query.

APPENDICES

Appendix A

A Representative Subset of Trademark Images

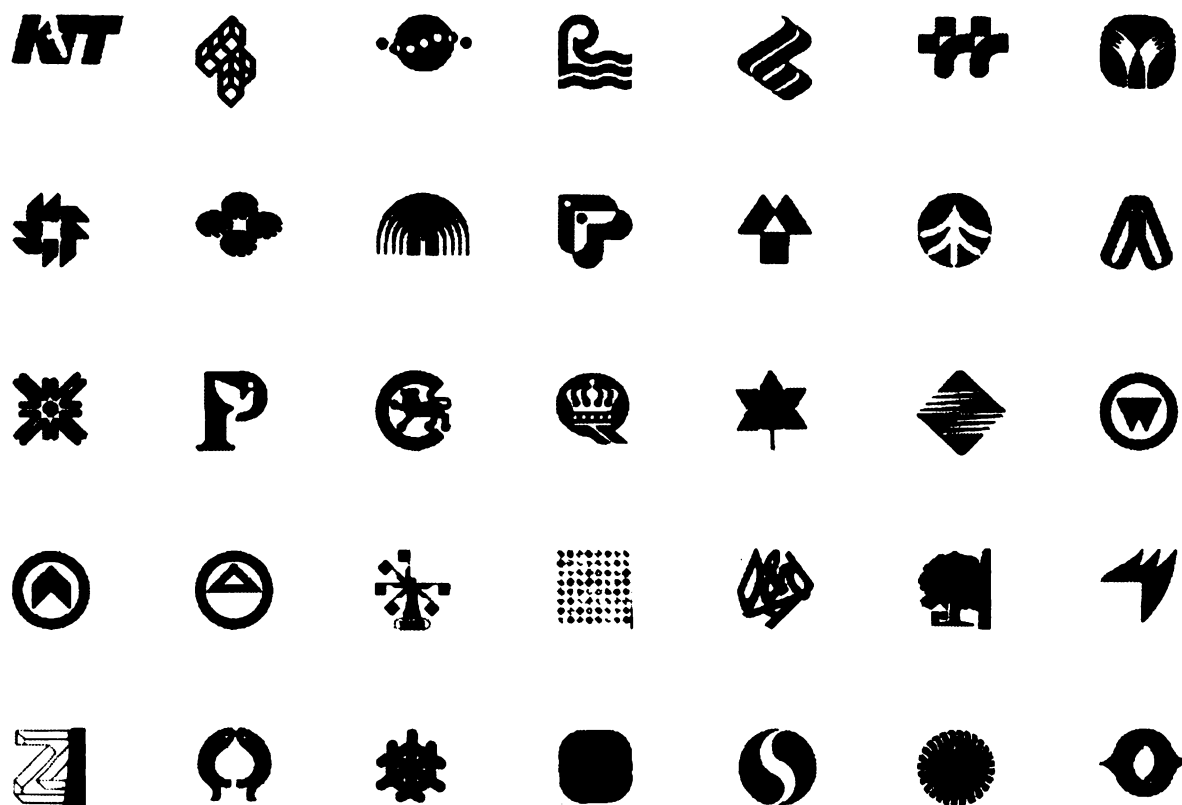
