International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:5, 2009

Bio-inspired Audio Content-Based Retrieval Framework (B-ACRF)

Noor A. Draman, Campbell Wilson, and Sea Ling

***Abstract*—**Content-based music retrieval generally involves analyzing, searching and retrieving music based on low or high level features of a song which normally used to represent artists, songs or music genre. Identifying them would normally involve feature extraction and classification tasks. Theoretically the greater features analyzed, the better the classification accuracy can be achieved but with longer execution time. Technique to select significant features is important as it will reduce dimensions of feature used in classification and contributes to the accuracy. Artificial Immune System (AIS) approach will be investigated and applied in the classification task. Bio-inspired audio content-based retrieval framework (B-ACRF) is proposed at the end of this paper where it embraces issues that need further consideration in music retrieval performances.

***Keywords*—**Bio-inspired audio content-based retrieval framework, features selection technique, low/high level features, artificial immune system

I. Introduction

THE increasing size of digital music collections whether for personal or commercial use has driven the need for a mechanism to automatically organize the music. One can always group their collections based on genre of music, artists or even based on gender. Music genre is a label used to differentiate types or categories of songs [1]. By using this kind of labels, one can easily organize his or her collections into groups of songs such as rock, jazz, rap or many others. This kind of information is useful and essential in music retrieval application as it helps to simplify the searching and retrieval parts of such application.

Identifying a genre by recognizing the songs played can be done easily by someone who works in the music industry like composers, song writers or instruments players. However it will be different with people who are not in this industry and not having such capability. Many studies on how to identify songs according to a certain class have started and researchers have found various techniques and methods to perform music recognition.

Noor A. Draman is with the Caulfield School of IT, Monash University, P O Box 197, Caulfield East, Victoria 3145, Australia (phone: +614 11784989; fax: +61 3 9903 1247; e-mail: noor.draman@infotech.monash.edu.au).

Campbell Wilson is with the Caulfield School of IT, Monash University, P O Box 197, Caulfield East, Victoria 3145, Australia (e-mail: campbell.wilson@infotech.monash.edu.au).

Sea Ling is also with the Caulfield School of IT, Monash University, P O Box 197, Caulfield East, Victoria 3145, Australia (e-mail: chris.ling@infotech.monash.edu.au).

We have seen a trend in this field as researchers mostly involved and focused on introducing new techniques and features for music classification whether it is based on genre, gender, instruments or others. This resulted many available tools readily and freely online for any other researchers to use. Basically, the studies involved investigating the content of the songs where it began with analyzing and extracting features from the sounds. The features then will be categorized to find series of similarity and these findings can be used to represent different kinds of song information.

Classification method is another major focus of study many researchers have been involved in. Features introduced combined with classification technique often provided certain level of performances. Throughout the years since the research in music analysis started, we have seen a tremendous development in music classification. Different techniques of extraction and classification were experimented and produced since then. Thus it is significant to conclude that features selected and classifiers applied have some sort of relevancy in classification performances.

With the whole range of features extracted and introduced, one cannot easily decide which features are more appropriate than the rest in recognizing certain type of music. To decide which one is significant to a particular type of song genre would need an approach to be applied. That is the reason why there is another emerging focus of study emerged which is developing and introducing techniques to select the right and significant features. Deciding and selecting the right features to represent different information of songs would be a crucial task alongside the extraction and classification tasks in music recognition.

These are the issues we are investigating throughout our research. We will discuss the highlighted issues and their consequence on music recognition performances. Some discussion on AIS classification approach will be presented here. The remainder of this paper is organized as follows: Section II gives some coverage and overview of background works in music analysis. Section III then will discuss music genre recognition where we divide it into three parts, feature extraction, features selection and genre classification. Section IV will discuss the proposed B-ACRF framework in this music analysis and retrieval research. Some concluding remarks will be outlined in the last section.

II. Audio Content-Based Retrieval

Content-based music retrieval is an area which involves steps and processes, algorithms and techniques and also experiments and evaluations. In theory, we can categorize all

these processes and steps into three main tasks which are identifying, grouping and searching.

Classifying music into different classes of genre would normally involve identifying and grouping tasks where these two tasks, once completed can contribute to the searching task performances. Identifying and grouping are important in music analysis as they determine what type of songs they are analyzing and which type of songs they belong to. Identifying songs in content-based analysis means analyzing music to capture features that can represent each type or genre of songs. Grouping on the other hand means clustering songs in a group or class of songs that have similarities in features. Meanwhile searching means looking for similar songs from the clustered songs based on query.

