A Fast Algorithm for Music Search by Similarity in Large Databases based on Modified Symetrized Kullback Leibler Divergence

Christophe Charbuillet, Geoffroy Peeters
{charbuillet,peeters}@ircam.fr
IRCAM-CNRS STMS
1 place Igor Stravinsky, 75004 Paris

Stanislav Barton, Valerie Gouet-Brunet
{stanislav.barton, valerie.gouet}@cnam.fr
CNAM/CEDRIC
292 rue Saint Martin, F75141 Paris Cedex 03

Abstract

State of the art on music similarity search is based on the pairwise comparison of statistical models representing audio features. The comparison is often obtained by the Symetrized Kullback-Leibler Divergence (SKLD). When dealing with very large databases (over one million items), usual search by similarity algorithms - sequential or exhaustive search - cannot be used. In these cases, optimized search strategies such as the M-tree reduces the search time but requires the dissimilarity measure to be a metric. Unfortunately, this is not the case of the SKLD. In this paper, we propose and successfully test on a large-scale a modification of the Symetrized Kullback-Leibler Divergence which allows to use it as a metric.

1. Introduction

Music similarity systems are about to become essential tools for music diffusion and promotion. Nowadays, online systems mainly rely on audio meta data (i.e. tag, ID3[1]) or user preferences. This is the case for example for the systems proposed by iTunes®, www.dezzer.fr, www.last.fm or www.amazon.fr. The main principle of this approach is based on the assumption that songs within the same genre, label and artist are more similar than others. Another approach recently developed consists in measuring the audio similarity based on signal analysis, able to allow transverse database recommendation (because two different artists from different labels can produce quite similar music styles). Several methods were proposed to mimic human similarity perception. In [7] the authors propose a similarity measure based on the alignment cost of two items by Dynamic Time Warping (DTW), efficient for cover-version detection. M. Casey and M. Slaney use a vectorial representation of the song characteristics and compared the so-called audio shingle by the L_2 norm [2] in order to detect couples of songs containing similar portions. Systems dedicated to music recommendation are based on statistical modeling of short term audio features. The model used can be a Gaussian Mixture Model (GMM) as proposed in [1, 9], or a single Gaussian Model with full covariance matrix [13, 10, 14] which provide similar performances. The measure used to compare the models is the Symetrized Kullback-Leibler Divergence (SKLD) [8] or alternatively the Earth Mover's Distance based on the SKLD when models are GMM [9]. When dealing with databases of millions items, usual search by similarity algorithm - sequential or exhaustive search - cannot be used. When music characteristics are stored on a feature and compared by the euclidean distance, efficient index structure, like Local Sensitive Hashing (LSH) [6] can be employed [2]. Unfortunately, the systems presented above do not satisfy this condition. The scalable indexation of Gaussian Model compared by the SKLD still remains an open problem. To our knowledge the only publication dealing with this specific problem is [15] in which the author proposed an approximate method based on the Fast Map Algorithm [12].

The main difficulty of indexing these models is the fact that the SKLD does not hold the triangle inequality and consequently, is not a metric. This fact prevent the use of efficient index structure like M-tree [5], PM-tree [18] and Pivot based approaches [3].

In this paper, we show that the function $\mathcal{T} : x \rightarrow \sqrt{log(x+1)}$ turns the Symetrized Kullback-Leibler Divergence into an exact metric when the statistical models compared are Gaussian. This includes multidimensional normal distributions with diagonal or full covaraince matrix. Besides, this transformation preserves the original similarity ordering (i.e. $a > b \Leftrightarrow \mathcal{T}(a) > \mathcal{T}(b)$). Consequently the transformed SKLD can be used by metric access methods for fast similarity search.

Based on this new distance, we propose a modification

[1] www.id3.org

978-1-4244-8027-2/10/$26.00 ©2010 IEEE

of the use of the M-tree index structure allowing performing both exact and approximate nearest neighbor search.

Results obtained with a database composed of 1 million of timbral gaussian models show that similarity retrieval can be performed an order of magnitude faster than sequential search, with high accuracy.

This paper is organized as follow: in section 2, we present the basis of the semi-metric modification framework for data indexing. In section 3, we show how this approach can be applied to the SKLD divergence and propose a new metric for single Gaussian model comparison. Then experiments using a M-tree structure are proposed in section 4. Finally we will present our conclusions and our perspectives in section 5.

