The Dissimilarity Representation as a Tool for Three-Way Data Classification: A 2D Measure

Diana Porro-Muñoz[1,2], Robert P.W. Duin[2], Mauricio Orozco-Alzate[3],
Isneri Talavera[1], and John Makario Londoño-Bonilla[4]

[1] Advanced Technologies Application Center (Cenatav), Cuba
{dporro,italavera}@cenatav.co.cu
[2] Pattern Recognition Lab, TU Delft, The Netherlands
r.duin@ieee.org
[3] Universidad Nacional de Colombia Sede Manizales, Colombia
morozcoa@bt.unal.edu.co
[4] Instituto Colombiano de Geología y Minería (Ingeominas), Colombia
jmakario@ingeominas.gov.co

Abstract. The dissimilarity representation has demonstrated advantages in the solution of classification problems. Meanwhile, the representation of objects by multi-dimensional arrays is necessary in many research areas. However, the development of proper classification tools that take the multi-way structure into account is incipient. This paper introduces the use of the dissimilarity representation as a tool for classifying three-way data, as dissimilarities allow the representation of multi-dimensional objects in a natural way. As an example, the classification of three-way seismic volcanic data is used. A comparison is made between dissimilarity measures used in different representations of the three-way data. 2D dissimilarity measures for three-way data can be useful.

Keywords: Object representation, classification, multi-dimensional data, dissimilarity representation.

1 Introduction

In many research areas e.g. chemometrics, image analysis, signal analysis, objects obtained from measurement equipments are represented by multi-dimensional arrays instead of a vector of features. Consequently, the variables from one dimension of the array can be related and analyzed together with the variables of the other dimensions. The structure in which a set of objects with this representation is organized is called multi-way data.

Multi-way data analysis [1, 2] is the extension of multivariate analysis when the analyzed data is arranged in this multi-way structure. However, the most common is the three-dimensional array. The analysis of such data is often used for extracting specific information and exploring the interrelations in the data. It has been shown that this data may not be analyzed optimally by two-way analysis, because it does not respect the multi-way design. Nevertheless, most of the applications and methods for multi-way analysis are for exploratory and

regression purposes. Classification has been studied much less. This might be caused by the lack of classification tools able to operate on multi-dimensional spaces and taking all the information available into account.

In recent studies [3, 4, 5], the advantage of learning from dissimilarities between the objects instead of traditional features has been shown, in what is known as the Dissimilarity Representation (DR) [3]. This representation was mainly designed for classification. It is based on the important role that pairwise dissimilarities between objects play. Classifiers may be built in the dissimilarity space generated by a representation set. In this way, the geometry and the structure of a class are determined by the user defined dissimilarity measure, in which application background information may be expressed. It is important to remark that, any traditional classifier that operates in feature spaces can also be used in the dissimilarity space.

In this paper, we introduce the use of the DR as a tool for classifying three-way data in such a way that, objects are analyzed in their 2D representation. Thus, the relations between the objects are analyzed in the dissimilarity space. Moreover, the relationship between the dimensions can be included if the proper dissimilarity measure is selected. The key in this process is to find the dissimilarity measure that takes into account the information embedded in the data. Information about the data that is missing in the actual representation e.g. shape and connectivity, can also be taken into account in the dissimilarity measure.

Traditionally, signals are analyzed in the time domain or by their spectrum in terms of energy spread over its frequency components (Fourier transform) [6]. Recent studies have also shown that training the classifiers in the dissimilarity space is a feasible and more reliable alternative for automatic classification of seismic signals than the frequency-based one [4]. Nevertheless, these representations alone may not be optimal for seismic signal analysis, since the changes of spectral energy in time are not considered. Due to this limitation, the use of a time-frequency representation like spectrograms or scalograms, may be advantageous. Although these types of 2D object representations are raising popularity for the analysis of seismic signals [7], they have not intensively been exploited as such in automatic classification systems [8].

Hence, although the proposed approach can be applied to any three-way data in the form (*objects × variables × variables*), we will based the demonstration of its feasibility on a problem of classifying three-way seismic volcanic signals. With the purpose of comparing how it works for three-way data with different characteristics and some suitable dissimilarity measures, two three-way seismic volcanic data will be generated from the spectrogram and scalogram techniques. A 2D dissimilarity measure is also proposed. Additionally, results are compared with 1D feature representation using the time integrated spectra to show the advantages of the proposed approach in this case.

