Defect Image Classification and Retrieval with MPEG-7 Descriptors

Jussi Pakkanen, Antti Ilvesmäki, and Jukka Iivarinen

Helsinki University of Technology
Laboratory of Computer and Information Science
P.O. Box 5400, FIN-02015 HUT, Finland
{jussi.pakkanen, jukka.iivarinen}@hut.fi

Abstract. In this paper the visual content descriptors defined by the MPEG-7 standard are applied to defect image classification and retrieval. A pre-classified defect image database is used in evaluation. The experiments are done with a KNN classifier and with a PicSOM content-based image retrieval system. Results indicate that the MPEG-7 features work with a high level of success, especially the Color Structure and Homogeneous Texture descriptors seem to perform well.

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. Content-based image retrieval (CBIR) systems [10,1] 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.

This paper describes experiments with the visual descriptors of the MPEG-7 standard [9,8] using a pre-classified defect image database. The experiments are done with a KNN classifier and with a PicSOM CBIR system. The PicSOM system [5,6] has been developed in our laboratory at Helsinki University of Technology to be a generic CBIR system for large, unannotated databases. The test database consists of images of pre-classified defects. Some earlier work on the CBIR and classification of these kinds of images can be found e.g. in [4,3].

2 PicSOM and MPEG-7 Visual Descriptors

PicSOM is a system designed for content-based image retrieval (CBIR) in large, unannotated image databases [5,6]. It uses concepts of unsupervised clustering, self-organizing maps, and relevance feedback. The system works by first calculating a chosen number of feature sets for each image in the database. Then a tree-structured self-organizing map (TS-SOM) is trained with each feature set, and each image is associated to the closest map unit in each TS-SOM. After this the database can be searched. TS-SOM is a tree-structured vector-quantizer that

J. Bigun and T. Gustavsson (Eds.): SCIA 2003, LNCS 2749, pp. 349–355, 2003.

has SOMs at each of its hierarchical levels. This kind of tree-structure makes training and searching much faster when compared to a normal, nonhierarchical SOM of the same size.

Several types of features can be used in PicSOM for image querying. These include features for color, shape, texture, and structure description of the image content. When considering defect images, there are two types of features that are of interest: 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. Internal structure features are used to characterize the gray level and textural structure of defects.

The MPEG-7 standard ISO/IEC 15938, formally named “Multimedia Content Description Interface”, defines a comprehensive, standardized set of audiovisual description tools for still images as well as movies. The aim of the standard is to facilitate quality access to content, which implies efficient storage, identification, filtering, searching and retrieval of media.

The feature vectors corresponding to the MPEG-7 visual descriptors were extracted with the MPEG-7 Experimentation Model (XM) software version 5.5. We have used the following still image features:

- *Dominant color* represents the most dominant colors.
- *Color layout* 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* slides a structuring element over the image, the numbers of positions where the element contains each particular color is recorded and used as a descriptor.
- *Scalable color* is a 256-bin color histogram in HSV color space, which is encoded by a Haar transform.
- *Edge histogram* calculates the amount of vertical, horizontal, 45 degree, 135 degree and non-directional edges in 16 sub-images of the picture.
- *Homogeneous texture* 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.

For more in-depth explanations and definitions of the different descriptors, their intended range of use and the precise algorithms see [7,2,9,8].

3 Experiments

A pre-classified database of approx. 1300 grayscale images was used to test the descriptors. The database, supplied by ABB Oy, contains defect images of varying size and type (spots, holes, wrinkles, streaks, tears, slime residues etc.) from a real, online process. The size of the images varied considerably.

