Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Pattern Recognition Letters 24 (2003) 2009–2019

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Improving similarity measures of histograms using smoothing projections

Joni-Kristian Kamarainen *, Ville Kyrki, Jarmo Ilonen, Heikki Kälviäinen

Laboratory of Information Processing, Department of Information Technology, Lappeenranta University of Technology, P.O. Box 20, FIN-53851 Lappeenranta, Finland

Received 31 October 2002; received in revised form 8 January 2003

Abstract

Selection of a proper similarity measure is an essential consideration for a success of many methods. In this study, similarity measures are analyzed in the context of ordered histogram type data, such as gray-level histograms of digital images or color spectra. Furthermore, the performance of the studied similarity measures can be improved using a smoothing projection, called neighbor-bank projection. Especially, with distance functions utilizing statistical properties of data, e.g., the Mahalanobis distance, a significant improvement was achieved in the classification experiments on real data sets, resulting from the use of a priori information related to ordered data. The proposed projection seems also to be applicable for dimensional reduction of histograms and to represent sparse data in a more tight form in the projection subspace.

Keywords: Ordinal histograms; Similarity measures; Distance functions

1. Introduction

At the core of many algorithms in pattern recognition is the metric or similarity measure used to determine the distance or similarity between two features. There is no generic method for selecting a similarity measure or a distance function. However, a priori information and statistics of features can be used in selection or to establish a new measure (Aksoy and Haralick, 2001; Hafner et al., 1995; Jin and Kurniawati, 2001; Mitra et al., 2002; Sebe et al., 2000). In practice, a similarity measure is often an underlying property of an algorithm, and thus, the use of a measure is implicit. Still, the role and meaning of selecting a proper similarity measure in any algorithm should not be neglected.

In general, there are many distance functions that can be used to measure similarities, distances, between most types of features. Perhaps the two most common ones are the Euclidean and Manhattan distances. However, different distance functions should be used depending on the type of a particular feature attribute. For example, the

* Corresponding author. Tel.: +358-5-621-2844; fax: +358-5-621-2899.

E-mail addresses: joni.kamarainen@lut.fi (J.-K. Kamarainen), ville.kyrki@lut.fi (V. Kyrki), jarmo.ilonen@lut.fi (J. Ilonen), heikki.kalviainen@lut.fi (H. Kälviäinen).

0167-8655/03/$ - see front matter
doi:10.1016/S0167-8655(03)00039-4

Euclidean and Manhattan distances are not suitable for symbolic (nominal) attributes and in such cases some other metrics should be used, e.g., value difference metric (Stanfill and Waltz, 1986; Wilson and Martinez, 1997) or conceptual distance (Kodratoff and Tecuci, 1988).

The accuracy of the most common distance functions, such as Euclidean, can be significantly improved if a priori information of the features is used. Our motivation is to study different distance functions for measuring similarity of ordered histograms. An ordered (also called ordinal, Cha and Srihari (2002)) histogram is a histogram where adjacent bins contain related information, for example a gray-level histogram or a color spectrum (wavelength distribution of light intensity). It should be noted that most measurements are of the ordinal type. A priori information of ordered bins can be used to construct a more robust similarity measure by combining information from neighboring bins. The correlation information of adjacent bins could also be inspected using statistical similarity measures, such as Mahalanobis distance, but if the data is sparse, the correlation information can be distorted. Thus, we are interested in a similarity measure which could also be used with an ordinal sparse data set of few examples in a high dimensional space, when statistical properties of data cannot be reliably inspected. Cha and Srihari (2002) present that with ordinal data, only landmover distance gives good results as their method is equal to the landmover distance. However, this study shows that the landmover can be significantly outperformed. The authors have introduced a subspace projection of the data, called the neighbor-bank projection (Kamarainen et al., 2001), where the data is projected to a subspace which reduces the dimension of the data by combining adjacent bins, and also represents sparse data in a more tight, smoothed, subspace. In this study, the most common similarity measures are evaluated in the context of ordered histograms and the interesting properties of the neighbor-bank projection are inspected. The experimental results show that the similarity measures can be significantly improved by the neighbor-bank projection, and furthermore, statistical properties of sparse data are more evident in the neighbor-bank subspace.

2. Similarity measures for ordered histograms

In this section, a set of similarity measures is proposed for ordered histograms. An ordered histogram or distribution is a histogram where adjacent histogram bins contain related information. For example, in a gray-level histogram the neighbor dimensions represent pixel intensity values that are almost the same. In many natural smooth distributions, such as in color spectra, the same characteristics are present. Similarity measures for ordered histograms can be built upon common distance functions, but by utilizing a priori information by smoothing projections the results of similarity measures can be improved.

2.1. Distance functions

Most commonly used distance functions are shown in Fig. 1 for two feature vectors $\boldsymbol{p} = (p_0, \ldots, p_{L-1})$ and $\boldsymbol{q} = (q_0, \ldots, q_{L-1})$. The most familiar distance functions are the Euclidean and Manhattan distances induced by L_2 and L_1 norms, respectively. They are both special cases of the Minkowsky distance (L_p). As the degree of the norm (p) increases, the weight of large differences between single attribute values increases. Both, the Euclidean and Manhattan, distances are calculated separately for each dimension, and thus, they are not good measures for similarity between two histograms, where attributes are correlated and ordered. For ordinal data, the cumulative Euclidean and landmover distances can be used. The cumulative Euclidean and landmover distances measure the spatial concentration of the values in the feature vector and the order of the feature attributes affects the value of the distance. Therefore, with ordered data these measures are likely to provide better results than the standard Euclidean and Manhattan distances.

