International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 7, Number 1 (2017), pp. 35-45

http://www. irphouse.com

Tracking Conductor's Hand Movement For Generating Artificial Music in Virtual Orchestra System

Harsimrat Kaur and Sonit Singh

Department of Electronics and Communication Engineering
Lovely Professional University, Punjab, India.

Abstract

Music having its long history and existence in all cultural contexts plays an important role in the social and personal lives of human beings. Common characteristics of music are tempo, beats, dynamics, articulation, nuances, etc. Conducting is the art of directing a musical performance by the way of visible gestures. Conductors act as a guide to the orchestras/choirs they conduct. The primary role of conductor is to unify musicians, set the tempo, beat, articulation, nuances etc in the music in order to the maximize the overall performance of the orchestra. In this paper, automatic recognition of orchestra conductor's hand movement is proposed for generating music which will provide the sense of real orchestra. Also, the proposed methodology is compared with related work found in the literature. The results of the proposed approach are promising and further improvements can be done in order to have completely virtual orchestra systems.

Keywords: Virtual orchestra system, artificial music, music conductor, electronic music, orchestral musical patterns, segmentation, feature extraction, tracking, kalman filter.

I. INTRODUCTION

Music being an abstract form of art with a long history is being realized today increasingly with the electronic means. Capturing and analyzing an orchestra conductor's movements has become one of the emerging topics of research from the

few past years because of its varied applications in the field of education and entertainment. Conducting is an ability of directing a musical presentation using visible gestures. Conductor is a person who acts like a leader, guide or director of music and sometimes also called as concertmaster. The conducting requires understanding of the elements of musical expression and the ability to communicate them effectively to a group. The conductor communicates his intentions to the orchestra with gestures in order to control the various musical parameters such as tempo, dynamics, nuances (staccato and legato), phrasing, etc. Despite having no direct control over the sound being produced, the conductor is able to drastically affect how a piece of music sounds solely through the use of gestures. A conductor is also in-charge of the rehearsals that lead to the successful performance of concert. The primary responsibility is to unify performers, adjust tempo, and execute clear preparations and beats and to control the interpretation and speed of music. Automated conductor system is the method which involves generating artificial music with computer controlled instruments. The software is responsible to synthesize the music [7] according to the conductor's movements like real musicians generating music.

This type of system provides a tool to study musical expression using gestures. Providing an explanation of gesture enables thorough study of movements and makes users to understand the gesture information. The musical information is translated into sound using the physical gestures. These methods will enable precise understanding of the functions of the orchestral conducting gestures and is also used to combine dynamic audio development to live concerts. The fully functional system is also used in the educational environment in which the teacher shows students a specific style for conducting technique or to conduct a specific piece of music or to record specific gestures, the student can then practice. The resemblance or the difference between the movements is calculated and displayed to the user, which helps to practice orchestra conducting.

In this paper, an approach has been proposed to track and recognize conductor's hand gestures. Further artificial music has been generated without the help of human musicians according to the recognized conductor's hand movements. The proposed method is divided into various stages namely: preprocessing of the video frames, segmentation of the conductor's hand, tracking of the hand gestures, recognizing (gesture meaning analysis) of hand gestures and finally music synthesis (producing music according to the conductor's hand movement). The process begins with preprocessing in the form of background modeling and filtering operation in order to adapt the gradual variation in the background and various other effects such as improper lightning conditions, field of view of the camera, etc. Further segmentation is done in order to detect the conductor's hand and then tracking of hand movements of orchestral conductor is done. After tracking feature vector is formed in the form of

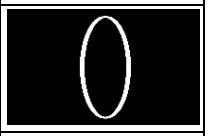
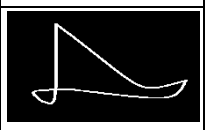
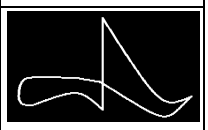
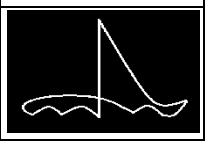
a vector in the form of the tracked position coordinates. The feature vector is then feed to the classifier in order to recognize the hand gestures and accordingly music is being generated by the system which resembles the actual musicians playing music in real-time.

