

Evaluation of Features for Defect Image Retrieval

Jukka Iivarinen, Antti Ilvesmäki, and Jussi Pakkanen
Laboratory of Computer and Information Science
Helsinki University of Technology
P.O. Box 5400, FIN-02015 HUT, Finland
e-mail: jukka.iivarinen, jussi.pakkanen@hut.fi

ABSTRACT

In this paper different feature descriptors are evaluated for content-based retrieval of defect images. Especially the visual descriptors of the MPEG-7 standard are of interest. The evaluation is done with a small, pre-classified defect image database using the K -nearest neighbor (KNN) classification. The best features are then applied to a content-based image retrieval (CBIR) system called PicSOM. The results indicate that some of the MPEG-7 descriptors work very well with our defect images, but there is a need for some additional shape descriptors.

KEYWORDS

defect classification, MPEG-7 descriptors, shape descriptors, content-based image retrieval

1 Introduction

An increase in the amount of image data that has to be stored, managed and searched has spurred the research and development of systems that automatically compare and classify images on the basis of their content. Technological advances, such as the proliferation of digital cameras, inexpensive disk space and fast network connections have made huge databases of digital images possible, but finding the right image from such a database is a complex problem. Content-based image retrieval (CBIR) systems [1, 2, 3] aim to solve this problem by extracting features that describe the relevant aspects of the images, such as color or shape in a compact manner. Unlike the original images, the features can be fairly easily compared and meaningfully classified because they represent a higher level of abstraction.

The classification of paper defect images is a real-world problem that would benefit from a fast, efficient CBIR system capable of evaluating the images. A single paper web inspection system may produce hundreds or even thousands of defect images in a day, but only the most severe ones require immediate action. Sorting through the images manually would be extremely time-consuming and expensive. This paper examines the suitability of different feature descriptors (e.g. MPEG-7 descriptors) for classifying paper web defect images. It extends on previous research done in the Laboratory of Computer and Information Science at Helsinki University of Technology in the fields of defect image retrieval [4, 5].

2 Features for CBIR

Two types of features are of interest when considering defect images: shape features and internal structure features. Shape features are used to capture the essential shape information of defects in order to distinguish between differently shaped defects, e.g. spots and wrinkles. In addition to shape a defect has some kind of internal structure that consists of dark and light areas, holes, etc. Internal structure features are used to characterize the gray level and textural structure of defects.

In this study we have experimented with our old features that we used in our defect classifier already in 1998 [6] and with the features of the MPEG-7 standard [7, 8, 9]. The MPEG-7 standard ISO/IEC 15938, formally named “Multimedia Content Description Interface”, defines a comprehensive, standardized set of audiovisual description tools. It is not aimed at any one application in particular, instead it supports a broad range of descriptors for different types of multimedia information. The aim of the standard is to facilitate quality access to content, which implies efficient storage, identification, filtering, searching and retrieval of media. In addition to the still image descriptors demonstrated in this paper, the standard includes various descriptors for video and audio content. Furthermore, the standard enables the definition of new types of content descriptors. The existing MPEG-7 descriptors provide a framework for both high level semantic abstractions and low level abstractions such as shape, color and movement. For paper defect image retrieval, only the low level visual descriptors that can be extracted without human interaction are of interest. The feature vectors corresponding to the MPEG-7 visual descriptors were extracted with the MPEG-7 Experimentation Model (XM) software version 5.5.

2.1 Shape Features

There are a plenty of shape descriptors available [10, 11] that can be divided into two main categories: region-based and contour-based methods. Region-based methods use the whole area of an object for shape description, while contour-based methods use only the information present in the contour of an object. There are some techniques, for example, Fourier transforms and moments, that can be ap-

plied using both approaches with only small changes in algorithms. Using only contour information in shape analysis can be beneficial:

- information inside the object's contour is lost when dealing only with the contour (whether this is an advantage or a disadvantage, depends on the application)
- it takes less space to store different objects (data compression)
- shape descriptors are faster to calculate because there are less image pixels to process (although the overhead that comes from contour tracking must be included in total computation time)
- variations in a contour are more easily detected

In case of surface defects the main interest is in the shape of a defect's contour. A natural approach is thus to consider contour-based shape descriptors that can better capture the information that is present in contours. One such set of shape descriptors, simple shape descriptors (SSD), was proposed in [12] where it was noted that each of the proposed descriptors was insufficient for a complex recognition task, but a combination of them had good recognition capabilities and also low computation costs. The other two shape descriptors that are used in this study, the edge histogram (EH) and the region-based shape (RS), are from the MPEG-7 standard. Different kinds of edge histograms have been popular in various CBIR applications since they are powerful, fast to extract, and they do not need a segmentation mask. The region-based shape is a standard region-based shape descriptor that (like the SSD) needs a segmentation mask. So the shape descriptors utilized in this paper are:

- *Simple shape descriptors (SSD)* calculate six features from defect's border: convexity, principal axis ratio, compactness, circular variance, elliptic variance, and angle. They are translation, rotation (except angle), and scale invariant shape descriptors.
- *Edge histogram (EH)* calculates the amount of vertical, horizontal, 45 degree, 135 degree and non-directional edges in 16 sub-images of the picture, resulting in a total of 80 histogram bins.
- *Region-based shape (RS)* descriptor utilizes a set of 35 Angular Radial Transform (ART) coefficients that are calculated within a disk centered at the center of the image's Y channel.

In addition to the EH and the RS descriptors, the contour-based shape descriptor defined by MPEG-7 could also be of use. However, it produces vectors with varying amounts of components which requires a unique method for similarity matching. This makes it useless for our CBIR application.

2.2 Internal Structure Features

As already noted, we are using texture and gray level features to characterize defect's internal structure. For texture description, we tested two methods: the co-occurrence matrix method (texture features (TEX)) and the homogeneous texture (HT) from the MPEG-7 standard. The co-occurrence matrix method, known also as the spatial gray-level dependence method, has been widely used in texture analysis since the early 1970's [13]. It is based on repeated occurrences of different gray level configurations in a texture. It works well for a large variety of textures, especially for stochastic textures. The co-occurrence matrix is usually reduced to a set of features to decrease calculation time and memory requirements. The homogeneous texture implements another common technique called Gabor filtering. So, the texture features are:

- *Texture features (TEX)* are calculated from the co-occurrence matrix. The co-occurrence matrix is formed to each defect and four texture features, energy, contrast, entropy, and mean, are calculated from it.
- *Homogeneous texture (HT)* descriptor filters the image with a bank of orientation and scale tuned filters that are modeled using Gabor functions. The first and second moments of the energy in the frequency domain in the corresponding sub-bands are then used as the components of the texture descriptor.

There is another texture descriptor in MPEG-7, called the texture browsing. However, XM 5.5 had problems calculating the texture browsing descriptor correctly. Because the homogeneous texture represents essentially the same qualities of an image (directionality, coarseness and regularity) but with more precision, the issues with the texture browsing were not further examined and the descriptor was omitted.

For gray level description, we tested the simple gray level histogram (GLH) and three descriptors from the MPEG-7 standard. The gray level (or color) features are:

- *Gray level histogram (GLH)* is a 256-bin distribution of gray levels in a defect.
- *Color layout (CL)* specifies a spatial distribution of colors. The image is divided into 8×8 blocks and the dominant colors are solved for each block in the YCbCr color system. Discrete Cosine Transform is applied to the dominant colors in each channel and the DCT coefficients are used as a descriptor.
- *Color structure (CS)* slides a structuring element over the image. The numbers of positions where the element contains each particular color are stored and used as a descriptor.
- *Scalable color (SC)* is a 256-bin color histogram in HSV color space, which is encoded by a Haar transform.

