Media Mapping:
Using Georeferenced Images and Audio to provide supporting information for the Analysis of Environmental Sensor Datasets.

Abstract
Field based environmental monitoring projects often fail to gather supporting temporal information on the surroundings, yet these external factors may play a significant part in understanding variations in the collected datasets. For example when sampling air quality the values may change as a result of a bus passing the sampling point, yet this temporal local information is difficult to capture at a consistently high resolution over extended time periods. Here we develop an application which runs on a mobile phone able to capture visual and audio data with corresponding time and location details. We also develop a desktop analysis tool which synchronises the display of this dataset with those captured from environmental sensors. The result is a tool able to assist researchers in understanding local changes in environmental datasets as a result of changes in the nearby surrounding environment.

1. Introduction

The analysis of temporal datasets in Geographic Information Systems (GIS) is often hampered by a lack of supporting relevant information on local conditions at the time of data capture. Being able to explain unpredictable variations in temporal datasets may depend on being able to understand the nature of the local environment at a very local scale. For example a passing vehicle may be the cause of a noted spike in airborne particulate matter, but unless this situational information is recorded the spike may never be explicitly explained. Sensor networks are able to supply background information revealing the wider situation, but a co-located synchronized set of sensors are required to understand the local situation during
mobile data capture. Here we develop an application able to assist researchers in storing information on the local environment at the time of data capture.

The solution uses a mobile phone to store audio, visual, and location details against time such that during analysis the researchers are able to view the local environment at the time of data capture. A custom playback and analysis tool was also developed to combine this situational data with other time stamped data captured from the same location, allowing researchers fast access to the relevant contextual information during the analysis of the environmental data. In this paper we describe the function of the mobile and desktop applications, their design, and report on their usefulness in an air pollution monitoring study conducted in an urban area. The tools proved useful in explaining local spikes in air pollution data due to local events documented in the supporting images and audio data. The audio stream was also useful for allowing the researcher to take spatially and temporally attributed verbal notes in the field.

2. Background

Many studies have examined the relationship between air pollution and modes of urban commuting (Fruin et al. 2008, Briggs et al. 2008, O'Donoghue et al. 2007, Gulliver and Briggs 2007, Van Roosbroeck et al. 2006, Kingham et al. 1998). Air pollution data is highly temporal, changing across time and space, affected by global and local events. Local events, such as a bus passing the recording equipment, are hard to document and traditionally paper based records are kept. However manual references to such events are hard to integrate into any analysis, and sampling frequency is often inconsistent. Yet these factors are important as has been strongly argued by Briggs et al (2008) who state that local factors could be the cause of differences reported between studies including such things as “building configuration, road layout, monitoring methods, averaging periods, season, meteorological conditions, vehicle, driving and walking behaviours, and the strength of in-vehicle sources” (Briggs et al. 2008, 20).

To provide a better understanding of the surrounding environment at the time of data capture a second set of sensors can be used to automatically capture contextual information. This contextual information is not the primary data for research purposes, but a supporting dataset for the analysis phase. Technological advancements have made it possible to sense the environment more accurately, at higher sampling densities than ever before. Sensor networks may consist of electronic devices (Culler and Mulder 2004, Microsoft Corporation 2006), or citizens volunteering local environmental information via the internet (Goodchild 2007). En mass citizens may provide data without realizing it, such as the monitoring of mobile phones to estimate population movement (Ratti et al. 2006), or to estimate travel delays on motorway sections (Astarita et al. 2006).

The majority of electronic wireless sensor networks are static, distributed across a region at fixed sites, feeding information to a central facility which com-
bines the data to build a picture of the surrounding conditions. For the purpose of mapping a commuter’s exposure to air pollution the local level changes are also important, therefore a set of sensors should remain co-located with the air pollution sampling equipment. The data samples must also be at a high enough temporal resolution to record significant local events, and the data streams need to remain synchronized during data capture.

Mobile computing devices and smartphones have been proven useful in environmental monitoring enabling participants to collect and share data in realtime (Rudman et al. 2005). The MESSAGE consortium (Polak and Hoose 2008) have undertaken a number of projects using mobile phones as personal environmental sensors and data loggers. The mobile phones were equipped with a payload of environmental sensors able to record carbon monoxide, carbon dioxide, traffic volume, and nitrogen dioxide levels. Researchers exploring the city could feed data in real time to a data centre for processing, revealing current city wide air pollution trends. The individual trip data could also be replayed and mapped so that air pollution trails could be reviewed to visualize areas of poor air quality in the city. However a key aspect missing from this research was the facility to store video or image data from the user’s surroundings. Therefore any analysis carried out at a later date would lack documented contextual detail. Although audio was used, it only provided an estimate of traffic volume and could not differentiate between vehicle types, or allow field researchers to take temporally and spatially attributed verbal notes.