In music research area, there could be various topics of interest that researchers are involved in such as identifying vocal of artist, separating instruments from vocal sound, identifying an instrument in a music, genre classification and many others. Having said that, the most important aspect concerns would be capturing features from the music sound and as a result, the development of new features and techniques of extraction are becoming more active now than previously. Most of the works can be found in [4] - [15].

In our research, we are not only focusing on classification of music which consisted identifying and grouping processes, but we also want to focus on the searching part as well. As we have mentioned earlier, the performance of searching task are very much dependent on the identification and grouping of tasks, so we might as well includes those tasks and combined them in a framework. Further explanation about the framework will be in Section IV.

There are frameworks from previous research that have been implemented to do the content-based retrieval especially for image, audio even for music data. They can be found in [16] - [21].

Few frameworks that in our opinion are very closely related to ours will be discussed here. In [18], the testing part of different query length is the main concern where it can show the reliability of the proposed framework in retrieving songs based on different queries. Meanwhile in [19], the strength of the proposed framework is very much referring to the availability of comparing different methods of features extraction, matching and presentation. Having the presentation part would contribute to the chances of user giving any feedback to the application developed based on the proposed framework.

Reference [21] discussed an outstanding music application tools developed based on the proposed framework. It is however more suitable for a commercial kind of development project which is very contradictory to what we proposed. The Marsyas framework in [21] is yet another exceptional framework developed where the group has implemented a free tool or software for other researchers to use. The framework is very similar to our proposed framework only without the process of selecting significant features to represent genre of songs. All frameworks discussed are somewhat different with what we are proposing in B-ACRF.

III. Music Genre Recognition

We basically have divided music genre recognition into three phases which are feature extraction, feature selection and genre classification. This is slightly different as proposed and mentioned in [2]-[3]. While classification stage implies a series of techniques that can be used such as machine learning techniques, the first and second stage plays a crucial role is determining the performance of genre classification.

A. Content-based Features Extraction

Experiments have shown that sounds from music can be classified according to its characteristics to differentiate the genres. The characteristics of low level features can be obtained by analyzing the content of music. Several low level features have been introduced such as timbre, pitch, rhythm, wavelet basics, energy based features that include band energy ratio and spectral basis, rhythm pattern and repeating patterns. Among the listed features, timbre, pitch and rhythm would be the focus of our study.

Timbre related features generally represent the quality of a sound that is purposely used to differentiate sounds particularly human voice and instruments. Physical characteristics like spectrum can be used to analyze this particular element. Spectra analysis normally relates to the energy distribution of a sound where it can differentiate sound from different sources. For example sounds from flute and violin when played together may be identical in pitch, loudness and duration but they are easily recognized by their timbre content. The different sounds they produce are actually distinguished by different vibration pattern and different spectrum.

Rhythmic content feature initially can be use to gather information such as the rhythm regularity, beat and tempo of music. Both beat and tempo have certain characteristics that can be utilized to identify types of songs. Tempo relates to the sound being repeated over time which can be useful in fast music like dance techno, meanwhile beat is a measure of signal strength between frequencies. Rhythm has also been investigated in studies to distinguish vocal and instrument sounds.

Pitch analysis can describes melody and harmony of a music which can be used to identify types of music genre. Generally it can recognize pieces of music especially classical and jazz better than other types of music because the pitch content change frequently in these two type of songs compared to the others. On the other hand octave property can be used to describe singing voices that can differentiate one singer from another.

B. Features Selection

Selecting and using the right features to classify music according to a certain genre of music would be crucial as it will determine the performance of classification task. It can reduce data dimensions that need to be processed and also reduce the execution time. There are certain criteria that we need to look at as mentioned in [22]. The first criteria would be the similarity of perceived objects must be map to a nearby points in feature space and the second criteria would be the

features selected should contain all important information which is vital in identifying the type of songs.

There are various techniques of selecting features available. In general, those techniques can be divided into two categories which are filter and wrapper methods [23]. Wrapper method normally depends on classifiers where it utilizes the classifier to evaluate selected features and the result provided normally is classifier-specific. Wrapper is said to produce better results but is computationally more complex than filters. The disadvantages of using this method are that it is normally for large database of dataset only and need to be trained every time a new classifier is used.

On the other hand, filter is reasonably appropriate to be applied in our research as it is independent of classifiers since during selecting significant features, the process will not involve the classifier. It is also less computationally complex which means it will produce faster results than wrapper method. However combination of both methods has been made and can be seen in [24].

