TESTING SUPERVISED CLASSIFIERS BASED ON NON-NEGATIVE MATRIX FACTORIZATION TO MUSICAL INSTRUMENT CLASSIFICATION

Emmanouil Benetos Constantine Kotropoulos

Aristotle Univ. of Thessaloniki
Dept. of Informatics
Box 451, Thessaloniki 541 24, Greece
E-mail: {empeneto, costas}@aiia.csd.auth.gr

Thomas Lidy and Andreas Rauber

Vienna University of Technology
Dept. of Software Technology and Interactive Systems
Favoritenstrasse 9-11/188, A-1040 Vienna, Austria
E-mail: {lidy, rauber}@ifs.tuwien.ac.at

ABSTRACT

In this paper, a class of algorithms for automatic classification of individual musical instrument sounds is presented. Two feature sets were employed, the first containing perceptual features and MPEG-7 descriptors and the second containing rhythm patterns developed for the SOMeJB project. The features were measured for 300 sound recordings consisting of 6 different musical instrument classes. Subsets of the feature set are selected using branch-and-bound search, obtaining the most suitable features for classification. A class of supervised classifiers is developed based on the non-negative matrix factorization (NMF). The standard NMF method is examined as well as its modifications: the local and the sparse NMF. The experiments compare the two feature sets alongside the various NMF algorithms. The results demonstrate an almost perfect classification for the first set using the standard NMF algorithm (classification error 1.0%), outperforming the state-of-the-art techniques tested for the aforementioned experiment.

1. INTRODUCTION

The need for musical content analysis arises in different contexts and has many practical applications, mainly for automatic music transcription, effective data organization and annotation in multimedia databases, and internet search. Automatic musical instrument classification is the first step in developing such applications. It is a research area which can also be applied to general sound recognition tasks. However, despite the massive research which has been carried out in the automatic speech recognition, limited work has been done on musical content identification.

The experiments carried out so far can be broadly classified into two categories: classification of isolated instrument tones and classification of sound segments. Classifiers using isolated tones have a limited use in a practical application, while sound segment classifiers could be effectively used in music retrieval systems. Using 2-second segments and employing a back propagation neural network for 7 instrument classes, a classification accuracy of 99% is reported in [7]. Samples were extracted from the MIS Database from UIOWA [1] that is used in this paper as well. In addition, Synak et al [8] used MPEG-7 temporal descriptors and various spectral features for sound segments consisting of 18 instrument classes and developed 2 classifiers. The first classifier uses the k-NN algorithm, while the second one uses decision rules based on rough sets theory. They achieve a recognition rate of 68.4% at best.

Non-negative matrix factorization (NMF) is a subspace method for basis decomposition [4]. Its various modifications have been used in several classification experiments, where the training procedure is performed by applying an NMF algorithm to a data matrix containing the training vectors of all the available classes. This technique results to an unsupervised training approach. NMF classification experiments report encouraging results compared to other unsupervised classifiers, but also indicate that a supervised NMF classification approach is also needed to obtain comparable results with other supervised classifiers.

In this work, the problem of automatically classifying musical instrument segments is addressed. Recordings from the UIOWA database [1] were used that form 6 instrument classes. Two different feature sets were employed, the first covering perceptual descriptors as well as spectral descriptors defined by the MPEG-7 audio standard [2]. The first and second moments of the features were considered, creating a feature set of 41 dimensions as explained in Section 4.2. The second feature set uses rhythm patterns developed in the SOMeJB project, which are used for music archives organization [14]. The resulting feature dimension was 1440 for each recording. Branch-and-bound selection was applied to the feature set in order to select the subset that maximizes the classification accuracy [11]. The audio files were split into a training set and a test set using 70% of the available data for training and the remaining 30% for testing. For classification, NMF is used by training individually a classifier for each class and projecting the test data onto each trained class matrix. The class label of each test recording is determined by using the cosine similarity measure (CSM). Several variants of the NMF algorithm were employed, such as the standard NMF method, the local, and the sparse NMF. The results indicate that the 6-feature subset from the first feature set and the standard NMF algorithm yields a correct classification rate of 99.0%, outperforming the traditional unsupervised NMF classification methods and other statistical model-based classifiers employed for the aforementioned experiment [9].

