Virtual Archaeology Review, **8**(16): 12-21, 2017
http://dx.doi.org/10.4995/var.2017.5872

Received: June 3, 2016
Accepted: September 22, 2016

DESIGNING NATURAL GESTURE INTERACTION FOR ARCHAEOLOGICAL DATA IN IMMERSIVE ENVIRONMENTS

DISEÑO DE GESTOS NATURALES INTERACTIVOS CON DATOS ARQUEOLÓGICOS EN ENTORNOS DE INMERSIÓN

Niccolò Albertini[a], Andrea Brogni[a,*], Riccardo Olivito[a], Emanuele Taccola[b], Baptiste Caramiaux[c], Marco Gillies[c]

[a] Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy. niccolo.albertini@sns.it; andrea.brogni@sns.it; riccardo.olivito@sns.it

[b] Università di Pisa, Lungarno Pacinotti 43, 56126 Pisa, Italy. emanuele.taccola@unipi.it

[c] Goldsmiths University, New Cross, London, SE14 6NWL, UK. b.caramiaux@gold.ac.uk; m.gillies@gold.ac.uk

Abstract:

Archaeological data are heterogeneous, making it difficult to correlate and combine different types. Datasheets and pictures, stratigraphic data and 3D models, time and space mixed together: these are only a few of the categories a researcher has to deal with. New technologies may be able to help in this process and trying to solve research related problems needs innovative solutions. In this paper, we describe the whole process for the design and development of a prototype application that uses an Immersive Virtual Reality system to acces archaeological excavation 3D data through the Gesture Variation Follower (GVF) algorithm. This makes it possible to recognise which gesture is being performed and how it is performed. Archaeologists have participated actively in the design of the interface and the set of gestures used for triggering the different tasks. Interactive machine learning techniques have been used for the real time detection of the gestures. As a case study the agora of Segesta (Sicily, Italy) has been selected. Indeed, due to the complex architectural features and the still ongoing fieldwork activities, Segesta represents an ideal context where to test and develop a research approach integrating both traditional and more innovative tools and methods.

Key words: cyber-archaeology, gesture recognition, virtual reality (VR)

Resumen:

Los datos arqueológicos son heterogéneos y difíciles de correlacionar entre diferentes tipos. Las fichas técnicas y las imágenes, los datos estratigráficos y los modelos 3D, el tiempo y el espacio mezclados entre sí: son solamente algunas de las categorías que el investigador tiene que tratar. Las nuevas tecnologías pueden ser capaces de ayudar en este proceso, llenando el vacío entre la historia y el futuro, y tratar de resolver las necesidades de investigación con soluciones innovadoras. En este trabajo se describe todo el proceso que conlleva el diseño y desarrollo de un prototipo de aplicación, que utiliza un sistema de realidad virtual inmersiva para acceder a datos 3D de la excavación arqueólogica a través del algoritmo Gesture Variation Follower (GVF), que permite reconocer lo que el gesto está realizando y cómo se realiza. Los arqueólogos han participado activamente en el diseño de la interfaz y el conjunto de gestos utilizados para la activación de las diferentes tareas. Se han utilizado avanzadas técnicas de aprendizaje automático para la detección en tiempo real de los gestos. Como caso de estudio se eligió el ágora de Segesta (Sicilia, Italia). De hecho, debido a las características arquitectónicas complejas y las actividades de trabajo de campo todavía en curso, Segesta representa un contexto ideal donde poner a prueba y desarrollar un enfoque de investigación integrando ambas herramientas tradicionales y los más innovadores métodos.

Palabras clave: ciber-arqueología, reconocimiento de gestos, realidad virtual (RV)

1. Introduction

We say archaeology and we think of history and events. We see ancient remains and we imagine palaces and battles, involving historical personalities and entire nations. Studying and working in this field opens the fantastic opportunity to travel back in time but also to be immersed in a very complex environment, made of pictures, numbers, remains, books and places, where hypothesis and research are driven by details. As scholars, we could see this huge amount of data as a heterogeneous database, strongly interconnected with and linked to their representation. The growing amount of scientific information creates problems of interpretation and understanding which are often solved by using graphical methods, since the human brain activates one-third of its neurons when processing visual information. Manipulating and visualizing this kind of data has always been a limiting factor in research and scientific history is full of new graphical conventions enabling new perspectives and understandings of data.

Given the increasing computational and graphical power of modern computers, the focus of research has been shifted from 'how to represent' to 'what should be represented', in every research field, from engineering to chemistry, from cultural heritage to medicine (Brickmann, Exner, Keil, & Marhöfer, 2000).

*Corresponding author: A. Brogni, andrea.brogni@sns.it

Archaeology is a field that more than others may take advantage from this innovation. Imagination may leave space for exploration, if we can efficiently transfer data from an excavation or other research to a virtual reconstruction. This process has to start from scholars, defining their needs and the limit of the available tools, and giving directions to technicians for designing solutions. It is not just making 3D models, but integrating new interactive technologies into the whole working process.

In particular, our ability to understand the archaeological materials relies on our ability to interact with and manipulate them. The theory of embodied cognition (Kirsh, 2013) stresses the importance of our interactions with the world in our cognitive understanding of it. The ability to rotate an artefact in our hand helps us to see it better from different perspectives, the ability to hold a tool helps us to understand how it can be used. Relative to our ability to manipulate physical artefacts, our manipulations of virtual artefacts are very unnatural and poor: indirect rotation via sliders and buttons and limited degrees of freedom. We propose that one of the great benefits of virtual reality is the ability to enable more natural forms of interaction that will better support archaeologists in understanding their digital artefacts.