2 Metric and semi-metric spaces

2.1 Definitions

Lets $O_i, O_j, O_k \in \mathbb{U}$ be three multimedia objects and $\delta : \mathbb{U} \times \mathbb{U} \mapsto \mathbb{R}$ a dissimilarity measure. $\delta(.,.)$ is a metric distance if :

$$\delta(O_i, O_j) = 0 \Leftrightarrow O_i = O_j \quad (1)$$
$$\delta(O_i, O_j) > 0 \Leftrightarrow O_i \neq O_j \quad (2)$$
$$\delta(O_i, O_j) = \delta(O_j, O_i) \quad (3)$$
$$\delta(O_i, O_j) + \delta(O_j, O_k) \geq \delta(O_i, O_k) \quad (4)$$

This conditions are called the reflexivity (1), non-negativity (2), symmetry (3) and triangle inequality (4). When the last condition is not respected the dissimilarity measure is called *semi-metric*.

2.2 Turning semi-metric to metric

The triangle inequality condition is fundamental for metric acces methods. It allows to lower and upper bound the distance between the query object and the database object using the precomputed distances between reference objects and database objects. This bounds are given by:

$$\delta(O_i, O_k) \leq \delta(O_i, O_j) + \delta(O_j, O_k) \quad (5)$$
$$\delta(O_i, O_k) \geq \|\delta(O_i, O_j) + \delta(O_j, O_k)\| \quad (6)$$

where O_i is the query object, O_k a object within the indexed database and O_j a reference object.

It is used for example by pivot based search methods [14] to filter the irrelevant objects from the database. A similar approach is used by M-tree index structures for discarding irrelevant nodes [5].

When the triangle inequality is not hold no general assumption can be done on distances between data objects. Therefore, specific index structures must be built, taking into account the properties of the considered dissimilarity measure. Another approach, recently proposed by T. Skopal [16, 17] is to transform the semi-metric dissimilarity into a metric one by the use of a so-called Triangle-Generating modifier (TG-modifier) function. We will now resume the principle of this approach. A TG-modifier function T, able to turn the semi-metric Δ into a metric one must satisfy:

$$T : \mathbb{R}^+ \mapsto \mathbb{R}^+ \quad (7)$$
$$T(0) = 0 \quad (8)$$
$$a < b \Leftrightarrow T(a) < T(b) \quad (9)$$
$$T(\Delta(O_i, O_k)) \leq T(\Delta(O_i, O_j)) + T(\Delta(O_j, O_k)) \quad (10)$$

The two first equations (7,8) guarantee the reflexivity and the non-negativity of the original semi-metric. The condition (9) is needed to ensure that the similarity order is preserved. Then equation (10) impose to $T[\Delta(.,.)]$ to hold the triangle inequality. To illustrate this principle by a simple exemple, let us consider the dissimilarity $\Delta(\mathbf{a}, \mathbf{b}) = (\mathbf{a} - \mathbf{b})^2$ with $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. One can observe that the triangular inequality does not hold : $\Delta(1,2) + \Delta(2,3) \not\leq \Delta(1,3)$. Consequently, Δ is a semi-metric. The function $\sqrt{(\cdot)}$ turn the dissimilarity $\Delta(.,.)$ into the euclidean distance and therefore is a TG-modifier of Δ.

Actually, there exist an infinity of functions which able to transform Δ into a true metric distance. This is the case of $T_\gamma : x \rightarrow x^{\frac{1}{\gamma}}\ \forall \gamma \in [2, +\infty[$. But as we will see in the following section, some of these functions are more suitable for data indexability.

2.3 Consequences on indexability

The distance distribution between the database objects (called pairwise distance distribution) allows picturing out the topology of the data and consequently their indexability. In [3], the authors propose a simple measure base on the Distance Distribution Histogram (DDH) to evaluate the intrinsic dimensionality of the data:

$$\rho(\mathbb{S}, \delta) = \frac{\mu^2}{2\sigma^2} \quad (11)$$

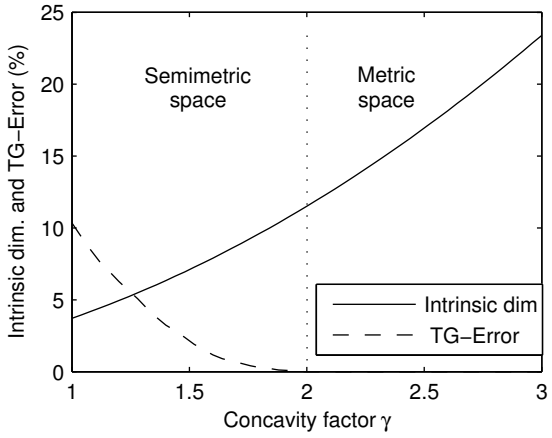
where $\mathbb{S}$ is a dataset, δ a dissimilarity measure and μ and σ are the mean and standard deviation of the DDH. A low intrinsic dimensionality (< 10) indicates that it exists