Previous distance functions consider the attributes to be non-correlated. For cross-correlated attributes, statistical properties of the data set can be used to reduce the effect of the correlations. One distance function with such a statistical factor, the correlation matrix, is the Mahalanobis distance. Calculation of the correlation matrix in the Mahalanobis distance needs quite much data, the

Euclidean:

$$d(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=0}^{L-1} (p_i - q_i)^2}$$

Manhattan:

$$d(\mathbf{p},\mathbf{q}) = \sum_{i=0}^{L-1} |p_i - q_i|$$

Minkowsky:

$$d(\mathbf{p},\mathbf{q}) = \left(\sum_{i=0}^{L-1} |p_i - q_i|^r\right)^{1/r}$$

Cumulative Euclidean:

$$d(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=0}^{L-1}\left(\sum_{u=0}^{i} p_u - \sum_{u=0}^{i} q_u\right)^2}$$

Landmover:

$$d(\mathbf{p},\mathbf{q}) = \sum_{i=0}^{L-1}\left|\sum_{u=0}^{i} p_u - \sum_{u=0}^{i} q_u\right|$$

Mahalanobis: [a]

$$d(\mathbf{p},\mathbf{q}) = (\mathbf{p}-\mathbf{q})^T \mathbf{C}^{-1} (\mathbf{p}-\mathbf{q})$$

Modified G:

$$d(\mathbf{p},\mathbf{q}) = 2\left\{\left[\sum_{f=p,q}\sum_{i=0}^{L-1} f_i \log f_i\right] - \left[\sum_{f=p,q}\left(\sum_{i=0}^{L-1} f_i\right)\log\left(\sum_{i=0}^{L-1} f_i\right)\right] - \left[\sum_{i=0}^{L-1}\left(\sum_{f=p,q} f_i\right)\log\left(\sum_{f=p,q} f_i\right)\right] + \left[\left(\sum_{f=p,q}\sum_{i=0}^{L-1} f_i\right)\log\left(\sum_{f=p,q}\sum_{i=0}^{L-1} f_i\right)\right]\right\}$$

[a] **C** is the covariance matrix

Fig. 1. Common distance functions.

exact amount depends on the length of the feature vectors and the variation between dimensions. If there is no enough data or no enough variation in the data, numerical calculation of the inverse of the covariance matrix may become an ill-posed problem. A more comprehensive study of the Mahalanobis distance with a limited sample set size can be found in (Takeshita et al., 1993). Another statistical method, the log likelihood ratio G, measures the degree that observed data fits to an expected distribution (Sokal and Rohlf, 1969).

2.2. Smoothing projections

New similarity measures can be introduced based on the previous distance functions and a priori information concerning ordered histograms. Because in an ordered histogram the closely situated elements correlate more strongly than elements which are further apart, a feature vector can be projected to a smaller number of dimensions without significant loss of information. This kind of smoothing projection together with any distance function induces a new similarity measure. In the case of statistical methods, such as the Mahalanobis distance, the statistical properties may be more evident in the projected space. A hypothesis is made that using a smoothing projection, the statistical properties of the samples are more evident and the similarity measure is improved (Kamarainen et al., 2001).

Dimensionality of a histogram is reduced by a linear projection to a subspace, called the neighbor-bank subspace. A set of discrete sampled $\cos^2$ functions can be used to form the neighbor-bank subspace. Histograms are projected on a set of

$\cos^2$ functions (see Fig. 2). Let L be the length of an original histogram $\boldsymbol{p} = (p_0, \ldots, p_{L-1})^{\mathrm{T}}$ and N be the number of banks indexed with k from 0 to $N-1$. Then for $B_k(i)$ the discrete neighbor-banks of $\cos^2$ function can be constructed as

$$B_k(i) = \begin{cases} \cos^2\left(\pi\left(\frac{i}{L}\frac{N-1}{2} + \frac{k}{2}\right)\right) & \text{if } \frac{L}{N-1}(k-1) \\ & \leqslant i \leqslant \frac{L}{N-1}(k+1), \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

By constructing a transformation matrix

$$\boldsymbol{B} = \begin{pmatrix} B_0(0) & B_0(1) & \cdots & B_0(L-1) \\ B_1(0) & B_1(1) & \cdots & B_1(L-1) \\ \vdots & \vdots & \ddots & \vdots \\ B_{N-1}(0) & B_{N-1}(1) & \cdots & B_{N-1}(L-1) \end{pmatrix}, \quad (2)$$

the projection can be performed by matrix multiplication as

$$\boldsymbol{r} = \boldsymbol{Bp}, \quad (3)$$

where $\boldsymbol{r}$ is the projection. If $\boldsymbol{r}$ belongs to space R and $d(\cdot,\cdot)$ is a metric in space R, it can be shown that the distance measure of the projections $\mathrm{d}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \mathrm{d}(\boldsymbol{Bp}_1, \boldsymbol{Bp}_2)$ is a pseudo-metric in space P for all $\boldsymbol{p}_1, \boldsymbol{p}_2 \in P$. For a measure to be a pseudo-metric, three conditions have to be satisfied

$$\mathrm{d}(x,x) = 0, \quad (4)$$

$$\mathrm{d}(x,y) = \mathrm{d}(y,x) \quad \text{(commutativity)}, \quad (5)$$

$$\mathrm{d}(x,z) \leqslant \mathrm{d}(x,y) + \mathrm{d}(y,z) \quad \text{(triangle inequality)}. \quad (6)$$

