FEOGARM: A Framework to Evaluate and Optimize Gesture Acquisition and Recognition Methods

Simon Ruffieux[1], Elena Mugellini[1],Denis Lalanne[2], Omar Abou Khaled[1]

[1] *University of Applied Sciences Western Switzerland*
[2] *University of Fribourg, Switerzland*

***Abstract*—This paper discusses the general concept and implementation steps towards a framework facilitating the development and evaluation of gesture acquisition and recognition systems through a common benchmarking standard. In particular, a complete ground-truth of annotated gestures will be acquired with multiple high-resolution acquisition devices and as such will facilitate, using machine-learning techniques, the development of non-intrusive and light capture devices dedicated to gestures recognition. This article presents our concept, the current achievements and the first steps towards such a framework and corpus.**

Keywords-gesture recognition; open-source corpus; framework; algorithms evaluation and optimization

I. INTRODUCTION

A new era of Human-Computer Interaction (HCI) is emerging with the large variety of cheap and powerful sensors becoming available on the market. The algorithms enabling robust fine-grained hand gestures tracking and recognition are yet still to be improved to reach the commercial markets. Therefore the need for solutions to compare quantitatively different algorithms and provide means to evaluate and optimize them with valid methodologies is increasing. Similar solutions have been developed for different contexts; the Opportunity project [1] and TUM Kitchen dataset [2] focus on activity recognition and the HumanEva project [3] focuses on full-body human pose estimation; the latter has been particularly well accepted by the research community [4].

The goal of this project is to develop a framework and an open-source corpus focusing on hands and arm gestures and including the ground truth data required to evaluate both tracking and recognition algorithms. Compared to the previously cited projects, FEOGARM will put a particular emphasis on evaluation and benchmarking methods for hand and arm gesture recognition algorithms. The framework should also facilitate the usage of the corpus by providing mechanisms to plug transparently different algorithms and specific tools to compare the obtained results. By providing developers with a simple solution to obtain qualitative and quantitative feedbacks, this project should also promote and encourage discussions, collaborations and competition through a common benchmarking standard.

II. A FACILITITATING FRAMEWORK

The developed framework should provide a robust architecture in order to acquire in a synchronized way the data from various sensors, to create a fully annotated open-source corpus similarly to the work from Bannach & al [5]. The main difference being the evaluation and benchmarking tools that are specifically designed for gesture recognition algorithms.

A. Capture environment

A specific capture environment has been set-up to record the base corpus. This dedicated room contains the required computer hardware to manage multiple types of distributed pervasive and wearable sensors and a screen to display information to the user. The sensors comprise a set of Xsens MTw inertial motion units (IMU) placed on the dominant arm of the user; Kinects™ and PlayStation Eyes™ placed in front and above the subject. Physiological devices such as EMG sensors are set on the arm of the subject and a light EEG on his head. Note that a "glove-sensing device" is also planned for a next step of the project. The IMU and glove-device will be used to produce the required ground-truth tracking data while maintaining as much as possible a natural appearance of the recorded subjects.

The different types of sensors should be easy to add, monitor and characterize through simple visual interfaces using blocks and connection mechanisms. The consistency and synchronization of the recorded data should be ensured through software or/and hardware specific methods; inter-sensors synchronization is a particularly important topic that should be strongly ensured considering a locally distributed network architecture.

B. Open-source corpus of hand gestures

The corpus contains recorded sequences of the upper-body with a particular focus on one-hand and arm gestures. The exact gestures and modalities still have to be precisely defined, depending on the chosen recording scenarios. The corpus will be made freely available and contain sequences acquired under different conditions. Each sequence of the corpus will contain various types of information such as meta-data, raw data, extracted features and annotations as listed in Fig 1.

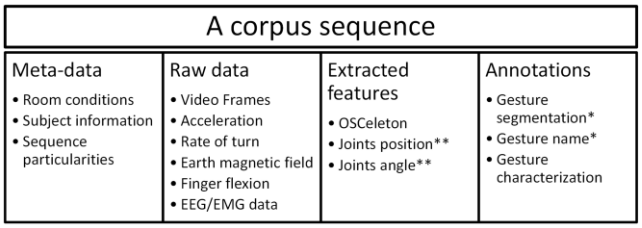

Figure 1: List of information available in each sequence of the corpus. The starred items correspond to the ground truth data for recognition (*) and tracking (**).

All the information is timestamped and meant to provide useful additional information to the raw data of the sensors. The extracted features and the annotations form the basis for the ground truth, for the evaluation of tracking and recognition respectively.

C. *Benchmarking and evaluation of algorithms*

To facilitate evaluation and benchmarking of algorithms, specific tools will be developed to provide a connection between algorithms and the framework by replaying all sensors data transparently, similarly to what has been developed for the Opportunity framework [6]. FEOGARM will provide two types of benchmarking tools. The first type, for recognition algorithm, will use the annotated ground-truth to provide results through statistical errors measurements; notably type I and II errors. The second type, for tracking algorithms, will take advantage of the extracted features ground-truth to infer a measure from joint-angles differences over-time.

