Journal of Imaging Science and Technology® 52(2): 020201–020201-11, 2008.

Distance Measures in the Training Phase of Self-Organizing Map for Color Histogram Generation in Spectral Image Retrieval

Oili Kohonen and Markku Hauta-Kasari
Department of Computer Science and Statistics, University of Joensuu, 80101 Joensuu, Finland
E-mail: oili.kohonen@cs.joensuu.fi

Abstract. *The usefulness of different distance measures in the training phase of self-organizing map (SOM) for color histogram generation for spectral image retrieval purposes is examined. The calculation of the best-matching unit (BMU) in the training phase of SOM is done by using Euclidean distance, Kullback–Leibler distance, Jeffrey divergence, and CIEL*a*b* color difference as distance measures. One-dimensional SOMs are generated for two different data sets consisting of 1269 Munsell color chips and 1, 440, 000 color spectra collected from a real spectral image database. The suitability of the introduced measures is first evaluated by calculating the average color differences between the Munsell data set and its BMUs in the SOMs trained by Munsell data. The achieved results are validated by a practical application, in which the queries from a real spectral image database are performed. Furthermore, the ability of SOMs trained by different distance measures to distinguish between spectral images of real human skin and magazine prints of human skin is examined. The achieved results are promising and indicate that two-dimensional self-organizing maps, which are trained by using Euclidean distance and Jeffrey divergence as distance measure and color histograms that correspond the spectral images as training data, could be used for classifying spectral images. © 2008 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.(2008)52:2(020201)]

INTRODUCTION

The benefits of spectral imaging have been utilized in many research areas and a variety of spectral imaging systems have been developed worldwide.[1–5] This has increased the amount of available spectral images and the conceivable problems with data formats are expected to be tackled by the ongoing standardization of spectral image formats.[6] Thus, both the amount and the size of existing spectral image databases are expected to increase and efficient retrieval methods for spectral images are therefore needed.

In conventional content-based image retrieval systems, such as MARS[7] and QBIC,[8] a color histogram is a widely used representation for color features. In a red-green-blue (RGB) image each pixel is represented by three values and a histogram representation is generated over a quantized color space. In spectral image, each pixel contains a color spectrum which is a power distribution of electromagnetic radiation. The dimension of spectra varies from tens to hundreds or even thousands, depending on the used application. There is no natural order for spectra but the use of self-organizing maps[9] enables to gain an order for multidimensional data.

Miyazawa et al.[10] ordered the color spectra by using self-organizing maps (SOMs), which have also been used for color histogram generation for spectral images in order to organize a spectral image database.[11] In the organizing method presented in Ref. 11, the spectral images were represented by best matching unit (BMU) histograms and the distances between these histograms were calculated by different distance measures, such as Euclidean distance (ED), Jeffrey divergence (JD)[12] and Kullback–Leibler distance (K–LD).[13] However, in both of these applications the used distance measure in a training phase of SOM was the conventional, widely used Euclidean distance (ED). In addition to ED, Huang et al.[14] evaluated two other distance measures for training SOM in respect to RGB color histogram generation. The measures were L_1-norm and Battacharyya distance which were used for both calculating the BMU and updating the map units. The results achieved indicated that the common ED may not yield the best results.

In Ref. 11 decent results were achieved by using ED, JD and K–LD for BMU-histogram comparison. In this article, the usefulness of JD, K–LD and the CIEL*a*b*[15] color difference ΔE^*_{ab} as distance measure in SOM training for spectral data in addition to conventional ED measure is examined.

SELF-ORGANIZING MAPS

The self-organizing map[9] algorithm is an unsupervised learning algorithm which defines a mapping from a high-dimensional input data space into a lower-dimensional space. SOM preserves the most important topological relationships and, therefore, the inputs that are located near to each other in a high-dimensional space are also located near to each other in a new, lower-dimensional space. SOM consists of arranged units, which are represented by weight vectors and are connected to adjacent units by a neighborhood relation, which dictates the topology of the map.

The initialization of the map units is done by either linear or random initialization. In random initialization the weight vectors $w_i(0)$ are constructed by using the arbitrary values, whereas in linear initialization the weight vectors are

Received Aug. 8, 2007; accepted for publication Jan. 8, 2008; published online Apr. 10, 2008.
1062-3701/2008/52(2)/020201/11/$20.00.

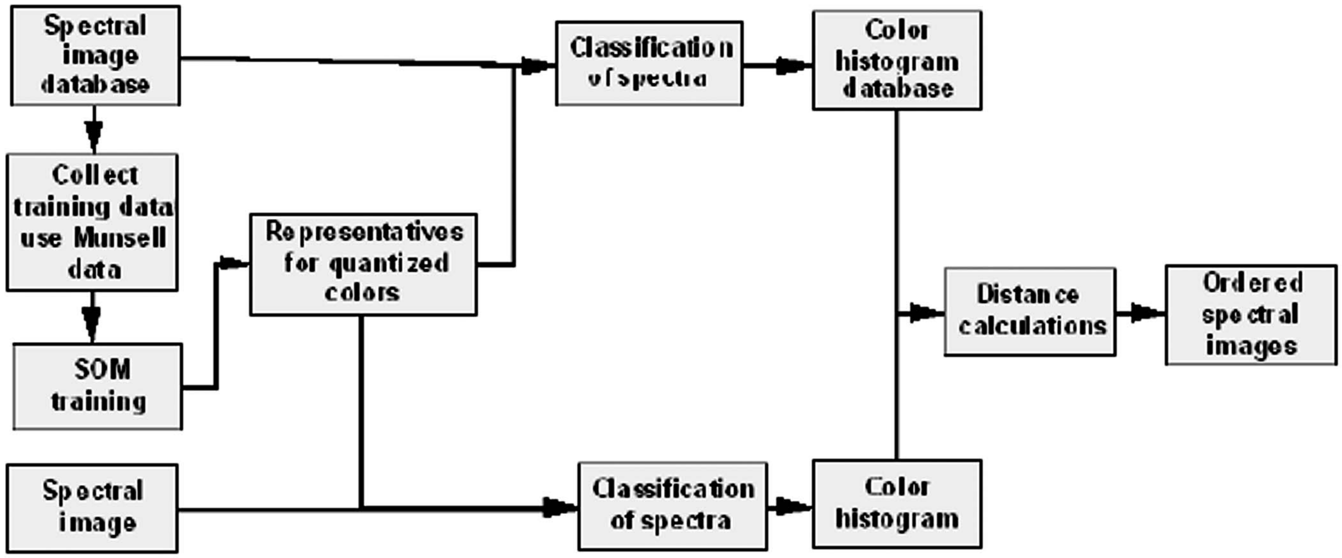

Figure 1. A diagram representing a query in a spectral image database.