Because $\mathrm{d}(\boldsymbol{r}_1, \boldsymbol{r}_2)$ is a metric in R, it satisfies (4)–(6), and in addition

$$\mathrm{d}(x,y) = 0 \iff x = y. \quad (7)$$

Next, the properties of $\mathrm{d}(B\cdot, B\cdot)$ are inspected in P. The first condition (4) is satisfied for $\mathrm{d}(\boldsymbol{Bp}_1, \boldsymbol{Bp}_2)$ since

$$\mathrm{d}(\boldsymbol{Bp}, \boldsymbol{Bp}) = \mathrm{d}(\boldsymbol{r}, \boldsymbol{r}) = 0 \quad \text{by (7)}. \quad (8)$$

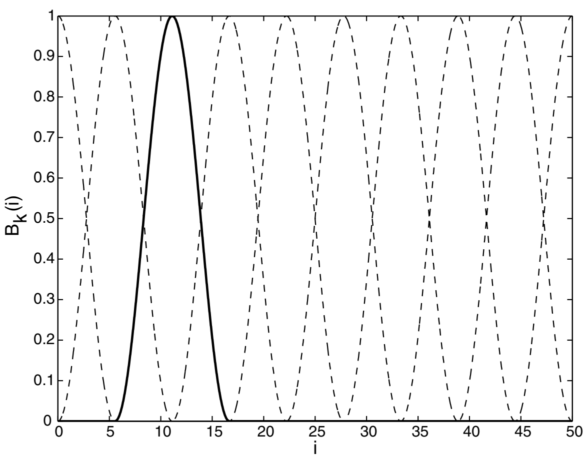
It should be noted that for a pseudo-metric it is also possible that $\mathrm{d}(\boldsymbol{Bp}_1, \boldsymbol{Bp}_2) = 0$ for $\boldsymbol{p}_1 \neq \boldsymbol{p}_2$. The second condition (5),

$$\mathrm{d}(\boldsymbol{Bp}_1, \boldsymbol{Bp}_2) = \mathrm{d}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \mathrm{d}(\boldsymbol{r}_2, \boldsymbol{r}_1) = \mathrm{d}(\boldsymbol{Bp}_2, \boldsymbol{Bp}_1). \quad (9)$$

Triangle inequality also holds, as for any $\boldsymbol{Bp}_1$, $\boldsymbol{Bp}_2$, $\boldsymbol{Bp}_3 \in R$ where $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, $\boldsymbol{p}_3 \in P$,

$$\mathrm{d}(\boldsymbol{Bp}_1, \boldsymbol{Bp}_2) = \mathrm{d}(\boldsymbol{r}_1, \boldsymbol{r}_2) \leqslant \mathrm{d}(\boldsymbol{r}_1, \boldsymbol{r}_3) + \mathrm{d}(\boldsymbol{r}_3, \boldsymbol{r}_2) = \mathrm{d}(\boldsymbol{Bp}_1, \boldsymbol{Bp}_3) + \mathrm{d}(\boldsymbol{Bp}_3, \boldsymbol{Bp}_2). \quad (10)$$

After the projection of data to the subspace spanned by $(B_0(i), \ldots, B_{N-1}(i))$, any standard distance metric can be used. The advantage of using $\cos^2$ functions is that the sum over the banks is 1



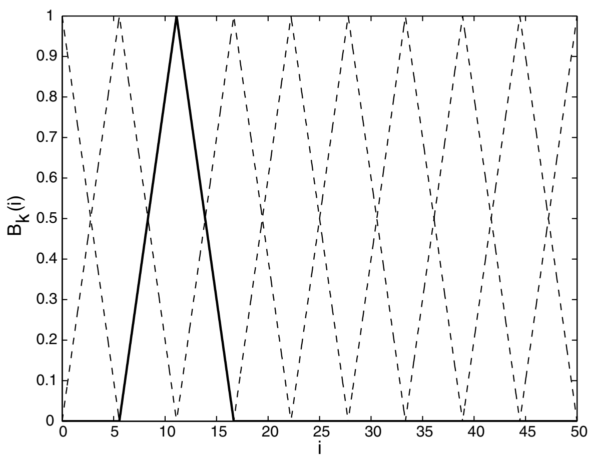


Fig. 2. $N = 10$ neighbor-banks of *cos*^{2} and *triangle* functions for discrete histograms of size $L = 50$ (the third neighbor-bank is highlighted).

over the whole interval of i (see Fig. 2). All attributes are thus equally weighted, although the property of equal weighting is not mandatory. A set of triangle functions also provides the same property of equal weighting. For triangle functions the neighbor-banks can be constructed from

$$B_k(i) = \begin{cases} 1 - \frac{N-1}{L}\left|i - k\frac{L}{N-1}\right| & \text{if } \frac{L}{N-1}(k-1) \leqslant i \leqslant \frac{L}{N-1}(k+1), \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Furthermore, in neural network classifiers, dimensionality of input must often be reduced, which may be performed using, for example, the PCA transform (Haykin, 1994). The neighbor-bank projection could also be considered as a dimensional reduction method which can be used instead of the PCA transform. However, information loss in the dimensional reduction is small only when the a priori assumption of ordered histograms holds.