2   Three-Way Volcanic Data

In several research areas, different multi-way array configurations can be found e.g. several sets of variables measured on different objects. These data would be

appropriately represented by higher order generalization of vectors and matrices. However, the most common design would be defined as $\mathbf{Y} \in \mathbb{R}^{n \times m \times l}$. Each horizontal slice ($\mathbf{m} \times \mathbf{l}$) of the block represents the data of one object; each vertical slice ($\mathbf{n} \times \mathbf{l}$) holds for the data of a specific type of variable and the front to back slices ($\mathbf{n} \times \mathbf{m}$), variables of other type.

The two three-way data to be used in this paper correspond to seismic signals from the ice-capped Nevado del Ruiz volcano in the Colombian Andes. This volcano is currently studied by the Volcanological and Seismological Observatory at Manizales. Signals from the Olleta crater station were selected for the experiments. Signals were digitized at 100.16 Hz sampling frequency by using a 12 bit analog-to-digital converter. The a-priori classification of the signals is done by visual inspection. The dataset is composed of 12032-point signals of two classes of volcanic activities: 235 of Long-Period (LP) earthquakes, and 235 of Volcano- Tectonic (VT) earthquakes.

The differences in 1D spectral content of these signals allow the discrimination between the events [9]. That is why spectral-based classification is often used for this type of data. However, with this representation we are not able to analyze how the frequency content changes in time. An intuitive way to represent this time-frequency relationship for all the signals would be in a three-way array $\mathbf{Y}$ as defined above. In the seismic volcanic three-way array configuration (*signals* $\times$ *time* $\times$ *frequency*), the signals are organized in the vertical axis, time in the horizontal and frequency in the depth axis.

To obtain the 2D time-frequency representation of each signal we used two techniques. The first one is the Short-Time Fourier Transform (STFT) by which the spectrograms are obtained [6]. With this technique, the time localization can be obtained by windowing the data at different times and computing the Fourier transform on that part of the signal. Consequently, it can be known what frequency intervals are present in a time interval of the signal, but not with much precision. The narrower it is, the better the time resolution will be and the poorer frequency resolution.

Another way to obtain the time-frequency representation of the signals is by the Continuous Wavelet Transforms (CWT), with which the scalograms are computed. This technique is based on the computation of continuous wavelet transforms over the entire signal for different scales [6]. It was originally introduced as a time-scale representation, but it can also be interpreted as a time-frequency representation as scales and frequencies are inversely proportional. In the scalograms we have better time resolution and poorer frequency resolution at high frequencies, and better frequency resolution and poorer time resolution at low frequencies. Consequently, this technique could lead to a more accurate time-frequency description of signals with low and high frequencies, as is the case of the treated data.

3   Dissimilarity Representation from Three-Way Data: A 2D Measure

The Dissimilarity Representation (DR) [3], was proposed as a more flexible representation of the objects than the feature representation, with the purpose of

having more information about the structure of the objects. It is seen as a link between the statistical and structural approaches, as both types of patterns can be described by the (dis)similarity measure. The DR is also based on the role that (dis)similarities play in a class composition, where objects from the same class should be similar and objects from different classes should be different (compactness property). Hence, it should be easier for the classifiers to discriminate between them.

Using the DR, classifiers are trained in the space of the proximities between objects, instead of the traditional feature space. Thus, in place of the feature matrix $\mathbf{X} \in \mathbb{R}^{n \times q}$, where $\mathbf{n}$ runs over the objects and $\mathbf{q}$ over the variables, the set of objects is represented by the matrix $\mathbf{D}(\mathbf{X}, \mathbf{R})$. This matrix contains the dissimilarity values $d(x_i, r_j)$ between each object x_i of $\mathbf{X}$ and the objects r_j of the representation set $\mathbf{R}(r_1, ..., r_h)$. We build from this matrix a dissimilarity space. Objects are represented in this space by the column vectors of the dissimilarity matrix. Each dimension corresponds to the dissimilarities with one of the representation objects.

For a t-dimensional array $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times ... \times I_t}$, the theory of the DR is the same. In fact, one of the advantages of the DR is that it can be generated from any representation of the objects e.g. vectors of numbers, graphs, as long as we have a proper dissimilarity measure. This applies also to the multi-way data. Originally, each object is represented by a $(t\text{-}1)$-dimensional array of numerical values and all the objects together conform the t-dimensional array. Hence, to obtain the dissimilarity space, a mapping $\phi(\cdot, R) : \mathbb{R}^{I_1 \times I_2 \times ... \times I_{t-1}} \to \mathbb{R}^h$ is defined, such that for every object y, $\phi(y, R) = [d(y, r_1), d(y, r_2), ..., d(y, r_h)]$. Classifiers are then built in this space, as in any feature space.

