Distinguish Musical Symbol Printed using the Linear Discriminant Analysis LDA and Similarity Scale

Ansam Nizar Younis
M.Sc Student
Mosul University/Iraq
Department of Computer Science

Fawzia Mahmoud Remo, PhD
Assistant Professor
Mosul University/Iraq
Department of Computer Science

ABSTRACT
Music exists in all areas of our daily lives , A moment does not pass without hearing a musical tone that expresses a specific event or any other sound like the sounds of animals, the sounds expressed by sadness or joy emanating from the sound systems in the human vocal cords and throat different rhythms, and others. The language of music written by signs, symbols and lines is one of the most important methods to save tunes and musical tones so that we can read the melody and retrieve it again when needed. A new method has been proposed in this research to highlight printed music labels, Where a computer system was built to read the images of various musical labels and then perform a series of sequential processes as a preliminary processing of the image. Then use Linear Discriminant Analysis (LDA) algorithm for the purpose of extracting the important characteristics of the process of discrimination from the images of different types of different marks and as a result of reducing the size of data entered, thus providing the time and capacity of the treasury during the treatment and discrimination symbols. The structure similarity index SSIM is then used which allows measuring the similarity between the input image and training images. The quality of the input signal is evaluated for the second signal, which can be considered to be of optimal quality, This metric has been used to identify different musical labels. The linear discrimination analysis algorithm with the structural similarity algorithm achieved very good performance and low executive time. A classification accuracy of 89.5% was obtained, and the search for any marker took about 0.784990 seconds.

General Terms
Linear Discriminant Analysis (LDA), Structure similarity index (SSIM) ,preprocessing.

Keywords
staff lines ,clef, bar lines.

1. INTRODUCTION
Music is an expressive and communicative language that is produced by the culture and history of the age in which it originated. Music can be defined as an art consisting of a series of consecutive sounds and silence across certain time periods. It can produce music from one sound or from one instrument, or from many different instruments and sounds. The music consists of several elements and the most important terms are rhythm, melody, musical score, harmony, form, instrumentation, musical expression, effectiveness, color of music, and finally musical weaving [1]. Music can be used in daily life and in many areas of the home, such as television and computers, in mobile phones, in work, in transportation, and also in the treatment of psychiatric patients, treatment of patients with autism, and promote wellness in patients in general [2]. Music notation is one of the most important methods used since ancient times to preserve musical melodies in a way that can be reproduced and retrieved later to be transmitted among generations. Musical notes and melody symbols are usually placed above the music staff or in empty spaces between the runway lines [3].

2. COMPUTER AND MUSIC
There is no doubt about the importance of music and the extent of the great impact in the psyche of humans and animals and sometimes on the growth of plants alike, So it is necessary to exploit the modern and sophisticated devices to serve music and these devices are represented by the computer, where specialists seek to expand and achieve the processes of discrimination and classification of the general formula of music that facilitate the work before the interested in the field of music. [4]

3. MUSICAL NOTES AND STAFF LINES
Numerous musical symbols and labels have been used, which have evolved over time. Symbols and musical notes are placed on straight lines called the staff, In Common Notation, the music is recorded in letters other than the alphabet where the musical tags are in a time-specific symbol format. The location of each letter or musical symbol on the ladder affects the tone of that letter. There are five horizontal lines parallel and equal in length and dimensions, including four spaces called the staff (or the musical ladder). The lines of the musical ladder begin from the bottom to the top, and the line on the musical scale takes its musical mark according to the music key located at the first left of the ladder (clef). Where the musical scale is divided into equal time scales by vertical lines on the bar lines, Most of the musical marks are placed on a runway or in the gap between the lines, and extra lines can be added for very high or very low marks. The shape of a tag represents its time period while the location of the mark on the music track represents its intensity or degree. Standard musical labels consist of only seven signs, while the eighth sound is repeated for the first sound: du-ri-mi-fa-sol-la-si-do. The name of the music tag is determined by the key used at the beginning of the musical scale [3] [5]. Note Figure (1).