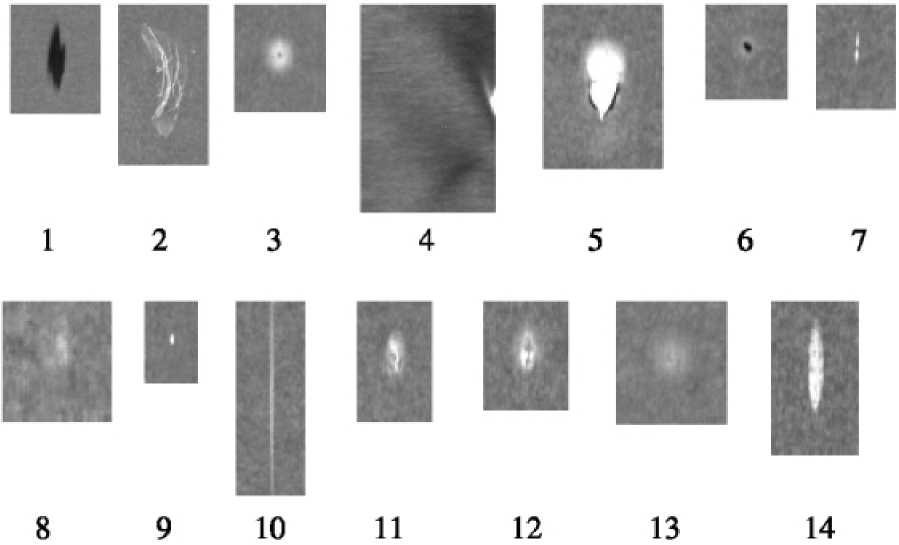

Fig. 1. Example images from each defect class.

The database had been pre-classified into 14 distinct classes, with 100 images in 12 of the classes, 76 images in class number 11 and 32 images in class number 12, adding up to a total of 1308 images. Example images from each class are depicted in Figure 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 different classes.

3.1 KNN Cross-Validation

The performance of the descriptors was first evaluated by performing a leave-one-out k-nearest neighbor (KNN) cross-validatory classification with k=5. The results are given in Figure 2. The best descriptors are Homogeneous Texture and Color Structure, with Edge Histogram and Color Layout following. Scalable Color and Dominant Color perform quite appallingly, as was expected. No descriptor is the best in all classes. Classes 11 and 12 are recognized miserably because of several reasons: the size of the classes is smaller than the others, they are very similar with each other and many images in them also look a lot like the images in classes 3 and 5. The outcome of all this is that their performance is sometimes worse than random guessing.

A simple KNN combination of all the descriptors was also tested. This was done by retrieving the classes of the five nearest images for each descriptor and the winning class was determined by counting the total occurrences of each class. This method was able to perform better than any single descriptor alone. The