II. ORCHESTRAL MUSICAL PATTERNS

Although the music is already written on the manuscript in front of every musician, the conductor makes the musicians aware of change in dynamics in the piece. Beat of the music is indicated with the conductor's right hand, with or without baton. The hand traces a shape in the air in every bar depending on the time signature, indicating each beat with a change from downward to upward motion. There are different rhythmic patterns for example 4/4, this is made up of 4 beats and is counted; 1,2,3,4, repeat. The orchestral conductor would create 4 points with his/her hands or a stick called baton; down, left, right, up repeat. These represent the timing of the 4 bests in the bar; down (1), left (2), right (3), up (4) repeat [8]. The dynamics of the conductor relate to the dynamic of the music. Music arrangement also known as mixing is a part professional music production practice.

The beat of the music is typically indicated by the cyber conductor's right hand. Conductor may use or may not use the baton. Conductor's hand traces shape of the pattern in air depending upon the time and signature. The time signature is compound if the tempo is slow. Sometimes the cyber conductor indicates "subdivisions" of the beats. The subdivision is done by adding smaller movements min same way as it is already moving for the beat to which it belongs. The changes in the tempo are indicated when the speed of the beat is changed. The dynamics may be communicated by the size of conducting movements. If the shape of the pattern is larger, then it represents louder sounds. The change in the size of conducting movements will result in the changes of character of the music. The dynamics needs to be fine-tuned and can be done through various gestures. The way in which the beats are grouped together is known as "Meter". The common meters are shown in the Table I. The time signatures 2/4, 3/4, 4/4 fall under Simple meters and the time signature 1, 3/8, 6/8, 9/8, 12/8 fall under compound meters.

Table I: Various orchestral conducting patterns

S.No.	Time Signature	No. of beats per minute	Beat pattern
1	1 (Compound meter)	1	
2	3 by 4 (Simple Meter)	3	
3	4 by 4 (Simple Meter)	4	
4	2 by 4 (Simple Meter)	2	
5	6 by 8 (compound Meter)	6	

III. RELATED WORK

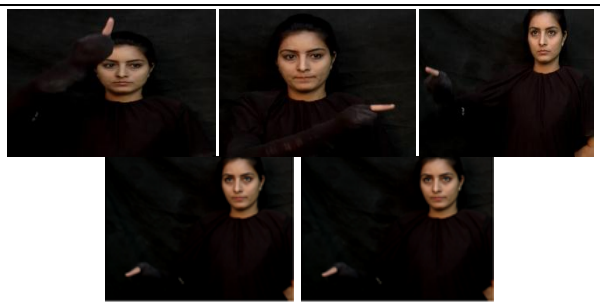
Typical approaches include optical or camera based approaches, non-optical or sensor based approaches and multimodal approach combining both camera and sensors based approaches. Optical based systems use cameras and use computer vision techniques to track and recognize the conductor's hand gestures and extracting relevant information from this. Optical systems have also been used to track positional markers attached to the conductor's hand. Two cameras have also been used to create 3 D image of the baton in the hand of the orchestra conductor which reconstructs the baton's position and consequently used to control computer music system. Apart from this, single camera is used to note the amount of motion by doing the difference between consecutive frames of the incoming video captured by the camera. Sensor based systems involve embedding various sensors in the conductor's clothing or installing system of sensors in the form of magnetic positional sensors or accelerometer [16]. This allows non-visually observable data to be acquired in the form of position vector for the movement of arms and hands of the orchestra conductor. Each approach has its pros and cons. The optical approach is non-intrusive in nature but captures only visual information which is limited by the field of view of the camera. The sensor based approach provides greater range of data in the form of positional vectors but being intrusive in nature, affects the performance of the orchestra conductor [12]. The type

of approach to be adopted highly depends upon various factors such as accuracy, variety of data, expense, size and complexity of the system.

In [4], the focus is made to analyze the movements of hands of conductor. Multiple wiimotes and 3D acceleration are used to explain the movements. Another method for tracking 3D position [13] is applied; data segmentation, three analysis modules and sonification method. A similar approach has been followed in [13] where two receivers were used to track the 3D position of infrared baton in conductor's hand. As compared to single handedly approach; multi-modal approaches have also been followed. In [14] a data glove is used to track the position of hand movement and also camera is used to capture its orientation. Different approaches of segmentation like thresholding based, background subtraction based [15], temporal difference, skin color based are used. The work in [15] have the limitation that it is unable to overcome the lightning effect. In [16], a baton is used to automate the system and to generate artificial music. The baton motion is analysed by the applicability of accelerometer with six degree of freedom.

