Journal of Biomedical Informatics 63 (2016) 249–258

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Open source platform for collaborative construction of wearable sensor datasets for human motion analysis and an application for gait analysis

César Llamas [a], Manuel A. González [b], Carmen Hernández [a], Jesús Vegas [a,*]

[a] Departamento de Informática, Universidad de Valladolid, Spain
[b] Departamento de Física Aplicada, Universidad de Valladolid, Spain

ARTICLE INFO

Article history:
Received 19 February 2016
Revised 26 August 2016
Accepted 31 August 2016
Available online 1 September 2016

Keywords:
Open source hardware
Sensor network
Open source software
Human motion analysis
Gait analysis dataset

ABSTRACT

Nearly every practical improvement in modeling human motion is well founded in a properly designed collection of data or datasets. These datasets must be made publicly available for the community could validate and accept them. It is reasonable to concede that a collective, guided enterprise could serve to devise solid and substantial datasets, as a result of a collaborative effort, in the same sense as the open software community does. In this way datasets could be complemented, extended and expanded in size with, for example, more individuals, samples and human actions. For this to be possible some commitments must be made by the collaborators, being one of them sharing the same data acquisition platform.

In this paper, we offer an affordable open source hardware and software platform based on inertial wearable sensors in a way that several groups could cooperate in the construction of datasets through common software suitable for collaboration. Some experimental results about the throughput of the overall system are reported showing the feasibility of acquiring data from up to 6 sensors with a sampling frequency no less than 118 Hz. Also, a proof-of-concept dataset is provided comprising sampled data from 12 subjects suitable for gait analysis.

1. Introduction

The human motion analysis (HMA) has been in the focus of the research community for the last three decades. Its range of interest extends from medicine to psychology and its application spans over security, multimedia, e-Health, recreational informatics, robotics and so, attracting both industry and research groups. Human pose estimation, motion tracking and activity recognition have been studied in depth by the computer vision community during past decades [1,2]. The recent exceptional development of microelectronics has permitted new approaches based on unobtrusive wearable sensors [3], mobile devices [4] and 3D depth sensors [5]. Wearable sensors present some advantages over image recognition based systems on certain environments where quality of video tracking is severely compromised, a finer grain data is needed or some privacy items related with video are involved.

The term dataset is a quite generic designation that denotes an organized set of data usable as source data to some model. The construction of a dataset poses a task not much different from that of many other research efforts, so the usual approach consisting of Materials, Methods, and Results is also applicable. Any data acquisition approach is part of the method of the research. We think that the problems regarding this issue in the field of HMA are challenging enough to be studied.

There has been a noticeable increase in the accuracy attained in the most recent datasets that integrate consistently different data sources [6], like MHAD from Ofli et al. that gathers synchronous video, 3D depth sensors as well as accelerometers and gyroscopes [7]. At the same time there exists some other very specialized datasets oriented to smartphones and small devices like [4] or specifically oriented to a set of tasks like, for example, human identification through gait recognition [8,9]. An important group of datasets is the one addressed to assess neurological and motor impairments. Mancini et al. propose a platform devised to this kind of task [10] and notice some datasets for the detection and diagnosis of parkinson disease. In this case, Mancini focuses his proprietary platform that consists a set of hardware sensors and an also proprietary software oriented to the medical community. Kirtley C. maintains the "CGA normative gait database" mostly for medical use.[1]

* Corresponding author.
E-mail addresses: cllamas@infor.uva.es (C. Llamas), manuelgd@termo.uva.es (M.A. González), chernan@infor.uva.es (C. Hernández), jvegas@infor.uva.es (J. Vegas).

[1] CGA normative gait database. Available at http://www.clinicalgaitanalysis.com/data/. (Last accessed in February, 2016).

http://dx.doi.org/10.1016/j.jbi.2016.08.025
1532-0464/

Almost any review on this subject includes an up to date list of datasets relevant to some area of interests. In [11], Zang and Sawchuk tabulate some significant datasets concerning wearable devices. Almost all of them include three axis accelerometers and gyroscopes placed on several parts of the human body ranging from one place up to 12 and involving from one subject up to 43 individuals and some other data relevant for the dataset. Most of them consider gait analysis as a part of the modeling task. In a recent review, Moore et al. [9] presents extensive results on recent datasets related to this task. It could be seen that the number of subjects, the specificity of the task and the quantity and variety or sensors are tightly correlated. Therefore devising a comprehensive dataset suitable intended for a too broad use is a task full of predicaments. Consequently, it is not unusual that any research effort must take into account the availability of a platform for data acquisition. It is a given fact that to develop a platform to acquire a dataset is serious hinder to every research group with a tight budget. Developing this kind of systems is a challenging task is as detailed extensively in the literature [12,13] and is implies to take decisions that affect the research activities. As a matter of fact one of our main objectives, we bear in mind is the achievement of a system complying with some of the former qualities in what we include autonomy and wearability through the use of affordable inertial sensors.

