Searching Technique in a Spectral Image Database

Markku Hauta-Kasari[1], Kanae Miyazawa[1],
Jussi Parkkinen[1], and Timo Jaaskelainen[2]

[1] Department of Computer Science, University of Joensuu,
P.O.Box 111, FIN-80101 Joensuu, Finland
{Markku.Hauta-Kasari, Kanae.Miyazawa, Jussi.Parkkinen}@cs.joensuu.fi
[2] Department of Physics, University of Joensuu,
P.O.Box 111, FIN-80101 Joensuu, Finland
Timo.Jaaskelainen@joensuu.fi
http://cs.joensuu.fi/spectral/

Abstract. We propose a technique for searching images from a spectral image database. In the current stage, the tec hnique uses spectral color information. The technique is based on the Self-Organizing Map (SOM), which is used to create the Best-Matching Unit (BMU) histograms for every spectral image in a database. Searching program calculates BMU-histogram for the wan ted spectral image, and then the histogram similarities between the wan ted BMU-histogram and the BMU-histogram database are calculated. The search result is shown as images that are ordered according to the histogram similarity. The searching technique and the current experimental results are reported.

1 Introduction

Spectral color representation has received a great deal of attention recently. Spectral imaging devices are under a rapid development and their use in color related applications is growing. The advan tages of the spectral color representation compared to the three-dimensional color representation, *e.g.* RGB, are as follows: 1) Spectral color av oids the problem of metamerism, in which the objects with similar color under one illumination may look different under another illumination. 2) The spectral measurement can be extended into ultra violet (UV) or infrared (IR) regions, which are interesting in many applications, for example, IR-region in the vegetation research.

The spectral color representation is evolving from traditional remote sensing into newer fields such as telemedicine, e-commerce, and electronic museums. Also the display devices that use more than three primary colors are under development. By combining the spectral imaging device and the multiprimary display, the degree of metamerism is becoming low er and the colors can be shown inside a larger color gamut than the gamut of three primary colors [1]. Recently, a new CIE T echnical Committeecalled "Division 8 TC 8-07: Multispectral imaging"

was established. The committee will define, for example, the standard data formats for spectral images. This also shows that the use of spectral imaging in high accurate color imaging applications is expanding.

Spectral images can contain tens or hundreds of spectral channels, depending on a spectral resolution needed. Typical numbers for channels are 31 and 61, when the spectra are measured from 400 to 700 nm, at 10 nm or 5 nm intervals, respectively. Table 1 shows the memory requirements for different type of image formats. It can be seen that the analysis of spectral information needs powerful computational methods.

Table 1. Memory requirements for different image formats.

Image type (8 bits / pixel)	256×256 pixels,	512×512 pixels
Gray level image	64 KB	256 KB
RGB image	192 KB	768 KB
Spectral image (10 nm resol.) [a]	2 MB	8 MB
Spectral image (5 nm resol.) [a]	4 MB	16 MB
Moving spectral image (10 s, 20 nm resol.) [a,b]	246 MB	983 MB

[a] 400-700nm

[b] 24 frames / second

There are only a few spectral image databases publicly available at the moment. However, the rapid development of spectral imaging systems is expected to bring more and more spectral images available. Fast methods for searching images in these databases will be needed in the future.

In this study, we propose a technique for searching images in a spectral image database. For RBG color image databases, there exists already techniques such as PicSOM [2] and QBIG [3]. However, in order to avoid the problem of metamerism and to preserve the measured spectral color information, we cannot convert the spectral images into RGB-system. Therefore, we are developing a new searching technique that can be used in spectral image databases. In the present stage we use spectral information and the texture features are under development.

The technique developed in this study uses Self-Organizing Map (SOM) [4] to generate the Best-Matching Unit (BMU) histograms for the database images. In searching technique, the wanted image is searched from the database according to the BMU-histogram similarity. The spectral images are ordered based on the BMU-histogram similarity and the ordered result is shown to the user. After creating BMU-histograms for the database, our technique for searching images is fast. As a test material we used 76 spectral images measured from real world objects.