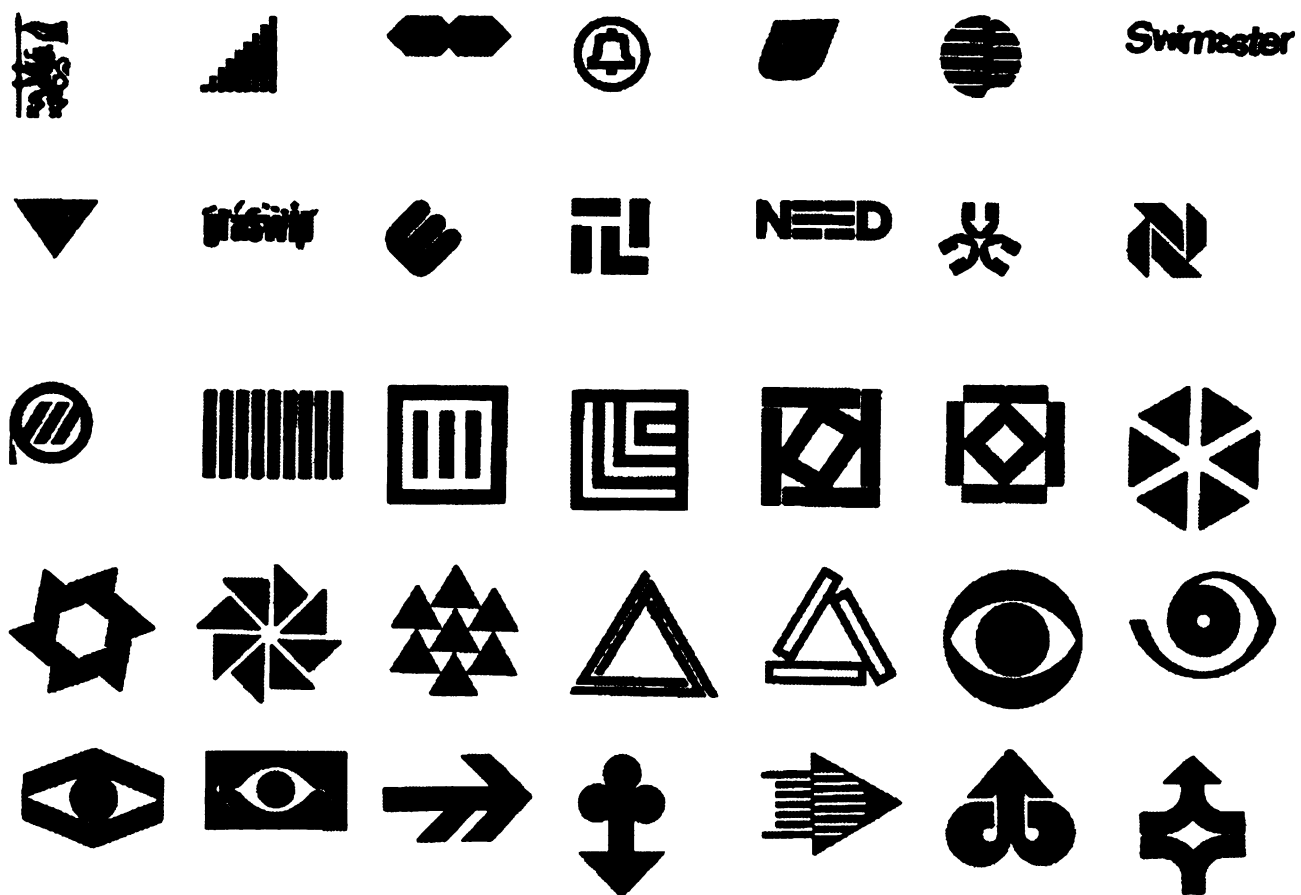
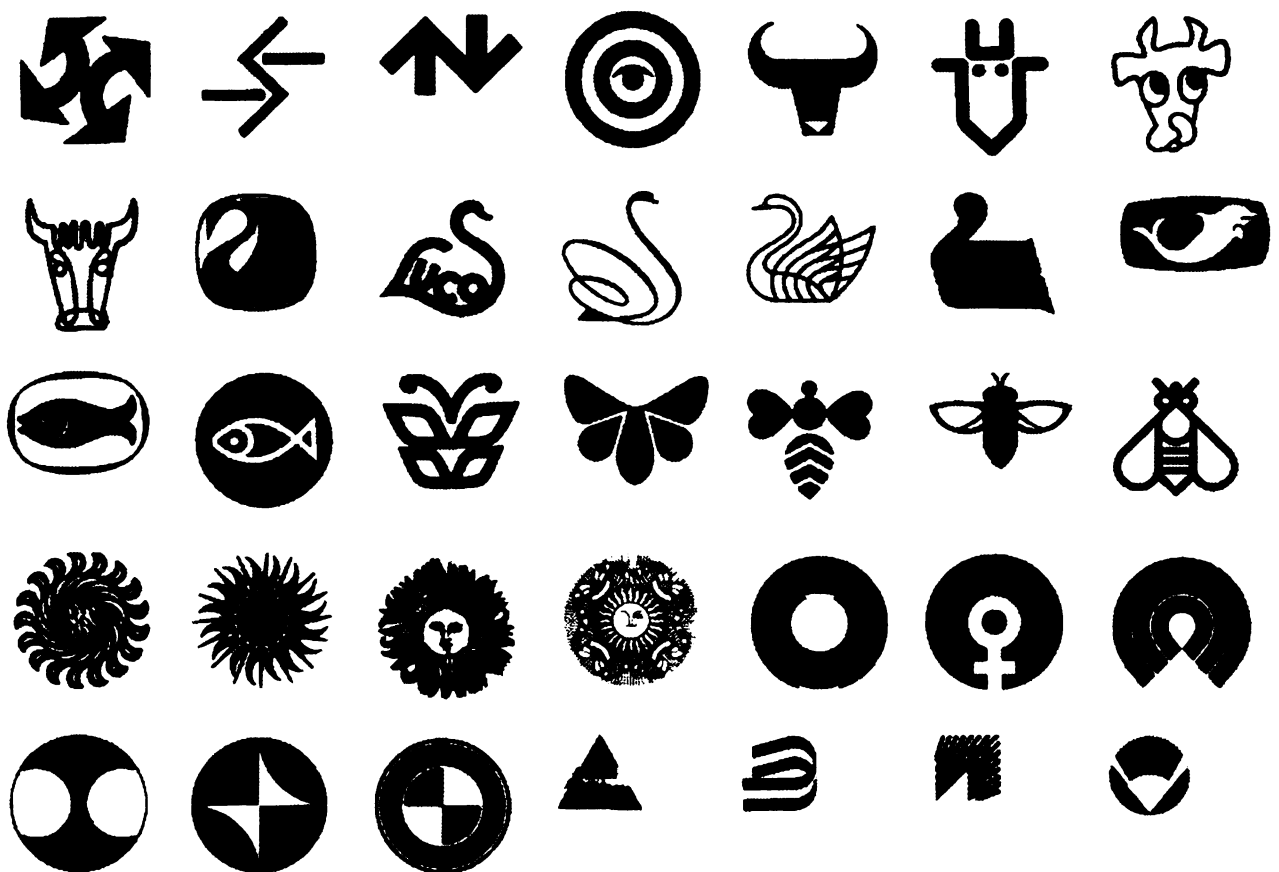