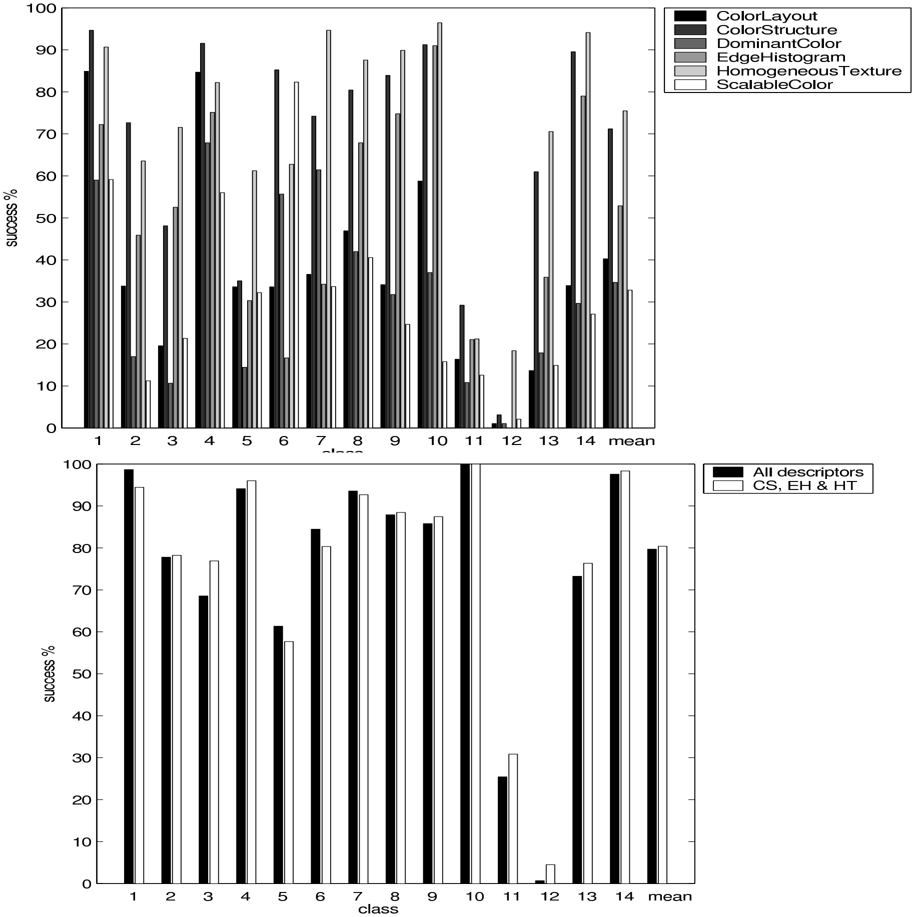

Fig. 2. Classification results for all the descriptors and classes (upper) and results of a KNN combination of all the descriptors and the combination of Color Structure, Edge Histogram and Homogeneous Texture (lower).

results are in Figure 2. When using the combination of the three best descriptors, Homogeneous Texture, Color Structure and Edge Histogram, the average performance is slightly better.

3.2 CBIR Experiments

PicSOM has a built-in method of evaluating the retrieval results. You tell it which defect class you want to test and which features to use. It then seeks a single example image belonging to that class to start the query. Then it seeks the best matches to the query and selects all returned images that belong to the desired class. All these selected images are then used as a search criterion for another search. Again the found correct images are added to the query criterion and a new query is performed. This process continues until a pre-specified number of iterations is reached.

The results contain recall and precision values for each iteration. Recall means the percentage of correct images retrieved so far. When recall reaches 1, all desired images have been found. Precision tells how many of all returned images belong to the correct class. A minimum requirement of any CBIR system is that precision should be higher than the a priori probability of the queried class. Otherwise the system is worse than just selecting images in the database at random.

The obtained results are very good. Most of the classes are found with high precision values using relatively few queries. Even when using just one feature, the obtained results are clearly above the a priori probability. The classes 11 and 12 were very difficult as was discovered in the KNN experiment.

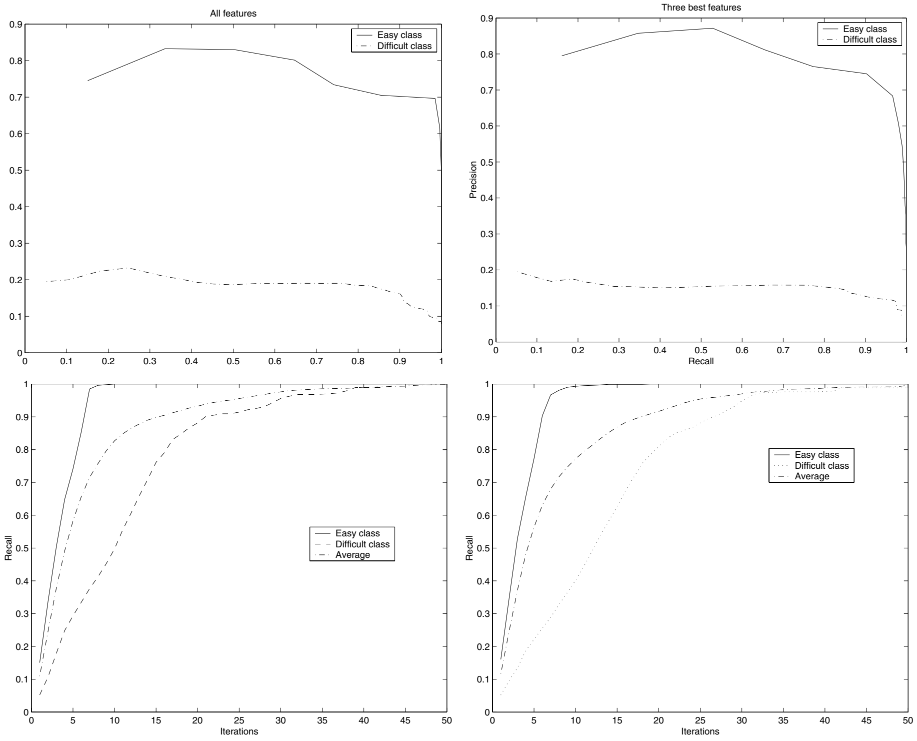

Fig. 3. Example precision/recall graphs when using all descriptors and the three best descriptors (upper row). Recall values as a function of the number of query iterations when using all descriptors and the three best descriptors (lower row).