The elements of $\mathbf{R}$ are called prototypes, and have preferably to be selected by a prototype selection method [3]. These prototypes are usually the most representative objects of each class, $\mathbf{R} \subseteq \mathbf{X}$ or $\mathbf{X}$ itself, resulting in a square dissimilarity $\mathbf{D}(\mathbf{X}, \mathbf{X})$. $\mathbf{R}$ and $\mathbf{X}$ can also be chosen as different sets. As dissimilarities are computed to $\mathbf{R}$, a dimensionality reduction is reached if a good, small set can be found, resulting in less computationally expensive classifiers.

The issue to be addressed in this problem is how to obtain the dissimilarities from the multi-way representation. Many ideas can arise to do this transformation. Focusing in three-way data we propose as a first approach, to take each object matrix y of $\mathbf{Y}$, and compute the dissimilarities between them by a 2D dissimilarity measure. Some 2D measures have been proposed in [10] for face and palm-print recognition. However, the selection of the suitable measure for the problem at hand is a very important aspect in the DR approach. To deepen in this task we will focus in our case of study on three-way representation of seismic volcanic signals by spectrograms and on scalograms. Thus, each object is represented by a matrix (2D). A comparison is made about the characteristics of each data and the dissimilarity measure to be used. A 2D dissimilarity measure is also proposed.

In many types of data e.g. spectral data, it is necessary to take into account the shape information and connectivity between the measure points. Such is the case

of the time-frequency three-way representation where shape changes are present in the spectral (frequency) direction and connectivity in the time direction. When this representation is obtained by scalograms, the CWT already retrieves these functional characteristics from the data. The observations in the signal can be seen as continuous single entities, instead of sets of different variables. Based on the results obtained with the 2D assembled matrix distance (AMD) (See Eq. 1) proposed in [10], it seems to be a good option for this case. As the information to be included about the data can be already found in its representation by wavelets, it might be enough to use this measure.

$$d_{AMD}(y_a, y_b) = \left(\sum_{k=1}^{l} \left(\sum_{j=1}^{m} (y_{a,j,k} - y_{b,j,k})^2 \right)^{p/2} \right)^{1/p} \tag{1}$$

The weight p is used to emphasize either small or large differences between the elements, in dependence of the problem at hand. If $p < 1$, all the differences will be reduced, thus the larger ones will not interfere much in the measure. On the other hand, if $p > 1$, the larger differences will be more pronounced, resulting in a heavy influence on the measure. However, when the information is not taken into account in the representation of the data, the dissimilarity measure has to take care of it. Thus, considering the results obtained with the Shape measure (manhattan distance on the first Gaussian derivatives) for simple spectra [5], we propose to make use of the derivatives into the AMD measure. In such a way, we can take the ordering information into account as well as the shape of the spectra. A principle of the DR approach is that instead of a single representation of a problem, one may also consider either a complex representation, built from many dissimilarity representations, where different aspects of the data are described in various ways [3]. Based on this and the previously stated, we define the 2DShape dissimilarity measure as follows:

1. Compute the matrix D^1

$$D_{a,b}^1 = \left(\sum_{k=1}^{l} \left(\sum_{j=1}^{m} (y_{a,j,k}^{\sigma} - y_{b,j,k}^{\sigma})^2 \right)^{p/2} \right)^{1/p} , \quad y_{i,j,\cdot}^{\sigma} = \frac{\mathrm{d}}{\mathrm{d}j} G(j, \sigma) * y_{i,j,\cdot}$$

2. Compute the matrix D^2

$$D_{a,b}^2 = \left(\sum_{j=1}^{m} \left(\sum_{k=1}^{l} (y_{a,j,k}^{\sigma} - y_{b,j,k}^{\sigma})^2 \right)^{p/2} \right)^{1/p} , \quad y_{i,\cdot,k}^{\sigma} = \frac{\mathrm{d}}{\mathrm{d}k} G(k, \sigma) * y_{i,\cdot,k}$$

