

3D MOTION DATA RECOGNITION
AND ITS APPLICATION ON
INTERACTIVE DANCING GAME

DENG LIQUN

DOCTOR OF PHILOSOPHY
CITY UNIVERSITY OF HONG KONG
SEPTEMBER 2012

CITY UNIVERSITY OF HONG KONG
香港城市大學

**3D Motion Data Recognition and Its Application on
Interactive Dancing Game**
三維人體動作識別及其在交互舞蹈遊戲上
的應用

submitted to
Department of Computer Science
電腦科學系
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
哲學博士學位

by

Liqun Deng
鄧利群

September 2012
二零一二年九月

Abstract

Human motion capture (mocap) has gained an increasing attention in many applications, such as advanced human-machine interaction, computer animations, digital films, interactive games etc. A motion capture system is able to record and digitalize a sequence of human postures representing by the 3D coordinates of a set of body joints across time. However, as the motions are usually captured continuously, it is a time- and labor-consuming work to segment and label the data manually. On the other hand, in some collaborative applications, such as human-machine interaction, it is also required that the input motions should be learnt automatically and in real time, which in turn can drive the computer with corresponding react. Hence, efficient methods to recognize motion data are necessary. In this thesis, we develop new methods to address the mocap data recognition problem, which includes three sub-problems, i.e., *isolated motion patterns recognition*, *sequential motion pattern recognition* and *real-time motion stream recognition*.

For *isolated motion patterns recognition*, the challenges are mainly caused by the high dimensionality and great variation of the data. Principle component analysis (PCA) is an efficient tool to reduce the dimension and extract features. However, it cannot retain the temporal information of the data points in samples when applied to time series data such as mocap data. Motivated by this, we propose two singular value decomposition (SVD) based methods named *segmental SVD* (SegSVD) and *bidirectional segmental SVD* (Bi-SegSVD). They first segment the motion data into a certain number of sub-segments, and then process them with SVD in an accumulative manner along the forward direction for SegSVD and both forward and backward directions for

Bi-SegSVD. Based on the segmental features, we calculate the similarity of two samples using a weighted dynamic time warping (DTW) based measure. The measure is further extended into a kernel function for support vector machine (SVM) classifier to classify the motion patterns.

In *sequential motion pattern recognition*, an input motion is composed of multiple motion patterns with their categories and boundaries unknown in advance. Thus, an additional challenge, i.e., to detect the start and end points of the embedded patterns, is imposed on this task. To address this problem, two new approaches are proposed. First, we exploit an *open-end DTW* (OE-DTW) based scheme motivated by the fact that OE-DTW is efficient in matching complete patterns with incomplete ones. By regarding the input motion as a complete pattern, and taking each of the template patterns as an incomplete pattern, we apply OE-DTW to find their optimal matched part, according to which, the embedded patterns are detected and recognized sequentially. Second, we take advantage of the SegSVD structure that is with multiple levels and detect the end points of the embedded patterns by referring to top levels of the template patterns using a new penalty based level matching scheme.

In *real-time motion stream recognition*, it requires not only to identify and recognize the embedded patterns in input motions, but also to detect the unwanted motions, and the task should be finished in real time. Motivated by the fast speed and efficiency of the content-based indexing techniques, we introduce an body partition index map based approach for this problem. Noting that human motions are composed of the sub-motions of upper limbs, legs and torso, we partition the motions into five parts according to the five body partitions, and process the submotions with standard clustering techniques separately. A generalized model for each motion class is trained by integrating the projected cluster node strings of the training trials and five body partition index maps are constructed. During recognition, the input frames are projected into the clusters and then used to look up the index maps. With a flexible voting scheme and a set of end point detection conditions, the input motions are segmented and recognized as legal patterns or unwanted motions in real time.

Finally, we apply the real-time recognition approach and develop an *interactive dancing game system*, in which users' dance motions are lively captured and recognized, and according to the recognition result, the corresponding interactive motions are determined and used to drive the avatar's animation. Hence, it provides an immersive environment that users can dance with the avatars (i.e. virtual partners) collaboratively.

Keywords: motion capture, motion pattern classification, segmentation and recognition, real-time recognition, interactive dancing game, singular value decomposition, dynamic time warping, index map

Contents

Abstract	i
Acknowledgement	iv
1 Introduction	1
1.1 Human Motion Capture	1
1.1.1 Existing Mocap Systems	2
1.1.2 Mocap Data Representation	3
1.2 Mocap Data Recognition	5
1.3 Related Work	7
1.3.1 Feature generation	8
1.3.2 Motion Segmentation	10
1.3.3 Recognition Techniques	11
1.3.4 Content-based Motion indexing	13
1.3.5 Mocap based Dance Applications	14
1.4 Summary of Contents and Contributions	14
1.5 Outline of Dissertation	17
2 Isolated Motion Pattern Recognition with Segmental SVD Features	19
2.1 Introduction	20
2.2 Segmental SVD Features	24
2.3 Constructing Kernel Function	27
2.3.1 Weighted similarity measure	27

Contents	II
2.3.2 Composite kernel function	31
2.4 Performance Evaluation	31
2.4.1 Dataset description	32
2.4.2 Existing methods used for comparison	34
2.4.3 Classification performance	36
2.5 Summary	44
3 Sequential Motion Pattern Recognition with Two Novel Methods	46
3.1 Introduction	46
3.2 Backgrounds	48
3.2.1 DTW and OE-DTW	48
3.2.2 Normalization and Global Constraint on DTW	50
3.3 The OE-DTW based Approach	51
3.3.1 <i>K-Repetition</i> Global Constraint	51
3.3.2 Flexible Endpoint Detection	52
3.3.3 Steps of Recognizing Sequential Patterns	53
3.4 The SegSVD based Approach	55
3.4.1 Continuous SegSVD	55
3.4.2 Similarity Measure for Continuous SegSVD	55
3.4.3 Segmentation & Recognition of Sequential Patterns	58
3.5 Performance Evaluation	59
3.5.1 Data Acquisition	59
3.5.2 Experimental Results	60
3.6 Summary	64
4 Real-Time Motion Recognition with Body Partition Index Maps	66
4.1 Introduction	66
4.2 Generalized Model and Index Maps	69
4.2.1 Motion Preprocessing	69
4.2.2 Training GMs	70

Contents	III
4.2.3 Constructing Index Maps	71
4.3 Recognition with GMs and Index Maps	72
4.3.1 Isolated Motion Recognition with GMs	72
4.3.2 Continuous Motion Recognition with Index Maps	74
4.4 Experiments and Results	78
4.4.1 Data Preparation	78
4.4.2 Results on Isolated Motion Recognition by GMs	79
4.4.3 Results on Continuous Motion Recognition with Index Maps .	82
4.5 Summary	87
5 Application: An Interactive Dancing Game System	88
5.1 Introduction	88
5.2 Determining Interactive Dance Motion	90
5.3 Interactive Dancing Game	92
5.4 Performance Evaluation	94
5.5 Summary	95
6 Conclusion and Future Directions	97
List of Publications	100
Bibliography	102

List of Tables

2.1	Summary of the datasets used in experiments.	32
2.2	Comparison of classification accuracy (%) on AUSLAN Dataset. . . .	36
2.3	Comparison of classification accuracy (%) on Japanese vowels dataset.	39
2.4	Distribution of training and testing motions on HDM05 dataset. . . .	40
2.5	Comparison of classification accuracy (%) on HDM05 dataset.	40
3.1	The 19 basic moves of <i>A-go-go</i> dance used in our dataset. The average number of frames of each move pattern is shown inside the parentheses.	61
3.2	Parameters of the test motion sequences.	62
3.3	Recognition rate on different sets of templates.	64
3.4	Recognition rate on selected set of templates.	64
4.1	Description of DTW, LCS, SW and PDTW.	73
4.2	Recognition performance of GMs with DTW, LCS, SW and PDTW on both CityU and HDM05 dataset.	80
4.3	Performance comparison for continuous motion recognition.	84
4.4	Continuous recognition result of our proposed method for each motion class.	85
5.1	Questions for questionnaire and result.	95

List of Figures

1.1	Existing mocap systems. (a) Computer vision based system, (b) Optical mocap system, (c) Mechanical mocap system, (d) Magnetic mocap system, (e) Inertial mocap system, and (f) XBOX Kinect mocap system.	2
1.2	Optical mocap system in CityU. (a) The mocap area, (b) Capturing with markers on body, (c) A captured posture visualized by linking the markers.	4
1.3	(a) Human character model used in CityU mocap system and (b) the corresponding joint hierarchical structure for BVH format.	4
2.1	Framework flow of (a) universal subspace based (b) individual subspace based feature extractions approaches, respectively.	21
2.2	Two example 3D human motions. (a) Squat Down, (b) Stand Up. . . .	22
2.3	SegSVD structure of a MTS sample.	24
2.4	Bi-SegSVD structure of a MTS sample.	26
2.5	Similarity calculation of two MTS samples with SegSVD representations.	28
2.6	An example of aligning the levels by DTW.	29
2.7	Human character model used in HDM05 dataset.	34
2.8	Three-fold cross validation data sets on AUSLAN dataset.	35
2.9	Classification accuracy obtained under different β values (a) and number of levels (L) (b) on AUSLAN dataset under Set $\mathcal{C}$	37

2.10	Classification accuracy obtained under different β values (a) and number of levels (L) (b) on Japanese vowels dataset under Set $\mathcal{C}$	39
2.11	Classification accuracy obtained under different β values (a) and number of levels (L) (b) on HDM05 dataset under Set $\mathcal{C}$	42
3.1	An example of DTW alignment.	49
3.2	Two sets of weights for DTW calculation.	50
3.3	(a) A negative example of the alignment between two patterns without global constraint, (b) Sakoe-Chiba band, (c) Itakura Parallelogram, and (d) Ratanamahatana-Keogh band.	51
3.4	An example of alignment with K -Repetition ($k = 3$).	52
3.5	An example of OE-DTW. The solid line shows the result with original OE-DTW, while the dotted line shows the result with our proposed flexible end-point detection scheme with $\varepsilon = 0.02$	53
3.6	An example of recognizing the sequential patterns in a continuous stream.	54
3.7	Performance comparison of our methods and CDP on sequential pattern recognition.	62
3.8	Evaluating the performance of K -Repetition and determining value of k	63
4.1	Two repetitions of the same dance motion performed by two different players.	67
4.2	The overview of our proposed framework.	68
4.3	Five body partitions of a human skeletal model.	69
4.4	An example of training GM with two example motion trials.	71
4.5	An illustration of an index map.	72
4.6	The flowchart of real-time recognition of an input motion.	75
4.7	An example for illustrating the end point detection and segmentation (here $\kappa = 2$).	78
4.8	Recognition performance using GMs with whole body matching vs body partition scheme on isolated motions.	81

4.9	Comparison of classification performance for isolated motion recognition.	82
4.10	Confusion matrix for our approach on substitution errors in continuous motion recognition.	85
4.11	An example of recognizing a sequence with two legal moves and one unwanted move. The axis on the top represents the original frames and the x-axis represents the key frames (here $\kappa = 2$).	86
5.1	The overview of the system with block diagram (The shaded area illustrates the indexing map based recognition engine presented in the previous chapter).	90
5.2	An example of determining interactive move $MoveA'$	91
5.3	An exemplar scenario of the system in training mode.	92
5.4	An exemplar scenario of the system in freestyle mode.	93
5.5	Recognition accuracy on isolated partial motions.	94