iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: V Month of publication: May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5349

A survey on Image Musicalization using Image Emotion Analysis

Soorya Vinod[1], Swathi Uttarkar[2], Umme Kulsum[3], Mansi R Setty[4], Prof. Prasad AM [5]
[1, 2, 3, 4]*Dayananda Sagar College of Engineering*

Abstract: In this survey paper, we summarize various other papers and as a result of these reviews we propose a project with the most suitable paper that gives methodologies to musicalize images based on their emotions. We choose our base paper which proposes to extract visual features, inspired by principle-ofart, to recognize image emotions. We musicalize the images by comparing these emotions to the emotions extracted by music data. The final part deals with what to do with this converted music data where we address network security involving the study of Encryption and Decryption of the music data so generated.

I. INTRODUCTION

Conversion of image data to musical/audio data has applications in a variety of fields, ranging from sensing augmentation, helping visually impaired in analyzing pictures, network security and multimedia applications. In the proceeding paragraphs, we review some of the most relevant contributions for the same.

Some of the papers have methods and algorithms for the same which are tested practically over a variety of data sets, and we conclude by the end of this paper, which method is the most appropriate to be used and why. We also review papers that talk about optimization of these algorithms and the input data set.

Since we use the converted musicalized image data in the network security space, we see methodologies for doing so and also to convert the music back to its original form at the decryption end.

Following sections have a brief review of literature about Image Musicalization using image emotion analysis:

A. This paper introduces a first-time approach for conversion of image data to sound. In this approach from the input image data it considers the music parameters pitch and rhythm in order to translate images to sounds. This paper uses the methods like Auditory Perception which provisions mapping data to specific dimensions of sound. But it is not suitable for complex parameters of music as this method uses only the parameters such as loudness, pitch, intensity, tone etc. The other method is the voice print technique. In this method, the user is trained for the outcomes and the non - trained users were unable to relate to the produced sound. The proposed method in the paper uses the parameters such as rhythm and pitch of transformation image to sound. It supports various instruments like piano, guitar and trumpet.

B. This paper talks about a method for producing an optimized and resized version of the input image. It deals with searching for the most aesthetic part of an image and cropping accordingly using various guidelines of aesthetic images which are rule of thirds, Diagonal dominance and Visual balance. Here, computationally the rule of thirds, Diagonal dominance and Visual Balance scores are calculated and an aesthetic score function is generated as the combination of these three scores. The score generated is then used to optimize the image using a method called Particle Swarm Optimization (PSO). The objective of referring this paper is that it helps in refining the input image data in our project to a standard quality before converting it to sound as it helps in maintaining uniformity in the input data.

C. In the paper, we talk about the relationship between visual arts and music. With the increasing amount and interest in sonification techniques and image processing techniques using machine learning, it is now a real possibility to transcribe visual experiences into appropriate auditory experiences. The use of these advances in technology, however, has been focused on the general sonification of images and increased accessibility through short sound feedback. On that basis, we decided to design a sonification algorithm designed to transcribe visual artworks and communicate its wider cognitive and emotional meaning appropriately, with the goal of setting guidelines for the sonification of artworks based on their characteristics.

Table 1: Initial mapping pairing strategies considered between visual and auditory parameters

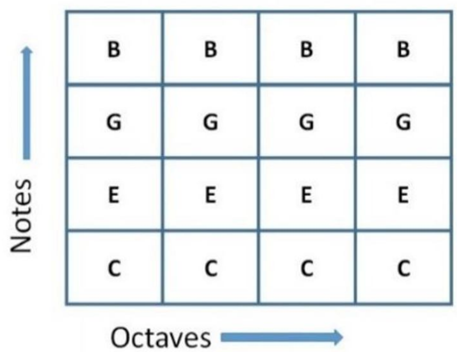
Visual Parameters \ Auditory Variables	Pitch	Tempo	Mode	Timbre	Loudness	Musical Composition
Hue	X					
Brightness		X			X	
Saliency	X		X			
Size	X					
Art Style				X		X

To compose and differentiate sound outputs for different artworks, the sonification algorithm would use mapping strategies between a variety of visual and musical parameters (Table 1).

Data processing techniques, including shape detection and sentiment analysis machine learning algorithms, would add further accuracy to the sonification program following the proposed experiments. Future tests and implementation for art galleries would be evaluated using the completed algorithm, which should achieve the goals set at the beginning of the project.