Figure A.1: Examples of trademarks present in the database.



Examples of trademarks present in the database.



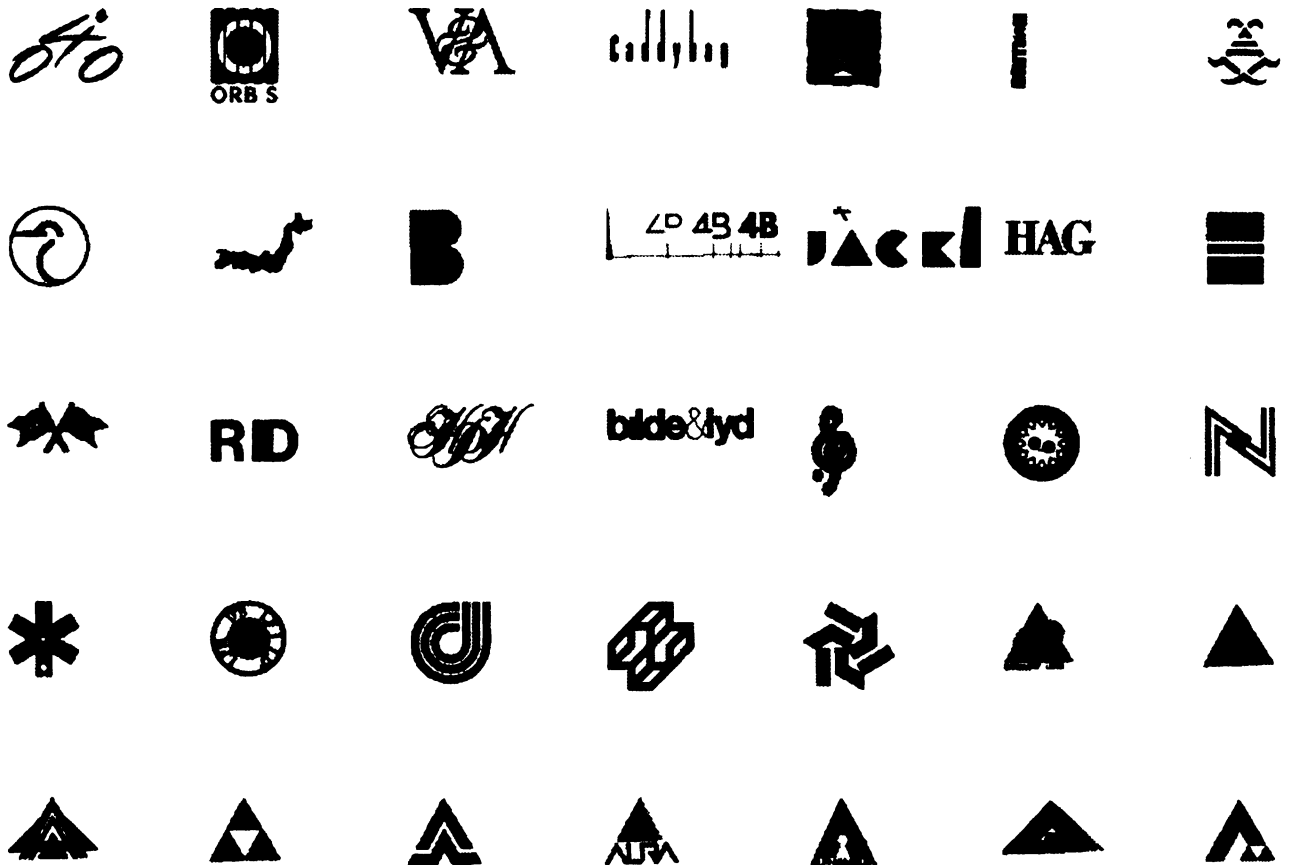
Examples of trademarks present in the database.