IV. PROPOSED METHODOLOGY

Gesture comprise a space of motion articulated by the body, face and/hands [6]. Hand gestures are the most meaningful, most expressive and very commonly used [12]. The types of gestures are hand gesture, head gesture and body gesture. Gesture recognition can be done either on an image or video. Due to the non-availability of the standard dataset, the experimental work is based on self-made video dataset shown in Fig. 1 which contains different patterns of orchestral conducting. We have considered total five different patterns like rebound, 2by2, 3by4, 4by4 and 6by8.

S.No.	Dataset	Source
1.		Frames taken from self made video, includes 5 patterns in it.
2.		Video from net for practice

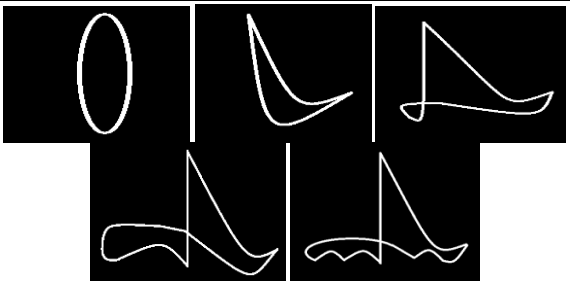
		
3.		Synthetic data, this also includes 5 patterns.

Fig. 1 Sample patterns of orchestral conducting in the generated dataset.

Our proposed methodology is divided into five stages: segmentation, tracking, feature extraction and classification. The each stage is explained below:

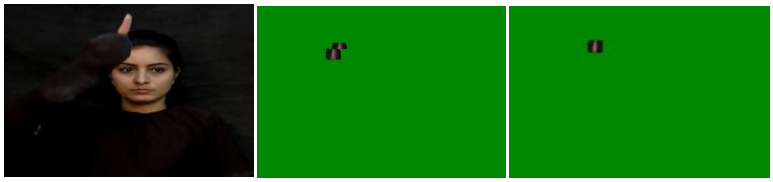
Segmentation: Segmentation is the process of dividing an image into the multiple segments. The aim of segmentation is to change the representation into some other form which is more meaningful and much easier to analyze. In this step the RGB color space is converted into YCbCr because in RGB (Red, green, blue) color space is additive in nature. The luminance and chrominance components are mixed in the RGB color space, which can't be good in case of color based segmentation. Because of this, we convert RGB color space into YCbCr color space according to the Equation in which Y represents luminance component whereas Cb and Cr represents the chrominance component.

$$Y=0.299R+0.587G+0.114B$$

$$Cr=R-Y$$

$$Cb=B-Y$$

In thresholding or color based segmentation technique, each pixel is allocated range of values in which the pixel lies. Based on the threshold applied, the foreground is separated in the form of conductor's hand or baton tip. The result of the segmentation is shown in Fig. 2 (b). The output after segmentation may not be proper due to occurrence of various small blobs because of improper segmentation. In order to improve the output of segmentation, we have applied morphological operation in the form of dilation and erosion. This improves the detection result which is shown in Fig. 2 (b).



(a) (b) (c)

Fig. 2 (a). Original Frame, (b). Segmented frame with some noise, (c). Segmented frame without noise after applying morphological operations.

The results of the segmentation (detection) on various conducting patterns are shown in the Fig. 3.

S.No.	Conducting Pattern	Result of conductor's hand segmentation(Results after detection)
01.	Rebound (Compound Meter)	
02.	2 by 2 (Simple Meter)	
03.	3 by 4 (Simple Meter)	
04.	4 by 4 (Simple Meter)	
05.	6 by 8 (Compound Meter)	

Fig. 3 Results of after segmenting the input video frames.