D. In this paper, the authors try to summarize different types of methods for mapping different dimensional images to sound. They collectively find out various methods for sonification. Methods like query based which ask the user for information about images, tour based methods which has previously routed the information of the images for the model. They use parameters like paths which is like a pipeline which defines the order of the information of images using path points, spans which define the boundary or the neighborhood for the points to depict the data from. There are many other parameters which support these models and different types of paths to identify the images. This information collected about images are mapped or connected to sounds via functions. Then they create a prototype system for this information collected by the model. So, this paper mainly deals with finding different methods of sonification and its journey.

E. This paper presents a unique method of sonifying various graphic objects, such as color pictures, based on an approach that is a combination sound communication with speech. A special color model is introduced in the paper, which supports this approach. The color model introduced has properties that can be used in a convenient way to make the graphical information available for hearing.



The semantic color model offers an effective basis for complex graphic object sonification. The client has a limited set of common primary colors to work with. There are only two primary colors presenting any color.

Such properties make the system interesting and convenient for both the sound and verbal modes of information delivery.

Obviously, the different arrangements of these features are not invariant. To describe specific semantics and emotions the elements must be carefully arranged and constructed into meaningful regions and images. The rules, techniques or instructions for the arrangement and orchestration of art elements in a work of art are known as the principles of art. Hence our base paper includes principles of art.

F. In this paper image is represented using audio sound series. Initially the image input is represented in an array of 4 rows and columns. For every sound element assign intensity and an audio. The audio is assigned to position of sound element. From these elements image intensity is computed. Audio for the element is the audio for element being played at that computed intensity. On merging 16 audios generated for the elements the audio for image is obtained. One finding says that mostly more sound is generated from images which is impossible for a person to detect minute differences among audios when more noise is present. This can be brought down by constraining the generated amount of noise. This leads to a less noisy result.

The below figure shows representation of octaves, notes for the divisions of sound element for the image.

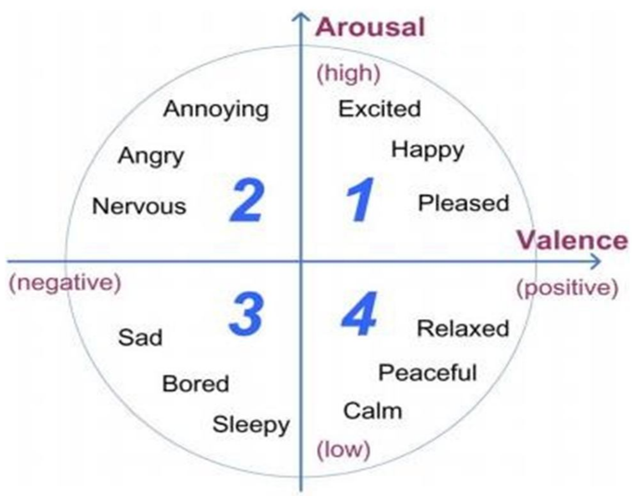


Fig. 1. Thayer's arousal-valence emotion plane.

G. In this paper, it is mentioned and explored how the roundness and complexity of shapes are the two important factors to understand an image's emotions. According to this paper, valence, arousal and dominance are the three dimensions that affect any image and these three dimensions have been defined accordingly. SVM regression with RBF kernel is used to model the valence- arousal values and it is found from the MSE values that visual shapes provide use in understanding valence instead of color, texture and composition. Line segments, angles, orientation, continuous lines, curves are the shape parameters used in this paper to capture emotion. The paper also stresses on why and how the IAPS dataset is important to use in this process.

H. This paper on music emotion recognition, talks on how the regression technique is better than Arousal-Valence modelling, Fuzzy Approach and System Identification Approach of detecting music emotions. The graph depicts a valence arousal plane with various emotions which holds true for music as well as images. The second half of our project is feature extraction from the music dataset and feeding music emotions, which are then used to compare to the image emotions and hence produce the final output of music for the respective image. The approach mentioned in this paper produces near accurate detection of music emotions which forms the music subsystem of our project.

I. In this paper the design of the SUM tool, a graphical user interface software library inside PWGL's computer-aided composition environment, aimed at image and sound integration. It allows image and sound to be integrated via a graphical interface. It offers a greater temporary technique to spatial composition via sonification – facts representation thru auditory manner. SUM believes in a multi-dimensional spatiotemporal technique to sonification, which distinguishes it from other strategies like sonart. One of the excellent applications of the sum tool is image sonification, the sum tool helps the sonification of any shade-coded picture with its image-based input and user-defined mapping. Which means that it's miles feasible to sonify any bitmap photograph in step with its personal color key. Picture and sound integration in more than one spatial and temporal dimensions are supported through the sum tool. As seen in this paper, we can use it to jot down and perform a multi-dimensional visible musical score from many perspectives. Sum's flexible structure typically lets in audiovisible representation of more than one such relationships, from an urban gadget to a musical rating.