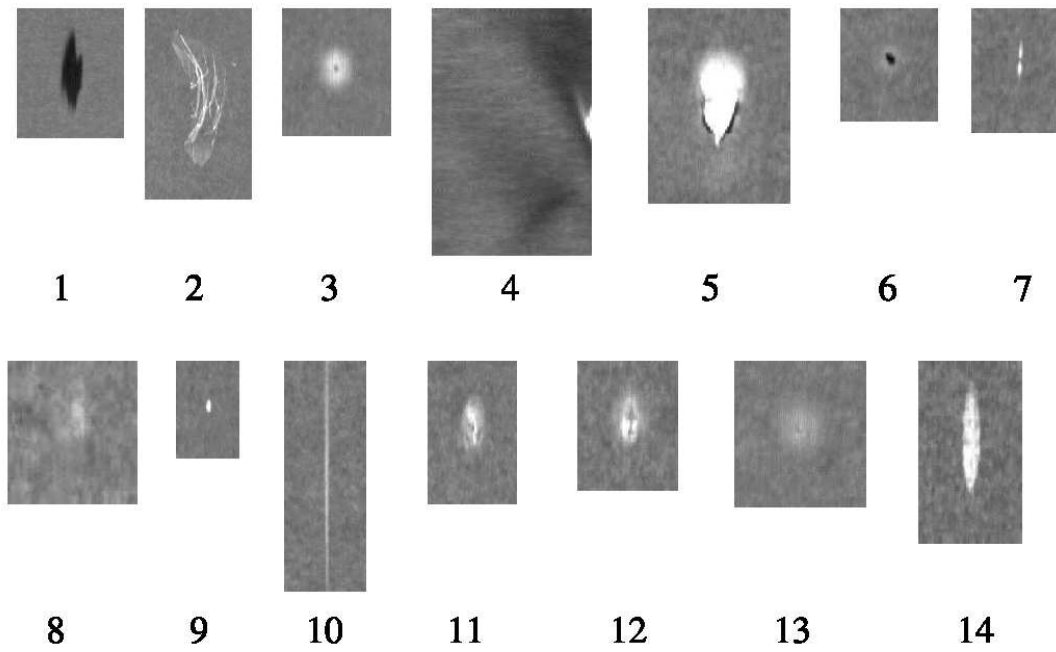


Figure 1. Examples of the defect classes. Note how several of the classes are very similar, especially classes 11 and 12.

3 Evaluation of Features

Our image database has 1308 paper defect images that were obtained from a real, online paper web inspection system. The images are preclassified into 14 different classes with 100 images in 12 of the classes, 76 images in class number 11 and 32 images in class number 12. Example images from each class are depicted in Fig. 1. The classes are based on the cause and type of a defect, and can therefore contain images that are visually dissimilar in many aspects. On the other hand, some classes are very hard to tell apart. This makes their classification with general-purpose visual descriptors an extremely challenging problem. Even a human cannot always differentiate between the images.

The size of the images vary according to the size of a defect. Some images are only 200 by 200 pixels in size while others are several thousand pixels high. All the images are in gray-scale with 256 gray levels and they are segmented after acquisition so that each image has a gray level image and a binary segmentation mask that indicates defect areas in the image. Segmentation of these kind of images is a very hard problem, and it cannot be done with simple gray-level thresholding techniques [14]. The image database with the defect segmentation masks were provided by ABB Oy. All further processing is done only for defect areas; we are thus omitting the noninteresting background (or normal, intact surface). It is important to get rid of the background since it usually covers most of the image. Thus it may spoil feature extraction and then the whole CBIR process; the point here is to find similar defects, not similar images. Especially the shape features (SSD and RS)

need to have the segmentation mask, otherwise they could not be extracted.

3.1 Evaluation Results

The performance of different descriptors was evaluated by performing a leave-one-out K -nearest neighbor (KNN) cross-validators classification. Euclidean distance and $K = 5$ was chosen for all calculations. Some experiments were also performed using L1-norm as the distance (as suggested by the MPEG-7 standard), but no significant difference in the performance was detected. We also tested some combinations of different feature sets. The combined results were obtained by retrieving the classes of the five nearest images for each feature set and determining the winning class by counting the total occurrences of each class.

Classification results are shown in Table. 1. Clear winners can be found among the texture and the color features. These are the homogenous texture (HT) that really outperforms the TEX features, and the color structure (CS). The performances of the HT and the CS are really good, almost 80%. If we look at their results more closely, we see that the HT works significantly better than the CS with classes 3, 7, 9, and 10, but the CS is better with classes 2, 5, and 13. So none of them is better with all the classes.

The shape descriptor results are not so clear. The edge histogram (EH) is the best one with the simple shape descriptors (SSD) following. But the SSD is much better than the EH with classes 2 and 8, and equally good with classes 4, 6, 9, 10, and 14. So the SSD seems to be quite useful for

Table 1. Classification results.

		Classification rates (%) of different classes														avg
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Shape features	EH	61	25	68	91	55	23	76	37	64	69	79	11	78	91	61
	SSD	50	44	47	88	37	24	41	77	65	72	14	4	30	91	52
	RS	54	27	37	52	35	26	43	34	45	57	38	0	35	88	42
Texture features	HT	96	54	71	89	74	82	97	65	91	94	75	0	64	93	79
	TEX	30	59	38	98	39	14	69	41	48	19	5	7	53	40	42
Color features	CS	97	75	60	88	82	78	70	59	77	62	76	15	86	95	76
	GLH	98	73	39	92	57	49	1	89	26	42	13	0	40	79	53
	SC	75	6	28	69	39	74	90	17	6	18	70	0	46	60	44
	CL	85	11	11	64	41	78	23	32	14	33	63	4	31	39	39
EH,HT&CS		99	79	77	98	84	89	98	74	95	96	81	11	91	98	87
All MPEG-7		98	68	75	96	84	85	97	77	95	94	71	11	90	99	85
SSD,TEX&GLH		96	81	58	100	66	48	64	92	68	78	13	0	60	94	70

classification of these kinds of images. The region shape (RS) performed poorly.