The paper is organized as follows. In Section 2, we describe our spectral image database. The technique for searching spectral images from the database

is presented in Section 3. The experimental results are collected in Section 4 and Section 5 gives the discussion.

2   Spectral Image Database

We collected 76 spectral images for our database. The images were measured at the University of Joensuu, Finland, Chiba University, Japan, Saitama University, Japan, Lappeenranta University of Technology, Finland, Marine Biological Laboratory, Maryland, USA, and the University of Bristol, UK. The spectral images were filtered so that every image contains 61 spectral components from 400 nm to 700 nm, at 5 nm intervals. The spectral images contain objects such as plants, forest, corals, fruits, GretagMacbeth colorchecker, skin, printed magazine pictures, and paintings. Some images have clear background, for example, black, and some images have more homogenous color. The smallest spatial pixel resolution is 40×160 pixels and the largest 508×764 pixels.

The filtered images were saved as Matlab binary mat-format, as three-dimensional matrices, in which there are two spatial dimensions and a spectral dimension. The total size of the spectral image database is over 3 gigabytes. We developed the data structure for the database for indexing images. To add images to our data structure is convenient afterward.

3   Searching Technique

Our technique for searching images in a spectral image database is shown in Figure 1, in which the numbers correspond to the order of procedure. We first select spectra randomly from each spectral image as a training data for SOM. The amount of spectra to be selected was chosen empirically. The size of the map is much smaller than the number of spectra used as training data. Then the one-dimensional SOM is trained. After training, the map contains spectra that has a self-organized order [5]. For each pixel in a spectral image, we calculate the BMU from the map. BMU is defined as a winner unit w_c for the input vector x: $\|x - w_c\| = \min_i \|x - w_i\|$. This results the BMU-image where each pixel value is scalar, $i.e$ the index number i of the map unit. Using these indices, we calculate the BMU-histogram, which tells about the distribution of the spectra in a spectral image. The BMU-histogram is normalized by the number of pixels in an image. This is done for every spectral image in the spectral image database and finally the BMU-histograms for the whole database are saved to a file. This BMU-histogram database is convenient to use to represent the spectral image database, because, for example, for 76 spectral images and for the map having 50 units, the size of the histogram database matrix is 76×50.

The search in the spectral image database is done as follows. The user selects wanted spectral image. The BMU-image is generated using the same map that was used to generate the BMU-histogram database. Then BMU-histogram is formed. Next, BMU-histogram of the wanted image and the BMU-histogram database are compared. In this study, we compare directly the BMU-histograms

by calculating the difference componentwise and calculating the sum of the differences. If the wanted image is included in a database, the difference is 0 and the other images are ordered based on the difference value. Finally the results are shown as RGB images to a user. Note that the results can be also shown using a multiprimary color display.
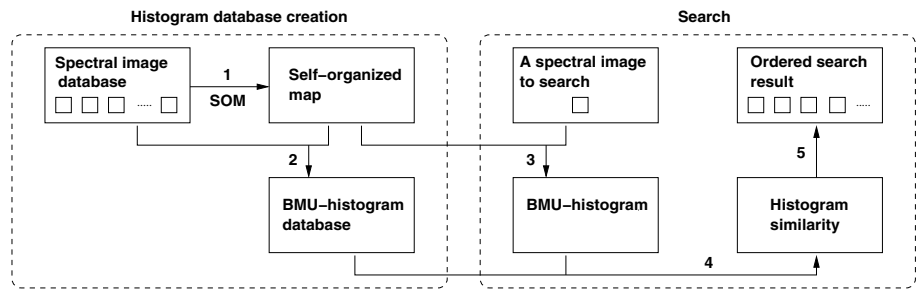

Fig. 1. Dataflow of the searching technique.

