Geobrowsing Behaviour in Google Earth – A semantic Video Content Analysis of On-screen Navigation

Pablo ABEND, Tristan THIELMANN, Ralph EWERTH, Dominik SEILER,
Markus MÜHLING, Jörg DÖRING, Manfred GRAUER and Bernd FREISLEBEN

Abstract

In this paper, a semantic approach to the analysis of the recorded on-screen navigation within virtual globes is presented. The research question addressed is which specific geovisual browsing patterns can be identified during the use of Google Earth. In order to explore and visualize geobrowsing behaviour systematically, we have developed a novel system for the analysis of Google Earth tours, called *Videana*. The software's functionality comprises the detection of 'text bubbles', the visualization of dominant/average colour values, and the allocation of 'virtual camera' movements. Based on 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.

1   Introduction

Since the tools of web mapping have become available to laymen, "new vistas of acquiring, assembling, and publishing geographic information" have been promised (GARTNER 2009). It seems that a new type of "armchair geographer" (CARTWRIGHT 2008) is using nearly immersive hybrids between maps and new media applications to explore the globe virtually. In accordance with PEUQUET & KRAAK (2002), this usage form of geographic knowledge depositories by laymen can be called 'geobrowsing'.

In geobrowsers such as Google Earth, NASA World Wind and Microsoft Bing Maps 3D, it is not only possible to individually select a certain part of the representation without the constraints of the sheet line and to zoom in to change the scale, but also to tilt to bird's eye views in order to look at buildings in 3D. There are several 'devices' within Google Earth's 3D Viewer for virtual navigation: While geobrowsing, one can use the so-called move joystick to change the direction of movement. The look joystick, in contrast, is used to direct one's gaze "as if you were turning your head" (GOOGLE INC. 2012).

However, 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 "Maps 2.0" (CRAMPTON 2010) in times of sophisticated multimedia

cartography and 3D virtual worlds poses new challenges to the empirical usage evaluation of geographic visualizations. For this reason, based on a multiple case study, this paper will focus on (semi-)automatic methods of "geovisual analytics" (ANDRIENKO et al. 2010) that can be used to gain new insights into how people navigate on-screen.

To this end, the paper starts with the setting of the case study and the research questions addressed in it. Thereafter, specific attention will be devoted to the technical solution to the desire to obtain a software application to analyse captured Google Earth video sequences. This geovisual analytical tool will be introduced in detail in the middle section, with the analysis of the multiple case study and its results making up the final part of the paper, before ending with the conclusions.

2 Research Overview and Questions

Google Earth was selected as the object of analyses for this paper because it is the most popular geobrowser worldwide. Its investigation 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" (JONES 2007) 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.

As digital applications for geobrowsing have become ubiquitous, this paper suggests a software approach to grasp the different possibilities of navigating within the interface and to cover all possible user operations. A computational analysis has the advantage of semantically enriching the research data after analysis and of visualizing these metadata on timelines and in diagrams. This procedure enjoys several advantages:

- Semantic annotation makes results intersubjectively comprehensible, since visibility is granted by the plotting of the metadata in different timelines which are displayed underneath the video.
- A geovisualization tool can conduct batch analysis procedures. This means that large quantities of data can be processed without human input and the approach can be therefore included in a quantitative research design.
- The semantic analysis of captured tours enables researchers to gain usage data independent of the system provider, in this case Google. In this respect, Google Earth is only a 'testing ground', since the software works for other virtual globe tools as well.

The aim of this paper is to find out which specific geovisual browsing patterns can be identified with the aid of a software toolkit designed for video content analysis during the use of Google Earth. Derived from the results of the geographic literacy research, the following two general epistemological research questions arise: (a) Is 'geobrowsing' revolutionizing the possibilities of ordinary maps and globes (JENSEN 2010)? Moreover, to what extent can we distinguish between "conventional vs. new maps" (CARTWRIGHT 2009) that use elements from the 'real world' to give a 'sense of place' at all? (b) Can the assumption be verified that Google Earth is a powerful educational tool encouraging applied geographical knowledge, as claimed by Google Inc. (http://www.google.com/educators/geo.html)?