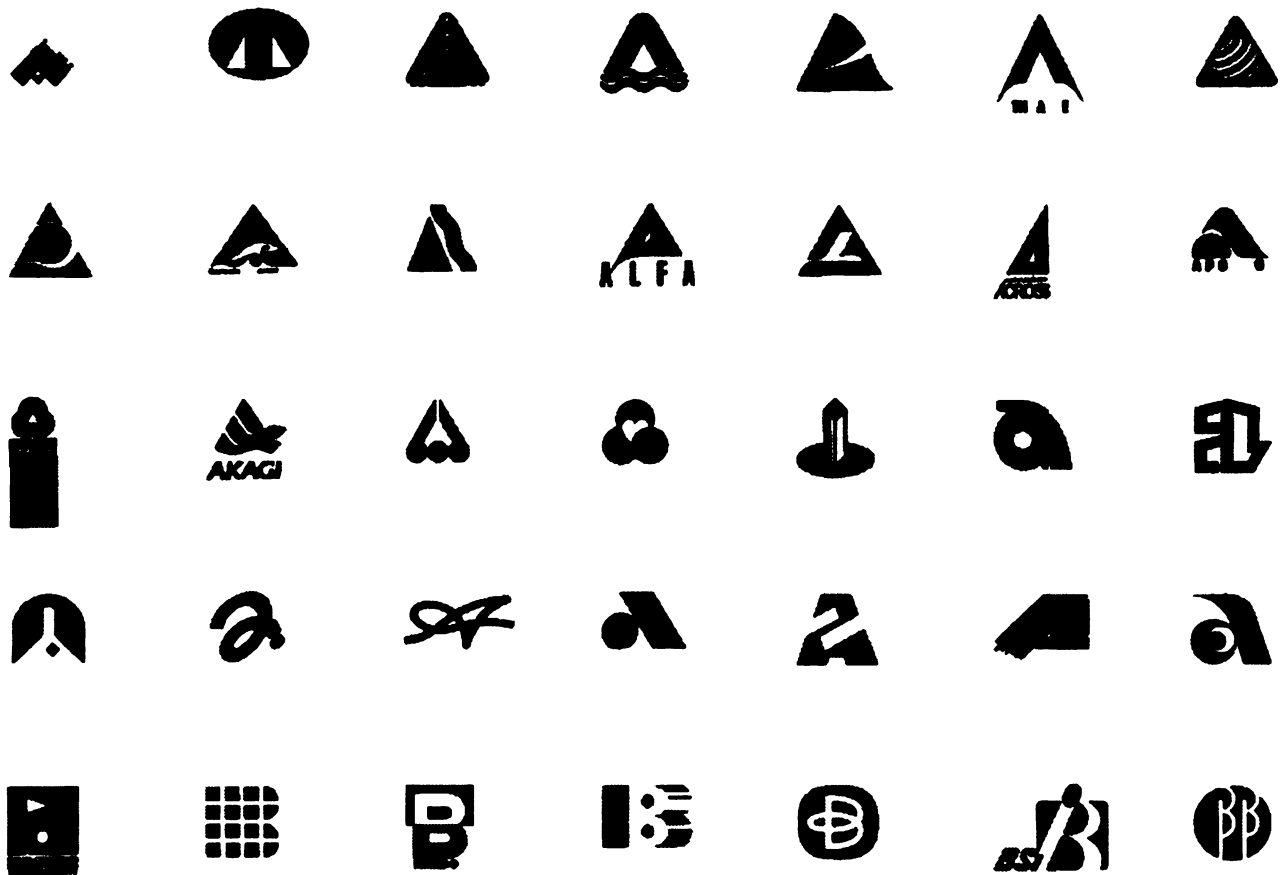
Examples of trademarks present in the database.