Another research group developed a prototype system known as GeoMobSense (Kanjo et al. 2007). This toolkit allows private users to equip their own phones with the necessary facilities to log data from connected sensors. The phones themselves are used to display information, as well as log sound levels, while separate data loggers are used to record the environmental data. The resulting datasets can be exported and displayed on Google Earth, and other GIS applications. Again this toolkit fails to store continuous image sequences, or to save a spatially attributed audio file. Therefore any post-capture analysis is hampered by a lack of documentation on the surrounding situation during the field study.

Multimedia files provide a useful companion dataset to prompt recall of events which occurred during data capture. They provide an extra channel of information useful when linked to GISs (Cartwright et al. 2007), however the video, image, or sound files are normally linked to a point, as in Media Mapper (Red Hen Systems 2009). This creates a one way relationship, only allowing the corresponding multimedia clip to be found when the user clicks on a map location, essentially using the map as document retrieval interface. The content of the multimedia file itself is not spatially attributed, and the ability to jump to the corresponding position in the video file for a given map location has to be performed manually.

There have been a number of attempts to more closely link the multimedia content to space through dynamically geo-referencing multimedia files. Spatial information is encoded into the file through an appropriate technique such that at any point in the video or audio the corresponding location may be referenced. One
technique to do this is to use the audio channel in a digital video camera to store digital sound data which composes of location information. This requires specialist equipment able to turn GPS location information into audio data, in a similar way that a modem is able to turn computer data into audio to send it across a telephone line. An example of this technology is CamNav Mapper (Blueglen Ltd 2009). This allows a user to search through a video file, and at all times be able to display the corresponding recording location in a GIS. However the connection is uni-directional, meaning the video is able to provide location information to the GIS, but it is not possible to initiate a search for the corresponding part in the video from the GIS.

Jaejun (2002) developed an application which supports bi-directional searching, permitting the user to find relevant video information from selecting a GIS location, or for finding the filming location by searching the video file. Similarly Zeiner (2005) developed an application able to fuse GPS location and video using timestamp information collected from synchronized clocks. For still images a set of points are created in the GIS, however for video recorded while moving a track denotes which parts of the video correspond to which geographic location. They also explore the use of data standards in providing geo-multimedia tools via the World Wide Web, with particular focus on the overlap between web mapping standards, metadata standards, and video streaming standards. These tools are not however designed with the ability to integrate other temporal datasets such as required for environmental research.

Other studies have used or developed analysis tools to visualize environmental datasets with local situational data (Kaur et al. 2006, Terwoert 2009, Arnold et al. 2004). However while these often include the ability to link photographic images with environmental data, they appear to lack a tightly integrated mapping facility.

For our research the requirement was for an application which could provide a high level of integration between temporal multimedia and location datasets, with the ability to support additional datasets collected from synchronized sensors. The capture device needed to be small, lightweight, and mobile such that it could be carried by a pedestrian or cyclist easily for extended periods of time. The datasets for location, audio and image needed to be tightly coupled such that no synchronisation issues could occur during long field trials. The analysis tools needed to be easily operated by an untrained GIS user, such that they could search through the datasets to interact with any of the captured data streams while maintain sync with the other linked data sources. We therefore looked at developing the data logging tool on a GPS equipped smartphone, which is a highly portable programmable device available at low cost. As a result of using a single programmable device the GPS, audio, and image datasets are tightly coupled, removing the need to synchronize clocks, and guaranteeing data streams remain in sync indefinitely. Additionally the multimedia files are georeferenced at the time of data capture removing the need for any post-capture data processing. In contrast to other applications which use laptops to capture and process media, the smartphone approach offers a
robust, small, and very portable platform which may be easily carried by pedes-
trians. Finally our analysis tool supports a tri-directional search mechanism, such
that users may drive the search by moving through the audio media, mapping in-
terface, or by interrogating charts of the additional sensor data. This means the
user is able to easily capture and analyze data using any of three mediums (loca-
tion, time, graph value). In the next section we look in more detail at the applica-
tions developed during this research.

