GEO 511 - Master's Thesis

Influence of Stereoscopic Displays on the Terrain Reversal Effect in Satellite Imagery

Martina Meyer
09-753-807

Supervisor
Dr. Arzu Çöltekin

Faculty Member
Prof. Dr. Sara I. Fabrikant

Geographic Information Visualization & Analysis (GIVA)
Department of Geography, University of Zurich

Submitted 09/30/2015

Abstract

Nowadays satellite images are freely accessible for everyone with internet access. In this images a common perceptual phenomenon called terrain reversal effect can occur. It affects how people perceive depth in the terrain. In non-stereo images the terrain reversal effect is commonly linked to shadow as a depth cue (e.g. Saraf et al. 1996; Bernabé-Poveda & Çöltekin 2014, etc.). The aim of this study was to find out, if the effect persist, gets stronger or weaker when another depth cue is added, namely *stereopsis*. A comparative user study with the two visualization types non-stereo and stereo had been set up. Anaglyph glasses had been used as a stereoscopic display. Additionally the influence of level of expertise, task type as well as terrain type on the terrain reversal effect had been observed. The results revealed that stereopsis helps to weaken the terrain reversal effect. Especially in highly rugged terrain the use of anaglyph glasses was valuable, whereas in subtle change terrain no improvement could be observed. The influence of level of expertise lead to no significant results. Nevertheless with higher level of expertise, a higher accuracy rate was achieved. When looking on task type significant results could be observed. If participants were given a more complex task their accuracy rate increased. The terrain type analysis revealed that the illusion is also present in subtle change landscape.

Keywords: Terrain reversal effect, stereoscopic displays, anaglyph glasses, level of expertise, task type, terrain type

Acknowledgement

I would like to thank my advisor Dr. Arzu Çöltekin who supported me through the whole time of the study and gave me valuable advises. I also appreciate the professional and the familiar cooperation.

Furthermore I am very thankful to people from the GIVA department, especially Annina Brügger, who supported me with words and deeds.

I would like to thank all the participants for their time taking part in the experiment and therefore making the study possible.

Furthermore I like to thank Patrizia Vollmar, Nadine Friedmann and especially Gabi and Stephanie Umbricht for reading and correcting the thesis.

Additional thanks go to my friends and family for their mental support.

And I am really thankful to Roger Meierhofer for his active support during the whole time.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Modern online map providers (e.g. Google Earth) offer satellite images to anyone with internet access for free. These satellite images and aerial photographs are, amongst others, common tools for the public as well as for researchers who have to deal with geographical tasks. In the field of physical geography for example, one can monitor changes or identify soil types using satellite imagery through image classifications as well as photo-interpretation (e.g. Taylor et al. 2000; Hengl & Rossiter 2003). It is therefore important that the terrain can be interpreted correctly by the user, i.e. that it is visualized correctly (Saraf et al. 2007).

In satellite images as well as shaded relief maps the interpretation of three-dimensional spatial relationships on terrain visualizations can sometimes be prone to perceptual errors. An important perceptual problem is the so called terrain reversal effect. This perceptual phenomenon causes "an illusion in various three-dimensional geographic visualizations where landforms appear inverted, e.g. valleys can be perceived as ridges and vice versa." (Bernabé-Poveda & Çöltekin 2014). The main reason for this inversion seems to lie in the direction of the illumination (e.g. Saraf et al. 1996; Bernabé-Poveda & Çöltekin 2014; Gil et al. 2014).

Different names have emerged over time to describe this phenomenon, such as relief inversion effect (e.g Imhof 2007), relief inversion fallacy (Kettunen et al. 2009) or false topographic perception phenomenon (FTPP) (e.g. Wu et al. 2013). Most of these terms have appeared in connection to shaded relief maps, where the effect is quite well-known. Therefore this master's thesis focuses on the terrain reversal effect in satellite imagery. In this research area literature mainly focuses on methods to correct the effect (e.g. Saraf et al. 1996; Rudnicki 2000; Bernabé-Poveda et al. 2005; Gil et al. 2005), rather than its occurrence and prevalence (Biland 2014).

In this study the influence of stereopsis on the terrain reversal effect will be investigated. Stereopsis is the principle behind stereo images and allows to perceive a further depth cue in an image. In this master's thesis anaglyph glasses are used to simulate this depth cue. An empirical user study will be designed to investigate if visualization type (non-stereo and stereo images) makes any difference in experiencing the terrain reversal effect. More specifically, if in the presence of stereoscopic displays (in this study anaglyph glasses) the illusion becomes stronger or weaker when looking at satellite imagery.

There are indications in different studies (Bernabé-Poveda & Çöltekin 2014 and Biland 2014) that the level of expertise might influence the perception of the terrain reversal effect. Therefore the influence of level of expertise on the illusion will be investigated. Three different level of expertise (expert, geographers and non-geographers) will be defined. It will be examined if experience with satellite image as well as expert knowledge in the field of remote sensing can weaken the illusion.

Another aspect, which will be deepened in the project and that was rarely found in the body of literature, is the task type. In most of the previous experiments, the studied task type was more on a basic perceptual level, where the focus laid on a single shape in the landscape, e.g. is A higher than B (e.g. Bernabé-Poveda & Çöltekin 2014, Biland 2014). In this project, there will be also tasks, where the interpretation of the area plays a higher role, so there will be more interpretative tasks, e.g. in which direction does the water flow. It will be examined if the task type (in this study perceptual vs. more interpretative) make a difference in experiencing the terrain reversal effect.

In literature it is assumed that the illusion is most severe in rugged terrain, especially in high altitude areas of hilly terrain (e.g. Saraf et al. 1996; Wu et al. 2013). But this has not been empirically tested yet. This thesis distinguish three different terrain types (subtle change, middle rugged and highly rugged) to investigate if the terrain reversal effect is mainly a phenomenon in highly rugged terrain, as assumed in the literature, or if the effect also appears in other terrains.

1.2 Research Questions

In satellite images the perceptual phenomenon called terrain reversal effect is rarely investigated concerning occurrence and prevalence (Biland 2014). This master's thesis mainly focuses on understanding the influence of stereopsis on this effect in a comparative study with two visualization types: non-stereo and stereo (Research Question 1). The influence of level of expertise (Research Question 2), task type (Research Question 3) and terrain type (Research Question 4) on the illusion will also be investigated.

Research Question 1: Visualization Type

Terrain reversal effect is commonly linked to shadow as a depth cue. Does the effect persist, get stronger or weaker when another depth cue is added, namely stereopsis? In other words is there a different result for non-stereo and stereo viewing conditions? Have stereoscopic displays an influence on the accuracy of land form identification, confidence in task success and response time?

Research Question 2: Level of expertise

Does level of expertise make a difference in experiencing the terrain reversal effect in terms of accuracy of land form identification, response time and confidence in task success?

Research Question 3: Task Type

Does task type (perceptual vs. interpretative) make a difference in experiencing the terrain reversal effect in terms of accuracy of land form identification, response time and confidence in task success?

Research Question 4: Terrain Type

Does terrain ruggedness make a difference in experiencing the terrain reversal effect? Is the illusion experienced as strongly for landscapes with subtle changes as it is for rugged terrains?

1.3 Thesis Structure

The thesis is organized in six chapters. In chapter 2 an overview of the latest research relevant for this master's thesis is given. The main focus lies on terrain reversal effect and visualization type (non-stereo and stereo). In chapter 3 the used methods for the experimental part are described. The results of the experiment are presented in chapter 4. In chapter 5 the main findings from the results chapter are discussed and limits of the study are presented. Last but not least chapter 6 gives a conclusion of the main outcomes and ideas for future research.

2 Literature Review

This chapter gives an overview of the latest research relevant for this master's thesis. Four different sections are discussed: terrain reversal effect, visualization type (non-stereo and stereo images), level of expertise (experts vs. non-experts), task type and terrain type . The main foci are on the terrain reversal effect and visualization type.

2.1 Terrain Reversal Effect

The terrain reversal effect is a perceptual phenomenon which causes "an illusion in various three-dimensional geographic visualizations (satellite images, shaded relief maps, etc.) where landforms appear inverted, e.g. we perceive valleys as ridges and vice versa" (Figure 2.1). (Bernabé-Poveda & Çöltekin 2014). The phenomenon has emerged under different names over time, such as relief inversion effect (e.g. Imhof 2007), relief inversion fallacy (e.g. Kettunen et al. 2009) or false topographic perception phenomenon (FTPP) (e.g. Saraf et al. 1996; Wu et al. 2013). The illusion can be very strong and can evidently lead to misinterpretation in map reading (Bernabé-Poveda & Çöltekin 2014).

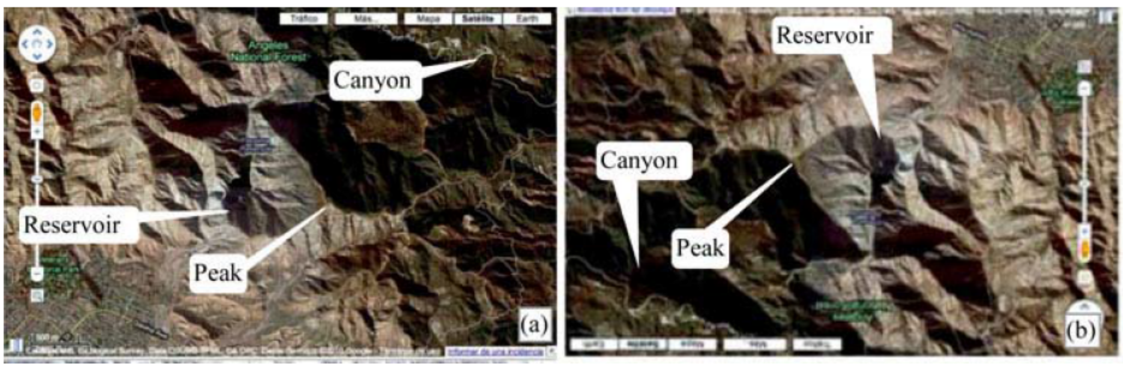


Figure 2.1: (a) North-oriented image of a natural region with canyon, reservoir and peak. Rivers appear to follow ridges and the peak is seen as a concave form. (b)A 180° rotation of the same region (without any other alterations) can remove the illusion (Bernabé-Poveda & Çöltekin 2014).

The literature about terrain reversal effect in satellite images focuses mostly on methods of correcting it rather than its occurrence and prevalence (Biland 2014). A compilation of various correction technics (Saraf et al. 1996; Rudnicki 2000; Saraf et al. 2005; Bernabé-Poveda et al. 2005; Gil et al. 2005; Saraf et al. 2007; Gil et al. 2010; Bernabé-Poveda et al. 2011; Wu et al. 2013) are described in Gil et al. (2014). Changing the illumination direction in the image is the basic idea of the described methods (Gil et al. 2014). Gil et al. (2014) mentioned that defining general methods for correcting the terrain reversal effect is difficult, on the one hand because the effect is very subjective, on the other hand because the environment and context of the images can vary a lot. One very simple correction method is turning the image by 180° (Saraf et al. 1996) (Figure 2.1b). However, because map users are used to north-oriented maps this method is often uncomfortable or even confusing (Bernabé-Poveda et al. 2005).

The terrain reversal effect is (along with satellite images) also known to occur shaded relief maps (Bernabé-Poveda & Çöltekin 2014). However in contrast to satellite images, in shaded relief maps the artificial light source can be controlled quite easily by cartographers (Bernabé-Poveda & Çöltekin 2014). A long standing cartographic tradition says that to portray shaded relief maps, an upper left illumination (northwest on a north-oriented map) should be used globally (Imhof 2007). This direction can still vary locally in manual shading (Imhof 2007). Interestingly, an upper left illumination is rarely the direction of the sunlight in the natural world (Imhof 2007, Patterson 2011). Figure 2.2 shows the Bright Angel Canyon in Arizona which appears as a proper canyon when illuminated from the northwest (left image) and as a mountain chain when illuminated from the south-east (right image) (Patterson 2011). These findings are backed up by the results from the light direction experiment conducted by Biland (2014). In his study, incident light at 337.5° yielded the highest accuracy and confidence rates among all investigated light directions.

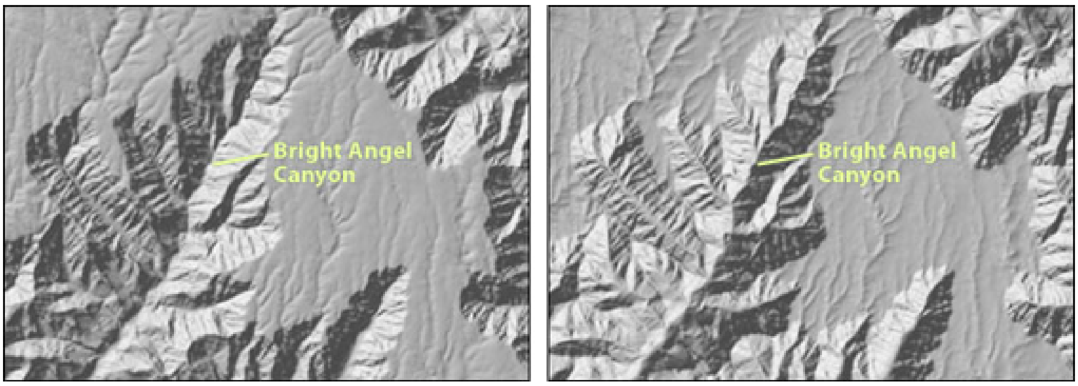


Figure 2.2: On the left image the Bright Angle Canyon is illuminated from the northwest and appears correctly as a canyon. On the right image the canyon is illuminated from the southeast and appears as a mountain chain (Patterson 2011).

Patterson (2011), Imhof (2007) and other authors see the preference for this upper left illumination in ergonomic considerations and cultural preference. Right-handed people, like most of us, need their light source above and to the left of themselves while writing and drawing, otherwise they would be disturbed by their own shadows. To minimize the possibility of smearing, people tend to work from the upper left to the lower right.

These explanations are not yet empirically tested. Imhof (2007) mentioned that their is a "considerable disagreement on the question of the best direction of illumination".

Most authors see the direction of illumination as the main reason for inversion in satellite imagery (Saraf et al. 1996; Saraf et al. 2005; Saraf et al. 2007; Bernabé-Poveda et al. 2005; Bernabé-Poveda et al. 2011; Bernabé-Poveda & Çöltekin 2014; Gil et al. 2014). They refer to literature from perceptual psychology (e.g. Okoshi 1976). In perceptual psychology the concept of inversion is well known (e.g. Gregory 1997; Kleffner & Ramachandran 1992; Howard 2002; Howard & Rogers 2002, Liu & Todd 2004; Hill & Johnston 2007). Although their research in this field focuses on individual objects (e.g. hollow-face illusion (Hill & Johnston 2007)) and not on complex landscapes, their findings are helpful in understanding the phenomenon in satellite imagery. Toutin (1998) and Saraf et al. (2007) mentioned that the psychological depth cues are the most significant causative factors of terrain inversion in satellite images.

Depth cues, which help our brain to perceive the third visual dimension, can be classified into two categories: physiological (e.g. accommodation, convergence, binocular parallax, etc.) and psychological (e.g. linear perspective, texture gradient, occlusion, lighting and shading, etc.) depth cues (Okoshi 1976; McAllister et al. 1995; Howard & Rogers 2002; Reichelt et al. 2010). Physiological depth cues are either monocular (i.e. information from one eye is sufficient to perceive depth) or binocular (i.e. information from both eyes is needed to perceive depth). Psychological depth cues are monocular (Mehrabi et al. 2013). Figure 2.3 gives an overview of the major physiological and psychological depth cues (Geng 2013).

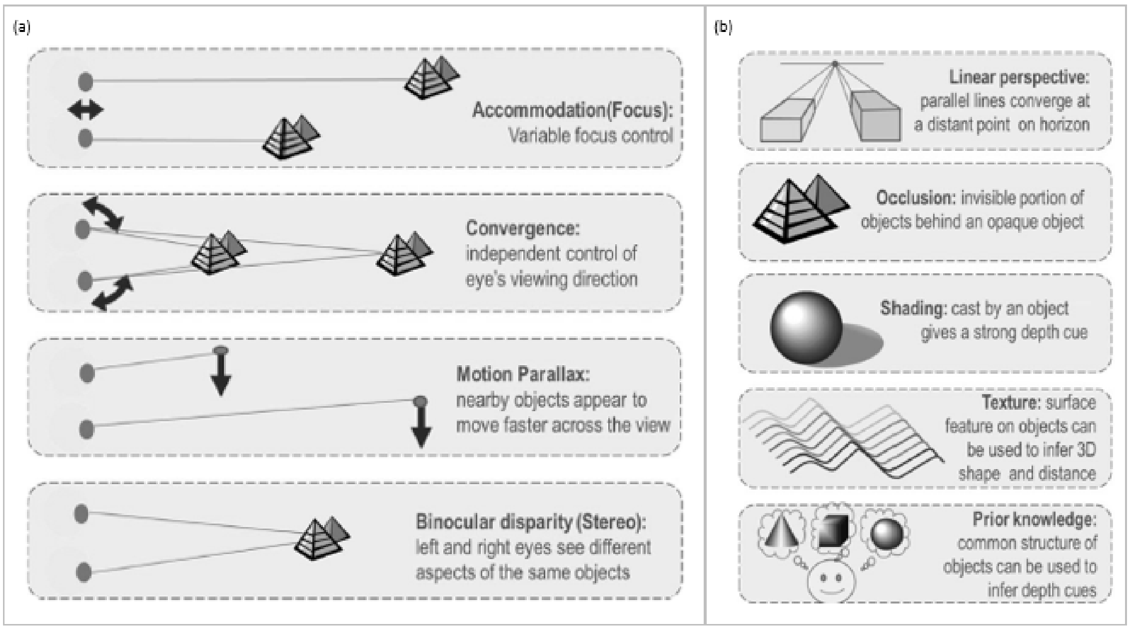


Figure 2.3: (a) Major physiological depth cues. (b) Major psychological depth cues (Geng 2013).

Lighting and shading thereby seems to be the main reason why we perceive depth in satellite imagery (Gil et al. 2014). In perceptual psychology the concept of shape from shading is widely spread (e.g. Ramachandran 1988; Kleffner & Ramachandran 1992; Sun & Perona 1998; Langer & Bülthoff 2001; Jenkin et al. 2004; Liu & Todd 2004; Gerardin et al. 2010; Morgenstern et al. 2011). Shape from shading theory, as the name suggests, says that we can perceive three-dimensional shapes because of shading (e.g. Ramachandran 1988). Kleffner & Ramachandran (1992) mentioned two assumptions in connection with the extraction of shape from shading information. A single light source illuminating the whole scene and the light is shining from "above" in relation to retinal coordinates. The first assumption is illustrated by Figure 2.4. People will never see both rows as either convex or concave.

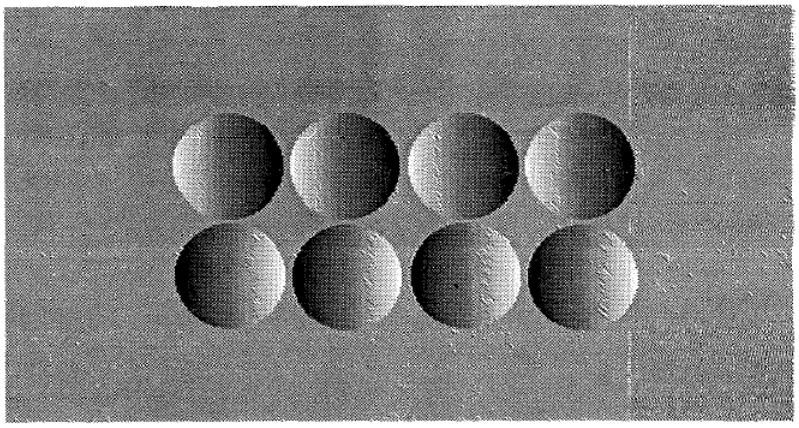


Figure 2.4: People will never see the rows as either both convex or both concave (Kleffner & Ramachandran 1992).

The second assumption (the light is shining from "above") is still discussed in literature. Sun & Perona (1998) conducted an experiment, which revealed, that people prefer lighting direction more shifted to the left than directly overhead. Gerardin et al. (2010) explored the neural basis of the "light from above left" preference. Their research revealed that in the visual system the illumination is processed before the representation of 3D shape is. In their studies, Morgenstern et al. (2011) found out that the light-from-above factor plays a more limited role than previously thought. Human vision relies more on lighting cues to recover 3D shape. It is also more likely that surfaces are perceived as convex rather than concave (e.g. Hill & Bruce 1994; Langer & Bülthoff 2001; Liu & Todd 2004). This phenomenon is called global convexity bias. Possible reasons can be found in natural environment. More objects are globally convex rather than concave (Johnston et al. 1992 and Hill & Bruce 1993). In the light direction experiment from Biland (2014) the overhead illumination bias (with the light-from-above-left preference) was the most driving factor for the prevalence of terrain reversal effect. Additionally, effects produced by specific terrain characteristics as well as global convexity or familiarity might have influenced the illusion too. For the experiment, Biland (2014) used shaded relief maps and not satellite images. But the results could be relevant for satellite images too.

It is believed that the terrain reversal effect occurs mostly on satellite images of the northern hemisphere (Toutin 1998; Rudnicki 2005; Bernabé-Poveda et al. 2005; Saraf et al. 2007). Figure 2.5 shows the annual location of the sun in relation to the horizon in the northern hemisphere. The sun is always located "towards the south" (Bernabé-Poveda et al. 2005). Saraf et al. (2005) sees the reason for the occurrence of the terrain reversal effect in the nothern-hemisphere in the position of the satellites and the time of observation during which the pictures are taken. Most observation satellites have a polar or quasi-polar orbit and are synchronized with the sun. Because of ideal image conditions, these satellites take pictures between 9:00 and 11:00 AM solar local time. On these pictures the sun is shining from south-west onto the northern hemisphere, which presumably causes the terrain reversal effect (Gil et al. 2014). This prediction is supported by an experiment from Bernabé-Poveda & Çöltekin (2014) where participants had to identify terrain on satellite images of the northern and southern hemisphere. First, they were shown the north-oriented picture and after, the same ones, but 180° rotated. They made significantly more mistakes on the north- oriented pictures of the northern hemisphere than on the rotated ones. For the images of the southern hemisphere the results were just the opposite: The north-oriented pictures were interpreted correctly whereas the ones rotated by 180° were misread. The light direction experiment from Biland (2014) also showed that the very southern lighting directions are indeed absolutely susceptible to relief inversion. These lighting directions are inherent in satellite images of the northern hemisphere, therefore the terrain reversal effect is very prevalent among them. A preliminary subjective analysis from Bernabé-Poveda & Çöltekin (2014) showed, that it is more likely that the effect is perceived in pictures of the northern hemisphere than in the ones of the southern hemisphere.

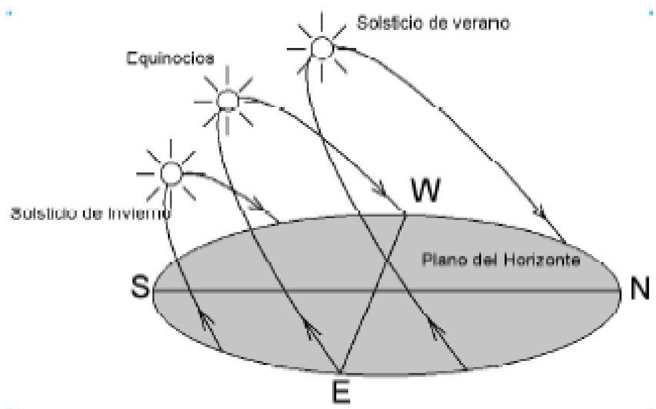


Figure 2.5: Annual location of the sun in relation to the horizon in the northern hemisphere (Bernabé-Poveda et al. 2005).

The terrain reversal effect also exists on extraterrestrial imagery like the Moon (e.g. Saraf et al. 2011; Wu et al. 2013), especially surrounding impact craters or long, narrow ridges. The effect is stronger on lunar images, i.e. the lunar image is more simple and more graphical. The reason for this is the lack of an atmosphere and vegetation and the presence of numerous craters (Saraf et al. 2011). There are also no recognizable clues (like streets, trees, etc.). Only the knowledge of the position of the sun allows a definite interpretation (e-mail communication with Hauber, DLR, 26/11/14). Figure 2.6 shows a typical example: On the left image, which is north-oriented, the crater is perceived as a hill. On the right, the same image is shown, but due to a 180° rotation, the crater can now be perceived as such(Wu et al. 2013).