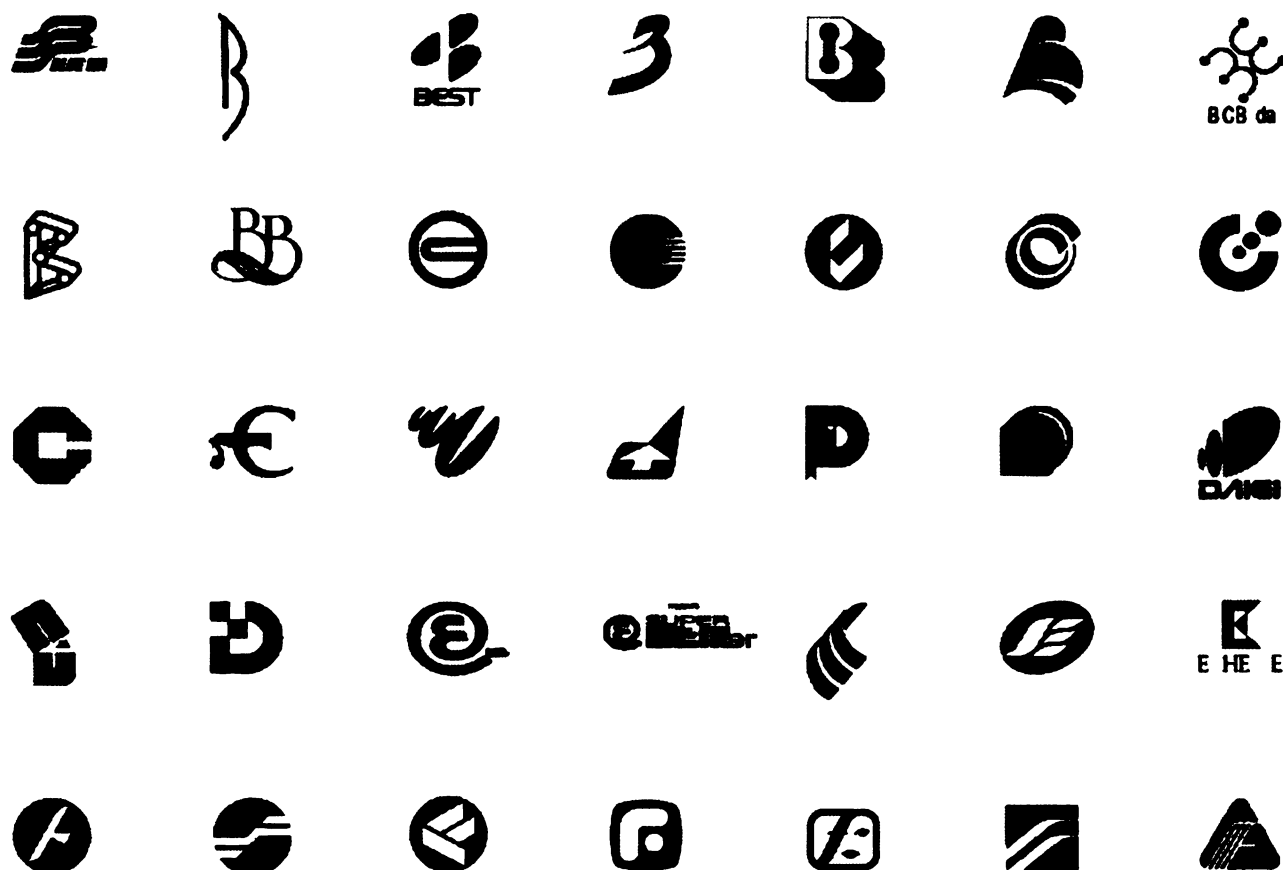
Examples of trademarks present in the database.