This paper presents a virtual reality representation of archaeological data that includes a gestural interface that enables researchers to interact with data immersively, using their hand movements, without the intermediary of desktop interface devices such as mice or touch screens. This work comes straight from archaeological researchers and their curiosity toward virtual technologies. The idea is to transfer actions from the excavation to the virtual environment, without going through the desktop computer. The aim is not to have buttons and sliders in 3D, a virtual replica of the desktop interface, but placing the archaeologist in a working replica of the excavation area, where they can access and manipulate the data, edit and consult information, moving around the reconstruction and the actual area. A very important issue will be to design such interactions with these virtual reality representations in a way that they feel natural and familiar to archaeologists, in their imaginative concept of acting on the excavation area. A key aspect of our system is that the process of gesture design is very rapid, involving recording one example of each gesture (though they may be re-recorded to refine the interface after testing). This makes it possible to record highly personalised gestures for each individual using the system and so adapt it to ideosynchratic ways of working, cultural differences or disabilities.

This paper describes the whole design process leading to the prototype application, used by the archaeologist to access the excavation data from inside the virtual reconstruction of the site, visualized in 3D in a CAVE-like system (Cruz-Neira, Sandin, Defanti, Kenyon, & Hart, 1992). An innovative gesture recognition library has been used for letting them record and re-use natural and simple gestures for the tasks they designed.

1.1. Motivations

The aim of the application is to contribute to the development of a new approach to the exploration and analysis of stratigraphic archaeological data. The application represents a further step in an already existing project, which is now enriched by new and more elaborate interaction tools, based on the requirements and practical experiences of archaeologists.

The first experiment was in fact based on a leap-sensor attached to a pair of 3D glasses, to be used within a CAVE-like system. This enabled interaction with the models through an interface consisting of a small set of gestures (Olivito, Taccola, & Albertini, 2015). Although activated by hand movements, the gestures looked quite artificial because the user had to continuously interact with a selection menu, which mimicked a desktop or touch-screen interface. In addition, since the 3D glasses and the leap-sensor were not wireless, they severly limited the freedom of movement within the CAVE.

In this new phase, thanks to a new wireless tracking system, it has been possible to develop hand gestures only by using 3D glasses and movement trackers.

The design of the gestures and their functionality is the result of a continuous debate between archaeologists and system developers, whose main goal has been to satisfy two specific requests by the archaeologists. On the one hand, to use a limited set of gestures that is easily memorizable because they are natural. On the other hand, to reproduce through these gestures the movements an archaeologist usually carries out, not only during the field activity but also in the cognitive process activated during the excavation and the interpretation phases.

The case study is the agora of Segesta, in the north-western corner of Sicily, Italy. Here, archaeologists have discovered the remains of a huge Late-Hellenistic portico (*stoa*) which bordered the ancient public square of the city (the north side was 82 m long ca., the west and east side were 20 m long). In May 2014, during the three-week campaign of excavation on the site, archaeologists started a project of photogrammetric documentation of the digging activity. In the first phase of this project six 3D image-based models of the stratigraphic sequence were created (Fig. 1), following the progressive development of field activities.

The excavation levels included the most superficial layers, up to the traces of medieval re-occupation of the area and those relating to the collapse and abandonment of the late Hellenistic *stoa*. The complexity of the stratigraphic sequence and the architectural context represent an excellent field test for the different requirements and questions that an archaeologist has to deal with during the research. As a consequence, although the case study does not fill the full range of possible interactions, it still constitutes a valid basis that could be applicable to all the contexts in which a stratigraphic investigation is carried out.

In this sense, the application aims at supporting the traditional tools used in the archaeological field, so as to enrich the interpretative process due to the use of 3D data and to simulate the activity of excavation, which is by its nature destructive. It does this while operating within an immersive virtual environment that allows for full embodiment and interaction with the digital models that is as natural as possible. A further benefit is the possibility of re-creating, from an emotional and perceptive side, the mental dynamics which an archaeologist processes during the field activity and after. This can help to validate, re-examine, or even modify the interpretations elaborated in the very moment of the excavation, so as to fill a gap that traditional tools leave open.

Figure 1: The six image-based models realized during the fieldwork activity, illustrating the stratigraphic sequence.

2. Background

2.1. Virtual reality (VR) for archaeology

Usually, in the archaeological practice, the data related to the excavation are managed by 2D investigation, collection, and consultation tools. The use of 3D image-based recording procedures, facilitated by easy-to-use and low-cost software (even open-source) has increased. This has recently stimulated a wide debate on the necessity to employ, in addition to traditional tools, a new methodology able to take advantage of all the possibilities offered by 3D acquisition methods, particularly by interacting with them in real time (Pietroni & Pescarin, 2010; Pietroni & Rufa, 2012).

This approach, known as Virtual- or Cyber- Archaeology, has produced some very interesting examples (Forte & Siliotti, 1997; Forte, 2010; Forte, 2014). Among these, is the case of Çatalhöyük where a team from Duke University conducted investigations in which 3D documentation played a key role. The result of this activity is available on different levels of interaction and immersion: by using portable devices that allow users to visualize and interact with the models in 3D (i.e. Z-space and Oculus Rift), but especially by using the DiVE (Duke immersive Visualization Environment), a 3 x 3 x 3 m CAVE-like system in which the user can interact with the models by a “magic wand” interface. Although the application presented in this paper aims at continuing the work that was begun with the case of Çatalhöyük, nevertheless it is significantly different from the above-mentioned case study. Indeed, our application is almost exclusively based on natural hand gestures which do not require any further device, apart from 3D glasses and hand-tracking sensors. As a result, other digital devices (e.g. tablets) can be handled by the user during the activity within the virtual immersive environment.