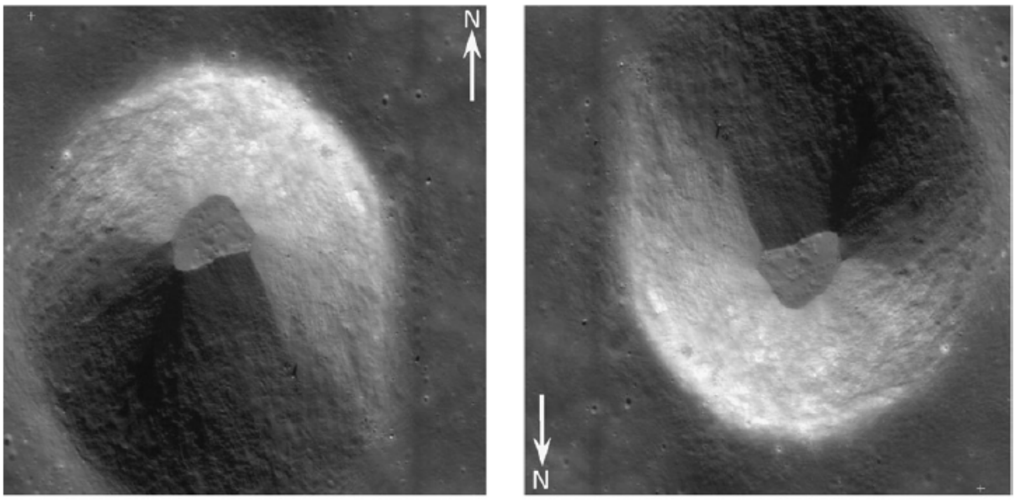


Figure 2.6: On the left image (north-oriented) we perceive the crater as a hill. On the right image (180° turned) we perceive the crater as crater. (Wu et al. 2013)

Literature has shown that there are factors which minimize or even made the perception of the terrain reversal effect impossible. Obvious hints in satellite images like land cover (snow, vegetation, rivers and such) can reduce or cancel the terrain reversal effect (Bernabé-Poveda & Çöltekin 2014; Biland 2014). For example when somebody perceives a ridge but also sees a river flowing along this ridge, they might realize that the "ridge" in in fact a valley. In other words the top-down cognitive signal might be suppressing the bottom-up perceptual signal (Bernabé-Poveda & Çöltekin 2014). So actually the illusion does not really disappear, but people notice discrepancies and therefore might decide correctly according to their knowledge. Furthermore when people who are familiar with the region shown on the satellite image (e.g. Matterhorn) are immune to the inversion effect (Bernabé-Poveda & Çöltekin 2014). Another well-known inversion effect in perceptual psychology is called the hollow-face illusion (Schröder 1858; Gregory 1997; Hill & Johnston 2007). For this experiment they used a hollow mask which had a convex and a concave side. Interestingly, people always saw the mask as convex which is the familiar way of seeing a face. The necker-cube effect can also influence the illusion (Kornmeier & Bach 2005). In his study Biland (2014) named this phenomenon terrain flipping. A valley can be perceived as a ridge and in the other moment as a valley.

2.2 Visualization Type

The existing body of research on terrain reversal effect focuses on non-stereo and not on stereo images. However, depth inversion does occur in stereo as well (Frisby & Mayhew 1979; Yellott & Kaiwi 1979 (both cited in Howard & Rogers 2002)). With a simple shape, i.e., a 3-D skeletal cube, Howard (2002) showed that the reversals occur more frequently when the cube is viewed monocularly than when it is viewed binocularly.

As mentioned in chapter 2.1, depth cues, which help our brain to perceive the third visual dimension, can be classified into two categories: physiological and psychological ones (Okoshi 1976). In non-stereo images lighting and shading seems to be the main reason why we perceive depth in satellite images (see chapter 2.1). When using stereo images, a further depth cue is given. The principle behind stereo images is stereopsis (binocular parallax) (Hubona et al. 1999; Mehrabi et al. 2013). Because our eyes are positioned approximately 50 to 75 mm apart (Dodgson 2004) they see images from slightly different angles (Anderson & Nakayama 1994; Westheimer 1994; Benoit et al. 2008). These two slightly different images are merged in the brain and provide 3D perception (Julesz 1960; Lambooij et al. 2011; Mehrabi et al. 2013). Using anaglyph glasses simulates these depth cues. With the glasses, consisting of two differently coloured lenses (e.g. cyan and red), each eye sees one of the two differently colour filtered images. In Figure 2.7 it is illustrated, that the left eye only sees the left-eye view and the right eye only sees the right-eye view.

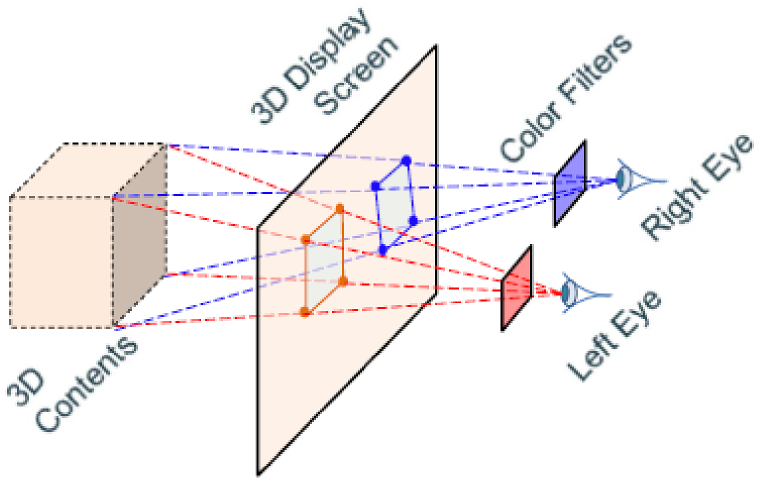


Figure 2.7: Principle of anaglyph glasses. Left eye only sees the left-eye view and the right eye only sees the right-eye view (Geng 2013).

Anaglyph glasses have various advantages: e.g. their images are easy to generate (e.g. with the software StereoPhoto Maker), they are economical, several viewers can use them and hard copies can be made (Sexton & Surman 1999; Mehrabi et al. 2013). The main disadvantages are that most of the colour information gets lost the during colour reproduction process and a prolonged use of anaglyph glasses can cause headache or even sickness (Kooi & Toet 2004; Lambooij et al. 2009; Mehrabi et al. 2013).

Anaglyph glasses are just one of the techniques for stereoscopic viewing and display. Figure 2.8 gives an overview of the most common stereoscopic viewing and display techniques (Çöltekin 2006). Two main categories can be distinguished: time multiplexed (the two images are sent in a sequence) and time parallel (the two images are sent simultaneously). In addition to stereoscopic display techniques, there are also real 3D display techniques, where all of the depth cues are simulated. By moving around additional information can be found about the object of observation. (Mehrabi et al. 2013).

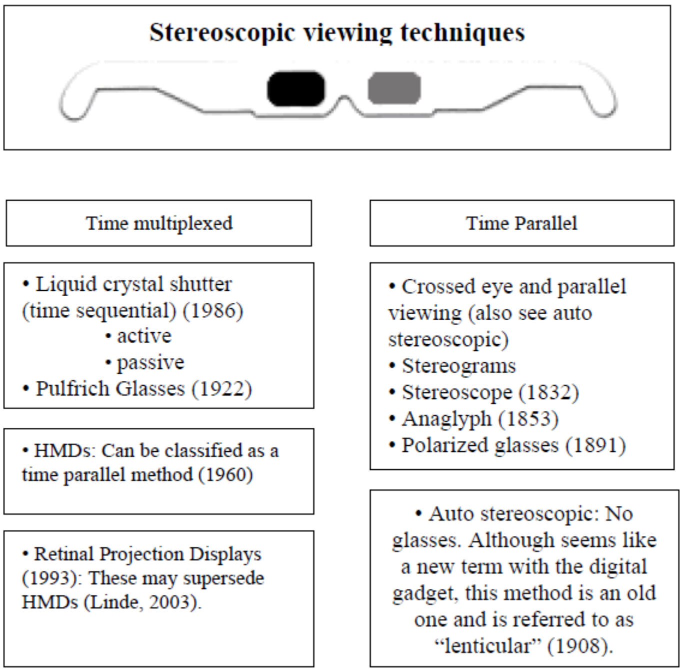


Figure 2.8: An overview of stereoscopic vision techniques (Çöltekin 2006).

When talking about stereopsis, stereo blindness must also be considered (e.g. Richards 1970; Fielder & Moseley 1996; Ukai 2006). Ware (2000) (cited in Çöltekin 2006) mentioned that stereo blindness does exist and that as many as 20% of the population may have it. Interestingly they are often unaware of their disability. This shows that people can still function perfectly well without stereoscopic disparity because there are many other depth cues that the brain uses to cope within 3D space. For testing stereopsis different stereotests are available (Fricke & Siderov 1997).

Nowadays, many satellites are able to generate stereo satellite imagery (e.g. IKONOS, Terra Aster, etc.) (Shaker et al. 2010). A single satellite can capture the images consecutively along the same orbit within a few seconds (along the track imaging technique) (see Figure 2.9). Or the same/different satellites capture the image from various orbits on divers dates (across the track imaging technique) (Shaker et al. 2010). The primary advantage of stereo satellite imagery is the ability to extract vector features and geographical features in 3D (like buildings, roads, man-made structures and other terrain features) (Dial et al. 2003; IKONOS 2015). In literature, different studies can be found, where stereo satellite imagery was used (Shaker et al. 2010; Nichol et al. 2006; Stearns & Hamilton 2007). Substantial research work has been conducted using stereo satellite imagery to map the Earth's surface (Shaker et al. 2010). Other studies investigated the application of high-resolution stereo satellite images to detailed landslide hazard assessment (Nichol et al. 2006) or to quantify rapid volume loss from outlet glaciers (Stearns & Hamilton 2007).

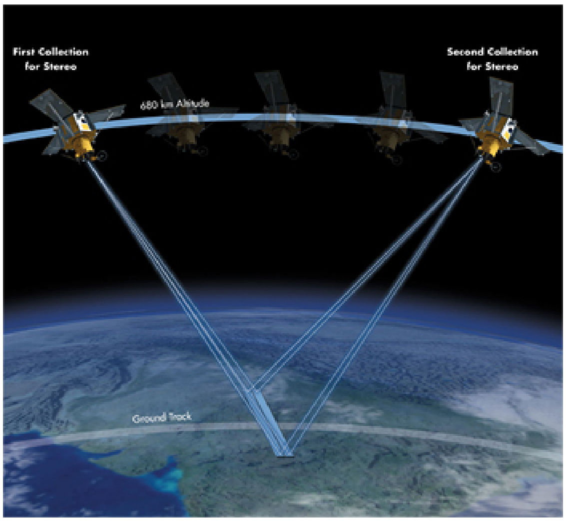


Figure 2.9: Image capturing method for creating IKONOS stereo satellite imagery. (IKONOS 2015).

Different studies have compared 2D and 3D visualizations (e.g. John et al. 2001; Smallman et al. 2001; Tory et al. 2006; Shepherd 2008; Carvalho 2011; Seipel 2013; Niedomysl et al. 2013). In the geographical field, such studies are, amongst others, executed in the field of wayfinding (Kray et al. 2003) and visualization of data (S. I. Fabrikant et al. 2014). Most of these studies see in 3D views a perspective or oblique view of an object or a scene displayed on a computer monitor. The image is actually two-dimensional but the viewing angle provides a three-dimensional perspective (e.g. John et al. 2001). Even if in stereo satellite images an orthogonal view is given, their findings can be still useful for the study. For tasks involving 3D spatial data, 2D as well as 3D visualization strategies are valuable and are appropriate for different tasks (Tory 2003). 2D views are often used to determine accurate relationships because there is no depth ambiguity (Tory 2003). They can also enable analysis of details, precise navigation and distance measurements (since only one dimension is ambiguous) (Smallman et al. 2001; John et al. 2001). Usually map presentations are shown in 2D layouts. But with only one perspective of the view being accessible in 2D some important information can be hidden. According to the literature 2D is better for tasks involving metric judgements. (Carvalho 2011). In comparison to 2D, 3D views enable additional display space (z-axis) where more information can be represented (Carvalho 2011; S. I. Fabrikant et al. 2014). 3D views are typically used to get a qualitative understanding of the data and to present this understanding to others (Springmeyer et al. 1992). 3D (perspective) displays are also well suited for gaining an overview of a 3D space, understanding 3D shapes and for approximate navigation (Tory 2003; Carvalho 2011). Because we live in a 3D world, different literature suggests that we should be able to extract more information from 3D displays than from 2D displays (e.g. Wise 1999).

In literature there are studies which demonstrate that stereoscopic visualization may provide a user with a higher sense of presence in displayed environments. Reasons can be seen in better depth perception, which can lead to better comprehension of distance (relative and egocentric), of ambient layout, object presence, etc. (Livatino & Privitera 2006). Stereo images enables the user to elicit depth information and certain other aspects of detail in what might otherwise be interpreted as a flat 2D image (Sexton & Surman 1999). Carvalho (2011) tested how 2D differs from a non-stereo and stereo 3D visualization. Participants had to interpret the distance between two points shown in 2D and 3D stereoscopic maps. The results were not significantly different but the participants were more confident using stereoscopic 3D maps. Forsberg et al. (2009) investigated line and tube representations of integral curves with both, non-stereo and stereo viewing. In their study, participants liked the combination of a clear visualization and paired with the stereoscopic viewing, although stereo viewing did not generally improve their accuracy. In a study from Seipel (2013), the assessment of distances in a geographical context in 2D and 3D presentations of maps was investigated. The 3D presentation was also in stereo viewing conditions. It revealed, that 3D visualisation in stereo viewing conditions, leads to visual discomfort and significantly fewer correct answers. Seipel (2013) also mentioned, that stereo cues tend to result in increased times for solving the tasks. In a study from Sexton & Surman (1999), presentations of intrinsic 2D content (maps) in 3D context did not benefit from cues provided by stereo viewing conditions.

In conclusion, the use of stereoscopic viewing methods depends on the task performance. In several studies stereo images improved the results, whereas in other studies the stereo viewing condition lead to less accurate performance. The main advantage of stereo images is certainly that they enable the user to elicit depth and certain aspects of detail in what might otherwise be interpreted as a flat 2D image (Sexton & Surman 1999).

2.3 Level of Expertise

In the field of terrain reversal effect in satellite images two different studies exist were level of expertise is briefly discussed. In the two-stage online user experiment from Bernabé-Poveda & Çöltekin (2014) people experienced the terrain reversal effect regardless of their background but a weak positive effect based on expertise could be observed which need to be properly investigated. The image type experiment from Biland (2014) showed that especially observers with an expertise knowledge in satellite imagery achieve better results in accuracy than unaccustomed people. Biland (2014) suggests that the experts had learned to interpret the land cover information so that they were not dependent solely on the overhead illumination bias.

The term of expertise can have different meanings. Therefore it is important to define the right context when using this term (Nyerges 1995 (cited in Slocum et al. 2001)). In geovisualization for example, the term of expertise can be used for user experience with the tool, the problem domain or computers in general (Slocum et al. 2001). In this master's thesis experts come from the field of remote sensing, because they have learned how to interpret satellite images. The assessment of satellite imagery can be carried out through different methods like visual image interpretation, photogrammetry assessment or digital image evaluation (Albertz 2009).

In the field of psychological research, scientists like Smith (1962) (cited in Toutin 1997) have indicated that performance at searching tasks is much improved if one knows something beforehand about what is to be looked at. Toutin (1997) mentioned that "because psychological factors play a major role in perception, the remote sensing expert can "go beyond the information given" in the display of an image. Thus, a viewer with some a priori knowledge of the data and of the terrain and with a good understanding of the processing has a more qualitative experience." It has been suggested (Hoffman 1990 (cited in Toutin 1997) that researchers might devote more time for studying and integrating these qualitative aspects of the remote sensing process. Carvalho (2011) showed in his study that the skilled participants, related to their 3D vision, 3D game playing and familiarity, have overall better results compared to non-skilled users.

In the research of cognitive skills it has become a maxim that experts remember better when presented with new information relating to their domains of expertise. A common interpretation of these research findings is that experts have built up a "rich repertoire of schemata, which enable them rapidly and efficiently to encode appropriate new information" (Gilhooly et al. 1988). Gilhooly et al. 1988 investigated this maxim on expert memory to the case of map reading. Therefore topographic contour maps were used which are abstract representations of three-dimensional landscapes in two-dimensional form. Analyses of verbal protocols showed that the skilled subjects made more use of specialist schemata, whereas the unskilled subjects spent more time in reading place names. In this specific example the maxim was confirmed.

In conclusion, different studies have shown that experts might have an other approach to solve tasks in their field of expertise than non-experts. But it is also important to mention that besides training and experience (nurture), abilities and talents (nature) also exist (Ericsson & Lehmann 1996). In connection to the terrain reversal effect in satellite images, preliminary studies have shown, that the level of expertise might have an influence on the results (Bernabé-Poveda & Çöltekin 2014; Biland 2014).

2.4 Task Type

The terrain reversal effect in satellite images has only been investigated through "perceptual, base shape recognition" tasks (Bernabé-Poveda & Çöltekin 2014; Biland 2014). In a study from Bernabé-Poveda & Çöltekin (2014) participants had to identify 3D landforms as well as judge 3D spatial relationships. In Figure 2.10 the two examples are illustrated. Participants had to decide, if A is higher than B, and if the line AB looks like a valley or a ridge. In both questions the focus was mainly on the area, immediate around the given letters in the image and not on a broader area. The tasks could be solved on a perceptual level and needed no further interpretation.

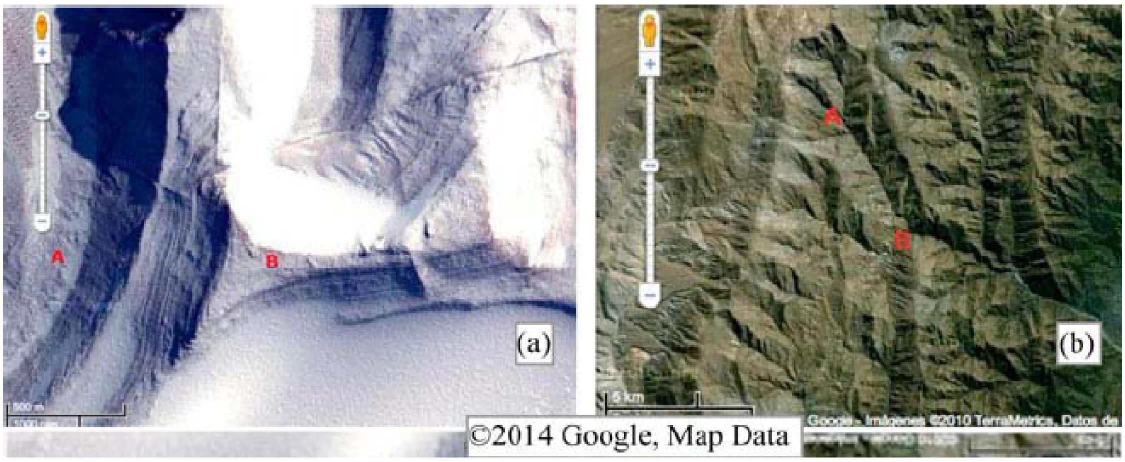


Figure 2.10: Two perceptual tasks. (a) Point A is located in a place higher/lower than place B (b) Line AB looks like a valley/ridge. (Bernabé-Poveda & Çöltekin 2014)

2.5 Terrain Type

In the literature of terrain reversal effect in satellite imagery, the focus is lying on highly rugged landscape, because they assume, that the effect is more severe in rugged terrain, especially in high altitude areas of hilly terrain (Saraf et al. 1996; Saraf et al. 2005; Saraf et al. 2007; Saraf et al. 2011; Wu et al. 2013). In such landscapes it is believed that the topographic relief leads to "changes in the solar illumination to the slopes and to the viewing geometry of the terrain" (Saraf et al. 1996).

Saraf et al. (1996) mentioned that "the conditions may be such that one of the slopes of a valley is illuminated, whereas the other slope is deprived of any direct illumination from the Sun, causing counter perception". In the case of hills the problem of differential illumination remains the same. When viewing satellite images observers unconsciously evaluate the relative distance of various objects from some imaginary datum plane, i.e. by trying to perceive a depth in the image (Saraf et al. 1996). Perceiving depth in satellite images can be influenced by different factors: perception of the individual (e.g. binocular disparity) or natural factors (like topographic relief and sun elevation). These simultaneously and sometimes individually influence the appearance of topography. (Saraf et al. 1996). As mentioned in section 2.1, most authors see the direction of illumination as the main reason why we perceive depth in satellite images and therefore perceive this inversion (Saraf et al. 1996; Saraf et al. 2005; Saraf et al. 2007; Bernabé-Poveda et al. 2005; Bernabé-Poveda et al. 2011; Bernabé-Poveda & Çöltekin 2014; Gil et al. 2014).

In Figure 2.11 Saraf et al. (1996) give a possible explanation why the effect might be more severe in rugged terrain. In normal situations, when looking at a flat topography, the mental model always fixes the position of source on the opposite side of the normal (a). In a sinusoidal topography, one side of the ridge/valley faces the Sun. That means that one face is illuminated, whereas the other remains shadowed (b). When the target objects are proportionately smaller than the viewing height, a valley/ridge will appear as "strips of two different grey levels juxtaposed against each other".

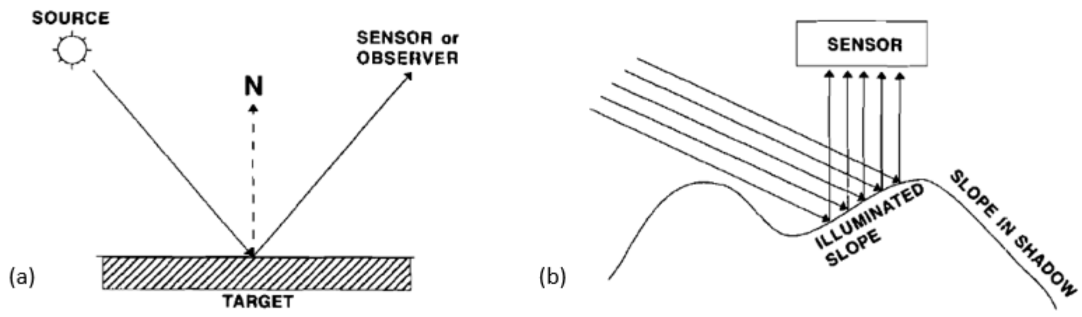


Figure 2.11: Flat and hilly topography. (a) The source-target-observer position in case of a flat target with respect to the normal (N). (b) Sun-target-sensor geometry in case of a sinusoidal topography; here the sunfacing slopes are illuminated, whereas the opposite slopes are shadowed (Saraf et al. 1996).

3 Methods

3.1 Participants

An experimental user study was designed to answer the research questions. A total of thirty-three participants took part in the experiment. The participants were divided into three groups depending on their level of expertise in the field of remote sensing. The groups were defined as expert, geographer and non-geographer.

An *expert* was defined as a person who has a geographical education at university level, has experience with satellite images and especially works with them frequently. Therefore MSc and PhD students from the remote sensing research field were chosen.

A *geographer* was defined as a person who has a geographical education at university level, is familiar with satellite images but has less experience with them than an expert and does not use them on a regular basis.

A *non-geographer* was defined as a person who has no geographical education at university level. He might use satellite images in his leisure but not professionally.

The geographers and the experts were recruited personally from the University of Zurich, the non-geographers were recruited the personal environment.

3.2 Materials

3.2.1 Eye Movement Laboratory

The user study took place in the windowless Eye Movement laboratory at the Institute of Geography at the University of Zurich. Using the same room for all participants ensures that they are all tested under the same conditions (same room, constant lighting conditions and room temperature, same viewing distance to screen, etc.). The lab is equipped with a permanently installed computer and a Tobii[1] TX300 eye- tracker. The eye-tracker works with its own software called 'Tobii Studio'. For the pre- and post-questionnaire as well as for the spatial ability test an Acer [2] Aspire E15 laptop was used. The arrangement of the equipment is shown in Figure 3.1. Throughout the experiment participants had to change seats between the laptop and the permanently installed computer with the eye tracker several times.