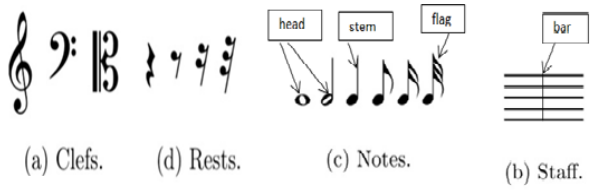

Fig 1:Some of musical notes and staff lines.

4. CHARACTERISTICS EXTRACTION AND CLASSIFICATION PROCESS

Characteristics extraction means to reduce the size of data entered by converting large and multi-dimensional vectors into less-dimensional vectors, so that the data size is less so that the representation of information effectively and leaving information that does not benefit us in discrimination. The process of extraction of features is a very important step in the methods used for the classification process and artificial intelligence algorithms, where the process of extraction of properties from the classification process is easy process where the process of extraction of qualities after the pre-treatment in recognition systems. And examples of the process of classification: fingerprint identification, face recognition, classification of images of blood cells and others. [6]. There are several types of attributes that can be used in the detection and classification of them:

- Chromatic qualities.
- Textural Characteristics.
- Geometrical or Shape-based Characteristics.
- Topological Characteristics.[7]

There are several types of analysis methods, such as the Principal Component Analysis (PCA), the Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA). [8]

5. LINEAR DISCRIMINANT ANALYSIS (LDA)

Is a statistical technique that is usually used to reduce the dimensions of data and then classify it. Also called the method of linear analysis of Fisher relative to the British statistical scientist (Ronald Aylmer Fisher) who introduced this method. This technique has been widely used in pattern recognition and image retrieval, in which data are usually large dimensions, so linear discrimination analysis is used to reduce these dimensions, reducing the time taken to distinguish. The advantage of linear discrimination analysis in reducing variability between single-label data and increasing variability between different species contributed to the classification process better than other methods [9], as shown in Figure (2).Where the method is isolated between the classes without changing the location of the original data set.

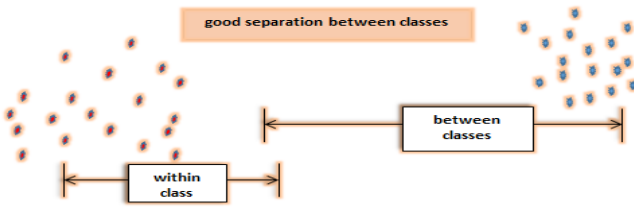

Fig 2: The good separation between two classes

LDA Is a new technique that helps us to understand and distribute data set features in a smooth way. Where he tries to model the differences between categories, so the analysis of linear discrimination has the most powerful technique in the recognition of patterns through the application of the characteristics of linear discrimination. The method of collecting images from one class is used together, separating the images from different categories. To determine the image of the input test, the test image is usually compared with each of the training images available, and the test image is known as the closest image of the training found. [9]

LINEAR DISCRIMINATION ANALYSIS ALGORITHM

Let's say that the number of tag images in the training group (R) and these images fall within (K) of the different classes (different marks) X_1, X_2, ..., X_K, Represents the vector of image i within class j, where j = 1,2, ..., K, i = 1,2, ..., N_j:

Step 1: The images of the marks used in the training are converted from binary matrices to mono-vectors.

Step 2: Calculate the vector mean μj and the Covariance Matrix Sj for the class j, j = 1 ... K according to the equations:

$$\mu_j = \frac{1}{N_j}\sum_{i=1}^{N_j} x_{ji} \qquad \dots\dots\dots(1)$$

$$S_j = \frac{1}{N_j}\sum_{i=1}^{N_j}\left(x_{ji}-\mu_j\right)\left(x_{ji}-\mu_j\right)^T \qquad \dots\dots..(2)$$

Where: Nj represents the number of symbol images in each category. Xji represents the vector of image i within class j. μj represents the vector of the mean of the symbol.

Step 3: Calculate the variance matrix SW The Within-Class Covariance Matrix according to the equation:

$$SW = \sum_{j=1}^{K} P(K_j)S_j \qquad \dots\dots\dots.(3)$$

Where: P (Kj) = 1 / K and represents a prior probability of the category j, The Prior Class Probability K represents the number of items (symbols).