J. In this paper, the concept of musical vision is introduced, a gadget that makes use of the music to symbolize a pixel-based photography.

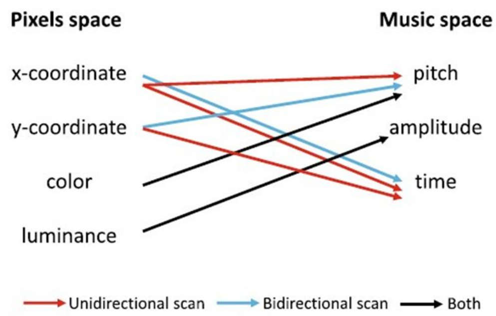
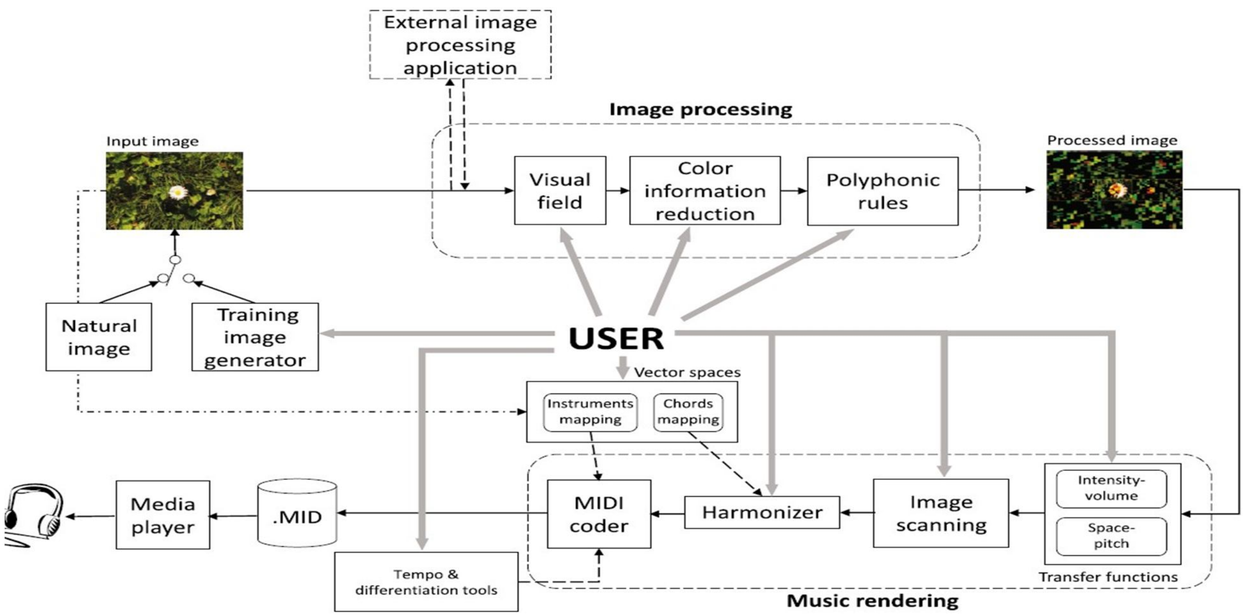


Fig. 3 Correspondences between the attributes of the pixel and music spaces depending on the type of image scanning pattern

Musical vision is an extremely flexible, sonifying device that converts colored pictures into periodic sounds by imitating the human eye in terms of their subject of vision. MV can code any color into music in preference to codifying a confined range of colors or hues. Furthermore, flexibility is the main strength in this proposal. MV allows more than one aspect of preprocessing image data and music rendering techniques to be configured.