4   Experiments

We programmed the technique in a Matlab-software environment. The SOM calculations were done using a SOM_PAK-software package [6] and a SOM toolbox for Matlab [7]. First, 10000 spectra from each 76 spectral image were selected randomly as a training data for SOM. Total number of training samples was 760000 spectra with 61 components in each spectrum. The one-dimensional SOM with 50 units was trained using 400000 epochs in order phase with 0.9 as a learning rate. Then 1600000 epochs were used in the fine tuning phase with 0.02 as a learning rate. The number of randomly selected spectra and the number of units in map were selected empirically. The resulted map is shown in Figure 2 a). The positions of the map units in CIELAB-color coordinate system are shown in Figure 2 b).

The map shown in Figure 2 a) shows the spectrum corresponding to each unit in a map.The order of the spectra in CIELAB-system in Figure 2 b) is discussed in Ref. [5] in detail. Finally, the BMU-image and BMU-histogram for each spectral image in database were created, and the BMU-histogram database was saved to a file.

Next, we show an example of the search in a spectral image database. We selected a spectral image of skin (palm) as a wanted image. The BMU-image and the BMU-histogram were created using the same self-organized map that was used to make BMU-histograms for the database. Figure 3 shows the BMU-image and its BMU-histogram.

The histogram similarities were calculated and the spectral images were ordered according to the similarity. The ordered output for 76 spectral images is
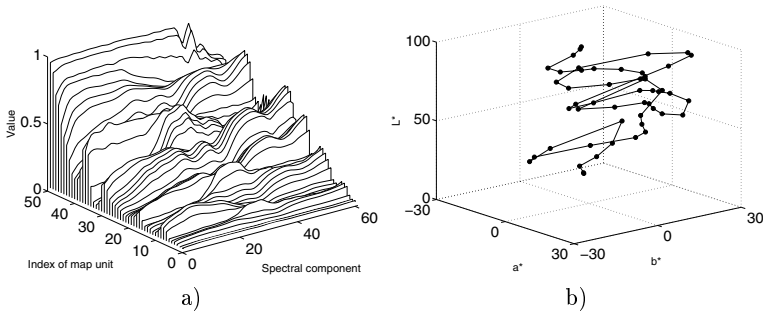
a)   b)

Fig. 2. a) Self-organized map, b) positions of map units in CIELAB-color coordinate system.

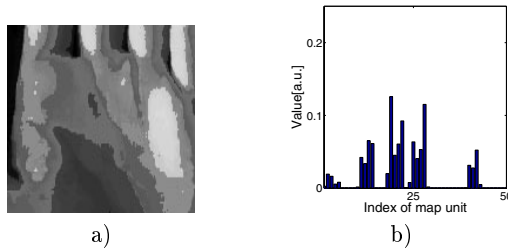

a)   b)

Fig. 3. a) BMU-image and b) BMU-histogram of palm spectral image.

shown as RGB images in Figure 4. Figure 5 shows the histogram differences for the database of 76 spectral images. The image indices are ordered corresponding to the histogram difference.

To approximate the computational efficiency of the method for a larger database, we simulated this situation as follows. We generated synthetically 1000 spectral images with the size of 87×147 pixels. The total size of the training data for SOM was 1 GB and the time required to train the SOM was 30 minutes. After the training of SOM, the histogram database file was the size of 400 KB. The time required to search images from the database took 1 second.