Step 4: The Overall Mean μ is calculated according to the equation:

$$\mu = \frac{1}{R}\sum_{j=1}^{K}\sum_{i=1}^{N_j} x_{ji} = \frac{1}{R}\sum_{j=1}^{K} N_j\mu_j \qquad \dots\dots\dots(4)$$

Where: R represents the total number of images in the training group.

Step 5: Calculation of the SB-Between-Class Covariance Matrix by the following equation:

$$SB = \sum_{j=1}^{K} P(K_j)\left(\mu_j-\mu\right)\left(\mu_j-\mu\right)^T \qquad \dots\dots\dots(5)$$

where μ represents the overall Mean of the symbol.

Step 6: The Eigen vectors of the binary matrix are obtained from the following equation:

$$W = \frac{SB}{SW} \qquad \dots\dots\dots.(6)$$

Step 7: Calculate the weight vector for each image within the training group and the images of the marks used in the test according to the following equations:

$$Y = XW^T \dots\dots\dots(7)$$

Where: x represents the vector image of the tag used in the training .y: represents The vector of weights of the image of the mark used in the training or: the weight vector of the image of the mark used in the examination. [10] [11]

Step 8: Discrimination
Many images are distorted for a variety of reasons, such as JPGE image storage, which is a method of compression with the loss of a lot of information, so matching operations need precision and effort and the method used is efficient. When an unknown tag image is entered into the discriminating system, it is compared to a set of known images in which the feature is trained and the most closely matched image of the tag is returned, The matching process is usually measured using some distance metric measures such as Euclidean distances or by matching patterns such as correlation and others.

STRUCTURE SIMILARITY INDEX (SSIM)
Reference samples often show strong dependencies, especially when samples are close to each other in terms of location. This dependency contains valuable and useful information and recipes useful in the classification process. This information revolves around the visual structure of the object. This information can therefore be used to compare the structural similarity of the input image with the available images. It is possible to measure the quality of the image changed from the original image and use this measure to find the most similar picture between the test image and training images The SSIM method is used to measure the similarity of two relatively similar images, one of which has some distortions or minor differences and the second is of perfect quality. The method is simple, low-complex and provides accurate and good predictions in the testing process. Where the similarities are measured in terms of luminance, contrast and structure, for both images, by applying the following equation:

$$SSIM(A,B) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{ab} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)} \dots\dots (8)$$

where:

μa: the average of A.

μb : the average of B.

σ_a^2:the variance of A.

σ_b^2: the variance of B.

σ_{ab}: the covariance of A and B.

$C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$ is Two variables to enable stability in the partition process with a non-strong denominator.

L:The range of the pixel-values(which is equal to $2^{\#\text{bits per pixel}} - 1$).

By default k^1=0.01,k^2=0.03.

THE PROPOSED ALGORITHM
The images of the musical tags used in this search are illustrated in Table (1), where 150 training images of 15 different markers are collected, each with 10 different models in size, direction, colors, head orientation, clarity, etc.

8_1 Preprocessing of training images
The initial process is represented by the following steps:

The process of reading all the images from the folder and placing them in a single large three-dimensional matrix.

Resize images to be 25 * 25.

Convert all images from jpg to gray scale.

Convert each two-dimensional image into a single-dimension vector of 1 * 625.

Convert all gray images to binary images.

Return images to gray color quality again but color 1 and 255 exclusively.

Delete white vertical or horizontal lines (of color 255) that are around the mark.

We return the size to 25 * 25 which was affected by the deletion of lines.

The colors unite after the resizing process to 1 and 255.

In the end we will have pictures of the tag-free blanks (white lines).

8_2 Apply the LDA algorithm
The LDA algorithm is then applied to extract the properties and find the vector of the projection line.

8_3 The Preprocessing of the test image
It is in several stages to include:

The process of reading the image from the folder and placing it into a single matrix.

Resize the image to 25 * 25.

Convert the image from type (jpg) to type (gray scale).

Convert the two-dimensional image into a single-dimension vector of 1 * 625.

