SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
VOL. 2, NO. 1, APRIL 2019

A New Genre Classification with the Colors of Music

Zehra DURDAG[1], Pakize ERDOGMUS[2]
[1]Duzce University, Engineering Faculty, Computer Engineering, Duzce, 81620 TURKEY
zdurdag14032009@hotmail.com,+90 546 482 3316
[2]Duzce University, Engineering Faculty, Computer Engineering, Duzce, 81620 TURKEY
pakizeerdogmus@duzce.edu.tr

Received: 26/03/2019; Revision; 28/03/2019 Accepted; 05/04/2019 Published online; 25/04/2019

Abstract

The aim of this study is to bring a new perspective for the classification of the songs, revealing of the colors of music. The first effort is to transform songs into images. The colorful images have been attained with Short time Fourier transform, discrete cosine transform and time to spatial transformation and some extra processing. It has been observed that the images of different music genres obtained with the same method have different colors. But some of them have similar colors and patterns, which making difficult to classify. Pre-trained deep convolutional network have been trained with these images. For five Turkish musical genres, nearly up to 60% classification accuracy has been achieved and for ten musical genre of a benchmark musical dataset nearly up to 54% classification accuracy has been achieved. In future studies, it has been planned to create the images using timbral texture and rhythmic contents, for increasing the accuracy.

Keywords: Convolutional neural networks, music genre classification, discrete cosine transfo

Müzik Türlerinin Derin Öğrenme Ağları ile Sınıflandırılması

Öz

Bu çalışmada günümüzde görüntülerin sınıflandırılmasında yüksek başarı sağlayan derin öğrenme ağları ile müzik türlerinin sınıflandırılması hedeflenmiştir. Bu çalışmanın amacı, müziklerin renklerini ortaya çıkararak şarkıların sınıflandırılması için yeni bir bakış açısı getirmektir. Bu amaçla ilk olarak, müzik türlerinden seçilen parçalar görüntülere dönüştürülmüştür. Renkli görüntüler kısa zaman fourier dönüşümü, ayrık kosinüs dönüşümü ve uzamsal dönüşüm yöntemleri ve bazı önişlemlerle elde edilmiştir. Farklı türlere ait görüntülerin renklerinin farklı olduğu görülmüştür. Ancak bazı türlerde sınıflandırmayı zorlaştıracak benzer renkler ve desenler görülmüştür. Önceden eğitilmiş derin konvolüsyon ağı bu görüntülerle eğitilmiştir. Türkçe müziklerden seçilen, arabesk, pop, türk halk müziği, türk sanat müziği ve rock müzikleri ile eğitilen ağda, yaklaşık % 60'a kadar bir sınıflandırma doğruluğu elde edilmiş ve yine literatürde müzik türü sınıflandırılmasında kullanılan genel bir veri tabanı ile yapılan testlerde, on farklı müzik türü için yaklaşık% 54'e kadar bir sınıflandırma doğruluğu elde edilmiştir.

Anahtar Kelimeler: Müzik Türlerinin Sınıflandırılması,Derin Öğrenme,Ayrık Kosinus Dönüşümü, Hızlı Fourier Dönüşümü,Kısa Zamanlı Fourier Dönüşümü

1. Introduction

With the development of computer based systems, the number of files has increased since a lot of documents, music files, image files were created easily.
So many documents have also been creating messy folders. One way to alleviate the problem is to manage the files automatically. After automatic classification of files, files are distributed to the folders storing related files. Automatic classification of files requires intelligent systems, which

understand the content of the files. Genre classification of audio data [1], paintings [2], web pages [3] and music [4] are some of the studies.

Music genre is a label defined by human categorizing the music. With these labels, music or songs are classified either manually or automatically. But it is a fact that there is not strict definition among genres [5, 6].

Since 1990's, with the increasing abilities of microprocessors for multi-media, genre classification of music has been started to have been studied. In some of the first studies, a genre classification system for TV sound signals, genre identification for videos and TV program genre classifications [7,8,9] have been realized using spectrogram analysis, audio pattern extraction and distinctive cinematic aspects respectively.

The next studies have been focused on musical genre classification, music information retrievals, music annotation and indexing [10].