2.3. Histogram similarity example

Let us consider the three histograms on left in Fig. 3. The distances from histograms 2 and 3 to histogram 1 are calculated using the previously defined distance functions. Visually histogram 2 seems to be much more similar to histogram 1 than to histogram 3. However, using the previously described distance metrics the similarity is not that obvious.

The distances between the original histograms in Fig. 3 are shown in Table 1. In addition to the distances, the ratio between distances to histograms 2 and 3 is shown. The greater the ratio, the better the ability of the corresponding metric in discriminating the histograms. The Euclidean, Manhattan and G-statistics distance functions produce exactly the same distance to histograms 2 and 3. This is because they operate separately in different dimensions and the a priori information of neighbor correlations remains totally unused. On the other hand the cumulative Euclidean and landmover distances show a significant difference between the distances. They both measure the overall similarity in the shape of the histograms and thus histogram 2 is measured as being much closer to histogram 1. Next, the number of dimensions of the histograms is reduced by projecting them to the set of 10 $\cos^2$ functions shown in Fig. 2. The result can be seen on right in Fig. 3. The same distances calculated in the smoothing neighbor-bank subspace are also shown in Table 1. Now, the results for all distance metrics are as expected and histogram 1 is measured to be closer to histogram 2 than to histogram 3.

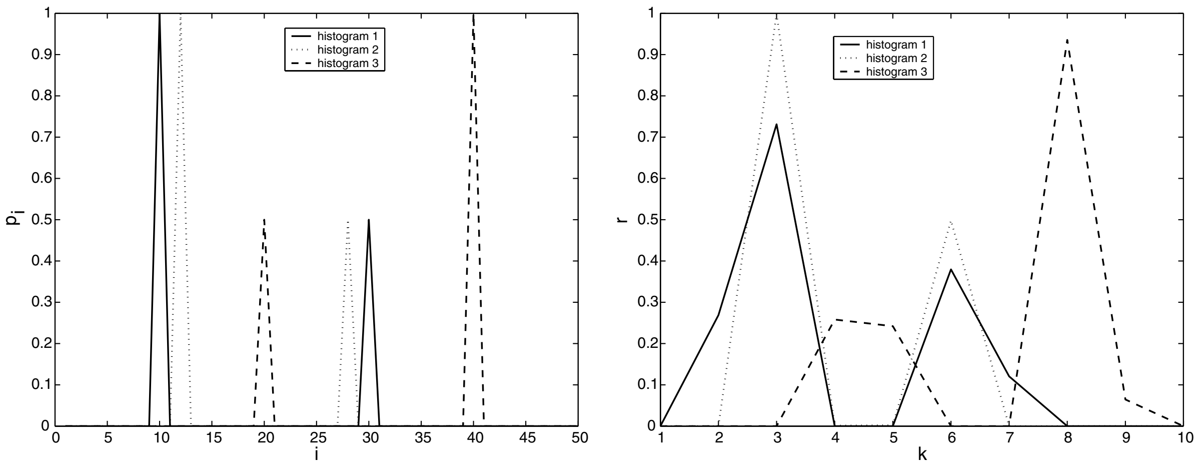


Fig. 3. Original values of example histograms 1, 2, and 3 of $L = 50$ discrete values and the projected histograms on the $N = 10$ neighbor-banks.

Table 1
Distances from histogram 1 to histograms 2 and 3 for original data and projected data

Metric	Original			Projected		
	Histogram 2	Histogram 3	Ratio	Histogram 2	Histogram 3	Ratio
Euclidean	1.581	1.581	1.00	0.4152	1.331	3.21
Manhattan	3.000	3.000	1.00	0.7783	3.000	3.85
Landmover	2.000	16.67	8.34	0.2615	2.970	11.36
Cumulative Euclidean	1.111	10.00	9.00	0.0385	1.623	42.16
G-statistic	0.4405	0.4405	1.00	0.0314	0.3295	10.49

3. Experiments

In this section, the discussed properties of the distance functions and the proposed smoothing projection, the neighbor-bank projection, are demonstrated with experiments conducted on real data sets. All data sets and the basic neighbor-bank functionality can be downloaded from the project homepage at http://www.it.lut.fi/project/metrics/.

3.1. Improving covariance information

In the first experiment, a color spectra data set of three different types of trees (pine, spruce, and birch) was used. The samples of the data set were classified to the correct tree classes. The number of elements in the feature vector was 93 representing wavelength intensities from 390 to 850 nm. There was 350 spectra per class. One hundred random vectors were used in training and the rest for the testing of classifiers. The data set has been formerly introduced in (Jaaskelainen et al., 1994).