The remainder of the paper is organized as follows. The two extracted feature sets are discussed in Section 2. Section 3 is devoted to the NMF method and its extensions, as well as the supervised NMF classifier. Section 4 describes the data set used, the feature selection strategy, and the experiments performed to assess the performance of the proposed classifier and the feature sets. Finally, conclusions are drawn in Section 5.

This work has been supported by the FP6 European Union Network of Excellence MUSCLE "Multimedia Understanding through Semantics, Computation, and Learning" (FP6-507752).

Table 1: List of feature set 1.

1	Zero-Crossing Rate
2	Delta Spectrum (Spectrum Flux)
3	Spectral Rolloff Frequency
4	Mel-Frequency Cepstral Coefficients
5	MPEG-7 AudioSpectrumCentroid
6	MPEG-7 AudioSpectrumEnvelope
7	MPEG-7 AudioSpectrumSpread
8	MPEG-7 AudioSpectrumFlatness
9	MPEG-7 AudioSpectrumProjection Coefficients

2. FEATURE EXTRACTION

In an audio classification system, the intention of the feature extraction step is to adequately and sufficiently describe the semantics of the audio content. In our approach, two feature sets were created, the first combining features describing the temporal and spectral sound structure. The second set is a time-invariant representation of fluctuation patterns on critical bands according to perception of the human auditory system.

2.1 Feature Set 1

In the first set, a combination of features originating from general audio data classification and the MPEG-7 audio framework is used. The complete list of extracted features is presented in Table 1.

The scalar features 1-3 are proposed in systems concerning general audio data (GAD) classification and speech recognition. They can be treated as a short-term description of the textural shape of the audio segments. The mel-frequency cepstral coefficients (MFCCs) form a feature vector. They are widely used in audio processing applications providing a description of the spectral shape of the audio signal. 13 MFCCs were used for each audio frame of 10 msec duration. The features 5-8 are proposed by the MPEG-7 audio standard [2]. They belong to the basic spectral descriptors category. As 9th feature we used the projection coefficient to a single basis. AudioSpectrumProjection coefficients are part of the MPEG-7 spectral basis descriptors.

2.2 Feature Set 2

Rhythm Patterns form the second feature set, developed as part of the SOMeJB project [12], [13], [14], [15] whose main focus is the description and automatic organization of music archives containing different styles or genres of music. The feature set is also suitable to differentiate between various classes of instruments, the reason being, that the feature set not only focuses on the description of rhythm in narrow sense, but also on fluctuations within the different pitch regions.

The algorithm for extracting the Rhythm Patterns is a two stage process: First, from the audio spectrum the specific loudness sensation according to the human auditory system is computed. Then, those values are transformed into a time-invariant domain resulting in a representation of modulation amplitudes per modulation frequency on several frequency regions (critical bands). In the following, we will give an outline of all the steps involved in the feature extraction process. An overview of the procedure is depicted in Figure 1.

The algorithm processes audio tracks in standard digital PCM format with 44.1 kHz sampling frequency as input. Each audio track is segmented into pieces of 6 seconds length. A short time Fast Fourier Transform (STFT) is applied to retrieve the energy per frequency band (the spectrum) every 11.5 ms, resulting in a spectrogram of the 6 second segment. The frequency bands of the spectrogram are summarized to 24 critical bands, according to the Bark scale [16].