Today, there are a lot of applications presented in the virtual markets. These applications offer individual playlists making genre classification. Several methods have been used for genre classification. As most of the studies based on genre classification, there are two main efforts. Firstly features, abstracting the real large data with few parameters have been selected. Secondly a good classifier for features has been selected discriminating the genres. Three sets of features for presenting timbral texture, rhythmic content and pitch content are proposed [5]. Statistical pattern recognition methods such as Gaussian classifier and K-nearest neighbor have been selected for classification in this study. In another study, the pitch contour of the song refrain, the pitch contour of all the song notes and the duration contour of all the song notes have been selected [11] and neural networks have been used for classification. In [12] Beat spectrum, LPC (linear predictive coding), zero crossing rates, spectrum power and me1 frequency cepstral coeficients have been used as features and musical genres have been classified with support vector machines. In [13], linear discriminant analysis (LDA) and support vector machines used for classification of genres using the same features in [5].Music genre classification using data mining algorithm [14] and Gaussian mixture model [15] have also been used. In recent years, the new classifiers have also been used for musical genre classification. Deep convolutional neural networks (CNN) have been used for the classification of musical genres [20, 21].

2. Genre Classification Process

In this study, it has been aimed to present the features of music as visually. In the literature there are quite a lot of studies trying to find which features presents the musical genre best. But there is no study looking these features as an image. There is only one latest study using spectrogram images for speech music discrimination with deep convolutional networks [25]. So according to our research, this study is the first effort using STFT, DCT and time to spatial transformation as image for genre classification. With this aim, GTZAN dataset [5] has been used [24] for musical genre classifications. And a new dataset containing five Turkish musical genres has been prepared for this study. The whole part of the music data have been used for conversion.

2.1 Transforming Music In to Images

In this process, four different methods have been used for transformation. In the first method, Short Time Fourier Transform coefficients has been transformed into proper mxmx3 matrix, normalized the values, resized to 227X227X3 images and saved.

In the second method and third methods discrete cosine transform (DCT) have been used. In the second method, a fixed number of DCT (1024X1024X3) coefficients have been taken and transformed to the matrix, normalized and resized to 227X227X3. In the third method a floating number of DCT coefficients have been taken, which summarizes the data best. Lastly only musical data have been transformed to the images.

Arabesque, Turkish Folk Music, Pop Music, Rock Music and Turkish classic music songs have been selected. Transformation process for DCT has been given in Figure 1.

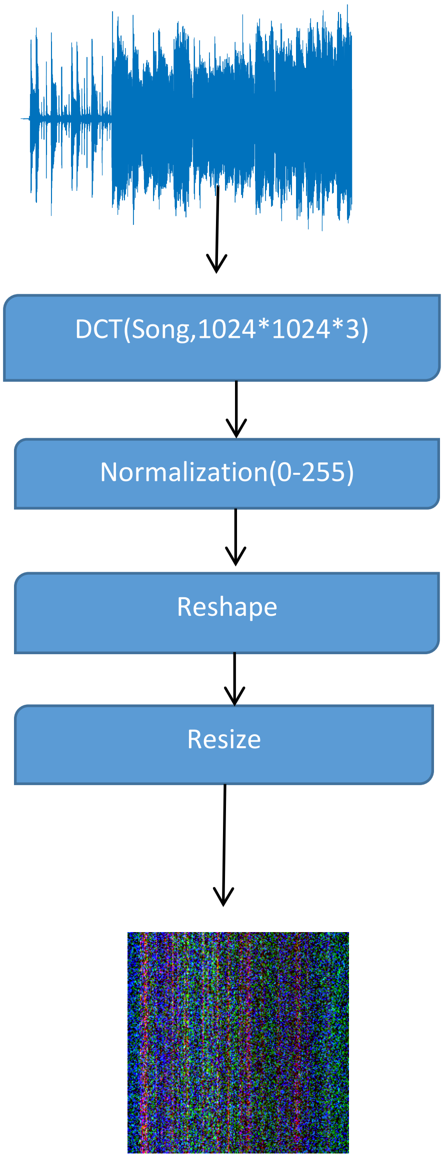


Figure 1. Music to Image transformation with DCT

In each process, different main colors have been attained for each genre. Table 1 shows sample images (belong to the same song) attained with four methods.

