Tai Chi Motion Recognition Using Wearable Sensors and Hidden Markov Model Method

Dennis Majoe[1], Lars Widmer[1],
Philip Tschiemer[1]and Jürg Gutknecht[1],

[1] Computer Systems Institute, ETH Zurich, Switzerland
{dennis.majoe@inf.ethz.ch}

Abstract. This paper reports ongoing work in which low cost sensors are combined with high level gesture recognition for automatic training and analysis of Tai Chi. Various feature extraction methods are developed for the sensors and Hidden Markov Modeling is used to classify Tai Chi movements.

Keywords: Motion Capture, Hidden Markov Model based gesture recognition, embedded system, wearable computer.

1 Introduction

Tai Chi Chuan is a low impact exercise comprising several gesture like movements. In order to develop an automatic training system one must be able to detect what the student is trying to do and then help improve upon this attempt by suitable advice and encouragement. Therefore a prime requirement for such a system is to be able to recognise the movements the student is doing.

2 Related Work

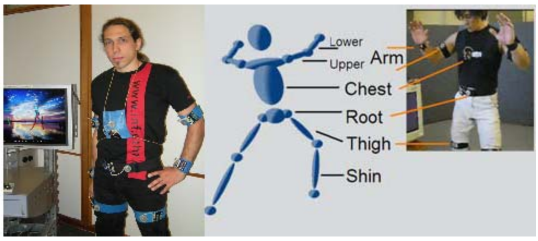

Fig. 1. On the left a student wears the 10 sensors and the computer displays his avatar in real time. The right shows the avatar bones model and relationship of sensor to limb.

Previous work in Tai Chi research [1],[2], and [3] has sensed the earth's gravity and the earth's magnetic field to accurately measure limb orientation wirelessly and ergonomically.

The current second generation sensors are smaller and work with a USB master allowing for higher data rates than before and higher accuracy.

2.1 Tai Chi Motion Recognition

Tai Chi is made up of a set of specific movements each running in a sequence. For example the most commonly taught simplified 24 movement (Yang) form, may be further deconstructed down to about 70 sequential sub movements each lasting about 2.5 seconds. When the sub movement is analysed it may be thought of as a complex human gesture that incorporates the hands, torso and legs moving in harmony.

The sensor technology used is intended to provide 3D generated avatars that may be rendered on a computer screen, so that the teacher and student may be seen either singly or superimposed. The avatar is created using forward kinematics [4] and provides 3D positional data for any part of the body. This data may be used for gesture recognition using the well known Hidden Markov Model approach [5].

In the review of alternative approaches [6] [7] [8] [9] we concluded that we must identify the best way to use limb data so that limb positions could be combined to account for correlated limb activity. We also need to constrain the encoding of data by explicit human body modeling and identifying clustering during limb moves. Finally we needed to use an ensemble approach to combine multiple classification viewpoints and study what combination of features provided the best classification results.

3 Feature Extraction

The HMM approach requires that a feature extraction layer generate abstract codes, an alphabet, that relates to the observation sequence. In full body gesture recognition one must choose wisely how the position of each body limb is included if one wants to keep the number of observation codes low.

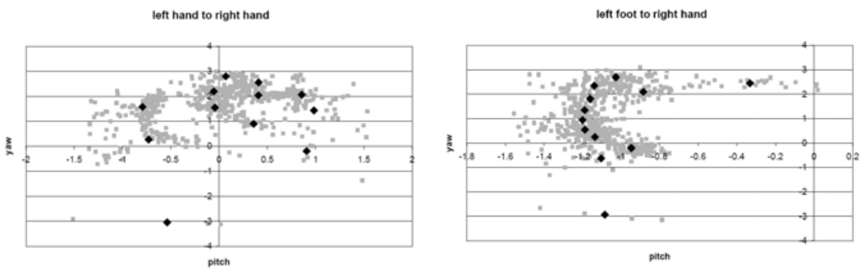

Fig. 2. The figure shows the k-means clustering in which 12 observation reference points are generated from all the available training data for the two dimensional feature methods A and B.

In order to ensure that a small number of codes represent the limb movement well the data points of the limb may be processed with a K-mean clustering algorithm. Figure 2, shows the 12 reference points that are generated from the overall mass of training data. These reference points are used to compare test data and the test data with the smallest linear distance to the reference point is assigned that point's assigned observation code. The clustering ensures that if the measurement occupies a certain region more than others on average then more codes are attributed to separate that space. If regions are sparsely populated, then fewer codes are used to specify that space.

When one must codify two limbs (or more) moving in some correlated action, then the number of codes rises rapidly with limb count. One way to solve this is to describe the interaction of the two limbs using a unique value. An example would be to take the angle between two arms, or two legs, or leg and arm, as a one dimensional feature, describing the two limbs by only M codes. Another method is to use two classifiers each trained to detect one limb but then combined in ensembles such that they must both give a positive result at the same time for a correct correlated movement..

To see how such methods could improve recognition a competitive framework was set up in which three different approaches were used to derive feature data each one becoming more complex in terms of dimensionality and calculation resources.

Each feature method should be used to drive identical HMM based machine trained classifiers with the same number of states and the same observation code ranges.

Each method would be evaluated based on how often the classifiers provided correct reporting as well as how often false reporting occurred across a number of different Tai Chi movements.

The methods are as follows. The first measures the one dimensional angle subtended by any two limb ends and the torso, as if the limbs form a scissors. The second method measures the two dimensional vector made by any two limb ends. Imagine one vector being the line drawn between your two wrists. The third method uses the 3D space positional data of any limb. This is pure 3D trajectory data.

Table 1. Different features generated using positional data from the hands and feet and torso.

Dimension	Feature Method A	Feature Method B	Feature Method C
1	Angle Left hand Right hand	Angle Left foot Right Foot	Angle Left hand right foot
2	Vector left hand right hand	Vector left hand right hand	Vector left hand right foot
3	x,y,z position of left hand	x,y,z position of right hand	x,y,z position of right foot

4 HMM applied to the Tai Chi data sets

Our approach was to train a 5 state HMM with each of the 9 different feature extraction methods. The HMMs would be trained on 5 different Tai Chi movements as shown in Figure 3. This means that we would generate 45 different classifiers with each classifier running with a particular feature extraction method and trained to detect one type of Tai Chi movement.

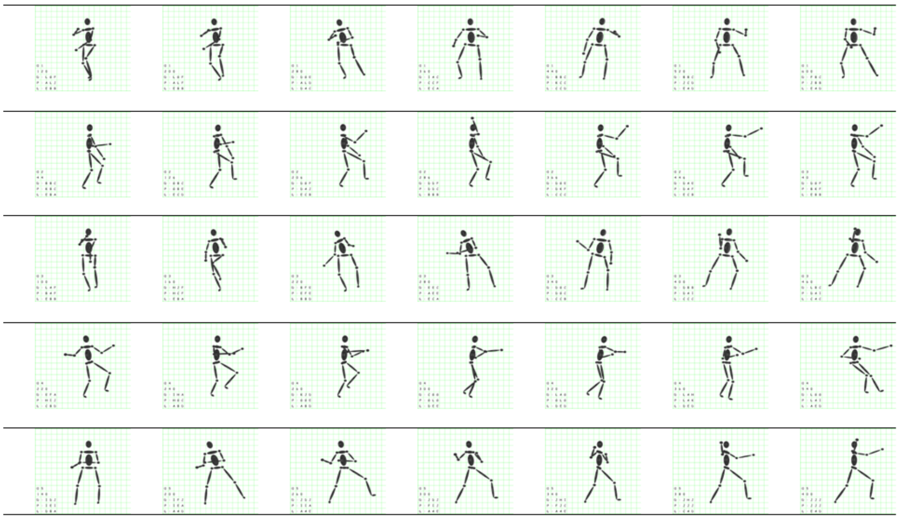

Fig 3. Five different Tai Chi sub movements shown as seven motion captured avatar screen shots. “Part mane”, “Spread Wings”, “Brush Knee”, “Repulse monkey”, “Single Whip”

The validation of the classifiers would be done by taking the stored probability matrices and using them to run the HMM forward algorithm and Viterbi algorithm on the test data that had been processed through each of the nine feature methods. The classifiers would be run using all the test data.

4.1 Results analysis

Figure 4 shows the forward algorithm classifier outputs for each of the five Tai Chi sub movements in each row. For column one, we see the different classifiers generating spikes when they report a classification.

The 1D feature extraction HMM classifiers generate a correct output about 80% of the time, however a large number of false positives, 12%, are obtained. One can see that in many cases if method A fails or false reports, methods B and C work correctly. So a majority voting technique could enhance this method. Finally in the third column the classifiers behave almost perfectly never making a false report and always generating a reasonable forward probability. In fact method B (see the 5 column spikes in the very middle of each group of 15) generates 99.7% recognition rates implying no other classifier is needed for majority voting. If this result holds for all future movements, it removes the need to double up on classifiers and this significantly reduces the computational requirements of a wearable with low processing power. For the second column the 2D one can see that there is a lot of false reporting by the “Spread Wings” classifier mistaking “Part Mane” as the same. However correct outputs are obtained 86% of the time.

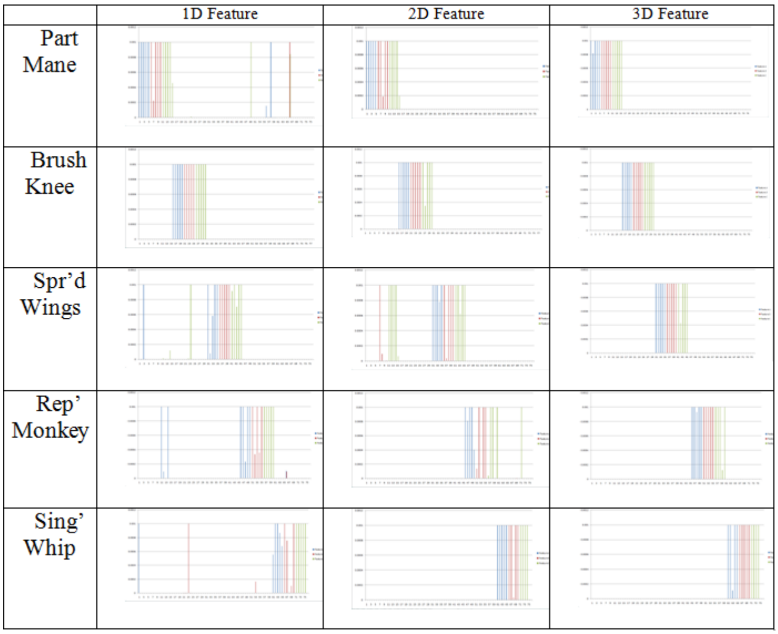

Fig. 4. Shows a graphical depiction of the confusion matrix for all classifiers and all data.

3D Feature	VITERBI STATE ESTIMATION	F'WD^2.
Part Mane A	0 0 0 1 2 2 2 2 2 2 2 2 4 4 4 4 4 4 4 4 4 4 4 4	1.99e-006
	0 0 0 0 1 2 2 2 2 2 2 2 2 2 2 2 2 4 4 4 4 4 4 4	1.00e-008
	0 0 0 1 1 2 2 2 2 2 2 2 4 4 4 4 4 4 4 4 4 4 4 4	9.95e-006
	0 0 0 0 1 1 2 2 2 2 2 4 4 4 4 4 4 4 4 4 4 4 4 4	4.73e-005
	0 0 0 0 0 1 1 2 2 2 2 2 2 2 2 4 4 4 4 4 4 4 4 4	2.48e-006
Part Mane B	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00
Part Mane C	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.00e+00

3D Feature	VITERBI STATE ESTIMATION	F'WD^3.
Spred Wings A	4 4 4 4 4 4 4 4 4 2 2 2 3 0 1 1 1 1 1 1 1 1 1 1	5.8e-004
	4 4 4 4 4 4 4 2 2 2 2 3 0 1 1 1 1 1 1 1 1 1 1 1	3.2e-004
	4 4 4 4 4 4 4 4 4 4 4 4 4 2 2 2 2 2 3 3 3 0 1 1	2.1e-008
	4 4 4 4 4 4 4 2 2 2 2 2 3 0 1 1 1 1 1 1 1 1 1 1	2.0e-005
	4 4 4 4 4 4 4 4 4 4 4 2 2 2 2 3 3 0 1 1 1 1 1 1	1.7e-005
Spred Wings B	4 4 4 4 4 4 4 4 4 4 2 0 0 0 0 0 0 0 0 0 0 0 0 0	0.0e+000
	4 4 4 4 4 4 4 4 4 4 2 0 1 1 0 0 0 0 0 0 0 0 0 0	0.0e+000
	4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 2 0 1 1 1 1	1.6e-247
	4 4 4 4 4 4 4 4 4 2 0 1 1 1 1 1 1 1 1 1 1 1 1 1	3.5e-235
	4 4 4 4 4 4 4 4 4 4 4 4 4 4 2 0 1 1 1 1 1 1 1 1	2.4e-243
Spred Wings C	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.0e+000
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.0e+000
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.0e+000
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.0e+000
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.0e+000

Fig. 5. Shows the Viterbi output listing the states for an accurate classification and a false one..

Figure 5 shows the Viterbi estimated states generated during the classification using the 3D feature extraction, for two sub movements, for limb relationship A, B and C. What one can see is that for the data list on the left those state lists associated with a high forward probability look very similar whereas low probabilities are associated with 0 only lists. This is due to the fact that transitions through that state list concur with the highest probabilities of emissions of similar observation codes.

In the data on the right however the state list associated with the “Spread Wings” classifier A lists states such as 4,2,2,2,3,0,1 while classifier B lists states 4,2,0,1, and does not list only zeroes as in table 4. The inference is that the maximum likelihood state sequence 4,2,3,0,1 must be achieved for a valid classification, whereas the missing state 3 in the sequence 4,2,0,1 heavily penalizes the outcome. Similarly in the case of “Part Mane” the sequence 0,1,2,4 must be traversed for a valid classification. By checking that the Viterbi traverses at very least the sequences 4,2,3,0,1 and 0,1,2,4, we can use the state list to classify the two different movements with high confidence.

5 Conclusions

The application of HMMs to Tai Chi has been presented here integrated with low cost motion capture and digital avatars lending it to future immersive environments. Future work must concentrate on the Dynamic Time Warping problem associated with this type of problem.

References

1. K. Kunze, M. Barry, E. Heinz, P. Lukowicz, D. Majoe , J. Gutknecht, "Towards Recognizing Tai Chi - An Initial Experiment UsingWearable Sensors", 3rd International Forum on Applied Wearable Computing, Mobile Research Center, Universität Bremen, Bremen, Germany, 2006.
2. D. Majoe, I. Kulka, and J. Gutknecht, “Qi energy flow visualisation using wearable computing,” Pervasive Computing and Applications, 2007. ICPCA 2007. 2nd International Conference on, pp. 285–290, 26-27 July 2007.
3. D. Majoe, M. Estermann, N. Ranieri, J. Gutknecht, “Ergonomic Low Cost Motion Capture for every day health exercise” Pervasive Computing and Applications, 2008. ICPCA 2008. Third International Conference Volume: 2, On page(s): 627-6326-8 Oct. 2008
4. Wikipedia, "Forward kinematic animation", http://en.wikipedia.org/wiki/Forward_kinematic_animation
5. L.R. Rabiner. "A tutorial on HMM and selected applications in speech recognition". In Proc. IEEE, Vol. 77, No. 2, pp. 257-286, Feb. 1989.
6. Kanav Kahol, Priyamvada Tripathi, Sethuraman Panchanathan, “ Computational Analysis of Mannerism Gestures.” Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3 - Pages: 946 – 949, 2004.
7. AF Bobick , AD. Wilson, "A state based technique for the summarization and recognition of gesture.", Proceedings 5th international conference on Computer Vision 1995.
8 M. Band, N. Oliver, A. Pentland, “Coupled hidden Markov models for complex action recognition.” Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997
9 HongLi Zhu, Peng Ying Du, Jian Xiang. “3D Motion Recognition based on Ensemble Learning”, Proceedings of the Eighth International Workshop on Image Analysis for Multimedia Interactive Services, 2007.