Combining Positional Information with Visual Media

Neill Campbell Henk Muller Cliff Randell

Department of Computer Science, University of Bristol, UK

cliff@cs.bris.ac.uk, http://wearables.cs.bris.ac.uk/

Abstract

By integrating visual media with positioning information obtained with our wearable computer, we create new opportunities for using visuals both in the field and at the workstation. The position information we store with each visual is direction, pitch, roll, location, focal length, and zoom. This information allows any system to reconstruct the frustum of the visual, and, if height data is available to reconstruct which parts of the earth are visible in the visual. This enables position based lookup and 3D mosaicing of visuals to reconstruct a 3D model.

1 Introduction

We are interested in the association of media files with contextual information gathered by our wearable computer. We have previously explored the combination of audio notes, or reminders, with locations [1]. In this paper we turn our attention to visuals.

Classic photographs and videos store only visual information. Recent advances in photo and video technology have increased the amount of information which can be stored with visuals. As an example, Advanced Photo System (APS) cameras store information about the lighting conditions on a magnetic strip with the photo, and digital cameras store focal length, and shutter time with each photo.

In this document we extend this idea, and annotate each visual (that is: each still, or each frame of a video) with positional information. We have supplemented our wearable computer system, which was already equipped with a GPS receiver, by mounting an electronic compass on a digital camera. The positional information obtained allows us to uniquely identify *where* a photograph was taken, and in which direction, thereby enabling us to develop applications which include indexing; and, in the near future, compositing; image reconstruction; and creation of 3D worlds.

The availability of images indexed by location to the users of wearable computers can provide them with different perspectives of the surrounding environment, and the development of a standard format also enables us to progress towards applications involving augmented reality.

2 Information stored with each visual

We have chosen a set of attributes which will enable us to precisely determine the theoretical field of view contained in each visual. In addition we have selected readily available data which may also have some relevance e.g.time/date. With each visual we thus store the following information:

- Absolute position (Latitude, Longitude, Altitude; from GPS location sensor)
- Direction (horizontal direction, pitch, roll from an electronic compass, or alternatively obtained from solid state accelerometers),
- Absolute Time/Date (also from GPS).
- Focal length and Depth of Field (which part of the scene is in focus),
- Zoom (the Field of View; the horizontal and vertical angle of the visible scene),
- Flash (whether the flash is used; indicates that only the first few metres are visible).

To store this data, we have extended the PNG format [2] with a chunk for position information. The chunk (tentatively named pOSd for position and direction), consisting of a 22 byte block storing the above information as fixed point numbers. This is more restricted and far simpler than the EXIF format [3].

3 Determining the frustum

The basic operation that can be performed using the positional information is to reconstruct the frustum of the camera. The frustum is the part of space which is covered on the visual. This process is shown in Figure 1: the absolute position of the camera (for example a triple X,Y,Z or latitude

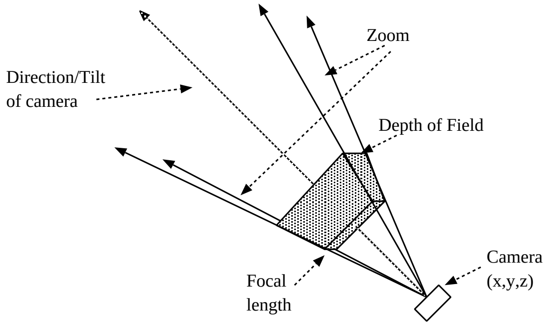

Figure 1. Determining the frustum

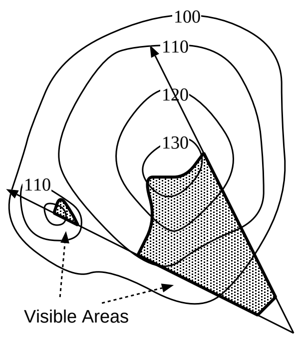

Figure 2. Clipping the frustum using a height map

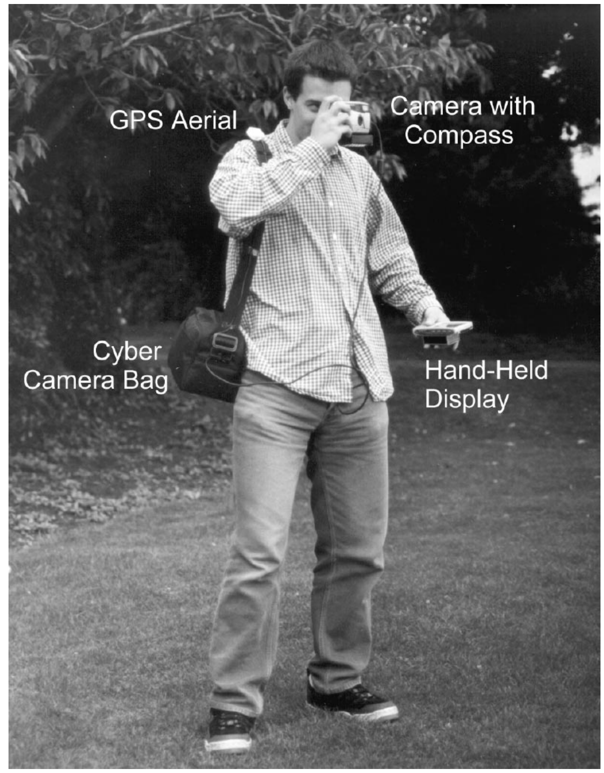

Figure 3. Using the 'Cyber Camera Bag'

longitude, altitude) determines where the photograph was taken from; the zoom, direction, pitch and roll information determines the field of view; and the focal length, with the depth of field, determines how much of the scene is on the photograph, in terms of depth.