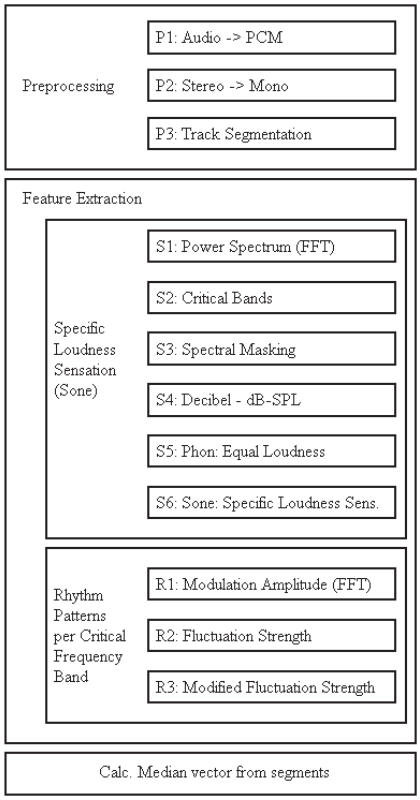

Figure 1: Block diagram of Rhythm Pattern extraction.

The data is then transformed into the logarithmic decibel scale. For transformation into the unit Phon the algorithm incorporates the so-called equal-loudness curves, which account for different loudness sensation of humans in different frequency regions. Afterwards a conversion into the unit Sone is done, reflecting the specific loudness sensation of the human auditory system according to loudness levels. At this point, we retrieved the specific loudness sensation over time on 24 critical frequency bands. Still, we have a time-dependent signal, although reduced to 511 sample values at the time axis due to the window size in the STFT.

In order to obtain a time-independent representation of the data, another Fourier Transform is applied. The idea is to regard the varying energy on a frequency band of the spectrogram as a modulation of the amplitude over time. With the second Fourier Transform, the spectrum of this modulation signal is retrieved. It is a time-invariant signal that denotes the modulation frequency on the abscissa, and the magnitude of modulation on the ordinate. The notion of rhythm ends above 15 Hz, where the sensation of roughness starts and goes up to 150 Hz, the limit where only three separately au-

dible tones are perceivable. The algorithm captures modulation frequencies up to 43 Hz, however the algorithm is set to cut off the information above a modulation frequency of 10 Hz. Subsequently, modulation amplitudes in that range are weighted according to a function of human sensation depending on modulation frequency, accentuating values around 4 Hz, followed by the application of a gradient filter and Gaussian smoothing.

The final feature vector contains a time-invariant representation of fluctuation strength according to human sensation between 0.168 Hz and 10 Hz of modulation frequency on 24 critical frequency band regions. A feature vector for each 6 second segment of a piece of audio is calculated. In order to summarize the characteristics of an entire piece of audio (especially music) we average the feature vectors derived from its segments by computing the median.

3. NON-NEGATIVE MATRIX FACTORIZATION

Non-negative matrix factorization (NMF) has been proposed as a novel subspace method in order to obtain a parts-based representation of objects by imposing non-negative constraints [4]. The problem addressed by NMF is as follows. Given a non-negative $n \times m$ data matrix $\mathbf{V}$ (consisting of m vectors of dimensions $n \times 1$), it is possible to find non-negative matrix factors $\mathbf{W}$ and $\mathbf{H}$ in order to approximate the original matrix:

$$\mathbf{V} \approx \mathbf{WH} \tag{1}$$

where the $n \times r$ matrix $\mathbf{W}$ contains the basis vectors and the $r \times m$ matrix $\mathbf{H}$ contains in its columns the weights needed to properly approximate the corresponding column of matrix $\mathbf{V}$ as a linear combination of the columns of $\mathbf{W}$. Usually, the component number r is chosen so that $(n+m)r < nm$, thus resulting in a compressed version of the original data matrix.

To find an approximate factorization in (1), a suitable objective function has to be defined. The generalized Kullback-Leibler (KL) divergence between $\mathbf{V}$ and $\mathbf{WH}$ is the most frequently used objective function. Various algorithms that incorporate additional constraints in deriving (1) have been proposed that are briefly reviewed subsequently.

