Geobrowsing the Globe:
A Geovisual Analysis of Google Earth Usage

Pablo Abend, Tristan Thielmann, Ralph Ewerth, Dominik Seiler, Markus Mühling,
Jörg Döring, Manfred Grauer, and Bernd Freisleben

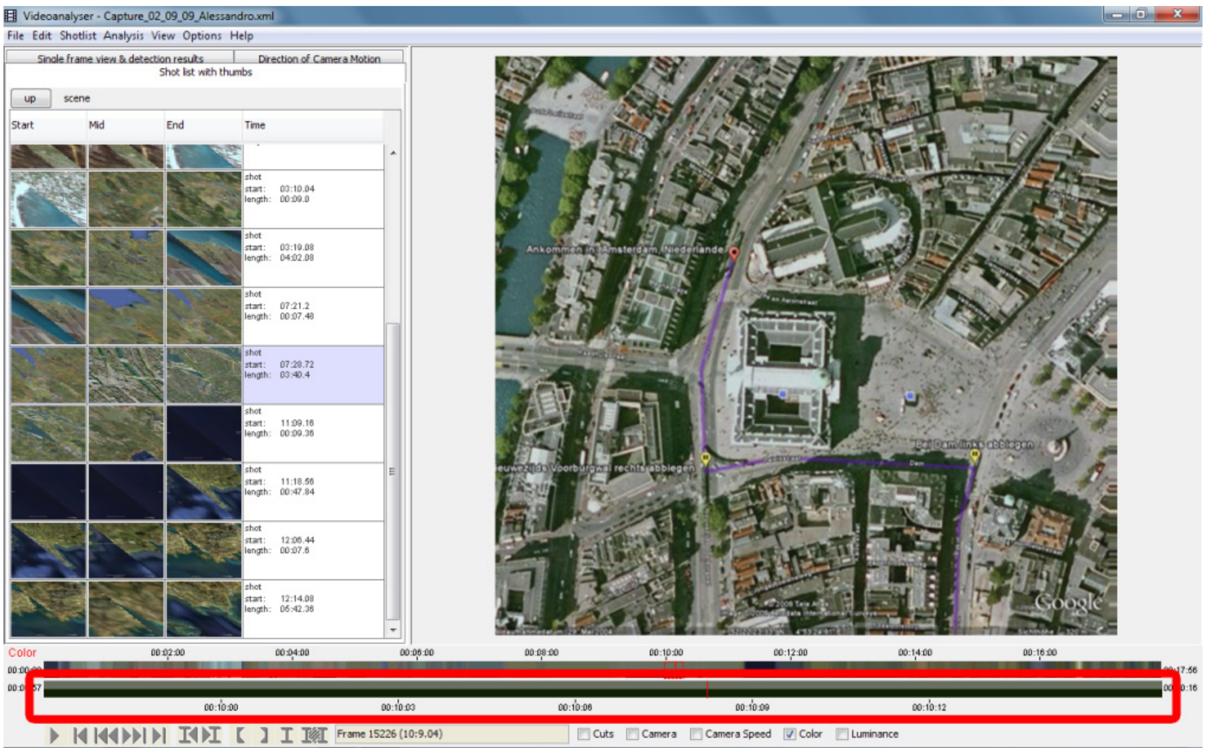

Fig. 1. The city of Amsterdam. Capture of free geobrowsing (male, 30 years), Sept. 2nd 2009, 09:57 min/17:56 min.

Abstract— In this paper, a semantic approach to the analysis of the recorded on-screen navigation within virtual globes is presented using the example of Google Earth. In order to explore and visualize geobrowsing behaviour systematically, we have extended the video analysis software *Videana* for the analysis of Google Earth tours. The software's functionality comprises the detection of 'text bubbles', the visualization of dominant/average colour values, and the allocation of 'virtual camera' movements. On the basis of a multiple case study this paper demonstrates that on-screen navigation behaviour is largely defined by the morphology of the landscape and, to a lesser extent, by the navigational aids and the additional multimedia information provided. Top view and orientation towards True North are most often retained. Users generally prefer satellite views rich in contrast where they can identify map contours. Thus, an established form of map use exists that has also been applied to virtual globes.

Index Terms—Geovisual analytics, Google Earth tours, semantic video analysis, geobrowsing, and virtual globes.

1 Introduction

The design of geovisualization tools that support the analysis of the everyday usage of geobrowsers such as Google Earth has been neglected thus far, although the exposure of easily handled Web Mapping 2.0 [1] in times of sophisticated multimedia cartography and 3D virtual worlds poses new challenges to the empirical usage evaluation of geographic visualizations [2].

- *Pablo Abend, Tristan Thielmann, and Jörg Döring are with the DFG Research Project "Media Geography" at the University of Siegen, E-Mail: info@mediengeographie.de.*
- *Ralph Ewerth[1], Markus Mühling, and Bernd Freisleben[1] are with the University of Marburg, Dominik Seiler and Manfred Grauer[1] are with the University of Siegen. Email: {ewerth, muehling, freisleb}@informatik.uni-marburg.de and {d.seiler, grauer}@fb5.uni-siegen.de*

  [1]Members of DFG Research Project "Methods and Tools for Computer-Based Media Analysis"

For this reason, based on a multiple case study, this paper will focus on (semi-)automatic methods of "geovisual analytics" [3] that can be used to gain new insights into how people navigate on-screen. In accordance with Peuquet and Kraak [4], this usage form of geographic knowledge depositories by laymen can be called 'geobrowsing'.

With the emerging Geospatial Web [5] it seems that a new type of "armchair geographer" [6] is using nearly immersive hybrids between maps and new media applications to explore the globe virtually. However, is this really revolutionizing the possibilities of ordinary maps and globes [7]? Moreover, to what extent can we distinguish between "conventional vs. new maps" [8] that use elements from the 'real world' to give a 'sense of place' at all? These questions will be addressed within this paper.

2 Research method and technique: a software for analyzing Google Earth tours

The investigation of Google Earth itself using film and media scientific methods is an obvious approach if we take seriously the diversity of navigational options and changes in perspective that the "virtual camera" [9] provides in geobrowsers, especially since the data obtained has the form of tours that can either be generated with the on-board functionality of the "Movie Maker" tool in Google Earth or with the help of screen-capturing software. These user-generated movies are increasingly found on video-sharing platforms like YouTube. The search term "Google Earth tours" returned 1.400 Google search results on October 13th 2010.

In previous work [10], the software *Videana* has been described as a development to relieve media scholars from the time-consuming task of annotating videos and films manually. The software provides a number of methods for analyzing videos automatically, including algorithms for shot boundary detection ("cut" detection), camera motion estimation, face detection, and text localization and OCR (optical character recognition). The graphical user interface of *Videana* presents the detection results in separate timelines (Fig. 1). In order to support the analysis of Google Earth tours, we have extended *Videana* with the following algorithms to detect semantic objects and events that occur in Google Earth videos:

- text bubble detection,
- dominant colour detection,

Basically, some geobrowsing information of a Google Earth tour can be extracted from related KML or KMZ files (Keyhole Markup Language, http://www.opengeospatial.org/standards/kml), e.g. camera motion information. KML is a format used to describe geobrowsing data at the client's site for the software Google Earth and Google Maps. It is a standard of the Open Geospatial Consortium. But KML files do not contain all information that is required for the research addressed by the current work. For example, these files do not store the data indicating when a user opens a Google Earth bubble in order to obtain detailed information about a place the user has visited. Moreover, the metadata provided in these files 'only' specify graphical representations on the base map (3D models, polygons, images, text information, etc.) and the general map view and appearance (tilt, heading etc.), but not the map itself [12].

The algorithms that are relevant for analyzing Google Earth tours are described briefly below:

2.1 Text bubble detection

Google text bubbles stand out from the background by their monochrome colour. They are detected by contour processing, which is a useful tool for shape analysis. Therefore, the image is transformed into a binary representation using a predefined high threshold, in order to separate white areas from the rest of the image. We use the border following algorithm (contour processing) of Intel's Open Source Computer Vision Library (http://sourceforge.net/projects/opencvlibrary) to assemble edge pixels into contours. The shapes of the resulting hierarchy of contours, basically being sequences of points, are analysed in the following way. First, it is assured that the contour is closed. Then, it is verified that the size of the contour area exceeds an adaptive threshold and that the line segments of the contour are either parallel to the x- or y-axis. Furthermore, we search in the upper right corner of the current contour for a 'hole contour' representing the close button. If all of these conditions apply, the corresponding shape is accepted as a Google text bubble.

2.2 Colour detection

Average and dominant colour detection can help to identify colour contrast as an important component of map complexity [13]. *Videana* provides two additional time lines and four different diagram types in order to visualize colour and brightness information for a video. The luminance information can be directly derived from the colour space (YCbCr, Y: luminance, Cb and Cr: colour information for blue and red) used by the MPEG decoder. The frame data is converted from YCbCr colour space to RGB (red, green, blue) colour space. Each colour channel is quantized to 16 (2^4) levels, and a colour histogram is created for a frame that consists of 4096 (2^{12}) bins. The dominant colour in a frame is determined by the colour that is related to the bin with maximum frequency.

Colour, as well as the luminance information, is visually represented by the corresponding time line synchronized with the video frames. The colour video time line is divided into two parts; the upper part represents the mean RGB colour value, while the lower part represents the determined dominant colour. In addition, colour and luminance can be visualized as diagrams. Currently, four different diagrams are available: mean and variance of frames, mean RGB value and the dominant colour of frames.

2.3 Camera motion detection

Camera motion is often used as an expressive element in film production. There are different types of camera motion: rotation around one of the three axes, translation along the x- and y-axis, and zooming in and out can be considered as equivalent to translation along the z-axis. When rotational and translational motion is summarized, a simple distinction can be made between horizontal motion (pan), vertical motion (tilt), and zoom. In Google Earth videos, ('virtual camera') motion is directly related to the browsing activity of the user, and zooming in and out corresponds to the level of detail in the information that is needed by the user.

The approach for the detection of camera motion [11] uses motion vectors encoded in MPEG videos. The advantage of using these vectors is that the runtime for the extraction of these motion vectors is very low compared to the decoding of a whole image and the calculation of the optical flow field, i.e., the calculation of motion for each pixel. Unreliable motion vectors are removed by an effective method in a pre-processing step, called outlier removal. The parameters of a 3D camera model are estimated by means of these remaining motion vectors using an appropriate optimization algorithm.

3 Analysis and results

3.1 Task execution

We conducted the analysis on the basis of the usual environment of the participants. The ephemeral interactions required the collection of data using screen capturing, in order to obtain visual protocols of concrete interactions within the browser window. Within the multiple case study, up to now 23 Google Earth tours were captured as MPEG files between July 2009 and June 2010.

3.2 Form of landscape

The results of the software facilitate the differentiation between the properties of the chosen territory. If the dominant colour shown in the lower timeline equals the average chromaticity depicted in upper timeline, then either water, forest or urban areas are visible. If the landscape shown is divided between the depiction of water and land, as in coastal regions, the dominant colour is darker and tends towards the blue, while the average RGB value shows as a lighter grey or brown colour, depending on the population density.

Further, it is possible to differentiate between looking at a city and a closer view of a building from a distance of below 400 meters. Aerial and satellite images of cities from higher altitudes lead to a grayish colour value showing up in the colour timeline. By zooming in closer to focus on a specific building, the dominant colour changes towards black because of the shadow of the buildings and contrasts with the average value (Fig. 1).

3.3 Depth of navigation

Videana comes with text bubble detection, showing where the users clicked on additional information and thus opened highlighted white primed pop-ups. This method permits the determination of whether the user is following hyperlinks and is searching for additional information – i.e., at what 'depth' the user is navigating. Since all embedded media are accessed via pop-ups or 'bubbles', we can turn to the results of the software analysis to review to what extent linked media objects can be commonly understood as changing the way in which we use maps.

3.4 Direction of movement

Every movement within the interface has a certain direction which seems to be primarily determined by the desire to move to a certain place in the terrain/on the 'cartographic' image. What *Videana* shows is that the three navigational means, horizontal, vertical movement and zoom, often show up in combination. Whenever this was detected, the image was tilted, indicating a 'flight' over the depicted landscape in the bird's eye view. In contrast, horizontal and vertical movements used simultaneously without zooming, indicate the use of a-perspectival views that coincides with the orthogonal map view.

However, the simultaneous occurrence of horizontal, vertical and zoom movement also opens up further options for interpretation: The main difference between a human subject and a programmed series of movements is that a human in front of a screen uses the different forms of navigation sequentially, whereas the software is capable of performing different operations such as zooming, rotating and scrolling simultaneously. Therefore, automated navigation via the search field can be distinguished from manual navigation by analysing the specific patterns created by the default movements.

Overall, a differentiation can therefore be made, based on the camera motion detection, between (a) which direction of navigation users prefer, (b) whether a perspectival or a-perspectival view is being used, and (c) whether the virtual navigation is automated or autonomous.

4 Conclusions

Using *Videana* adapted for Google Earth allows the detection of the pattern of use in geobrowsing behaviour, its annotation and visualization. This permits the analysis of the visible spatial and navigational semantics of virtual globe tools based on a broad set of data.

Overall, the multiple case study conducted within the framework of this paper reveals the multitude of geobrowsing behaviours, especially contours, form/shape and colour as the three most important visual attributes of attention-guiding geovisualizations [14] [15] that can be detected automatically.

Several conclusions can be drawn based solely on the analyses of form of landscape, 'deepness' of navigation, and direction of movement:

(a) The tendency towards a preference for geospatial visualizations that exhibit a strong difference between average and dominant colour values indicates that Google Earth users spend more time over coastlines, buildings, and landmarks than on more uniform landscapes, e.g., forests, oceans, etc. As areas rich in contrast are generally preferred, it appears as though the user has the ability to identify map contours within satellite imagery.

(b) The (virtual) camera motion analysis clearly demonstrates that the test persons can hardly comprehend horizontal or vertical movements alone, but generally navigate using mixed, ambiguous directions of movement. This is mainly due to the fact that virtual navigation imitates various real-life transport modes, following roads or city streets, travelling by ship on a river, or flying over a forest. Users tend to click only on the infrastructure they are 'using'; open countryside and buildings remain untouched.

The chance to tilt the view and thus to look at 3D objects and textures from a first-person viewing perspective is only taken advantage of by users who already have some experience with, and knowledge of, 3D graphics programs. However, even these users repeatedly return to the top view typically offered by paper maps. The north orientation is also most often retained. Even if users rotate the map, they generally return to an orientation towards True North after a short period of time.

(c) Taking into consideration the fact that the persons under investigation clicked on very few text bubbles, the research questions posed at the outset can be answered as follows:

It can be concluded that navigation behaviour is largely defined by the morphology of the landscape and to a lesser extent by the navigational aids and the additional multimedia information provided. Taken together with the test result showing that users constantly return to an a-perspectival view and orientation towards the North, this shows that an established form of map use exists that has also been applied to geobrowsers. From a user perspective, it appears doubtful at this stage whether making a basic distinction between conventional maps and new geobrowser maps is meaningful.

Acknowledgments

This work is financially supported by the Deutsche Forschungsgemeinschaft (PAK 519).

References

[1] G. Gartner, "Web Mapping 2.0," *Rethinking Maps. New Frontiers in Cartographic Theory*, M. Dodge, R. Kitchin, and C. Perkins, eds. Routledge, London, pp. 68–82, 2009.

[2] M. Goodchild, "What does Google Earth Mean for the Social Sciences?," *Geographic Visualization: Concepts, Tools and Applications*, M. Dodge, M. McDerby, and M. Turner, eds. Wiley, Chichester, UK, pp. 11–23, 2008.

[3] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M.-J. Kraak, A. MacEachren, S. Wrobel, "Geovisual Analytics for Spatial Decision Support: Setting the Research Agenda," Special Issue of *The International Journal of Geographical Information Science*, Vol. 21, No. 8, pp. 839–857, 2007.

[4] D. Peuquet and M.-J. Kraak, "Geobrowsing: Creative Thinking and Knowledge Discovery Using Geographic Visualization," *Information Visualization*, Vol. 1, No. 1, pp. 80–91, 2002.

[5] A. Scharl and K. Tochtermann, *The Geospatial Web – How Geo-Browsers, Social Software and the Web 2.0 are Shaping the Network Society*. Springer, London, 2007.

[6] W. Cartwright, "Delivering Geospatial Information with Web 2.0," *International Perspectives on Maps and the Internet*, M. Peterson, ed. Springer, Berlin, pp. 11–30, 2008.

[7] J. L. Jensen, "Augmentation of Space: Four Dimensions of Spatial Experiences of Google Earth," *Space and Culture*, Vol. 13, No. 1, pp. 121–133, 2010.

[8] W. Cartwright, "Applying the Theatre Metaphor to Integrated Media for Depicting Geography," *The Cartographic Journal*, Vol. 46, No. 1, pp. 24–35, 2009.

[9] M. Jones, "Vanishing Point: Spatial Composition and the Virtual Camera," *Animation*, Vol. 2, No. 3, pp. 225–243, 2007.

[10] R. Ewerth, M. Mühling, T. Stadelmann, J. Gllavata, M. Grauer, and B. Freisleben, "Videana: A Software Tool for Scientific Film Studies," *Proceedings of International Workshop on Digital Tools in Film Studies Siegen*, Transcript Verlag, Bielefeld, Germany, pp. 101–116, 2009.

[11] R. Ewerth, M. Schwalb, P. Tessmann, and B. Freisleben, "Estimation of Arbitrary Camera Motion in MPEG Videos," *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. I, IEEE Press, Cambridge, UK, pp. 512–515, 2004.

[12] E. Stefanakis and K. Patroumpas, "Google Earth and XML: Advanced Visualization and Publishing of Geographic Information," *International Perspectives on Maps and the Internet*, M. Peterson, ed. Springer, Berlin, pp. 143–152, 2008.

[13] D. Fairbairn, "Measuring Map Complexity," *The Cartographic Journal*, Vol. 43, No. 3, pp. 224–238, 2006.

[14] A. MacEachren, *How Maps Work: Representation, Visualization, and Design*. 2nd ed., Guilford Press, New York, 2004.

[15] O. Swienty, T. Reichenbacher, S. Reppermund, and J. Zihl, "The Role of Relevance and Cognition in Attention-guiding Geovisualisation," *The Cartographic Journal*, Vol. 45, No. 3, pp. 227–238, 2008.