Presented at the Research Congress 2013
De La Salle University Manila
March 7-9, 2013

HAND SHAPE IDENTIFICATION FOR GUITAR CHORD IDENTIFICATION USING CENTROID DISTANCE AND FOURIER DESCRIPTORS

Gabriel Phillipe A Abiera and Bien Justin L. Mandac
Center for Automation Research, CCS, DLSU—Manila
2401 Taft Avenue, 1004 Malate, Manila, Philippines

Abstract: With the dawn of e-learning capabilities, the reach of technology in allowing the teaching and tutorial of lessons has never been easier. As such, using intelligent systems to monitor progress in lessons such as guitar lessons make e-learning with guitar playing more interactive. Using images of a person playing a guitar as input, the first step of being able to do this is to be able to discriminate the hand from the rest of the environment. Using a combination of skin tone segmentation and image differencing, the hand blob is obtained. This is then transformed into a 1-d signal and Fourier Descriptors are then extracted using Fourier Transforms. Results show that while some shapes are identified correctly, it is important to make the hand shape stay as 1 blob in order to make it easier for the system to identify the shape.

Key Words: Centroid distance function; Image processing;

1. INTRODUCTION

With the dawn of e-learning capabilities, the reach of technology in allowing the teaching and tutorial of lessons has never been easier. But to make e-learning more successful, it is necessary that such systems are able to interact and be able to identify the need for remediation by the system. This allows such e-learning apparatus to adjust its teaching parameters and be able to correct any errors that the student is encountering. In some matters, this is easy but in hands-on and practical skills, there are necessary technologies that should enable such a system. In the case of guitar learning, it is necessary for an e-learning system to be able to identify if a student who is learning correctly is playing the right chords.

An automated guitar chord identification system based on the principles of image processing can address such a scenario presented above. By providing the ability to the system to recognize the specific hand gesture per chord, such a system will be able to identify whether a student is performing a chord automatically. This can be integrated into a simple tutorial-type software that an interested student can install on a computer equipped with a regular web camera which is ubiquitous to users today.

The basic premise and involvement in this task is the ability of the system to be able to discriminate different chord positions based on the finger shapes of the user on the guitar fret board. This ability would enable the system to identify whether a correct chord is being played correctly along with its correct form. This will allow the system to react via a feedback of correct or not to the user, allowing the user to adjust his position in order to attain the correct position and be able to master it. To develop such a system, it is then first necessary to create a system that is able to discriminate between the different positions of the hand and identify if its shape corresponds to a known guitar chord.

2. METHODOLOGY

To identify the hand shape that corresponds to the guitar chord to be identified, the system subjects the input images consisting of a guitar player with the fretboard clearly in the image. The algorithm would then subject said input image to a preprocessing, then to the extraction of image signature and finally for signature matching.

a. Preprocessing of Input

The first step of the system is to be able to discriminate the hand from the rest of the environment. Currently, this focuses on discriminating the hand from the fretboard and assumes that the camera only has view of the fretboard similar to the following input image:

Figure 1. Sample A Chord

The input images were obtained in a controlled setup. The guitar was stabilized in a way that forward and backward movements of the guitar body and fretboard are restricted as much as possible. This was done so as to minimize any differences in scale between images of the same set of chords hence, simplifying preprocessing. Moreover, images of a particular set were manually cropped and aligned using commercially available image processing software.

Current methodologies for discriminating the hand from the board involves the use of two methodologies. The first involves the use of image differencing wherein the fretboard is taken as a background image. This image is then used as a basis; that is, it is subtracted from the input images in order to find the general location of the hand. The tolerances for this procedure are set such that the system has a large amount of leeway and is not necessarily strict in locating the hand. As such, it can be said that this step aims to locate the hand on a general area. Another methodology used by the system is the use of segmentation by detecting skin tone. The skin tone algorithm detects for the skin tone based on a certain tone value. It is also helped by an illumination compensation using the greyworld algorithm. This step is used as is from an implementation of [Jain 2012]. By performing image binarization on the products of these two methods, and performing and an AND operation, an intersection of the two images is obtained. This image is further refined by using morphological operations (ie erosion and dilation) that remove holes and smaller blobs that could have been created by the process above.

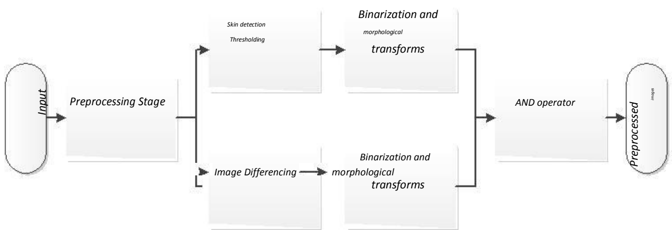

Figure 2. Preprocessing algorithm

The resulting image below then becomes the candidate image for processing.