2.2. Natural interaction in VR

VR and 3-D visualization enable researchers to perceive data in new and more natural ways. In some cases, graphs are better than tables, and different graphs emphasize different aspects of the same dataset. It all depends on what data we are observing and what we are looking for in the data. 3D data should be easier to understand if visualized in a tridimensional and stereoscopic system.

An immersive visualization adds new modalities to the perception of data, adding to sight the proprioception of our own body (Taylor, 2009) and inducing natural reactions during the exploration of the virtual objects (Lackner, 1988; Brogni, Caldwell, & Slater, 2011). This will help to recognize shapes and dimensions, distances and spatial relationships between elements of the scene, bringing our space perception closer to everyday reality. In the case of archaeology, for example, seeing an artefact in a 3D reconstruction and comparing it with others, is more intuitive and efficient than watching pictures or columns of numbers. The whole design should enable researchers to understand the data and test hypothesis, acting with a focus on the task itself and not on how to perform the task.

Motion capture devices improve the capability of virtual reality systems, in terms of performance and affordability, enabling so-called natural interfaces: interfaces that are effectively intuitive and easy to learn, but transparent to the user. Natural interaction does not mean ‘interacting in a human way’ or ‘imitating the physical world,’ but it means designing an interaction that is going to be invisible and effective for the user in the task they are working on. Gestures and hand manipulations are actions that we perform every day without thinking, and in many cases, they can be far more effective than traditional computer interfaces. For example, to rotate a 3-D model, it is more ‘natural’ (and effective) to grab it with the user’s hand, rotate it, and perceive the rotation with their proprioceptive system, than to use a slider or other interface widget. The user’s experience of doing similar actions every day will help make the interaction more effective and quick to learn. Merging 3-D stereoscopic visualization and natural interaction could bring enormous advantages in different areas, where the visualization and the perception of data are important for the results.

A key part of this is allowing users to interact with the visualized data in real time in a natural and comfortable way. Being able to switch between different representations, highlight specific features, hide the less relevant aspects and interactively compare different datasets (or different representations of the same entity) helps the processes of understanding and insight.

2.3. Gestural interaction

Physical gestures and actions are a natural way to interact with our external environment. However, designing gesture interaction still involves important challenges such as defining a relevant and application-compatible gesture vocabulary, as well as designing an accurate gesture recognizer that allows for taking into account expressive components of the performed gestures.

2.4. Gesture design

Gesture design has been investigated following two distinct approaches. The first approach is designer-centered. Previous works consider ergonomics and technical constraints (Nielsen, Störring, Moeslund, & Granum, 2003). Ergonomics form the underlying metrics for characterizing ballistic movements like pointing (Grossman & Balakrishnan, 2005) or reaching (Nieuwenhuizen, Aliakseyeu, & Martens, 2010). Specific features of recognition systems can steer the design of gesture vocabularies to guarantee high recognition success rates (Nielsen et al., 2003). The second approach is user-centered. Long, Landay, Rowe, & Michiels (2000) asked users to rate similarity between shape-based gestures to define a vocabulary avoiding ambiguity. Wobbrock, Morris, & Wilson (2009) asked participants to perform gestures corresponding to a given command in order to arrive at a tabletop gesture vocabulary. Kane, Wobbrock, & Ladner (2011) sought to better understand the difference between gesture vocabularies created by sighted and blind people. Bragdon, Nelson, Li, & Hinckley (2011) extend this approach by looking at environmental demands on attention. Ouyang & Li (2012) have developed an evolving gesture library collected by a large user population.

A key challenge for gesture design is variability between people. People may perform gestures in different ways for a wide variety of reasons. These might be due to physical characteristics such as body shape or gender difference as well as different levels of physical ability due to age or disability. Difference might also be due to non-physical factors such as learned idiosyncratic ways of performing gestures. One approach to handling variation is to have a sufficiently general set of gestures (and a sufficiently general recogniser) that most people are able to do them. However, these general gestures might not be confortable for all people. Also, there may not be a single gesture that is applicable to all people. Gestures can be highly culturally specific: a vertical head nod might be an appropriate gesture in most western countries but a very different gesture is used in India. Also, a type of gesture might be impossible for people with certain disabilities, and they might require a very particular gesture performed in a particular way. For this reason it is often better to allow easy personalisation of gestures by individuals, an approach we take in this paper.

2.5. Gesture recognition

There exist a number of techniques for gesture recognition. Template-based methods typically rely on a single exemplar in order to define a gesture class. These include Dynamic Time Warping, the $1 Recognizer (Wobbrock, Wilson, & Li, 2007) and $N Recognizer (Anthony & Wobbrock, 2010). These methods are mainly used in systems designed to recognize simple shapes and are not robust to noise and missing values. Methods based on machine learning make use of multiple examples to derive gesture classes handling uncertainties within such classes due to different types of noise. Established methods include Hidden Markov Models (HMM) (Lucchese, Field, Ho, Gutierrez-Osuna, & Hammond, 2012) for time-dependent signals and k-Nearest Neighbor (Varona, Jaume-I-Capò, Gonzàlez, & Perales, 2009; Gillies, Kleinsmith, & Brenton, 2015) or Support Vector Machines (SVM) for static pattern. A training procedure is needed to estimate model parameters that fit best the data. Building up comprehensive databases, however, are time-consuming and are not well suited for user-centric approaches.