There exist some other hinders to the construction of some datasets, in addition to the availability of the sensor platform. The most important one is to have access to a significant amount of individuals to achieve significant conclusions [14]. For example, in the case of clinical studies with a low frequency of rare disease cases. Then the cooperation in the construction of the dataset is a must. This is not different from what happens in other experimental sciences where the cooperative support of the research community it is essential to collect experimental data as shown in [15,16]. This kind of interacting communities can be embraced by the more general concept of Open Science [17]. Moreover, Open Science could serve as an accelerator to this kind of research where the availability of the collaboration based data collection is critical, as states Woelfle [18]. The HMA problem is not unfamiliar to this circumstance and the collaboration and the use of common open tools could be advantageously used in the creation of specific data sets. This could be a crucial aspect, especially in the case of small research workgroups with small budgets. Also, where a very specific research topic is intended, a common platform which it could be easily replicated by different workgroups could be very useful.

In this paper, we introduce a portable sensoring platform using wearable sensors, made of open source hardware and software, suitable for the creation of a broad class of collaborative datasets for human motion analysis. This platform could be useful as a part of the research method of a workgroup researching about human motion analysis. The implementation of this platform presented also in this paper includes the hardware and software elements needed to achieve the construction of a data set about HMA, right out-of-the-box.

This kind of foundation allows a fast prototyping, especially in the research phase, with flexible general purpose hardware and software components.

Our design takes advantage of existing Open Source hardware and software. Nowadays, open-source interest makes possible the creation of open-source scientific devices made of free open-source software on top of open-source microcontrollers [19]. This, combined with 3D printing facilities, allow us to dispose of highly sophisticated scientific equipment while driving down the cost of research tools.

The rest of the paper is organized as follows: Section 2 summarizes some findings about this area, and digs in the requirements of other similar initiatives. Section 3 describes the architecture and the design proposed by our team. Section 4 exposes the implementation of our system and the main performance details of it. In addition, proof-of-concept dataset is presented to test and illustrate the capabilities of our proposal. Section 5 discusses our main findings and prospects some suggestions for the future designs.

2. Background

2.1. Collaborating in dataset development

The collaboration is a necessary part of the everyday activity of most research groups, and participating in research networks is a must for almost every application in finding resources from public research programs. As stated formally, for the collaboration in the construction of a common dataset to be possible, the teams must agree into some common items given the teams share a common description for the problem:

- an application domain defined in terms of baseline task the dataset is addressed for, including physical and methodological restrictions on the data and way in what they are collected;
- a formally stated common metadata description;
- an curation procedure including roles of the participants, protocols, data status and storage and retrieving conditions of the dataset; and finally
- a statement of ethical and legal issues related to acquisition and operation of the dataset.

As far as we know from the bibliography on HMA [1–3] all the datasets are achieved through comprehensive projects performed by one specific team without a collaborative basis, or the intention of to be extended by other researches in a consistent and systematic manner. There exist, nonetheless, some relevant cases of collaboration in the Open Science initiative as HUPO Plasma Proteome Project [20] and Human Genome Project [21]. Both are publicly well known cases of successful collaboration in the construction of high level datasets resulting from partition and aggregation efforts. However, the datasets we are interested in (HMA) comprise low level data suitable to be filtered and transformed with low level techniques. In our experience to share the same acquisition platform among the involved groups is an invaluable asset to meet this restriction pertaining the very first assumption of the list given above.

2.2. Prototyping sensor platforms

Along the last years of developing of sensor networks and sensor platforms, independently of its field of application, there seems to be some agreement on the main issues and constraints of these systems. In [22] the authors list the challenges established in the development of the pervasive healthcare systems, being a main fact that several issues are concerned to the developing of the sensor platform. Some of the topics highlighted by the authors are the need of interoperability and the intention of simplifying the deployment and enhance the scalability. These goals are completely aligned with the conclusions of Akyildiz et al. [23], that stated that flexibility, affordability and flowing deployment are some the desired goals related to the success of sensor networks. Moreover, the authors identify the main constraints and requirements in the development of these systems, such as fault tolerance, scalability, cost, hardware, topology change, environment and power consumption.

Lo and Yang, in [24] state that most hardware platforms for pervasive healthcare applications are proprietary designs. They hold