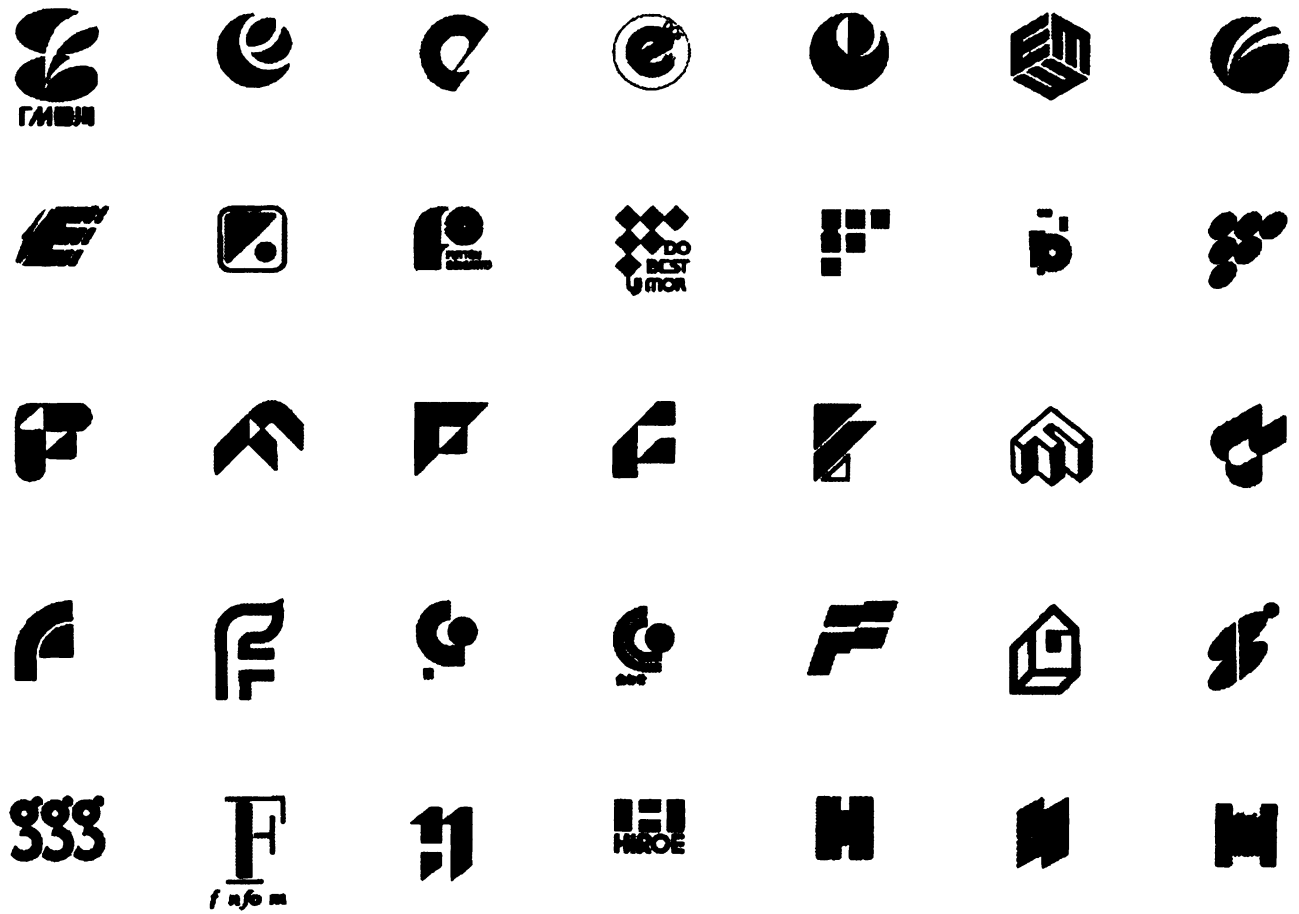
Examples of trademarks present in the database.



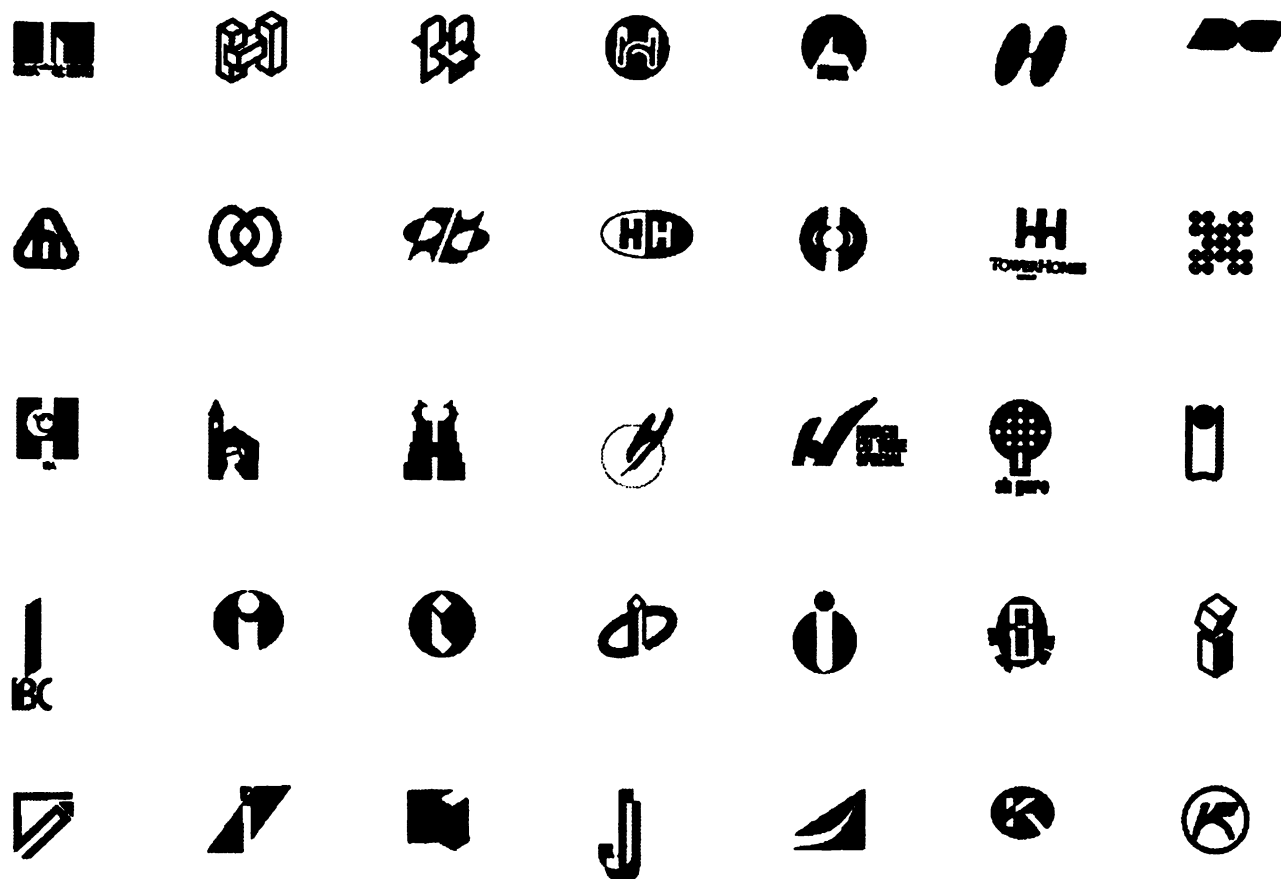
Examples of trademarks present in the database.