[1]Tobii: http://www.tobii.com/

[2]Acer: http://us.acer.com/ac/en/US/content/home

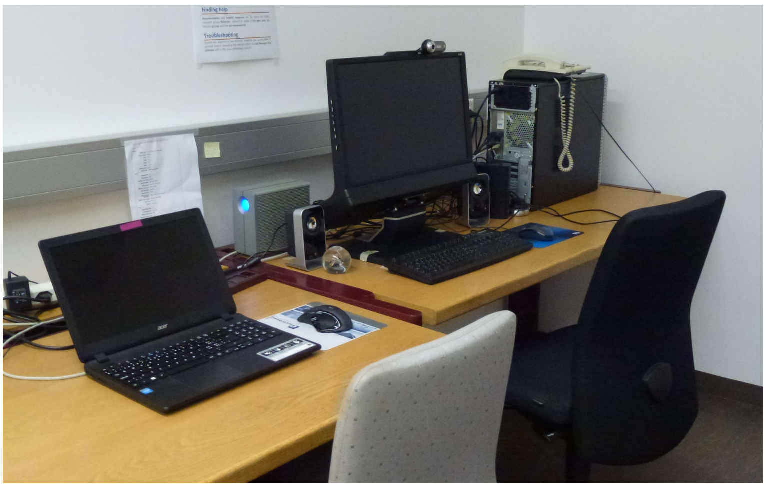


Figure 3.1: Eye movement laboratory. On the right the computer with the integrated eye tracker and on the left the additionally used Acer Aspire E15 laptop.

3.2.2 Pre- and Post-Questionnaire

To create the pre- and post-questionnaire an online survey tool called Survey Monkey[3] was used. This tool allows the design of surveys, collects responses and analyzes the results. In the appendix the complete pre- and post-questionnaire can be viewed.

Pre-Questionnaire

In the pre-questionnaire participants had to answer eleven different questions about their gender, age, level of education, handedness, visual impairments (such as visual aids and colour blindness) and hours of sleep. They also had to rate their experience in different study related fields (e.g. satellite imagery, graphic design, Google Earth or other online map providers, cartography, etc.) on a 5-point Likert Scale ranging from no experience (1) to professional (5) (see appendix).

Post-Questionnaire

In the post-questionnaire participants were asked if they had a preference between non-stereo and stereo images, and if they had participated in a similar experiment before. They also had to rate their level of discomfort with anaglyph viewing on a 5-point Likert Scale (from low to high) and how boring/ tiring they experienced the task to be, again with a 5-point Likert Scale: low to high. Participants were also asked if they noticed a contradiction between land form and land cover (e.g. snow, vegetation, river). The aim of this question was to investigate, if the participants experienced the terrain reversal effect. At the end of the experiment, the terrain reversal effect was been briefly explained

[3] Survey Monkey: https://www.surveymonkey.com/

in text form and with illustration. Afterwards participants were asked if they had noticed the described phenomenon.

3.2.3 Ability Tests

Spatial Ability Test

Spatial ability is a broad term. So broad in fact that measures of spatial ability sometimes equate with measures of intelligence (McGrew & Flanagan 1998 (cited in Lobben 2007)). Geographers are generally interested in investigating spatial ability and its influence on environmental task (e.g. reading a map, environmental perception, etc.). Geographic spatial ability can be understood as the ability to think geometrically (e.g. perceive three-dimensional structures two-dimensionally) (Lobben 2007). In this user study a Pattern Folding test was used to investigate the spatial ability of the participants. The aim of the test is to interpret three-dimensional shapes based on a two-dimensional projection (ADA 2007). In Figure 3.2 an example of the Pattern Folding Test is shown. On the left part of the image an unfolded figure is presented. On the right part of the image four different folded figures are shown but only one is the correctly folded version. The test is composed of 15 items. Each participant had a time limit of six minutes.

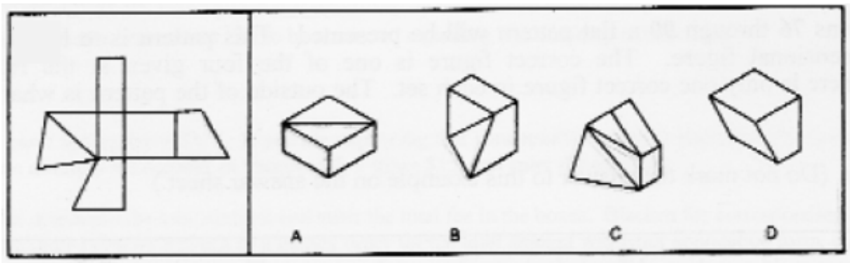


Figure 3.2: Pattern Folding Test. On the left, an unfolded figure is presented. On the right, four figures are shown, but only one is the correctly folded version of the left figure. (ADA 2007).

Stereoscopic Vision Test

To test if the participants had a stereoscopic vision, the Lang-Stereotest was used. Only using stereoscopic vision are participants able to see the stereoscopic images from the main experiment. The test was invited by a Swiss ophthalmologist called Joseph Lang in the 1980s. It is based on two principles: on random dots and on cylinder gratings. Random dots are used in stereograms. Monocularly seen no clues regarding the form are visible, whereas in binocular vision stereoscopic forms are recognizable. The disadvantage of most random dots is that they require glasses. The other principles consists of cylindrical screens. They were invented by W. R. Hess, a Swiss ophthalmologist and physiologist, in 1912. A system of fine parallel cylindrical strips separates the images for both eyes. Two fine strips of pictures, one seen by the right, the other seen by the left eye, are beneath each cylinder. Combining these two methods enables the omission of glasses what facilitates the handling. (Lang 1983; Lang 1988). In Figure 3.3 the Lang-Stereotest is shown. If participants were able to see the cat, the star and the car they had stereoscopic vision and were able to participate in the experiment.

Figure 3.3: Lang-Stereotest. A method used to test if people have stereoscopic vision. When participants can see the star, the cat and the car on the test plates they have stereoscopic vision.

3.2.4 Main Experiment

The main experiment consisted of two parts: a non-stereo and a stereo part. The two sections were designed with the online tool Survey Monkey and the software Tobii Studio was used as a web-stimuli. Participants were told to solve the two parts as quickly as possible. The experimental parameters in each section were the same, except for the visualization type, i.e. the same questions and natural regions were used only once shown non-stereoscopically and once stereoscopically. For the stereo part participants had to put on the anaglyph glasses to see the stereoscopic images on the screen. The anaglyph glasses consisted of one cyan and one red lens. When using the glasses an additional depth perception is produced. The principle behind anaglyph glasses is explained in section 2.2.

3.3 Experimental Design

A mixed 2 (task type) x 3 (level of expertise) x 6 (stimuli) factorial design was developed for the experiment (Martin 2008). The factors are task type (Task Type 1 and Task Type 2), level of expertise (non-geographer, geographer and experts) and stimuli (non-stereo subtle change, non-stereo middle rugged, non-stereo highly rugged, stereo subtle change, stereo middle rugged, stereo highly rugged). It is a mixed factorial design because the stimuli and task type have a within-subject design (all participants received the same stimuli and task type), whereas the level of expertise has a between-subject design (three different groups of participants) (Martin 2008). For Task Type 1, two different questions were created, for Task Type 2, four different questions. In Figure 3.4 the schematic representation of the mixed factorial design is illustrated.

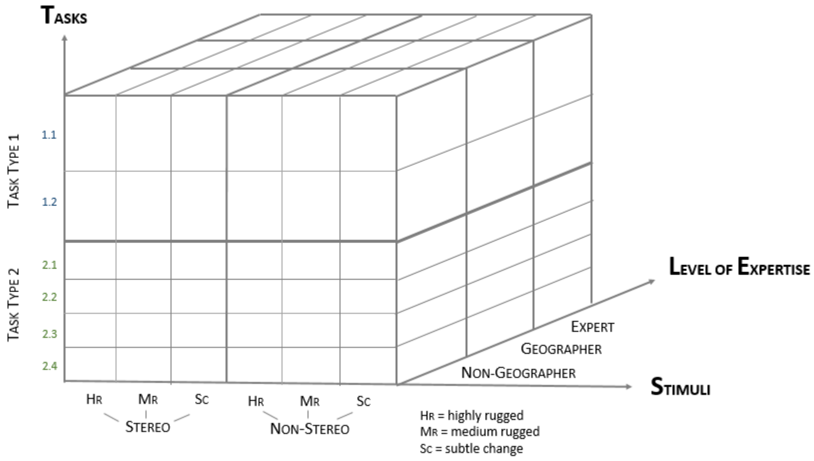


Figure 3.4: A 2 x 3 x 6 factorial design was developed for the experiment. The factors are task type (Task Type 1 and Task Type 2), level of expertise (non-geographers, geographers and experts) and stimuli (non-stereo subtle change, non-stereo middle rugged, non-stereo highly rugged, stereo subtle change, stereo middle rugged, stereo highly rugged)

3.3.1 Experimental Procedure

Before starting with the experiment, a pilot test was conducted. It revealed that the word order of the Likert Scale questions should not be changed because it can lead to confusions and mistakes.

After the pilot test, participants were recruited personally from the University of Zurich and the personal environment. Each participant who was willing to take part in the study, received an e-mail with some information about the experiment, a doodle link and the consent form (see appendix). They were kindly asked to fill in the doodle link as soon as possible and to read and sign the consent form before coming to the eye-movement lab. This enabled a saving of time. The experimental sessions were conducted during May 2015.

When the participants arrived, they were welcomed and handed in the signed consent form or signed it right then and there. To ensure anonymity, each participant received a code number. To make sure that all the participants received the same information brief instructions were given verbally. They were either given in German or in English, depending on the preference of the participant. The written instructions and tasks were all in English. If participants had any questions during the experiment they were free to ask them at any point during the experiment. However only formal questions (e.g. about the correct understanding of the task or the English words) were answered. Content related questions could only be answered after the experiment was finished. Otherwise some answers would have revealed too much information about the aim of the experiment. Before starting with the pre-questionnaire part, the Lang-Stereotest was conducted to test if participants had stereoscopic vision (see section 3.2.3). All participants could see the star, the car and the cat on the test plate. They then were allowed to continue with the pre-questionnaire part on the laptop (see section 3.2.2). For the main part of the experiment, participants had to move to the permanently installed computer with the eye tracker (section 3.1). As mentioned in section 3.2.4, the main part of the experiment was divided into two parts: a non-stereo and a stereo part. Half of the participants started with the non-stereo part, half with the stereo part. This counterbalancing should minimize an order effect (Martin 2008). The eye tracker was used for both parts. Therefore a brief instruction for the eye tracker and a calibration was made. Participants were advised to sit in a comfortable position on the chair because during the recording they were just allowed to move their eyes and not their whole body. Between the two main parts, a spatial ability test was integrated (see section 3.2.3). Therefore participants had to move back to the laptop. Finally participants had to fill in the post-questionnaire on the laptop (see section 3.2.2). To thank the participants for taking part in the experiment, some sweets were handed out at the end of the study. On average the whole experiment lasted about one hour. In Figure 3.5 the complete experimental procedure is illustrated.

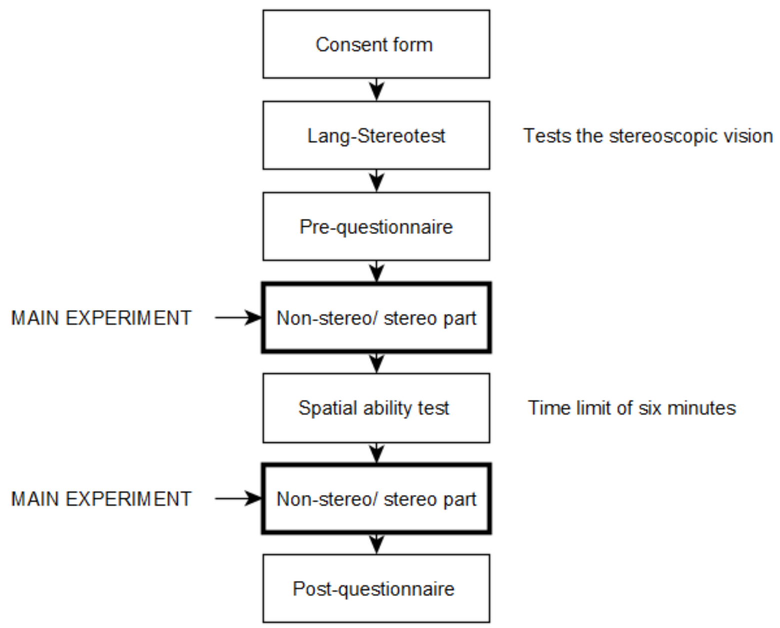


Figure 3.5: Experimental procedure

3.3.2 Independent Variables

Stimuli

To create the stimuli the software Google Earth[4] was used. Google Earth provides access to satellite images and different tools like zooming or location marking as well as information about elevation and scale. A total of thirty different natural locations displaying the effect were subjectively chosen by the researcher (see appendix). Some of these regions had also been used by Biland (2014). The zooming tool allowed to control that the effect really occurred in the image. This was possible because it enabled an oblique view of the environmental region (see Figure 3.6). Google Earth's altitude above sea level indication was also used as a control tool (Bernabé-Poveda & Çöltekin 2014; Biland 2014). Each of the thirty chosen regions were repeated eight times: image orientation (north-oriented and 180° rotated), visualization type (non-stereo and stereo) and task type (Task Type 1, Task Type 2). This made a total of 240 items per participant. In the appendix an overview of the image compilation is illustrated.

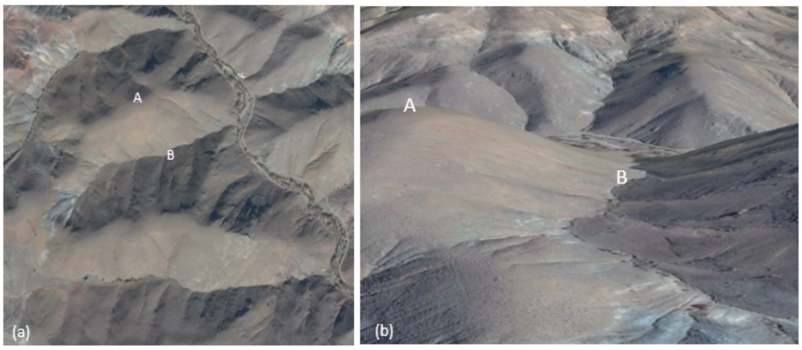


Figure 3.6: Aerial and oblique view in Google Earth. (a) The left image occurs in aerial view and a terrain reversal effect can be experienced. (b) In the oblique view the real landscape is seen.

The environments were selected from the *northern hemisphere* because it is believed that the terrain reversal effect occurs mostly on the northern hemisphere in satellite images (Toutin 1998 (cited in Saraf et al. 2007); Rudnicki 2005; Bernabé-Poveda et al. 2005; Saraf et al. 2007). This assumption is supported by an experiment from Bernabé-Poveda & Çöltekin (2014) and Biland (2014)(see section 2.1).

In their study, Bernabé-Poveda & Çöltekin (2014) mentioned, that when people are familiar with a region, they might know the correct answer. Therefore known *landmarks* were avoided and most of the images were chosen from the Asian continent.

Another aspect which was considered during sampling selection, was the *content* of the image. This had also been done by Bernabé-Poveda & Çöltekin (2014) and Biland (2014). It is important to note that only a single variable changes from one condition to

[4] Google Earth: https://www.google.com/earth/

the next and that different factors are counterbalanced. An even amount of land form types (half convex, half concave) and land cover cues (with and without cues like snow or river), was tried to be obtained. Through counterbalancing the land form orientation (0° N, 45° NE, 90° E and 135° SE), every changing light condition was tried to be considered. With the terrain type (subtle change, middle rugged and highly rugged) a similar overall altitude variation is guaranteed. The viewing height is dependent on the different terrain types. All the images had the same aerial perspective.

In all the selected satellite images the terrain reversal effect occurred. To avoid learning effects, *images without illusion* were also necessary. Therefore the same images were turned by 180°. This correction method is found throughout literature (see 2.1) and was first described by Saraf et al. (1996). Figure 3.7 gives an example of this method. In the left image the illusion occurs whereas in the 180° image, the illusion has disappeared. A total of 120 images were oriented north, the other 120 images were rotated 180°.

Figure 3.7: Same image. On the left side, the image is with terrain reversal effect, on the right side, the image is turned 180° and the effect disappears.

To investigate if the effect is most pronounced in highly rugged environments, like it is mentioned in literature (e.g. Saraf et al. 1996), three different terrain types were defined: subtle change, middle rugged and highly rugged. For the *terrain type*, the height difference between the lowest and the highest point of the observed valley/ ridge was measured. If it was lower than 50 m, it was defined as subtle change. When it was between 50 m and 500 m it was determined as middle rugged. With over 500 m difference it was defined as highly rugged. These three categories had been defined based on qualitative judgements. For each terrain type ten different regions were selected. To get a better understanding of the three terrain types an example for each category is shown in Figure 3.8. To experience the various terrain types better they are shown in an oblique view. The images are from Google Earth [5].

To investigate if *stereopsis* makes any change in experiencing the terrain reversal effect, non-stereo images as well as stereo images were generated. Each part consisted of 120

[5] Google Earth: https://www.google.com/earth/

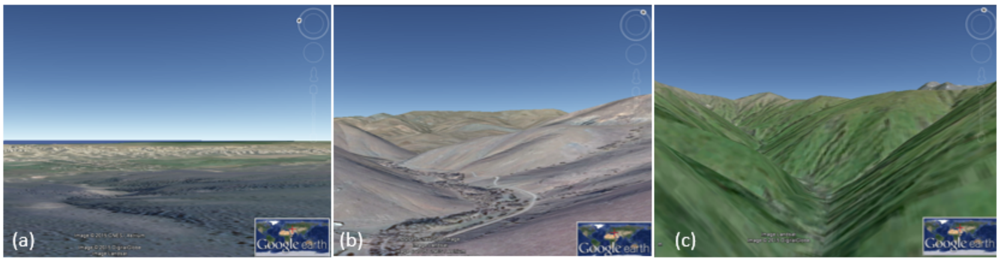


Figure 3.8: Oblique view of the different terrain types. Examples for a) subtle change b) middle rugged c) highly rugged terrain type.

items. The experimental parameters in each section were the same, except for the visualization type, i.e. the same questions and natural regions were used, only once shown non-stereoscopically and once stereoscopically. For creating the stereo images, the software StereoPhoto Maker[6] was used. The software is connected to Google Earth and generates a left and a right image of the chosen region (see Figure 3.9). These two slightly different images are merged by the program and allow for three-dimensional perception when seen with anaglyph glasses (see section 2.1).

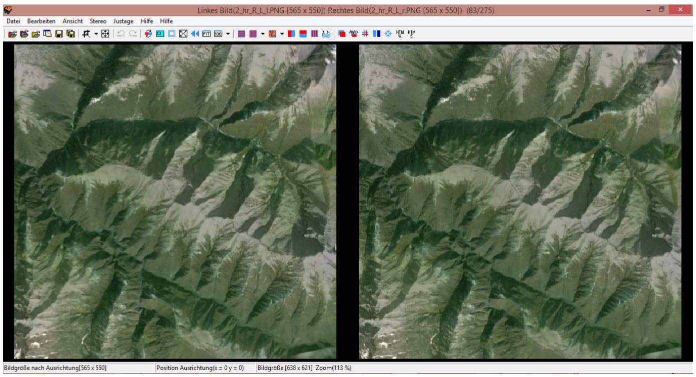


Figure 3.9: StereoPhoto Maker. Two slightly different images of the same region. Subsequently the program merges these images, which allows for three-dimensional perception when seen with anaglyph glasses.

Task Type

The terrain reversal effect in satellite images has only been investigated through perceptual, base shape recognition tasks (Bernabé-Poveda & Çöltekin 2014; Biland 2014). Participants had to identify 3D landforms as well as judge 3D spatial relationships (Bernabé-Poveda & Çöltekin 2014). For this study, an additional task was included, so that two different types of tasks could be distinguished: perceptual tasks (Task Type 1) and more interpretive tasks (Task Type 2). A total of two different questions for Task

[6]StereoPhoto Maker: http://stereo.jpn.org/eng/stphmkr/

Type 1 and four different questions for Task Type 2 were generated. Various questions for each task type allow for validation of the results because observing the same thing twice is more reassuring than observing it once. Generally the tasks needed to be solvable and lead to "correct" answers. Because the terrain reversal effect is an illusion, there actually is no right or wrong answer. Each participant perceives the effect different. For the analysis the correct answer was defined as the real shape of the observed object without illusion. So even if people see a "ridge" but it is actually a valley, valley would be the correct answer. Participants were told to answer the questions as quickly as possible. This should prevent them from interpreting too much especially during Task Type 1 questions. Likert Scales were integrated in the questions. They enabled to indirectly measure the confidence rate of the participants. One fourth of the questions did not have a Likert Scale.

Task Type 1

Task Type 1 contained perceptual, base shape recognition items. The items were adopted from Bernabé-Poveda & Çöltekin (2014). The first item asked participants to rate the height of A in regard to B. They could choose between: A is clearly higher than B (1), A is higher than B (2), ambiguous (3), B is higher than A (4), B is clearly higher than A (5) (see left image from Figure 3.10). By the second question they had to say if the line "ABC" appears as: clearly a valley (1), a valley (2), ambiguous (3), a ridge (4), clearly a ridge (5) (see right image from Figure 3.10).

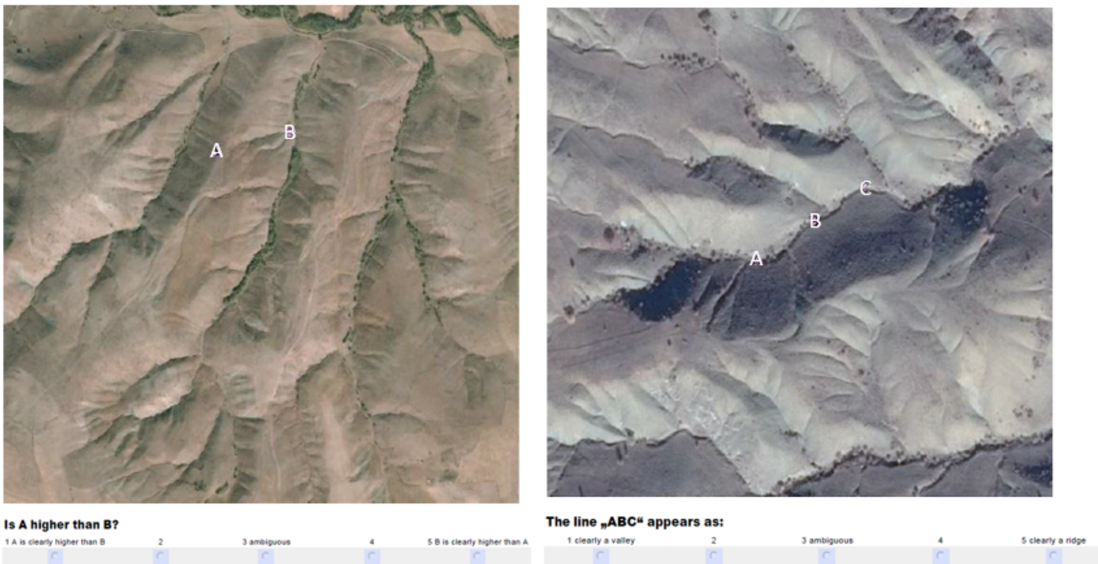


Figure 3.10: Examples for Task Type 1. On the left image participants were asked if A is higher than B. They could choose between: A is clearly higher than B (1), A is higher than B (2), ambiguous (3), B is higher than A (4), B is clearly higher than A (5). On the right images they were asked, if the line "ABC" appears as: clearly a valley (1), a valley (2), ambiguous (3), a ridge (4), clearly a ridge (5)

Task Type 2

Task Type 2 contained items which forced the participants to look at a wider area to be able to answer the item correctly. They should study the region more holistically, and not only perceive locally if something is convex or concave. So the items were interpretative. But the items should have still been simple enough to be answered within a certain time limit. Because there were different levels of expertise, the items should have been understandable for all three levels.