3. Combine both dissimilarities matrices $D = \frac{1}{\omega_1} D^1 + \frac{1}{\omega_2} D^2$

The variables $y_{i,j,\cdot}$ and $y_{i,\cdot,k}$, stand for the k-th columns and the j-th rows of the i-th matrix (object); $\forall i = 1, 2, ..., n$. Their expression correspond to the computation of the first Gaussian derivatives of spectra, where $*$ denotes convolution

and σ stands for a smoothing parameter [5]. The dissimilarities in step 1 and step 2 correspond to the first and second directions respectively, as indicated by the notation e.g. spectra and time. In the combination step, we included a weight for scaling. We defined $\omega_c = var(D^c)$, to scale each dissimilarity matrix by its columns (prototypes) variance. This measure can also be used in three-way data where there are no variations in shape in one of the directions. In this case, it is enough to use the AMD measure in step 1 or step 2, such that only the differences in area are compared.

A good example where the proposed measure can be applied is in the time-frequency representation obtained by spectrograms. The connectivity in the time direction is not taken into account as the Fourier transform is computed separately in the different parts of the windowed signal. Besides, instead of having continuous points in time, we have time intervals.

4   Experimental Results and Discussion

To show how the proposed approach works, we selected a data of seismic volcanic signals. We make a comparison between the results with the three-way data obtained by the spectrograms and scalograms. This comparison is not only made in terms of the dissimilarity measures, but in the information we get from the three-way representation. A comparison is also made between the classification on the dissimilarity spaces derived from 2D and the 1D spectral representation of the data. This way, we show the advantages of using the 2D representation over the 1D e.g. time-frequency (spectral) based classification over the spectral-based. The Average Classification Errors (ACE) for the DR on both spectral and three-way data from spectrograms and scalograms are shown, using different sizes of the representation set.

For the experiments, a dataset with 235 objects per class (VT and LP) is considered. For the 1D (spectral) representation we have computed the spectrum by using a 12032-point Fast Fourier Transform (FFT). Thus, the whole signal is analyzed in both 1D and 2D representations. To compute the spectrograms, trying to make a trade-off between time and frequency resolution, a 256-point (windows size) short time Fourier transform was calculated with 50% of overlap. The values for these parameters were selected empirically. However, it is important to determine the best combination for these parameters as they can influence the results. Further research studies should be done in optimizing these parameters and its influence in the solution of the problem. From this technique, we get a $470 \times 129 \times 93$ three-way data.

For the scalograms computation we used the Morlet wavelet, based on the literature [7, 8] and an interactive Matlab tooolbox for the analysis of seismic volcanic data [11]. Taking into account the inversely proportional relation frequency-scales, we selected the scale values related to the major frequency components in the signals. In a previous study on signals of the same volcano and station (although they are not the same samples)[12], the authors concluded that most of the discriminative information is contained between 7.5 Hz and 25

Hz approximately. A narrow band around 40 Hz, associated to an always present peak, was also selected.

However, a 10-component PCA was also made on the spectral representation of the signals and analyzed the modeling power of all the frequencies present. From this analysis we arrived to the same range of frequency values selected in the reference. Nevertheless, some important peaks were also detected from 0.1 Hz to 7 Hz. Hence, range of scales=[1/(0.1:02:2,3:0.4:25,39:0.3:42)] was used to analyze those frequencies. A $470 \times 72 \times 12032$ three-way array was obtained. Before computing the three representations, the raw signals were normalized to zero-mean and unit-variance.

A Fisher Linear classifier was computed in the dissimilarity space. Experiments were repeated 10 times. Training and test objects were randomly chosen from the total data set, in a 10-fold cross-validation process. Different sizes of the representation set $[10, 20, 50, 75, 100, 125, 200, 250, 300]$ were randomly selected. For the generation of the dissimilarity space, the Manhattan (MD), Euclidean (ED) and Shape measures were computed on the spectral representation. These measures have performed well for spectral data [4, 5]. In a 5-times 10-fold cross-validation from a range of values $[1 - 50]$, the best results were achieved with $\sigma = 15$. For the two measures analyzed in Sec. 3, we used values of $p = [0.5, 1, 2]$ so we can investigate the effect of parameter p (small or big differences) in our classification problem. For the spectral direction in the 2DShape measure, we selected $\sigma = 3$ and for the time direction $\sigma = 2$.