Previous approaches on analysing usability issues in geovisualization have focused on maps and fail to grasp entirely the variety of possibilities for interaction with the geographic content currently available with virtual globe tools. The most comprehensive map-design research "How Maps Work" (MACEACHREN 2004) at least covers in detail visual-cognitive and semiotic research on the design and interpretation of maps and other geovisualisations. In MacEachren's book chapter "How Maps Are Seen" it is assumed, while referring among other to Marr's model of vision (MARR 1982) and a study of GRATZER & MCDOWELL (1971) that edges are important elements of visual scenes, that "edges defined by the sky-line, ridgelines, shorelines, and vegetation boundaries all received particular attention" (MACEACHREN 2004: 102). Based on these study results arises (c) the question whether this observation can be applied to Google Earth as well, and if yes, which kind of edges attract the foveal attention of Google Earth users? A further three research questions focus on the navigational semantics of geobrowsing behaviour:

MacEachren shows that scanning of the visual screen based on a model by TREISMAN et al. (1990) involves five independent visual pathways: luminance, motion, binocular disparity, colour, and texture. This model "posits a series of 'feature maps' with one set for each pathway plus one for size and one for orientation. Each set of feature maps consists of individual layers (analogous to the structure of GIS)" (MACEACHREN 2004: 106) that can be combined into a common "depth map" (MARR 1982). With the assumed continued validity of this model idea based on neuropsychological findings, this serves as an explanation, why some geobrowsing tasks are conducted in parallel while others require serial visual search and virtual navigation processes. This, in turn, leads to (d) the research questions, whether and in as far Treisman's feature integration theory (TREISMAN 1988) can also be applied to the navigations behaviour in Google Earth and which independent visual pathways can be detected.

In a test of two hundred twenty-five subjects using 15 perspective block diagrams ROWLES (1978) found that test persons were able to judge relative height accurately, when the point of view for the map perspective was nearly overhead or as low as 15°, although the view from 15° resulted in occlusion of map sections and made the geovisualisation less useful. Therefore, the respective (e) final research question concerning the perception of depth in two-dimensional map displays reads: Which perspective view is favoured in Google Earth?

There are a number of different research methods that could be used to explore everyday geobrowsing usage. In the following, we will concentrate on the use of a visual analytic tool that permits the (semi-)automatic construction of protocols on geobrowsing behaviour, independent of the use of specific icons, joysticks or sliders on the screen or steering tools like the mouse or keyboard.

3   Research Method and Technique

3.1   Software for analysing Google Earth tours

In previous work (EWERTh et al. 2009), we have presented our video content analysis software (called *Videana*) that partially relieves media scholars from the time-consuming task of annotating videos and films manually. The software provides a number of methods that we have developed for analysing videos automatically, including algorithms for shot

boundary detection ("cut" detection) (EWERTH & FREISLEBEN 2009), camera motion estimation (EWERTH et al. 2004), face detection, text localization and OCR (optical character recognition) (GLLAVATA et al. 2004), and semantic video indexing (MÜHLING et al. 2012). We have tested our algorithms in the corresponding tasks of TRECVID evaluation for several years now (TRECVID: TREC Video Retrieval Evaluation, http://trecvid.nist.gov). The TRECVID evaluation series is the de-facto benchmark for the evaluation of algorithms for video content analysis and retrieval. Our approaches for camera motion estimation and semantic video indexing belonged to the best systems in the related tasks in recent years. The graphical user interface of *Videana* presents the detection results in separate timelines. 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: a text bubble detection and a colour detection.

In general, some geobrowsing information of a Google Earth tour can be extracted directly 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. However, KML files do not contain all information that is required for the research addressed by the current work (WERNECKE 2011). 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.