Examples of trademarks present in the database.



Examples of trademarks present in the database.



Examples of trademarks present in the database.

Bibliography

- [1] Stephen W. Smoliar and HongJiang Zhang. Content-based video indexing retrieval. *IEEE Multimedia*, Summer:62–72, 1994.
- [2] Borko Furht. Multimedia systems: An Overview. *IEEE Multimedia*, Spring:47–59, 1994.
- [3] J. K. Wu, A. D. Narasimhalu, B. M. Mehtre, C. P. Lam, and Y. J. Gao. CORE: A content-based retrieval engine for multimedia information systems. *Multimedia Systems, Springer-Verlag*, 3:25–41, 1995.
- [4] William I. Grosky. Multimedia information systems. *IEEE Multimedia*, Spring:12–23, 1994.
- [5] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3:231–262, 1994.
- [6] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *SPIE Vol 2185: Storage and Retrieval for Image and Video Databases II*, pages 34–47, February 1994.

- [7] C. P. Lam, J. K. Wu, and B. Mehtre. STAR - a system for trademark archival and retrieval. In *Proceedings of the 2nd Asian Conference on Computer Vision*, volume III, pages 214–217, December 1995, Singapore.
- [8] A. K. Jain and A. Vailaya. Image retrieval using color and shape. In *Proceedings of the 2nd Asian Conference on Computer Vision*, volume II, pages 529–533, December 1995, Singapore.
- [9] USPTO. *Basic Facts About Registering a Trademark*. US Patent and Trademark Office, http://www.uspto.gov/web/trad_reg_info/basic_facts.html, 1995.
- [10] USPTO. *Trademark Manual of Examining Procedures*. US Patent and Trademark Office, <http://www.uspto.gov/web/info/ftp.html>, 1995.
- [11] T. Igarashi, editor. *World Trademarks and Logotypes*. Graphic-sha, Tokyo, 1983.
- [12] T. Igarashi, editor. *World Trademarks and Logotypes II : A Collection of International Symbols and their Applications*. Graphic-sha, Tokyo, 1987.
- [13] *Collection of Trademarks and Logotypes in Japan*. Graphic-sha, Tokyo, 1973.
- [14] Theo Pavlidis. Algorithms for shape analysis for contours and waveforms. In *Proceedings of Fourth International Joint Conference on Pattern Recognition*, pages 70–86, Kyoto, Japan, November 7-10 1978.
- [15] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

- [16] R. Schettini. Multicolored object recognition and location. *Pattern Recognition Letters*, 15:1089–1097, November 1994.
- [17] B. M. Mehtre, M. S. Kankanhalli, A. D. Narsimhalu, and G. C. Man. Color matching for image retrieval. *Pattern Recognition Letters*, 16:325–331, March 1995.
- [18] Larry S. Davis. Shape matching using relaxation techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(1):60–72, January 1979.
- [19] G. Cortelazzo, G. A. Mian, G. Vezzi, and P. Zamperoni. Trademark shapes description by string-matching techniques. *Pattern Recognition*, 27(8):1005–1018, 1994.
- [20] P. W. Huang and Y. R. Jean. Using 2D C^+ -strings as spatial knowledge representation for image database systems. *Pattern Recognition*, 27(9):1249–1257, 1994.
- [21] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, September 1993.
- [22] D. J. Kahl, A. Rosenfeld, and A. Danker. Some experiments in point pattern matching. *IEEE Transactions on Systems, Man, and Cybernetics*, 10(2):105–116, February 1980.
- [23] J. K. Cheng and T. S. Huang. Image registration by matching relational structures. *Pattern Recognition*, 17(1):149–159, 1984.