Gesture recognition methods that use a single template for all people provide little possibility to personalise gestures. Machine learning methods allows for variation by generalising from a large data set of people performing a gesture. However, these datasets can still be biased and limited, for example, to a data set trained on North American people might not recognise Southern European gestures. Nor will a general database be able to adapt to the specific requirements of a disabled person. This kind of personalisation requires a method that allows rapid learning of a gesture from few, or even one examples. Template methods such as the $1 Recognizer (Wobbrock et al., 2007) and $N Recognizer (Anthony & Wobbrock, 2010) allow for this, as do methods such as Caramiaux, Montecchio, Tanaka & Bevilacqua's (2014) Gesture Variation Follower (discussed below).

Most recognition systems output however results in discrete time, typically upon gesture completion. There are some systems that continuously report estimation on gesture classes or characteristics (Mori et al., 2006; Bevilacqua et al., 2010), allowing for prediction and "early recognition" meaning that the ability to recognize before the end of the gesture execution.

Still, all these methods considered variation in input data as variability, in other words noise. For example, Varona et al. (2009) normalise gestures on a number of dimensions to eliminate variation. However, variation is a key feature to allow for user-center approach in gesture interaction design as it allows for exploration. Adaptation procedures that modify the class description during recognition having been described for both template-based methods (Kratz, Morris, & Saponas, 2012) and those using statistical learning (Wilson & Bobick, 1999; Wilson & Bobick, 2000).

Some work makes direct use of gesture variation in the interaction process (Wilson & Bobick, 1999; Fdili Alaoui, Caramiaux, Serrano, & Bevilacqua, 2012). These approaches remain largely unexplored and confined to the study of gesture in subjective art performance contexts. A recent promising approach has been proposed in Caramiaux et al., (2014) where input gesture variations are explicitly taken into account in the model and estimated continuously while the gesture is performed. This method is called Gesture Variation Follower (GVF) and will be further detailed in Section 4.1.

2.6. Participatory design

The approach used in this paper is inspired by User Centred Design and in particular Participatory Design (Muller & Kuhn, 1993). User Centred Design (Norman, 1990) is an approach to designing technology where the needs of users are the key driving force behind any design decisions. Participatory design is an approach to achieving this by directly involving users throughout the process. Participatory design is key to multidisciplinary endeavors like digital heritage because it ensures that stakeholders like archaeologists are able to determine the design of their tools, even if they are eventually implemented by computer scientists. Participatory design generally involves working directly with users in workshop settings to create example designs. However, it is currently difficult to create real software prototypes in this way as users do not have the implementation skills and developing software takes a long time, so design ideas are generally limited to paper sketches that capture the look of a graphical user interfaces but do not have the feel of a full interactive system, and certainly not a virtual reality one. A variety of techniques have been used to get over this problem. Expressing designs in terms of a video can capture more of the feel of interacting with a system (Mackay & Fayard, 1999), but do not allow for actual interaction with the prototype. Generic interfaces can be used as a “design probe” (Tanaka, Bau, & Mackay, 2013), which can encourage users to think about different modes of interaction, but are still limited to the capabilities of the object. Rough mock ups and wizard of oz prototypes allow for what Buxton calls “User Experience Sketches” (Buxton, 2010), but may not have the feel of a real interactive system. Recent advances have made it possible to do interaction design in real time so that it is possible to create working prototypes of novel interfaces during participatory design sessions. They key enabling technology has been Interactive Machine Learning (Fiebrink, Cook, & Trueman, 2011), the use of statistical machine learning algorithms to allow users to design interactive systems by giving examples of interaction rather than by programming. This has the advantage that the creation of the system can be much quicker and that it can be done by end users who do not have programming skills. This makes it doubly suited to use in participatory design sessions. Caramiaux, Altavilla, Pobiner & Tanaka (2015) have successfully used interactive machine learning during participatory design sessions for novel sonic interfaces.

3. The gestural interface

3.1. Gesture Variation Follower

The Gesture Variation Follower is a technique able to recognize which gesture is being performed and how it is performed (Caramiaux et al., 2014). The former feature is called realtime recognition (or early recognition), the latter adaptation. These two features have been specifically designed in order to enable both discrete and continuous interactions.

In practice, GVF relies on two successive phases: a training phase and a recognition phase. In the training phase, the user provides the system with a set of gestures to be recognized. Each gesture is given through a single example, making the training of the system simple and light. Once at least one gesture example has been recorded, the user can switch to the recognition phase. In this phase, the user’s performed gesture is recognized as being one of the recorded example and variations with respect to the recognized example are estimated.

The current implementation of the GVF accepts free space gestures given by their positions and adapts to both temporal and geometric variations. The temporal variations are: the real-time alignment of the performed gesture onto the recognized example; its relative speed with respect to the same example, an estimated speed equal to 1 means that the performed gesture has the exact same speed at that moment than the example. The geometric variations are the relative scales (along each axis) and orientations (three angles of rotations for free space gestures). Therefore, estimated scale equals to 1 means that the gesture has the same size as the recorded example.

GVF relies on a tracking formulation of the gesture and its variations. The tracking formulation involves latent variables and observable variables. Latent variables represent the gesture identifier, position, speed, scale and orientation. Observable variables are the value of the incoming gesture given by the motion capture system. A non-linear function links observable variables and latent variables.

The goal is to estimate in realtime the latent variables. To do so, a Bayesian formulation of the tracking problem is proposed in order to take into account uncertainties stemming from the noise in capturing the data and the uncertainty in the hypothesis of the model. Therefore, instead of estimating the values of the latent variables, we estimate a probability density function (PDF) over its possible values. Several techniques exist to estimate such probability density in real-time, in the current implementation in GVF we use particle filtering. Particle filtering is a way to estimate a complex probability distribution of a random variable by sampling a large amount of the variable’s potential values and by weighting each of these values according to a given likelihood.