Converts the gray image to a binary image.

Return the image to the quality of gray colors again but color 1 and 255 exclusively.

Delete white vertical or horizontal lines (255) that are around the mark.

Cancel the staff lines if the test image if it contain the lines.

We return the size to 25 * 25 which was affected by the deletion of lines.

The colors unite after the resizing process to 1 and 255.

In the end we have a picture of the mark free of blank spaces (white lines) and staff lines if any.

Note the Preprocessing for the image in figure (3).

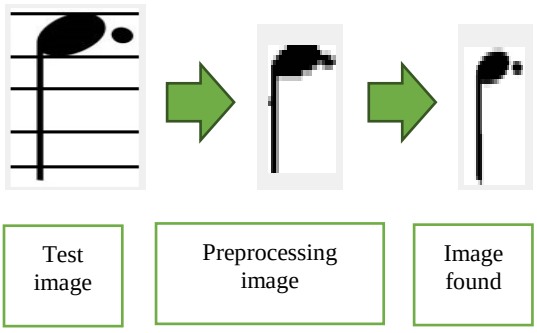

Fig 3: Stages of distinguishing one of the musical marks

8_4 Applying projection

The images are dropped which in the training group and the testing image is scanned on the weight line vector .

8_5 Check the similarity between the test and the original images

The SSIM scale is finally used, according to the equation 8.

Note the steps in the following figure (4).

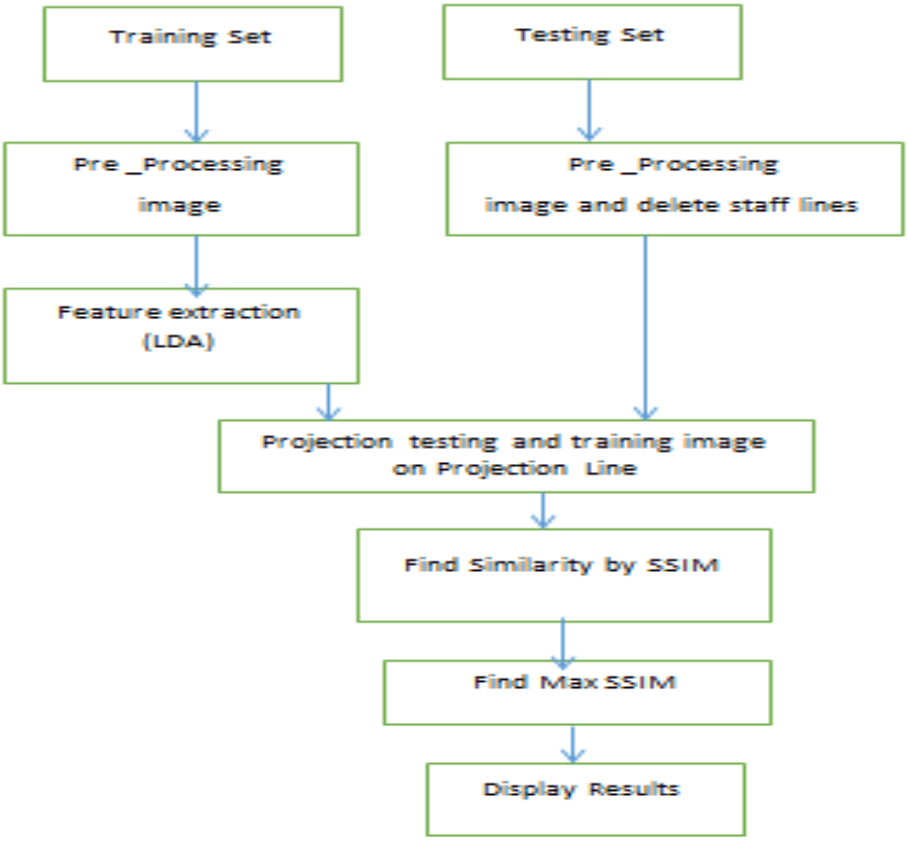

Fig 4:The main steps in the proposed algorithm

6. RESULTS

The algorithm was applied to the training and examination images, where the training pictures were 7 pictures for each score, totaling 105 training pictures , In addition, 5 images were examined for each marker. The total number of scan images was 75. Thus, the total number of pictures (training and examination) was 180. A very good percentage of discrimination was obtained as shown in Table (1).We used the tow equations (9 and 10)for calculates the results. The program can be developed in the future to process images of music tags taken with a mobile phone camera to be more useful for those with specializations and connect the software to networks to be accessible to everyone.