represented by the points of rectangular array defined along the subspace spanned by two eigenvectors of the autocorrelation matrix of input data. The use of linear initialization orders the weight vectors and therefore, the training can be started directly from the convergence phase. In the case of random initialization, the weight vectors are unordered after initialization. However, the random initialization can be used because the algorithm orders the unordered weight vectors in the long run.[9]

After initialization phase, the map is trained iteratively as follows. One input data vector x is chosen from the training data at each training step. This input data vector is compared to each unit of the map and the distances between the input data vector and map units are calculated. The unit producing the smallest distance is called the best matching unit (BMU) for an input data vector x.

A typically used distance measure in a training phase is Euclidean distance. However, it is possible to apply many other distance measures, such as Kullback–Leibler distance and Jeffrey divergence. In the cases of these mentioned measures (ED, K–LD and JD) the best matching unit is defined mathematically for m-dimensional data as follows:

$$D_{\mathrm{ED}}(x,w_{\mathrm{BMU}}) = \min_i \sum_{j=1}^{m} (x(j) - w_i(j))^2, \tag{1}$$

$$D_{\mathrm{K\text{-}LD}}(x,w_{\mathrm{BMU}}) = \min_i \sum_{j=1}^{m} x(j)\log\left(\frac{x(j)}{w_i(j)}\right), \tag{2}$$

$$D_{\mathrm{JD}}(x,w_{\mathrm{BMU}}) = \min_i \sum_{j=1}^{m} \left(x(j)\log\left(\frac{x(j)}{H(j)}\right) + w_i(j)\log\left(\frac{w_i(j)}{H(j)}\right)\right). \tag{3}$$

In the above defined equations w_i and w_{BMU} indicate the weight vectors of i$^{\mathrm{th}}$ unit and the best matching unit, respectively. H is the mean histogram of the input data vector x and weight vector w_i,

$$H = \frac{x + w_i}{2}. \tag{4}$$

The color difference ΔE^*_{ab} for two color spectra is defined in CIEL*a*b* color coordinate system as Euclidean distance

$$\Delta E^*_{\mathrm{ab}} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}, \tag{5}$$

where ΔL^*, Δa^* and Δb^* correspond the differences in luminance, red/green and yellow/blue values, respectively. When using the spectral data as a training data for SOM, the CIEL*a*b* color difference can be used as a distance measure. In that case the best matching unit is defined for input data vector x as follows:

$$D_{\Delta E^*_{ab}}(x,w_{\mathrm{BMU}}) = \min_i \Delta E^*_{ab}(x,w_i). \tag{6}$$

For calculating the L^*, a^* and b^* values from color spectra one needs to define the illuminant under which the calculations are to be performed. According to CIE recommendations the calculations should be performed under either standard illuminant D_{65} or standard illuminant A whenever possible.

After finding the BMU, the weight vectors of the SOM are updated using the following equation:

$$w_i(t+1) = \begin{cases} w_i(t) + \alpha(t)[x(t) - w_i(t)], & \text{if } i \in N_{\mathrm{BMU}}(t) \\ w_i(t), & \text{otherwise} \end{cases} \tag{7}$$

in which t denotes the time. $N_{\mathrm{BMU}}(t)$ is a decreasing neighborhood around the BMU and $\alpha(t)$ is a decreasing learning rate, for which holds $0 < \alpha(t) < 1$. In the sequential training, the data are presented to the map one vector at a time and

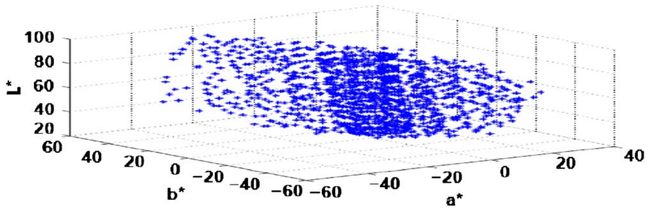

Figure 2. The used Munsell data in the CIEL*a*b* color coordinate system.

the weight vectors of the map are moved gradually towards to the training vectors. Despite applying different distance measures for finding the BMUs, the Euclidian distance based updating law is preserved in Eq. (7).

QUERYING METHOD FOR SPECTRAL IMAGES

In our querying method the spectral images are represented by color histograms constructed as follows. The training data for SOM algorithm is either generated by collecting 10,000 spectra randomly from each spectral image in a database or the Munsell data consisting of 1269 reflectance spectra of Munsell color chips[16] are used as a training data. Munsell data are well known for their uniform color space on trichromatic basis such as in CIEL*a*b*. The data samples the color space reflecting the properties of human color perception and has been used for many spectral applications such as such as color filter design for spectral image acquisition system,[17] studies on spectral spaces and color spaces,[18] the optimal sampling of color spectra[19] and the conversion between the reflectance spectra and Munsell data,[20] and surface reflectance estimation.[21]

The nap is initialized by the random initialization, after which the map is trained by the training data. After training phase, the weight vectors of the map units are considered to be representatives for quantized colors and the number of weight vectors in the trained map corresponds to the number of bins in color histograms to be generated as follows. Each pixel in a spectral image is classified to that quantized color, whose representative produces the smallest distance. A histogram over the quantized colors is generated and normalized by the number of pixels in an image. This is done