Four different questions were created. Two of the questions were about the physical law of gravitation. Participants had to tell if they thought a rock would roll/ water would flow from A to B or from B to A. They could choose between: clearly from A to B (1), from A to B (2), ambiguous (3), from B to A (4), clearly from B to A (5). The other two questions were more from the "planning field". Participants had to tell where they thought would be a good location for a cable car base station and were they thought they would have an overview of the landscape. For these questions they could choose between A, B and C as well as multiple answers. In Figure 3.11 the questions from the "planning field" are illustrated. As can be seen in the right example, a cable car base station would not be appropriate in this region. Participants were instructed to just choose the most logical answer, even if this were not transferable to reality.

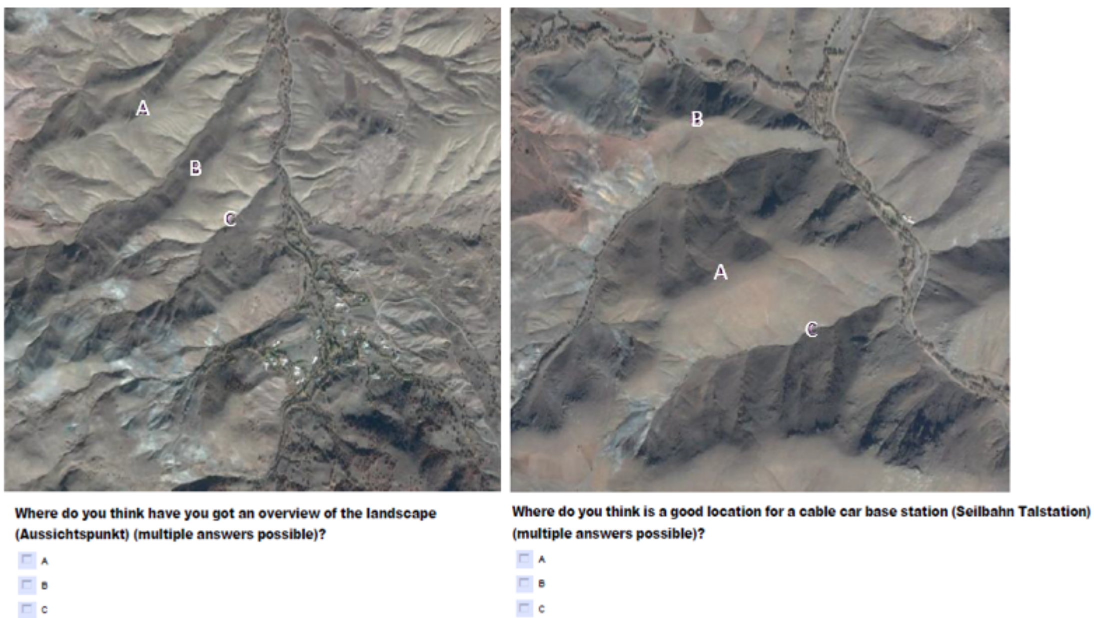


Figure 3.11: Examples for Task Type 2. On the left: Where do you think would you have an overview of the landscape (multiple answers possible)? On the right: Where do you think is a good location for a cable car base station (multiple answers possible)

Level of Expertise

As mentioned in section 3.1, participants were divided into three groups depending on their level of expertise in the field of remote sensing. The groups were defined as expert, geographer and non-geographer. For more detail see section 3.1.

3.3.3 Dependent Variables

Accuracy (Effectiveness)

The accuracy of each participant was scored as a percentage. Therefore the total amount of correct answers was taken in regards to the total number of possible correct answers. Because the terrain reversal effect is an illusion actually there is no right or wrong answer. Each participant perceives the effect different. For the analysis the correct answer was defined as the real shape of the observed object without illusion. So if people saw a "valley" and it was actually a valley, it was counted as the correct answer.

Each question required an answer, otherwise participants could not move on to the next question. This ensured that all questions were answered.

Response Time (Efficiency)

The response time was calculated using the software Tobii Studio. It was the defined as the time participants needed to answer a question. Time started once the question appeared on the screen and ended when the participants clicked on their answer and moved on to the next question. For the final analysis, the sum of mean response times per participant divided by the total number of participants was calculated. Because the terrain reversal effect is a perceptual phenomenon, time was analyzed for all answers, not only for those which were solved "correctly".

Confidence

To measure the confidence rate of the participants, a Likert Scale was integrated in most of the questions. The Likert Scale ranged from 1 to 5. With the use of the scale participants rated indirectly if they had felt confident while giving their answers. When one, two, four or five were chosen, participants were confident with their answers. For the evaluation all answers, except the ambiguous ones were considered. In Figure 3.12 an example of an applied Likert Scale is shown.

Figure 3.12: Example of an item with a Likert Scale. The Likert Scale ranges from 1 to 5. When one, two, four or five was chosen, participants were confident about their answers.

Preference

The preference for non-stereo or stereo images was investigated by asking the participants at the end of the experiment (in the post-questionnaire part (see section 3.2.2)) if, why and which visualization type they preferred.

3.4 Statistics

All analysis were performed with the statistical software IBM SPSS Statistics 21[7]. The theoretical background was provided by the book "Discovering Statistics using SPSS" from Field (2009) as well as documents from the statistical course GEO 246[8] from the Department of Geography at the University of Zurich.

Different statistical tests were used, depending on the characteristics of the data. To decide which statistical test was most appropriate, decision trees from GEO 246 were consulted (see appendix).

The descriptive data are presented in means (M) and standard error of the mean (SEM). When the Wilcoxon Signed Rank Test was used, the median (Mdn) was reported. Field (2009) mentioned that in this case the median is a more appropriate value than the mean.

Depending on the test, the results are reported different. The results are reported as in Field (2009) suggested. When the p-value was less than 0.05, results were considered as statistically significant.

To visualize the data, the software Microsoft Excel 2013[9] was used. It allows to generate diagrams with integrated error bars. The accuracy and the confidence rate are presented in percentage [%], whereas the response time is reported in seconds [s].

[7] IBM SPSS Statistics 21: http://www-01.ibm.com/software/analytics/spss/products/statistics/
[8] Statistical course GEO 246: http://www.vorlesungen.uzh.ch/FS14/lehrangebot/fak-50000008/sc-50503822/cga-50503822010/cg-50017196/sm-50529412.modveranst.html
[9] Microsoft Excel 2013: http://www.microsoft.com/de-ch/

4 Results

In this chapter, the results from the experiment are presented. First, the main findings from the pre- and post-questionnaire are reported. Then, the main effects (effect of image orientation, visualization type, level of expertise, task type and terrain type) are described, and following that, the interactions between the visualization type and the other factors (level of expertise, task type and terrain type) are examined. Therefore accuracy, confidence rate and response time were all analysed. As a reference, the 180° turned images were evaluated. But these results are only marginally discussed, because the main focus lies on the images with an illusion. All diagrams from the section 4.2 and the section 4.3 can be found in the appendix, the most relevant ones are integrated in the text.

4.1 Participants

Thirty-three participants, eighteen women and fifteen men, took part in the study. They were between twenty and sixty years old, but the majority lied within the range of twenty and thirty years of age (94%). The participants were divided into three groups, depending on their level of expertise (see section 3.1). A total of 10 experts (5 women and 5 men), 11 geographers (7 women, 4 men) and 12 non-geographers (6 women, 6 men) participated in the study. Table 4.1 provides an overview of the three groups. The participants had different level of education: less than high school (3%), high school (9%), bachelor degree (48%), master degree (33%) and doctoral degree (6%).

Level of expertise	Women	Men	Total
Expert	5	5	10
Geographer	7	4	11
Non-Geographer	6	6	12
Total	18	15	33

Table 4.1: Level of expertise. The table provides an overview of the three different participant groups. It also shows the amount of women and men in each group.

Level of Experience

In the pre-questionnaire the participants had to rate their experience in different fields. The results are shown in Figure 4.1. Not surprisingly, experts rated their experience with satellite imagery as highest, followed by the geographers and the non-geographers. Also the experience with photo-interpretation and online map providers like Google Earth had been rated highest by the experts, followed by the geographers and the non-geographers. In the field of cartography, geographers rated their experience as highest, followed by the experts and the non-geographers. In the field of photography, graphic design and fine arts, non-geographers rated their experience as highest.

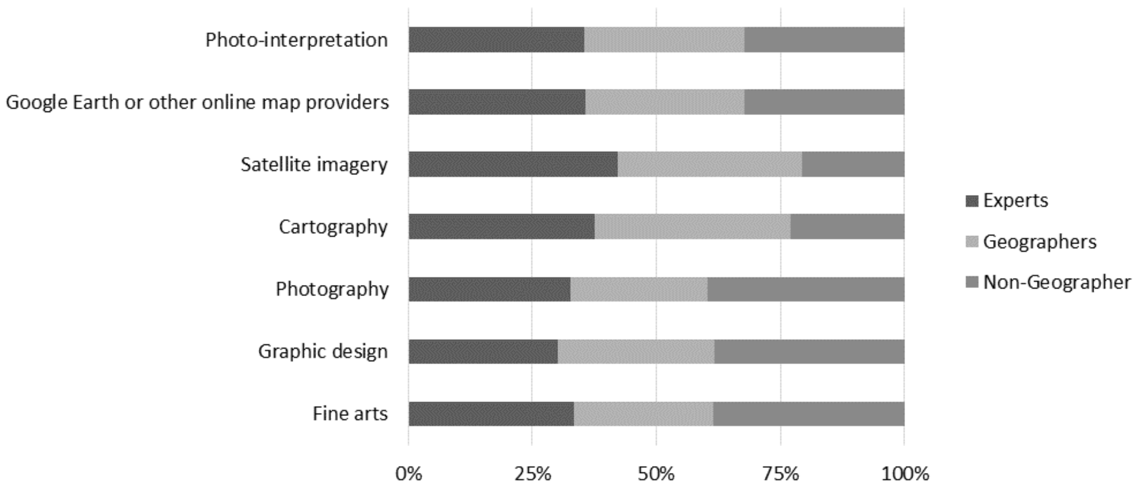


Figure 4.1: Level of experience for seven study related fields. Participants had to rate their level of experience. The results are grouped by level of expertise (experts, geographers, non-geographers.

Anaglyph Viewing

In the post-questionnaire participants were asked if and why they preferred non-stereo or stereo images. Overall, the evaluation showed no preference for a visualization type. 45% of the participants preferred the non-stereo images, 48% the stereo images and 6% neither (Figure 4.2). Also within the level of expertise groups the preference was similar. The main reasons for preferring non-stereo images were the familiarity with these images in everyday life and wearing the anaglyph glasses was considered exhausting and uncomfortable. The main reason for preferring stereo images was the better recognition of the terrain shape because the 3D effect was stronger. Some of the participants also mentioned that they felt more secure in giving the answers for stereo images and that they were more comfortable using stereo images because reality is normally perceived in 3D. The participants were also asked about the level of discomfort with anaglyph viewing. Most participants rated the discomfort level with anaglyph glasses as low (39%) or medium (36%). 24% rated the discomfort level with anaglyph glasses as high.

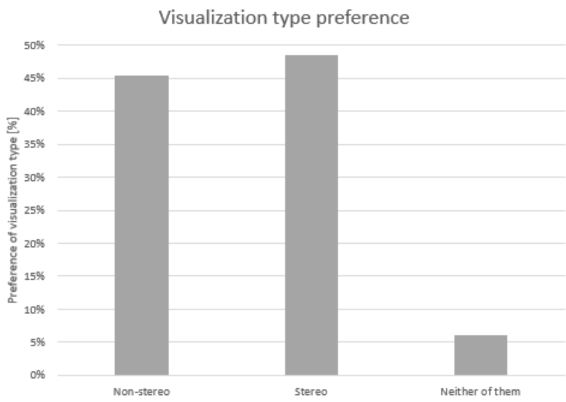


Figure 4.2: Preference between non-stereo and stereo images or neither [%].

Terrain Reversal Effect

In the post-questionnaire participants were asked if they perceived a contradiction between land form and land cover (e.g. snow, vegetation, river). 76% of the participants noticed such a difference, 24% did not. They also had to explain the observed contradiction. Most perceived contradictions were unnatural distributions of snow and vegetation between valley and ridges (more snow in the valley than on the ridge or more vegetation on the ridge than in the valley). Another observed contradiction were rivers on a ridge. When a contradiction was perceived, 52% answered the question based on their perception and 36% based on their interpretation. Only 12% could not remember how they had answered the question (Figure 4.3).

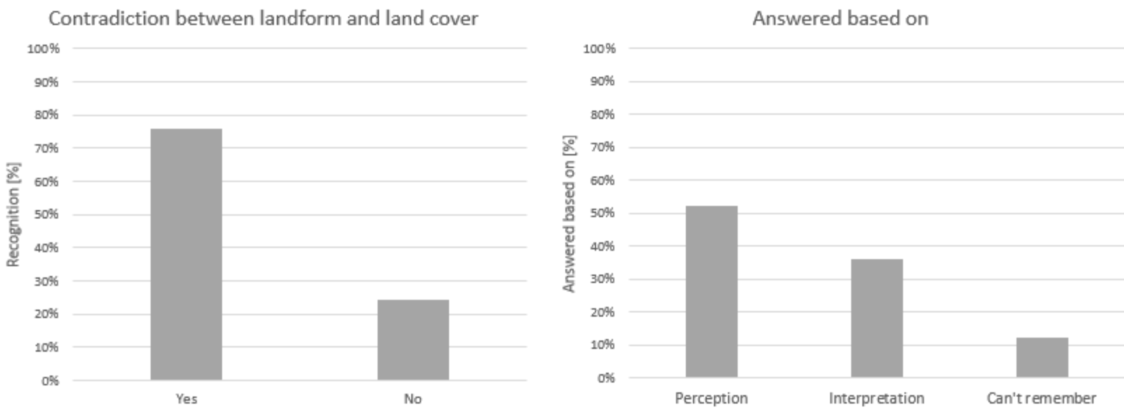


Figure 4.3: Participants who recognized contradiction between land form and land cover and those who did not in % (left). When they recognized a contradiction, they answered either based on their perception (52%), their interpretation (36%) or that they could not remember (12%) (right).

At the end of the experiment, the terrain reversal effect was briefly explained in text form and with illustrations. Afterwards participants were asked if they had noticed the phenomenon. 85% mentioned that they had noticed the effect, 15% had not (Figure 4.4). They were also asked if they had previous participated in a similar experiment. From all of the thirty-three participants only one attended a similar experiment before.

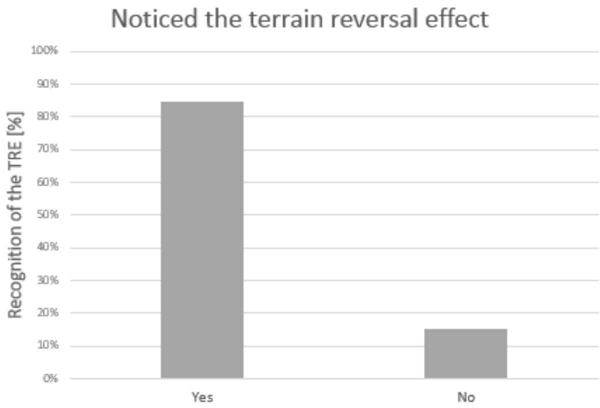


Figure 4.4: Percentage [%] of participants that have recognized and not recognized the terrain reversal effect.

4.2 Main Effects

4.2.1 Image Orientation

Accuracy

A paired-samples t-test showed on average, participants had a significantly higher accuracy with the 180°-rotated images (M = 90.76, SE= 0.869) than with the north-oriented (0° turned) images ((M= 28.97, SE= 3.345), t(32)= - 16.940, p= .000). Figure 4.5 shows the accuracy of the north-oriented and 180°-turned images.

Confidence

A paired-samples t-test showed, no significant effect on the confidence rate of the north-oriented images (M= 94.09, SE= .966) and the 180°-rotated images (M= 94.33, SE= .646), t(32)= -.230, p= .820. Figure 4.5) shows the confidence rate of the north-oriented and 180°-turned images.

Response Time

A Wilcoxon Signed Ranks Test showed that there was no significant effect on the response time of the north-oriented images (*Mdn*= 1016.00) and the 180°-rotated images (*Mdn*= 983.00), z= -1.188, p= .235. Figure 4.5 shows the response time of the north-oriented and 180°-turned images.

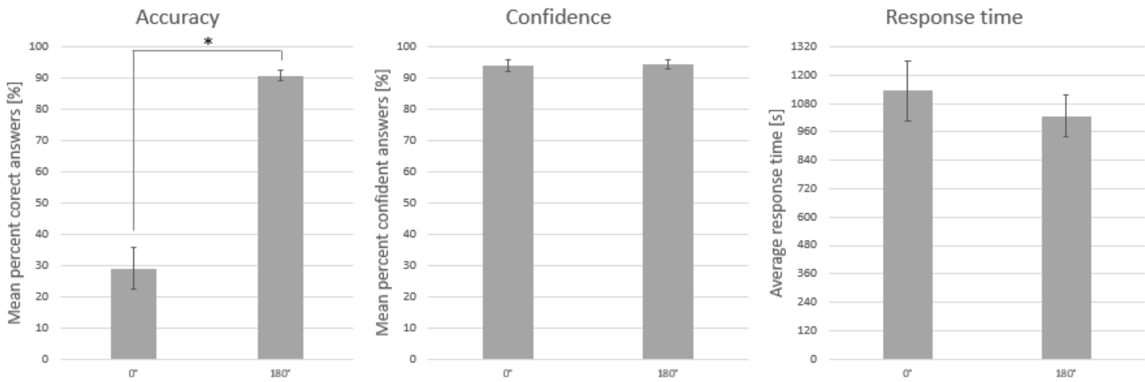


Figure 4.5: Mean accuracy, mean confidence and mean response time of north-oriented and 180°-turned images.

4.2.2 Visualization Type

Accuracy

A Wilcoxon Signed Ranks Test showed that the accuracy of the stereo images was significantly higher (*Mdn*= 32.00) than the non-stereo images (*Mdn*= 15.00), z= -3.774, p= .000. This result would point to the fact that people are better at solving the given tasks with the stereo images than with the non-stereo images (see Figure 4.6).

Looking at the 180°-turned images, a significant difference between the accuracy of non-stereo images (*Mdn*= 92.00) and stereo images (*Mdn*= 93.00) is also visible. A Wilcoxon Signed Ranks Test showed that the accuracy of the 180°-turned stereo images is significantly higher than of 180°-turned non-stereo images, z= -2.840, p= .005.

In both conditions (north-oriented and 180°-turned images) participants are more accurate using stereo than non-stereo images.

Confidence

A Wilcoxon Signed Ranks Test showed no significant difference between the confidence rate of the non-stereo images (*Mdn*= 96.00) and the stereo images (*Mdn*= 96.00), z= -1.880, p= .060. Thus means that the visualization type has no influence on the confidence rate of the participants.

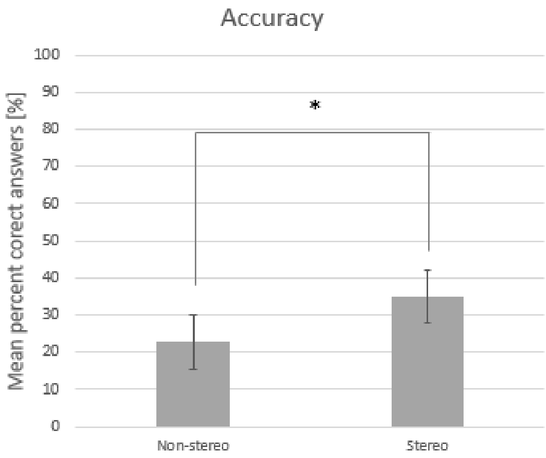


Figure 4.6: Mean accuracy of the two visualization types non-stereo and stereo.

Looking at the 180°-turned images there is, in contrast to the north-oriented images, a significant difference between the confidence rate of non-stereo images (*Mdn*= 93.00) and the stereo images (*Mdn*= 98.00). A Wilcoxon Signed Ranks Test revealed that the confidence rate of the stereo images is significantly higher than of the non-stereo images, z= -3.779, p= .000.

The north-oriented images had no influence on the confidence rate between non-stereo and stereo images, whereas the 180°-turned images had one. Participants felt more confident using stereo images.

Response time

A paired-samples t-test showed no significant effect between the response time of non-stereo images (M= 576.94, *SE*= 40.783) and stereo images (M= 556.39, *SE*= 30.087), $t(32)$= .626, p= .536. This means that the visualization type has no influence on the response time of the participants.

Looking at the 180°-turned images, there was also no significant effect between non-stereo images (M= 517.03, *SE*= 28.488) and stereo images (M= 510.33, *SE*= 24.555), $t(32)$= .224, p= .824. The visualization type does not seem to have an influence on the response time of the participants when looking at the 180°-turned images.

In both conditions (north-oriented and 180°- turned images), visualization type had no influence on the response time of participants.

4.2.3 Expertise

Accuracy

A Kruskal-Wallis Test showed no significant effect on the accuracy between the three levels of expertise, $H(2)= 1.852$, $p= .396$. However, non-geographers ($M= 23.17$, $SE= 2.467$) were least accurate, followed by the geographers ($M= 29.09$, $SE= 6.757$) and the experts ($M= 35.80$, $SE= 7.544$) but without significance (see Figure 4.7).

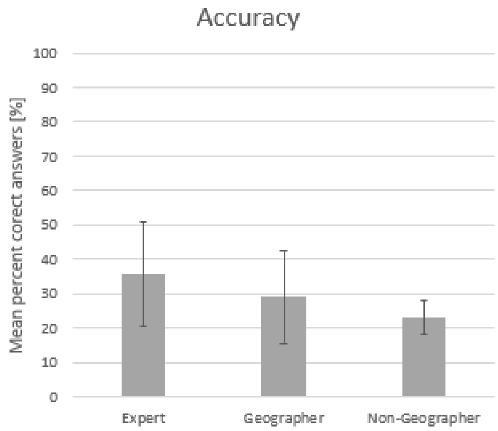


Figure 4.7: Mean accuracy of the three level of expertise experts, geographers and non-geographers.

Looking at the 180°-turned images a Kruskal-Wallis Test also showed no significant effect on the accuracy between the three levels of expertise, $H(2)= 2.526$, $p= .283$. In comparison to the north-oriented images geographers ($M= 92.64$, $SE= .754$) and non-geographers ($M= 90.33$, $SE= 1.144$) tended to be more accurate than experts ($M= 89.20$, $SE= 2.351$) but without significance.

In both conditions (north-oriented and 180° turned images), level of expertise had no influence on the accuracy.

Confidence

A Kruskal-Wallis Test showed no significant effect on the confidence rate between the three levels of expertise, $H(2)= 1.868$, $p= .393$. However, geographers ($M= 95.64$, $SE= 1.350$) were most confident followed by the experts ($M= 94.80$, $SE= 1.373$). Least confident was reported by the the non-geographers ($M= 92.08$, $SE= 2.013$) but without significance.

Looking at the 180°-turned images, a one-way ANOVA test showed no significant effect on the confidence rate between the three levels of expertise non-geographer ($M= 93.17$, *SE*= 1.375), geographer (*M*= 95.55, *SE*= .731) and expert (*M*= 94.40, *SE*= 1.056) , *F*(2,30)= 1.194, *p*= .317.

In both conditions (north-oriented and 180°-turned images), level of expertise had no influence on the confidence rate.

Response Time

A Kruskal-Wallis Test showed no significant effect on the response time between the three levels of expertise, *H*(2)= 1.684, *p*= .431. However, non-geographers (*M*= 1017.92, *SE*= 63.148) were faster than experts (*M*= 1124.90, *SE*= 102.863) and geographers (*M*= 1267.18, *SE*= 149.196) but without significance.

Looking at the 180°-turned images, a one-way ANOVA test revealed that there was no significant effect on the response time between the three levels of expertise, *F*(2,30)= .688, *p*= .510. However, non-geographers (*M*= 962.42, *SE*= 52.529) were faster than geographers (*M*= 1085.45, *SE*= 96.934) and experts (*M*= 1041.20, *SE*= 78.686) but without significance.

In both conditions (north-oriented and 180°-turned images), level of expertise had no influence on the response time.

4.2.4 Task Type

Accuracy

A Wilcoxon test showed that the accuracy for Task Type 2 was significantly higher (*Mdn*= 27.00) than for Task Type 1 (*Mdn*= 22.00), *z*= -3.800, *p*= .000. This means that people are more accurate in solving Task Type 2 than Task Type 1 (see Figure 4.8).

