

Survey on Music Conducting Gestures using Dynamic Time Warping

Ms. Snehal Kshirsagar.

Department of E&TC Engg, MITCOE, Kothrud, PUNE, Savitribai Phule Pune University, India

Prof. M.N. Annadate.

Department of E&TC Engg, MITCOE, Kothrud, PUNE, Savitribai Phule Pune University, India

Abstract- Automatic recognition of musical patterns plays a crucial part in Musicological research and can become an indispensable tool for the search and comparison of music extracts within a large multimedia database. In today's life musical band performer requires conductor for the musical performance. Based on this study we present an efficient way for recognition isolated musical patterns in a monophonic environment, using of Dynamic Time Warping a probabilistic framework.

In this automatic recognize hand gestures pattern for the musical performance by using the RGB-D Camera. Each pattern, to be recognized, is converted into a sequence of frequency jumps by means of a Dynamic Time Warping. The important Characteristic of Context Dependent Dynamic Time Warping is that it exploits the correlation without using local point. Other is context of Traditional Music, which exhibits certain characteristics that make the classification task easy and accuracy is evaluated over the existing approaches.

Index Terms— Distance learning, dynamic time warping (DTW), meter-mimicking, musical conducting, pattern recognition

I. INTRODUCTION

Gesture recognition is a process of capture standing or classifying meaningful moment like humans hand, arm, face and head. In this process detect the moment of speedily and find out meaning full moment. The inspiration of technology in microelectronics has inspired research in the field of accelerometer-based gesture recognition. Three-axis accelerometers are being increasingly added into many personal electronic devices like the Apple iPhone, Apple iPod touch, Apple iPad, wiimote, and Lenovo laptops, to name a few.

The musician is performed the some moment of hand and arm which every moment are meaningful in musical field. Conductors have own set of gestures to represent their particular style. However, it is their responsibility to make understood by the musicians who perform each particular musical work through their hand movement. The conductor is drawing the pattern of hands moment and arm moment in 3D space around the pattern. This moment of pattern is periodic which grouped into different sets. Related to music in gesture field has three important pattern.

- Duple.
- Triple.
- Quadruple.

Duple: The duple is simplest way of pattern which has horizontal or vertical lines with two moments. In conductor draw the duple pattern which recognition in easily format. The process of duple format has only two moves allowed in any direction which related $\frac{2}{2}$, $\frac{2}{4}$ and $\frac{2}{8}$ format.

Triple: The second part of conductor is triple which have three moment in musician conductor. The its is a complex way of to recognition of sign. Which have used to $\frac{3}{2}$, $\frac{3}{4}$ and $\frac{3}{8}$ format.

Quadruple: The quadruple has $\frac{4}{2}$, $\frac{4}{4}$ and $\frac{4}{8}$ Format which has four moment sign in vertical and horizontal side. It is a more complex to detect musical sign.

Time signature is metric pattern which has characteristic path drawn by moment of hand and arms of musician. This path is shown in respectively duple, triple and quadruple format. We have developed a system for musical student teach in music class. Which student is practiced in home that time system detecting moment of student and find out moment is correct or not. It is worth noting that the focus of this paper is on gesture patterns used in training sessions for beginners. In this case, the evaluation considers the ability to reproduce basic patterns, and ability to keep time under control. The complex gestures including expressiveness and artistic movements are out of the scope of this work.

To capture moment used to RGB-D camera track the 3D hand position. Divide the image as a time section and find out the related duple, triple or quadruple. The proposed system Dynamic time warping to match timing and remove background and outline of system.

II. SYSTEM OVERVIEW

1. Previous Work

In the computer graphics literature, time warping of motion capture data has been a key component in many animation systems [4]. To account for the variations of human motion performed by different subjects, one popular strategy is to augment DTW with certain regression models. The human motion is articulated, and capturing such highly articulated motion from monocular video sensors is a difficult task. This difficulty limits the performance of video-based human action recognition, as indicated in the studies in the past decade [6].

[5] Accelerometer based gesture recognition has been discussed in many publications. Most of these literatures involve statistical methods and frequency domain approaches hence resulting in floating point and gesture recognition are uWave and FOSVM. uWave requires only one gesture to start with and employs feedback dependent template adaptation.