Finally there some combined classification rates in Table 1. When all the MPEG-7 feature sets were combined the classification rate of 85% was obtained. This is not an optimal approach since these feature sets have a lot of redundant information. When using a combination of the best three feature sets (HT, CS, and EH), the classification rate of 87% was obtained. This can be considered a really good result for our database. We also tested the performance of our old features (SSD, TEX, and GLH). Their combined classification rate was only 70% which is still quite acceptable but is far worse than with the best MPEG-7 feature sets.

We also ran the same tests without the segmentation masks to see how the masks would affect the classification rates. The shape features (SSD and RS) were not tested since their calculation needs these masks. The results were quite surprising. For example, when using the combination of the best three MPEG-7 feature sets (HT, CS, and EH), the classification rate dropped from 87% to 80%. When all the five MPEG-7 descriptors (excluding RS) were combined, the classification rate dropped only 3%. So it is possible to use these features successfully even without the segmentation masks.

4 Content-Based Retrieval of Defect Images

We have realized a prototype CBIR system for defect images [4, 5] where we can test different feature descriptors. It is based on a noncommercial, generic content-based image retrieval system called PicSOM [15, 16]. PicSOM has been developed in Laboratory of Computer and Information Science at Helsinki University of Technology to be a generic content-based image retrieval (CBIR) system for large, unannotated databases. It builds on the concepts of unsupervised clustering, self-organizing maps, and rele-

vance feedback. PicSOM works in the following way. First a number of feature sets is extracted for each image in the database. Then a tree-structured self-organizing map (TS-SOM) [17] is trained with each feature set, and each image is associated to the closest map unit in each TS-SOM. Then the database can be searched.

In this paper we have applied the best feature sets (according to the experiments of the previous section) to PicSOM. Some example query results using the four feature sets (HT, EH, CS, and SSD) are shown in Figure 2. The leftmost image in each row was used as an initial query image. The four rightmost images are the best matches after 3-4 queries. At each query PicSOM returned 5 images of which the best ones were selected. This way we were trying to imitate the actual use of the PicSOM system where the user gives one image and tries to find within few queries similar images from the database.

Due to the nature of the PicSOM system, we cannot easily get exact classification percentages. We need some kind of a measure that is comparable to those. It is reasonable to assume that a human would do at most ten iterations of searching before quitting. Therefore we examined the recall rates after five and ten query iterations. These values are not exact classification rates, but they describe quite well the system's performance. The more images of the correct class the system returns the better the recall value. A value of 100% means the query target has been fulfilled optimally, since all images of the desired class have been found. It should be noted that in the optimal case recall could be 100% after five queries (since PicSOM was configured to return 20 best matches per query and almost all classes have 100 images).

The results of these tests can be seen in Figure 3 where the retrieval rates are depicted with the three best feature sets (HT, EH, and CS) with and without the additional shape feature set (SSD). The percentages are very much like the ones obtained with KNN. The recall rates are approximately 55% after five iterations and 80% after