Looking at the 180°-turned images, a significant difference between the accuracy of Task Type 1 (*Mdn*= 90.00) and Task Type 2 (*Mdn*= 93.00) was found. A Wilcoxon test showed that the accuracy for Task Type 2 is significantly higher than for Task Type 1, *z*= -2.949, *p*= .003.

In both conditions (north-oriented and 180°-turned images), participants were more accurate for Task Type 2 than Task Type 1.

Confidence

A Wilcoxon test showed no significant difference in the confidence rate of Task Type 1 (*Mdn*= 95.00) and Task Type 2 (*Mdn*= 97.00), *z*= -0.778, *p*= .436. This means that the task type had no influence on the confidence of the participants.

Looking at the 180°-turned images, there is, in contrast to the north-oriented images, a significant difference between Task Type 1 (*Mdn*= 97.00) and Task Type 2 (*Mdn*= 93.00). A Wilcoxon test showed that the confidence rate of Task Type 1 was significantly higher than of Task Type 2, z= -2.340, p= .019.

For the north-oriented images, no influence on the confidence rate between the two task types could be seen, whereas the 180°-turned images an influence was found. For those images participants felt more confident with Task Type 1 than Task Type 2.

Response Time

A paired-samples t-test showed, that there was a significant effect between the response time of Task Type 1 (*M*= 534.00, *SE*= 30.435) and Task Type 2 (*M*= 599.52, *SE*= 33.693), $t(32)$= -8.051, p= .000. This means that participants are faster with Task Type 1 than Task Type 2. (see Figure 4.8).

Looking at the 180°-turned images there was also a significant effect between the response time of Task Type 1 (*M*= 456.42, *SE*= 20.4661) and Task Type 2 (*M*= 570.76, *SE*= 24.972), $t(32)$= -14.082, p= .000. Participants were faster with Task Type 1 than Task Type 2.

In both conditions (north-oriented and 180°-turned images), participants were faster with Task Type 1 than Task Type 2.

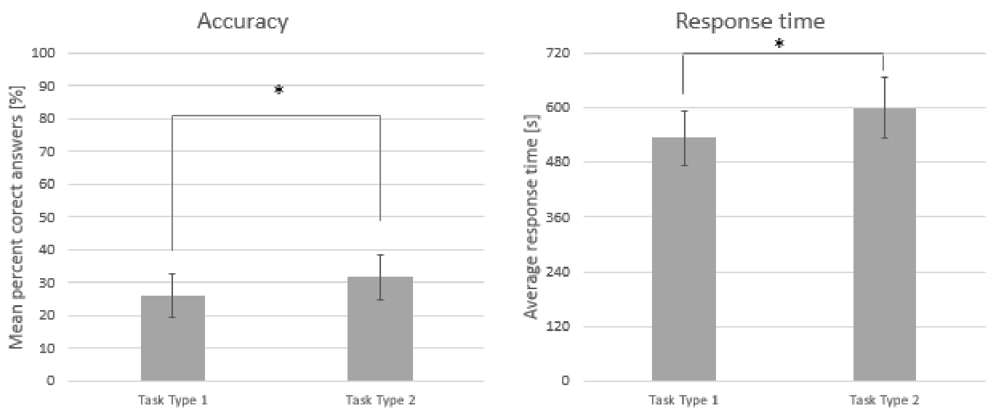


Figure 4.8: Mean accuracy and response time for Task Type 1 and Task Type 2.

4.2.5 Terrain Type

Accuracy

A Friedman test showed a significant effect on accuracy between the three different terrain types subtle change (*M*=24.42, *SE*= 3.985), middle rugged (*M*= 21.73, *SE*= 3.036) and highly rugged (M=41.33, SE= 3.673), $\chi^2(2)$= 37.181, p= .000. Wilcoxon test was used to compare each possible combination. The test results showed a significant difference between the accuracy of subtle change (Mdn= 20.00) and highly rugged (Mdn= 38.00) terrain, z= -4.677, p= .000 and between middle rugged (Mdn= 18.00) and highly rugged (Mdn= 38.00) terrain, z= -4.689, p= .000. There was no significant interaction between subtle change (Mdn= 20.00) and middle rugged (Mdn= 18.00), z= -1.030, p= .303 (see Figure 4.9).

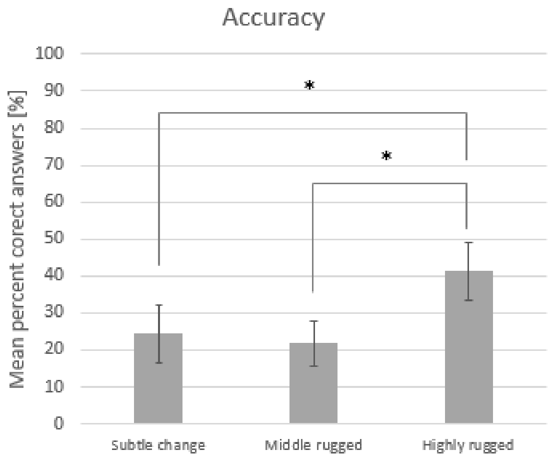


Figure 4.9: Mean accuracy of the three terrain types subtle change, middle rugged and highly rugged.

Looking at the 180°-turned images, a Friedman test also showed that there was a significant effect on accuracy between the three different terrain types subtle change (M=90.09, SE= 1.568), middle rugged (M= 88.58, SE= .989) and highly rugged (M=94.15, SE= .730), $\chi^2(2)$= 17.968, p= .000. Wilcoxon test was used to compare each possible combination. The tests showed a significant difference between the accuracy of subtle change (Mdn= 92.00) and highly rugged (Mdn= 95.00) terrain, z= -2.741, p= .006 and between middle rugged (Mdn= 90.00) and highly rugged (Mdn= 95.00) terrain, z= -3.914, p= .000. There was no significant interaction between the accuracy of subtle change (Mdn= 92.00) and middle rugged (Mdn= 90.00), z= -1.595, p= .111.

In both conditions (north-oriented and 180°-turned images) subtle change and highly rugged terrain as well as middle rugged and highly rugged terrain had a significant interaction effect.

Confidence

A Friedman test showed a significant effect on confidence rate between the three different terrain types subtle change (M=91.97, SE= 1.398), middle rugged (M= 95.94, SE= .839) and highly rugged (M= 94.24, SE= 1.259), $\chi^2(2)$= 7.000, p= .030. Wilcoxon test was used to compare each possible combination. The tests showed that there was a significant difference between the confidence rate of subtle change (*Mdn*= 93.00) and middle rugged (*Mdn*= 97.00) terrain, z= -2.521, p= .012. There was no significant interaction between the confidence rate of subtle change (*Mdn*= 93.00) and highly rugged (*Mdn*= 97.00) terrain, z= -1.582, p= .114 and between the confidence rate of middle rugged (*Mdn*= 97.00) and highly rugged (*Mdn*= 97.00) terrain, z= -1.753, p= .080.

Looking at the 180°-turned images, a Friedman test also showed, that there is a significant effect on confidence rate between the three different terrain types subtle change (M=89.82, *SE*= 1.017), middle rugged (M= 96.30, *SE*= .555) and highly rugged (M=96.82, *SE*= .816), $\chi^2(2)$= 36.162, p= .000. Wilcoxon test was used to compare each possible combination. The tests showed significant differences between the confidence rate of subtle change (*Mdn*= 90.00) and middle rugged (*Mdn*= 97.00) terrain, z= -4.478, p= .000, and between the confidence rate of subtle change (*Mdn*= 90.00) and highly rugged (*Mdn*= 100.00) terrain, z= -4.463, p= .000. There was no significant interaction between confidence rate of middle rugged (*Mdn*= 97.00) and highly rugged (*Mdn*= 100.00), z= -.667, p= .505.

In north-oriented images, subtle change and middle rugged terrain show a significant interaction effect. In 180°-oriented images, subtle change and middle rugged terrain, as well as subtle change and highly rugged terrain a significant interaction effect was found.

Response Time

A Friedman test showed no significant effect on response time between the three different terrain types subtle change (M= 386.12, *SE*= 21.781), middle rugged (M= 371.30, *SE*= 22.952) and highly rugged (M= 376.00, *SE*= 20.503), $\chi^2(2)$= 4.200, p= .122.

Looking at the 180°-turned images, a Friedman test showed a significant effect on response time between the three different terrain types subtle change (M= 330.61, *SE*= 14.438), middle rugged (M= 354.45, *SE*= 16.758) and highly rugged (M= 342.21, *SE*= 14.807), $\chi^2(2)$= 13.879, p= .001. Wilcoxon test was used to compare each possible combination. The tests showed a significant interaction between response time of subtle change (*Mdn*= 319.00) and middle rugged (*Mdn*= 334.00), z= -3.449, p= .001. There was no significant difference between the response time of subtle change (*Mdn*= 319.00) and highly rugged (*Mdn*= 337.00) terrain, z= -1.564, p= .118 and between the response time of middle rugged (*Mdn*= 334.00) and highly rugged (*Mdn*= 337.00) terrain, z= -.965, p= .335.

For the north-oriented images no significant effect on response time between the three different terrain types. In 180°-oriented images, subtle change and middle rugged terrain show a significant interaction effect.

4.3 Interactions

4.3.1 Visualization Type and Level of Expertise

Accuracy

A Kruskal-Wallis test showed the accuracy of non-stereo images was not significantly affected by the level of expertise, $H(2) = 2.388$, $p = .303$. The accuracy of stereo images was also not affected by the level of expertise, $H(2) = 1.040$, $p = .595$. However, experts (*Mdn*= 20.50) were most accurate using non-stereo images, followed by the geographers (*Mdn*= 13.00) and the non-geographers (*Mdn*= 12.50) but without significance. Experts (*Mdn*= 37.50) were also most accurate using stereo images, followed by the non-geographers (*Mdn*= 27.00) and the geographers (*Mdn*= 30.00) but without significance (see Figure 4.10).

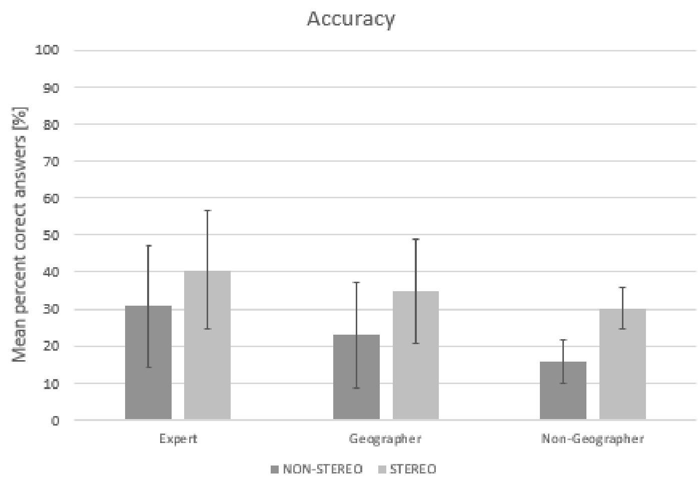


Figure 4.10: Mean accuracy of the interaction between visualization type and level of expertise.

Looking at the 180° turned images, a Kruskal-Wallis test showed the accuracy of non-stereo images was not significantly affected by the level of expertise, $H(2) = 1.359$, $p = .507$. The accuracy of stereo images was also not affected by the level of expertise, $H(2) = 1.136$, $p = .567$. However, geographers (*Mdn*= 92.00) were most accurate using non-stereo images, followed by the experts (*Mdn*= 90.50) and the non-geographers (*Mdn*= 88.00) but without significance. Geographers (*Mdn*= 93.00) were also most accurate using stereo images, followed by the experts (*Mdn*= 92.50) and the non-geographers (*Mdn*= 92.50) but without significance.

In both conditions (north-oriented and 180°-turned images), concerning the accuracy there was no significant effect between the visualization type and the level of expertise.

Confidence

A Kruskal-Wallis test showed the confidence rate of non-stereo images was not significantly affected by the level of expertise, $H(2)= .303$, $p= .859$. The confidence rate of stereo images was also not affected by the level of expertise, $H(2)= 2.908$, $p= .234$. However, geographers (Mdn= 98.00) were most confident using non-stereo images, followed by the experts (Mdn= 97.00) and the non-geographers (Mdn= 96.00) but without significance. Using stereo images, geographers (Mdn= 98.00) felt most confident, followed by the experts (Mdn= 96.00) and the non-geographers (Mdn= 91.00) but without significance.

Looking at the 180°-turned images, a Kruskal-Wallis test showed that the confidence rate of non-stereo images was not significantly affected by the level of expertise, $H(2)= 2.477$, $p= .290$. The confidence rate of stereo images was also not affected by the level of expertise, $H(2)= .855$, $p= .652$. However, experts (Mdn= 96.00) felt most confident using non-stereo images, followed by the geographers (Mdn= 93.00) and the non-geographers (Mdn= 92.00) but without significance. Using stereo images, the confidence rate for the three levels of expertise were nearly the same. Geographers (Mdn= 98.00) and non-geographers (Mdn= 98.00) felt most confident followed by the experts (Mdn= 97.00) but without significance.

For both conditions (north-oriented and 180°-turned images), there was no significant effect between the visualization type and the level of expertise concerning the confidence rate.

Response Time

A Kruskal-Wallis test showed the response times for non-stereo images were not significantly affected by the level of expertise, $H(2)= .759$, $p= .684$. The response times for stereo images were also not affected by the level of expertise, $H(2)= 1.601$, $p= .449$. However, non-geographers (Mdn= 485.00) were fastest using non-stereo images, followed by the experts (Mdn= 555.00) and the geographers (Mdn= 556.00) but without significance. Using stereo images, the experts (Mdn= 504.50) were fastest, followed by the non-geographers (Mdn= 514.00) and the geographers (Mdn= 632.00) but without significance.

Looking at the 180°-turned images, a Kruskal-Wallis test showed the response times for non-stereo images were not significantly affected by the level of expertise, $H(2)= .783$, $p= .676$. The response times for stereo images also were not affected by the level of expertise, $H(2)= .382$, $p= .826$. However, non-geographers (Mdn= 462.50) were fastest using non-stereo images, followed by the experts (Mdn= 490.50) and the geographers (Mdn= 584.00) but without significance. Using stereo images, the non-geographers (Mdn= 479.00) were fastest, followed by the experts (Mdn= 519.00) and the geographers (Mdn= 544.00) but without significance.

For both conditions (north-oriented and 180°-turned images), there was no significant effect between the visualization type and the level of expertise concerning the response time.

4.3.2 Visualization Type and Task Type

Accuracy

A Friedman test showed a significant effect on accuracy between interactions of visualization type and task type, $\chi^2(3)= 41.876$, $p= .000$. Wilcoxon tests were used to follow-up on these findings. The tests revealed a significant difference between the accuracy of stereo Task Type 1 ($Mdn= 23.00$) and stereo Task Type 2 ($Mdn= 37.00$), $z= -4.194p = .000$ between the accuracy of non-stereo Task Type 1 ($Mdn= 13.00$) and stereo Task Type 1 ($Mdn= 23.00$), $z= -3.216$, $p = .001$, as well as between non-stereo Task Type 2 ($Mdn= 20.00$) and stereo Task Type 2 ($Mdn= 37.00$), $z= -4.021$, $p = .000$. There was no significant effect between the accuracy of non-stereo Task Type 1 ($Mdn= 13.00$) and non-stereo Task Type 2 ($Mdn= 20.00$), $z= -1.910$, $p = .056$. (see Figure 4.11).

Looking at the 180°-turned images, a Friedman test showed a significant effect on accuracy between interactions of visualization type and task type, $\chi^2(3)= 16.643$, $p= .001$. Wilcoxon tests were used to follow up this findings. The tests showed, that there is a significant difference between the accuracy of non-stereo Task Type 1 ($Mdn= 90.00$) and non-stereo Task Type 2 ($Mdn= 93.00$), $z= -2.689$, $p= .007$. There was also a significant effect on the accuracy of stereo Task Type 1 ($Mdn= 93.00$) and stereo Task Type 2 ($Mdn= 93.00$), $z= -2.358$, $p= .018$, as well as between the accuracy of non-stereo Task Type 1 ($Mdn= 90.00$) and stereo Task Type 1 ($Mdn= 93.00$), $z= -2.384$, $p = .017$. There was no significant effect between the accuracy of non-stereo Task Type 2 ($Mdn= 93.00$) and stereo Task Type 2 ($Mdn= 93.00$), $z= -1.811$, $p= .070$.

In the north-oriented images, a significant effect between stereo tt1 and stereo tt2, between non-stereo tt1 and stereo tt1, as well as between non-stereo tt2 and stereo tt2 was found. In the 180° rotated images, significant effects occurred between non-stereo tt1 and non-stereo tt2, between stereo tt1 and stereo tt2 and between non-stereo tt1 and stereo tt1.

Confidence

A Friedman test showed no significant effect on confidence rate between interactions of visualization type and task type, $\chi^2(3)= 1.962$, $p= .580$. However, participants were most confident using non-stereo Task Type 1 images ($M= 95.24$, $SE= .805$), followed by non-stereo Task Type 2 images ($M= 94.27$, $SE= 1.402$). Using stereo Task Type 1 images ($M= 93.48$, $SE= 1.190$) and stereo Task Type 2 images ($M= 92.27$, $SE= 1.641$) they felt least confident but without significance.

Looking at the 180°-turned images, a Friedman test showed a significant effect on the confidence rate between interactions of visualization type and task type, $\chi^2(3)$= 49.567, p= .000. Wilcoxon tests were used to follow-up on these findings. The tests showed a significant difference between non-stereo Task Type 1 (Mdn= 97.00) and non-stereo Task Type 2 (Mdn= 100.00), z= -4.000, p= .000, as well as between non-stereo Task Type 2 (Mdn= 100.00) and stereo Task Type 2 (Mdn= 93.00), z= -4.446, p= .000. There were no significant effects between the confidence rate of stereo Task Type 1 (Mdn= 97.00) and stereo Task Type 2 (Mdn= 93.00), z= -1.760, p= .078, nor between the confidence rate of non-stereo Task Type 1 (Mdn= 97.00) and stereo Task Type 1 (Mdn= 97.00), z= -1.306, p = .191.

In the north-oriented images no significant effects on the confidence rate between visualization type and task type were found. In the 180°-rotated images significant effects were found between non-stereo Task Type 1 and non-stereo Task Type 2, and between non-stereo Task Type 2 and stereo Task Type 2.

Response Time

A Friedman test showed a significant effect on accuracy between interactions of visualization type and task type, $\chi^2(3)$= 18.283, p= .000. Wilcoxon tests were used to follow-up on these findings. The tests showed a significant difference between the response time of non-stereo Task Type 1 (Mdn= 241.00) and non-stereo Task Type 2 (Mdn= 276.00), z= -4.360, p= .000 as well as between stereo Task Type 1 (Mdn= 244.00) and stereo Task Type 2 (Mdn= 278.00), z= -3.538, p = .000. There were no significant effects between the response time of non-stereo task type 1 (Mdn= 241.00) and stereo task type 1 (Mdn= 244.00), z= -.018, p= .986, nor between the response time of non-stereo Task Type 2 (Mdn= 276.00) and stereo Task Type 2 (Mdn= 278.00), z= -.652, p= .514 (see Figure 4.11).

Looking at the 180°-turned images, a Friedman test showed a significant effect on the response time between interactions of visualization type and task type, $\chi^2(3)$= 28.127, p= .000. Wilcoxon tests were used to follow-up on these findings. The tests showed a significant difference between the response time of non-stereo Task Type 1 images (Mdn= 210.00) as well as non-stereo Task Type 2 images (Mdn= 263.00), z= -4.540, p = .000, and between the response time of stereo task type 1 (Mdn= 211.00) and stereo Task Type 2 (Mdn= 279.00), z= -4.942, p= .000. There were no significant effects between the response time of non-stereo task type 1 (Mdn= 210.00) and stereo Task Type 1 (Mdn= 211.00), z= -.599, p = .549, nor between non-stereo Task Type 2 (Mdn= 263.00) and stereo Task Type 2 (Mdn= 279.00), z= -.420, p= .675.

For both conditions (north-oriented and 180°-turned images) significant effects between non-stereo Task Type 1 and non-stereo Task Type 2, and between stereo Task Type 1 and stereo Task Type 2 were found.

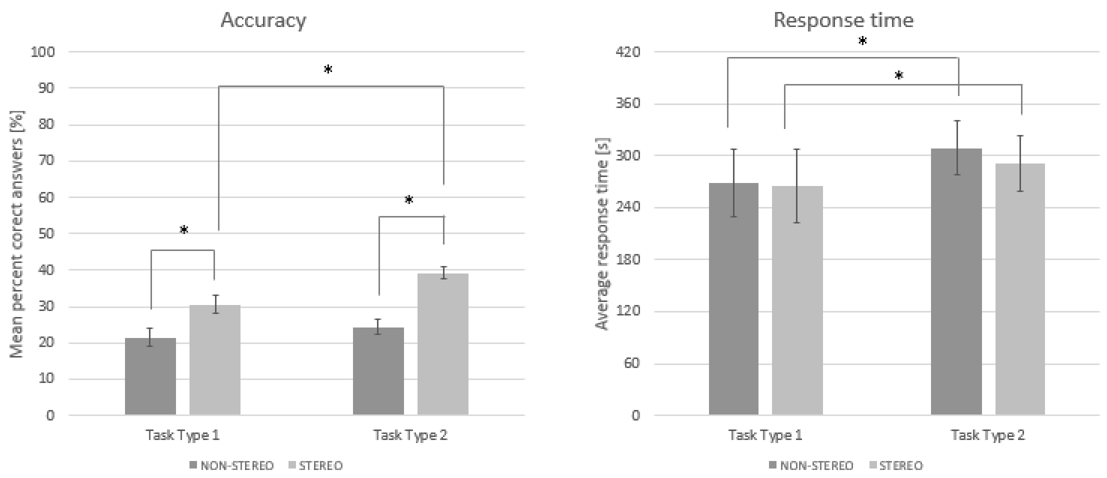


Figure 4.11: Mean accuracy and response time of the interaction between visualization type and task type.

4.3.3 Visualization Type and Terrain Type

Accuracy

A Friedman test showed a significant effect on accuracy between interactions of visualization type and terrain type, $\chi^2(5)$= 71.335, p= .000. Wilcoxon tests were used to follow-up these findings. The tests showed a significant difference between the accuracy of non-stereo subtle change (Mdn= 20.00) and non-stereo middle rugged images (Mdn= 10.00), z= -2.826, p= .005, between non-stereo middle rugged (Mdn= 10.00) and non-stereo highly rugged images (Mdn= 20.00), z= -3.026, p= .002, between stereo subtle change (Mdn= 15.00) and stereo highly rugged images (Mdn= 60.00), z= -4.867, p= .000, between stereo middle rugged (Mdn= 20.00) and stereo highly rugged images (Mdn= 60.00), z= -4.856, p= .000, between non-stereo middle rugged (Mdn= 10.00) and stereo middle rugged images (Mdn= 20.00), z= -3.376, p= .001, and between non-stereo highly rugged (Mdn= 20.00) and stereo highly rugged (Mdn= 60.00), z= -4.798, p= .000. There were no significant effects between the accuracy of non-stereo subtle change (Mdn= 20.00) and non-stereo highly rugged (Mdn= 20.00), z= -1.071, p= .284, between stereo subtle change (Mdn= 15.00) and stereo middle rugged (Mdn= 20.00), z= -.912, p= .362, and between non-stereo subtle change (Mdn= 20.00) and stereo subtle change (Mdn= 15.00), z= -.255, p= .799 (see Figure 4.12).