Block diagram of the Musical Vision system architecture. The gray arrows show the sonification controls available to the user

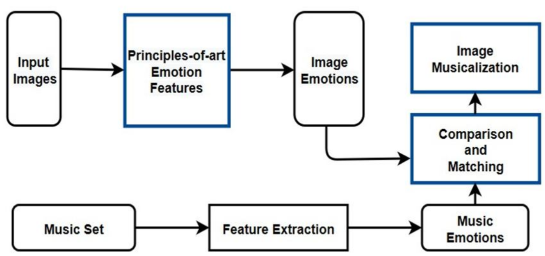
Almost every aspect of sonification is controlled by the Musical Vision user. The consumer of musical vision, as referred to above, has an excessive level of manipulate of the sonification parameters, however the versatility of the machine makes it usable for specific purposes of image sonification. This makes it possible, relying on the want for photo definition fidelity, to render melodies of very one-of-a-kind length and complexity**.**

K. In this paper, focus is on prediction of the probability distribution of categorical image emotions and the experiments that are carried out, demonstrate the superiority of categorizing images using principles-of-art over state-of-the-art approaches. There are two emotion representation models- CES and DES of which prediction in CES form is highlighted here. They have introduced three baselines and then one main algorithm to predict image emotions. Global weighting, K-nearest Neighbor Weighting and SoftMax Regression are the three baseline models and Shared Sparse Learning is the algorithm used which is iteratively reweighted least squares to get the optimized result.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

L. This paper introduces us to the two emotion representation models and the principles of art approach which shall be used by us for the development of our project. On research of most recent work on image emotion analysis uses "visual features based on elements of art" (EAEF). But these lowlevel elements have weak link to emotions i.e., it is said to be vulnerable. It is also not invariant to different arrangement of elements. This results in poor performance for emotion recognition for the given image. As low-level features include color, space, value, texture, line space and form, these low-level features are not capable of representing high level emotions. Therefore, we prefer PAEF (principles of art-based emotion features). They include the parameters like emphasis, balance, harmony, movement, gradation and variety. They are more semantic than EAEF and also that symmetry and variety are human understandable when compared to texture and line. PAEF effectiveness is evaluated by two ways namely AIC and Emotion score prediction. AIC depends on emotion category such that various combinations of principles express various emotions. In Affective image classification balance and harmony positive emotions often whereas variety and emphasis classify all 8 categories of emotions. Therefore, the results of experiments on AIC have demonstrated that the PAEF has superior performance.

M. In this paper, they use the emotions or the feelings of the visuals or images to collect data from them to convert them later into sound or music as a form of security encryption or for visually impaired people to get to know the visuals in depth.



They use emotion based models to categorically collect graphical image information or by using regression machine learning techniques. They collect certain parameter which not stable to changes as the image are at a stable state, for videos they collect information which can change as we can select the music accordingly.

They use parameters like balance, hue , contrast, etc for measuring the images.

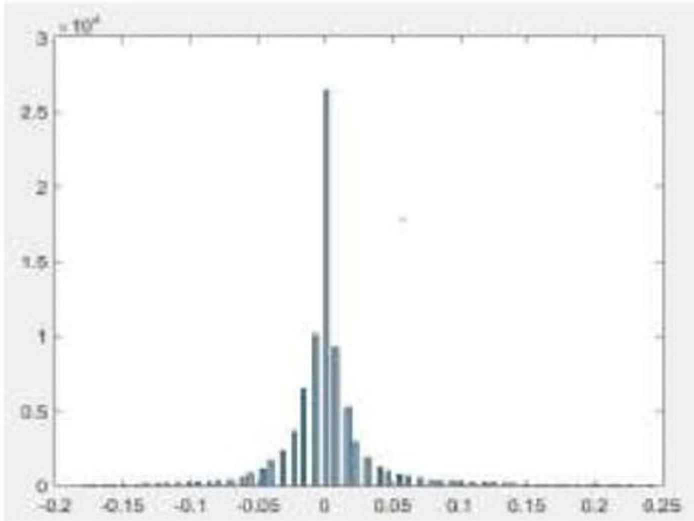
They use different datasets for music and different dataset for images which have pre defined parameters in them. They compare these two collections based on the similarity levels to conclude which music is apt for which image.

They used human intervention to know the comaprisons as they told humans to select music for the images at first. Based on this huge data, they regressed the model. They then selected random image for which the model gave the apt music which had most similarities with.

As emotions of the images are unique to the images, they selected this method. As a result, it turned to be very efficient compared to the primitive methods used in previous papers.