3. Application Development

In this study we developed two applications to assist in the process of recording
the surrounding situational conditions. The first application runs at the time of
data capture in the field on a smartphone equipped with Assisted GPS (A-GPS),
and stores both an audio and visual record of the surroundings. A-GPS is particu-
larly useful for urban based research as it provides a faster startup location solu-
tion throughout a greater range of urban environments, such that a position could
be found more quickly and maintained more consistently. Furthermore the phone
selected for this research had a high sensitivity GPS chipset, enabling locations to
be calculated across a high proportion of the city, including inside some single sto-
rey buildings. The second application developed for this research runs on a desk-
top computer and assembles independent data streams against a common timeline,
such that the user may easily browse through multiple datasets whilst maintaining
sync between them, at all times being able to refer to the corresponding situational
image and audio data. We discuss each application in more detail in Sections 3.1
and 3.2, starting with the mobile data capture tools.

3.1. Mobile Data Capture Application

The main design criterion for the data capture device was that it would be used
in urban studies everyday over an extended period of a few months. It therefore
had to be small, robust, light weight, offer a large data storage capacity, be able to
capture audio, imagery, and run on battery for at least 90 minutes to ensure an en-
tire urban commute could be captured in a single session. We decided to use a
Nokia N82 smartphone to carry out these tasks as they can be programmed easily
using the Python language, incorporate a high sensitivity A-GPS able to function
adequately in urban canyons, and have a high quality camera. Furthermore they
are able to use micro-SD cards for data storage, are smaller than any laptop or net-
book computer, and have good battery life.

Nokia Series 60 smartphones can be programmed in three main languages, which
are C, Java, and Python. Python is very suitable for rapid development and allows
the developer to access core phone hardware through supported Application Pro-
gram Interfaces (APIs). The hardware access required for this project included
GPS hardware, audio, and screen display. Our initially application was designed to record a continuous video and audio feed to the micro-SD card at 15 frames per second, while logging GPS locations every second. The phone supports the ability for Python applications to request the position of the current playhead in the video file during recording, enabling the application to log the GPS position information along with the current video position to ensure a tight coupling of the location and video data streams. The data capture application performed well, and as it used MP4 compression a full one hour video with GPS log files occupied only around 70MB. However, the continuous video capture depleted a fully charged battery in 60 minutes, and made the mobile phone run fairly hot. After discussions with the air pollution research team, we looked at an alternative solution to record still images at regular intervals.

The next iteration of our application, and the one used in field trials, records an audio file continuously but captures still images at the rate of 1 image every 3 seconds. The audio file is recorded to a WAV file at 8kHz, again the playhead position is stored with each GPS update such that for every image the location and position in the audio file is known explicitly.

To link the audio and image files with GPS, a common timeline primary key is required. GPS time was considered unsuitable for the base timeline as the user may lose the GPS signals when moving inside buildings. Therefore, the Python time from the phone clock was considered more reliable, and forms the baseline to which all other datasets are synced. GPS time is however stored as an additional attribute in the log file in case it is required later.

Specialist sensors were used to sample the air quality, recording the concentration of particulate matter at various sizes (PM$_{10}$, PM$_{2.5}$, and PM$_{1}$), ultrafine particles (UFP), temperature, and carbon monoxide levels. These devices all have internal clocks which were synced to the nearest second to the clock on the mobile phone before each journey. During transit, the mobile phone was mounted facing forwards on either the car dashboard, bike handlebars, or the strap of a rucksack to ensure the camera could capture a clear view of the oncoming route.

![Error! Reference source not found.](image) shows the data collected with each journey. The smartphone tags each image with a unique identification number based on the Python time in seconds, ensuring that images correspond to a single log entry. Log entries record the position of the playhead in the sound file, Python time, GPS time, and the GPS latitude and longitude. Additionally, we stored the cell tower identification value so that approximate locations can be determined if GPS positioning is lost. As well as recording GPS position, we also record speed, GPS accuracy, heading, the number of satellites visible, and number used for the position solution. This enables us to carry out analysis on the location accuracy if required at a later date.
3.2. Data Analysis and Playback

To be able to efficiently analyse the large volume of time series data collected it was necessary to build a custom application which allowed non-GIS users the ability to review the data easily from a single interface. To do this we wrote a custom application using C# .NET which made use of a number of open source libraries for charting, map display, and coordinate projection (Figure 2). One of the key criteria in our application design was that all the datasets should be synced and remain in sync across all dataset viewers while the user explored the data. For example selecting a map point should display the relevant air pollution data, the corresponding street image, and move the sound file playback to the correct location so any relevant audio notes and background noises could be heard. In the following section we discuss each of these data visualisation elements, and describe the methods through which the datasets are kept in sync.
3.2.1. Mapping