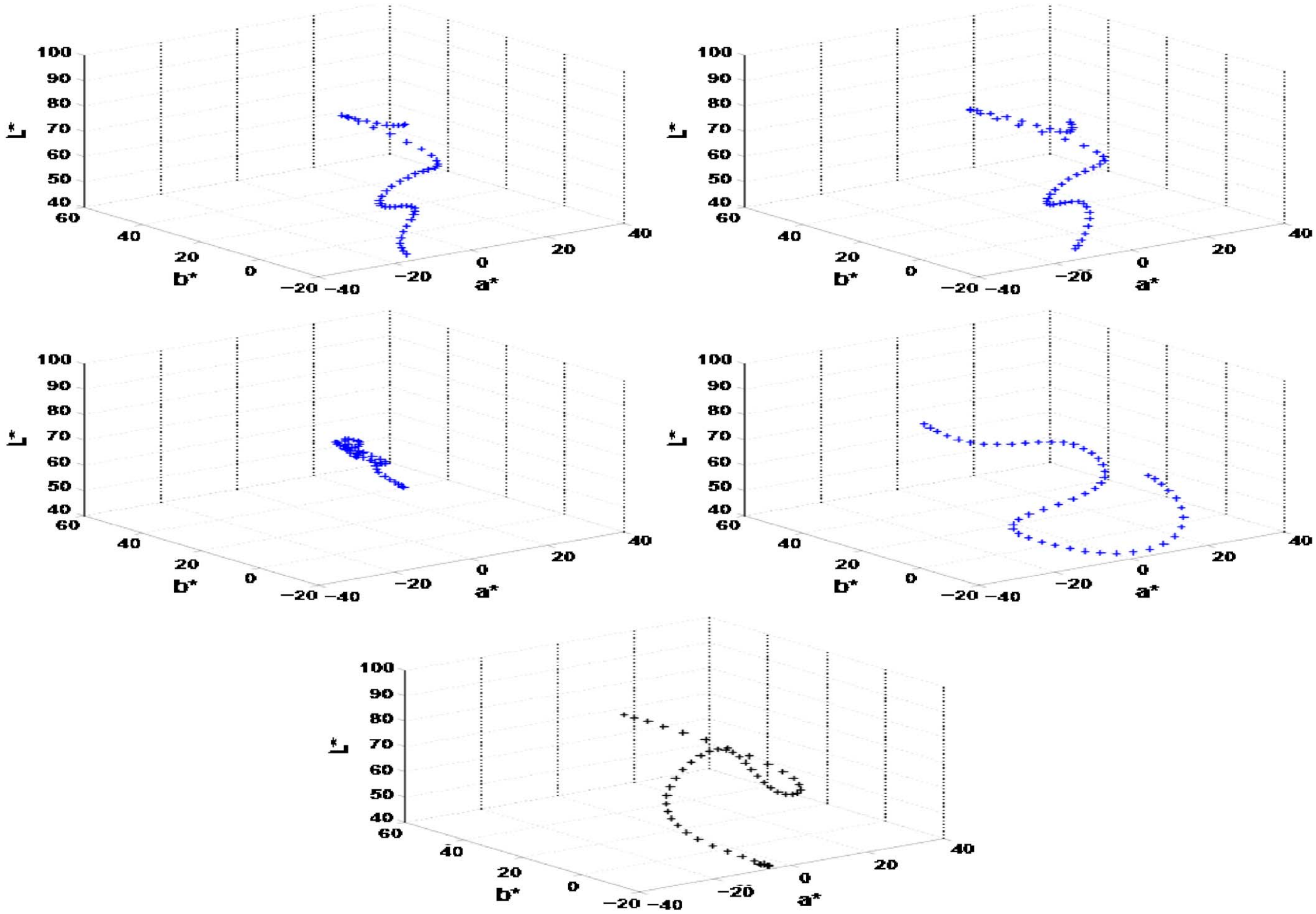

Figure 3. The weight vectors of one-dimensional, Munsell data trained maps in CIEL*a*b* color coordinate system. The used training measures from left to right and from top to down are ED, JD, K–LD, ΔE^*_{ab} under illuminant D_{65} and ΔE^*_{ab} under illuminant A. The calculations for transforming the weight vectors into CIEL*a*b* color coordinate system have been performed for CIE 1931 two degree standard colorimetric observer under illuminant D_{65}.

Table I. Average, maximum and minimum color differences between the Munsell data samples and the representatives of that quantized colors in the cases of one-dimensional SOMs.

Training by Euclidean distance						
	Under illuminant D_{65}			Under illuminant A		
Classification by	ΔE^*_{ave}	ΔE^*_{max}	ΔE^*_{min}	ΔE^*_{ave}	ΔE^*_{max}	ΔE^*_{min}
ED	18.7	58.6	0.5	17.6	49.7	1.0
JD	18.5	59.3	0.5	17.4	49.8	1.0
K-LD	30.9	83.4	3.6	31.2	86.8	2.8
ΔE^*	16.5	50.5	0.5	15.1	44.6	1.0
Training by Jeffrey Divergence						
	under illuminant D_{65}			under illuminant A		
Classification by	ΔE^*_{ave}	ΔE^*_{max}	ΔE^*_{min}	ΔE^*_{ave}	ΔE^*_{max}	ΔE^*_{min}
ED	18.1	60.8	1.4	17.0	53.5	0.3
JD	17.8	59.5	1.4	16.9	55.1	0.3
K-LD	30.6	85.4	3.8	30.7	82.6	4.1
ΔE^*	15.8	51.0	1.3	14.7	44.2	0.3
Training by Kullback-Leibler Distance						
	under illuminant D_{65}			under illuminant A		
Classification by	ΔE^*_{ave}	ΔE^*_{max}	ΔE^*_{min}	ΔE^*_{ave}	ΔE^*_{max}	ΔE^*_{min}
ED	28.1	57.4	2.2	28.1	59.8	0.9
JD	27.5	57.3	2.3	27.4	59.8	1.6
K-LD	30.0	76.7	2.3	30.4	74.9	1.6
ΔE^*	26.0	54.4	0.8	25.8	59.8	0.9
Training by CIEL*a*b* color difference						
	under illuminant D_{65}			under illuminant A		
Classification by	ΔE^*_{ave}	ΔE^*_{max}	ΔE^*_{min}	ΔE^*_{ave}	ΔE^*_{max}	ΔE^*_{min}
ED	17.8	62.3	0.5	17.4	61.2	0.5
JD	17.4	56.6	0.5	17.2	59.8	0.5
K-LD	31.0	89.2	0.5	32.8	96.1	4.4
ΔE^*	14.2	47.1	0.5	14.0	40.8	0.5

for each spectral image in a spectral image database and the generated histograms are saved into a histogram database.