N. This paper proposes a method for encryption of sound/audio signals which is a secured mode for transferring of audio using encryption. This combats the unintended users who then cannot decode user's confidential and precious data. Data is scrambled on encryption and then a secret message is produced which is then read or can also be decrypted by a password or a key. Here we use diffusion and confusion principles to produce the encoded data. Confusion is making bond between encrypted output and the key as complex as possible whereas diffusion is that many of the characters of cipher-audio changes if single character of plain-audio is changed which is the input. Cryptography is implemented using hashing algorithm. This is one of the most secured hash functions. DWT is used together with Inverse DWT. These increases time efficiency. For efficiency and security, hashing and encryption are two frequently used terms. Plain audio is the input for hash function here and then alphanumeric string of constant size (called a message digest or hash value) is generated. Then encryption is performed on music file that is to be stored initially in LX2 matrix form. Then we perform confusion and diffusion to make sure that the audio is not vulnerable. To improve timing efficiency, both the mentioned methods are included with decryption and encryption. Later we analyze key using statistical analysis such as peak sound to noise ratio, correlation analysis and histogram. The data distribution for matrix of plain audio is as shown.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com



Histogram of original audio.

The below shown figure says that data is uniform and closer which says it is a good encryption and that distributions that are closely packed are hard to be decrypted without the external keys.

II. CONCLUSION

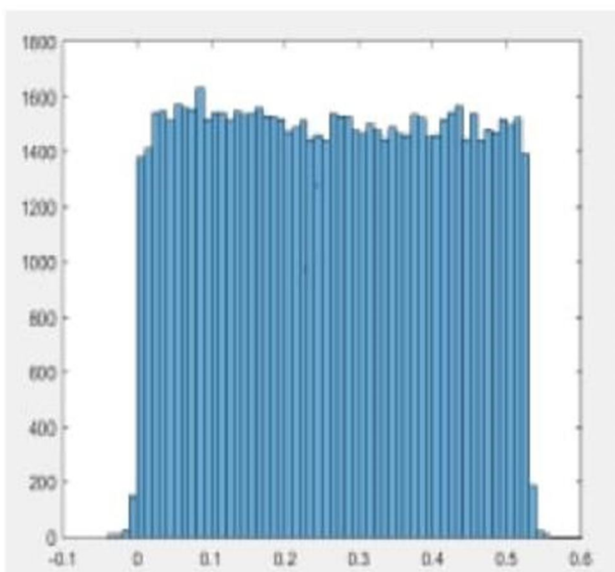
We choose [13] as our reference base paper as they use regression and classification techniques instead of just color-based classification technique to extract emotion from the images, as emotions depict better understanding of images rather than just colors and lines as they are invariant to changes.

Also, the paper has many advantages over other papers like:

First, the method proposed exceeds performance compared to the other two methods as it considers the prediction over average emotions.

Secondly, the music data chosen in this method is far better than the ones selected randomly for Image Musicalization.