a high disparity between the distances, including close objects, able to be clustered. When ρ is high, all the objects within the database are quite equidistant. In this case, the indexability of the data is known to be very poor.

When the dissimilarity function is semi-metric it is possible to measure this "distance" to a real metric by measuring the amount of triangular inequality violations. The TG-error [16], is defined as the probability of the event $\Delta(O_i, O_k) > \Delta(O_i, O_j) + \Delta(O_j, O_k)$ when O_i, O_j, O_k are randomly sampled from the database $\mathbb{S}$.

In order to illustrate these two concepts let us consider the semi-metric $\Delta(\mathbf{a}, \mathbf{b}) = (\mathbf{a} - \mathbf{b})^2$ with $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and the TG-modifier function class: $T_\gamma : x \rightarrow x^{\frac{1}{\gamma}}$. The objects considered are 5 dimensional vectors, uniformly distributed in $[0, 1]$. The parameter γ can be viewed as a *concavity factor* (the greater γ parameter, the more concave the function is). Fig. 1 shows the influence of γ on the intrinsic dimension and on the TG-Error of the dissimilarity measure $T_\gamma[\Delta(\mathbf{a}, \mathbf{b})]$.

Figure 1. Influence of the concavity factor on intrinsic dimensionality and TG-Error



We can notice than the intrinsic dimensionality monotonically increases with γ whereas the TG-error rate decrease. When $\gamma \geq 2$ the TG-Error falls to zero and the dissimilarity becomes a true metric. We can conclude that the better function within the function class T_γ is T_2 (euclidean distance) because it shows the lower intrinsic dimensionality and still holds the triangle inequality.

Another aspect of the this framework is the possibility to allow a small probability of triangle inequality violation in order to perform a faster but approximative similarity search. This can be done by setting the concavity factor according to the desired similarity search precision (possibly at query time). This idea was also introduced by T. Skopal in [17].

3 Turning the Symetrized Kullback-Leibler Divergence into metric

3.1 Definitions

The Kullback Leibler Divergence (KLD) [8], also known as the relative entropy is defined by:

$$KLD(P||Q) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} dx \quad (12)$$

where P and Q are two random variables and p, q are their probability density functions. A variant, widely used in music retrieval is the so-called Symetrized Kullback Leibler Divergence (SKLD) :

$$SKLD(P||Q) = \frac{KLD(P||Q) + KLD(Q||P)}{2} \quad (13)$$

The SKLD is a semi-metric. If P and Q are normally distributed this dissimilarity becomes :

$$SKLD(P||Q) = \quad (14)$$
$$\frac{1}{4}\left(tr(\Sigma_p^{-1}\Sigma_q) + (\mu_p - \mu_q)^T\Sigma_p^{-1}(\mu_p - \mu_q) - N\right)$$
$$+\frac{1}{4}\left(tr(\Sigma_q^{-1}\Sigma_p) + (\mu_q - \mu_p)^T\Sigma_q^{-1}(\mu_q - \mu_p) - N\right)$$

where Σ is the covariance matrix, μ is the mean vector and N is the dimension of the multivariate distributions.

3.2 SKLD TG-modifier and concavity controler

The SKLD TG-modifier we propose is:

$$\mathcal{T} : x \rightarrow \sqrt{\log(1 + x)} \quad (15)$$

As we will see in the next section, this function turns the SKLD into a true metric. Unfortunately, there is no direct mean to control its concavity. To overcome this drawback, we introduce an algorithm able to approximate any continuous function and to control its concavity.