- [24] S. P. Smith and A. K. Jain. Chord distribution for shape matching. *Computer Graphics and Image Processing*, 20:259–271, 1982.
- [25] W. Richards and D. D. Hoffman. Codon representation on closed 2d shapes. *Computer Vision, Graphics, and Image Processing*, (31):265–281, 1985.
- [26] C. T. Zahn and R. Z. Roskies. Fourier descriptors for plane closed curves. *IEEE Transactions on Computer*, C-21(1):269–281, 1972.
- [27] S. A. Dudani, K. J. Breeding, and R. B. McGhee. Aircraft identification by moment invariants. *IEEE Transactions on Computers*, C-26(1):39–45, January 1977.
- [28] M. Tuceryan and A. K. Jain. Texture analysis. In C. H. Chen, L. F. Pau, and P. S. P. Wang, editors, *Handbook of Pattern Recognition and Computer Vision*, pages 235–276. World Scientific Publishing Company, 1993.
- [29] R. M. Haralick. Statistical and structural approaches to texture. In *Proc. IEEE* 67, pages 786–804, 1979.
- [30] M. Tuceryan and A. K. Jain. Texture segmentation using Voronoi polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:211–216, 1990.
- [31] D. Blostein and N. Ahuja. Shape from texture: Integrating texture-element extraction and surface estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(12):1233–1251, December 1989.

- [32] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of Royal Statistical Society Ser.B*, 36:192–236, 1974.
- [33] A. Pentland. Fractal-based description of natural scenes. *IEEE Trans. on Pattern Anal. Mach. Intell.*, 9:661–674, 1984.
- [34] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Opt. Soc. Am. Series, A* 7:923–932, 1990.
- [35] J. M. Coggins and A. K. Jain. A spatial filtering approach to texture ananlysis. *Pattern Recognition Letters*, 3:195–203, 1985.
- [36] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24:1167–1186, 1991.
- [37] B. Holt and L. Hartwick. Visual image retrieval for applications in art and art history. In *SPIE Vol. 2185: Storage and Retrieval for Image and Video Databases II*, pages 70–81, February 1994.
- [38] Mikihiro Ioka. A method of defining the similarity of images on the basis of color information. Technical Report RT-0030, IBM Tokyo Research Lab, 1989.
- [39] J. Francos. Orthogonal decomposition of 2-d random fields and their applications for 2-d spectral estimation. In *Signal Processing and its Applications*, pages 287–327, North-Holland, 1993.

- [40] R. W. Picard and T. P. Minka. Vision texture for annotation. Technical Report TR-302, MIT, 1994. In *Multimedia Systems: Special Issue on Content-based Retrieval* 3:3-14,1995.
- [41] B. M. Mehtre and M. S. Kankanhalli. Content-based image retrieval using a composite color-shape approach. Technical Report TR95-190-0, National University of Singapore, August 1995.
- [42] B. M. Mehtre and M. S. Kankanhalli. Shape measures for content based image retrieval: A comparison. Technical Report TR95-195-0, National University of Singapore, September 1995.
- [43] J. Weng. SHOSLIF: A framework for sensor-based learning for high-dimensional complex systems. In *IEEE Workshop on Architectures for Semantic Modeling and situation analysis in Large Complex Systems*, August 27-29 1995. Invited Paper.
- [44] T. P. Minka and R. W. Picard. Interactive learning using a ‘society of models’. Technical Report TR-349, MIT, 1995.
- [45] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, November 1986.
- [46] W. Niblack and J. Yin. A pseudo-Distance measure for 2D shapes based on turning angle. *Proc 2nd IEEE Int. Conf. on Image Processing (ICIP)*, III:352–355, 1995.

- [47] Y. Cui, D. Swets, and J. Weng. Learning-based hand sign recognition using SHOSLIF-M. *Proc. 5th Int. Conf. on Computer Vision (ICCV)*, pages 631–636, 1995.
- [48] A.K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Trans. Pattern Anal. and Machine Intell.*, 18:267–279, March 1996.
- [49] A. Vailaya, Yu. Zhong, and A. K. Jain. A Hierarchical System for Efficient Image Retrieval. Vienna, August 1996. Accepted: 13th International Conference on Pattern Recognition.
- [50] Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable template. *J. American Statistical Association*, 86(414):376–387, June 1991.
- [51] Britt Kroepelien. *Fra Stil til Algoritme*. PhD thesis, University of Bergen, 1995.
- [52] B. Kroepelien, A. Vailaya, and A. K. Jain. A Case Study in Norwegian Silver Authentication. Vienna, August 1996. Accepted: 13th International Conference on Pattern Recognition.
- [53] Jorunn Fossberg. *Norske Soelvestempler*. Universitetsforlaget, Oslo, Norway, 1994.
- [54] Thw. Krohn-Hansen and Robert Kloster. *Bergens Gullsmedkunst fra Laugstiden*. Gullsmedlauget i Bergen og Vestlandske Kunstinddustrimuseum, Bergen, Norway, 1957.