3.1 Standard NMF

The standard NMF enforces the non-negativity constraints on matrices $\mathbf{W}$ and $\mathbf{H}$. Thus, a data vector can be formed by an additive combination of basis vectors. The proposed cost function is the generalized KL divergence:

$$D(\mathbf{V}||\mathbf{WH}) = \sum_{i=1}^{n}\sum_{j=1}^{m}[v_{ij}\log\frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij}] \tag{2}$$

where $\mathbf{WH} = \mathbf{Y} = [y_{ij}]$. $D(\mathbf{V}||\mathbf{WH})$ reduces to KL divergence when $\sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} = 1$. NMF factorization is defined then as the solution of the optimization problem:

$$\min_{\mathbf{W},\mathbf{H}} \; D(\mathbf{V}||\mathbf{WH}) \quad subject\ to \;\; \mathbf{W},\mathbf{H} \geq 0, \sum_{i=1}^{n} w_{ij} = 1 \; \forall j \tag{3}$$

where $\mathbf{W},\mathbf{H} \geq 0$ means that all elements of matrices $\mathbf{W}$ and $\mathbf{H}$ are non-negative. The above optimization problem can be solved by using the iterative multiplicative rules [4].

3.2 Local NMF (LNMF)

Aiming to impose constraints concerning spatial locality and consequently revealing local features in the data matrix $\mathbf{V}$, LNMF incorporates 3 additional constraints into the standard NMF problem: 1) Minimize the number of basis components representing $\mathbf{V}$. 2) The different bases should be as orthogonal as possible. 3) Retain the components giving most important information. The above constraints are expressed in the following LNMF cost function:

$$\begin{aligned} D(\mathbf{V}||\mathbf{WH}) &= \sum_{i=1}^{n}\sum_{j=1}^{m}[v_{ij}\log\frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij}] \\ &+ \alpha\sum_{i=1}^{r}\sum_{j=1}^{r}u_{ij} - \beta\sum_{i=1}^{r}\sum_{j=1}^{r}q_{ii} \end{aligned} \tag{4}$$

where α, β are constants, $\mathbf{W}^T\mathbf{W} = \mathbf{U} = [u_{ij}]$, and $\mathbf{HH}^T = \mathbf{Q} = [q_{ij}]$. The minimization is similar to the one used in NMF (3) and a local solution can be found by using 3 update rules, where α and β are considered equal to 1 [5].

3.3 Sparse NMF (SNMF)

Inspired by NMF and sparse coding, the aim of SNMF is to impose constraints that can reveal local sparse features on data matrix $\mathbf{V}$. The following cost function is optimized for SNMF:

$$D(\mathbf{V}||\mathbf{WH}) = \sum_{i=1}^{n}\sum_{j=1}^{m}[v_{ij}\log\frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij}] + \lambda\sum_{j=1}^{m}||\mathbf{h}_j||_l \tag{5}$$

where λ is a positive constant and $||\mathbf{h}_j||_l$ the l-norm of the j-th column of $\mathbf{H}$. An SNMF factorization is defined as in (3), including also that $\forall i ||\mathbf{w}_i||_l = 1$. In SNMF, the sparseness is measured by a linear activation penalty, the minimum l-norm of the column of $\mathbf{H}$. A local solution of the minimization problem (5) can be obtained by the update rules proposed in [6].

3.4 Supervised NMF Classification

The major drawback of unsupervised NMF classification presented in [9] is the manner of learning parts-based patterns from the data, since no information about the class discrimination is incorporated into the NMF training procedure. In addition, the initial random values of matrices $\mathbf{W}$ and $\mathbf{H}$ can affect the convergence of the algorithm, as the value of NMF objective function defined in (2) may result in a local minimum, thus not yielding in an appropriate factorization.

In this paper, a supervised classifier where the NMF training procedure is performed for each data class individually is applied, thus resulting in a pair of matrices $\mathbf{W}$ and $\mathbf{H}$ for each class:

$$\mathbf{V}_i = \mathbf{W}_i\mathbf{H}_i, \quad i = 1, 2, \cdots, N \tag{6}$$

where $\mathbf{N}$ is the number of different classes, $\mathbf{V}_i$ the data matrix of class i. The number of components used for training each class is given by:

$$r_i = \left\lfloor \frac{n_i m_i}{n_i + m_i} \right\rfloor \tag{7}$$

where n_i and m_i are the dimensions of matrix $\mathbf{V}_i$. In a sense, this approach is an application of one-class classification,

where the training of each class is performed individually, by using a set of training data representing the respective class in the absence of counter-examples [10].