Lets $f : \mathbb{R}^+ \mapsto \mathbb{R}^+$ be the original function and $\widehat{f}_\gamma : \mathbb{R}^+ \mapsto \mathbb{R}^+$ its concavity controlled version. The main idea

Figure 2. Application of the concavity controlled algorithm to the function $\sqrt{\log(x+1)}$

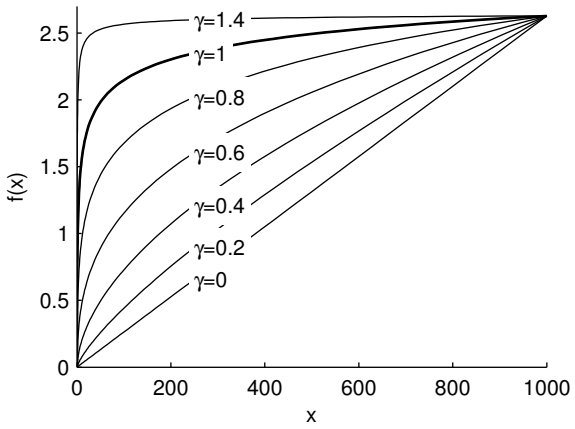
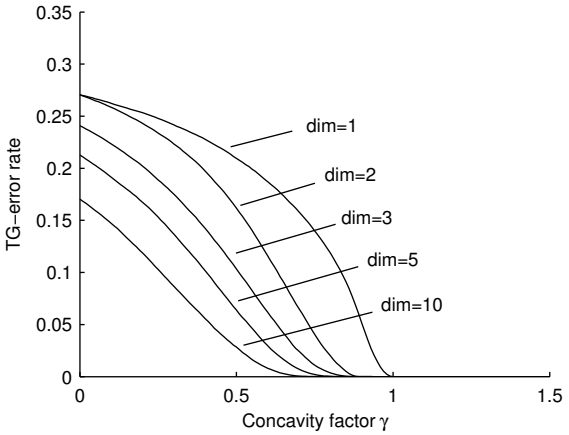


Figure 3. Evaluation of the metricity of $\mathcal{T}_\gamma(SKLD)$



of our algorithm consists in modelizing the original function by a piecewise linear function in which the gradient of the segments are defined recursively by :

$$\alpha_0 = f'(0) \tag{16}$$

$$\alpha_{i+1} = \frac{\alpha_i}{1+\gamma c_i} \tag{17}$$

where α_i is the gradient of the i^{th} segment of the function $\widehat{f}_\gamma$, γ is the concavity factor and c_i is the local concavity coefficient. These coefficients are obtained from the original function by:

$$c_i = \frac{\alpha_i}{\alpha_{i+1}} - 1 \tag{18}$$

$$= \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} \frac{x_{i+2} - x_{i+1}}{f(x_{i+2}) - f(x_{i+1})} - 1 \tag{19}$$

where f is the original function and x_i represents the abscise of the begin of the segment i. The function reconstruction consists of two steps. First the values of $\widehat{f}_\gamma$ in x_i are iteratively computed by:

$$\widehat{f}_\gamma(0) = f(0) \tag{20}$$

$$\widehat{f}_\gamma(x_{i+1}) = \widehat{f}_\gamma(x_i) + (x_{i+1} - x_i)\alpha_i \tag{21}$$

Then, for all $a \in \mathbb{R}^+$, $\widehat{f}_\gamma(a)$ is obtained by a linear interpolation:

$$\widehat{f}_\gamma(a) = \widehat{f}_\gamma(x_k) + (a - x_k)\alpha_k \tag{22}$$

$$\text{where } k = \min_i (a - x_i) \ \forall \ x_i < a$$

Fig. 2 shows the original SKLD TG-modifier $\mathcal{T}$ and its concavity controlled version for several values of γ. As we can observe, $\gamma = 0$ leads to a linear function whereas $\gamma = 1$ reproduces the original function. When $\gamma > 1$, the obtained function is more concave than the original one.

3.3 Numeric evidence of metricity

Pending a mathematical demonstration of the metricity of $\mathcal{T}(SKLD)$ we present here an empirical evidence based on a numerical experiments. This experiment consists in :

1- building a large database of uniformly distributed gaussian models in several dimensions;
2- computing N distance triplets $SKLD(O_i, O_j)$, $SKLD(O_j, O_k)$ and $SKLD(O_i, O_k)$;
3- modifying these distances by $\hat{\mathcal{T}}_\gamma$ and measure the TG-error.