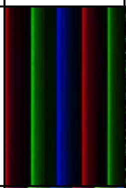
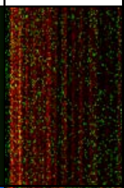
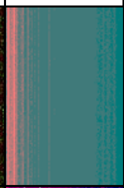
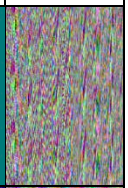
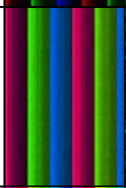
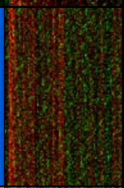
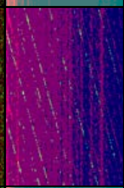
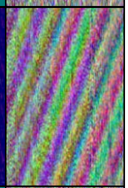
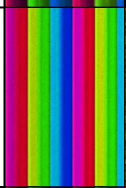
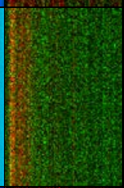
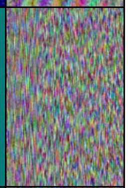
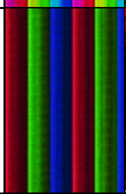
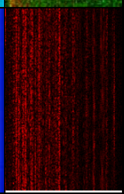
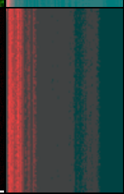
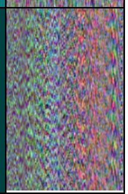
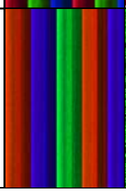
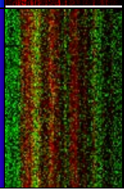
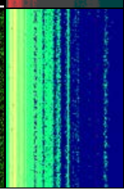
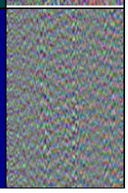
2.2 Training And Testing The Classification Accuracy Of Images

In the first test in our study, 150 music images for each genre have been used for training and testing. The 70% and 80% of examples have been used for training and 30 % and 20% of examples have been used for testing respectively.

For the classification of the musical genres, using these images, pre-trained deep convolutional neural network(AlexNET) has been trained and tested using Matlab© 2018 Deep Learning Toolbox. AlexNET has a CNN structure designed by Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever[26]. It has been used 25 layered default network structure in the study. 227X227X3 inputs and 5 outputs for Turkish Musical Classification and 10 outputs for GTZAN have been used. The net contains convolutions, max pooling, ReLU and fully-connected layer.The classification system has been given in Figure2

The results have been submitted in Table 2. Since the execution time of these methods is not comparable with the other studies, it has been omitted for this study.

Table 1. The colors of Turkish Music for Each Genre

Musical Genre	Applied Method			
	STFT	DCT Fixed	DCT Float	Time to Spatial
Arabesque				
Turkish Pop				
Turkish Rock				
Turkish Classical				
Turkish Folk				

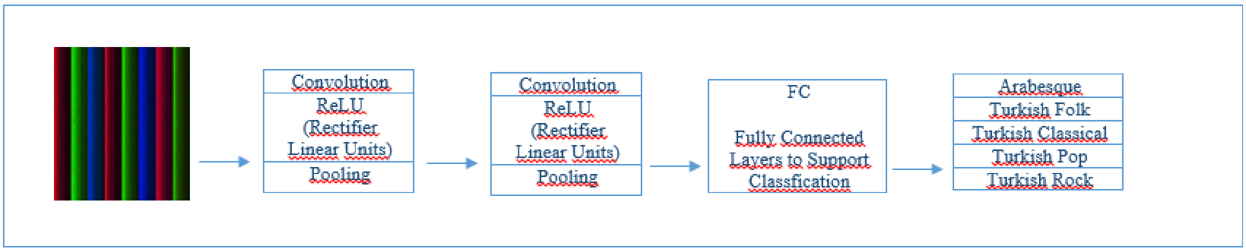


Figure 2. Genre Classification System

Table 2. Classification accuracy of Turkish music dataset

METHOD	ACCURACY ACCORDING TO TRAINING DATA PERCENTAGE	
	70%	**80%**
TIME TO SPATIAL	0.4209	0.4467
FIXED DCT	0.6061	0.5736
FLOATING DCT	0.4747	0.4619
STFT	0.5503	0.5404

In order to validate the classification success of the transformed images, GTZAN dataset has also been used. In this dataset, ten genres and 1000 music data have been used for classification. 100 music images for each genre have been used for training and testing. The images belong to these genres have been given in Table 3. The 70% and 80% of examples have been used for training and 30 % and 20% of examples have been used for testing respectively. The classification results have been submitted in Table 4.

