Ethno-Goggles: Supporting Field Capture of Qualitative Material

Paul Tennent, Andrew Crabtree, Chris Greenhalgh

Department of Computing Science,

University of Nottingham,

Nottingham, UK

{pxt, axc, cmg} @ cs.nott.ac.uk

Abstract. This paper discusses some early work in the creation of a system called Ethno-goggles, which is designed to support the capture of heterogeneous qualitative data in field studies. The aim of the system, which comprises several types of capture hardware, is to automate the process of digitization and synchronization of several different capture stream, thus reducing the logistical overhead involved in creating a corpus of effective field data for later analysis.

Introduction

When capturing field work, it has become common place to record potentially several types of media including video, audio, field notes, and more recently system log data (Crabtree et al., 2006). While each of these data may add value to an analysis, each may also add logistical overhead. Challenges include digitization of heterogeneous media; synchronization of that media; and attention-division during capture. (Crabtree et al., 2006) also present a number of challenges associated with capturing field experiments, specifically related to those experiments focusing on the analysis of mobile and ubiquitous computing systems. These challenges include:

- **Mobility**. Users of ubiquitous systems are often mobile. They move across extended physical areas, quickly at times, sometimes even running, which can make the documentation of action and capturing of video material difficult at best.

- **Small Displays**. Interaction frequently involves the use of small displays such as handheld computers and mobile phones. This makes it difficult to see users' interactions with the system.

- **Occlusion**. Closely related to the previous two issues, the situation of an interaction means that the environment, or even the users themselves will frequently occlude information that might otherwise be captured by video cameras. The same is true of audio cues, environmental noise can easily occlude key audio information.

- **Non-collocation**. When interacting with collaborative ubiquitous systems users are frequently interacting with other users who are not collocated with the user. Again this affects the analysts' ability to capture all the required information about the context of

the interaction, unless multiple analysts and capture devices are involved, and even then this presents the challenge of the synchronization of captured data.

- **Invisible interaction**. Users often interact with invisible sensor systems such as Global Positioning Systems or video tracking, which can make it challenging to understand why users are acting in a given way and how the sensing systems are actually behaving.

- **Distribution of Interaction**. Interaction may be distributed across different applications and devices. Interaction is thus not only located in different physical locations but may also be mediated through different applications and devices, which makes it difficult to develop a coherent description of interaction.

Here we present some early work on *Ethno-goggles*, a hardware addition to the maturing digital replay system (DRS) (Greenhalgh et al., 2007) to support the process of capture, and aims to address some of these issues. By combining the tool described here with existing DRS functionality, an analyst is supported through the entire process of capturing field work, through organizing her data, to analyzing them and producing effective output.

The ethno-goggles, comprise several pieces of hardware combined into a data collection toolkit. These include a glasses-mounted camera, in-ear microphones, a digital pen for recoding field notes, A GPS device, A laptop computer in a rucksack and finally a small PDA or mobile phone device for controlling the system remotely.

The paper will discuss both the technical aspects of the system: what it is and how it works, and some usage scenarios, along with the implications of those usages.

Attention Division

One issue with the process of capturing video for field work is attention. Let us assume for a moment that the analyst is working on the ethnomethodological tradition (Garfinkel, 1967). The task of performing an ethnography is a skilled process which requires a lot of attention. Indeed the very practice of ethnography is predicated on attention to detail. The result of this tends to be very low quality videos, because the ethnographer, as he should be, is focusing on the task of ethnography and not the video capture.

Ethno-goggles, by mounting its camera on one's glasses and its microphones in one's ears, takes the approach of "see-what-I-see" and "hear-what-I-hear", which should help to alleviate this issue. The effect of this is that the analyst, has access to a complete record of both what they looked at during the analysis and exactly what they heard. Of course the camera cannot perfectly match exactly what the looked at, as it simply captures at an angle of ninety two degrees to the front of the analyst's head, while human vision is closer to two hundred degrees horizontal and one hundred and thirty five degrees vertical. There is still therefore required some notion of actively pointing one's head at what one wishes to record. However because by looking through the lenses one is effectively looking through a viewfinder there is no need to divide attention a camera and the task in hand. Figure 1 shows the normal human field of vision.

Figure 1 – A representation of the fields of view of normal eyesight. The human sees roughly 200 degrees on the horizontal plane and 135 degrees in the vertical. The glasses mounted camera however manages only around 92 degrees in each plane, meaning that peripheral vision is largely lost from the capture. This issue is only counteracted by an awareness of the camera itself in the wearer.

The microphones are slightly more simple. They have a recording range similar to that of the human ear, and being located in-ear, they will capture the analyst's aural experience completely.