3.2. Virtual reality gesture design tool

The GVF algorithm was used to develop an interface for designing gestures in immersive VR. The interface was implemented in the Unity3d game engine and used within a CAVE-like immersive VR system. Users’ hand movements are tracked using an Optitrack optical motion capture system that is situated inside the CAVE and whose coordinate system is aligned with that of the CAVE.

We used Optitrack because, compared to other systems (such as Kinect), it is much more accurate. In addition, it provides an application which gives the tracking data directly, without the need to interpolate between the various cameras that would serve to cover the entire area of the CAVE.

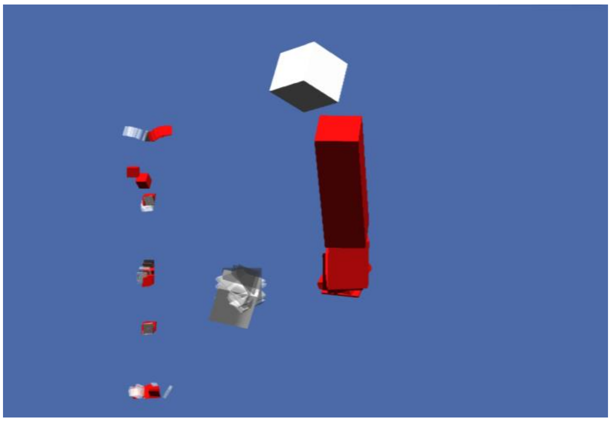
The position and orientation of each hand is used as input to the GVF. Each hand is represented as x,y,z coordinates of position and x,y,z Euler angles for rotation resulting in 6 dimensions per hand and therefore 12 dimensions overall. A single 3D scale and rotation is estimated for each gesture and applied to both hands.

The aim of the design tool was to make gesture design and personalization as easy as possible, with users only having to record a single example of each gesture. This enables each user to design and perform gestures in a way that is most natural and comfortable for them and avoids the problem of trying to build gesture vocabularies that generalize across different individuals, professional specialisations, cultures and physical abilities.

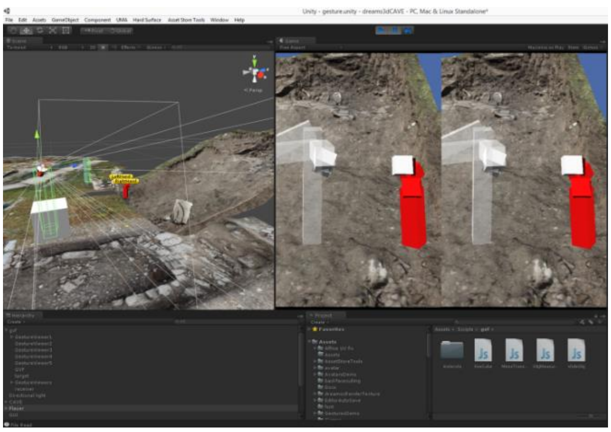
Visual feedback is provided to users in order to support them in training and using gestures. Gestures are represented as graphical trails as shown in Figure 2. These visualizations aim to support users in recording gestures and refining the gesture interfaces, as described below.

The gesture design process consists of a number of steps:

1. **Recording Phase:** Users are able to record an example of each gesture they wish to perform. As they record a gesture for the first time, a graphical trail is displayed in 3D, following their hands (Fig. 2 a). This allows them to see the overall shape of the gesture. Once the gesture has been recorded, a thumbnail is created (Fig. 2b). This is a small version of the trail that is displayed in the virtual environment, in front of, and above where the user would normally stand (roughly at eye height). The thumbnails allow the user to see all of the available gestures.
2. **Recognition Phase:** Once multiple gestures have been recorded, they can enter recognition mode, in which the gestures are recognized in real-time. During this mode, the current recognized gesture is forwarded to other modules of the application (see next section). In addition, the GVF is able to provide a measure of how much of the gesture the user has currently performed. In recognition mode, a trail is drawn in the same way as during learning, allowing users to see the shape of the gesture that they are currently performing.



(a)



(b)

Figure 2: The visualizations of the gestures: a) when performing gestures, users can see trails of their movements showing them the shape of the gesture; b) at the top left of the screen there are thumbnails of the gesture vocabulary.

3. **Testing Phase:** Recognition mode is used for testing the effectiveness of the gesture recognition to see if the system correctly recognizes gestures when they are performed again. If tests are successful, the system is ready for use, but any errors can be corrected by returning to the recording phase and re-recording the original gestures and possibly redesigning the gestures vocabulary to make the gestures easier for the system to recognize. This debugging process requires users and designers to understand why gestures are incorrectly classified and how to correct them. This difficult process is supported by our visualization. As gestures are recognized, the thumbnails are made partially transparent, each one having opacity proportional to the current estimated probability of that gesture. This makes gestures gradually fade out if they are not recognized, leaving only the recognized gesture visible. This makes it possible to easily see cases where two gestures are being confused by the algorithm and therefore support users in understanding why the system may not be working as expected. This will help users detect cases where the gestures should be rerecorded because they are too easily confused. In this sense, they are analogous to the interface proposed by Gillies et al. (2015) as a means of supporting users in building a conceptual model of a machine learning algorithm in order to better debug it and the work of Bau & MacKay (2008) who provide dynamic visual suggestions to support gesture design.

This design process can be performed in depth on initial design of the application to find a good set of gestures that work that are easily recognized and work well for archaeologists in general. However, since the process is rapid, it is also possible to repeat it for each new user, allowing them to record personalized versions of the gesture.

3.3. The application

The main idea behind the design of the application was to allow archaeologists to think about the data and not the interface. In particular, the application allows the researchers to interact with the data in a natural way by manipulating them with gestures (see Fig. 3). A short movie of the working application can be downloaded from http://polipapers.upv.es/public/journals/32/var5872.mp4.