Table 3. The colors of GTZAN database for each genre

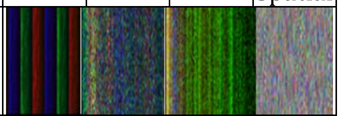
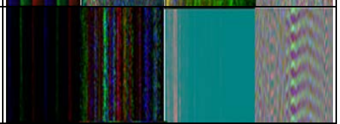
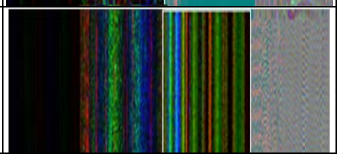
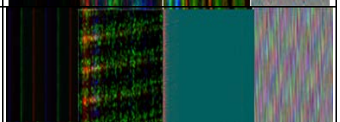
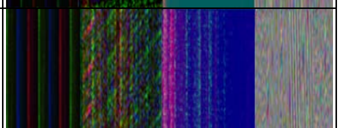
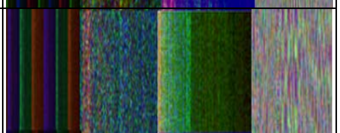
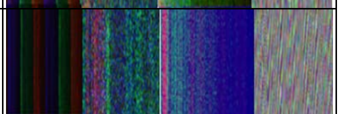
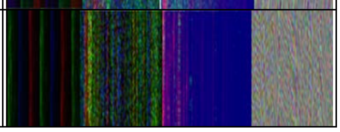
Musical Genre	Applied Method			
	STFT	DCT Fixed	DCT Float	Time to Spatial
Blues(77)				
Country(30)				
Classical(89)				
Hiphop(87)				
Pop(65)				
Metal(52)				
Rock(40)				
Jazz(54)				

Table 3. The colors of GTZAN database for each genre (Continue)

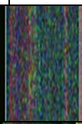
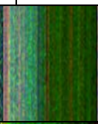
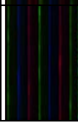
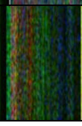
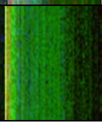
Musical Genre	Applied Method			
	STFT	DCT Fixed	DCT Float	Time to Spatial
Disco(67)				
Reggae(77)				

Table 4. Classification accuracy of GTZAN dataset

METHOD	ACCURACY ACCORDING TO TRAINING DATA PERCENTAGE	
	70%	**80%**
TIME TO SPATIAL	0.3600	0.3650
FIXED DCT	0.3733	0.3950
FLOATING DCT	0.3733	0.3700
STFT	0.5433	0.5350

3. Conclusion

In this study, the first and foremost aim is to convert musical data to images. The idea behind these efforts is that “The harmony in the music gives us a visual harmony.” Another reason for us to convert musical data to the images is that we wanted to use the power of the AlexNet pre-trained network which was trained more than one million images. According to the 1750 images attained with the transformations, results verify this idea. Another idea is that “The images of music data can also discriminate the different genres”. The results are promising. The figures can be used for online musical applications. In future studies, we are planning to design a musical genre classification system.

References

[1] M. Kobayakawa, M. Hoshi and K. Yuzawa, "Music Genre Classification of MPEG AAC Audio Data," *2014 IEEE International Symposium on Multimedia*, Taichung, 2014, pp. 347-352. doi: 10.1109/ISM.2014.25

[2] E. Cetinic and S. Grgic, "Genre classification of paintings," 2016 International Symposium ELMAR, Zadar, 2016, pp. 201-204. doi: 10.1109/ELMAR.2016.7731786G.

[3] C. Jebari and M. A. Wani, "A Multi-label and Adaptive Genre Classification of Web Pages," *2012 11th International Conference on Machine Learning and Applications*, Boca Raton, FL, 2012, pp. 578-581. doi: 10.1109/ICMLA.2012.106

[4] Karatana and O. Yildiz, "Music genre classification with machine learning techniques," *2017 25th Signal Processing and Communications Applications Conference (SIU)*, Antalya, 2017, pp. 1-4.doi: 10.1109/SIU.2017.7960694

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

[6] Pachet, Francois & Cazaly, D. “A classification of musical genre” in Proc RIAO Content-Based Multimedia Information Access Conf, Paris, France, March, 2000.