The most recent gesture recognition system that is solely accelerometer-based is the uWave [4]. An uWave is a user-site system that supports personalized gesture recognition. uWave functions by utilizing only one training sample, which stored in a template, for each gesture pattern. The core of the uWave is DTW and the system's database undergoes two types of adaptation: positive and negative adaptation [3].

These works are generally principally oriented towards recognition tasks. Determining the time progression of a gesture during a performance is generally not explicitly covered with the exception of score. Several authors have proposed systems based on HMM (or Semi-Markov Models). Nevertheless, in such cases, the Markov structure is essentially built from a symbolic representation given by the musical score, and not from continuous gesture data [8].

2. Literature survey:

Jeen-Shing Wang, Member, IEEE, and Fang-Chen Chuang [2] Users can use the pen to write digits or make hand gestures, and the accelerations of hand motions measured by the accelerometer are transmitted to a computer for online trajectory recognition through wireless. The algorithm first extracts the time and frequency domain features from the

acceleration signals. The kernel-based class is most important features by a hybrid method. This is separability for selecting significant features and linear discriminant analysis for reducing the dimension of features.

The problem of gesture recognition using the theory of random projection by formulating the whole recognition problem as a minimization problem. The gesture recognition system operates a single 3-axis accelerometer data and comprises two main stages: a training stage and a testing stage. For training, the system employs DTW as well as affinity propagation to create model for each gesture the system projects of all candidate traces and also the unknown trace onto the same lower dimensional subspace for recognition for instant. Ahmad Akl, Student Member, IEEE, Chen Feng, Student Member, IEEE, and Shahrokh Valaee, Senior Member, IEEE [3]

GTW has shown promising preliminary results, there are still unresolved issues. First, the GaussNewton algorithm for time warping converges poorly in the area where the objective function J_{gtw} is non-smooth. Second, generalized time warping is subject to local minima. A well known strategy to escape from local minima in image alignment has been to adopt a coarse-to-fine approach for optimizing GTW at different temporal scales. Third, although the experiments show admissible time warping results with fixed bases, it is more desirable to automatically learn the monotonic bases. a technique for temporally aligning multiple multi-modal sequences. The GTW algorithm offers a more flexible and efficient framework than the state-of-art DTW algorithms because parameterize the time warping function as a linear combination of monotonic bases. Feng Zhou Fernando De la Torre [4].

Hand gesture recognition is becoming increasingly popular for applications in computing environment. Accelerometer based methods have proven itself to be competitive in terms of both recognition and portability accuracy. But there are only a few algorithms which have generated accuracies in user independent gesture recognition. A novel accelerometer based user independent hand gesture recognition Shah Muhammed Abid Hussain and A. B. M. Harun-ur Rashid [5].

The depth maps captured by the depth cameras are very noisy and the 3D positions of the tracked joints may be completely wrong if serious occlusions appear, which increases the intraclass variations in the actions. In this paper, an action ensemble model to represent each action and to capture the intra class variance. In addition, novel

features that are easy for depth data are proposed. They are difficult to noise, invariant to translational and temporal misalignments, and capable of characterizing both the human-object interactions and human motion. Jiang Wang¹ Zicheng Liu² Ying Wu¹ Junsong Yuan [6].

The main component of this combined system is an HMM-based recognizer which considers contextual and dynamic information for a better modeling of writing units. A state-tying process based on decision tree clustering. Decision trees are built according to a set of expert-based questions on how characters are written and questions are divided into precise questions, yielding smaller ones and global questions, yielding larger clusters. Such clustering enables us to reduce the total number of models and Gaussians densities by 10 Anne-Laure Bianne-Bernard, Fare`s Menasri, Rami Al-Hajj Mohamad [7].

Frédéric Bevilacqua, Bruno Zamborlin, Anthony Sypniewski, Norbert Schnell [8] The system outputs continuously parameters relative to the gesture time progression and continuously. These parameters are comparing the performed gesture which stored reference gestures. The method relies on a detailed modeling of multidimensional temporal curves which Compared to standard HMM systems, the learning procedure are simplified using prior knowledge allowing the system to use a single example for each class.