C. Genre Classification

In this section we will discuss Artificial Immune System (AIS) approach specifically on negative selection algorithm. According to [25] AIS is inspired by human immunology system in which the immune functions, models and principles of immunology are observed.

AIS is an adaptive system, emulating human body immunology system to solve problems. It is concerned with abstracting the whole concept of immune systems to computational systems in solving problems from mathematics, engineering and information technology point of view. Previous work on AIS in music genre classification can be found in [26] - [29].

Negative selection algorithm was first introduced by [30] where the idea was inspired by negative selection of T-cells in thymus. The algorithm focused on recognizing self or non-self cells where it will eliminate the T-cells that are not recognize by the thymus. Detailed explanations on how negative selection algorithm works can be found in [25].

The suggestion of using negative selection algorithm in this research was mainly driven by the concept of recognizing self or non-self cells where it is closely related to emphasize the individual features that represent each genre of songs. Originally, the very first version of NSA is about defining only two classes or two set of cells which are the self and non-self cells. Then it was modified into defining more than two set strings of classes by [29]. The modification was made with the intention to suit with their pattern matching problem. They implied the definition of self cells must be equal to the number of classes defined in their pattern problem. As we have more than two genres of songs to be classified, our research is basically very closely related to [29].

There are quite a number of other techniques normally used and applied in classification work. Among them, techniques that are repeatedly investigated in research are machine learning, statistically methods and neural network. Both features extraction and classification processes are complementing each other. While the extracted features contained information that represent different genre of songs, classifiers on the other hand determine the accuracy of the classification part of the songs. Some of the works that apply machine learning technique can be found in [8], [14], [16] and 17].

Statistical classifiers can also be considered as part of machine learning methods and they are mostly employed in pattern recognition and speech analysis application. Neural network on the other hand is an artificial mathematical model that works in an adaptive technique by changing the structure based on internal and external information. This happened through the network during a learning process. With the ability to model non-linear statistical data, they can be used to search similar patterns that exist in the data.

TABLE I
COMPARISONS OF CLASSIFICATION ACCURACY BASED ON CLASSIFIERS

Ref	Dataset	(Classifiers)			FST	%
		ML	ST	AIS		
[9]	Western		√			61.00
[10]	Western	√				79.00
[13]	Western	√				64.40
[25]	Malay			√	Yes	87.05
[26]	Western			√	Yes	93.00
[31]	Korean	√			Yes	83.00

The comparison was made between different classifiers applied, ML (machine learning), ST (statistical-based) and AIS (artificial immune system). Among them, two studies applied the FST (features selection technique) and they showed higher percentage of accuracy compared to other results. The % refers to accuracy percentage of the classification process.

The difference between [10], [13] and [9] is that both [10] and [13] include other features as well in their experiments which are called repeating pattern and wavelet basis features. To summarize, AIS approach combined with feature selection technique could increase the accuracy rate for almost 20%.

In music analysis, neural network method using abstraction of acoustical events that occur while the song is playing has been discussed in [5].

Table I shows a comparison of works in classification where mainly the works are chosen mostly based on features similarities. We want to highlight the extracted features used in those study can be used to represent not only the western songs but also other type of songs as well [25][30]. In addition we also want to compare different classifiers used in the works ranging from machine learning, statistical based and artificial immune system approach. The datasets used were from western group of songs, Korean traditional songs and Malay traditional songs.

IV. B-ACRF: THE PROPOSED FRAMEWORK

Content-based analysis and retrieval of multimedia data is basically related to information demand. An exact match between demand and supply in multimedia databases is unavailable since the information is differently structured.

This requires, besides an accurate and comprehensive model of the available information, specific functionality to handle the queries and the presentation of the results. In content-based retrieval, information should be extracted from features, concepts or combination of both. Retrieving any particular multimedia data generally would need a very comprehensive application in order to get good results. We will highlight a few important issues that need to be considered in developing such applications.

Related works in music analysis have shown that different types of features and techniques used to extract the features provided different levels of performances. Nonetheless those studies emphasized more on classification. To say which feature is better than the other actually depends on the type of data analysed. According to [9] [10], certain genre of music tends to have different levels of pitch. For example jazz and classical music are said to have a higher pitch than pop or rock songs. Using others apart from pitch related features to identify jazz or classical songs are not highly recommended for such genres. Also, rather than using individual features to identify the songs, combinations of features in identifying certain types of songs can also increase performances [10] [13].