The query in a spectral image database is done as follows. A desired spectral image is selected to be a query image. Images spectra are classified to the quantized colors whose representatives are the weight vectors of the same map that was used for the generation of the existing histogram database. The distances between the histogram of the query image and the histograms in the histogram databases are calculated and the images in the spectral image database are ordered by these distances. A desired amount of images closest to the used query image are shown to the user as ordered output images in RGB format. A diagram of the querying method is shown in Figure 1.

EXPERIMENTAL RESULTS

Altogether ten one-dimensional SOMs were trained in the performed experiments. Half of the maps were trained by using the data collected from the spectral image database and the other half was trained by the Munsell data, which have been used for many spectral applications such as color filter design,[17] studies on spectral spaces and color spaces,[18] finding the optimal sampling of color spectra,[19] and for conversion between the reflectance spectra and Munsell data,[20] and surface reflectance estimation.[21] Furthermore, five introduced distance measures: Euclidean distance (ED), Jeffrey divergence (JD), Kullback–Leibler distance (K–LD), and CIEL*a*b* color difference (ΔE^*_{ab}) under standard illuminants A and D_{65} were used for finding the best matching unit at each training step. In addition to one-dimensional SOMs, three two-dimensional SOMs were trained by using ED, JD and K–LD as distance measures for finding the best matching units during the training phases. The training data for these SOMs consisted of those corresponding histogram databases, which were generated by using the weight vectors of the earlier trained one-dimensional SOMs as representatives for quantized colors. Each of these histograms corresponded to one spectral image and therefore the training data for two-dimensional SOMs consisted of only 144 samples.

The performed calculations were based on the use of

Figure 4. The used spectral image database in a RGB format.

the SOM toolbox for Matlab.[22] The initializations of the maps were done by using the random initialization. To avoid the differences caused not by the training, but the initialization phase, the same random seed was used for each map. In all cases, the neighborhood function was hexagonal and learning rate linear. All maps were trained by using the sequential algorithm, in which the weight vectors are updated after each training sample. The amounts of the used training steps were 1,000,000 and 500,000 in the cases of one- and two-dimensional SOMs, respectively.

One-dimensional SOMs Trained by Munsell Data

In the first part of the experiments, the one-dimensional SOMs consisting of 50 units were trained by using the reflectance spectra of 1269 Munsell color chips[16] as a training data. The used distance measures in the training phases were ED, JD, KL–D and CIEL*a*b* color difference ΔE^{*}_{ab}, which was implemented under both standard illuminants A and D_{65}. The used Munsell data and the weight vectors of the generated maps are presented in CIEL*a*b* color coordinate system in Figures 2 and 3. In these cases the calculations were performed for CIE 1931 two degree standard colorimetric observer under illuminant D_{65}.

After training of the maps, each sample of Munsell data is classified to the closest quantized color. The classification for each map is done in respect to distance measure used during the training phase of the map. If the map was trained by using ED as distance measure, the distances between the Munsell data and representatives of the quantized colors were calculated in respect to ED and so on. After classification, the CIEL*a*b* color differences between each sample of

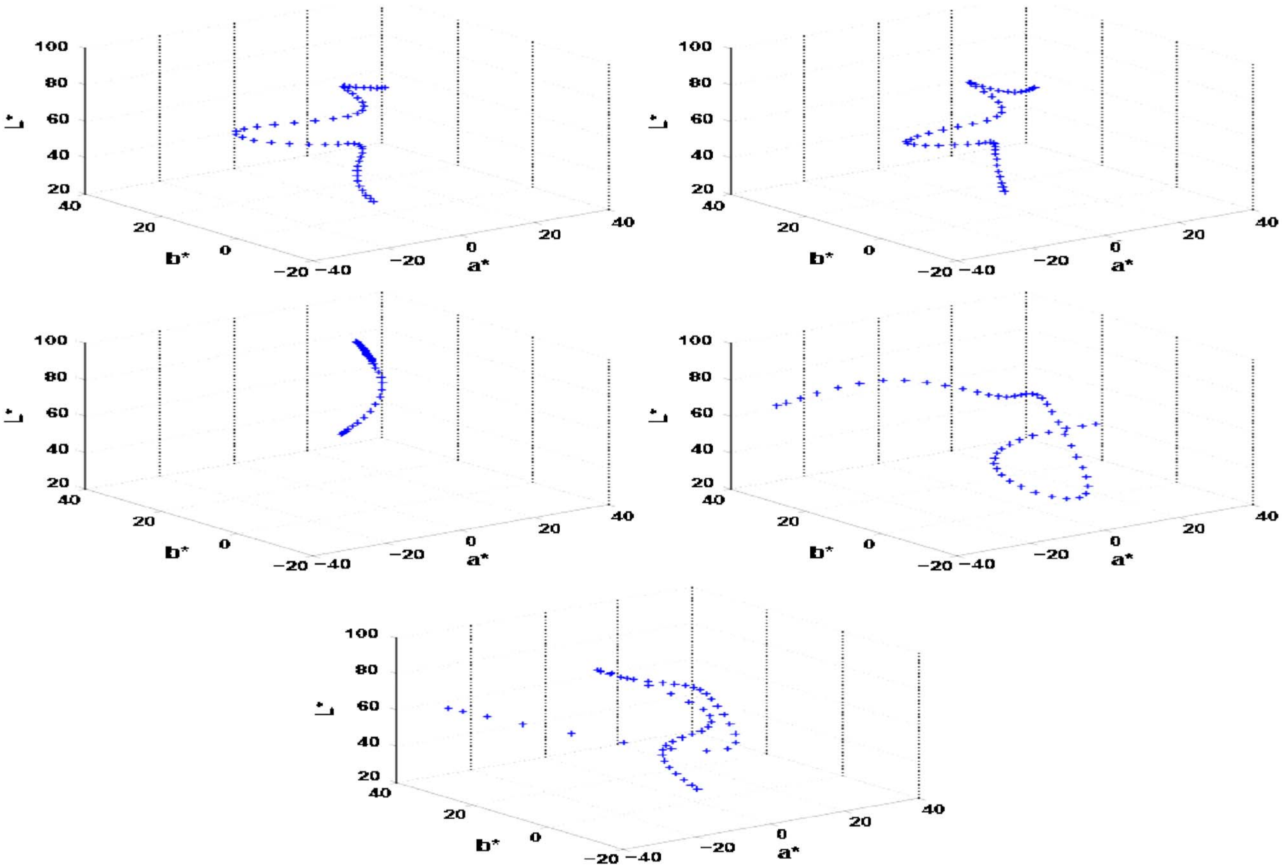

Figure 5. The weight vectors of one-dimensional maps in CIEL*a*b* color coordinate system for data collected from spectral image database. The training measures from left to right and from top to down are ED, JD, K–LD, ΔE^*_{ab} under illuminant D_{65} and ΔE^*_{ab} under illuminant A.