The correlations between the detailed results of an algorithm evaluation and the data, features and annotations from the corpus might also yield a characterized evaluation of algorithms performances. Such specific evaluation might provide useful feedbacks on the particular strengths and flaws of evaluated algorithms.

III. Design and Implementation

After a study of the existing platforms, the decision has been taken to build our framework on top of the Context Recognition Networked Toolbox (CRNT) [7]. This toolbox provides the basic architecture and tools corresponding to our needs and it is actively used in the pervasive research field [1], [5], [6]. The idea is to develop or extend CRNT drivers and specific processing and evaluation modules, a database for the corpus, an API to handle external communications and graphical interfaces to facilitate the usage of the toolbox as illustrated in Fig. 2 below.

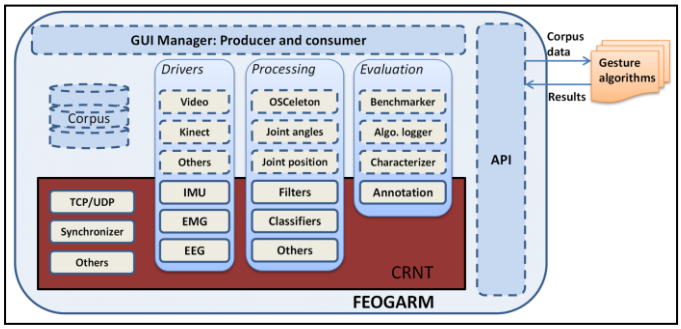

Figure 2: Design view of the FEOGARM framework articulated around CRNT. In dashed lines are the elements that will be developed.

The developed framework will provide two main interfaces: producer and consumer. The producer interface targets the creation of datasets and includes all required plugins for acquisition, recording, annotation and visualization of data during experiments. The consumer interface targets developers who are in the process of developing their algorithm(s) and require complete datasets for training and testing purposes or developers which already developed their algorithm(s) and require benchmarks and comparisons against others to validate their approach. Note that consumers may also introduce new data to the corpus such as specific annotations or the measures resulting from the evaluation of their algorithm(s).

The actual implementation focus is set on developing drivers to manage video streams, notably for Kinect sensors, directly from the CRN Toolbox, and also to find efficient ideas and solutions to synchronize heterogeneous devices. The local-networking possibilities of the toolbox for our requirements are also being assessed. The first recording trials with the previously presented setup are planned for the autumn 2011.

IV. Conclusions and Future Work

This paper laid the foundation for a platform which will support the development and evaluation of gestures tracking and recognition algorithms. The main characteristics and goals of the framework have been defined along with specific details concerning its design. The presented framework should, at term, replace specific frameworks developed independently in research laboratories working on gesture recognition and establish incentives for the community to interact and discuss via a common framework allowing quantitative benchmarking measures and characterized evaluations of algorithms.

In the future steps of the project, more types of sensors will be progressively added such as EEG, EMG and data-glove(s) sensors to add modalities and improve the accuracy of the ground-truth. The model of the corpus and gesture characterization will be further defined to provide accurate feedbacks on algorithms to developers. The long-term goal is also to study the possibility to infer an optimal non-intrusive device dedicated to gestures recognition by taking advantage of the developed framework and its corpus.

Acknowledgment

This research has been supported by Hasler foundation within the framework of “Living in Smart Environment project”.

References

[1] D. Roggen et al., “Collecting complex activity datasets in highly rich networked sensor environments,” in *Networked Sensing Systems (INSS), 2010 Seventh International Conference on*, 2010, no. 0, pp. 233–240.

[2] M. Tenorth and J. Bandouch, “The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition,” *Vision Workshops (ICCV*, pp. 1089-1096, Sep. 2009.

[3] L. Sigal, A. O. Balan, and M. J. Black, “HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion,” *International Journal of Computer Vision*, vol. 87, no. 1-2, pp. 4-27, Aug. 2009.

[4] L. Sigal and M. J. Black, “Guest Editorial: State of the Art in Image- and Video-Based Human Pose and Motion Estimation,” *International Journal of Computer Vision*, vol. 87, no. 1-2, pp. 1-3, Oct. 2009.

[5] D. Bannach, K. Kunze, and J. Weppner, “Integrated tool chain for recording and handling large, multimodal context recognition data sets,” *Proceedings of the 12th*, pp. 357-358, 2010.

[6] M. Kurz and A. Ferscha, “Sensor abstractions for opportunistic activity and context recognition systems,” in *5th European Conference on Smart Sensing and Context (EuroSSC 2010)*, 2010, pp. 135–149.

[7] D. Bannach, O. Amft, and P. Lukowicz, “Rapid Prototyping of Activity Recognition Applications,” *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 22-31, Apr. 2008.