Quite a number of features representing different types of song genre have been acknowledged now and researchers can always select which features are relevant to their works. Choosing the right and appropriate features to be analysed and used sometimes will takes time but solution to this problem has been rectified since there are techniques that can be used to select significant features. The techniques are known as the feature selection techniques and we have come across few works that applied these techniques in their research. In Table I, we have indicated whether these works have applied the technique (FST) and a comparison of their performance is made.

Another important aspect to consider is choosing the appropriate classifier in classification process. Table I lists the most common classifiers used in related music analysis works. AIS however can be considered as a reasonably new approach explored in this field although it has been widely used in other areas like pattern recognition and image analysis studies. Apart from being quite new, it has been shown that using AIS helps to get better performance than using any other types of classifiers. This can be seen from Table I as the accuracy rate of using AIS classifier increased almost 20% compared to works.

Having discussed all those important issues, we are proposing a framework that includes all the factors previously mentioned. Fig. 1 shows the proposed framework of music retrieval by content which consisting of three phases, 1) feature extraction, 2) classification and 3) retrieval. The proposed phases are generally the basic steps of information retrieval. However, to our knowledge, there is still no framework that emphasized on the previously discussed issues that can contribute to a better performance in retrieval. Our purpose with B-ACRF is to do that.

In the extraction phase, we are focusing on timbre, pitch and rhythm related features to represent different genre of songs from different regions or countries. To extract these related features we use Marsyas version 0.2 which is developed in [7] and we add feature selection task here. Combination of both elements has been applied in the first phase with the intention to find a set of significant features from the whole range of features extracted earlier. The selected features will improve the classification performance as they will reduce execution time while maintaining classification accuracy [23].

The selected features then will be converted into binary set of strings before we can proceed with the classification work in the second phase. In the classification phase we applied Negative Selection Algorithm (NSA) from AIS as classifier to identify and group songs into different types of genre. The reason for using NSA in classification has been explained in Section III.

There are two modules involved which are called censoring and monitoring in the process of recognizing and generating detectors. From the figure, defining self cells and generating detectors are part of censoring module where we generate detectors based on how many song genres we want to classify. During generating detectors, antigen will be compared to evaluate the affinity binding (similarity value) with the detector candidates. We apply exclusive-or operator (XOR) operation to determine and evaluate the affinity binding between them where the threshold value will be used as benchmark [25].

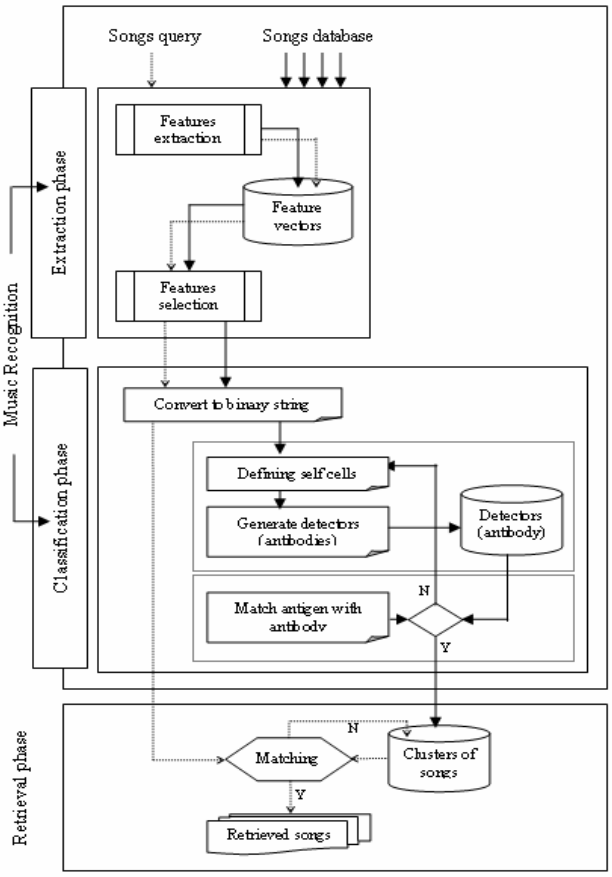

Fig. 1 Proposed Audio Content-Based Retrieval Framework

In XOR operation, values "0" and "1" are used and will be counted to decide whether it exceeds the threshold value or not. As NSA considers non-self cells as detectors, the non-matching antigen-detector will be based on the "1" value. The higher the "1" than "0" value during comparison, the better results of indicating the non-self cells will be. If indicated, the cell then will be considered as a detector and stored for further analysis in the next module.