The mapping element makes use of the Piccolo Zoomable User Interface (ZUI) graphics framework. This open source library is not specifically designed for map display but offers powerful functions enabling rapid development of ZUIs. Piccolo is able to create scene graphs consisting of both vector and raster nodes, which may be easily animated to change location, shape or colour. It also provides the functionality to smoothly zoom and pan around the data, capturing mouse and keyboard input. Events can be assigned to graphical objects such that a user may interact with map items and trigger a custom action. In our application we use this functionality to link map objects to custom search functions, so for example clicking on a GPS track point moves the charting tool along to display the air quality values at that location, and moves the audio playhead forwards or backwards to the corresponding audio sample.

The mobile phone logs GPS data using the WGS84 coordinate system, however as Piccolo is unaware of geographic coordinate systems it requires projected datasets. We used New Zealand Map Grid (NZMG) as our projected coordinate system for all datasets. The background mapping layers included Quickbird satellite imagery, and Land Information New Zealand road centrelines transformed using GRASS to NZMG. The vector datasets, typically ESRI Shapefiles, were converted to a BNA text format using OGR libraries before being loaded and displayed by Piccolo, while the raster datasets (JPEG) were natively supported.
ESRI World files were required with each raster image to provide pixel resolution and location coordinates values, so that raster datasets could be loaded into the correct position.

The GPS data, once downloaded from the mobile application to the desktop application, is transformed into the NZMG coordinate system using the most accurate NTv2 projection transformations provided by the Proj4 libraries. These points were then displayed as Piccolo vector nodes over the base mapping, each node with a hidden tag holding the corresponding capture time primary key. This tag corresponds to the phone capture time in seconds, and is also recorded in the log file against the playback position and location details. These tag data enable the software to instantly locate corresponding relevant information in other datasets when a user clicks on a map object ensuring the system is very responsive.

3.2.2. Audio

The audio playback is handled using Microsoft DirectSound libraries, which enable the rapid development of software able to control audio datasets. Here we use basic playback control features (i.e. play, pause, forward, rewind), and the functionality to read the current playback position during playback. From this we can calculate the current playback time in seconds from the start of the sampling period, and therefore find the corresponding log entries which hold the geographic location and relevant image filenames. As the audio file is captured at 8000 samples per second it has the highest resolution of any of the datasets. Therefore any searches performed from the user cueing or reviewing this dataset require an additional step to find the most relevant (i.e. closest) timestamp in the log files. This was performed by simply ranking the difference in time from the audio position to all log entries, the first item being selected as the nearest. This functionality allows the user to review the audio file and also see corresponding imagery, location and chart data.

Due to memory issues when loading a 90 minute long audio file, a buffered playback method was required. Only a small section of the file is loaded into a memory buffer, and this buffer is constantly filled from the disk file as audio playback progresses. The performance impacts of this technique were minimal, and rapid audio reviewing and cueing are still possible.

3.2.3. Street Images

The smartphone captures an image every 3 seconds in JPEG format, and labels it with the appropriate timestamp. The log file holds a list of these timestamp filenames along with their corresponding location and other GPS details. When the user selects a location on the map the tag (with each Piccolo node object) holds the timestamp details, and therefore the corresponding image can instantly be loaded without performing any search other than for the filename in the file sys-
tem. When the user controls the audio playback the nearest timestamp information is found in the log file, and from this the corresponding picture name can be generated. Similarly when the user moves through the chart information the nearest timestamp in the log file is found and used to determine the appropriate image to display.

The image display is supported using native .NET libraries, with functionality to rotate the image sequence which is useful if the phone has been placed on its side during image capture.

3.2.4. Charting

The environmental sensor datasets are charted using the open source ZedGraph libraries. These provide sophisticated charting capabilities, and allow the programmer to link into many key events such as when the user pans across the chart, clicks in the charting area, or changes the scale on a chart axis. In our case the x-axis was allocated to time in seconds since the start of sampling. Ideally the smartphone would be turned on first to ensure the audio file timeline starts before the environmental data, but negative time values are also supported. The y-axis displays the values from the relevant sensor. The display automatically scales the y-axis according to the values in the entire loaded dataset, although manual scaling is also possible to zoom into values for smaller sample periods.

The ZedGraph library was used to create line graphs for each of the environmental sensor types. These can remain static to allow the researcher to view the entire datasets while replaying map, image and audio datasets. Alternatively the graphs can be dynamically linked to the playback such that they pan along the x-axis (time) automatically as the data log is replayed. In this case the current playback time is shown on the far left of the chart.