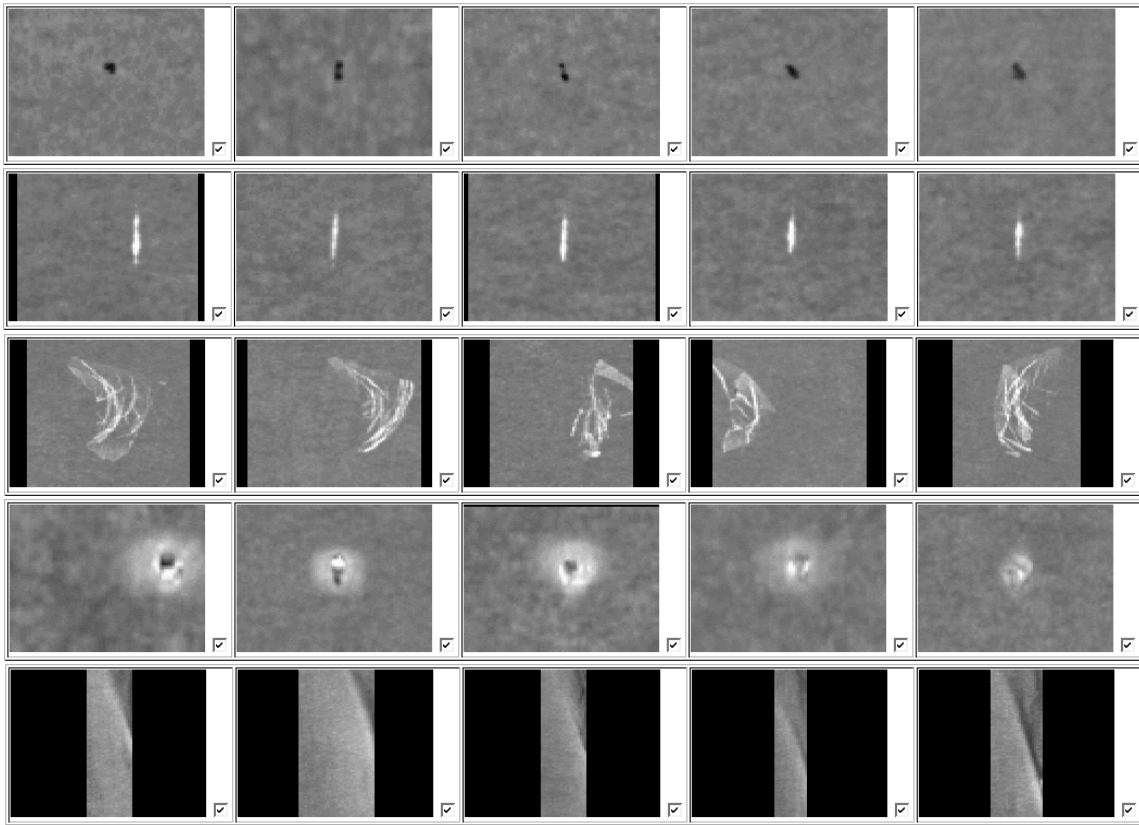


Figure 2. Some example query results.

ten iterations. These results can be considered quite good, given the data set's difficult nature. When we compare the two graphs in Figure 3, we find that adding the problem specific shape feature set (SSD) improves the query results approximately 6%, from 77% to 83%. This is a very noticeable improvement. It should be noted that the classes that are easy for the KNN classifier are also easy for PicSOM. However, PicSOM seems to do much better on the very difficult class 12. This can not be proved conclusively, though, given the different nature of the recall values and classification percentages.

5 Conclusions

In this paper we have evaluated several feature descriptors for content-based image retrieval of surface defect images, and applied them to the PicSOM CBIR system. The evaluation results show that some of the MPEG-7 descriptors can be used as features in machine vision problems. Especially the homogeneous texture, the color structure, and the edge histogram work well with our defect images. The region-based shape descriptor performs quite poorly, and thus there is a need for our own problem specific shape feature set that improves noticeably the obtained retrieval rates.

Acknowledgments

The authors wish to thank the PicSOM group at Helsinki University of Technology and our industrial partner ABB Oy (J. Rauhamaa). The financial support of the Technology Development Centre of Finland (TEKES's grant 40111/02) and the Academy of Finland (project New Information Processing Principles, 44886) is gratefully acknowledged.

References

- [1] Yong Rui, Thomas S. Huang, and Shih-Fu Chang. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1):39–62, 1999.
- [2] Alberto Del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann Publishers, Inc., 1999.
- [3] Björn Johansson. A survey on: Contents based search in image databases. Technical report, Linköping University, Department of Electrical Engineering, <http://www.isy.liu.se/cvl/Projects/VISIT-bjojo/>, December 2000.
- [4] Jukka Iivarinen and Jussi Pakkanen. Content-based retrieval of defect images. In *Proceedings of Ad-*