Digitization and Synchronization

Papers such as (Tennent & Chalmers, 2005) have explored methods of synchronizing data after an experiment has taken place. The problem with synchronizing multiple heterogeneous media types if that after clips are digitized and stored on a computer, they lose their reference to a real, or absolute time and become simply durations.

An event that took place at 13:05 on November 5th 2007, and was captured on a clip that started at 12:55 on the same date, is referenced as occurring 10 minutes into the video, rather than at a particular time. A second video of the same event that started 5 seconds before, would have the event at 9.55. This in itself is challenging, Now consider synchronizing that with a sensor log that started at 12:57:15 (Sensor logs often have a very high frequency, so times can be extremely accurate). On top of that we may have one or more audio streams, and possibly any number of other media types. So the same event is happening at different points in different media clips. This scenario may seem needlessly convoluted, but it is not uncommon. DRS provides tools for laying media clips on a reference timeline, allowing the user to effectively synchronize them, but the process is still in essence manual.

Any synchronization method that requires manual synchronization is time consuming and subject to human error. Add to that the lengthy process of actually digitizing the data in the first place, along with transcoding formats into some kind of reliable and usable standard, and an analyst has already wasted a impressive amount of time before even beginning to explore the data. This is one of the most fundamental issues of data capture that has evolved out of the increasing use of digital media that the ethno-goggles project aims to address.

Crucially, with ethno-goggles, everything is connected to a running version of DRS, which captures, digitizes and synchronizes in real time, meaning that there is no time wasted on the collation process. The upshot of this process is that when the analyst begins to review his data, they are already fully synchronized and in a suitable form within DRS for him to begin

exploring them. Of course it is likely that additional resources may be required - extra video footage from other cameras, logs from the system being analyzed etc., however by using ethno-goggles we have a good start for the process. Also because the capture toolkit is WiFi enabled, and modular in nature, it is possible to automatically send system logs etc to the recorder, if the system being analyzed is aware of the device.

Another major synchronization challenge is that of the notebook. Traditionally when recording notes, the recoded data becomes temporally flat - that is, while notes may (though not always) follow each other in correct chronological order, there is typically no more accurate temporal information provided. While an analyst may scribble down times while taking the notes, these lack any major accuracy and again, will add to the logistical overhead of the capture. Indeed field notes present a second problem of digitization. In order to include them in a corpus of digital data, they must either be transcribed or scanned - both of these are time consuming practices. Ethno-goggles circumvents both these problems by digitizing and synchronizing what the analyst writes as it is written. DRS provides a special viewer for examining the output of the pen, in synchrony with any other recorded media. The output is recorded graphically, so without performing some form of OCR on the handwriting, it is not possible to textually search within it, however, it does provide a similar effect to a scanned notebook, with the additional benefit of temporal synchronization.

The use of GPS as part of the system also allows support for spatial synchronization. A suitable viewer within DRS would allow an analyst to select data based on the location it was collected rather than the time at which the were collected. One example of the use of this feature would be the construction of simple spatial distributions of events.

Capturing a User's Experience

Up till now we have explored only the process of capturing an analyst's experience, but a similar collection of equipment can easily be used to capture an accurate picture of an experimental participant's experience. While in the user's case there my be less need for an electronic notepad, there is certainly an argument for capturing their audio/visual sensory experience using the glasses and microphones and indeed in some cases GPS and or additional system logs.

Ethno-goggles is constructed in a modular fashion allowing additional sensors, or anything that generates live output to be attached to the system and automatically synchronized. Support for WiFi, BlueTooth and USB should ensure a strong range of compatibility with other sensors, with DRS's already strong log-translation components used to generate suitable databases, and its forthcoming log visualization tools to generate suitable views over those databases. So we have a system which supports multiple configurations of hardware, all working on a simple premise, record it and record when you recorded it, thus automating the often complex synchronization issues.

Hardware

Ethno-goggles is a collection of hardware designed to capture as much data about the analysts' experience as possible. At its heart is a rucksack-mounted laptop, running in clamshell mode and running the DRS Software. WiFi provides a communication medium for those devices which have the capability, with all others wired directly in. It is designed to be as unobtrusive as possible, using 'spy' like equipment as capture devices in an attempt to minimize the potential effect of the analyst on the participants' behavior, as well as improve the synchrony between what the analyst sees and hears and what is recorded. Figure 2 shows a representation of the system.