The Bayesian classifier was used in the experiment since its success strongly depends on the correct covariance information. For comparison the same test was performed using 1-NN (1-Nearest Neighbor) classifier and various distance functions. From the results shown in Fig. 4, it can be seen that the better results were achieved for smaller number of banks for both, Bayesian and 1-NN with the Mahalanobis distance, classifiers. In 1-NN, the best classification accuracy was achieved by the Mahalanobis distance using pro-

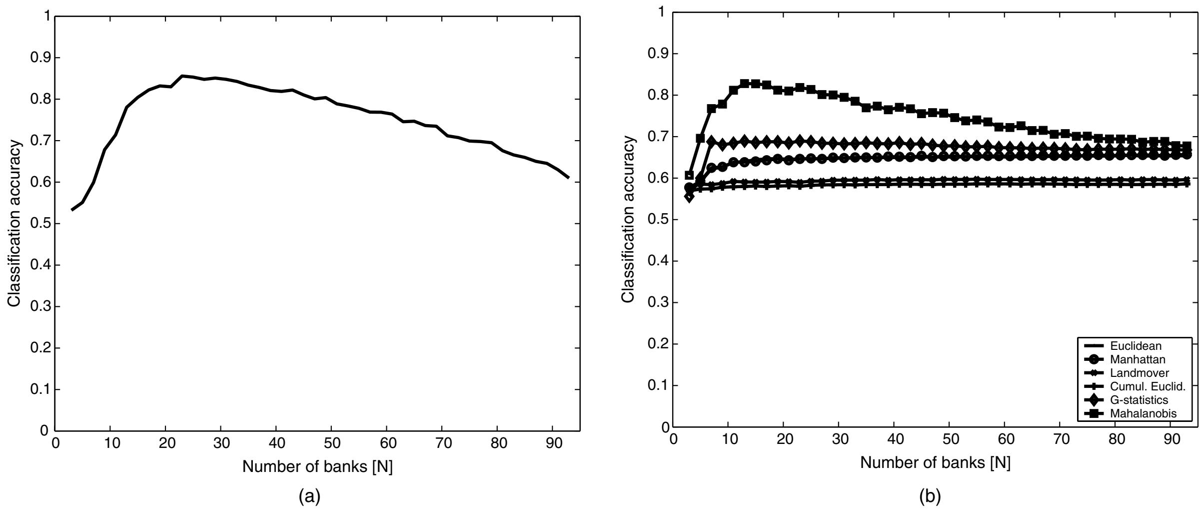


Fig. 4. Results of 1-NN and Bayesian classification of neighbor-bank projected forest color spectra: (a) Bayesian classifier and (b) 1-NN.

jection only to 12 neighbor-banks. Results did not significantly change with smaller number of banks even for other distance functions.

A somewhat surprising finding is that the use of original data gives significantly worse results compared to the projected data. If the number of neighbor-banks equals the dimensionality of the original data (in this case $N = 93$), the data is not changed in the projection. Thus the classification accuracies for the original distance functions can be seen in the rightmost edge of the plots in Fig. 4. The result can be explained by the sparseness of the data and the ability of the projection to enhance the covariance information.

It is important not to confuse the projection with smoothing. No better results could be achieved by plain smoothing as shown in Fig. 5 where the original forest spectra were smoothed by filters similar to $\cos^2$ neighbor-banks (pseudo-inverse of the covariance used due to the singularities) before applying the Bayesian classification method. The classification accuracy was significantly worse for filtered data than for projected in Fig. 4.

3.2. Comparing distance functions

Next, the accuracy and behavior of the different distance functions for neighbor-bank projected data was examined. Three different data sets were classified using 1-NN classifier and various distance functions. The first data set was the same forest data set used in the first experiment. In the second data set sawn timbers were classified on the basis of whether there were branches. Spectra were of four different species of trees (aspen, birch, pine, and spruce) and each species included trees with and without branches, thus giving eight different classes. Wavelength interval was 380–2700 nm with a resolution of 1 nm, the feature vector length being 2321. There were about 40 samples per class. Third data set consisted of gray-level histograms from images of vacuum tank degassing. The level of bubbling of molten steel was classified based on the gray-level histograms of the images. The histograms were divided into five classes. Data from the vacuum tank degassing process has been introduced in more detail in (Kämäräinen et al., 2000). The gray-level histograms were from images of one size and represented in 8 bits, and thus, contained a total of 256 different gray-levels. The sum of the histogram bins was the number of pixels in the original image. In the bubbling data there were approximately 30 examples in every class.

The experiment was conducted using a 1-NN classifier and leave-one-out evaluation. By using leave-one-out method the sparseness of the data was avoided and the actual accuracies of different distance functions were thus comparable. The results of the experiments can be seen in Figs. 6–8 for neighbor-banks created using $\cos^2$ and *triangle* functions.