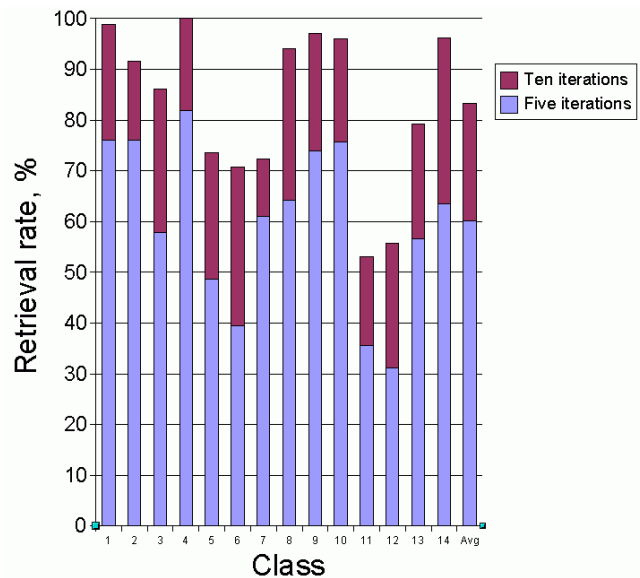
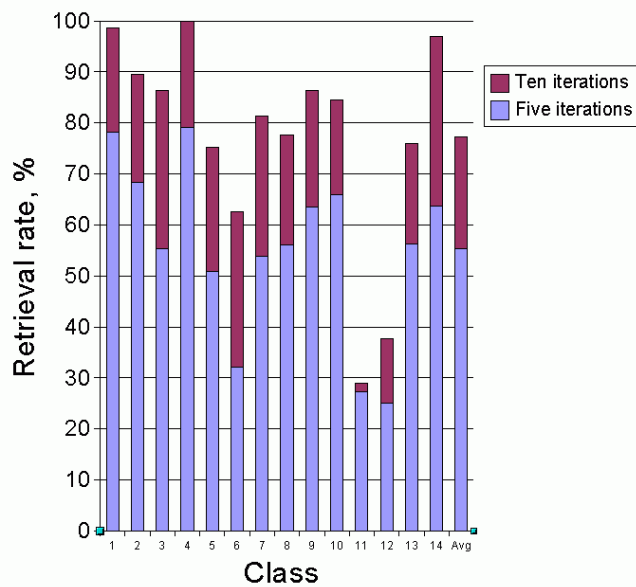


Figure 3. Query results with PicSOM. In the first graph the three best MPEG-7 descriptors (HT, CS, and EH) are used. The second has these features, and also the additional shape feature (SSD).

vanced Concepts for Intelligent Vision Systems, pages 62–67, Ghent, Belgium, September 9–11 2002.

- [5] Jussi Pakkanen and Jukka Iivarinen. Content-based retrieval of surface defect images with mpeg-7 descriptors. In *Proceedings of Sixth International Conference on Quality Control by Artificial Vision 2003*, pages 201–208, Gatlinburg, Tennessee, USA, May 19–23 2003.
- [6] Jukka Iivarinen and Ari Visa. An adaptive texture and shape based defect classification. In *Proceedings of the 14th International Conference on Pattern Recognition*, volume I, pages 117–122, Brisbane, Australia, August 16–20 1998.
- [7] B .S Manjunath, Jens-Rainer Ohm, Vinod V. Vasudevan, and Akio Yamada. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), June 2001.
- [8] MPEG-7. MPEG-7 visual part of the experimentation model (version 9.0). ISO/IEC JTC1/SC29/WG11 N3914, 2001.
- [9] MPEG-7. MPEG-7 multimedia content description interface – part 3 visual. ISO/IEC JTC1/SC29/WG11 W3703, 2001.
- [10] S. Marshall. Review of shape coding techniques. *Image and Vision Computing*, 7(4):281–294, November 1989.
- [11] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing, Analysis and Machine Vision*. Chapman & Hall Computing, London, 1993.
- [12] Jukka Iivarinen, Markus Peura, Jaakko Särelä, and Ari Visa. Comparison of combined shape descriptors for irregular objects. In *Proceedings of the 8th British Machine Vision Conference*, volume 2, pages 430–439, University of Essex, UK, September 8–11 1997.
- [13] R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, November 1973.
- [14] Jukka Iivarinen, Katriina Heikkinen, Juhani Rauhamaa, Petri Vuorimaa, and Ari Visa. A defect detection scheme for web surface inspection. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(6):735–755, September 2000.
- [15] Jorma Laaksonen, Markus Koskela, Sami Laakso, and Erkki Oja. PicSom - content-based image retrieval with self-organizing maps. *Pattern Recognition Letters*, 21(13-14):1199–1207, 2000.
- [16] Jorma Laaksonen, Markus Koskela, Sami Laakso, and Erkki Oja. Self-organising maps as a relevance feedback technique in content-based image retrieval. *Pattern Analysis and Applications*, 4(2+3):140–152, 2001.
- [17] Pasi Koikkalainen and Erkki Oja. Self-organizing hierarchical feature maps. In *Proceedings of 1990 International Joint Conference on Neural Networks*, volume II, pages 279–284, San Diego, CA, 1990.