5   Discussion

We presented a searching technique in the spectral image database. The main advantages of this technique are the following: after generating the histogram database, the searching technique is fast and it preserves the spectral color information. The use of SOM to quantize and order the spectra in database was motivated by the fact that SOM orders the spectra highly corresponding to the lightness value (L^*) in CIELAB-system [5]. BMU-histograms have also connection to lightness based order, which means that the histograms can be later
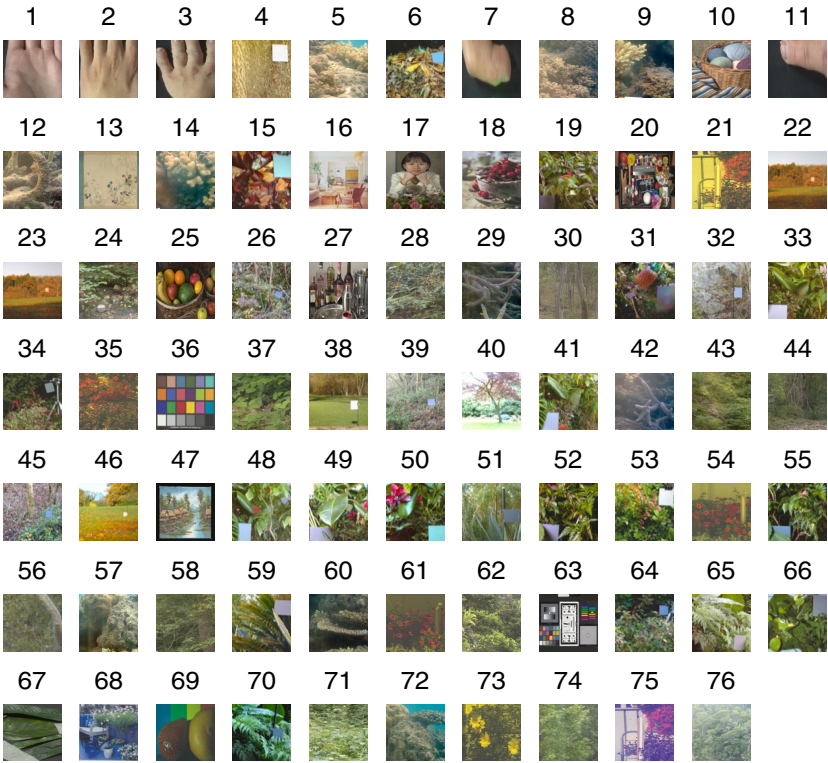
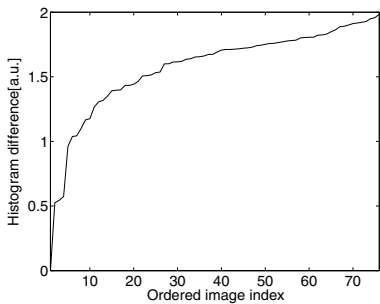**Fig. 4.** Ordered output of the search.

Fig. 5. Ordered histogram differences for 76 images.

analyzed using histogram features defined for gray level images. Next step in this research is to test more features for BMU-histogram similarity calculation and to add textural features to the technique.

References

1. Ajito, T., Obi, T., Yamaguchi, M., Ohyama, N.: Six-primary color projection display for expanded color gamut reproduction. In Proceedings, International Symposium on Multispectral Imaging and Color Reproduction, Chiba, Japan, October 21-22 (1999) 135–138
2. Laaksonen, J., Koskela, M., Laakso, S., Oja, E.: Self-Organizing Maps as a Relevance Feedback Technique in Content-Based Image Retrieval. Pattern Analysis & Applications **4** (2001) 140–152
3. QBIC - IBM's Query By Image Content: http://wwwqbic.almaden.ibm.com/ (January 20, 2003)
4. Kohonen, T.: Self-Organizing Maps. 3rd extended edn. Springer, Berlin Heidelberg New York (2001)
5. Miyazawa, K., Hakkarainen, J., Parkkinen, J., Jaaskelainen, T.: Ordering of color spectra for digital image enhancement. In Proceedings, International Congress of Imaging Science (ICIS'02), Tokyo, Japan, May 13-17, (2002) 492–493
6. Kohonen, T., Hynninen, J., Kangas, J., Laaksonen J.: SOM_PAK: The Self-Organizing Map Program Package. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150 Espoo, Finland (1996)
7. SOM Toolbox, a software package for Matlab 5: http://www.cis.hut.fi/projects/somtoolbox/ (January 20, 2003)