

Interact: Gesture Recognition in the Cloud

Alexandros Preventis, Kostas Stravoskoufos, Stelios Sotiriadis, Euripides G.M. Petrakis

Intelligent Systems Laboratory

School of Electronic and Computer Engineering

Technical University of Crete (TUC)

Chania, Greece, GR-73100

Email: {apreventis, kgstravo, s.sotiriadis, petrakis}@intelligence.tuc.gr

Abstract—*Interact* is a cloud based gesture recognition solution based on the FI-WARE cloud platform, realized by means of cloud services communicating with each other through REST APIs. To promote the development of Future Internet (FI) applications, all functionalities of *Interact* are offered as cloud services. The system is sensor independent and is designed to work with the most popular motion sensors. To show proof of concept, a system prototype has been developed utilizing the LEAP Motion sensor.

I. INTRODUCTION

Gestures have always been an integral part of human social interaction. They are typically used in a real-life conversation to resolve or to express a meaning (for example raising a thumb up to signal that something is “ok”). There are communities of people with speaking or hearing disabilities that communicate explicitly through a set of gestures which form what we call a “sign language”. Communication outside of their communities for those people is often difficult, as most people are not familiar with sign languages. Also, the number of different sign languages (each country has its own) makes the communication even harder. Gesture recognition technology can address the problem of translating sign languages to text or speech, and at the same time enable new means of interaction and communication with devices introducing a more natural way to control machines and enhance user experience [1]. Interaction by gesture recognition techniques is achieved by processing data from Web cameras, motion sensors or other wearable devices. Gesture recognition systems suffer from certain disadvantages:

- 1) They are proprietary, featuring a limited number of gestures not allowing users to modify their gesture database to create their own gestures.
- 2) They are all sensor-specific. The user does not have a choice to use different sensors with the same system.
- 3) They cannot be accessed through the Web. They are tied to the local machine where the sensor they are obtaining data from is connected.
- 4) They are developed and tested on personal computers, with limited computational resources [2]. Thus, there are limitations on the amounts of data or the number of users they can serve.

This work focuses on gesture recognition by analyzing data obtained by motion sensors (e.g., the Microsoft Kinect¹, the ASUS Xtion PRO² and the Leap Motion³ sensor which can

track only the hands with more accuracy). The output of those sensors are streams of data with vectors of features describing the human body with every change in position or posture. As motion sensors are becoming more and more popular, with the concept of the Internet of Things (IoT) becoming a reality, their numbers are growing and huge streams of data are produced resulting to high demands on computational power, storage and network. To handle the volume of data produced by the sensors and address their exponentially growing demands in resources, IoT devices rely upon cloud computing service.

The functionality of *Interact* are offered as cloud services to promote the development of Future Internet (FI) applications. Specifically we have developed open cloud services for:

- Recognizing both, static and motion gestures. In static gestures the hands do not move when the gesture is performed (like raising a thumb to signal “ok”) while, in motion gestures hands are moving like waving or swiping. Gesture recognition accuracy is enhanced by machine learning.
- Creating private gesture collections where developers can store gestures and use them to develop FI applications. *Interact* allows multiple users to subscribe on the cloud and receive or process the communicated hand gestures. Subscribed users in difference places may take part in a lecture or develop an application that interprets the hand gestures.

Two use cases have been developed:

- 1) A “gesture to speech” service to demonstrate how *Interact* can be used to offer communication services for people with speaking inabilities and translates hand gestures to text or speech. Our system prototype features a gesture collection containing the American Sign Language (ASL)⁴ finger-spelled alphabet.
- 2) A “gesture to action” service to show how gestures can be transformed to system actions and commands. We use *Interact* to control the Sphero⁵ robotic ball through the cloud.

II. INTERACT

Interact is a motion sensor driven gesture recognition system, based on the FI-WARE⁶ core platform. FI-WARE

¹www.microsoft.com/en-us/kinectforwindows

²www.asus.com/Multimedia/Xtion_PRO

³www.leapmotion.com

⁴<http://www.nidcd.nih.gov/health/hearing/pages/asl.aspx>

⁵<http://www.gosphero.com>

⁶<http://www.fi-ware.org>

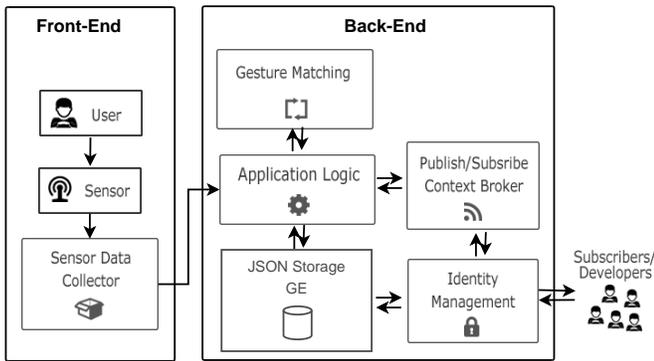


Fig. 1. Architecture of Interact

offers a set of cloud-based, reusable and modular components (APIs), referred to as Generic Enablers (GEs) implementing reusable and commonly share functions that make it easier to develop FI applications. GEs are stored in a public catalogue that developers can browse and select the appropriate APIs to use.