It can be observed in Fig. 1 that, the ACE on the dissimilarity space generated from the spectral data is around 25% and 30%. The error values for the Manhattan measure are slightly better than those of the Euclidean and Shape measures. Nevertheless, if the standard deviation is taken into account, the values for the three measures are very similar. The results with the Shape measure (derivative-based) are not as expected (based in previous works). Hence, these results could suggest that there is not more information to be captured from this representation. It is also possible that these measures are not robust enough for this problem, which somehow contradicts the previous studies [4, 5]. Further studies may be done to find a more proper measure for this type of data.
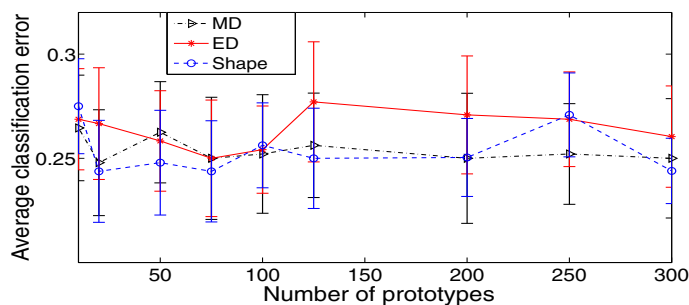

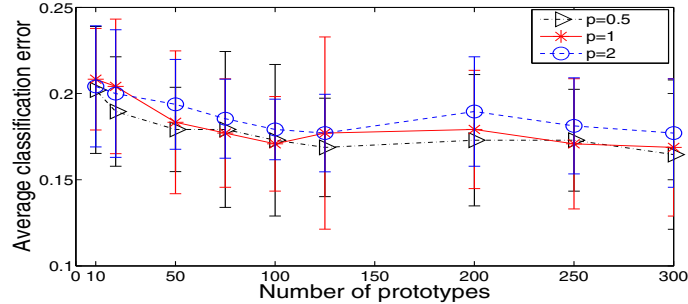**Fig. 1.** ACE on the 1D representation for different numbers of prototypes

Fig. 2. ACE on the three-way data from spectrograms (2D) for different numbers of prototypes
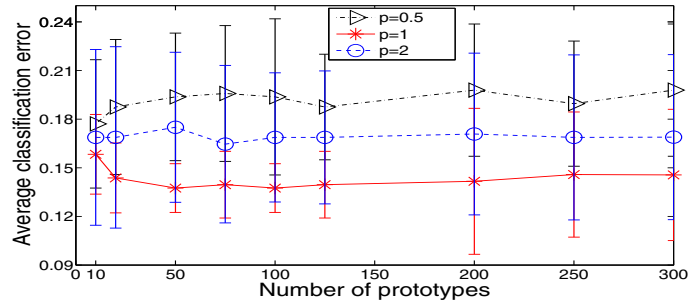

Fig. 3. ACE on the three-way data from scalograms (2D) for different numbers of prototypes

However, when we analyze the error of the DR from the three-way data we see a significant improvement in both Fig. 2 and Fig. 3. This ratifies the fact that the time-frequency relation is more discriminative than the spectra. In the case of the data obtained from the spectrogram, the ACE ranges from 15% to 20%. This also suggests that the proposed 2D measure is capable of capturing the information needed. Nevertheless, if we analyze the ACE of the three-way data obtained from the scalogram, it is slightly better. We can notice that it is also in a range of 13% to 20%, taking into account all the values of p. However, if we analyze only $p = 1$, the largest ACE value is around 15%. These results might be supported by the advantages of the CWT for analyzing this type of seismic volcanic signals. It is also evident that the AMD measure works well for this data, given that the shape and continuity information is already taken into account in the representation by wavelets. It might be possible to obtain better results if more precise scale values are chosen.

The selection of a dissimilarity measure for a representation depends on what we are looking for. In the case of the scalograms, the dissimilarity measure is very

simple. However, the computation of the scalogram is really expensive in cases like this, where there are too many important frequency components and the signals are so large. Nevertheless, we cannot forget the advantages of using this technique for the analysis of frequencies in exact moments in time. On the other hand, if we analyze the computation of the spectrograms, it is less expensive than of the scalograms. But, due to the lack of some information in it, a more complex dissimilarity measure is required to include this information. Besides, for the signal analysis it is less precise than the scalograms, as we can only know what intervals of frequencies are present in an interval of time. Thus, it is up to the specialists to decide which of them to use in dependence of their priority.