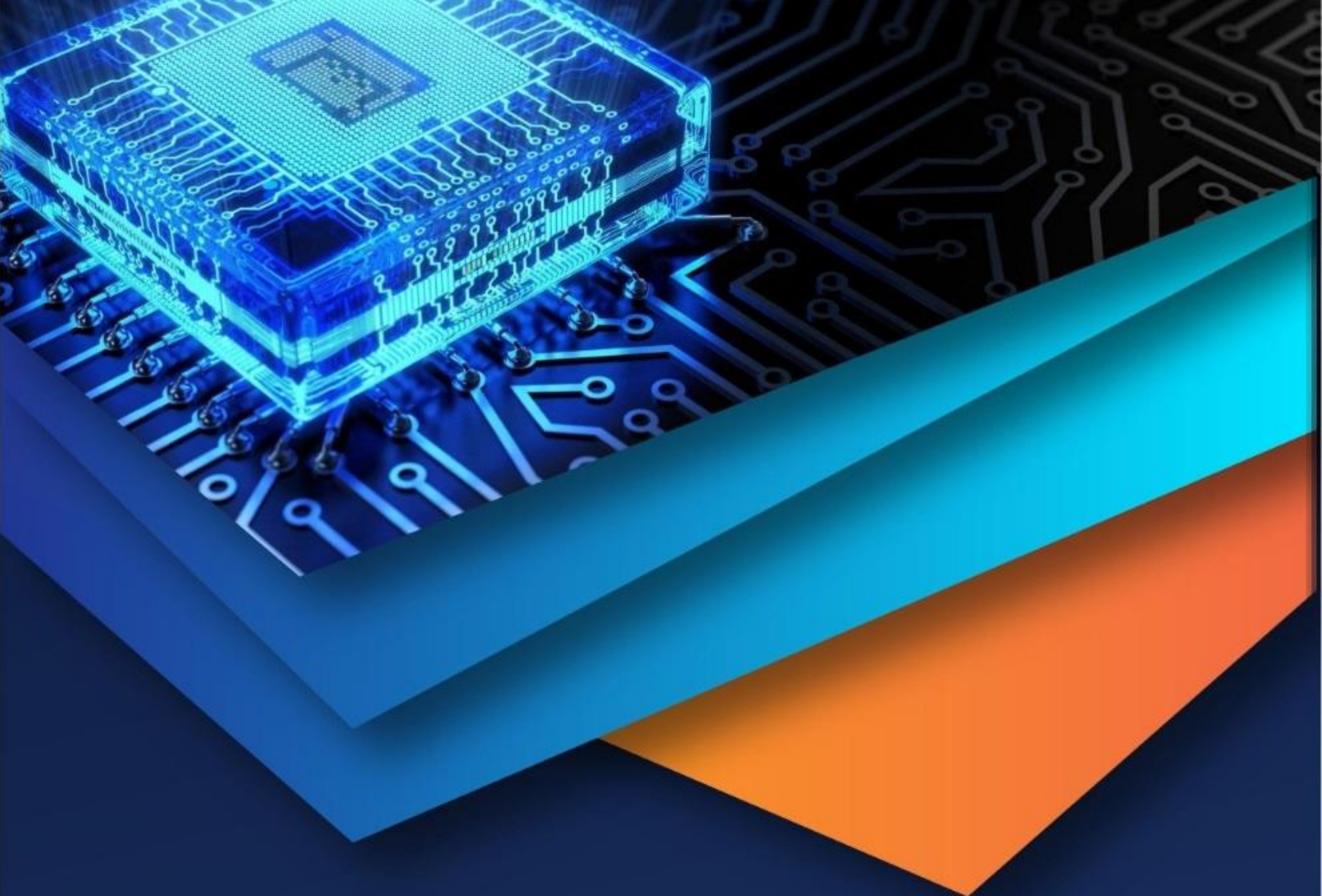
Finally, the method hence proposed has better performance and efficiency for the images with strong emotions and Humans



Histogram for encrypted audio. can pick up the images with less emotions, as emotions are subjective. All these reasons summarize why we find this paper as the best reference to go ahead with our project.

REFERENCES

[1] "An approach for image sonification" by suresh matta,dinesh k kumar,xinghuo yu, mark burry ieee 2004
[2] "Optimizing photo composition" by Ligang liu,renjie,chen lior wolf, daniel cohen-or(graphics forum, 2010)
[3] "Preliminary guidelines on the sonification of visual artworks: linking music, sonification & visual arts" by chihab nadri, chairunisa anaya, shan yuan, and myounghoon jeon (the 25th international conference on auditory display (icad 2019))
[4] "Path based model for sonification" by keith m. Franklin, jonathan c. Roberts (computing laboratory, university of kent at canterbury, england 2004)
[5] "Hybrid approach to sonification of color images" by ivan kopecek,radek oslejsek (third 2008 international conference on convergence and hybrid information technology ieee 2008)
[6] "Representing pictures with sound" Edward m. Schaefer ieee 2014
[7] "On shape and the computability of emotions" by xin lu,poonam suryanarayan,reginald b. Adams, jr. Jia li, michelle g. Newman,james z. Wang
[8] "A regression approach to music Emotion recognition" by yi-hsuan yang, yu-ching lin, ya-fan su, and homer h. Chen ieee 2008
[9] "Sum: from image-based sonification to computer-aided composition" by sara adhitya, mika kuuskankare
[10] "Musical vision: an interactive bioinspired sonification tool to convert Images into music" by antonio polo,xavier sevillano(journal on multimodal user interfaces © springer nature switzerland ag 2018)
[11] "Predicting discrete probability distribution of image emotions" sicheng zhao, hongxun yao, xiaolei jiang, xiaoshuai sun ieee 2015
[12] "Exploring principles-of-art features for image emotion recognition" By sicheng zhao, yue gao, xiaolei jiang, hongxun yao,tat-seng chua, xiaoshuai sun(proceedings of the 22nd acm international conference on multimedia 2014)
[13] "Emotion based image musicalization" by sicheng zhao, hongxun yao, fanglin wang, xiaolei jiang, wei zhang (international conference on multimedia and expo workshops (icmew) ieee 2014)
[14] "Dwt based audio encryption scheme" by chloe albin,dhruv narayan,ritika varu,thanikaiselvan v ieee 2018



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY