# Demo: Complex Human Gestures Encoding from Wearable Inertial Sensors for Activity Recognition

Mathias Ciliberto
University of Sussex
m.ciliberto@sussex.ac.uk

Luis Ponce Cuspinera
University of Sussex
l.ponce-cuspinera@sussex.ac.uk

Daniel Roggen
University of Sussex
daniel.roggen@ieee.org

## Abstract

We demonstrate a method to encode complex human gestures acquired from inertial sensors for activity recognition. Gestures are encoded as a stream of symbols which represent the change in orientation and displacement of the body limbs over time. The first novelty of this encoding is to enable the reuse previously developed single-channel template matching algorithms also when multiple sensors are used simultaneously. The second novelty is to encode changes in orientation of limbs which is important in some activities, such as sport analytics. We demonstrate the method using our custom inertial platform, *BlueSense*. Using a set of five BlueSense nodes, we implemented a motion tracking system that displays a 3D human model and shows in real-time the corresponding movement encoding.

## 1 Introduction

Inertial sensors and template matching algorithms have been used successfully for activity recognition in healthcare, well-being and sports applications [2]. Template matching algorithms can be embedded on low-power sensor nodes [6]. Nevertheless, they are generally designed to work with a single channel of data. In certain situation, such as in beach volleyball movements analytics, this can be a limitation as multiple sensors are required to be employed on different body parts in order to get and analyse the complexity of the movements. For this reason, using them can become challenging for complex gestures recognition.

Modern inertial platforms, such as XSens [5], Ethos [4] and our BlueSense, can provide orientation data, generally as quaternions. We present an encoding approach for complex gestures that elaborate the orientation data provided by several inertial sensors worn by the user. This method encodes the position and the orientation of the user's hand during a movement as a single stream of symbols, simplifying the application of single channel pattern matching algorithms. It is an extension of [7] with the novelty of including the rotation of the hand during the movements in the encoding. This will be important in future applications, such as sport analytics and specially in beach volleyball gesture recognition.

International Conference on Embedded Wireless Systems and Networks (EWSN) 2018
14–16 February, Madrid, Spain

ISBN: 978-0-9949886-2-1

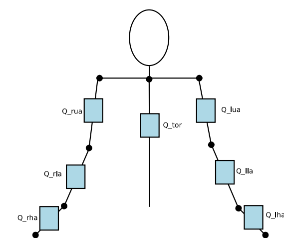
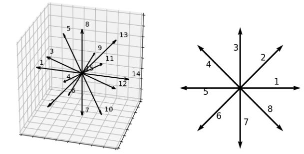



**Figure 1. Setup of BlueSense on the user's upper body.**



**(a)** *3D, 15 symbols* **(b)** *2D, 8 symbols*

**Figure 2. 3D and 2D codebooks examples. Different numbers of vectors can be used in each codebook, in order to reduce or increase the granularity of the displacement sampling.**

## 2 Gesture Encoding

The system described in [7] computes the position of the upper body joints using the 3D orientation of sensors placed on each limbs and the torso of the user, as displayed in Figure 1. Combining these positions, the algorithm finds the position of the hand in the 3D space. A gesture is then expressed as successive positions of the hand forming a trajectory. Then, the trajectory can be sampled at regular time intervals or after that a certain distance has been covered. For each sample, the vector difference between two contiguous positions is calculated. This vector is finally encoded to a symbol using a codebook: this is a set of 3D unit vectors equally distributed with respect to their direction (Figure 2). The symbol corresponding to the displacement vector is given by the closest codebook vector. The coded symbols are indexes in the codebook.

The method as described in [7] lacks of information about the rotation of the hand during the movement. Two different gestures can lead to the same displacement encoding, for ex-

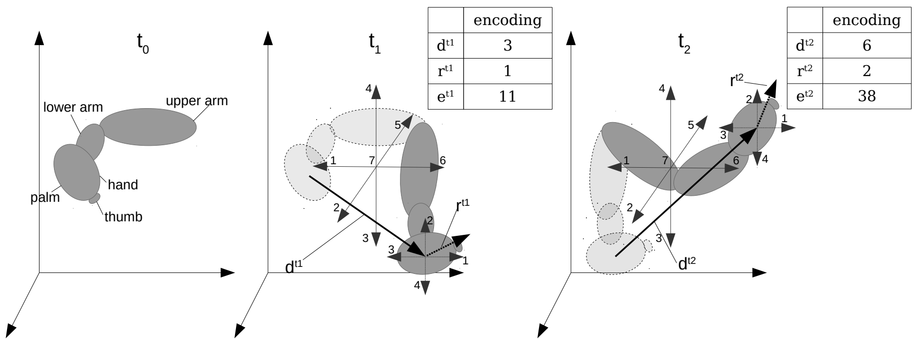



**Figure 3. Example of encoding of two movements of a right arm, from $t_0$ to $t_1$ and from $t_1$ to $t_2$. At every instant $t_i$, the displacement of the hand is encoded with the $d^{ti}$, the rotation of the hand is coded with $r^{ti}$ and the final encoding, computed using the Cantor pairing function, is represented by the symbol $e^{ti}$.**

ample if they have been performed once with the palm facing upwards and another time with the palm facing downwards. In order to overcome this issue, we introduced a second encoding for the rotation of the hand.