If we analyze the ACE on the DR from the spectral data and three-way data from scalograms, we can see that from 50 or more prototypes it is approximately stable. The explanation we give to this phenomenon, is that there is no more discriminating information to be found in more prototypes. On the other hand, if we analyze the ACE on the DR from the three-way data from spectrograms, we can see that the behavior is different. While increasing the number of prototypes, the ACE decreases. The more prototypes we add, the more information we have to discriminate between the classes. Nevertheless, due to the so-called peaking phenomenon, when the number of prototypes starts reaching the size of the training set, the errors will increase.

5 Conclusions

We introduced the use of the Dissimilarity Representation as a tool for classifying three-way data. In this approach, objects are analyzed with a 2D representation. The relationship between the different dimensions is analyzed in the 2D dissimilarity measure. Besides, information about the data that is missing in the original representation e.g. shape, can be considered in it. The good performance of classifiers on the 2D representation of the objects, compared with the traditional 1D, shows that this approach can be a good solution for the classification of data with a three-way structure.

Two 2D dissimilarity measures were analyzed for three-way seismic volcanic data to evidence the importance of the selection of a suitable dissimilarity measure for the problem at hand. We developed a new 2D dissimilarity measure that allows taking into account the shape and continuity information in the directions of the three-way array. This measure demonstrated to work well in cases like the three-way seismic volcanic data generated by the spectrograms. In this data, the shape and continuity variation is not represented itself. Consequently, this type of measure is needed to make use of that information. Nevertheless, in cases where there is not discriminative information in both directions, we cannot ensure that this measure is effective. The combination of the matrices from both directions could be influenced if one of them is not good. Further investigations should be done on this issue. In cases like the three-way data obtained by scalograms, more simple dissimilarity measures can be used e.g. AMD. The discriminative information is already embedded in the representation by wavelets.

Although this paper was more focused on the solution for three-way data, it can be extended to multi-way. Further studies will be done on this aspect.

Acknowledgment

References

[1] Porro-Muñoz, D., Talavera, I., Duin, R.P.W.: Multi-way data analysis. Technical report, CENATAV (2009)

[2] Kroonenberg, P.M.: Applied Multiway Data Analysis. John Wiley & Sons, Chichester (2008)

[3] Pekalska, E., Duin, R.P.W.: The Dissimilarity Representation For Pattern Recognition. In: Foundations and Applications, World Scientific, Singapore (2005)

[4] Orozco-Alzate, M., García, M.E., Duin, R.P.W., Castellanos, C.G.: Dissimilarity-based classification of seismic signals at Nevado del Ruiz Volcano. Earth Sci. Res. J. 10(2), 57–65 (2006)

[5] Paclik, P., Duin, R.P.W.: Dissimilarity-based classification of spectra: computational issues. Real Time Imaging 9(4), 237–244 (2003)

[6] Benbrahim, M., Daoudi, A., Benjelloun, K., Ibenbrahim, A.: Discrimination of seismic signals using artificial neural networks. In: Ardil, C. (ed.) WEC (2), Enformatika, Çanakkale, Turkey, vol. (2), pp. 4–7 (2005)

[7] Lesage, P., Glangeau, F., Mars, J.: Applications of autoregressive models and time-frequency analysis to the study of volcanic tremor and long-period events. Journal of Volcanology and Geothermal Research 114, 391–417 (2002)

[8] Curilem, G., Vergara, J., Fuentealba, G., Acuña, G., Chacón, M.: Classification of seismic signals at Villarrica Volcano (Chile) using neural networks and genetic algorithms. Journal of Volcanology and Geothermal Research 180(1), 1–8 (2009)

[9] Zobin, V.M.: Introduction to Volcanic Seismology. Developments in Volcanology, vol. 6. Elsevier, Philadelphia (2003)

[10] Zuo, W., Zhang, D., Wang, K.: An assembled matrix distance metric for 2DPCA-based image recognition. Pattern Recognition Letters 27, 210–216 (2006)

[11] Lesage, P.: Interactive Matlab software for the analysis of seismic volcanic signals. Computers & Geosciences 114, 391–417 (2009)

[12] Orozco-Alzate, M., Skurichina, M., Duin, R.P.W.: Spectral characterization of volcanic earthquakes at Nevado del Ruiz Volcano using spectral band selection/extraction techniques. In: Ruiz-Shulcloper, J., Kropatsch, W.G. (eds.) CIARP 2008. LNCS, vol. 5197, pp. 708–715. Springer, Heidelberg (2008)