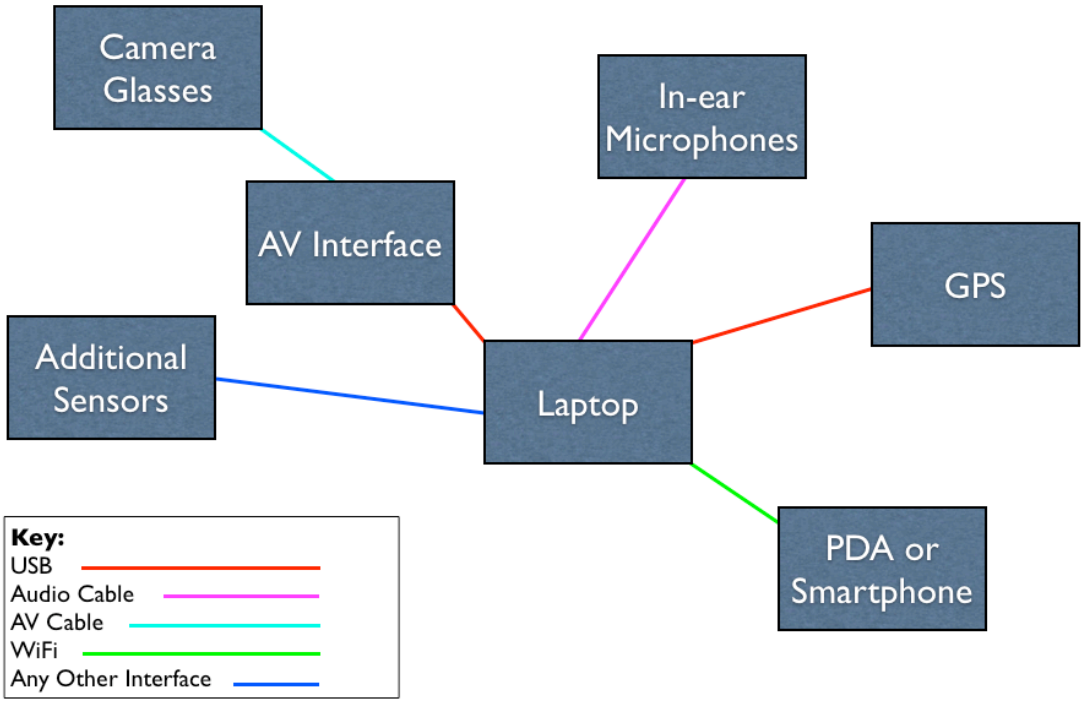

Figure 2 – A representation of the hardware architecture of the Ethno-goggles system.

Camera

Ethno-goggles gets its name from its signature hardware device: a glasses mounted camera (figure 3). Hailing from the world of espionage, the benefit of this device is simple; the analyst does not have to make a choice between looking at a camera and focusing on their analysis. Of course the analyst will usually be focusing on the subject rather than the camera. An examination of the video recoded during the system trials of Uncle Roy All Around You (Flintham et al., 2003) shows a generally poor quality of filming. Ethno goggles films are certainly equally, if not more shaky, but they do at least tend to focus on the subject rather than the ground.

Figure 3 – The glasses mounted camera used in the ethno goggles system. As can be seen here, the glasses are not overly intrusive.

Microphones

The second capture device incorporated into ethno goggles is a set of in-ear stereo microphones (figure 4). The principle here is to record exactly what the analyst hears. When recoding with a traditional camera microphone, even an omnidirectional one it is easy to miss things simply because of the positioning of the camera. Using in-ear microphones we capture everything that the analyst hears, this should help to marry the analysts remembered experience with the recording of the experience on which he will predicate his discussion.

Figure 4 – In ear microphones. These look similar to normal in-ear ‘bud’ style headphones, but provide outward facing microphones. An internal audio pass-through means they do not serve as a barrier to normal hearing.

Data Pad

The data pad is based on Nokia's SU-1B digital pen (Figure 5), and a pad of special paper. The paper is printed with microdots, read by a camera in the pen which allows it to establish exactly where on the paper it is at any given time. A Bluetooth connection is used allowing DRS to monitor and record every pen stroke and subsequently reconstruct the notes taken along with temporal information about when they were taken, allowing the analyst to easily find notes associated with a particular event or slice of time. Each individual note (delimited by a pause between pen strokes) can also be assigned metadata, and annotations during later analysis.

Figure 5 – The Nokia SU-1B digital pen used in the data pad component. It is slightly large and unwieldy for a pen, but usable nonetheless.

Remote Control

One crucial area of the system is control. With the laptop controlling the system packed in a rucksack, along with the necessary cabling and hardware interfaces, it is simply not practical to use it for controlling the system. Instead a mobile device is used. Rather than depending on one specific piece of hardware to achieve this, DRS when in record mode generates a small HTTP server, creating pages designed to be accessed on a small screen. This allows either a PDA or HTML capable smartphone, such as an iPhone to act as a remote control for the system. The interface displays battery information for the laptop, along with buttons to start and stop recording, and to test the input from each individual component.

Location Awareness

The system includes a GPS device, logged by the application. This allows location traces for the analyst to be recorded. One benefit of this is that if a mobile application is being analyzed, even if it does not itself record location traces, the analyst's location is recorded and can therefore be used to infer some location information for the participants, or indeed if the user is wearing the equipment, their location can be measured.