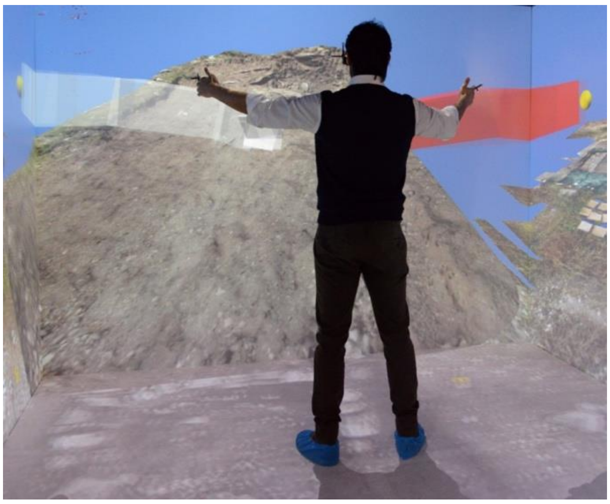


Figure 3: Using the gesture design system in immersive VR.

The first step was defining tasks and gestures. The archaeologists identified the tasks (e.g. comparison between layers of excavation, analysis of the finds). They also identified suitable gestures (see below).

The tasks are shown in Table 1. The first task is selecting the layer of excavation via two gestures to move up and down. When a layer of interest is chosen, particular finds are selected by touching them (this was not implemented via the gesture system). Once selected, objects need to be measured and then deselected.

Table 1: Correlation between gestures and tasks.

n	*Gesture*	*Task*
1	Rotation of the right forearm from inside to outside	Open contextual menu
2	Rotation of the right forearm from outside to inside	Close contextual menu
3	Slide up of the hand	Slide layer up
4	Slide down of the hand	Slide layer down
5	Arm opening from inside to outside	Measurement
6	Touch object	Selection

There is also a context menu that displays different data types depending on where you are. This needs to be opened and closed.

For the creation of the scene, we imported 3D models of layers stacked on each other and for each layer we imported 3D models of the most interesting finds, lined up with the excavation. The scene was integrated with a reconstruction of the entire agora, which can be useful for comparison with the current state of excavations.

4. Results

4.1. Designing gestures

The proposed system was used to design a gestural interface for the virtual recreation of the agora of Segesta. The system developers worked with two archaeologists over 3 sessions to design the interface. During the first phase, archaeologists, without the intervention of the system developers, designed a gesture vocabulary that would represent their behaviours and movements on the field. The first session did not use the gesture design software and primarily aimed to determine which functions of the system would be most suited to a gestural interface. The typology can be resumed in two main conceptual elements: slide/selection and analysis/measuring (see Table 1).

Then, the archaeologists linked two gestures to the task and created a diagram that described how the gestures can be integrated and used in the different sections of the application: enabling and disabling a menu, measuring an object, deselecting the currently selected object and moving from one level of the excavation to another. Selecting an object was considered an important feature, but it was decided not to use the GVF to implement it as there was a straightforward gesture that did not require gesture recognition library to implement: moving the hand inside the object.

The second session was the first to use the GVF tool in practice. The session started with a discussion of the possible gestures. The participants used both movements and language to talk through the possible interfaces. Acting out the gestures was a key part of this design phase, supporting our hypothesis that a gesture design tool based on performing example movements affords a natural way of working. The participants then attempted to record some of the gestures. They were able to record the gestures successfully and some were recognized effectively. However, some performance issues were identified which were solved before the final session.

The last session was the design of the actual interface. The participants began by reviewing the gestures they had designed in the previous session and re-recording them. They then tested them in practice. It quickly emerged that certain gestures were easily confused, for example moving the hand up to change layers (Fig. 4-2) and rotating and sweeping the hand to open the menu (Figs. 4-4 and 4-5). The open menu command was re-recorded to restrict it to only rotation and therefore removing the common positional movement. This shows that the system allowed identification of problems with gestures and re-recording of gestures. Despite these improvements, it was found that many gestures were often confused. This was because the application responded to gestures as soon as they were recognized, and therefore in the early stages of the gesture where it is still difficult to determine which gesture is performed and where the recognition may oscillate between different gestures. The application was therefore changed so that responses only happened once a user was 30% of the way through a gesture. On making this change recognition improved, other problems were solved by re-recording gestures to make them more easily distinguishable (e.g. by using a more clearly different initial movement or using different hands to perform different gestures). When the system was able to recognize individual gestures and correctly respond to them, it was tested by performing sequences of gestures in a way that simulated actual use of this system. The archaeologists were able to perform a number of tasks that were typical of their research methods such as: open and interact with the menu, visualize and slide between excavation layers by using an interaction that recalls the practical activity of ground removal, selection and analysis of objects.

4.2. The gesture vocabulary

The final set of gestures consists of two sets of two coupled gestures, in which one gesture is the inverse of the other: opening and closing the menu and moving up and down levels. There are also two single gestures: initiating measurement of the selected object and deselecting an object (Fig. 4 and Table 1).