the classified data and the representative of the color they were classified into were calculated. In the cases of ED, JD and K–LD trained maps, the color difference calculations were performed under both illuminant A and illuminant D_{65}. For the maps trained by ΔE^*_{ab} under illuminant A and illuminant D_{65} the color difference calculations were performed under illuminants A and D_{65}, respectively. The calculations were performed for CIE 1931 standard observer and the achieved results are shown in Table I. The results achieved by using the ED and JD trained maps are quite similar in each case. The K–LD performs miserably all the time whereas the ΔE^*_{ab} gives the smallest errors. When the ΔE^*_{ab} is used as a distance measure in the classification the average color difference between the Munsell data and the constructed SOMs is lowered with 1.5–3 units, depending on the used distance measure in the training phase and the used illuminant in the color difference calculations.

One-dimensional SOMs Trained by Data Collected from Spectral Image Database

The second part of the experiments was based on the use of the spectral image database of Joensuu University.[23] Altogether 1,440,000 spectra were collected randomly from 144 spectral images, which were filtered into equal format: 59 spectral components with the spectral range from 410 to 700 nm at 5 nm intervals. The used database contained both natural spectral images and synthetical images generated by the virtual texture coloring technique.[24] The used database is shown in a RGB format in Figure 4.

The collected spectra were used as a training data for five different one-dimensional SOMs which were trained by using the introduced measures for finding the BMUs during the training phases of SOMs. The weight vectors of the generated maps are presented in CIEL*a*b* color coordinate system in Figure 5. The generated maps were furthermore used for generating the color histograms for the querying method, introduced in chapter 3 and described more carefully in Ref. 11. Histograms for each spectral image in the database were generated by using these five maps and five different histogram databases were obtaned as a result. Examples of ordered outputs for spectral images in the cases of ED, JD, K–LD, ΔE^*_{ab} under illuminant D_{65} and ΔE^*_{ab} under illuminant A-trained SOMs are presented in Figures 6 and 7. The used query image is shown on top of the output rows and the first output image is the leftmost image of the output in each case. The number of output images is restricted to nine.

In Fig. 6 the reflectance images of two hands, whose color differs from each other a bit, are used as query images whereas in Fig. 7 the used query image is a reflectance image of a magazine print of human skin. The results are evaluated

Figure 6. The ordered outputs for spectral images when the used dissimilarity measures in training phases from top to down are ED, JD, K–LD and ΔE^*_{ab} under illuminant D_{65} and ΔE^*_{ab} under illuminant A.