Additional Sensors

GPS is one type of sensor included in the system, but like DRS in general, ethno-goggles is designed to handle heterogeneous data. It is difficult to predict what type of sensors may be attached to the system, but an extension of DRS's log file workbench means that as long as a device is producing textual logs, those logs can be automatically included, stored, synchronized and represented by DRS.

Potential Issues

The most obvious issue raised by a system such as this is battery life. While it may be trivial to change the battery in a handheld camera, many of the components of ethno-goggles contain independent batteries, and those that don't represent a draw on the battery of the laptop. The average lifespan of a small form-factor laptop battery is around three hours, and with the additional draw, and heavy computation from recording, the expected lifespan of the system is around 2 hours. Indicating the battery life on the remote control interface is crucial, as is the need for audio cues from that device when the battery is running low, as the system must be correctly shut down to avoid possible data corruption.

Another possible issue with this system is the potential instability of the camera image, because of the way people move their heads around. One possible approach to solving some of the camera instability issues may be to use a gyroscope mounted on the glasses and post process the video to compensate for actions like head tilting. A system of autoclassification such as that described for Replayer (Morrison et al., 2007) could also serve to exclude unusable areas of the video.

Early Use

While Ethno-goggles is still very early in its development stages, some initial tests on the hardware have been carried out, specifically relating to the glasses mounted camera, to determine whether or not the videos recorded by them will be acceptable for use in analysis. At present videos are being recoded from moth analysts and users of the *Rider Spoke* system (http://www.blasttheory.co.uk/bt/work_rider_spoke.html). Tests have also been carried out in preparation for the use of ethno-goggles in Nottingham University's English department in the construction of 'A Day in the Life of a Language.' A corpus of real language use in real-world situations.

Conclusion

In the introduction a number of challenges related to studying mobile computing systems were outlined. Ethno-goggles aims to counteract these issues in the following ways:

- **Mobility**. Even when running, a user of ethno-goggles will focus their head primarily on what they are supposed to be looking at. This is significantly easier to achieve than the same affect with a handheld camera.

- **Small Displays**. If the user is equipped with the ethnogoggles system then this issue can to some extent be negated. In order to interact even with a small display, the user must typically look at it. While there may be a problem with fields of vision not quite

matching a user who is aware of the recording system may actively counteract this problem.

- **Occlusion**. Again, equipping the user or analyst with this system will serve to reduce instances of occlusion more than simply pointing a handheld video camera. One is less likely to tolerate occlusion of ones eyesight than of a handheld camera.

- **Non-collocation**. This particular issue is more correctly circumvented by DRS itself, rather than specifically by the ethno-goggles. Post hoc collation of multiple synchronized datasets can allow the analyst to assemble non-collocated synchronized media streams for a more complete view over the data.

- **Invisible interaction** and **Distribution of Interaction**. The modular facility of Ethno-goggles to record different types of data from different devices and combine them in DRS with system logs servers to counteract the difficulties associated with each of the challenges.

While the ethno-goggles project requires specific hardware to use, however the benefits to the analyst are potentially high: removing the need for the time consuming tasks of hand synchronization and digitization of media files. The synchronized field notes may also help to reduce workload, and aid memory by displaying notes during playback at the time they were taken. and thus free up time to focus on the analytical rather than logistical tasks of analysis.

References

Andy Crabtree, Steve Benford, Chris Greenhalgh, Paul Tennent, Matthew Chalmers, and Barry Brown. Supporting ethnographic studies of ubiquitous computing in the wild. In DIS '06: Proceedings of the 6th conference on Designing Interactive systems, pages 60–69, New York, NY, USA, 2006. ACM.

Martin Flintham, Rob Anastasi, Steve Benford, Adam Drozd, James Mathrick, Duncan Rowland, Amanda Oldroyd, Jon Sutton, Nick Tandavanitj, Matt Adams, and Ju Row-Farr. Uncle roy all around you: mixing games and theatre on the city streets. In DIGRA Conf., 2003.

H. Garfinkel. Studies in Ethnomethodology. Prentice-Hall, Englewood Cliffs, NJ, 1967.

Chris Greenhalgh, Andy French, Paul Tennent, Jan Humble, and Andy Crabtree. From replaytool to digital replay system. In (ICESS2007) 3rd International Conference on e-Social Science, 2007.

Alistair Morrison, Paul Tennent, John Williamson, and Matthew Chalmers. Using location, bearing and motion data to filter video and system logs. Pervasive Computing, pages 109–126, 2007.

Paul Tennent and Matthew Chalmers. Recording and understanding mobile people and mobile technology. In (ICESS2005) 1st International Conference on e-Social Science, 2005.