Figure 3. Binarized Image obtained from Figure 1

The resulting image may include more than one blob as the distance between the fingers may be such that it will appear as two blobs instead of one.

b. 1-d Transformation

In order to identify the shape of the hand and interpret it, it is necessary that the shape is essentially identified. The method used in this case is a 1-d transformation that is based on the Centroid distance of the image. A 1-d transformation is a transformation that allows the 2-d shape information in the image to be transformed into a 1-d signal. In this case, that transformation is based on the distance of the centroid from the boundary of a blob. The centroid is the center of mass for the blob, that is, it can be taken as the point such that it is equally placed at the balancing point of the image. A distance from this center to the boundary is therefore taken as thethe transformation that will encode the shape details of the blob. This transformation changes a closed blob into a 1-d signal that in this case is based on an equi-angular sampled distance from each other.

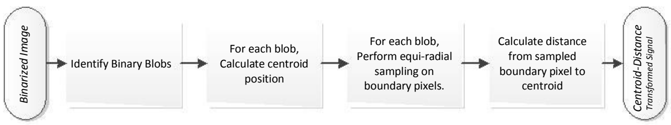

Figure 4. Centroid Distance Function Transform

Note that the values on this are normalized from the largest magnitude of distance sampled. This use of a transformation allows invariance with regards to translation. It also shows some noise resiliencies as the coefficients are not going to be affected heavily by simple noise. In this case, the sampling into a number of samples that is based on an equi-angular distribution of the points allows a normalization of the system, capturing (in this case, a sample every 360/256) ensuring that there are an equal number of boundary points per identification run. Thus there are 256 points in each Centroid distance function performed on a blob. The reason for using a 256 count is to be able to take advantage of the speedup associated when using power-of-two FFT.

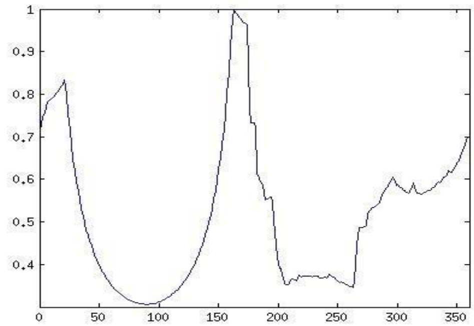

Figure 5. Normalized Centroid-Distance Transform of Figure 2.

c. Matching to Signature

A Fourier transform of this signal was then obtained, and because all inputs are sampled at the same equi-angular rate, then all samples will produce a Fourier coefficient equivalent to the number of samples/2. This allows for a comparison with another set of Fourier transforms on a controlled signature file that can be used to identify the shape. In this case however, it might not be necessary to compare all Fourier coefficients as most of the shape's general information is only recorded in the lowermost coefficients. The algorithm in this scenario uses 8 coefficients to determine a match in the signatures.

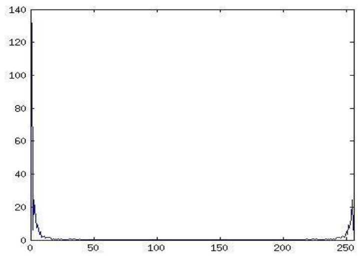

Figure 6. 256-point FFT Magnitude plot of Centroid Distance Transform from Figure 3.

A match is made when on the input when it receives the highest scores from one of the signatures loaded. A match is tested by first normalizing each of the Fourier coefficients with the DC component of the results to make it more scale-invariant. Then each value is subtracted from the recorded signature values. The residuals are then divided over the normalized signature Fourier coefficients. Each coefficient essentially has equal weights and the resulting residuals/normalized coefficients produce a certain percentage (if it exceeds 1.0, it is clipped to 1.0).

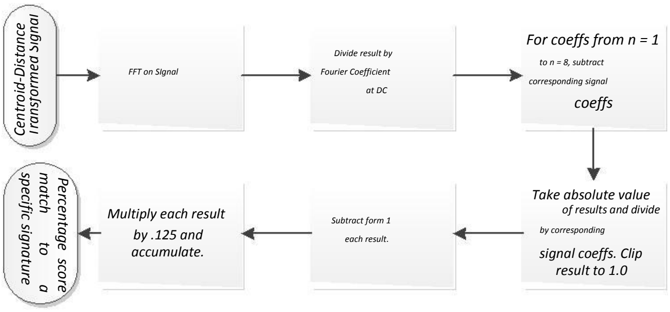

Figure 7. Score assignment algorithm

Thus each coefficient contributes around 12.5% to the score. This score is then subtracted from 1 to show its percentage of similarity to the coefficients. In cases where the input and output being matched contain more than 1 blobs, then the best score total is taken. Note that in such a scenario, the scores are weighted such that the blob that occupies the most area in the signature being compared against has the most weight on the matching.