If the user clicks within the chart area then application retrieves the corresponding time (x-axis value) and moves the playback position to that value. This allows the researcher to instantly find the current geographical location for any spikes noted in the environmental datasets. Also as the audio and image files remain in sync the researcher is also able to look and listen to information from the surroundings at that point.

For reporting purposes the system supports high quality output of the graphs, by simply double clicking on them. There is also functionality to allow the user to export snapshots of any interesting results, effectively using this tool to produce a filtered dataset. To do this they simply click a button during playback and the GPS, picture link reference, time and date, and graph data are exported to a text file. As each environmental sensor operates at a different sampling frequency the application interpolates values between readings (i.e. straight line between known values). When the data is output the interpolated value is used if an actual reading value is not present for the current playback position.
4. **User Interaction**

One of the key differences in the analysis application developed for this research to those reviewed earlier, is the ability to initiate a search from any of three linked interfaces. Figure 3 summaries the processes required to provide this facility.

![Diagram of User Interaction via Audio, Map, or Chart Interfaces](image)

**Figure 3: User Interaction via Audio, Map, or Chart Interfaces**
During normal playback the system carries out an audio update event every 4000 samples (500ms). This ensures that the displays of each dataset remain in sync, without impacting the performance on slower computers.

In the next section we look at a number of examples which demonstrate the usefulness of this application

5. Examples

In the following section we look at data collected in air pollution studies from Christchurch, New Zealand. The figures illustrate how useful the contextual information was in explicitly explaining peaks in the air quality datasets. In Figure 4 we can see that the air quality spikes in the top left graph occur just after the bus (pictured) pulls away from the field observer. In addition the location information and map allow the researchers to easily and quickly identify where in the city these interesting results occurred.

![Figure 4: Analysis of Peak Resulting from Bus Pulling Away](image)

In the next example the air quality levels are much better while exploring a pedestrian precinct, with low particulate matter concentrations (Figure 5). However following this a sharp rise in particulate matter can be noted. By clicking on the spike in the graph display area the research is able to see, from the map location and supporting images, that this spike correlates to when the field observer entered a multi-storey car park.
In the final example data was collected while travelling on a bus around the city. A series of spikes can be noted in the carbon monoxide levels are fairly regularly spaced intervals. By reviewing the audio data, just before each spike, it was possible to hear the sounds of the door opening, ticket machine being operated and so on, and therefore conclude that these spikes correlated to door opening events.
6. Conclusion

In this research we have demonstrated that a smartphone may be used as a suitable data capture tool, able to accurately and consistently capture location attributed audio and visual data while operating in an urban space. The onboard Assisted-GPS (A-GPS) was able to rapidly locate the user when the device was first turned on, and maintain position throughout exploration of outdoor urban areas. The audio and image quality was suitable for later analysis to identify key events which may be associated with changes in the local environment as a result of short temporal events (such as a bus pulling out) or changes in the nature of the physical environment (such as entering a car park). Python provided an excellent programming language for rapid application development on mobile phone, with the necessary functionality to use the phone’s hardware such that explicit georeferenced image and audio could be processed on board the phone, rather than during later post-processing in the desktop environment. This ensures that data sync between audio, location, and imagery datasets may be maintained indefinitely. The smartphone platform also proved to be rugged enough for daily trials over a period of several months, and had battery and storage facility to cater for long urban commutes (tested up to 120 minutes).

We also demonstrated that a simple desktop application able to maintain sync be-
tween mapping, environmental, audio, and visual datasets proved useful in the analysis phase of the research. The tool allowed non-GIS researchers the functionality to easily explore environmental datasets while maintaining links to the corresponding contextual information.

The project was completed using a number of open source tools and libraries, without which the development would not have been possible in the allotted time or within budget. Future versions of the application should include the ability to store bookmarks against time, allowing the user to add keywords or notes. This would allow the annotation of findings, useful for collaborative work practices. Also more powerful query tools would be useful such that map points may be highlighted or hidden based on the closest environmental data. Finally the ability to search audio files for speech would be useful in longer sampling runs such that the system could automatically identify where audio notes have been taken.

7. Acknowledgements

The authors would like to acknowledge the contributions made by Kreepa Shrestha, and Woodrow Pattinson who carried out the extensive field trials. Also Justin Harrison for setting up and supporting the environmental sensor equipment. This research would not have been possible without funding support from the Geospatial Research Centre (NZ) and New Zealand Transport Agency.

8. References