With the exception of the Mahalanobis and *G*-statistic functions, the distance functions gave practically the same classification percentages for most neighbor-banks and no distance function performed significantly worse for any data set. It seems that the information loss is tolerable for the classification task when the number of dimensions is heavily reduced by the neighbor-banks. When fewer banks were used, the Mahalanobis distance gave better results than the other distances except for the last data set where all distance functions performed considerably well. Also, for all data sets, the singularity of the covariance matrix was avoided with fewer neighbor-banks. For example,

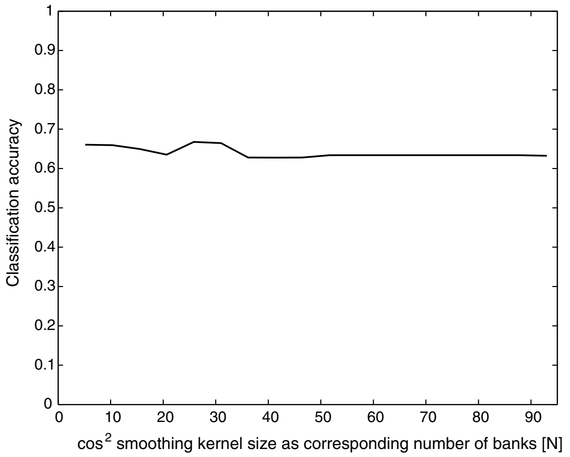


Fig. 5. Results of Bayesian classification of filtered forest color spectra.

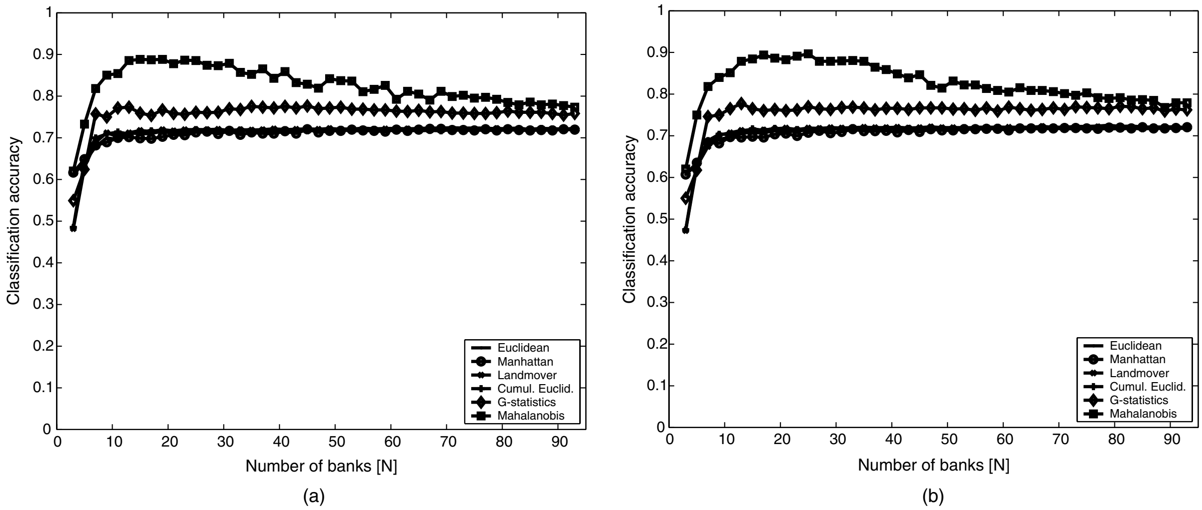


Fig. 6. Results of 1-NN classification of forest color spectra: (a) *cos²* neighbor-banks (b) *triangle* neighbor-banks.

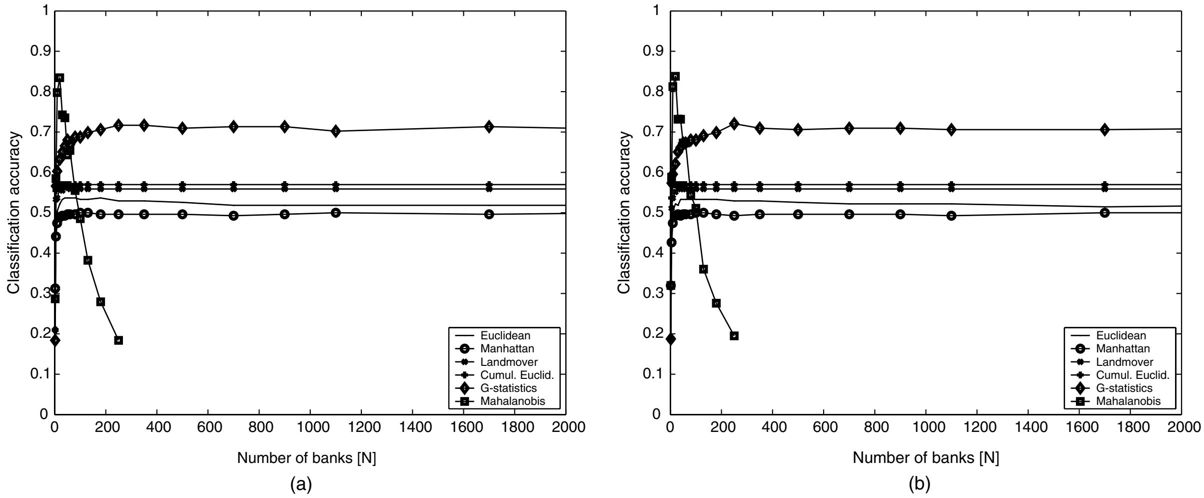


Fig. 7. Results of 1-NN classification of lumber color spectra: (a) *cos²* neighbor-banks (b) *triangle* neighbor-banks.