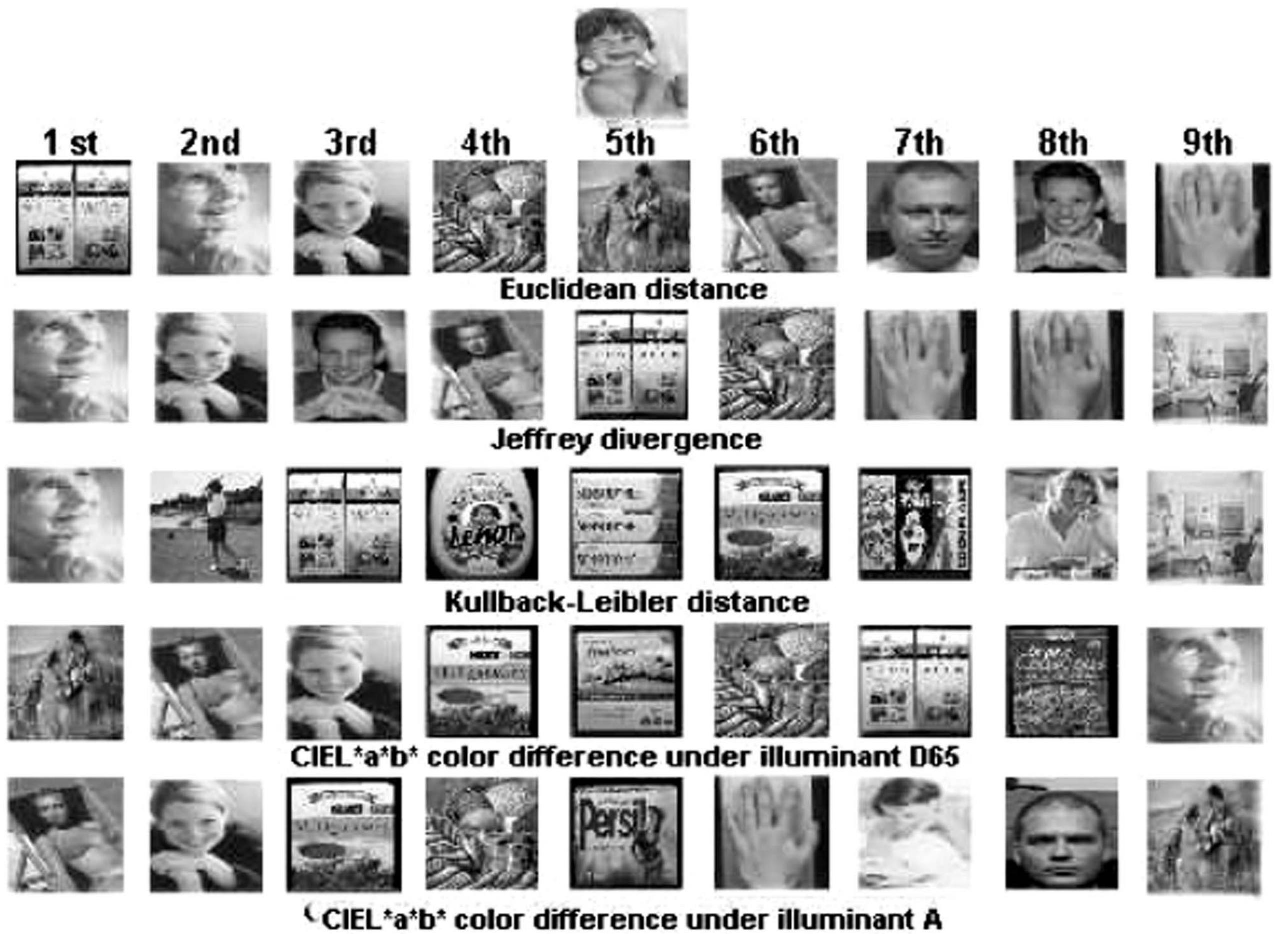

Figure 7. The ordered outputs when the used dissimilarity measures in training phases from top to down are ED, JD, K–LD and ΔE^*_{ab} under illuminant D_{65} and ΔE^*_{ab} under illuminant A.

visually by humans—the more skin containing images that one obtained as output the better the query output is. In the cases of natural skin samples the best results are achieved by using ΔE^*_{ab}-trained SOMs. When the magazine print of human skin is used as a query image the ΔE^*_{ab}-trained SOMs are not that good anymore and the best results are achieved by using SOM trained by JD as the distance measure. This can be explained by the different spectral properties of printed skin samples compared to previously used natural skin samples. An example of spectral power distributions of natural skin sample and printed skin sample can be seen in Figure 8. The same figure also includes the spectral radiances of the illuminants A and D_{65}. In general, the performance of the K-LD-based map was the weakest, as expected, based on the distribution of maps' weight vectors in the CIEL*a*b* color coordinate system.

Classification of Spectral Images

The last part of the experiments consisted of spectral image classification by histogram-trained SOMs. Three two-

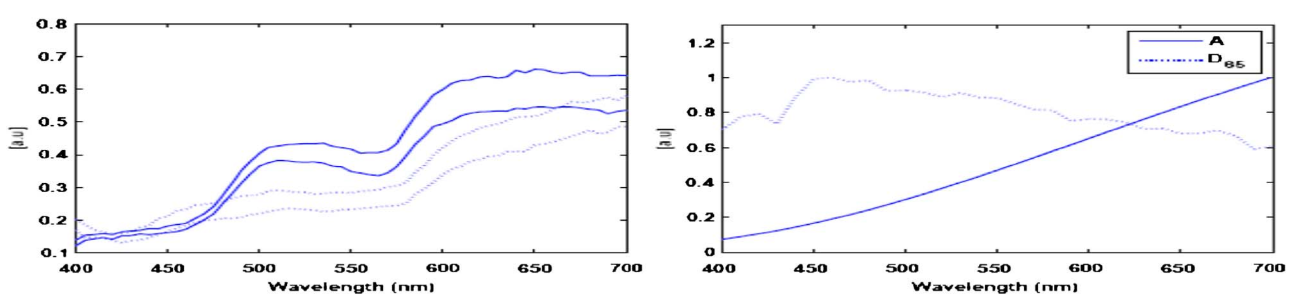

Figure 8. Left: Reflectance spectra of real human skin (dotted lines) and magazine print of human skin (solid lines). Right: Spectral radiances of illuminant A (solid line) and D_{65} (dotted line).

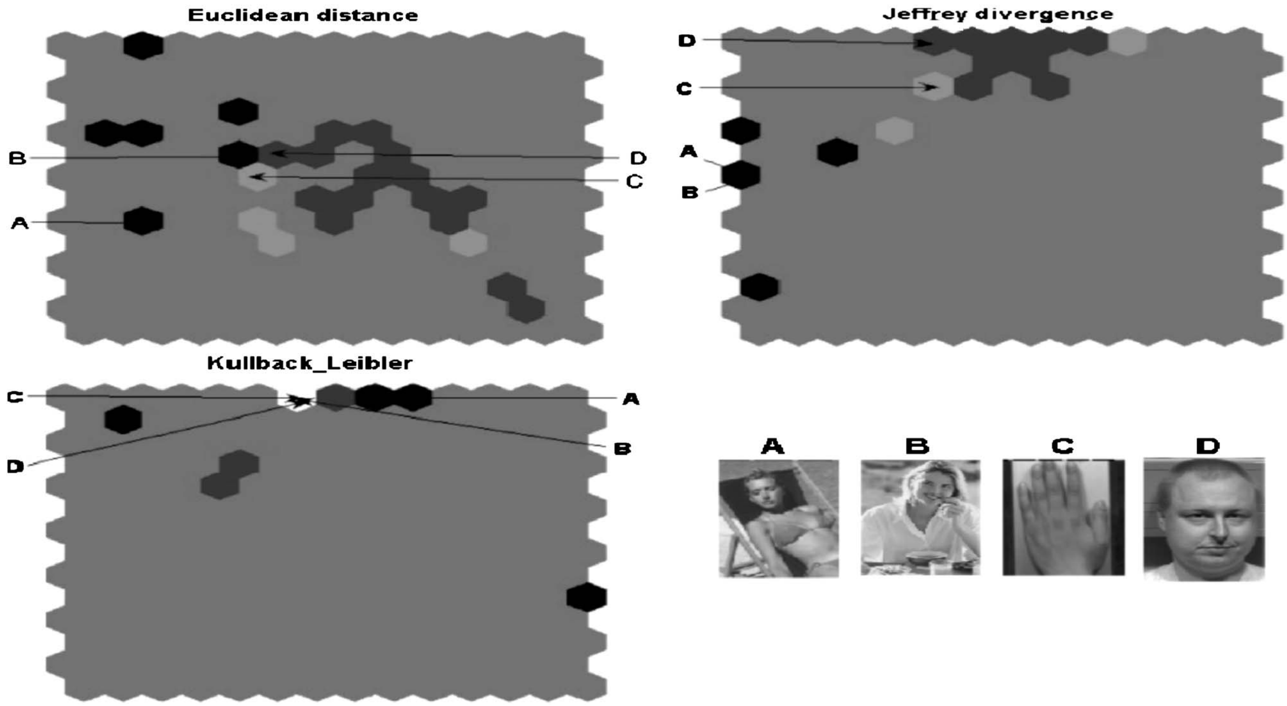

Figure 9. The placements of spectral images in the generated maps. The black, light gray and dark gray nodes represent the images of printed skin, real hands and real faces, respectively. The white color represents the node into which some images from each group were placed. The maps are trained by using ED (top left), JD (top right) and K–LD (down left) as distance measures.

dimensional (14×14 units) SOMs were trained by using the earlier calculated histogram databases as training data. The CIEL*a*b* color difference ΔE^*_{ab} can be used properly only for color spectra. Therefore, in this case the SOMs were generated by using only ED, JD and K-LD as distance measures. Moreover, the used histogram databases were the ones described in the previous section. The histogram databases were generated by using the one-dimensional SOMs which were trained by the data collected from the spectral image databases.