These gestures exemplify a number of gesture design strategies. The open and close menu gestures are relatively arbitrary, they do not use a particular metaphor, but were designed based on convenience of movement and the constraints of the system (the

Figure 4: The gesture vocabulary. 1) initiating measurement by stretching hands out; 2) moving to a lower layer by raising the hand; 3) moving to a higher layer by lowering the hand; 4) rotation of the right forearm from inside to outside for opening the contextual menu; 5) closing the menu with an inverse movement; 6) deselecting the current object with a sharp downward movement of both hands.

requirement to ensure that gestures are sufficiently different to be easily recognized). The measurement gesture used a fairly straightforward real world metaphor: stretching out the hands in a clear echo of the movement made when using a measuring tape. Thedeselection gesture was similar, a sharp downward movement of the hands similar to the movement that would be made to throw down and object that was being held or to shake off an object stuck to the hands. Both of these have clear metaphors from daily life and are generic in the sense that it easy to see them being applicable in many domains, not just archaeology. The gestures to move between levels, however, are much more specific to archaeology because they use a metaphor derived directly from the archaeological practice. The gesture to move down a layer is an upward movement of the hand as if lifting earth off the layer to reveal the layer below (the archaeologists explicitly described the movement in this way). Conversely, movement to the layer above is a downward movement replacing the earth and covering the current layer. This is particularly interesting as it contrasts directly with the gestures used by the engineers in the team when doing initial testing on the system. Unlike the archaeologists, they immediately and without thinking about it, used an upward gesture to move up a layer, similar to scrolling a in a graphical interface. This example clearly shows that, while many gestures are sufficiently generic to be similar and usable across a wide range of domains, certain gestures are specific to archaeologists. They are gestures that have clear metaphors to the archaeologists' physical practice of digging and handling artefacts (in fact, in later discussion the archaeologists raised the possibility of gestures mimicking the use of tools).

It is therefore important, when designing interfaces for heritage experts (or experts in other domains for that matter) that themselves be participants in the design and as far as possible the design the interface. Otherwise, the result will be generic interfaces that miss the particular physical metaphors that arise from expert practice.

5. Conclusions

The application design process forced a reflection on the interaction modalities that archaeologists could use in virtual environments. This has made it possible not only to visualize 3D models, but also to interact with them using natural gestures, in order to consult data associated with models.

The discussion between archaeologists and software developers during the gesture design process has highlighted two different kinds of criticalities: on the one hand, conceptual differences concerning the configuration of gestures linked to specific tasks. On the other hand, some difficulties during the gesture recording and recognition phase, due to conflicts between movements which were apparently similar but which were difficult for the application to recognize. Nevertheless, the GVF has given promising results, both for the efficiency of the application and its future developments. Archaeologists have emphasized the importance of having the possibility to mentally and virtually reproduce, even after a long time and working in a different place, the field activity, so as to be able to rethink and/or reformulate interpretations, originally elaborated during the excavation. Besides the ability to operate in a completely hand-free way, thanks to hand-tracking sensors, the system will allow the use of external devices (i.e. tablets) for a wider consultation of metadata linked to 3D models, which are at present hardly visualizable within the CAVE.

References

Anthony, L., & Wobbrock, J.O. (2010). A lightweight multistroke recognizer for user interface prototypes. *Proceedings of Graphics Interface 2010*, 245–252. http://doi.acm.org/10.1145/4713060.1839258

Bau, O., & Mackay, W. E. (2008). OctoPocus: a dynamic guide for learning gesture-based command sets. *Proceedings of the 21st annual ACM symposium on User interface software and technology*, 37–46. http://doi.org/10.1145/1449715.1449724

Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., & Rasamimanana, N. (2010). Continuous realtime gesture following and recognition. In S. Kopp & I. Wachsmuth (Eds.), *Gesture in Embodied Communication and Human-Computer Interaction: 8th International Gesture Workshop, GW 2009, Bielefeld, Germany, February 25-27, 2009, Revised Selected Papers* (pp. 73–84). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-12553-9_7

Bragdon, A., Nelson, E., Li, Y., & Hinckley, K. (2011). Experimental analysis of touch-screen gesture designs in mobile environments. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 403–412. http://doi.org/10.1145/1978942.1979000

Brickmann, J., Exner, E. T., Keil, M., & Marhöfer, J. R. (2000). Molecular Graphics - Trends and Perspectives. *Molecular modeling annual, 6*(2), 328–340. http://doi.org/10.1007/s0089400060328

Brogni, A., Caldwell, D. G., & Slater, M. (2011). Touching Sharp Virtual Objects Produces a Haptic Illusion. In R. Shumaker (Ed.), *Virtual and Mixed Reality - New Trends: International Conference, Virtual and Mixed Reality 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part I* (pp. 234–242). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-22021-0_26

Buxton, B. (2010). *Sketching user experiences: getting the design right and the right design*. San Francisco, CA: Morgan Kaufmann.

Caramiaux, B., Altavilla, A., Pobiner, S. G., & Tanaka, A. (2015). Form follows sound: designing interactions from sonic memories. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3943–3952. http://doi.org/10.1145/2702123.2702515

Caramiaux, B., Montecchio, N., Tanaka, A., & Bevilacqua, F. (2014). Adaptive Gesture Recognition with Variation Estimation for Interactive Systems. *ACM Trans. Interact. Intell. Syst., 4*(4), 1–34. http://doi.org/10.1145/2643204

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM, 35*(6), 64–72. http://doi.org/10.1145/129888.129892

Fdili Alaoui, S., Caramiaux, B., Serrano, M., & Bevilacqua, F. (2012). Movement qualities as interaction modality. *Proceedings of the Designing Interactive Systems Conference*, 761–769. http://doi.org/10.1145/2317956.2318071

Fiebrink, R., Cook, P. R., & Trueman, D. (2011). Human model evaluation in interactive supervised learning. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 147–156. http://doi.org/10.1145/1978942.1978965

Forte, M. (2010). *Cyber-archaeology*. Oxford, England: Archaeopress.