Teresa Marrin Nakra Yuri Ivanov Paris Smaragdis Chris Ault [9] The UBS Virtual Maestro to simulate the experience of orchestral conducting for the general public attending a classical music program. The system applies the Wii Remote, which users hold and move like a conducting baton to affect the dynamics and tempo of an orchestral video/audio recording. The accelerometer data from the Wii Remote is used to control playback speed and volume in real-time.

Luke Dahl [10], a study where participants perform “air drumming” gestures in time to rhythmic sounds. These movements are recorded, and the timing of various movement features with respect to the onset of audio events is analyzed. A novel algorithm for detecting sudden changes in direction is used to find the end of the strike gesture. We find that these occur on average after the audio onset and that this timing varies with the tempo of the movement.

Luke Dahl [11], To detecting physical movement for recognition of musical sign we have performed this. But

avoid this techniques used to air gesture. It has designed to detect sound of air, like we perform musical sign instrument then detect sound of air which generated by movement.

Aggelos Pikrakis [12], An efficient method to designed using dynamic time warping algorithm is converted sequence of frequency jumps by fundamental tracking algorithm. A recognition occurred in 95% in system.

II. REVIEW OF LITERATURE

No	Paper name/ Author	Method	Advantages	Disadvantages
1	J.-S. Wang and F.-C. Chuang, “An accelerometer-based digital pen with a trajectory recognition algorithm for handwritten digit and gesture recognition,” <i>IEEE Trans. Ind. Electron.</i> , vol. 59, no. 7, pp. 2998–3007, Jul. 2012.	Accelerometer, gesture, handwritten recognition linear discriminant analysis (LDA), probabilistic neural network (PNN).	accelerometer-based digital pen for handwritten digit and gesture trajectory recognition applications	Limitation of the proposed trajectory recognition algorithm is that it can only recognize a letter or a number finished with a single stroke
2	M. Mandanici and S. Sapir, “Disembodied voices: A Kinect virtual choir conductor,” in <i>Proc. 9th Sound Music Comput. Conf.</i> , 2012, pp. 271–276.	1. Handling different compositional algorithms 2. Exploring different interaction metaphors	gesture-based musical performance	complex and structured musical materials.
3	A. Akl, C. Feng, and S. Valaee, “A novel accelerometer-based	Affinity propagation, compressive sensing,	Gesturerecognition using the theory of random projection	Recognition efficiency, being user-dependent limits the applications of uWave.

	gesture recognition system," <i>IEEE Trans. Signal Process.</i> , vol. 59, no. 12, pp. 6197–6205, Dec. 2011.	dynamic time warping, gesture recognition, random projection (RP).	(RP) and by formulating the whole recognition	
4	F. Zhou, F. D. la, and T. Frade, "Generalized time warping for multimodal alignment of human motion," in <i>Proc. IEEE Conf. Comput. Vis. Pattern Recog.</i> , Jun. 2012, pp. 1282 –1289 .	Tele rehabilitation, multi-modal temporal alignment	generalized time warping (GTW), an extension of dynamic time warping (DTW) for temporally aligning multi-modal sequences from multiple subjects	Problem of aligning two sequences, and it is unclear how to extend it to the alignment of multiple sequences
5	S. Hussain and A. Rashid, "User independent hand gesture recognition by accelerated DTW," in <i>Proc. Int. Conf. Informat., Electron. Vis.</i> , 2012, pp. 1033–1037.	Gesture recognition; dynamic time warping; user independent gesture; accelerometer; hardware acceleration	A high recognition accuracy and possibility of simpler implementation also gives this algorithm an upper hand among competitive methods	Limited its use in matching sequences of same duration.
6	J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in <i>Proc. IEEE Conf.</i>	Gesture recognition; dynamic time warping; user independent gesture; accelerometer; hardware acceleration	A high recognition accuracy and possibility of simpler implementation also gives this algorithm an upper hand	Limited its use in matching sequences of same duration.