In monitoring module, the process of matching between detectors and antigen will take place to identify types of antigen (song genres). Again similar process as in censoring module will be repeated. Only in this module, the matching process will be looking for affinity measurement between antigen and known non-self cells (detectors) to determine types of song genres. Value "0" will be used to indicate the similarity during monitoring process. The more "0"s are found during affinity measurement, the more similar the antigen will be to the detector. The identified song then will be classed in the same group of song genre for retrieval purposes. Note that during classification, we use the selected features individually or combined with the purpose to look for the most significant representation that can identify each type of song genres accurately.

The last phase of our proposed framework will be focusing on retrieval part inclusively where we apply similarity technique to search for similar type of song genre from the storage with the query.

V. Conclusion

The availability of techniques and methods for feature extraction and classification in music analysis field today has shown that researchers in this area are very concerned in automating the extraction and classification processes. As the collections of digital songs keep increasing online, their studies contribute to a major breakthrough for internet users and others.

In this paper, we have explained the framework with important factors that need to be considered to get higher performance on retrieval part. We strongly believe that our discussion throughout this paper has given opportunities to other researchers in this area of studies to fill the gaps, to explore further and to provide solutions to the known and un-known problems that have yet to be discovered. Furthermore the discussion on proposed framework using the new approach in genre classification especially, hopefully will contribute to others and also, is going to be investigated further.

References

[1] R. Neumayer, "Musical genre classification using a multi layer perceptron", In *Proceedings of the 5th Workshop on Data Analysis (WDA'04), Tatranska Polianka, Slovak Republic, 2004*, pp 51-66.

[2] O. Kotov, A. Paradzinets and E. Bovbels, "Musical genre classification using modified wavelet-like features and support vector machines", In *Proceedings of the IASTED European Conference on Internet and Multimedia Systems and Application, Cambridge, United Kingdom,* 2007.

[3] U.Bagsci, "Automatic classification of musical genre using inter-genre similarity", *Journals of IEEE Signals Processing Letters,* August 2007vol. 14, no. 8, pp. 521-524.

[4] T. Lambrou, P. Kudumakis, R. Speller, M.Sandler and A Linney, "Classification of audio signals using statistical features on time and wavelet transform domain", In *Proc. Int. Conference on Acoustic, Speech and Signal Processing (ICASSP – 98), 1998,* vol. 6, pp 3621-3624.

[5] H. Soltau, T. Schultz and M. Westphal, "Recognition of music types", In *Proceedings of the 1998 IEEE International Conference on Acoustic, Speech and Signal Processing,* 1998, Denver, pp 1137-1140.

[6] A. Rauber and M. Fruhwirth, "Automatically analyzing and organizing music archives", In *Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries (ECDL),* Darmstadt, Germany September 2001, pp. 402-414.

[7] T. Lidy and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification", In *Proceeding of the 6th International Conference on Music Information Retrieval (ISMIR '05), 2005,* pp. 34-41.

[8] T. Li, M. Ogihara and Q. Li, "A comparative study on content-based music genre classification", *Proceedings of the 26th annual international ACM SIGIR*, 2003, Toronto, Canada, pp. 282-289.

[9] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals", *IEEE Transactions on Speech and Audio Processing*, vol.10, no.5, 2002, pp. 293-302.

[10] T. Li and M.Ogihara, "Toward intelligent music information retrieval", In *Proceedings of IEEE Transactions on Multimedia*, June 2006, vol. 8, No 3, pp. 564-574.

[11] M.McKinney and J. Breebaart, "Features for audio and music classification", In *Proceeding ISMIR*, 2003, pp. 151-158.

[12] C.R. Lin, N.H. Liu, Y.H. Wu and A.L.P. Chen, "Music classification using significant repeating patterns", In *Proceedings Database Systems for Advanced Applications*, 2004, pp. 506-518.

[13] I. Karydis, A. Nanopoulos and Y. Manolopoulos, "Symbolic musical genre classification based on repeating patterns", in *Proceedings of the ACM Multimedia Workshop on Audio and Music for Multimedia (AMCMM), Santa Barbara, California, USA,* 2006, pp. 53-58.

[14] F. Moerchen, I. Mierswa and A. Ultsch, "Understandable models of music collections based on exhaustive feature generation with temporal statistics", In *Proceedings of International Conference on Knowledge Discovery and Data,* Philadelphia, USA, 2006, pp. 882-891.