for the lumber data other distance functions converged to their maximal accuracy when enough divergent data was included in the neighbor-banks. On the other hand, the Mahalanobis was unusable after the covariance matrix became nearly singular for neighbor-bank projected data, even though the best classification accuracy was achieved by the Mahalanobis. The dissimilarity of distinct classes is quite high for the bubbling data set, since the accuracy was good for the Euclidean distance already for two neighbor banks. In the last data set, the Mahalanobis suffered from a singularity already for few banks in the calculation of the inverse of covariance matrix. The pseudo-inverse method should be applied. It should be noted that for the last data set where N banks were used, the total number of banks in the representation of N banks was $N - 1$, because in the all histograms the sum over the histogram bins was the same (the number of pixels in the image). This

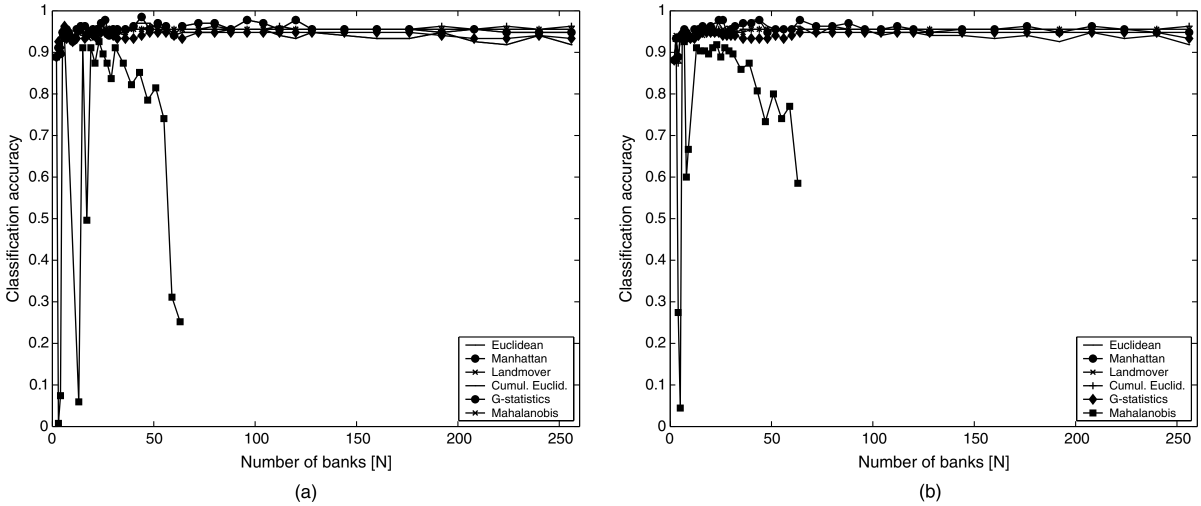


Fig. 8. Results of 1-NN neighbor-bank classification of bubbling image histograms: (a) *cos*^{2} neighbor-banks and (b) *triangle* neighbor-banks.

led to an ill-posed covariance matrix inversion. Thus, with the Mahalanobis distance only $N-1$ bank values were actually used out of the N possible.

It is not possible to define a rule to select the optimal number of neighbor-banks because the optimum depends heavily on the application and available data. The selection is influenced by the amount of correlation in the neighboring attributes and the sparseness of the data as well as the length of the original attribute vector L. However, Fig. 6 suggests that the selection is not very critical as the accuracy is quite stable around the optimum. For a particular application, the number of banks should be as low as possible and can be sometimes suggested by application knowledge. Furthermore, it is evident that the selection of the neighbor-bank type is not a fatal consideration for the success of the method since the results were almost the same for $\cos^2$ and *triangle* neighbor-bank projections in all three data sets.

3.3. Dimensional reduction

As previously demonstrated, the neighbor-bank projection can reduce the dimension of ordered histogram type data with a small loss of information. In this experiment the forest color data from the first experiment was classified by a multi-layer perceptron (MLP) neural network (Haykin, 1994). In MLP networks, the number of dimensions of the data must be limited to allow the training to be possible within a reasonable time. The neighbor-bank projection was compared to the PCA transform using both methods in the dimensional reduction of the data. The data was randomly divided into two halves. The first half was used to train the MLP neural network and the second half

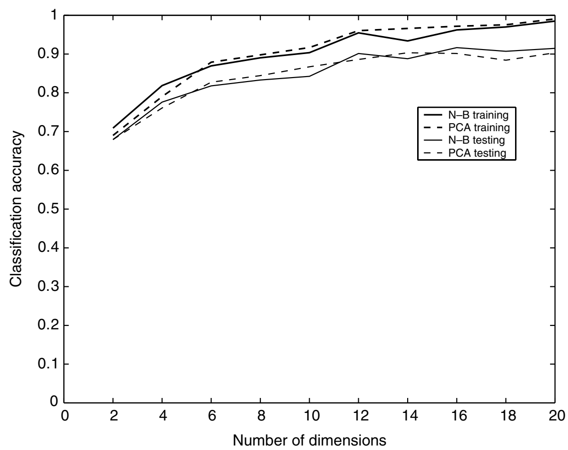


Fig. 9. Results of MLP neural network classification of forest color spectra.

to test the network. The classification results are shown in Fig. 9. The test was repeated several times to achieve more reliable results. The network had a hidden layer with six sigmoidal neurons and output layer of three sigmoidal neurons. As a dimensional reduction method for the forest color spectra, the PCA transform or the neighbor-bank projection do not seem to provide any distinct advantages over each other.