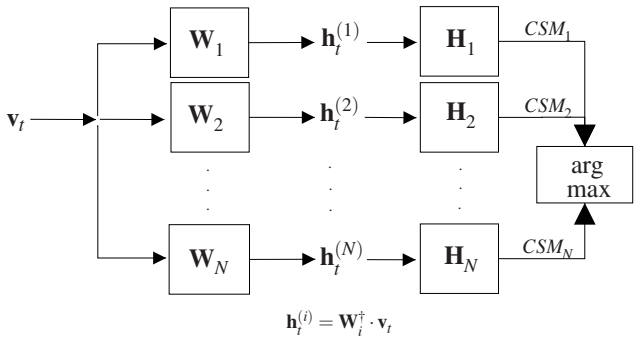

Figure 2: Testing using the supervised NMF classifier ($\mathbf{h}_t$ and $\mathbf{v}_t$ stand for $\mathbf{h}_{test}$ and $\mathbf{v}_{test}$ respectively).

During test procedure, each test sound is represented by the feature vector $\mathbf{v}_{test}$. Afterwards, $\mathbf{v}_{test}$ is projected onto each class basis matrix $\mathbf{W}_i$, yielding:

$$\mathbf{h}_{test}^{(i)} = \mathbf{W}_i^{\dagger} \cdot \mathbf{v}_{test} \tag{8}$$

For each class, the vector $\mathbf{h}_{test}^{(i)}$ is compared to each column vector of matrix $\mathbf{H}_i$ using the cosine similarity measure. The vector that maximizes the CSM for the matrix $\mathbf{H}_i$ is calculated as a measure of similarity for this class:

$$CSM_i = \max_{j=1,2,\ldots,r_i} \left\{ \frac{\mathbf{h}_{test}^{(i)T} \mathbf{h}_j^{(i)}}{\|\mathbf{h}_{test}^{(i)}\| \|\mathbf{h}_j^{(i)}\|} \right\} \tag{9}$$

where $\mathbf{h}_j^{(i)}$ represents the j-th column of matrix $\mathbf{H}_i$. Finally, the class label of the recording is determined by the the maximum CSM_i, i.e.:

$$l' = \arg \max_{i=1,2,\ldots,N} \{CSM_i\} \tag{10}$$

A block diagram of the testing procedure using the supervised NMF classification method is plotted in Figure 2.

4. EXPERIMENTAL RESULTS

4.1 Dataset

Audio files extracted from the Musical Instrument Samples database collected by the university of Iowa [1] were used. 300 audio files were extracted that belong to 6 different instrument classes: piano, violin, cello, flute, bassoon, and soprano saxophone. In detail, 58 piano recordings, 101 violin recordings, 52 cello recordings, 31 saxophone recordings, 29 flute recordings, and 29 bassoon ones were used. The 300 sounds are partitioned into a training set of 210 audio files and a test set of 90 audio files, which is typical for classification experiments. All recordings are discretized at 44.1 kHz and have a duration of about 20 sec.

4.2 Feature selection

Regarding the first feature set, for each feature described in Table 1, its mean and its variance were computed, resulting in 41 features in total. The feature dimension of the rhythm patterns of the second set is 1440, which is quite a large value for training classification algorithms.