Looking at the 180°-turned images, a Friedman test showed a significant effect on accuracy between interactions of visualization type and terrain type, $\chi^2(5)$= 30.182, p= .000. Wilcoxon tests were used to follow-up these findings. The tests showed a significant difference between the accuracy of non-stereo middle rugged (Mdn= 90.00) and non-stereo highly rugged (Mdn= 90.00), z = -2.654, p = .008, between stereo subtle change (Mdn= 95.00) and stereo highly rugged (Mdn= 95.00), z= -2.469, p = .014,

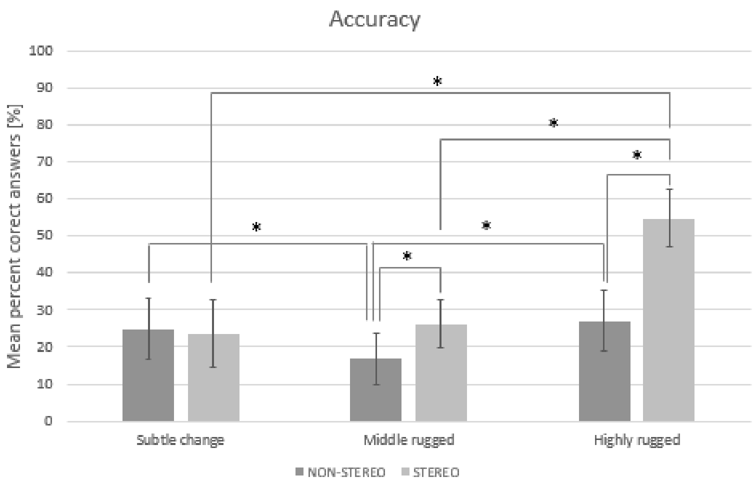


Figure 4.12: Mean response time of the interaction between visualization type and terrain type.

between stereo middle rugged (*Mdn*= 90.00) and stereo highly rugged (*Mdn*= 95.00), z= -4.147, p = .00, and between non-stereo highly rugged (*Mdn*= 90.00) and stereo highly rugged (*Mdn*= 95.00), z= -2.836, p = .005. There were no significant effects between the accuracy of non-stereo subtle change (*Mdn*= 95.00) and non-stereo middle rugged (*Mdn*= 90.00), z= -.747, p = .455, between non-stereo subtle change (*Mdn*= 95.00) and non-stereo highly rugged (*Mdn*= 90.00), z= -1.056, p = .291, between stereo subtle change (*Mdn*= 95.00) and stereo middle rugged (*Mdn*= 90.00), z= -1.819, p = .069, between non-stereo subtle change (*Mdn*= 95.00) and stereo subtle change (*Mdn*= 95.00), z= -1.325, p = .185 and between non-stereo middle rugged (*Mdn*= 90.00) and stereo middle rugged (*Mdn*= 90.00), z= -1.801, p = .072.

For both conditions (north-oriented and 180°-turned images), there were significant effects between visualization type and terrain type were found.

Confidence

A Friedman test showed a significant effect on confidence rate between interactions of visualization type and terrain type, $\chi^2(5)$= 17.385, p= .004. Wilcoxon tests were used to follow-up these findings. The tests showed a significant difference between the confidence rate of non-stereo subtle change (*Mdn*= 93.00) and non-stereo middle rugged (*Mdn*= 100.00), z= -2.701, p = .007, and between non-stereo middle rugged (*Mdn*= 100.00) and stereo middle rugged (*Mdn*= 93.00), z= -2.342, p = .019. There were no significant effects between the confident rate of non-stereo subtle change (*Mdn*= 93.00) and non-stereo highly rugged (*Mdn*= 100.00), z= -1.268, p = .205, between non-stereo middle rugged (*Mdn*= 100.00) and non-stereo highly rugged (*Mdn*= 100.00), z= -1.647, p = .100, between stereo subtle change (*Mdn*= 93.00) and stereo middle rugged

(*Mdn*= 93.00), z= -1.564, p = .118, between stereo subtle change (*Mdn*= 93.00) and stereo highly rugged (*Mdn*= 100.00), z= -1.084, p = .278, between stereo middle rugged (*Mdn*= 93.00) and stereo highly rugged (*Mdn*= 100.00), z= -.429, p = .668, between non-stereo subtle change (*Mdn*= 93.00) and stereo subtle change (*Mdn*= 93.00), z= -.053, p = .958, and between non-stereo highly rugged (*Mdn*= 100.00) and stereo highly rugged (*Mdn*= 100.00), z= -.806, p = .420.

Looking at the 180°-turned images, a Friedman test showed a significant effect on confidence rate between interactions of visualization type and terrain type, $\chi^2(5)$= 62.658, p= .000. Wilcoxon tests were used to follow-up on these findings. The tests showed significant differences between the confidence rate of non-stereo subtle change (*Mdn*= 87.00) and non-stereo middle rugged (*Mdn*= 93.00), z= -4.579, p = .000, between non-stereo subtle change (*Mdn*= 87.00) and non-stereo highly rugged (*Mdn*= 100.00), z= -4.015, p = .000, between stereo subtle change(*Mdn*= 93.00) and stereo highly rugged (*Mdn*= 100.00), z= -2.953, p = .003, between stereo middle rugged (*Mdn*= 100.00) and stereo highly rugged (*Mdn*= 100.00), z= -2.328, p = .020, and between non-stereo subtle change (*Mdn*= 87.00) and stereo subtle change(*Mdn*= 93.00), z= -3.890, p = .000. There were no significant effects between the confident rate of non-stereo middle rugged (*Mdn*= 93.00) and non-stereo highly rugged (*Mdn*= 100.00), z= -.023, p = .982, between stereo subtle change(*Mdn*= 93.00) and stereo middle rugged (*Mdn*= 100.00), z= -1.797, p = .072, between non-stereo middle rugged (*Mdn*= 93.00) and stereo middle rugged (*Mdn*= 100.00), z= -.410, p = .682, and between non-stereo highly rugged (*Mdn*= 100.00) and stereo highly rugged (*Mdn*= 100.00), z= -1.723, p = .085.

For both conditions (north-oriented and 180°-turned images), there were significant effects between visualization type and terrain type were found.

Response Time

A Friedman test revealed no significant effect on response time between interactions of visualization type and terrain type, $\chi^2(5)$= 4.535, p= .475.

Looking at the 180°-turned images, a Friedman test showed a significant effect on response times between interactions of visualization type and terrain type, $\chi^2(5)$= 19.317, p= .002. Wilcoxon tests were used to follow-up on these findings. The tests showed a significant differences between the non-stereo subtle change (*Mdn*= 147.00) and non-stereo middle rugged images (*Mdn*= 160.00), z= -3.996, p = .000, between non-stereo subtle change (*Mdn*= 147.00) and non-stereo highly rugged images (*Mdn*= 161.00), z= -3.815, p = .000, between non-stereo middle rugged (*Mdn*= 160.00) and non-stereo highly rugged images (*Mdn*= 161.00), z= -1.992, p = .046, between stereo subtle change (*Mdn*= 159.00) and stereo highly rugged images (*Mdn*= 166.00), z= -2.067, p = .039, and between stereo middle rugged (*Mdn*= 166.00) and stereo highly rugged images(*Mdn*= 166.00), z= -3.619, p = .000. There are no significant effects between the response times of stereo subtle change (*Mdn*= 159.00) and stereo middle rugged images (*Mdn*= 166.00), z= -1.711, p = .087, between non-stereo subtle change (*Mdn*= 147.00) and stereo subtle change images (*Mdn*= 159.00), z= -1.367, p = .172, between non-stereo middle rugged (*Mdn*= 160.00) and stereo middle rugged (*Mdn*= 166.00), z= -.679, p = .497, and between non-stereo highly rugged (*Mdn*= 161.00) and stereo highly rugged (*Mdn*= 166.00), z= -1.760, p = .078.

For the north-oriented images there was no significant effect on response times between visualization type and terrain type. For the 180°-turned images there were significant effects between visualization type and terrain type.

4.4 Summary of Results

Anaglyph Glasses and Terrain Reversal Effect

Overall, the evaluation revealed no *preference* for a visualization type. Participants preferred the non-stereo and stereo viewing conditions equally. Their discomfort level with anaglyph viewing was low to medium. Only few rated it as high.

A majority of the participants mentioned that they generally recognized *contradictions between land form and land cover* (e.g. snow, vegetation, river) as well as even the *terrain reversal effect*.

Overall Findings from the Main Effects and Interactions

The 180°-turned images were used as a reference for the north-oriented images. In all conditions (main effects and interactions) their accuracy rate was around or even above 90%.

Comparing the response times from all conditions (main effects and interactions) using the 180°-turned images, participants were always faster to answer these items than while using north-oriented images.

The confidence rate analysis revealed that participants were confident overall, i.e. image orientation as well as main effects and interactions had no influence on the confidence rate of the participants. Participants felt always more than 90% confident in giving their answers.

Main effects - Image Orientation, Visualization Type, Level of Expertise, Task Type & Terrain Type

The *image orientation* analysis revealed that participants had a significantly higher accuracy rate using the 180°-turned images than with the north-oriented images, regardless of visualization type, level of expertise, task type and terrain type (see Figure 4.5). There were no significant effects on the confidence rate and the response times between the two image orientations. However, participants were more confident and faster with the 180°-images (see Figure 4.5).

A significant difference between the two *visualization types* was revealed by the data analysis. Participants were better in solving the given tasks using the stereo images than the non-stereo images (see Figure 4.6). In regards to the confidence rate and the response times, there were no significant effects in.

Analyzing the accuracy between the three different *level of expertise* (expert, geographer, non-geographer) showed the results not be significant. In this specific study, non-geographers were least accurate, followed by the geographers and the experts (see Figure 4.7). Confidence rate as well as response times are also not significantly influenced by level of expertise.

A significant result was shown for the accuracy of the two different *task types*. Participants were more accurate with Task Type 2 than with Task Type 1 (see Figure 4.8). The confidence rate was not significantly affected by the task type, whereas the response times were. Participants were faster for Task Type 1 than Task Type 2 (see Figure 4.8).

While analyzing the *terrain type*, a significant effect was found between the accuracy of subtle change and highly rugged terrain as well as between middle rugged and highly rugged terrain. Highly rugged terrain lead to more accurate answers than subtle change and middle rugged terrain (see Figure 4.9). There was also a significant effect between the confidence rate of subtle change and middle rugged terrain. Middle rugged terrain lead to a higher confidence rate than subtle change terrain. In regards to the response times, there was no significant effect.

Interactions - Interactions of Visualization Type with Level of Expertise, Task Type and Terrain Type

The interaction between visualization type and level of expertise was not significant. Accuracy, confidence rate and response times of non-stereo as well as stereo images were not significantly affected by the level of expertise. However, in this specific study, each group was better at solving the given task using stereo images than using non-stereo images, whereas the expert were most accurate, followed by the geographers and the non- geographers (see Figure 4.10).

The interaction between visualization type and task type was significant regarding accuracy and response times. Within Task Type 1, participants were more accurate using stereo than using non-stereo images. The same results were observed for Task Type 2. Considering the stereo images, the accuracy was higher with Task Type 2 (see Figure 4.8). Looking at the response times, within the non-stereo images participants were faster for Task Type 1. The same was observed for the stereo images (see Figure 4.11). There was no significant effect on the confidence rate.

The interaction between visualization type and terrain type was significant regarding accuracy and confidence rate. The analysis revealed that in middle and highly rugged terrain, stereo increased the accuracy, whereas in subtle change landscapes the accuracy was nearly the same between the two visualization types. The results also showed that in highly rugged terrain stereo had the strongest effect. The accuracy rate was nearly twice as high as in non-stereo viewing conditions. In middle rugged terrain the accuracy in the stereo viewing condition was able to be increased by nearly half (see Figure 4.12). In all three terrain types participants felt more confident in the non-stereo viewing conditions. The highest confidence rate was reported for the middle rugged terrain. There was no significant effect on the response times.

5 Discussion

In this chapter the main findings from chapter 4 are summarized and discussed. The chapter is divided into sections according to the research questions from chapter 1. Each section is divided into the three dependent variables accuracy, confidence and response time. Ancillary to the research questions overall findings are also discussed. At the end of the chapter limitations of the study are examined.

5.1 Research Question 1: Visualization Type

To understand the influence of stereopsis on the terrain reversal effect, a comparative study with the two visualization types 'non-stereo' and 'stereo' was executed. Accuracy, confidence rate and response time were analysed.

Accuracy

The results revealed, that the visualization type does have an influence on the terrain reversal effect. Participants were more accurate in solving the given tasks using stereo than using non-stereo satellite imagery. Therefore it can be said that stereopsis has an influence on the terrain reversal effect. Two possible explanations for these findings are discussed.

In non-stereo satellite imagery the main way people perceive depth is through lighting and shading (e.g. Saraf et al. 1996; Bernabé-Poveda & Çöltekin 2014). This means, that individuals can perceive three-dimensional shapes through shading. Stereo images provide an additional depth cue. The principle behind this is stereopsis (Hubona et al. 1999; Mehrabi et al. 2013). Using anaglyph glasses participants are able to perceive this additional depth cue (Geng 2013). Once stereo images are used, people are no longer dependent solely on the shape-from-shading depth cue. To interpret volumetric shapes, multiple cues (e.g. shape-from-shading and stereopsis) are available. These might help perceive the real depth in the image and prevent people from falling for the illusion of the shadows. Another possible explanation is that through the anaglyph glasses the illusion is augmented, i.e. three-dimensional shapes become even more pronounced. In this case stereopsis would not lead to a correction of the shadow illusion. Instead the illusion would get so strong that individuals consciously or unconsciously feel something is wrong and end up giving the right answer. Livatino & Privitera (2006) also mentioned that stereoscopic visualizations can provide a user with a higher sense of presence in displayed environments. A higher given depth perception is given. A higher depth perception can lead, to higher comprehension of distance and object presence. This could also be the case for stereo satellite imagery. Because of the perceived depth trough the anaglyph glasses, participants feel higher sense of presence in the images. This is also a reason, why some participants liked the stereoscopic view more than the non-stereoscopic view. They mentioned that because the three-dimensional effect was stronger, the terrain shape was better recognizable.

Participants were also significantly more accurate using 180° turned stereo images than using 180° non-stereo images. As Sexton & Surman (1999) mentioned, stereo images help to distinguish depth and certain detailed aspects in what might otherwise be interpreted as a flat 2D image. Therefore using anaglyph glasses, the depth perception in the displayed environments are strengthened and lead to more accurate answers compared to non-stereo images.

Confidence

In a study from Carvalho (2011) participants felt more confident using stereoscopic three-dimensional than two-dimensional maps. In this study that effect could not be observed. Participants are equally confident in non-stereo and stereo viewing conditions. A possible explanation might be that the confidence rate was measured indirectly through an integrated Likert Scale in the questions. Participants were not directly asked if they felt confident while giving their answers. The analysis method might also have influenced the results. All answers, expect the ambiguous ones, were coded as confident.

Response time

Seipel (2013) mentioned that stereo cues tend to result in increased response time when solving tasks. This might be due to the additional depth cue in the stereo images, which leads to studying the images longer. Contrary to the study from Seipel (2013), no significant difference between the response time of non-stereo and stereo images were observed in this study. Participants were equally fast using the two visualization types. A reason might be that participants were told to solve the tasks as quickly as possible. There was no given time limit (e.g. twenty seconds per image) but it still might have influenced the response time. This was also reported by Wilkening & S. Fabrikant (2011)(cited in Bruegger 2015). They mentioned that time pressure leads to differences in decision making.

5.2 Research Question 2: Level of expertise

For this study, three different levels of expertise were defined: expert, geographer and non-geographer. The aim was to find out if level of expertise makes a difference in experiencing the terrain reversal effect. Accuracy, confidence rate and response time were considered.

Accuracy

While analyzing the accuracy between the three different levels of expertise, the results were not significant. A reason might be that the number of participants within the three groups was small, namely between ten and twelve individuals. Nevertheless, all different levels of expertise experienced the terrain reversal effect. This was also observed by Bernabé-Poveda & Çöltekin (2014). Most accurate were the experts, followed by the geographers and the non-geographers. At a preliminary stage, Bernabé-Poveda & Çöltekin (2014) were able to identify a weak positive effect due to expertise when investigating the terrain reversal effect. The same findings were found by Biland (2014).

One possible explanation is that experts created their own repertoire of schemata which enabled them to encode appropriate new information rapidly and efficiently (Gilhooly et al. 1988). In this study, experts were defined as coming from the field of remote sensing. This was decided because the terrain reversal effect was investigated using satellite imagery. So it was necessary for experts to have experience using satellite images and most important working with them frequently. Consciously or unconsciously they should be more aware of the illusion, because they have learned how to interpret satellite images. Biland (2014) suggested that experts have learnt to interpret the land cover information, so that they are not solely dependent on the overhead illumination bias. Experts also seem to be able to learn new information related to their field of expertise better than non-experts (Gilhooly et al. 1988). For this study that would mean that because the experts already have experience working with and interpreting satellite images displaying the terrain reversal effect, they are more aware of it. Carvalho (2011) also showed in his study that the skilled participants achieved better results overall then the non-skilled participants. Other researchers like Smith (1962) (cited in Toutin 1997) have indicated that performance for search tasks can be improved if the subject knows beforehand what they will look at. Experts from the remote sensing field have learned to use many cues in an image to interpret it correctly because they know what to look for(Toutin 1997). These cues can then give them hints and prevent them from falling for the terrain reversal effect.

Since the level of expertise was not significant the interaction between visualization type and level of expertise was not significant either. Still, experts were more accurate than geographers and non-geographers. Within the groups, participants were more accurate using stereo images than using non-stereo images.

Confidence

The results showed, that the confidence rate was not significantly influenced by level of expertise. Interestingly experts were not more confident than geographers or non-geographers. A possible explanation is, that the confidence rate in the experiment is related to the self-confidence of each of the participants. Some of the participants might generally be more confident than others, independent on their level of expertise.

Response time

Hoffman (1990) (cited in Toutin 1997) suggested that remote sensing experts might attend more time for studying satellite images. In this study, response time is not significantly influenced by level of expertise. A reason might be, because participants were told to solve the task as fast as possible. To investigate if experts are slower than non-experts, as suggested by Hoffman (1990) (cited in Toutin 1997), there should be no restriction of time.

5.3 Research Question 3: Task Type

Two different task types had been designed, a "perceptual" and a "more interpretative" task. The aim was to find out, if the different task types make a difference in experiencing the terrain reversal effect in terms of accuracy of land form identification, response time and confidence in task success.

Accuracy

A significant result could be shown between the accuracy of the two different task types. Participants are more accurate with Task Type 2, than with Task Type 1. A reason might be, that for the more interpretative questions from Task Type 2, participants are forced to observe and study a broader area in the image. Thereby they might get aware of discrepancy when looking at the environment. When the image is turned by 180° a significant effect can be seen too. Participants are more accurate with Task Type 2, than with Task Type 1. This leads to the assumption, that independent on the image orientation, participants are more accurate with Task Type 2. So when participants get forced to look on a broader area and interpret the satellite image, they seem to make less mistakes, irrespective of occurrence of illusion.

Also when looking at the interaction between the visualization type and task type, significant results can be seen. Within the two task types, stereo lead to more accurate answers than non-stereo. A reason for this result might be the additional depth cue provided by the anaglyph glasses (see section 5.1). There is also a significant effect between stereo Task Type 1 and stereo Task Type 2. Participants are more accurate using stereo Task Type 2 than using stereo Task Type 1 images. This supports the result from the main effect of the task type, where participants are more accurate with Task Type 1 than Task Type 2. Because each image had been shown in non-stereo and stereo viewing conditions, it can be said, that with Task Type 2 participants are more accurate.

Confidence

The confidence rate is not significantly affected by the task type. Interestingly when turning the image by 180° a significant result between Task Type 1 and Task Type 2 can be seen. Participants are more confident with Task Type 1 than with Task Type 2. Reasons for this could be that Task Type 1 questions were not as complex as the Task Type 2 questions, i.e. questions were easier to understand what lead to a higher confidence rate.

Response time

The response time between the two task types is significantly affected. Participants are faster with Task Type 1 than Task Type 2. A possible reason might be, that for answering Task Type 2 a broader area had to be considered and participants had to interpret the environment on a higher level to answer the questions. This can lead to an increasing response time.

5.4 Research Question 4: Terrain Type

Three different terrain types had been designed, subtle change, middle rugged and highly rugged. The aim was to find out, if the three terrain types make a difference in experiencing the terrain reversal effect in terms of accuracy of land form identification, response time and confidence in task success.

Accuracy

In literature it is assumed that the effect is more severe in rugged terrain, especially in high altitude areas of hilly terrain (Saraf et al. 1996; Wu et al. 2013). In this study even in subtle change environments the terrain reversal effect can be observed. Interestingly the subtle change landscape lead to the lowest accuracy rate, followed by the middle rugged and highly rugged terrain. A reason might be, that cues (like rivers) in the images helped more when a highly rugged landscape was shown.

The interaction between visualization type and terrain type revealed that in middle and highly rugged terrain, stereo increased the accuracy, whereas in subtle change landscapes the accuracy was nearly the same between the two visualization types. This findings are very interesting because they confirm the assumption that shape-from-shading together using stereopsis increases the depth perception in the image (see section 5.1). In highly rugged environments it is believed that the topographic relief leads to a change in the solar illumination to the slope and to the viewing geometry of the terrain (Saraf et al. 1996). The shape-from shading depth cue is therefore stronger in highly rugged terrain. But at this state of the study it is still not clear, if the illusion or the real shape of the environment gets stronger (see section 5.1). The results also show, that in highly rugged terrain stereo has the strongest effect. The accuracy rate is nearly twice as high than in non-stereo viewing conditions. It can be supposed that the highly rugged terrain had the most influence why participants were more accurate using stereo images than using non-stereo images (see section 5.1). This leads to the assumption that stereo images are most useful when the illusion occurs in highly rugged terrain. But also in middle rugged terrain the accuracy in the stereo viewing condition could be increased by nearly the half compared to the non-stereo images

Confidence

A significant effect between the confidence rate of subtle change and middle rugged terrain occurred. Participants were more confident with middle rugged than with subtle change terrain. A possible explanation might be, that in the subtle change images the objects were less pronounced because the shape-from shading effect is not as strong as in middle rugged terrain. Therefore participants were less confident when giving their answers.

The interaction between visualization type and terrain type revealed that in all three terrain types, participants were more confident in the non-stereo viewing conditions. This result is interesting because they were more confident using non-stereo images but at the same time were less accurate using non-stereo images. This might be because normally satellite images are viewed in non-stereo. Therefore participants were more confident using non-stereo images. Familiarity using non-stereo images was also a reason for preferring stereo-images.

Response time

Considering the response time, there is no significant effect. One reason could be, because participants were told to be as fast as possible. Without time restriction a significant effect between the different terrain types might have occurred.

Turning the image by 180°, a significant effect between subtle change and middle rugged can be seen. Participants were faster with subtle change images. A possible explanation might be that in middle rugged terrain the depth perception is stronger and therefore participants observed the image longer. But still this significant result was influenced by the instruction that people should answer as fast as possible.

5.5 Other findings

The results from the image orientation showed, that 180° rotation leads to correcting the effect regardless of visualization type, level of expertise, task type and terrain type. In over 90% of the 180° turned image the illusion disappeared. This finding is also mentioned in different literature as a correction method (e.g. Saraf et al. 1996, Bernabé-Poveda et al. 2005, Bernabé-Poveda & Çöltekin 2014, Biland 2014, etc.). In this study the

180° turned images were used as a reference to the north-oriented images. Additionally, it was important that participants had also images without illusion to avoid a learning effect.

Comparing the response time from all conditions (main effects and interactions), with the 180° turned images, participants were always faster in answering the question with the 180° turned images (without illusion) than with the north-oriented images (with illusion). A possible explanation is that in images with illusion, participants tended to study the image more, because they consciously or unconsciously were influenced by the effect.

The confidence rate analysis revealed, that participants had an overall high confidence rate, i.e. image orientation as well as main effects and interactions had no influence on the confidence rate of the participants. A possible explanation might be, that the confidence rate was measured indirectly through a integrated Likert Scale in the questions. Participants were not explicit asked, if they felt confident when giving their answers. The analysis method might also have influenced the result. All answers, expect the ambiguous ones, had been considered as confident.

Forsberg et al. (2009) observed that that participants liked the stereoscopic viewing more than the non-stereoscopic viewing. In this study there was no preference for a visualization type. Nearly half of the participants liked non-stereo and nearly half liked stereo, only a few had no preferences. Also within the three different level of expertise groups the preferences are similar. In literature different studies show, that stereo images lead to visual discomfort (Seipel 2013). This was also a reason, why some participants preferred non-stereo images.

Interestingly most participants recognized generally a contradiction between land form and land cover. And after briefly explaining in text form and figurative the terrain reversal effect at the end of the study, most of the participants told, that they have recognized the illusion. They might have realized the effect, but as the results showed, they have not been consistently aware of it. Possibly some single images looked obviously wrong (e.g. rivers on a ridge does not make sense).