The upper row of Figure 3 shows two precision/recall figures with all descriptors and the three best descriptors (Homogeneous Texture, Color Structure, and Edge Histogram). Each figure contains two graphs, one for an easy class and one for a difficult class, when using all features. The easy class has especially good average precision. The difficult one has clearly worse results, but they are still quite acceptable. In the easy class, all desired images are obtained with just seven iterations. This is very close to the optimal value of five[1]. The difficult class has noticeably worse performance. The results from the three best features are almost as good as using all of them. On average 80% of the images are found with just 10 queries. The results of other classes fall between these two.

Another result can be seen in the lower row in Figure 3. It shows the amount of found images as the number of iterations increases. The results are again slightly better when using all features. This is opposite to the result from the KNN classification where the results were slightly better when only the three best features were used. This is due to the different feature weighting of KNN and PicSOM. In KNN all the feature sets have equal weights, i.e., they all contribute equally to classification. In PicSOM, however, these weights are adapted according to the queries so that PicSOM *learns* what feature sets are important in each query session. So in PicSOM 'bad' features won't much affect the retrieval whereas in KNN they may lower classification rates.

All of the graphs have one thing in common: their precision increases after the first couple of iterations. This is a very desired feature of PicSOM. This increase indicates that PicSOM is able to learn what kind of images the user was searching. It should be noted that this learning is done based only on the feedback information gathered during the queries.

We have also a larger database that has 13000 defect images. This database is unclassified so we have only been able to do visual testing. The performance of PicSOM and MPEG-7 features seems to be very good with the larger database, too. PicSOM's tree-based structure yields efficient indexing and querying engines, and thus the system works fast and efficiently. The queries are completed almost in real time, even though the database size is ten-fold. The query results also seem very sensible when examined by visual inspection. The query results are very similar to the desired images and the system adapts if the query target is modified during the search. These results suggest that the system retains a high level of success when used with large databases.

4 Conclusions

In this paper six different MPEG-7 visual descriptors were applied to paper defect image classification and retrieval. The experiments suggest that the descriptors, especially Homogeneous Texture and Color Structure, can be successfully used for these tasks. Their performance on these rather difficult paper defect images is surprisingly good. Of course, these tests do not guarantee the usability of MPEG-7 descriptors in the general case, but they imply that MPEG-7 descriptors are worth experimenting with.

[1] PicSOM returns 20 images per round so the minimum number of rounds required to obtain all desired images is five (since almost all our classes have 100 images) .

Acknowledgments. The authors wish to thank the PicSOM group (J. Laaksonen, M. Koskela, S. Laakso, E. Oja) 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. A. D. Bimbo. *Visual Information Retrieval.* Morgan Kaufmann Publ., Inc., 1999.
2. M. Bober. MPEG-7 visual shape descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), June 2001.
3. J. Iivarinen and J. Pakkanen. Content-based retrieval of defect images. In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, pages 62–67, Ghent, Belgium, Sept. 9–11 2002.
4. J. Iivarinen and A. 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, Aug. 16–20 1998.
5. J. Laaksonen, M. Koskela, S. Laakso, and E. Oja. Picsom - content-based image retrieval with self-organizing maps. *Pattern Recognition Letters*, 21(13-14):1199–1207, 2000.
6. J. Laaksonen, M. Koskela, S. Laakso, and E. Oja. Self-organising maps as a relevance feedback technique in content-based image retrieval. *Pattern Analysis and Applications*, 4(2+3):140–152, 2001.
7. B. S. Manjuanth, J.-R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), June 2001.
8. MPEG-7. MPEG-7 multimedia content description interface – part 3 visual. ISO/IEC JTC1/SC29/WG11 W3703, 2001.
9. MPEG-7. MPEG-7 visual part of the experimentation model (version 9.0). ISO/IEC JTC1/SC29/WG11 N3914, 2001.
10. Y. Rui, T. S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1):39–62, 1999.