In order to reduce the feature vector dimension for both sets, a suitable feature subset for classification has to be selected. The optimal feature subset should maximize the ratio of the inter-class dispersion over the intra-class dispersion:

$$J = \mathrm{tr}(\mathbf{S}_w^{-1} \mathbf{S}_b) \tag{11}$$

where $\mathrm{tr}(\cdot)$ stands for the trace of a matrix, $\mathbf{S}_w$ is the within-class scatter matrix, and $\mathbf{S}_b$ is the between-class scatter matrix. Because the number of distinct subsets is $\frac{N!}{(N-D)!D!}$, where D is the desired subset size and N the feature dimension, the branch-and-bound search strategy is considered for complexity reduction. In this strategy, a tree structure of $(N-D+1)$ levels is created, where every node corresponds to a subset. The highest level corresponds to the full set, while each node corresponds to a D-dimensional subset at the lowest level. The branch-and-bound algorithm traverses the structure using a depth-first search with backtracking [11].

4.3 Performance Evaluation

Two separate experiments on the various NMF algorithms have been performed by using the two feature sets described in Section 2. Subsets of the feature sets were created using feature selection, in order to find the feature dimension that maximizes the classification performance. For the first set, 6 features were used from the total 41 (mainly the moments of the first two MFCCs). For the second set, 50 features were selected from the total 1440.

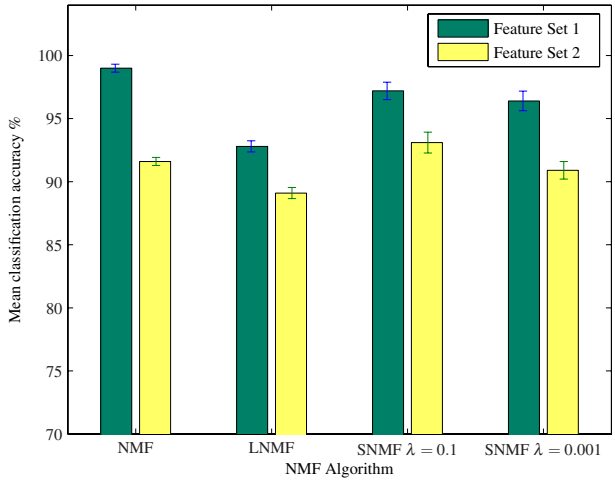

Figure 3: Mean classification accuracy for NMF algorithms.

Experiments were carried out using 7-fold cross validation and the mean value of the classification accuracy and its standard deviation for the three NMF algorithms and for the two feature sets is shown in Figure 3. The SNMF algorithm was tested using two different values for the parameter λ (0.001 and 0.1). The highest mean accuracy of 99.0% is achieved by the standard NMF algorithm when the subset of 6 features from the first feature set is used. Using the second feature set, the accuracy of NMF exceeds 91.5%. The achieved results outperform the classification accuracy for the aforementioned experiment using unsupervised NMF

classification [9]. In addition, the results using the first set outperform the supervised classifiers based on gaussian mixture models (GMM) and continuous hidden Markov models (HMM) also utilized in [9]. Generally, the performance of the classifier diminishes when the second feature set is utilized. The main reason is the large feature dimension of the rhythm patterns, which could be diminished by using statistical moments to describe the feature vector. The highest accuracy for the second set is achieved using the SNMF algorithm for $\lambda = 0.1$, being 93.1%. The LNMF is clearly outperformed by all algorithms, which may be explained due to the locality constraints LNMF imposes when applied to holistic descriptors. The SNMF overall displays better results than the LNMF, but its efficiency depends on the selection of parameter λ (performance is slightly better when $\lambda = 0.001$).

Table 2: Confusion matrix for standard NMF, Feature Set 1.

Instr.	Piano	Bassoon	Cello	Flute	Sax	Violin
Piano	**18**	0	0	0	0	0
Bassoon	0	**9**	0	0	0	0
Cello	0	0	**16**	0	0	0
Flute	1	0	0	**8**	0	0
Sax	0	0	0	0	**9**	0
Violin	0	0	0	0	0	**29**

Table 3: Confusion matrix for standard NMF, Feature Set 2.