5.6 Limitations

In this study the influence of stereopsis on the terrain reversal effect in satellite imagery had been investigated. The results showed, that participants are more accurate using stereo than non-stereo images. Task type as well as terrain type lead also to significant results concerning to the terrain reversal effect. However, there are limitations in this study. These limitations and possible influence on the results are discussed in the following.

Amount of participants

In this study a total of thirty-three participants took part. Within the groups of level of expertise the number of participants lied between ten and twelve. A higher amount of individuals would have increased the quality of the results and maybe lead to more statistically significant results. Especially for the groups from level of expertise, where no significant results occurred. Additionally the influence of outliers could have been weakened.

Randomization

The randomization command on Survey Monkey did not work properly. Thus every participant had the same order of questions in the main part of the experiment. This circumstance could have affected the results (Martin 2008). One possible effect could have been that each participant got tired towards the end. In this case the last questions might have been influenced through tiredness. Taking a closer look at the data, no pattern in the accuracy rate could be seen in this study.

Stimuli

For the main experiment the stimuli were subjectively chosen. Because the terrain reversal effect is a perceptual phenomenon, each individual experience the effect different. Therefore it is possible, that another person would not have seen an effect in the selected images. Trying to minimize this risk, the images had been shown to other persons. When selecting the stimuli, it was also tried to reach an even amount of land cover cues. This was also subjectively decided and other images might have lead to different results.

Necker cube effect

The necker cube effect might also have been an issue (Kornmeier & Bach 2005). It had also been mentioned by Biland (2014) and named as terrain flipping. When participants looked at a three-dimensional object in the satellite image the effect might have occurred. This means, that in one moment they had seen for example a valley and in the other moment it had switched into a ridge. This phenomenon could have had an influence on the results.

Global convexity bias

Another issue which could have influenced the results is the so called "global convexity bias". It implies that it is more likely to perceive surfaces as convex rather than concave (e.g. Hill & Bruce 1994, Langer & Bülthoff 2001, Liu & Todd 2004). Possible reasons can be seen in natural environment. More objects are globally convex rather than concave (Johnston et al. 1992 and Hill & Bruce 1993) (see section 2.1). Consequently participants may have perceived more landforms as ridges than as valleys.

Anaglyph glasses

In this study anaglyph glasses were used as a method for stereoscopic viewing. During the colour reproduction most of the colour information can get lost (Kooi & Toet 2004, Lambooij et al. 2009, Mehrabi et al. 2013) this might have influenced the illusion. Anaglyph glasses can also lead to visual discomfort (e.g. headache and sickness) (Kooi & Toet 2004, Lambooij et al. 2009, Mehrabi et al. 2013 what also could have influenced the results. But in this study only few had a high discomfort level with anaglyph glasses.

Task Type

In the main experiment images with letters had been used. This letters might have had an influence on the participants as well as on the illusion. The colour, the size and the spacing between the letter could have had affected the terrain reversal effect. There were also more tasks where a higher point had to be detected. This could have influenced the mistakes in the lower point tasks because participants might have been tended to look more on higher points.

Terrain Type

The classifications of the terrain types (subtle change, middle rugged and highly rugged) had been subjectively chosen (lower than 50 m, between 50 and 500 m and over 500 m). Another gradation might have lead to other findings.

Response time

Participants were advised to answer the questions as fast as possible. If they would have had no time limitation at all, the results might have been different. Ancillary to the response time, accuracy rate as well as confidence rate might have been increased. Participants might have been more accurate and more confident without this instruction.

Confidence rate

The confidence rate had been measured indirectly through a Likert Scale in the question. If participants would have been directly asked how confident they are when giving their answer, it might have change the results. Also the analysis method might have influenced the results. All answers, expect for the ambiguous ones, had been considered as confident.

6 Conclusion and Future Research

This study investigated the influence of visualization type (non-stereo and stereo), level of expertise (experts vs. non-experts), task type (perceptual and more interpretative) and terrain type (subtle change, middle rugged and highly rugged) on the terrain reversal effect. This chapter gives a conclusion to all of these four parts. At the end some suggestions for future research are mentioned.

6.1 Visualization Type

The results of this study indicate that stereopsis can help to weaken the terrain reversal effect. Participants were more accurate using stereo than using non-stereo images. A closer look on the interaction of visualization type and task type revealed, that in highly rugged terrain the accuracy rate using stereo images could be increased the most compared to the non-stereo images. But also in middle rugged terrain the accuracy rate could be increased. In subtle change landscapes the stereo viewing condition did not lead to a higher accuracy rate. So it can be concluded, that especially in highly rugged terrain the use of anaglyph glasses is valuable when in the non-stereo image the illusion occurs. But still, also using stereo images the illusion did not disappear in all images, but the accuracy rate increased. So this findings can be used as an approach for further research. Furthermore when using anaglyph glasses the effect of stereo blindness has to be taken into account. Individuals which have this blindness are not able to see stereo images. As Ware (2000) (cited in Çöltekin 2006) mentioned nearly 20% of the population may have it. So even if anaglyph glasses and other methods of stereoscopic viewing weaken the illusion, it is not a holistic solution. The confidence rate as well as the response time were not significant.

6.2 Level of Expertise

The results from the level of expertise analysis was not significant. This might be, because within the groups the number of participants was quite low. If the amount of participants would have been higher, results could have been significant. However in this study all different levels of expertise experienced the terrain reversal effect. This could also be observed by Bernabé-Poveda & Çöltekin (2014). It can be concluded that the effect is independently on the level of expertise, so all individuals independent on their profession experience the illusion. But it could be seen that experts were consciously or unconsciously better than geographers and non-geographers when confronted with the terrain reversal effect. This could also be observed by Bernabé-Poveda & Çöltekin (2014) and Biland (2014). This leads to the conclusion that the terrain reversal effect can be weaken through experience and expertise knowledge in the field of remote sensing. So if people are aware that the effect can occur in satellite images and if they know how to interpret satellite images, the influence of the illusion decreases. Biland (2014) also suggested that experts learned to interpret the land cover information, so that they not only dependent on the overhead illumination bias. The confidence rate as well as the response time were not significant.

6.3 Task Type

The results from the task type analysis revealed that participants are more accurate with questions from Task Type 2 than Task Type 1. So it can be concluded that when people need to study a broader area and if the questions are more complex the influence of the illusion can be reduced. Participants were also slower with questions from Task Type 2 what indicates that they were studying the environment longer to answer the questions. Task type had no significant influence on the confidence rate. When looking at the interaction between visualization type and task type participants were better with the stereo viewing condition. This would confirm that stereo has an alleviative influence on the illusion.

6.4 Terrain Type

So far the terrain reversal effect had only been described as a phenomenon which occurs mostly in highly rugged terrain (e.g. Saraf et al. 1996; Wu et al. 2013). This study could show, that the effect is also present in less rugged terrain. It could be shown that the accuracy rate in highly rugged images was even higher than for subtle change image. This could be because there were more obvious cues in the highly rugged images. But it can be concluded that people should also be aware of the effect in subtle change and middle rugged terrain. Participants were more confident with middle rugged than subtle change terrain. Considering the response time there was no significant effect.

6.5 Future Research

In this study anaglyph glasses had been used as a stereoscopic display. In future research it would be interesting to see how other stereoscopic techniques like polarized glasses (like used in 3D films in the cinema) would influence the terrain reversal effect. An advantage would be that the colour information do not get falsified. These method still needs glasses. To avoid spectacles the effect could be examined with auto stereoscopic 3D displays (e.g lenticular lenses).

For additional investigation it would be interesting to discuss the given answers at the end of the experiment with each participant. Thereby the leader of the experiment can better comprehend whether the participants recognized the real shape or if they just guessed. Another possible method in this direction can be a think aloud approach. During the experiment participants have to tell what they are doing. This can help for example to understand the process behind the different task types. Furthermore it can give a better understanding about how the solving strategy from the experts differs from the non-experts.

Bibliography

ADA (2007). *Dental admission testing program - Sample Test Items.* American Dental Association.

Albertz, J. (2009). *Einfuehrung in die Fernerkundung - Grundlagen der Interpretation von Luft- und Satellitenbildern.* 4. Auflage. WBG (Wissenschaftliche Buchgesellschaft), Darmstadt,

Anderson, B. L. & Nakayama, K. (1994). "Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion." In: *Psychological review* 101.3, p. 414.

Benoit, A., Le Callet, P., Campisi, P. & Cousseau, R. (2008). "Quality assessment of stereoscopic images." In: *EURASIP journal on image and video processing* 2008, Article–ID.

Bernabé-Poveda, M. A. & Çöltekin, A. (2014). "Prevalence of the terrain reversal effect in satellite imagery." In: *International Journal of Digital Earth* ahead-of-print, pp. 1–16.

Bernabé-Poveda, M. A., Manso, M.A. & Ballari, D (2005). "Correction of relief inversion in images served by a web map server." In: *Proceedings of the XXII. International Cartographic Conference (ICC). A Coruña. Spain.*

Bernabé-Poveda, M. A., Sánchez-Ortega, I. & Çöltekin, A (2011). "Techniques for Highlighting Relief on Orthoimaginery." In: *Procedia - Social and Behavioral Sciences* 21, pp. 346–352. ISSN: 18770428. DOI: 10.1016/j.sbspro.2011.07.028.

Biland, J. (2014). "Terrain Reversal Effect: Assessing the Impact of Incidence of Light in Shaded Relief Maps and Comparison of Landform Perception in Shaded Relief Maps." Master's thesis.

Bruegger, A. (2015). "Where are the ups and downs? Evaluating Elevation Representations for Bicycle Paths in City Maps." Master's thesis.

Carvalho, L. (2011). *Evaluation of 2D and 3D Map Presentation for Geo-Visualization.*

Çöltekin, A. (2006). "Foveation for 3D visualization and stereo imaging." Doctoral dissertation.

Dial, G., Bowen, H., Gerlach, F., Grodecki, J. & Oleszczuk, R. (2003). "IKONOS satellite, imagery, and products." In: *Remote sensing of Environment* 88.1, pp. 23–36.

Dodgson, N. A. (2004). “Variation and extrema of human interpupillary distance.” In: *Electronic Imaging 2004*. International Society for Optics and Photonics, pp. 36–46.

Ericsson, K. A. & Lehmann, A. C. (1996). “Expert and exceptional performance: Evidence of maximal adaptation to task constraints.” In: *Annual review of psychology* 47.1, pp. 273–305.

Fabrikant, S. I., Maggi, S. & Montello, D. R. (2014). “3D Network Spatialization: Does It Add Depth to 2D Representations of Semantic Proximity?” In: *Geographic Information Science*. Springer, pp. 34–47.

Field, A. (2009). *Discovering statistics using SPSS*. Sage Publications Ltd, 3rd ed.

Fielder, A. R. & Moseley, M. J. (1996). “Does stereopsis matter in humans?” In: *Eye* 10.2, pp. 233–238.

Forsberg, A. S., Chen, J. & Laidlaw, D. H. (2009). “Comparing 3d vector field visualization methods: A user study.” In: *Visualization and Computer Graphics, IEEE Transactions on* 15.6, pp. 1219–1226.

Fricke, T. R. & Siderov, J. (1997). “Stereopsis, stereotests, and their relation to vision screening and clinical practice.” In: *Clinical and experimental optometry* 80.5, pp. 165–172.

Frisby, J. P. & Mayhew, J. E. W. (1979). “Depth inversion in random-dot stereograms.” EN. In: *Perception* 8.4, pp. 397–399. ISSN: 0301-0066. DOI: 10.1068/p080397.

Geng, J. (2013). “Three-dimensional display technologies.” In: *Advances in optics and photonics* 5.4, pp. 456–535.

Gerardin, P., Kourtzi, Z. & Mamassian, P. (2010). “Prior knowledge of illumination for 3D perception in the human brain.” In: *Proceedings of the National Academy of Sciences* 107.37, pp. 16309–16314.

Gil, M. L., Armesto, J. & Cañas, I. (2005). “3-D expression of relief in Landsat Thematic Mapper and SPOT P images.” In: *Optical Engineering* 44.4, p. 7003.

Gil, M. L., Arza, M., Ortiz, J. & Ávila, A. (2014). “DEM shading method for the correction of pseudoscopic effect on multi-platform satellite imagery.” In: *GIScience & Remote Sensing*.

Gil, M. L., Ortiz, J., Rego, T. & Gelpi, L. (2010). “The Correction of the Pseudoscopic Effect on Quickbird Satellite Imagery.” In: *Survey Review* 42.318, pp. 318–326. ISSN: 0039-6265. DOI: 10.1179/003962610X12747001420465.

Gilhooly, K. J., Wood, M., Kinnear, P. R. & Green, C. (1988). “Skill in map reading and memory for maps.” In: *The quarterly journal of experimental psychology* 40.1, pp. 87–107.

Gregory, R. L. (1997). "Knowledge in perception and illusion." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 352.1358, pp. 1121–1127.

Hengl, T. & Rossiter, D. G. (2003). "Supervised Landform Classification to Enhance and Replace Photo-Interpretation in Semi-Detailed Soil Survey." In: *Soil Science Society of America Journal* 67.6, p. 1810. ISSN: 1435-0661. DOI: 10.2136/sssaj2003.1810.

Hill, H. & Bruce, V. (1993). "Independent effects of lighting, orientation, and stereopsis on the hollow-face illusion." In: *PERCEPTION-LONDON-* 22, pp. 887–887.

Hill, H. & Bruce, V. (1994). "A comparison between the hollow-face and 'hollow-potato' illusions." EN. In: *Perception* 23.11, pp. 1335–1337. ISSN: 0301-0066. DOI: 10.1068/p231335.

Hill, H. & Johnston, A. (2007). "The hollow-face illusion: Object-specific knowledge, general assumptions or properties of the stimulus?" In: *Perception* 36.2, pp. 199–223. ISSN: 0301-0066. DOI: 10.1068/p5523.

Hoffman, R. R. (1990). "Remote perceiving: A step toward a unified science of remote sensing." In: *Geocarto International* 5.2, pp. 3–13.

Howard, I. P. (2002). *Seeing in depth, Vol. 1: Basic mechanisms.* Vol. I. ISBN: 0973087307.

Howard, I. P. & Rogers, B. J. (2002). "Seeing in depth, volume 2: Depth perception." In: *Ontario, Canada: I. Porteous.*

Hubona, G. S., Wheeler, P. N., Shirah, G. W. & Brandt, M. (1999). "The relative contributions of stereo, lighting, and background scenes in promoting 3D depth visualization." In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 6.3, pp. 214–242.

IKONOS (2015). *Stereo Satellite Imagery.* URL: http://www.satimagingcorp.com/satellite-sensors/ikonos/ikonos-stereo-satellite-images/.

Imhof, E. (2007). *Cartographic Relief Presentation.* ISBN: 1589480260.

Jenkin, H. L, Jenkin, M. R., Dyde, R. T. & Harris, L. R. (2004). "Shape-from-shading depends on visual, gravitational, and body-orientation cues." In: *Perception* 33.12, pp. 1453–1461.

John, M. St., Cowen, M. B., Smallman, H. S. & Oonk, H. M. (2001). "The use of 2D and 3D displays for shape-understanding versus relative-position tasks." In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 43.1, pp. 79–98.

Johnston, A., Hill, H. & Carman, N. (1992). "Recognising faces: effects of lighting direction, inversion, and brightness reversal." In: *Perception* 21.3, pp. 365–375.

Julesz, B. (1960). "Binocular Depth Perception of Computer-Generated Patterns." In: *Bell System Technical Journal* 39.5, pp. 1125–1162.

Kettunen, P., Sarjakoski, T., Sarjakoski, L. T. & Oksanen, J. (2009). "Cartographic Portrayal of Terrain in Oblique Parallel Projection." In: *24th International Cartographic Conference. The World's Geo-Spatial Solutions.*

Kleffner, D. A. & Ramachandran, V. S. (1992). "On the perception of shape from shading." In: *Perception* 52.1, pp. 18–36. ISSN: 1943-3921. DOI: 10.3758/BF03206757.

Kooi, F. L. & Toet, A. (2004). "Visual comfort of binocular and 3D displays." In: *Displays* 25.2, pp. 99–108.

Kornmeier, J. & Bach, M. (2005). "The Necker cube—an ambiguous figure disambiguated in early visual processing." In: *Vision research* 45.8, pp. 955–960.

Kray, C., Elting, C., Laakso, K. & Coors, V. (2003). "Presenting route instructions on mobile devices." In: *Proceedings of the 8th international conference on Intelligent user interfaces.* ACM, pp. 117–124.

Lambooij, M., Fortuin, M., Heynderickx, I. & IJsselsteijn, W. (2009). "Visual discomfort and visual fatigue of stereoscopic displays: a review." In: *Journal of Imaging Science and Technology* 53.3, pp. 30201–1.

Lambooij, M., IJsselsteijn, W., Bouwhuis, D. G. & Heynderickx, I. (2011). "Evaluation of stereoscopic images: beyond 2D quality." In: *Broadcasting, IEEE Transactions on* 57.2, pp. 432–444.

Lang, J. I. (1983). "Ein neuer Stereotest." In: *Klin Monatsbl Augenheilkd* 182, p. 373.

Lang, J. I. (1988). "Eye screening with the Lang stereotest." In: *Am Orthop J* 38, pp. 48–50.

Langer, M. S. & Bülthoff, H. H. (2001). "A prior for global convexity in local shape-from-shading." In: *Perception* 30.4, pp. 403–410. ISSN: 0301-0066. DOI: 10.1068/p3178.

Liu, B. & Todd, J. T. (2004). "Perceptual biases in the interpretation of 3D shape from shading." In: *Vision research* 44.18, pp. 2135–45. ISSN: 0042-6989. DOI: 10.1016/j.visres.2004.03.024.

Livatino, S. & Privitera, F. (2006). "Stereo visualization and virtual reality displays." In: *proc. of Vis2006: IEEE Visualization 2006 conference.*

Lobben, A. K. (2007). "Navigational map reading: Predicting performance and identifying relative influence of map-related abilities." In: *Annals of the association of American geographers* 97.1, pp. 64–85.

Martin, D. (2008). *Doing psychology experiments.* Cengage Learning, 7th ed.

McAllister, D. F., McKay, S. & othera (1995). "Stereo computer graphics and other true 3D technologies." In: *Computers in Physics* 9.1, pp. 56–56.

McGrew, K. S. & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Allyn & Bacon.

Mehrabi, M., Peek, E. M., Wuensche, B. C. & Lutteroth, C. (2013). "Making 3D work: a classification of visual depth cues, 3D display technologies and their applications." In: *Proceedings of the Fourteenth Australasian User Interface Conference-Volume 139.* Australian Computer Society, Inc., pp. 91–100.

Morgenstern, Y., Murray, R. F. & Harris, L. R. (2011). "The human visual system's assumption that light comes from above is weak." In: *Proceedings of the National Academy of Sciences* 108.30, pp. 12551–12553.

Nichol, J. E., Shaker, A. & Wong, M. (2006). "Application of high-resolution stereo satellite images to detailed landslide hazard assessment." In: *Geomorphology* 76.1, pp. 68–75.

Niedomysl, T., Elldér, E., Larsson, A., Thelin, M. & Jansund, B. (2013). "Learning benefits of using 2D versus 3D maps: evidence from a randomized controlled experiment." In: *Journal of Geography* 112.3, pp. 87–96.

Nyerges, T. L. (1995). "Cognitive issues in the evolution of GIS user knowledge." In: *Cognitive aspects of human-computer interaction for geographic information systems.* Springer, pp. 61–74.

Okoshi, T. (1976). "Three Dimensional Imaging Techniques, 1976." In: *New York.*

Patterson, T. (2011). *Illumination.* URL: http://www.shadedrelief.com/retro/discussion.html.

Ramachandran, V. S. (1988). "Perceiving shape from shading." In: *Scientific American* 259.2, pp. 76–83. ISSN: 0036-8733. DOI: 10.1038/scientificamerican0888-76.

Reichelt, S., Häussler, R., Fütterer, G. & Leister, N. (2010). "Depth cues in human visual perception and their realization in 3D displays." In: *SPIE Defense, Security, and Sensing.* International Society for Optics and Photonics, 76900B–76900B.

Richards, W. (1970). "Stereopsis and stereoblindness." In: *Experimental Brain Research* 10.4, pp. 380–388.

Rudnicki, W. (2000). "The new approach to the relief shading applied in the satellite image maps." In: *Proceedings of High Mountain Cartography.*

Rudnicki, W. (2005). "Perspective Views and Panoramas in Presentation of Relief Forms in Poland." In: *Proceedings of the International Cartographic Conference (ICC)*, pp. 9–16.

Saraf, A. K., Das, J. D., Agarwal, B. & Sundaram, R. M. (1996). "False topography perception phenomena and its correction." In: *International Journal of Remote Sensing* 17.18, pp. 3725–3733.

Saraf, A. K., Ghosh, P., Sarma, B. & Choudhury, S. (2005). "Cover: Development of a new image correction technique to remove false topographic perception phenomena." In: *International Journal of Remote Sensing* 26.8, pp. 1523–1529. ISSN: 0143-1161. DOI: 10.1080/0143116031000101648.

Saraf, A. K., Sinha, S. T., Ghosh, P. & Choudhury, S. (2007). "A new technique to remove false topographic perception phenomenon and its impacts in image interpretation." In: *International Journal of Remote Sensing* 28.5, pp. 811–821.

Saraf, A. K., Zia, M., Das, J. D., Sharma, K. & Rawat, V. (2011). "False topographic perception phenomena observed with the satellite images of Moon's surface." In: *International journal of remote sensing* 32.24, pp. 9869–9877.

Schröder, H. (1858). "Ueber eine optische Inversion bei Betrachtung verkehrter, durch optische Vorrichtung entworfener, physischer Bilder." In: *Annalen der Physik* 181.10, pp. 298–311.

Seipel, S. (2013). "Evaluating 2D and 3D geovisualisations for basic spatial assessment." In: *Behaviour & Information Technology* 32.8, pp. 845–858.

Sexton, I. & Surman, P. (1999). "Stereoscopic and autostereoscopic display systems." In: *Signal Processing Magazine, IEEE* 16.3, pp. 85–99.

Shaker, A., Yan, W. Y. & Easa, S. (2010). "Using stereo satellite imagery for topographic and transportation applications: an accuracy assessment." In: *GIScience & Remote Sensing* 47.3, pp. 321–337.

Shepherd, I. D. H. (2008). "Travails in the third dimension: a critical evaluation of three-dimensional geographical visualization." In:

Slocum, T. A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D. R., Fuhrmann, S. & Hedley, N. R. (2001). "Cognitive and usability issues in geovisualization." In: *Cartography and Geographic Information Science* 28.1, pp. 61–75.

Smallman, H. S., John, M. S., Oonk, H. M. & Cowen, M. B. (2001). "Information availability in 2D and 3D displays." In: *IEEE Computer Graphics and Applications* 5, pp. 51–57.

Smith, S. L. (1962). "Color coding and visual search." In: *Journal of Experimental Psychology* 64.5, p. 434.

Springmeyer, R. R., Blattner, M. M. & Max, N. L. (1992). "A characterization of the scientific data analysis process." In: *Proceedings of the 3rd conference on Visualization'92*. IEEE Computer Society Press, pp. 235–242.

Stearns, L. A. & Hamilton, G. S. (2007). “Rapid volume loss from two East Greenland outlet glaciers quantified using repeat stereo satellite imagery.” In: *Geophysical Research Letters* 34.5.

Sun, J.R. & Perona, P. (1998). “Where is the sun?” In: *Nature neuroscience* 1.3, pp. 183–184.

Taylor, J. C., Brewer, T. R. & Bird, A. C. (2000). “Monitoring landscape change in the National Parks of England and Wales using aerial photo interpretation and GIS.” In: *International Journal of Remote Sensing* 21.13-14, pp. 2737–2752. ISSN: 0143-1161. DOI: 10.1080/01431160050110269.

Tory, M. (2003). “Mental registration of 2D and 3D visualizations (an empirical study).” In: *Proceedings of the 14th IEEE Visualization 2003 (VIS’03)*. IEEE Computer Society, p. 49.

Tory, M., Kirkpatrick, A. E., Atkins, M. S. & Möller, T. (2006). “Visualization task performance with 2D, 3D, and combination displays.” In: *Visualization and Computer Graphics, IEEE Transactions on* 12.1, pp. 2–13.