Interact is realized by means of cloud services communicating with each other through REST APIs [3]. It works with any motion sensor currently available, it offers open cloud services for gesture recognition and manipulation. To achieve sensor compatibility, *Interact* uses a data collector module that features sensor specific code in order to transform the data from each sensor to a compatible format (i.e., JSON⁷). It is hosted on the *Intellicloud*⁸ infrastructure of TUC⁹. The current prototype system utilizes the Leap Motion sensor. Hand gestures are represented by a set of features whose values are provided by the motion sensor. Gestures are organized in sensor databases with each database featuring access to specific gesture collections. Only the creator of a collection has privileges to add new gestures and access rights. Gestures are stored in the form of objects with key-value pairs that represent gesture features (e.g., number of fingers, rotation of hand etc.). When a gesture is captured, gesture data is obtained by the sensor and the system matches the input to the gesture records in the database of the specified sensor.

A. Architecture

Figure 1 illustrates the architecture of *Interact*. It consists of three basic blocks: (a) the front-end where the sensor data is collected, (b) the cloud back-end where the application logic is implemented and (c) the subscribers-end where applications operate, accessing the services of *Interact*. Each of these blocks consists of one or more modules that provide certain functionality:

- **Sensor Data Collector:** Collects data from the motion sensor and converts it to JSON Objects to be processed at the back-end.
- **Storage Service:** A REST API for storing and retrieving hand gestures.

- **Gesture Matching Service:** The gestures are compared with those in a given collection in order to determine their identity and assign them a meaning or translate them to commands.
- **Publish/Subscribe Service:** This service informs subscribers about changes in a domain of interest (e.g., new gestures have been added in a collection).
- **Identification Service:** This module is used for user authentication and access authorization. This applies to both users and developers of applications.
- **Application Logic:** It implements the application logic of *Interact*. It handles requests received from the front-end or consumer applications.

The modules implementing the functionality of *Interact* operate at the back-end except the Sensor Data Collector which operates at the front-end. They are implemented by means of the following GEs.

- 1) **JSON Storage GE:** This (FI-WARE) GE provides NoSQL¹⁰ database management services.
- 2) **Gesture Matching GE:** This GE is implemented at TUC. It takes as an input the gesture data and the selected collection and responds with the name of the class this gesture belong to. It implements a Naive Bayes Classifier. The system has been tested on the ASL finger-spelled alphabet. The alphabet consists of 26 static one-handed gestures, each one representing a letter of the English alphabet. The predicted accuracy of gesture recognition is 71.9%.
- 3) **Publish/Subscribe Context Broker GE:** This (FI-WARE) GE provides services to manage context information and context information availability.
- 4) **Identity Management GE:** This (FI-WARE) GE provides secure and private authentication from users to devices, networks and services and, user profile management.

III. CONCLUSIONS AND FUTURE WORK

Interact is sensor-independent gesture recognition system on the cloud. Improving recognition accuracy by incorporating more elaborate machine learning and implementing additional use-case scenarios using more sensors (e.g., Microsoft Kinect, Asus Xtion PRO) are issues for future research.

ACKNOWLEDGMENT

The research leading to these results has received funding from the 7th Framework Programme (FP7/2007-2013) of the EU under grant agreement no. 604691 (project FI-STAR).

REFERENCES

- [1] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 677–695, Jul. 1997.
- [2] N. Liu and B. C. Lovell, "MMX-Accelerated Real-Time Hand Tracking System," in *IVCNZ 2001*, 2001, pp. 381–385.
- [3] R. T. Fielding, "Architectural styles and the design of network-based software architectures," Ph.D. dissertation, University of California, Irvine, 2000.

⁷<http://www.json.org>

⁸[urlhttp://cloud.intellicloud.tuc.gr](http://cloud.intellicloud.tuc.gr)

⁹<http://147.27.50.135/interact>

¹⁰<http://nosql-database.org>