Tracking: For tracking conductor's hand gestures, Kalman filter is used which is a recursive, adaptive filter that is well known for its ability to track the object in a timely and accurate manner. The Kalman filter estimates the position of the conductor's hand in each frame of sequence. The input parameters of the Kalman filter are the position of the object in the image, the size of the object and the width and the length of the search window of the object. The variable parameters of the

Kalman filter are the state vector and the measurement vector. The equations for Kalman filter fall into two groups: time update equation and measurement update equation which are shown in the Fig. 4

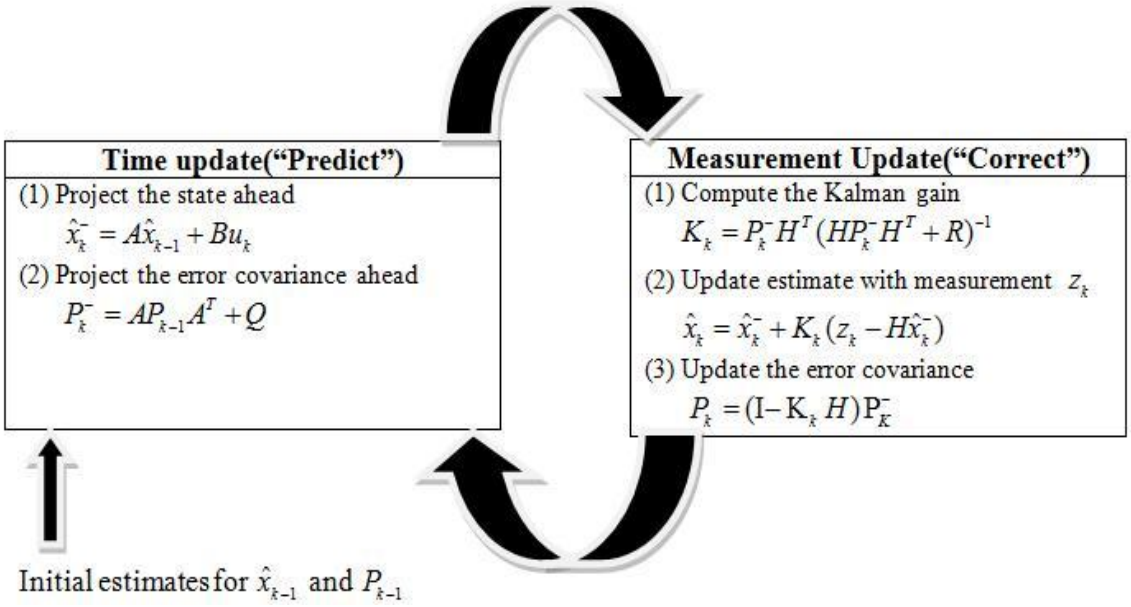
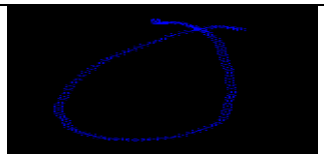
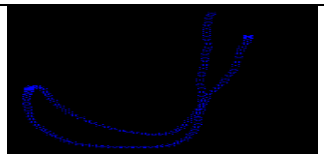


Fig. 4 A complete picture of the operation of the Kalman filter

The result of Kalman filter are shown in Fig. 5 which shows the bounding box tracked as the conductor's hand movement takes place. The tracking result in the form of tracked coordinates (Position x, Position y) in each frame is stored.

S.No	Conducting Pattern	Result of tracking using Kalman filter
01.	Rebound (Compound Meter)	
02.	2 by 2 (Simple Meter)	

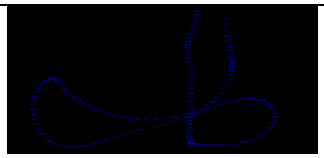
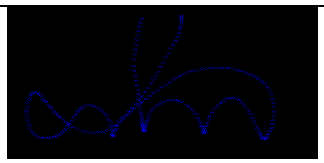
03.	3 by 4 (Simple Meter)	
04.	4 by 4 (Simple Meter)	
05.	6 by 8 (Compound Meter)	

Fig. 5 Results after tracking using Kalman filter.

Feature Vector Formation: In order to train the classifier, there is the need to provide feature vector as an input. As the tracking gives 2 coordinates (x coordinate, y coordinate) in each frame of video, the feature vector from the tracking result in 2 * n dimensional vector where n represents the number of frames in the video. In this proposed work, the feature vector formed is single dimensional. If there are n frames in the video, then the x coordinate is stored in single dimensional vector having range from 1 to n. Further, y coordinate is stored having range from n+1 to 2*n. This will result in single dimensional feature vector having length 2*n if there are n frames to be processed in the video.