After the map generation, the placements of spectral images in these two-dimensional SOMs were found by calculating the final BMUs for the histograms corresponding the images. The placements in the SOMs were examined more carefully for three different groups of images. The first group included images of magazine prints of human skin whereas the second and the third groups consisted of images of real human hands and faces, respectively. These images were placed into the generated maps and the placements can be seen in Figure 9. The maps are generated by using ED (top left), JD (top right) and K-LD (down left) as distance measures and the "colored" nodes in the map represent the placements of the images. The black, light gray and dark gray nodes correspond to printed skin, real hands and real faces, respectively.

More than one image may be placed into the same map unit and, therefore, some changes exist in the amounts of colored nodes between different maps. However, in the cases of maps trained by ED and JD as distance measures there exist no map units, which would include mappings from two different groups of images. In the case of K-LD there exists one map unit into which some images from all three groups are placed. This map unit is indicated by white color.

The histograms and the relationships between the placements of four different images, referred to as A, B, C and D, in these three maps, are shown in Figs. 9 and 10, respectively. In Figure 10 the rows from top down correspond images referred as A, B, C and D. The ordering of maps used to generate the color histograms is from left to right ED, JD, K-LD, and ΔE^*_{ab} under illuminant D_{65}. The color histograms generated by using the maps trained by color difference ΔE^*_{ab} under illuminant D_{65} as distance measure have not been place into the maps but these histograms are still shown in Fig. 10 to allow the comparing of the histogram shapes.

DISCUSSION AND CONCLUSIONS

The possibility of using the different distance measures in the training phases of self-organizing maps for color histogram generation for spectral image retrieval purposes was examined. One-dimensional self-organizing maps for both Munsell data and data collected from a real spectral image database were trained by using Euclidean distance (ED), Jeffrey divergence (JD), Kullback–Leibler distance (K–LD) and CIEL*a*b* color difference ΔE^*_{ab} under CIE standard illuminants A and D_{65} as distance measures in the training phases. The CIEL*a*b* color differences between the Munsell data and SOMs trained by Munsell data were calculated for two degree standard colorimetric observer under illuminants A and D_{65}. The smallest average color difference was achieved by using the color difference under illuminant

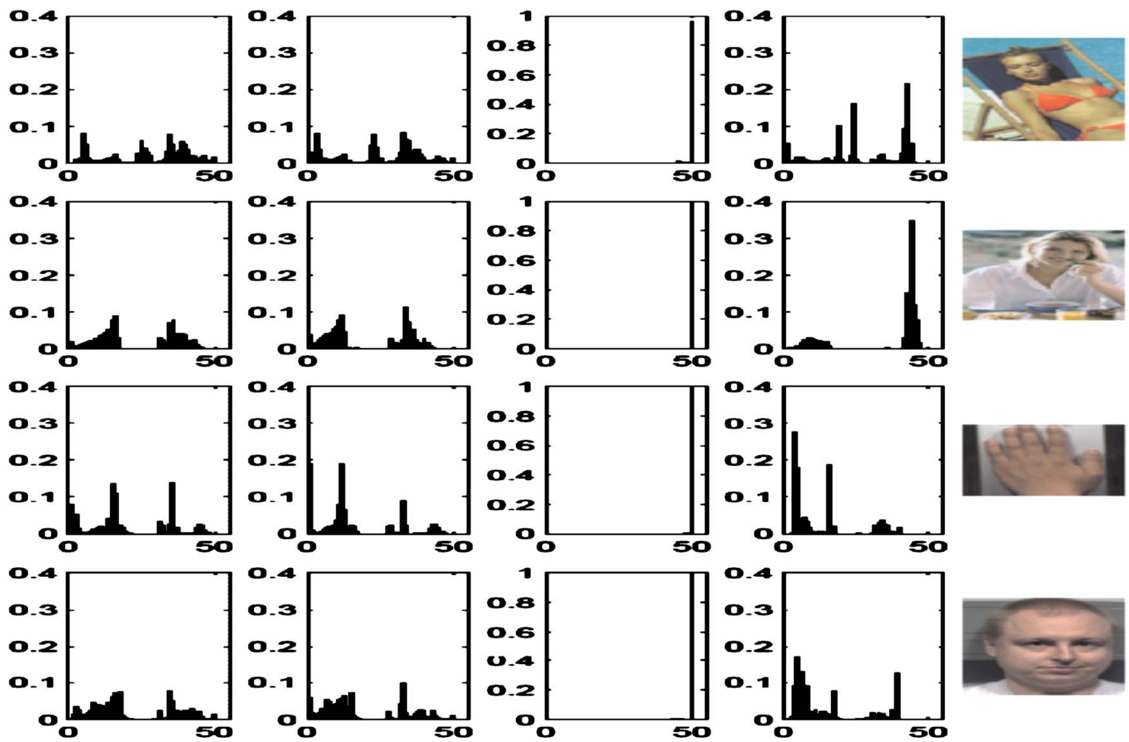

Figure 10. Histograms of four spectral images constructed from self-organizing maps trained by using ED, JD, K–LD and color difference ΔE^*_{ab} under illuminant D_{65} as distance measures.

A as a distance measure in a training phase of SOM. By ordering the measures to an ascending order by the achieved color differences, the following result was obtained K–LD, ED, JD, ΔE^*_{ab} under illuminant D_{65} and ΔE^*_{ab} under illuminant *A*. Compared to other measures both ΔE^*_{ab} measures were clearly better.