Instr.	Piano	Bassoon	Cello	Flute	Sax	Violin
Piano	**18**	0	0	0	0	0
Bassoon	0	**9**	0	0	0	0
Cello	0	0	**14**	0	0	2
Flute	3	0	0	**6**	0	0
Sax	0	0	0	0	**9**	0
Violin	0	0	2	0	0	**27**

Additional information about the performance of the standard NMF algorithm using the two sets is shown in Tables 2 and 3 in the form of a confusion matrix. The columns of the confusion matrix correspond to the predicted musical instrument and the rows to the actual one. For the first set, a single misclassification occurs. For the second set, most misclassifications occur for the flute, as well as for the violin and cello.

5. CONCLUSIONS

In this paper, a method of classifying musical instrument recordings by using supervised NMF classifiers using two different feature sets has been presented. The results indicate that the standard NMF algorithm used in conjunction with the first set can perform classification with a high accuracy compared to its variants (LNMF and SNMF).

In the future, NMF techniques will be applied to discriminate the whole spectrum of orchestral instruments and will be also used in general sound classification experiments. Finally, statistical moments of the rhythm patterns could be used instead of the feature vector in order to improve classification accuracy.

REFERENCES

[1] Univ. of Iowa Musical Instrument Sample Database, http://theremin.music.uiowa.edu/index.html.

[2] MPEG-7 overview (version 9), *ISO/IEC JTC1/SC29/WG11 N5525*, March 2003.

[3] H. G. Kim, N. Moreau, and T. Sikora, "Audio classification based on MPEG-7 spectral basis representations," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 716-725, May 2004.

[4] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Adv. in Neural Information Processing Systems*, vol. 13, pp. 556-562, 2001.

[5] S. Z. Li, X. Hou, H. Zhang, and Q. Cheng, "Learning spatially localized, parts-based representation," in Proc. *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-6, 2001.

[6] C. Hu, B. Zhang, S. Yan, Q. Yang, J. Yan, Z. Chen, and W. Ma, "Mining ratio rules via principal sparse non-negative matrix factorization," in Proc. *IEEE Int. Conf. Data Mining*, 2004.

[7] A. Livshin, and X. Rodet, "The importance of cross database evaluation in musical instrument sound classification: a critical approach," in Proc. *Int. Symp. Music Information Retrieval*, October 2003.

[8] A. Wieczorkowska, J. Wroblewski, P. Synak, and D. Slezak, "Application of temporal descriptors to musical instrument sound recognition," *J. Intelligent Information Systems*, vol. 21, no. 1, pp. 71-93, July 2003.

[9] E. Benetos, M. Kotti, C. Kotropoulos, J. J. Burred, G. Eisenberg, M. Haller, and T. Sikora, "Comparison of subspace analysis-based and statistical model-based algorithms for musical instrument classification," *2nd Workshop On Immersive Communication And Broadcast Systems*, October 2005.

[10] D. M. J. Tax, *One-Class Classification*, PhD thesis, Delft University of Technology, The Netherlands, 2001.

[11] F. van der Heijden, R. P. W. Duin, D. de Ridder, and D. M. J. Tax, *Classification, Parameter Estimation and State Estimation: An Engineerign Approach using MATLAB*. London UK: Wiley, 2004.

[12] A. Rauber and M. Frühwirth, "Automatically analyzing and organizing music archives," in Proc. *European Conf. Research and Advanced Technology for Digital Libraries*, Springer Lecture Notes in Computer Science, Darmstadt, Germany, September 2001.

[13] A. Rauber, E. Pampalk, and D. Merkl, "Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by musical styles," in Proc. *Int. Conf. Music Information Retrieval*, pp. 71-80, Paris, France, October 2002.

[14] A. Rauber, E. Pampalk, and D. Merkl, "The SOM-enhanced JukeBox: organization and visualization of music collections based on perceptual models," *J. New Music Research*, Vol. 32, No. 2, pp. 193-210, June 2003.

[15] T. Lidy and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification," in Proc. *Int. Conf. Music Information Retrieval*, pp. 34-41, London, UK, September 2005.

[16] E. Zwicker and H. Fastl, "Psychoacoustics - Facts and Models," *Springer Series of Information Sciences*, Vol. 22, Springer, Berlin, 1999.