$$\text{Detection Rate DR} = \frac{\text{Number of sample correctly detected}}{\text{Total number of samples}} * 100 \quad \ldots (9)$$

$$\text{Error Rate ER} = \frac{\text{Number of sample wrongly detected}}{\text{Total number of samples}} * 100 \quad \ldots (10)$$

7. REFERENCES

[1] Hansen,B., Whitehouse,D., Silverman,C.,2014. Introduction to Music Appreciation, .ePress Course Materials.

[2] Montagu,J.,2017. How Music and Instruments Began: A Brief Overview of the Origin and Entire Development of Music, from Its Earliest Stages, Hypothesis and Theory.

[3] Schmidt-Jones,G.,2008. Reading Music: Common Notation, ©2008 Catherine Schmidt-Jones.

[4] Mohammed, jasem, Saja, 2012 .Standard printed musical note recognition based on neural network, J. of university of anbar for pure science : Vol.6:NO.2 : 2012, ISSN: 1991-8941.

[5] Kornfeld,J.,2005," Music Notation and Theory for Intelligent Beginners", Jason Dullack.

[6] Lee,C., Landgrebe,D.,1993,"FEATURE EXTRACTION AND CLASSIFICATION ALGORITHMS FOR HIGH DIMENSIONAL DATA", TR-EE 93-1.

[7] Nowozin,S., May 8, 2006. Object Classification using Local Image Features, Technische Universita¨t Berlin.

[8] M. A. Anusuya, S. K. Katti, 2009. Speech Recognition by Machine: A Review, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009.

[9] Eleyan A., Demirel H., (2007). PCA and LDA based neural networks for Human Face Recognition, source: Face Recognition, Book edited by :Kresimir Delac and Mislav Grgic, PP:93-106, I-Tech Education and Publishing, Vienna, Austria.

[10] M. Bishop,C.,2006. Pattern Recognition and Machine Learning, © 2006 Springer Science+Business Media, LLC.

[11] R. Webb,A., Ltd.,Q., UK,M, 2002. Statistical Pattern Recognition, Second Edition, Copyright , John Wiley & Sons, Ltd.

[12] Chetouani,A., Deriche,M., Beghdadi,A.,2010. Classification of Image Distortions using Image Quality Metrics and Linear Discriminant Analysis ,EUSIPCO_2010.

[13] Wang,Z., Li,L., Wu,S., Xia,Y.,2015 .A New Image Quality Assessment Algorithm based on SSIM and Multiple Regressions, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.8, No.11 (2015), pp.221-230.

Table 1.Results after testing images

NO.	Symbol	Test1	Test2	Test3	Test4	Test5	Sum	DR	ER
1		Y	Y	Y	Y	N	4	80%	20%
2		Y	Y	Y	N	Y	4	80%	20%
3		Y	Y	Y	N	Y	4	80%	20%
4		Y	Y	Y	Y	N	4	80%	20%
5		Y	Y	Y	Y	Y	5	100%	0%
6		Y	Y	Y	Y	Y	5	100%	0%
7		Y	Y	Y	N	Y	4	80%	20%
8		Y	Y	Y	Y	Y	5	100%	0%
9		Y	Y	Y	Y	Y	5	100%	0%
10		Y	Y	Y	Y	Y	5	100%	0%
11		Y	Y	N	Y	Y	4	80%	20%
12		Y	Y	Y	Y	Y	5	100%	0%
13		Y	Y	N	Y	N	3	60%	40%
14		Y	Y	Y	Y	Y	5	100%	0%
15		Y	Y	Y	Y	Y	5	100%	0%
Total	89.5%								10,5%