The synthetic models are generated by :

$$\mu = \mathcal{U}_{[-1,1]}^{[N\times 1]} \tag{23}$$

$$E = \mathcal{U}_{[0,1]}^{[N\times N]} \tag{24}$$

$$\Sigma = EE^T \tag{25}$$

where μ and Σ are the mean and covariance matrix of the gaussian model and $\mathcal{U}_{[0,1]}^{[N\times N]}$ is a multivariate random distribution of dimension $N \times N$ uniformly distributed in $[0, 1]$.

For each dimension in $\{1, 2, 3, 5, 10\}$ we build 10000 models and measure the SKLD divergences of 100000

triplets. Fig. 3 shows the TG-error relative to the dissimilarity $\mathcal{T}_\gamma(SKLD)$ according to the concavity factor γ.

As we can observe, the TG-error rate is strictly equal to zero when $\gamma \geq 1$ implying that the triangle inequality hold. It is important to remember that $\mathcal{T}_{\gamma=1} = \sqrt{\log(1+x)}$. Assuming that the synthetic model generated fully describes the model space, we can argue that $\sqrt{log(SKLD(.,.)+1)}$ is a true metric when the statistical distributions compared are multivariate Gaussian.

4 Indexing experiments

In this section, we propose to apply the new metric presented above to the indexation of timbral model music. We first present the database, then we describe the index structure we used and its adaptation to approximate search. Then we present the performances of the system.

4.1 Databases

Our database is composed of timbral model of music, build on MFCC features [11] . First, the 13 dimensional MFCC vectors are extracted from audio signals by an analysis window of 20ms shifted with a step of 10 ms. Then, a set of models are extracted by modeling the MFCC sequence using a window of 10s shifted with a step of 3s. The statistical model used is a multivariate Gaussian model with full covariance matrix. One million models are extracted from 15000 music songs originally provided by the open European Archive [2]. The files, originally encoded in mp3 are down-sampled to 22050kHz before processing. This database can be viewed as a representative subset of the web music. It is fully diversified according to music genre and presents a large diversity in audio quality. To perform the query, an independent database (the query-database) is made from 100 models extracted from 100 new music files.

4.2 M-tree index structure

The Metric tree index structure [5] is one of the most popular metric index structures. It consists of hierarchical clusters of data, build from their pairwise distances only. While the original M-tree was strictly dedicated to true metric distances, we proposed here to use it in a quite different way.

Actually, the metric assumption is only needed by the search algorithm. The bulk-loading algorithm [4] (used to create a M-tree from scratch) only takes into account the similarity ordering of the data. Thus the use of the metric $\mathcal{T}_1(SKLD)$ or the use of the original semi-metric $SKLD$ will exactly lead to the same tree structure. Consequently, we propose to build the M-tree using the original $SKLD$ semi-metric and to modify the distances at query time, according to the desired results precision.

[2] http://www.europarchive.org/

We bulk-loaded the M-tree parameters with a min node utilization of 5 (u_{min}) and a max node utilization of 20 (u_{max}). We create two M-trees with 500k items and 1M items, respectively. The creation cost was of 1h and 2h on a [64bits,2GHz] computer for the two bases, respectively.

4.3 Results

In this part, we present an evaluation of the similarity system based on the k-NN query, i.e. the retrieval of the k first objects. Performances are presented among three criteria: the time of the query, the relative computation rate which relies on the comparison with the sequential scan cost, and the precision of the returned result. The precision is defined as the ratio between the correct nearest neighbors retrieved and the number of returned nearest neighbors. The presented results were obtained on two databases (500k items and 1M items) using 100 query objects which do not belong to these databases. Table 1 and Table 2 present the obtained results.

Table 1. 20 KNN performances on 500k objects

γ	Precision (%)	Computation rate (%)	Time (s)
0	87.3	9.08	0.398
0.1	91.3	10.9	0.388
0.2	94.9	13.4	0.47
0.3	97.1	16.9	0.583
0.4	98.7	21.8	0.756
0.5	99	28.7	1.03
0.6	99.3	38.2	1.43
0.7	99.6	50.8	1.89
0.8	99.6	65.3	2.36
0.9	99.6	79.4	2.8
1	100	91.5	3.21

First, we notice that exact nearest neighbors are effectively retrieved for $\gamma = 1$ but for a prohibitive cost of 91.5% and 90.8% of the sequential scan cost, for the 500k and 1M databases, respectively. Nevertheless, dramatic speed improvement can be obtained by allowing some precision error. For example, the system is about 10 times faster than the sequential scan when γ is set to 0.1, involving a precision higher thant 90% for the two databases.