4. Conclusions and discussion

In this paper, the properties of various distance functions were examined primarily in the context of ordered histogram type data. A new smoothing projection, the neighbor-bank projection, was also introduced. The smoothing projection seems to improve the accuracy of some of the studied distance functions and to have advantages when combined with methods utilizing the statistical properties of the data, such as the Mahalanobis distance and the Bayesian classifier.

For a sparse data set, where there are too few samples to accurately acquire statistical properties of the data, the introduced neighbor-bank projection seems to have useful properties. The neighbor-bank projection is a smoothing projection that reduces the dimensions of the data. Thus, the feature space can be projected to a subspace where the data is more tightly represented and the statistical properties are more evident. In the experiments the best results were achieved with much less neighbor-banks than the original dimension of the data. The improvement was significant especially for the Bayesian and 1-NN classifiers with the *G*-statistics and the Mahalanobis distances. For the non-statistical methods, the dimensional reduction had no significant effect. The increasing robustness of the covariance information also provided more robust calculation of the inverse of the covariance matrix, which suffered of singularities.

The neighbor-bank projection seems to have also desirable properties as a plain dimensional reduction method. In our experiment, the widely used PCA method did not show any advantages over the neighbor-bank for an ordered histogram type data. Furthermore, the neighbor-bank smoothing projection does not need to analyze the statistical properties of the data, unlike the PCA transform, and thus, the neighbor-bank transform can be applied in situations where the PCA transform is unusable.

The goal of this study was to improve similarity measures of ordered histograms based on common distance functions. A significant improvement was achieved using the developed neighbor-bank projection and statistical distance functions, such as the Mahalanobis distance. The implications of the results appear significant since most of the measurements are of the ordinal type. In further studies implications of the neighbor-bank projection to the performance of methods based on similarity measures will be addressed, e.g., significance of similarity measure selection in the training of self-organizing maps.

Acknowledgements

The financial support of the East Finland Graduate School on Computer Science and Engineering (ECSE), KAUTE foundation (Kaupallisten ja teknillisten tieteiden tukisäätiö), Jenny and Antti Wihuri fund, Foundation of the Lappeenranta University of technology, and Foundation of the Advance in Technology (Tekniikan edistämissäätiö) is gratefully acknowledged.

References

Aksoy, S., Haralick, R.M., 2001. Feature normalization and likelihood-based similarity measure for image retrieval. Pattern Recognition Lett. 22, 563–582.

Cha, S.-H., Srihari, S.N., 2002. On measuring the distance between histograms. Pattern Recognition 35, 1355–1370.

Hafner, J., Sawhney, H., Equitz, W., Flickner, M., Niblack, W., 1995. Efficient color histogram indexing for quadratic form distance functions. IEEE Trans. Pattern Anal. Machine Intell. 17 (7), 729–736.

Haykin, S., 1994. Neural Networks A Comprehensive Foundation. Macmillan College Publishing Company, New York.

Jaaskelainen, T., Silvennoinen, R., Hiltunen, J., Parkkinen, J.P.S., 1994. Classification of the reflectance spectra of pine, spruce, and birch. Appl. Opt. 33 (12), 2356–2362.

Jin, J.S., Kurniawati, R., 2001. Varying similarity metrics in visual information retrieval. Pattern Recognition Lett. 22, 583–592.

Kämäräinen, J.-K., Kälviäinen, H., Terho, K., Saarelainen, E., August 2000. Visual quality control of the vacuum tank degassing. In: IFAC Workshop on Future Trends Autom. Mineral Metal Process., pp. 379–384.

Kamarainen, J.-K., Kyrki, V., Kälviäinen, H., June 2001. Similarity measures for ordered histograms. In: Proc. 12th Scand. Conf. Image Anal. SCIA2001, Bergen, Norway, pp. 699–705.

Kodratoff, Y., Tecuci, G., 1988. Learning based on conceptual distance. IEEE Trans. Pattern Anal. Machine Intell. 10 (6), 897–909.

Mitra, P., Murthy, C., Pal, S.K., 2002. Unsupervised feature selection using feature similarity. IEEE Trans. Pattern Anal. Machine Intell. 24 (3), 301–312.

Sebe, N., Lew, M.S., Huijismans, D.P., 2000. Toward improved ranking metrics. IEEE Trans. Pattern Anal. Machine Intell. 22 (10), 1132–1142.

Sokal, R.R., Rohlf, F.J., 1969. Biometry. W.H. Freeman, New York.

Stanfill, C., Waltz, D., 1986. Toward memory-based reasoning. Comm. ACM 29, 1213–1228.

Takeshita, T., Nozawa, S., Kimura, F., 1993. On the bias of Mahalanobis distance due to limited sample size effect. In: Proc. Second Internat. Conf. Document Anal. Recognition, pp. 171–174.

Wilson, D.R., Martinez, T.R., 1997. Improved heterogeneous distance functions. J. Artificial Intell. Res. 6, 1–34.