3. RESULTS AND DISCUSSION

Preliminary tests of the identification algorithm show the following results:

Table 1. Matrix of correctly identified images from the result set.

	C	D	E	F	G	A	B
Identified as value	2	2	4	3	3	4	4

Note in this results that the in each sample pile, there are 4 correct images and 24 incorrect images per signature being tested (thus a total of 28 images tested per signature).Thus a score of 4 represents identifying all samples that are supposed to match the given signature in the sample set. Likewise a sample matrix result for a test image set of chords consisting of one chord per image is given below:

Table 2. Sample Matrix of One Set to showcase false positive cases.

	A signature	B signature	C signature	D signature	E signature	F signature	G signature
A	0.9187	0.3536	0.6182	0.5458	0.6559	0.6585	0.4139
B	0.5317	0.7964	0.6862	0.5289	0.5023	0.4348	0.4233
C	0.4235	0.4311	0.6467	0.5095	0.5269	0.6844	0.3374
D	0.3415	0.6300	0.5522	0.5893	0.3086	0.4469	0.3327
E	0.6708	0.5954	0.5994	0.3624	0.8689	0.5067	0.5230
F	0.3898	0.4395	0.5539	0.4979	0.3415	0.7399	0.3180
G	0.6346	0.5200	0.4890	0.5001	0.6133	0.5288	0.7099

Note that the columns represent a score determined per comparison to the corresponding signature file. Each row represents a known sample image being subjected to the identification routine. Thus, it is expected that the diagonal contain the match values if all matches are correctly made.

The results of the initial algorithm shows that most of the difficulty is in the parts where there are multiple blobs in a single image to identify the shape, both in the signature and in the input file. The design of the algorithm did not anticipate that the hand shapes would split into more than one blob and because the cropping removes some parts of the hand which are contiguous, it results in such a scenario. The results also show the need for further testing using more data. Currently the research has access to only two sets of data of sufficient quality. In order to test the identification routine and use the other data sets, it is required that a better segmentation and binarization pre-processing procedure is first devices.

The algorithm however shows the feasibility given that the shapes extracted from the images are far from ideal compared to the signatures and is thus potentially capable of delivering the expected results, given more tuning. However, based on the matrix of results from a sample set of images, it is clear that a better method of identifying a match score is needed instead of the current method.

4. CONCLUSIONS

As stated before, the following represented the initial algorithm design for the application stated. While preliminary test shows promise, it is necessary that a larger data set is acquired and tested through the algorithm in order to provide more proof. Likewise, there is a need to improve the preprocessing portion which segments the hand from the background. For example, it is ideal to keep only one blob to represent the hand and identify its gesture, instead of the current way of having multiple blobs represent the hand shape. In retrospect, the weighing system is not very efficient and optimal and certain boundary conditions are bound to skew these results.

To keep a single blob, it is recommended to readjust the cropping of the image from the fretboard in order to make the entire hand more visible. Likewise, there is a need to improve the binarization portion. One improvement could come in the form of the utilization of a colored glove which can then be used to possess the color of an easily isolatable color for which segmentation and binarization could be performed, similar to those used in green screens for film production. Because chords are sensitive to the location along the fretboard, it is necessary to identify the location of the hand and the figures in addition to merely identifying shape. A grid-based technique can be used for this, using the guitar strings and frets as a guide for the formation of the grids. Lastly, a means of automatically extracting the fretboard area from a given image is necessary for the system such that it would automatically segment the fretboard area of concern along with applying various transforms to said image prior to preprocessing for identification.

5. REFERENCES

Bhonsle, S., & Klinzmann, A. (2011). Centroid Distance Function and the Fourier Descriptor with Applications to Cancer Cell Clustering. Retrieved from http://math.uci.edu/icamp/summer/research_11/bhonsle/cdfd.pdf

Dengsheng, Z., & Guojin, L. (2002, November). A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures. *Signal Processing: Image Communication*, *17*(10), 825-848.

Jain, G. (2010, August 26). *Skin Detector*. Retrieved from http://www.mathworks.com/matlabcentral/fileexchange/28565-skin-detection

Ronsin, J., Yang, M., & Kpalma, K. (2008, July 15). A Survey of Shape Feature Extraction Techniques. Retrieved from HAL: http://hal.archives-ouvertes.fr/docs/00/44/60/37/PDF/ARS-Journal-SurveyPatternRecognition.pdf