Table 2. 20 KNN performances on 1M objects

γ	Precision (%)	Computation rate (%)	Time (s)
0	88.1	8.18	0.705
0.1	92.2	9.87	0.698
0.2	95.7	12.2	0.848
0.3	97.3	15.4	1.08
0.4	98.5	20	1.41
0.5	98.8	26.7	1.89
0.6	98.9	36.2	2.58
0.7	98.9	49	3.5
0.8	99.9	64.1	4.6
0.9	99.9	78.6	5.62
1	100	90.8	6.43

5 Conclusions

In this paper, we show that the function $\mathcal{T} : x \rightarrow \sqrt{log(x+1)}$ turns the Symetrized Kullback-Leibler Divergence into an exact metric when the statistical models compared are Gaussian. Furthermore, the proposed modification preserves the similarity ordering. Therefore, this new metric can be used by classical Metric Access Methods for fast similarity search. Further, we presented an algorithm called concavity controller able to modify the concavity of any mono-dimensional function. We showed how this algorithm can be applied to the proposed transformation function $\mathcal{T}$ to perform a precision controlled approximative k-NN search, with the possibility to set the desired precision at the query time. Results obtained by performing experiments using a M-tree structure show the relevance of our approach for fast music similarity search.

An interesting perspective of this work would be to test the proposed transformation function with the method proposed in [15].

6. Acknowlegments

The works presented in this paper were supported by the French project DISCO (2008-2010).

References

[1] J. Aucouturier and F. Pachet. Finding songs that sound the same. In *Proc. of IEEE Benelux Workshop on Model based Processing and Coding of Audio*, pages 1–8, 2002.

[2] M. Casey and M. Slaney. Song intersection by approximate nearest neighbor search. In *Proc. ISMIR*, pages 144–149, 2006.

[3] E. Chavez and G. Navarro. A probabilistic spell for the curse of dimensionality. In *ALENEX01, LNCS 2153*, pages 147–160, 2001. Springer.

[4] P. Ciaccia and M. Patella. Bulk loading the m-tree. In *Proceedings of the 9th Australasian Database Conference (ADC'98)*, pages 15–26, 1998.

[5] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. pages 426–435, 1997.

[6] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *SCG'04: Proceedings of the twentieth annual symposium on Computational geometry*, pages 253–262, 2004. ACM.

[7] D. Ellis and G. Poliner. Identifying cover songs with chroma features and dynamic programming beat tracking. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pages 1429–1432, 2007.

[8] S. Kullback. *Information theory and statistics*. Wiley Publication in Mathematical Statistics, 1959.

[9] B. Logan and A. Salomon. A music similarity function based on signal analysis. In *IEEE International Conference on Multimedia and Expo, ICME 2001*, pages 745–748, 2001.

[10] M. Mandel and D. Ellis. Song-level features and support vector machines for music classification. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR05), London, UK*, 2005.

[11] P. Mermelstein. Distance measures for speech recognition, psychological and instrumental. *Pattern Recognition and Artificial Intelligence*, pages 374–388, 1976.

[12] M. Ng and A. Yip. A fast map algorithm for high-resolution image reconstruction with multisensors. *Multidimensional Systems and Signal Processing*, 12:143–164, 2001.

[13] E. Pampalk. Audio-based music similarity and retrieval: Combining a spectral similarity model with information extracted from fluctuation patterns. In *Proceedings of the International Symposium on Music Information Retrieval, Victoria, BC, Canada*, 2006.

[14] T. Pohle and D. Schnitzer. Striving for an improved audio similarity measure. In *4th Annual Music Information Retrieval Evaluation Exchange*, 2007.

[15] D. Schnitzer, A. Flexer, G. Widmer, and A. Linz. A filter-and-refine indexing method for fast similarity search in millions of music tracks. In *10th International Society for Music Information Retrieval Conference (ISMIR), Kobe, Japan*, 2009.

[16] T. Skopal. On fast non-metric similarity search by metric access methods. *Lecture Notes in Computer Science*, 3896:718, 2006.

[17] T. Skopal. Unified framework for fast exact and approximate search in dissimilarity spaces. *ACM Transactions on Database Systems (TODS)*, 32(4):29, 2007.

[18] T. Skopal, J. Pokorny, and V. Snasel. Pm-tree: Pivoting metric tree for similarity search in multimedia databases. In *ADBIS (Local Proceedings)*, 2004.