	<i>Comput. Vis. Pattern Recog.</i> , Jun. 2012, pp. 1290–1297.	on	among competitive methods	
7	A.-L. Bianne-Bernard, F. Menasri, R.-H. Mohamad, C. Mokbel, C. Kermorvant, and L. Likforman-Sulem, "Dynamic and contextual information in HMM modeling for handwritten word recognition," <i>IEEE Trans. Pattern Anal. Mach. Intell.</i> , vol. 33, no. 10, pp. 2066–2080, Oct. 2011.	translational and temporal misalignments	Robust to noise, invariant to translational and temporal misalignments, and capable of characterizing both human motion and the human-object interactions	Limits the performance of video-based human action recognition
8	F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana, "Continuous realtime gesture following and recognition," in <i>Proc. 8th Int. Conf. Gesture Embodied Commun. Human Comput. Interaction</i> ,	Latin and Arabic handwriting recognition, context-dependent HMMs, neural-network combination.	building an efficient word recognition system resulting from the combination of three handwriting recognizers	Limited to capturing a close neighborhood (4 to 8 pixels) and the resulting observation probability densities

	2010, pp. 73-84.			
9	T. M. Nakra, Y. Ivanov, P. Smaragdis, and C. Ault, "The UBS virtual maestro: An interactive conducting system," in <i>Proc. Int. Conf. New Interfaces Musical Expression</i> , 2009, pp. 250-255.	conducting, gesture, interactive installations, Wii Remote	study of emotional expression as it is conveyed through conducting technique	Limitation in delays tracking in Wii data set.
10	N. Balakrishnan and D. Mitra, "Likelihood inference based on left truncated and right censored data from a gamma distribution," <i>IEEE Trans. Rel.</i> , vol. 62, no. 3, pp. 679-688, Sep. 2013.	Asymptotic variance-covariance matrix, coverage probability, expectation maximization algorithm, gamma distribution	Asymptotic variance-covariance matrix of the MLEs under the EM framework	Discrimination within this family by using likelihood-ratio and information

3. Conclusion

We analyze efficient scheme for the recognition of isolated musical patterns was presented. The scheme is based on the DTW novel extension and the performance of meter-mimicking patterns by the analysis of hand gestures evaluation. This system is used for the aid student for practicing the music, designed for the recognition of multivariate temporal musical gestures. By using camera gestures are captured and proposed framework used for the correctly classify a set of multivariate of the input data.

REFERENCES

1. M. Mandanici and S. Sapir, "Disembodied voices: A Kinect virtual choir conductor," in *Proc. 9th Sound Music Comput. Conf.*, 2012, pp. 271-276.

2. J.-S. Wang and F.-C. Chuang, "An accelerometer-based digital pen with a trajectory recognition algorithm for handwritten digit and gesture recognition," *IEEE Trans. Ind. Electron.*, vol. 59, no. 7, pp. 2998-3007, Jul. 2012.

3. A Akl, C. Feng, and S. Valaee, "A novel accelerometer-based gesture recognition system," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6197-6205, Dec. 2011.

4. F. Zhou, F. D. la, and T. Frade, "Generalized time warping for multimodal alignment of human motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1282-1289.

5. S. Hussain and A. Rashid, "User independent hand gesture recognition by accelerated DTW," in *Proc. Int. Conf. Informat., Electron. Vis.* 2012, pp. 1033-1037.

6. J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1290-1297.

7. A.-L. Bianne-Bernard, F. Menasri, R.-H. Mohamad, C. Mokbel, C. Kermorvant, and L. Likforman-Sulem, "Dynamic and contextual information in HMM modeling for handwritten word recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 2066-2080, Oct. 2011.

8. F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana, "Continuous realtime gesture following and recognition," in *Proc. 8th Int. Conf. Gesture Embodied Commun. HumanComput. Interaction*, 2010, pp. 73-84.

9. T. M. Nakra, Y. Ivanov, P. Smaragdis, and C. Ault, "The UBS virtual maestro: An interactive conducting system," in *Proc. Int. Conf. New Interfaces Musical Expression*, 2009, pp. 250-255.

10. L. Dahl, "Triggering sounds from discrete air gestures: What movement feature has the best timing?," in *Proc. Int. Conf. New Interfaces Musical Expression*, London, U.K., Jun. 2014, pp. 201-206.

11. L. Dahl, "Triggering sounds from discrete air gestures: What movement feature has the best timing?," in *Proc. Int. Conf. New Interfaces Musical Expression*, London, U.K., Jun. 2014, pp. 201-206.

12. aggelos pikrakis, sergios theodoridis, dimitris kamarotos, "recognition of isolated musical patterns using context dependent dynamic time warping", *iee transactions on speech and audio processing*, vol. 11, no. 3, may 2003