This extra encoding uses a 2D codebook (Figure 2) in order to represent the rotation of the hand. Defining a starting position encoding (for example the palm facing downwards parallel to the ground is encoded as *1*), it is possible to represent the rotation of the hand with the closest symbol in term of angular distance. The two symbols for position and rotation of the hand are eventually combined in order to obtain a single symbol. As natural numbers are used to index the codebooks vectors, this step is performed using a pairing function. An example, the Cantor pairing function, is presented in 1:

$$e(d,r) := \frac{1}{2}(d+r)(d+r+1)+r \tag{1}$$

where $d$ is the symbol for the displacement, $r$ is the symbol for the rotation of the hand and $e$ is the final encoding. The addition and product operators are the arithmetical operations of sum and multiplication.

An example of the whole system can be observed in Figure 3. A 7-symbols codebook for the displacement and a 4-symbols codebook for the rotation are used.

# 3 Demonstration

In order to visualize the data provided by BlueSense nodes, we created a motion tracking system which is presented in Figure 4. The system includes a set of 4 BlueSense sensor nodes and two programs: SensHub and the 3D human model. The former collects the orientation data through Bluetooth from the sensors, synchronizes them and forwards them to the 3D model as single line of text through a TCP connection. The latter is a 3D human model developed using the JMonkeyEngine 3D engine [1]. The TCP connection allows to place SensHub and the 3DHumanEngine on two different devices, potentially in two different locations.

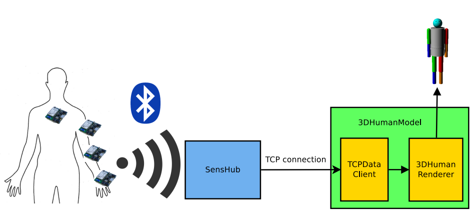



**Figure 4. Motion tracking system overview. The 3DHumanModel receives the data as a line of text through a TCP connection. Then it parses this text in order to extract the quaternions to animate the 3D model. At every rendering cycle, the 3D model is updated with the most recent orientation data.**

We were able to evaluate of the 3D model and the motion tracking system during the British Science Festival 2017 [3]. During the event, we deployed the system in a more complex simulation where people where asked to play a virtual beach volleyball game. We analysed the latency and the battery life of the sensors. The sensors are able to stream quaternions up to 500 Hz, but considering the framerate of the 3D rendering set to 60fps, we set the sample rate to 100 Hz. The latency is highly related to the hardware of the PC that runs the simulation. During the event it was acceptable for real time gaming. We also tested the battery life of the sensors that streamed the data for about 4 hours continuously.

In the future, we plan to employ the motion tracking system to support the training of beach volleyball players.

# 4 Acknowledgments

U.K. EPSRC First Grant EP/N007816/1

# 5 References

[1] Jmonkeyengine, 3d game engine. `http://jmonkeyengine.org/`, 2017. [Online; accessed 28-November-2017].

[2] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga. Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey. *Architecture of computing systems (ARCS), 2010 23rd international conference on*, pages 1–10, 2010.

[3] M. Ciliberto. ”How good are you in beach volleyball?”. `https://www.britishsciencefestival.org/event/how-good-are-you-at-beach-volleyball/`, 2017. [accessed 28-November-2017].

[4] H. Harms, O. Amft, R. Winkler, J. Schumm, M. Kusserow, and G. Tröster. Ethos: Miniature orientation sensor for wearable human motion analysis. In *Sensors, 2010 IEEE*, pages 1037–1042. IEEE, 2010.

[5] D. Roetenberg, H. Luinge, and P. Slycke. Xsens mvn: full 6dof human motion tracking using miniature inertial sensors. *Xsens Motion Technologies BV, Tech. Rep*, 2009.

[6] D. Roggen, L. P. Cuspinera, G. Pombo, F. Ali, and L.-V. Nguyen-Dinh. Limited-Memory Warping LCSS for Real-Time Low-Power Pattern Recognition in Wireless Nodes. *European Conference on Wireless Sensor Networks*, 8965:151–167, 2015.

[7] T. Stiefmeier, D. Roggen, and G. Tröster. Gestures are strings: efficient online gesture spotting and classification using string matching. *Proceedings of 2nd International Conference on Body Area Networks*, pages 16:1–8, 2007.