Forte, M. (2014). 3D archaeology. New perspectives and challenges. The example of Catalhoyuk. *Journal of Eastern Mediterranean Archaeology and Heritage Studies*, *2*(1), pp. 1–29. http://doi.org/10.13140/2.1.3285.0568

Forte, M., & Siliotti, A. (1997). *Virtual archaeology: re-creating ancient worlds.* London: Harry N Abrams B.V.

Gillies, M., Kleinsmith, A., & Brenton, H. (2015). Applying the CASSM framework to improving end user debugging of interactive machine learning. *Proceedings of the 20th International Conference on Intelligent User Interfaces*, 181–185. http://doi.org/10.1145/2678025.2701373

Grossman, T., & Balakrishnan, R. (2005). A probabilistic approach to modeling two-dimensional pointing. *ACM Transactions on Computer-Human Interaction, 12*(3), 435–459. http://doi.org/10.1145/1096737.1096741

Kane, S. K., Wobbrock, J. O., & Ladner, R. E. (2011). Usable gestures for blind people: understanding preference and performance. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 413–422. http://doi.org/10.1145/1978942.1979001

Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction, 20*(1), 1–30. http://doi.org/10.1145/2442106.2442109

Kratz, L., Morris, D., & Saponas, T. S. (2012). Making gestural input from arm-worn inertial sensors more practical. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1747–1750. http://doi.org/10.1145/2207676.2208304

Lackner, J. R. (1998). Some proprioceptive influences on the perceptual representation of body shape and orientation. *Brain*, *111*(2), 281–297. http://dx.doi.org/10.1093/brain/111.2.281

Long, A. C., Landay, J. A., Rowe, L. A., & Michiels, J. (2000). Visual similarity of pen gestures. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 360–367. http://doi.org/10.1145/332040.332458

Lucchese, G., Field, M., Ho, J., Gutierrez-Osuna, R., & Hammond, T. (2012). GestureCommander: continuous touch-based gesture prediction. *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, 1925–1930. http://doi.org/10.1145/2212776.2223730

Mackay, W. E., & Fayard, A. L. (1999). Video brainstorming and prototyping: techniques for participatory design. *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, 118–119. http://doi.org/10.1145/632716.632790

Mori, A., Uchida, S., Kurazume, R., Ichiro Taniguchi, R., Hasegawa, T. & Sakoe, H. (2006). Early recognition and prediction of gestures. *ICPR 2006. 18th International Conference on Pattern Recognition*, 560–563. http://doi.org/10.1109/ICPR.2006.467

Muller, M. J., & Kuhn, S. (1993). Participatory design. *Communications of the ACM, 36*(6), 24–28. http://doi.org/10.1145/153571.255960

Nielsen, M., Störring, M., Moeslund, T. B., & Granum, E. (2004). A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In A. Camurri & G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers* (pp. 409–420). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-24598-8_38

Nieuwenhuizen, K., Aliakseyeu, D., & Martens, J.-B. (2009). Insight into Goal-Directed Movements: Beyond Fitts' Law. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2009: 12th IFIP TC 13 International Conference, Uppsala, Sweden, August 24-28, 2009, Proceedings, Part I* (pp. 274–287). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1145/1753326.1753457

Norman, D. A. (1990). *The design of everyday things* (1st Doubleday/Currency Ed.). New York: Doubleday.

Olivito, R., Taccola, E., & Albertini, N. (2015). A hand-free solution for the interaction in an immersive virtual environment: the case of the agora of Segesta. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 40-5/W4*, 31–36. http://doi.org/10.5194/isprsarchives-XL-5-W4-31-2015

Ouyang, T., & Li, Y. (2012). Bootstrapping personal gesture shortcuts with the wisdom of the crowd and handwriting recognition. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2895–2904. http://doi.org/10.1145/2207676.2208695

Pietroni, E., & Pescarin, S. (2010). VR cooperative environments for the interpretation and reconstruction of the archaeological landscape. *Virtual Archaeology Review, 1*(2), 25–29. http://dx.doi.org/10.4995/var.2010.4680

Pietroni, E., & Rufa, C. (2012). Natural interaction in Virtual Environments for Cultural Heritage: Giotto in 3D and Etruscanning study case. *Virtual Archaeology Review*, *3*(7), 86–91. http://dx.doi.org/10.4995/var.2012.4394

Tanaka, A., Bau, O. & Mackay, W. (2013). The A20: Interactive instrument techniques for sonic design exploration. In: K. Franinović & S. Serafin (eds.), *Sonic Interaction Design* (pp. 255–270). Cambridge, MA: MIT Press.

Taylor, J. L. (2009). Proprioception. *Encyclopedia of Neuroscience* (pp. 1143–1149). Oxford: Academic Press. http://dx.doi.org/10.1016/B978-008045046-9.01907-0

Varona, J., Jaume-I-Capò, A., Gonzàlez, J., & Perales, F. J. (2009). Toward natural interaction through visual recognition of body gestures in real-time. *Interacting with Computers*, *21*(1-2), 3–10. http://doi.org/10.1016/j.intcom.2008.10.001

Wilson, A. D., & Bobick, A. F. (1999). Parametric hidden Markov models for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*(9), 884–900. http://doi.org/10.1109/34.790429

Wilson, A. D., & Bobick, A. F. (2000). *Realtime online adaptive gesture recognition. Proceedings 15th International Conference on Pattern Recognition.* http://doi.org/10.1109/ICPR.2000.905317

Wobbrock, J. O., Wilson, A. D., & Li, Y. (2007). Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. *Proceedings of the 20th annual ACM symposium on User Interface Software and Technology*, 159–168. http://doi.org/10.1145/1294211.1294238

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1083–1092. http://doi.org/10.1145/1518701.1518866