[15] R. Neumayer and A Rauber, "Integration of text and audio features for genre classification in music information retrieval", In *Proceeding of 29th European Conference on Information Retrieval, Rome, Italy,* 2007, pp. 724-727.

[16] F.M. El-Hadidy, H. J. G. de Poot and D. D. Velthausz, "Multimedia information retrieval framework: From theory to practice," in *Proc. 8th Working Conference on Database Semantics- Semantic Issues in Multimedia Systems*, Deventer, Netherlands, 1999, pp. 271-299.

[17] M. Gabbouj, **S.** Kiranyaz, K. Caglar, B. Cramariuc, F. Alaya Cheikh**,** 0. Guldogan, and E. Karaoglq, "MUVIS: A multimedia browsing, indexing and retrieval system," in *Proc of the WDC 2002 Conference on Advanced Methodr for Multimedia Signal Processing,* Capri, Italy 2003, pp. 1-8.

[18] K.-S. Park, W.-J Yoon, K.-K. Lee, S.-H. Oh and K.-M. Kim, "MRTB framework: a robust content-based music retrieval and browsing," in *IEEE Transactions on Consumer Electronics,* 2005, vol. 51, issue 1, pp. 117-122.

[19] M. Bosma, R. C. Veltkamp and F. Wiering, "MUGGLE: A music retrieval experimentation framework", in *Proceedings of 9th International Conference on Music Perception and Cognition*, Italy, 2006, pp. 1297-1303.

[20] X. Amatriain, M. de Boer, E. Robledo and D. Garcia , "CLAM: an OO framework for developing audio and music applications", In *Companion of the 17th annual ACM SIGPLAN Conference on Object-oriented programming, systems, languages and applications (OOPSLA '02)*, Washington, USA, 2002, pp. 22-23.

[21] G. Tzanetakis and p. Cook, "Marsyas: A framework for audio analysis", in *Organised Sound Journal*, vol. 4 issue 3, 1999, pp. 169-175.

[22] K. Kosina, "Music Genre Recognition," M.S. thesis, Technical College Hagenberg, Austria, 2002.

[23] J. Liang, S. Yang and A. Winstanley, "Invariant optimal feature selection: A distance discriminant and features ranking based solution", *Pattern Recognition Society*, 2007, pp. 1429-1439.

[24] R. Kumar, V.K. Jayaraman and B. D. Kulkarni, "An SVM classifier incorporating simultaneous noise reduction and feature selection:

Ilustrative case examples*", Pattern Recognition*, vol. 38, issue 1, 2005, pp. 41-49.

[25] L.N. de Casto and J. Timmis, "*Artificial immune system: A new computational intelligence approach",* Great Britain, Springer, 2002, pp. 76-79.

[26] S. Doraisamy, S. Golzari, N. M. Norowi, M. N. Sulaiman and N. I. Udzir, "A study on feature selection and classification techniques for automatic genre classification of traditional Malay music", in *Proceedings of Ninth International Conference on Music Information Retrieval (ISMIR'08)*, Philadelphia, Pennsylvania USA, 2008, pp. 331-336.

[27] D.N. Sotiropoulus, A.S. Lampropoulus and G.A. Tsihrintzis, "Artificial immune system-based music genre classification", in *New Directions in Intelligent Interactive Multimedia*, 2008, pp. 191-200.

[28] M. Caetamo, J. Manzolli and F. J. Von Zuben, " Application of an artificial immune system in a compositional timbre design technique", in *Proceedings of International Conference on Artificial Immune Systems*, Baff, Alberta, Canada, 2005, pp. 389-403.

[29] R.-B. Xiao, L. Wang and Y. Liu, "A framework of AIS based pattern classification and matching for engineering creative design", in *Proceedings of the First International Conference on Machine Learning and Cybernetics*, Beijing, China, 2002, pp. 1554-1558.

[30] S. Forrest, A.S. Perelson, L. Allen and R. Cherukuri, "Self-nonself discrimination in a computer", in *Proceedings of 1994 IEEE Computer Society Symposium on Research in Security and Privacy*, Oakland, CA, USA,1994, pp. 202-212.

[31] K.-K. Lee and K.-S. Park, "Robust feature extraction for automatic classification of Korean traditional music in digital library", in *Proceedings of 8th International Asian Digital Library*, Bangkok, Thailand, 2005, pp.167-170.