Therefore, putting too much emphasis on KML analysis would mean to miss a big part of the picture and to run the risk of losing the originally collected data since Google's 'reference map' is constantly subject to change. Video recordings grasp more of the usage since the sidebar and all the menu items can be recorded along with the browser window while in the KML file only the action within this frame is retained. Video recordings are also available for research even when no internet connection is available. This poses another problem: The situation in which the KML file was recorded is highly dependable on the internet connection. Delays caused by slow data traffics are not reproducible by the researcher when he looks at the data again on different hardware. Hesitation in action that might be due to the loading of the map tiles or 3D buildings is missed looking only at KML file data.

Beyond that KML is not the standard for applications such as NASA World Wind and other available globe software and it is a format with predefined parameters. This shortens the possibility for comparative analysis with other cartographic media. Other digital maps such as 'Flash maps' lack the functionality of recording tours in the KML format. In addition, certainly KML cannot serve as the starting point for the comparison of geobrowsing with interactions in other media such as 3D games. The possibility to analyse screen capture data along with KML files enables the researcher to use formats independent of the system provider and the various media content available.

We have developed some algorithms that are relevant for analysing Google Earth tours, they are briefly described below:
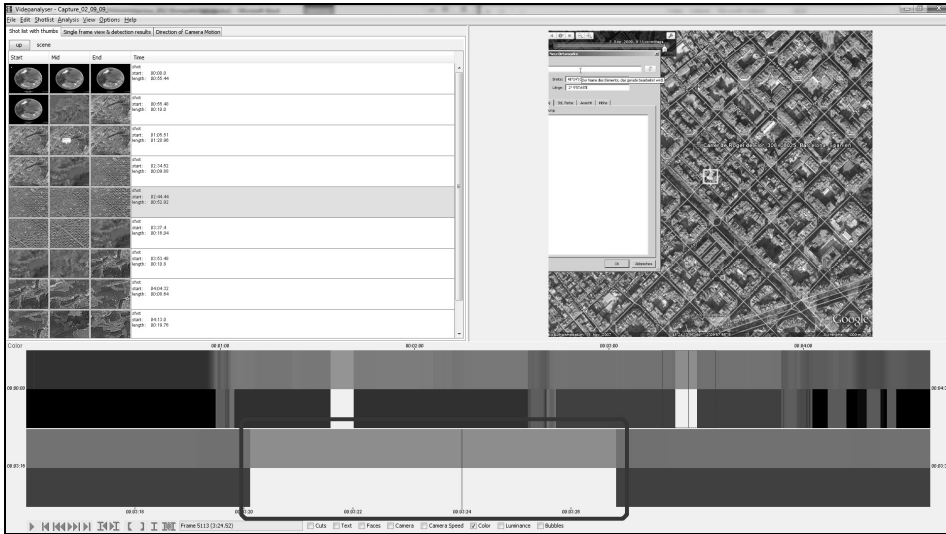
Fig. 1:   The city of Barcelona with a pop-up 'bubble'. Source: Capture of free geobrows-
ing Sept. 2, 2009 (female, 22 years, total time 17:56 min.). The lower timeline
shows an enlarged section of the selection marked in the upper timeline.

3.1.1 Text bubble detection

Google text bubbles stand out from the background by their white colour (Fig. 1). 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 algo-
rithm (contour processing) of Intel's Open Source Computer Vision Library to assemble
edge pixels into contours. The shapes of the resulting hierarchy of contours (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 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.

3.1.2 Colour detection

Average and dominant colour detection can help to identify colour contrast as an important
component of map complexity (FAIRBAIRN 2006). *Videana* provides two additional time-
lines and four different diagram types in order to visualize colour and brightness informa-
tion 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 de-
termined by the colour that is related to the bin with maximum frequency.

Colour and luminance information is visually represented by the corresponding timeline synchronized with the video frames. The colour video timeline 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.

3.1.3 Camera motion detection

The virtual camera in Google Earth serves as an input device with six degrees of freedom. There, different types of camera motion are: rotation around one of the three axes, translation along the x- and y-axis and zooming in and out, which 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 correspond to the level of detail in the information that is needed by the user.

The approach for the detection of camera motion (EWERTH et al. 2004) 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. Using the described system for camera motion estimation, we have achieved very good results at TRECVID's task for camera motion estimation. The recognition rate and precision for the detection of pan, tilt, and zoom were between 80% and 95% for these different kinds of motion.

3.2   Test arrangement

In order to obtain MPEG video recordings of everyday geobrowsing behaviour, the test persons were asked to explore Google Earth on a desktop computer with an activated screen capturing software. Checks were carried out in advance to ensure that all tested users had installed Google Earth version 5 and the standard configuration was set, without an activated additional information layer. In order to create a context for analysis that was as close to reality as possible, the data collection was conducted in the usual environment of the participants (at home or at work) and without limiting the time of their Google Earth usage.

The task of exploring Google Earth was given within two different settings: a) without initial instruction, and b) with the provision of a newspaper article (BILD, February 2, 2009) that outlines how people can dive underwater and fly over undersea landmarks with the new ocean layer of Google Earth 5. It was completely up to the users to geobrowse by mouse/fingertip travel or by following links/search results.

4   Analysis

4.1   Task execution

The inquiry period started in July 2009 with an explorative data analysis of two cases. This served as a pre-test of the design and as a foundation for the classification of further findings. In addition to the screen captures, this first test run involved an interview and a written field report. Between July and October 2009, the navigational activities of 15 test persons were logged, followed by five recordings in January and February 2010. The final three captures were produced in June 2010. The database comprises a total of 23 tours: of these, 11 participants are female and 12 male; 17 test persons carried out the entire have searched and browsed to places of their own choice, 6 subjects were led by the article on the Google Earth ocean layer. The recorded tours are of variable lengths ranging from 4 to 33 minutes. Within the geosemantic analysis we focused on the following aspects: type of landscape, 'depth' of navigation, and direction of movement.
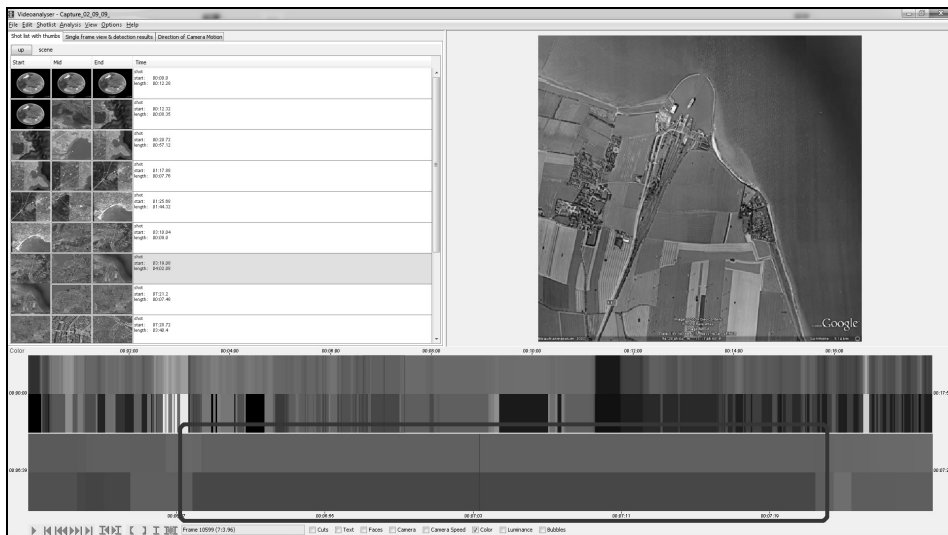

Fig. 2:   Coastal area in North Germany. Source: Capture of free geobrowsing on Sept. 2, 2009 (male, 30 years, total time 6:51 min.).

4.2   Type of landscape

The results 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 (in the dashed frame of Fig. 2).

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 greyish 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.

4.3 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. These white bubbles are also represented as white blocks in the dominant colour timeline (Fig. 1). 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.
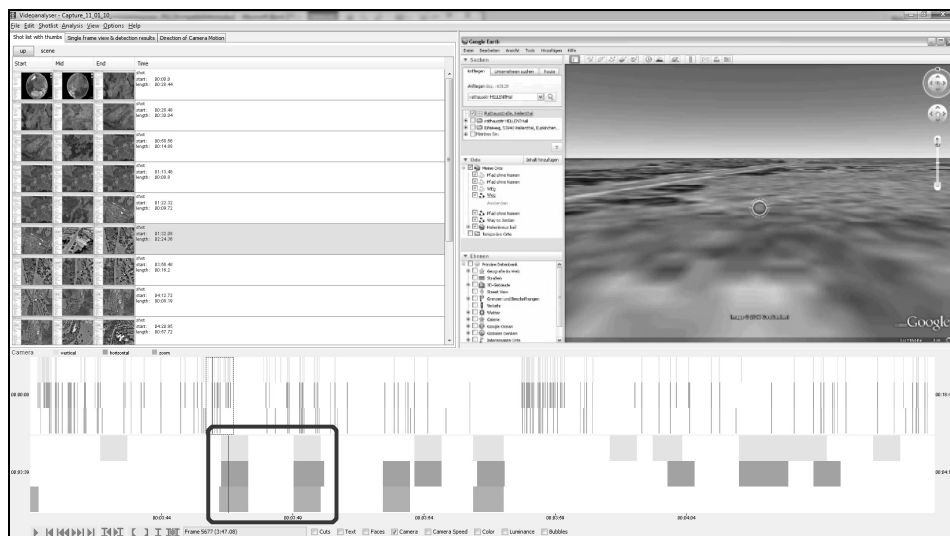

Fig. 3: Tilted view of a rural area near Cologne. Source: Capture of free navigation on January 11, 2010 (female, 36 years, total time 18:47 min.).

4.4 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 (Fig. 3). 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.

5   Results and Conclusions

Using *Videana* adapted for Google Earth allows the detection of the pattern of use in geo-browsing 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 geovisualisations (MACEACHREN 2004; SWIENTY et al. 2008) that can be detected automatically.

Several observations can be made based solely on the analyses of landscape type, 'deepness' of navigation, and direction of movement:

- 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 (Fig. 2).
- 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.

- The various different layers of extra information are used only marginally. Even if the test persons were informed of what an additional layer might offer, as in the case of the ocean layer, they returned to the standard configuration relatively rapidly. Icons are often only 'clicked' to display missing labels through roll-over. The information in the form of links placed below icons is used hardly at all.

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

(a) 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 navigational map use exists that has also been applied to geobrowsers. Therefore, when it comes to the literal handling of both map types, the relevance of making a strict epistemological distinction between conventional maps and new geobrowser maps appears doubtful at this stage.

(b) It is striking that even when distances are relatively short, and even within known contexts, navigation tends to be delegated to the system. The assumption that geobrowsing strengthens the interactive experience of geographical knowledge cannot be confirmed. The test persons often followed known patterns of movement that were repeated again and again. 'New explorations' were often broken off and the subjects returned to the role of 'armchair travellers', who want to use built environments that need no input at all to experience a Google Earth tour.

Users rely on the media presented and sabotage the accessing of their own geographical knowledge in the process. This is particularly problematic in as much as Google Earth's user interface has been consciously designed to make it extremely difficult, if not impossible, for users to distinguish between advertising, user-generated content and data provided by Google. It is precisely the existence of this mixture which renders it so important that there should be further independent usage analysis of such software.

(c) Furthermore, the analysis of the resulting hypotheses of the cognitive map-design research posted in the beginning demonstrates their reflection in geobrowsing behaviour. Territorial edges do attract foveal attention of Google Earth users since there is a preference for geospatial visualizations that comprise a noticeable difference between average and dominant colour values within a frame (Fig. 2).

(d) Treisman's five-pathway model for processing of visual scenes can be differentiated, when taking into consideration the results of the navigations behaviour in Google Earth. Horizontal motion (pan), vertical motion (tilt), and zoom each are individual pathways, while the map, the satellite image, the icons, the text bubbles etc. are perceived as a single texture – that is, a single 'feature map' operates with these different layers.

These findings are consistent with Phillips' explanation of the map reading process as a 'data reduction task' while the map reading subject processes and stores distinct snapshots of the map from retinal input. In this early stage of processing capacities to store information are limited. Given that the virtual camera in Google Earth simulates the human gaze, the on-screen navigation corresponds to the movement of the eye necessary to take those snap shots and integrate them in the overall structure of the map. Along with PHILLIPS

(1984) it can be concluded that Google Earth users constantly scroll and stop in order to acquire the processing of visual cues while geobrowsing. This conclusion is reached because in complex environments with lots of visual cues the timeline where the movement is depicted shows a staccato-like pattern whereas over homogenous landscapes that are easier to process users tend to fly over generating a cohesive pattern of uninterrupted movement.

(e) Additionally, in conclusion to (a) must be emphasized once again that the immersive elements of Google Earth are applied only rudimentarily. The observation that the users prefer an overhead perspective therefore implies according to ROWLES (1978) that the Google Earth users want to be aware of the height of their flights, and thus do not want to relinquish control of the territory.

6   Outlook

To enhance further the fitting for geographical evaluation tasks the KML integration needs to be driven forward. Within this next phase of development, the geovisual analytical toolkit will be tested with larger amounts of data. This includes spatial interactions such as the movements of players in Online Role-Playing Games that will be collected and compared to geobrowsing data obtained through screen-capturing and the extraction of geosemantic information from KML files directly.

References

ANDRIENKO, G., ANDRIENKO, N., DEMSAR, U., DRANSCH, D., DYKES, J., FABRIKANT, S. I., JERN, M., KRAAK, M.-J., SCHUMANN, H. & TOMINSKI, C. (2010), Space, Time and Visual Analytics. International Journal of Geographical Information Science, 24 (10), 1577-1600.

CARTWRIGHT, W. (2008), Delivering Geospatial Information with Web 2.0. In: PETERSON, M. (Ed.), International Perspectives on Maps and the Internet. Springer, Berlin, 11-30.

CARTWRIGHT, W. (2009), Applying the Theatre Metaphor to Integrated Media for Depicting Geography. The Cartographic Journal, 46 (1), 24-35.

CRAMPTON, J. (2010), Mapping: A Critical Introduction to Cartography and GIS. Wiley-Blackwell, Chichester.

DÖLLER, M. & LEFIN, N. (2007), Evaluation of Available MPEG-7 Annotation Tools. Proceedings of I-MEDIA '07 and I-SEMANTICS '07. Graz, 33-40.

EWERTH, R. & FREISLEBEN, B. (2009), Unsupervised Detection of Gradual Video Shot Changes with Motion-Based False Alarm Removal. In: 11th Conference on Advanced Concepts for Intelligent Vision Systems. Springer, Lecture Notes on Computer Science (LNCS), Bordeaux, France, Vol. 5807, 253- 264.

EWERTH, R., MÜHLING, M. & FREISLEBEN, B. (2007), Self-Supervised Learning of Face Appearances in TV Casts and Movies. Invited Paper (Best Papers from IEEE International Symposium on Multimedia 2006). International Journal on Semantic Computing, World Scientific, June, 185-204.

EWERTH, R., MÜHLING, M., STADELMANN, T., GLLAVATA, J., GRAUER, M. & FREISLEBEN, B. (2009), Videana: A Software Tool for Scientific Film Studies. In: ROSS, M., GRAUER,

M. & FREISLEBEN, B. (Eds.), Digital Tools in Media Studies: Analysis and Research. An Overview. Transcript Verlag, Bielefeld, Germany, 101-116.

FAIRBAIRN, D. (2006), Measuring Map Complexity. The Cartographic Journal, 43 (3), 224-238.

GARTNER, G. (2009), Web Mapping 2.0. In: DODGE, M., KITCHIN, R. & PERKINS, C. (Eds.), Rethinking Maps. New Frontiers in Cartographic Theory. Routledge, London, 68-82.

GLLAVATA, J., EWERTH, R. & FREISLEBEN, B. (2004), A Text Detection, Localization and Segmentation System for OCR in Images. In: Proc. of the 6th IEEE Int. Symposium on Multimedia Software Engineering. IEEE Press, Miami, USA, 310-317.

GOOGLE INC. (2012), Google Earth Tutorials: Navigating on the Earth. Available at http://support.google.com/earth/bin/answer.py?hl=en&answer=176674.

GRATZER, M. A. & MCDOWELL, R. D. (1971), Adaptation of an Eye Movement Recorder to Aesthetic Environmental Mensuration. Storrs Agricultural Experiment Station Research Report No 36. College of Agriculture and Natural Resources, University of Connecticut, Storrs, CT.

JENSEN, J. L. (2010), Augmentation of Space: Four Dimensions of Spatial Experiences of Google Earth. Space and Culture, 13 (1), 121-133.

JONES, M. (2007), Vanishing Point: Spatial Composition and the Virtual Camera. Animation, 2 (3), 225-243.

MACEACHREN, A. (2004), How Maps Work: Representation, Visualization, and Design. 2nd edition. Guilford Press, New York.

MARR, D. (1995), Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W. H. Freeman, San Francisco.

MÜHLING, M., EWERTH, R., ZHOU, J. & FREISLEBEN, B. (2012), Multimodal Video Concept Detection via Bag of Auditory Words and Multiple Kernel Learning. In: Proc. of the 18th International Conference on MultiMedia Modeling. Springer, Lecture Notes on Computer Science (LNCS), Klagenfurt, Austria, Vol. 7131, 40-50.

PEUQUET, D. & KRAAK, M.-J. (2002), Geobrowsing: Creative Thinking and Knowledge Discovery Using Geographic Visualization. Information Visualization, 1 (1), 80-91.

PHILLIPS, R. J. (1984), Experimental Method in Cartographic Communication: Research on Relief Maps. Cartographica, 21 (1), 120-128.

ROWLES, R. A. (1978), Perception of Perspective Block Diagrams. The American Cartographer, 15 (1), 31-44.

SWIENTY, O., REICHENBACHER, T., REPPERMUND, S. & ZIHL, J. (2008), The Role of Relevance and Cognition in Attention-guiding Geovisualisation. The Cartographic Journal, 45 (3), 227-238.

TREISMAN, A. (1988), Features and Objects: The Fourteenth Bartlett Memorial Lecture. Quarterly Journal of Experimental Psychology, 40A, 201-237.

TREISMAN, A., CAVANAGH, P., FISCHER, B., RAMACHANDRAN, V. & VAN DER HEYDT, R. (1990), Form Perception and Attention: Striate Cortex and Beyond. In: SPILLMANN, L. & WERNER, J. S. (Eds.), Visual Perception: The Neurophysiological Foundations. Academic Press, New York, 273-316.

WEBER, T., JENNY, B., WANNER, M., CRON, J., MARTY, P. & HURNI, L. (2010), Cartography Meets Gaming: Navigating Globes, Block Diagrams and 2D Maps with Gamepads and Joysticks. The Cartographic Journal, 47 (1), 92-100.

WERNECKE, J. (2011), The KML Handbook. Geographic Visualization for the Web. Addison-Wesley, Upper Saddle River, NJ.