The above process does not use any knowledge about the environment in which the photograph was taken. If more information is available, for example an altitude map of area where the photograph was taken, then the frustum can be clipped further. This is especially useful in outdoor scenes, where the depth of field is probably infinite. First of all, the frustum can be intersected with the elevation map, indicating which areas can be visible. In this case, the altitude map can be used to clip the visible area to exclude anything which is behind, say, a mountain range. Figure 2 shows how to clip the frustum with altitude information. The camera's absolute position together with an altitude map allows us to calculate which parts of the map are visible. We cannot do a precise clip, for parts of this area may be occluded by objects which are not on the map, for example a building, tree or person in the field of vision.

4 Configuration

For development purposes our wearable computer (a PC104 based 486, GPS Receiver with differential correction, and LiIon batteries) was installed in a "Cyber Camera Bag" (a heavy version of our unconsciously wearable "Cyberjacket"). This is worn by the photographer in a similar manner to a conventional camera bag into which a digital camera with electronic compass is connected. Figure 3 shows the equipment in use.

Focal length, Zoom, and Flash are already provided by existing digital cameras [4]. We retrieve the geographic position and magnetic variation from the GPS receiver. The prototype digital compass uses Hall Effect sensors and is mounted on the camera to retrieve directional information. Pitch and roll information will be supplied by three accelerometers.

Images are either stored in the camera, or added to the database on a remote server using a GSM 'phone link. A wireless hand-held web browser display [5] can be used to remotely search for, and display, the images also using the 'phone link.

5 Applications

5.1 Positional photo album

The first application is to create a photo album database with position based lookup. All frustums are stored, and given a location on earth you can query all photographs covering that point. Using the camera position data we are able to select the particular images which interest us. This allows the user to quickly collect, for example:

- photographs which show the whole of Stonehenge, yet exclude all the photographs taken *from* Stonehenge;
- photographs of the south face of the Everest - only those taken from the south.

A simple version of the photo-album can be stored locally on our wearable computer by downloading the images of the area we are interested in visiting. We are also currently investigating a web based version where an alta-vista type robot could create a location based index of photographs which can be searched remotely using the hand-held web browser.

The photo album stores the information by calculating the frustum for each photo (as outlined before), and storing it in a quad-tree. The root of the quad-tree spans the whole of the earth (using latitude and longitude information). For as long as the photo spans few cells in the quad tree, the quad tree is extended. Eventually, the frustum will be approximated by a number of squares in the quad-tree. The user can navigate by pointing at an area of a map, the software will navigate through the quad-tree and select appropriate photographs. (An octree would allow height information to be stored, but this adds little information.)

5.2 Intelligent compositing

The second application will be to give feedback to the wearer while taking the photograph, and to improve the composition of the photograph. The wearable will know where the horizon and sun are, and can therefore instruct the wearer that the photograph may be better if taken from a different angle. Alternatively such feedback could be used to automatically control the camera e.g. with a fill-in flash when facing towards the sun.

5.3 Reconstructing 3D images

A more challenging application which we are investigating at the moment is the reconstruction of a 3D model of the environment using the images collected. This 3D model can be used to create a virtual walk-through [6, 7].

Purely vision based 3D reconstruction techniques are now emerging [8]. A set of 2D images of an object are used to build up a 3D model. This reconstruction process requires only approximate positional information of the camera; the precise position and direction of the camera are calculated as part of the reconstruction process. We may be able to construct a mountain range given a set of photographs.

A simpler reconstruction technique will use an elevation map of the area concerned. The photograph can be analysed to produce a texture map, and the height map provides the 3D info. Although neither the position information nor the directional information will be a perfect match, we expect that we will still be able to reconstruct the original image and potentially create images from other viewpoints.

6 Other Formats

Positional information can also be stored with photographs from non-digital cameras or videos. It is possible to store the additional information on the magnetic strip of an APS camera. The incorporation of the 22 bytes of data into each frame of a video sequence could provide the basis of a system to enable the viewer to track the image source in real time. If multiple video sources are available this system could be used by a viewer to select different viewpoints of live action e.g. a football match.

References

[1] R. Hull, P. Neaves, and J. Bedford-Roberts. Towards situated computing. In *Proceedings of The First International Symposium on Wearable Computers*, pages 146–153, October 1997.

[2] T. Boutel. PNG (portable network graphics) specification. *http://www.w3.org/TR/REC-png.html*, October 1996.

[3] ISO technical committee 42. Photography - electronic still picture cameras - removable memory - part 2: Image data format - TIFF/EP. Technical Report ISO/DIS 12234-2, ISO, Nov. 1997.

[4] K. Murata. Kodak digital camera. *Software Developer's Kit Reference Manual*, pages 40,50,120, April 1997.

[5] P. Neaves and J. Bedford-Roberts. Dynamic connection of wearable computers to companion devices using near-field radio. In *Proceedings of The Second International Symposium on Wearable Computers*, pages 156–157, October 1998.

[6] L.-T. Cheng and J. Robinson. Dealing with speed and robustness issues for video-based registration on a wearable computing platform. *Proceedings of The Second International Symposium on Wearable Computers*, October 1998.

[7] S. Mann. Lookpainting: Towards developing a new camera user-interface. *http://wearcam.org/monet/node8.html*, April 1999.

[8] D.P. Gibson, N. Campbell and B. Thomas. The generation of 3D models without camera calibration. *Computer Graphics and Imaging*, pages 146–149, June 1998.