Classification: Classification is the process of identifying to which set of categories the observation belongs to. Classification means to categorize something according to shared characteristics. An algorithm that implements classification is known as a classifier.“Classifier” refers to the mathematical function, implemented by a classification algorithm that maps input to a category. Though there are several classifiers like SVM [10], Hidden Markov Model (HMM), Nearest Neighbor classifier, BPNN (Back-Propagation Neural Network) [3], etc in the literature, but the choice of classifier highly depends upon the type of data type, number of input parameters required to be tuned, training time, etc. In this proposed work, SVM is used as a classifier because of its simplicity and having property to classify multiple classes precisely.

V. EXPERIMENTAL RESULTS

In our proposed work, we have taken five orchestral patterns for testing. The proposed work shows 100% accuracy of recognition. Also, music is generated according to the orchestral pattern formed by the orchestra conductor's hand or baton. As it is the initial work in this direction of research, only drum beats are taken into account.

VI. CONCLUSION

In this paper, automation of orchestra conductor's hand movement has been proposed in order to generate music which will provide real-time virtual orchestra experience. Effort is being made to design an intelligent system which is capable of recognizing the patterns which are used by the conductors. First of segmentation of the conductor's hand is done in order to separate the conductor's hand or baton. After this, morphological operation is performed in order to improve the results of segmentation and to overcome smaller blobs. Tracking is done using Kalman filter. The feature vector is formed from the tracking result which takes both x and y coordinates in each frame of the video sequence. SVM is used as a classier and music is generated according to the orchestral conductor pattern formed by orchestra conductor's hand. The proposed work shows promising results in order to provide real time orchestra experience without the help of human musicians. In future, work can be extended to add more musical instruments and to take various musical characteristics such as phrasing, nuances and dynamics into account.

REFERENCES

[1]. Tapio Lokki, Lauri Savioja et.al, "Virtual concerts in virtual spaces-in real time", Laboratory of Acoustics & Audio signal processing, Helsinki University of technology.

[2]. Markter Maat, Rob M. Ebbers, Dennis Reidma & Anton Nijholt, "Beyond the Beat: Modelling intentions in Virtual conductor", Human Media Interaction, University of Twente, Netherland.

[3]. Tommy Ilmonen & Tapio Takala, "Conductor following with Artificial Neural Networks", ICMC proceedings, 1999.

[4]. David Bradshaw & Kia Ng, "Analysing conductors gestures with the wiimote", EVA London Conference, 22-24 July, 2008.

[5]. David Bradshaw & Kia Ng, "Tracking conductors hand movements using multiple wiimotes", International conference on automated solutions for cross media content and multi-channel distribution", IEEE, 2008.

[6]. Sushmita Mitra & Tinku Acharya, "Gesture Recognition: A Survey", IEEE, 2007.

[7]. Declan Murphy, Tue haste Anddersen and Kristoffer Jensen, 2003. "Conducting audio files Via Computer Vision".

[8]. Brock McElheran, "Conducting Technique for beginners and Professionals", Oxford University Press, USA 1989. ISBN 978-0193858305.

[9]. Dennis Reidsma, Anton Nijholt and Pieter Bos, "Temporal Interaction between an Artificial Orchestra Conductor and Human Musicians".

[10]. Nasser H. dardas and Nicolas D. Georganas.2011, "Real-Time Hand Gesture Detection, Recognition Using Bag-of-features and Support Vector Machine Techniques", Vol 60, No-11.

[11]. Eric Lee, Teresa Marrian Nakra, Jan Borchers, "You're the conductor: A realistic interactive conducting system for children", RWTH Aachen University, Germany.

[12]. Pragati Garg, Naveen Aggarwal, Sanjeev Sofat, "Vision based hand gesture recognition", World Academy of Science, Engineering & Technology, Vol: 3, 2009.

[13]. F. Tobey, I. Fujinaga, "Extraction of Conducting Gestures in 3D Space", Proceedings of the International Computer Music, 1996.

[14]. H. Morita, S. Hashimoto, S. Ohteru, "A Computer Music System that Follows a Human Conductor", Computer vol.24no.7, 1991, pp. 44-53.

[15]. Teresa Marrian Nakra, Daniel Tilden, "Improving upon musical analyses of conducting gestures using computer vision", College of New Jersey.

[16]. Tammi Ilmanen, Janne Jalkanen, "Accelerometer based motion tracking for orchestra conductor following", Virtual environment, Springer-Verlag/wien 2000.