[7] Kyu-Phil Han, Young-Sik Park, Seong-Gyu Jeon, Gwang-Choon Lee and Yeong-Ho Ha, "Genre classification system of TV sound signals based on a spectrogram analysis," in IEEE Transactions on Consumer Electronics, vol. 44, no. 1, pp. 33-42, Feb. 1998.

[8] Ba Tu Truong and C. Dorai, "Automatic genre identification for content-based video categorization," Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, Barcelona, Spain, 2000, pp. 230-233 vol.4.

[9] R. S. Jasinschi and J. Louie, "Automatic TV program genre classification based on audio patterns," Proceedings 27th EUROMICRO Conference. 2001: A Net Odyssey, Warsaw, Poland, 2001, pp. 370-375.

[10] K. Markov and T. Matsui, "Music Genre and Emotion Recognition Using Gaussian Processes," in IEEE Access, vol. 2, pp. 688-697, 2014.

[11] B. Colaiocco and F. Piazza, "A music retrieval system based on the extraction of non trivial recurrent themes and neural classification," Proceedings of the International Joint Conference on Neural Networks, 2003., Portland, OR, 2003, pp. 1110-1115 vol.2.

[12] Changsheng Xu, N. C. Maddage, Xi Shao, Fang Cao and Qi Tian, "Musical genre classification using support vector machines," 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)., Hong Kong, 2003, pp. V-429.

[13] Tao Li and G. Tzanetakis, "Factors in automatic musical genre classification of audio signals," *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, New Paltz, NY, USA, 2003, pp. 143-146.

[14] M. M. Panchwagh and V. D. Katkar, "Music genre classification using data mining algorithm," *2016 Conference on Advances in Signal Processing (CASP)*, Pune, 2016, pp. 49-53. doi: 10.1109/CASP.2016.7746136

[15] C. Kaur and R. Kumar, "Study and analysis of feature based automatic music genre classification using Gaussian mixture model," 2017 *International Conference on Inventive Computing and Informatics (ICICI)*, Coimbatore, 2017, pp. 465-468.

[16] S. Lippens, J. P. Martens and T. De Mulder, "A comparison of human and automatic musical genre classification," *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Que., 2004, pp. iv-iv. doi: 10.1109/ICASSP.2004.1326806

[17] D. Turnbull and C. Elkan, "Fast recognition of musical genres using RBF networks," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 580-584, April 2005. doi: 10.1109/TKDE.2005.62

[18] F. Pachet and P. Roy, "Exploring Billions of Audio Features," *2007 International Workshop on Content-Based Multimedia Indexing*, Bordeaux, 2007, pp. 227-235. doi: 10.1109/CBMI.2007.385416

[19] S. Sharma, P. Fulzele and I. Sreedevi, "Novel hybrid model for music genre classification based on support vector machine," *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Penang, 2018, pp. 395-400.

[20] A. R. Rajanna, K. Aryafar, A. Shokoufandeh and R. Ptucha, "Deep Neural Networks: A Case Study for Music Genre Classification," *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, 2015, pp. 655-660

[21] S. Vishnupriya and K. Meenakshi, "Automatic Music Genre Classification using Convolution Neural Network," *2018 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, 2018, pp. 1-4.

[22] A. Elbir, H. O. İlhan, G. Serbes and N. Aydın, "Short Time Fourier Transform based music genre classification," *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, Istanbul, 2018, pp. 1-4.

[23] R. Sarkar, N. Biswas and S. Chakraborty, "Music Genre Classification Using Frequency Domain Features," *2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT)*, Kolkata, 2018, pp. 1-4.

[24] http:// http://marsyas.info/downloads/datasets.html/ Accessed on: November. 28, 2018

[25] Michalis Papakostas, Theodoros Giannakopoulos, Speech-music discrimination using deep visual feature extractors, Expert Systems with Applications, Volume 114, 2018, Pages 334-344, ISSN 0957-4174.

[26] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.