The achieved results were confirmed in practical application by implementing the queries from a spectral image database. In respect to the used query images, the most meaningful outputs were achieved, when the maps were trained by ΔE^*_{ab} as distance measure instead of conventional ED. The results indicate that CIEL*a*b* color difference ΔE^*_{ab} might be a useful distance measure in a training phase of self-organizing map when the algorithm is used to organize the color spectra in the visible wavelength for spectral image retrieval. However, in the case where the CIEL*a*b* color difference ΔE^*_{ab} is used as a distance measure, the color quantization is not anymore done in the spectral space, and this somehow neglects the information content which is included in the spectral data only.

The spectral classification of both natural and artificial samples was done by using two-dimensional SOMs whose training was performed by using ED, JD and K–LD as distance measures. The histograms generated by using one-dimensional SOMs trained by the corresponding distance measures were used as the training data. In the case of maps trained by using ED and JD as distance measures, none of the map units contained images from two different groups and, in general, the images from different groups were placed into different parts of the map. In the case of ED-trained map, the placements within the image groups were more sparse compared to the map trained by JD. In the case of K–LD-trained maps the image classification did not really succeed and some images from all different groups were placed into the same map unit. The reason can be seen in Fig. 10, in which the histograms for each of the example images are very similar and most of the images spectra have been classified into the same quantized color in the case of K–LD. In this approach the first quantization was done in spectral space so the information content available in spectral data was fully utilized in the performed color quantization.

REFERENCES

[1] H. Laamanen, K. Miyata, M. Hauta-Kasari, J. Parkkinen, and T. Jaaskelainen, “Imaging spectrograph based spectral imaging system,” *Proceedings of Second European Conference on Color in Graphics, Imaging, and Vision, and Sixth International Symposium on Multispectral Color Science -CGIV 2004*, Aachen, Germany (IS&T, Springfield, VA, 2004) pp. 427–430.

[2] A. Ribes, H. Brettel, F. Schmitt, H. Liang, J. Cupitt, and D. Saunders, “Color and multispectral imaging with the CRISTATEL muitispectral system,” *Proc. of PICS Conference, Rochester, NY* (IS&T, Springfield, VA, 2003) pp. 215–219.

[3] H. Haneishi, S. Miyahara, and A. Yoshida, “Image acquisition technique for high dynamic range scenes using a multiband camera,” Color Res. Appl. **31**, 294 (2006).

[4] M. Hauta-Kasari, K. Miyazawa, S. Toyooka, and J. Parkkinen, “Spectral vision system for measuring color images,” J. Opt. Soc. Am. A **16**, 2352 (1999).

[5] G. D. Finlayson, S. D. Hordley, and P. Morovic, “Using the SpectraCube to build a multispectral image database,” *Proc. Second European Conference on Colour in Graphics, Imaging and Vision*, Aachen, Germany. (IS&T, Springfield, VA, 2004) pp. 268–273.

[6] T. Jetsu, P. Hertzog, T. Jaaskelainen, and J. Parkkinen, “Standardization of spectral image formats,” Pattern Recogn. Image Anal. **15**, 618 (2005).

[7] S. Mehrotra, Y. Rui, M. Ortega-Binderberger, and T. S. Huang, “Supporting content-based queries over images in MARS,” *Proceedings*

of Int. Conf. Multimedia Computing and Systems, Ottawa, Ontario, Canada (IEEE, Piscataway, NJ, 1997) pp. 632–633.

8 M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: the QBIC system," Computer **28**, 23 (1995).

9 T. Kohonen, *Self-Organizing Maps* (Springer-Verlag, Berlin, 2001).

10 K. Miyazawa, J. Hakkarainen, J. Parkkinen, T. Jääskeläinen, "Ordering of color spectra for digital image enhancement," *Proc. ICIS'02*, (SPSTJ, Tokyo, Japan, 2002) pp. 492–495.

11 O. Kohonen, M. Hauta-Kasari, J. Parkkinen, and T. Jääskeläinen, "Organizing spectral image database by using self-organizing maps," J. Imaging Sci. Technol. **49**, 431 (2005).

12 H. Jeffreys, "An invariant form for the prior probability in estimation problems," Proc. R. Soc. London, Ser. A **186**, 453 (1946).

13 S. Kullback and R. A. Leibler, "On information and sufficiency," Ann. Math. Stat. **22**, 79 (1951).

14 Y. Huang, P. N. Suganthan, S. M. Krishnan, and X. Cao, "Evaluation of distance measures for partial image retrieval using self-organizing map," *Proc. ICANN, LNCS 2130*, (Springer, New York, 2001) pp. 1042–1047.

15 D. Malacara, *Color Vision and Colorimetry, Theory and Applications*, (SPIE, Bellingham, WA, 2002).

16 *Munsell Book of Color-Matte Finish Collection* (Munsell Color, Baltimore, Md., 1976).

17 M. Hauta-Kasari, K. Miyazawa, S. Toyooka, and J. Parkkinen, "Spectral vision system for measuring color images," J. Opt. Soc. Am. A **16**, 2352 (1999).

18 R. Ramanath, R. G. Kuehni, W. E. Snyder, and D. Hinks, "Spectral spaces and color spaces," Color Res. Appl. **29**, 29 (2004).

19 J. Lehtonen, J. Parkkinen, and T. Jaaskelainen, "Optimal sampling of color spectra," J. Opt. Soc. Am. A **23**, 2983 (2006).

20 H. Laamanen, T. Jaaskelainen, and J. Parkkinen, "Conversion between the reflectance spectra and Munsell notations," Color Res. Appl. **31**, 57 (2006).

21 C.-H. Lee, B.-J. Moon, H.-Y. Lee, E.-Y. Chung, and Y.-H. Ha, "Estimation of spectral distribution of scene illumination from a single image," J. Imaging Sci. Technol. **44**, 308 (2000).

22 J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, "Self-organizing map in Matlab: The SOM toolbox," *Proc. Matlab DSP Conference* (Comsol Oy, Espoo, Finland, 1999), pp. 35–40.

23 http://spectral.joensuu.fi/ (January 26, 2008).

24 V. Botchko, S. Nakauchi, J. Parkkinen, and H. Kälviäinen, "Multispectral texture derivation in virtual coloring," in *Texture Analysis in Machine Vision*, edited by M. Pietikäinen (World Scientific, Singapore, 2000).