47 PicSOM: Self-Organizing Maps for Content-Based Image Retrieval

Erkki Oja, Jorma Laaksonen, Markus Koskela, and Sami Brandt

Content-based image retrieval from unannotated image databases has been an object for ongoing research for a long period. Digital image and video libraries are becoming more common and growing in size as more visual information is produced at a rapidly increasing rate. The technologies needed for retrieving and browsing this accumulating amount of information are still, however, quite immature and inadequate for many practical applications. Many projects have been started in recent years to research and develop systems for content-based image retrieval, of which best-known being the Query By Image Content (QBIC) [1] developed at the IBM Almaden Research Center.

We have started to develop methods to utilize the strong self-organizing power of the Self-Organizing Map (SOM) [2] in unsupervised statistical data analysis to facilitate content-based retrieval from large image databases [4,5]. Our experimental image retrieval system is named PicSOM, bearing similarity to the WEBSOM document browsing and exploration tool. PicSOM uses a World Wide Web browser as the user interface and a hierarchical version of the SOM algorithm called Tree Structured Self-Organizing Map (TS-SOM) [3] as the image similarity scoring method. The TS-SOM is a tree-structured vector quantization algorithm that uses two-dimensional

Figure 105: WWW-based user interface of PicSOM. The user has previously selected five aircraft images. The system is displaying the user ten new images to select of.

SOMs at each of its hierarchical levels. The implementation of PicSOM is based on a general framework in which the interfaces of co-operating modules are defined.

Figure 105 shows a screenshot of the current web-based PicSOM user interface, which can be found at *http://www.cis.hut.fi/picsom/*. A notable feature in PicSOM is its ability to use multiple reference images, while most other current systems are based on using a single reference image.

The retrieval approach in PicSOM is based on relevance feedback [6] adopted from traditional information retrieval techniques. In relevance feedback, implicit information of the human-computer interaction during previous queries is used to refine the response on subsequent rounds. The image queries are thus iteratively refined during the retrieval process, as the system exposes more images to the user and tries to adapt to the user's preferences regarding the similarity of images.

The system may use one or several types of statistical features for image querying. Separate feature vectors can be formed for describing, for example, the color content, various textures, and objects of the images. A separate TS-SOM is then constructed for each feature vector set and these maps are used in parallel to calculate the best-scoring similarity results. The feature selection is not restricted in any way and new features can be added to the system later on.

Combining the results from several maps can be done in a number of ways. A simple method would be to ask the user to enter weights for different maps and then calculate a weighted average. This, however, requires the user to give information which she normally does not have, as it is a difficult task to give low-level features such weights which would coincide with human's perception of images. Therefore, a better solution is to apply the relevance feedback approach, in which the results of multiple maps are combined automatically, using the implicit information from the user's responses during the query.

The rationale behind our approach is as follows: If the images selected by the user map close to each other on a certain TS-SOM map, it seems that the corresponding feature performs well on the present query and the relative weight of its opinion should be increased. This can be implemented by marking the images the user has seen on the maps. The units are given positive and negative values depending whether she has selected or rejected the corresponding images. The mutual relations of positively-marked units residing near each other can then be enhanced by convolving the maps with a simple low-pass filtering mask. As a result, areas with many positively marked images spread the positive response to their neighboring map units. The images associated with these units are then good candidates for next images to be shown to the user, if they have not been shown already.

Figure 106 shows a set of convolved feature maps during a query. The three images on the left represent three map levels on the TS-SOM for a RGB color feature, whereas the convolutions on the right are calculated on a texture map. The sizes of the SOM layers are 4×4, 16×16, and 64×64, from top to bottom.

The research will continue along several lines: To increase PicSOM's retrieval performance, we need to add better feature representations. These will include color histograms, color layout descriptions, shape features, and some more sophisticated texture models. As the PicSOM architecture is designed to be modular and expandable, adding new features is straightforward. We are also developing quantitative measures to compare the performance of different features and of PicSOM with

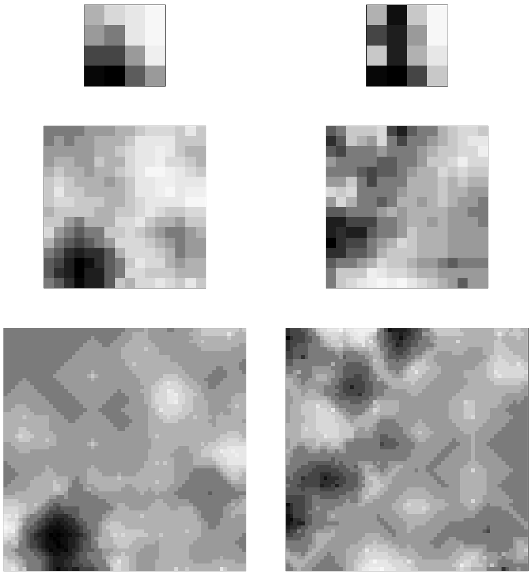

Figure 106: An example of convolved TS-SOMs for color (left) and texture (right) features. Black corresponds to positive and white to negative convolved values.

that of other content-based image retrieval systems. Quantitative measures of the image retrieval performance are, however, problematic due to human subjectivity. Generally, there exists no definite right answer to an image query as each user has individual expectations. Furthermore, to study our method's applicability on a larger scale we shall need larger image databases. A vast collection of images is available on the Internet, and we have preliminary plans to use PicSOM as an image search engine for the World Wide Web.

References

[1] M. Flickner, H. Sawhney, W. Niblack, et al. Query by image and video content: The QBIC system. *IEEE Computer*, pages 23–31, September 1995.

[2] T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer-Verlag, 1997. Second Extended Edition.

[3] P. Koikkalainen. Progress with the tree-structured self-organizing map. In A. G. Cohn, editor, *11th European Conference on Artificial Intelligence*. European Committee for Artificial Intelligence (ECCAI), John Wiley & Sons, Ltd., August 1994.

[4] J. Laaksonen, M. Koskela, and E. Oja. Content-based image retrieval using self-organizing maps. In *Third International Conference on Visual Information Systems*, Amsterdam, The Netherlands, June 1999.

[5] J. Laaksonen, M. Koskela, and E. Oja. PicSOM - a framework for content-based image database retrieval using self-organizing maps. In *11th Scandinavian Conference on Image Analysis*, Kangerlussuaq, Greenland, June 1999.

[6] Y. Rui, T. S. Huang, and S. Mehrotra. Relevance feedback techniques in interactive content-based image retrieval. In I. K. Sethi and R. J. Jain, editors, *Storage and Retrieval for Image and Video Databases VI*, volume 3312 of *Proceedings of SPIE*, pages 25–36, 1997.