Toutin, T. (1997). “Qualitative Aspects of Chromo-Stereoscopy for Depth Perception.” In: *Society* 63.2, pp. 193–203. ISSN: 00991112.

Toutin, T. (1998). “Depth perception with remote sensing data.” In: *Future Trends in Remote Sensing (edited by P. Gudmandsen: AA Balkema)*, pp. 401–409.

Ukai, K. (2006). “Human factors for stereoscopic images.” In: *2006 IEEE International Conference on Multimedia and Expo*. IEEE, pp. 1697–1700.

Ware, C. (2000). *Information visualization: perception for design*. Morgan Kaufmann.

Westheimer, G. (1994). “The Ferrier Lecture, 1992. Seeing depth with two eyes: stereopsis.” In: *Proceedings of the Royal Society of London B: Biological Sciences* 257.1349, pp. 205–214.

Wilkening, J. & Fabrikant, S.I. (2011). “The effect of gender and spatial abilities on map use preferences and performance in road selection tasks.” In: pp. 232–242.

Wise, J. A. (1999). “The ecological approach to text visualization.” In: *JASIS* 50.13, pp. 1224–1233.

Wu, B., Li, H. & Gao, Y. (2013). “Investigation and remediation of false topographic perception phenomena observed on Chang’E-1 lunar imagery.” In: *Planetary and Space Science* 75, pp. 158–166.

Yellott, J. I. & Kaiwi, J. L. (1979). “Depth inversion despite stereopsis: The appearance of random-dot stereograms on surfaces seen in reverse perspective.” In: *Perception*.

Appendix

Consent Form

The University of Zurich - **Participant Information Statement and Consent Form** **Landform identification study: A Case Study with Eye Movement Analysis**
April/May 2015
Participant No:

Purpose of study

You are invited to participate in a study regarding landform identification. We hope to learn more about how people identify landforms in monoscopic and stereoscopic satellite images.

Description of study and risks

If you decide to participate, we will ask you to begin by filling out a short background questionnaire including demographic information. This will be followed by a session at the computer where you will be asked to identify landforms in monoscopic and stereoscopic satellite images. During this process we will record your interactions with the computer using a webcam, audio recorder and eye tracking. The eye tracking device is non-contact, uses near infrared light and should not cause any discomfort. After the experiment we will ask you to fill out a second questionnaire.

The whole procedure should take approximately one hour and there are no particular risks or benefits to you from participating in this experiment.

Confidentiality and disclosure of information

Any information and that can be identified with you in connection with this study will remain confidential and will be disclosed only with your permission. If you give us permission by signing this document, we plan to publish the results of this research in scientific publications. In any publication, information will be provided in such a way that you cannot be identified.

Compensation

We do not provide any compensation for your participation in this experiment, nor are there any costs for you for your participation.

Feedback to participants

If you would like to be kept informed about the results of this research, please leave your name and contact details with the experiment leader. A copy of publications resulting from this research will be sent to you when available.

Your consent

Your decision whether or not to participate will not prejudice your future relations with University of Zurich. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without prejudice.

If you have any questions, please feel free to ask us. If you have any additional questions later, Dr. Arzu Coltekin (044 6355440, arzu@geo.uzh.ch) will be happy to answer them.

You will be given a copy of this form to keep.

Page 1 of 2

The University of Zurich - **Participant Information Statement and Consent Form (continued)** **Landform identification study: A Case Study with Eye Movement Analysis**
April/May 2015
Participant No:

You are making a decision whether or not to participate. Your signature indicates that, having read the information provided above, you have decided to participate.

……………………………………. ………………………………

Signature of Research Participant Signature of Experimenter

……………………………………. ………………………………

Please PRINT name Please PRINT name

…………………………………….

Date and Place

The University of Zurich - **Participant Information Statement and Consent Form (continued)** **Landform identification study: A Case Study with Eye Movement Analysis**
April/May 2015
Participant No:

REVOCATION OF CONSENT

Evaluating Interface Design for Interactive Geovisualizations: A Case Study with Eye Movement Analysis

I hereby wish to WITHDRAW my consent to participate in the research proposal described above and understand that such withdrawal WILL NOT jeopardize any treatment or my relationship with The University of Zurich.

……………………………………. ………………………………

Signature Date

…………………………………….

Please PRINT name

This section of Revocation of Consent should be forwarded to Dr. Arzu Coltekin, Geographic Information Visualization and Analysis, Dept. of Geography, University of Zurich, CH-8057, Zurich.

Page 2 of 2

Pre-Questionnaire

1_TRE_prequestionnaire

1. Participant number (to be inserted by the supervisor)

1. Gender:

- ◯ Male
- ◯ Female

2. Age group:

◯ 10-19 ◯ 20-29 ◯ 30-39 ◯ 40-49 ◯ 50-59 ◯ 60-69 ◯ 70-79

3. Level of education

- ◯ Less than highschool (Gymansium)
- ◯ Highschool (Gymansium)
- ◯ University: Bachelor degree or equivalent
- ◯ Univeristy: Master degree or equivalent
- ◯ University: Doctoral degree

prescription glasses / contact lenses

1. Do you use prescription glasses or contact lenses?

- ◯ Yes
- ◯ No

prescription glasses / contact lenses II

1. Do you use them now?

- ◯ Yes
- ◯ No

color blindness

1. Have you ever been told by a professional that you have color blindness?

- ◯ Yes
- ◯ No

1_TRE_prequestionnaire

color blindness II

1. If you know, please note what type of color deficiency you have and how strong it is (or type "I don't know")

visual impairment

1. Have you ever been told by a professional that you have any other visual impairment (Beeinträchtigung), such as stereoscopic vision deficiency (3D-vision impairment) or .. ? If yes, please note shortly.

1. I am

1 totally left-handed	2	3 ambiguous	4	5 totally right-handed
○	○	○	○	○

1. Please rate your experience in the following fields:

	1 No experience	2	3	4	5 Professional
Fine arts	○	○	○	○	○
Graphic design	○	○	○	○	○
Photography	○	○	○	○	○
Cartography	○	○	○	○	○
Satellite imagery	○	○	○	○	○
Google Earth or other online map providers	○	○	○	○	○
Photo-interpretation	○	○	○	○	○

Page 2

1_TRE_prequestionnaire

1. How many hours did you sleep last night?

- ◯ Less than 3 hours
- ◯ 3 hours
- ◯ 4 hours
- ◯ 5 hours
- ◯ 6 hours
- ◯ 7 hours
- ◯ 8 hours
- ◯ More than 8 hours

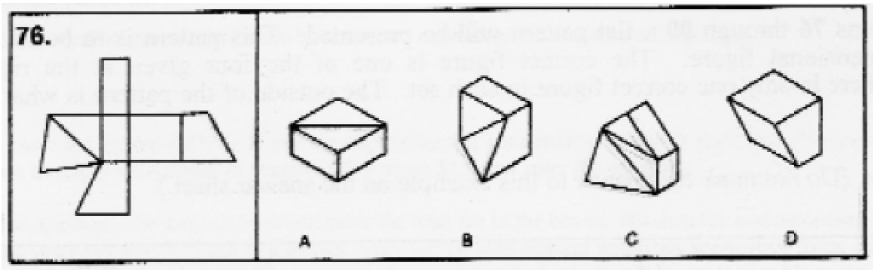
Thank you. We will now move on to the next section.

Spatial ability Test

3_spatial_ability_test

Participant number (to be inserted by the supervisor)

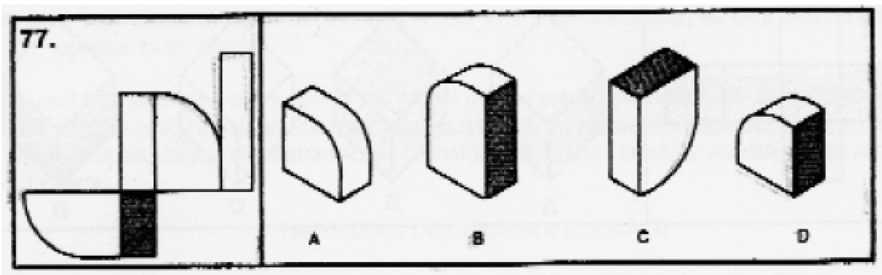
1



Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

2

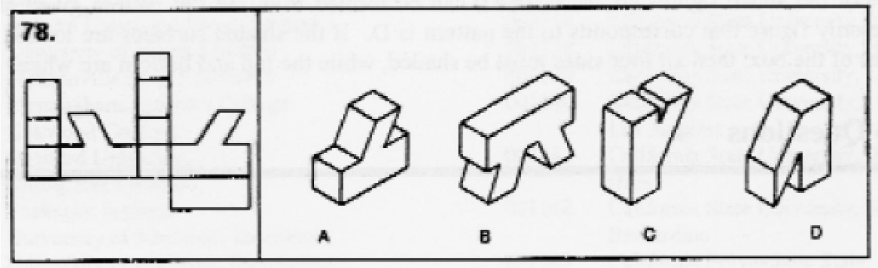


Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

3

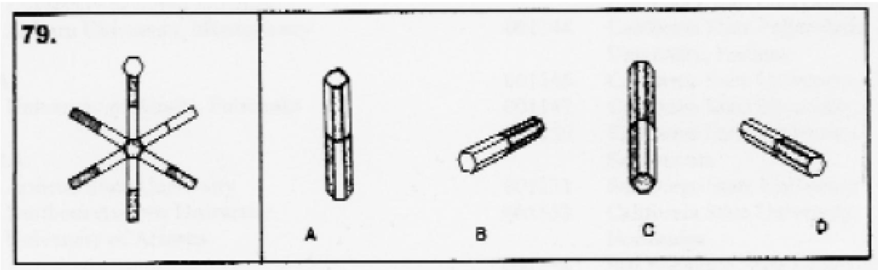
3_spatial_ability_test



Which object fits to the unfolded version on the left side?

- A
- B
- C
- D

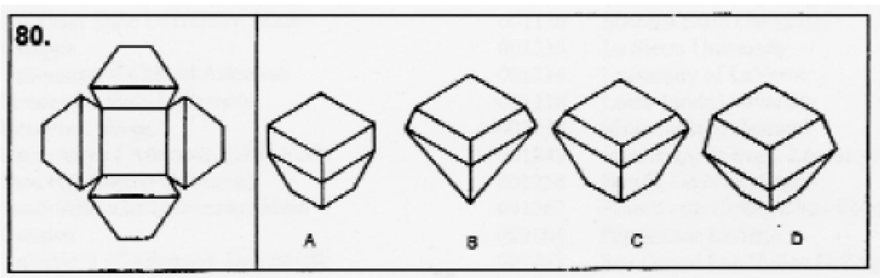
4



Which object fits to the unfolded version on the left side?

- A
- B
- C
- D

5

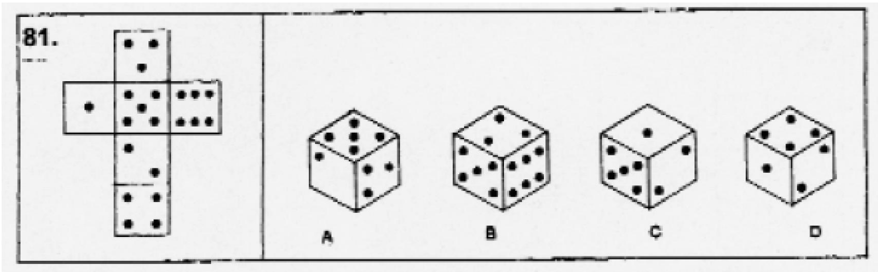


3_spatial_ability_test

Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

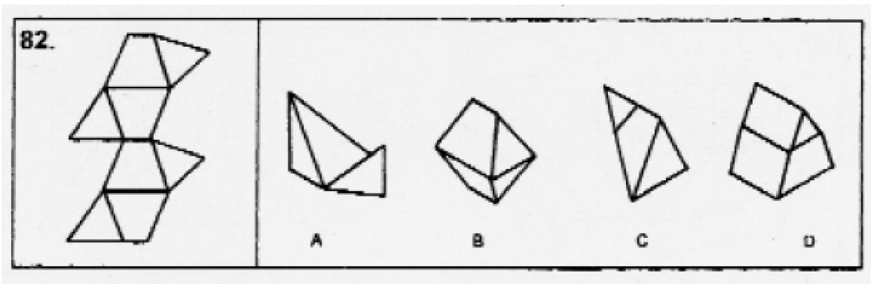
6



Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

7

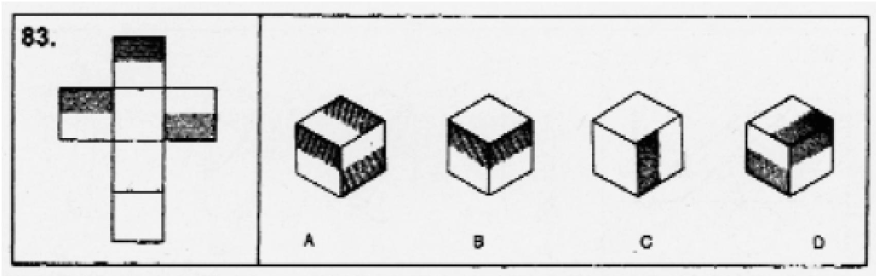


Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

3_spatial_ability_test

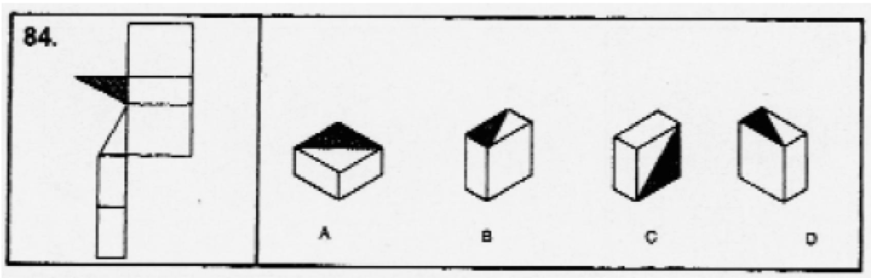
8



Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

9

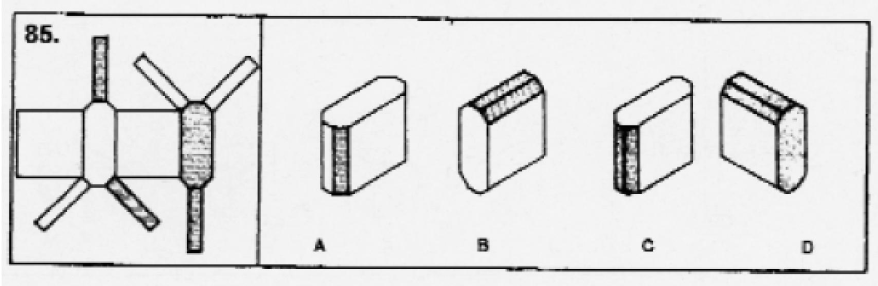


Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

10

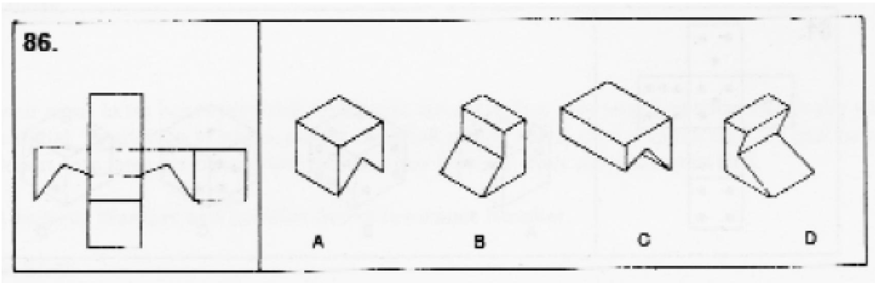
3_spatial_ability_test



Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

11

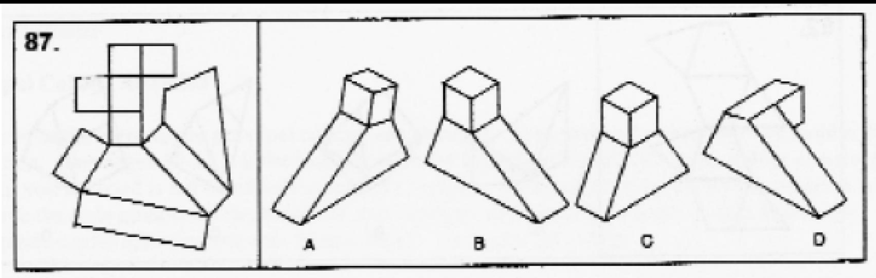


Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

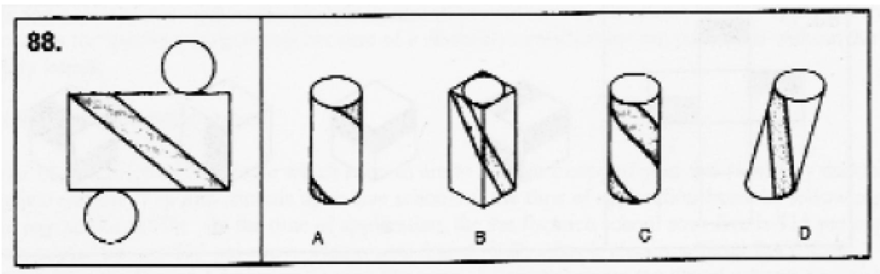
12

3_spatial_ability_test



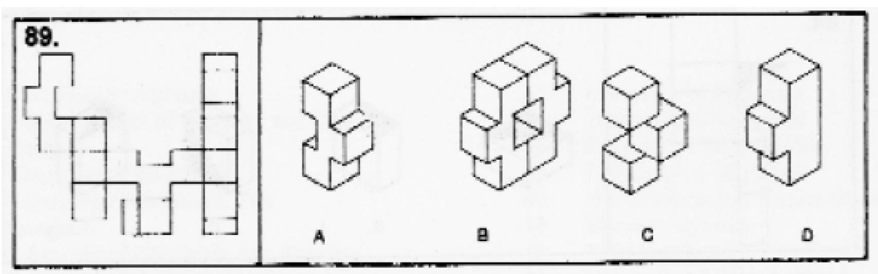
Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D



Which object fits to the unfolded version on the left side?

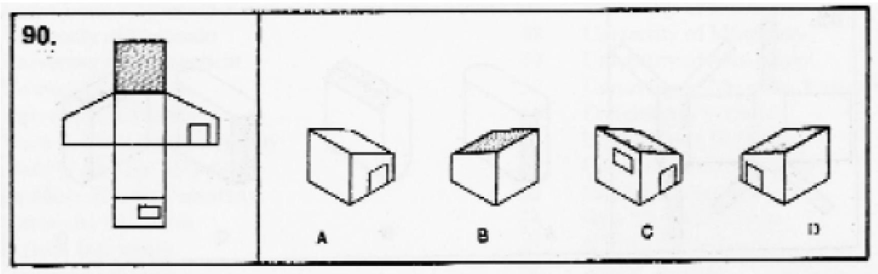
- ◯ A
- ◯ B
- ◯ C
- ◯ D



3_spatial_ability_test

Which object fits to the unfolded version on the left side?

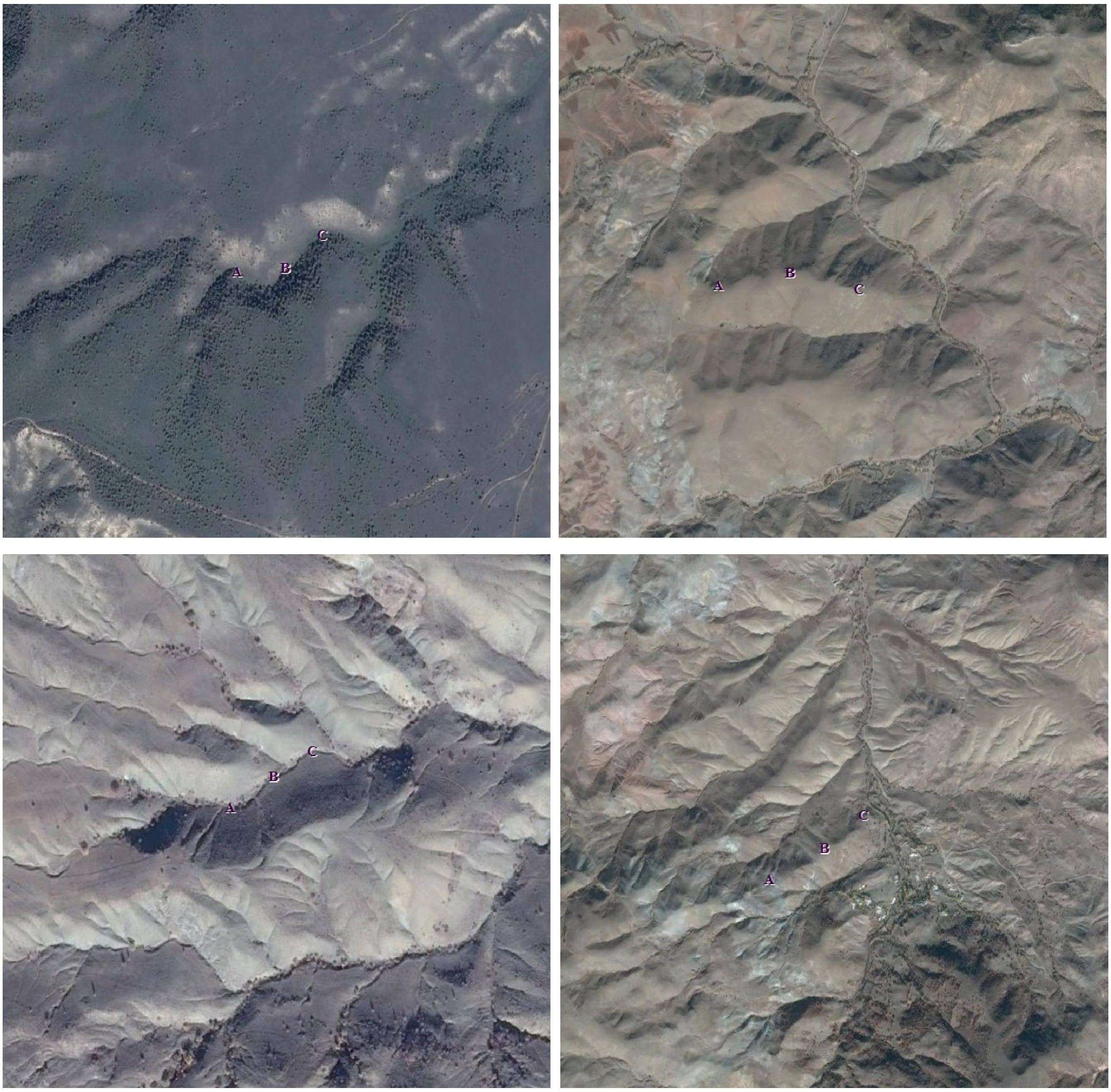
- ◯ A
- ◯ B
- ◯ C
- ◯ D

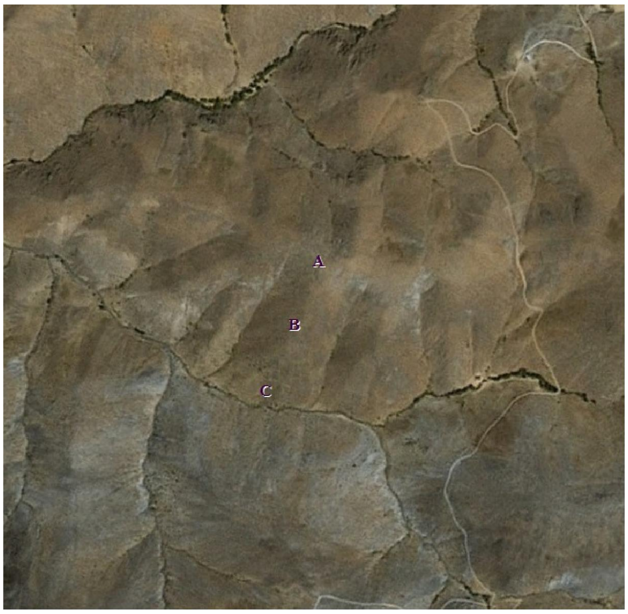


Which object fits to the unfolded version on the left side?

- ◯ A
- ◯ B
- ◯ C
- ◯ D

Terrain Type - Subtle Change




A
B
C


A
B
C

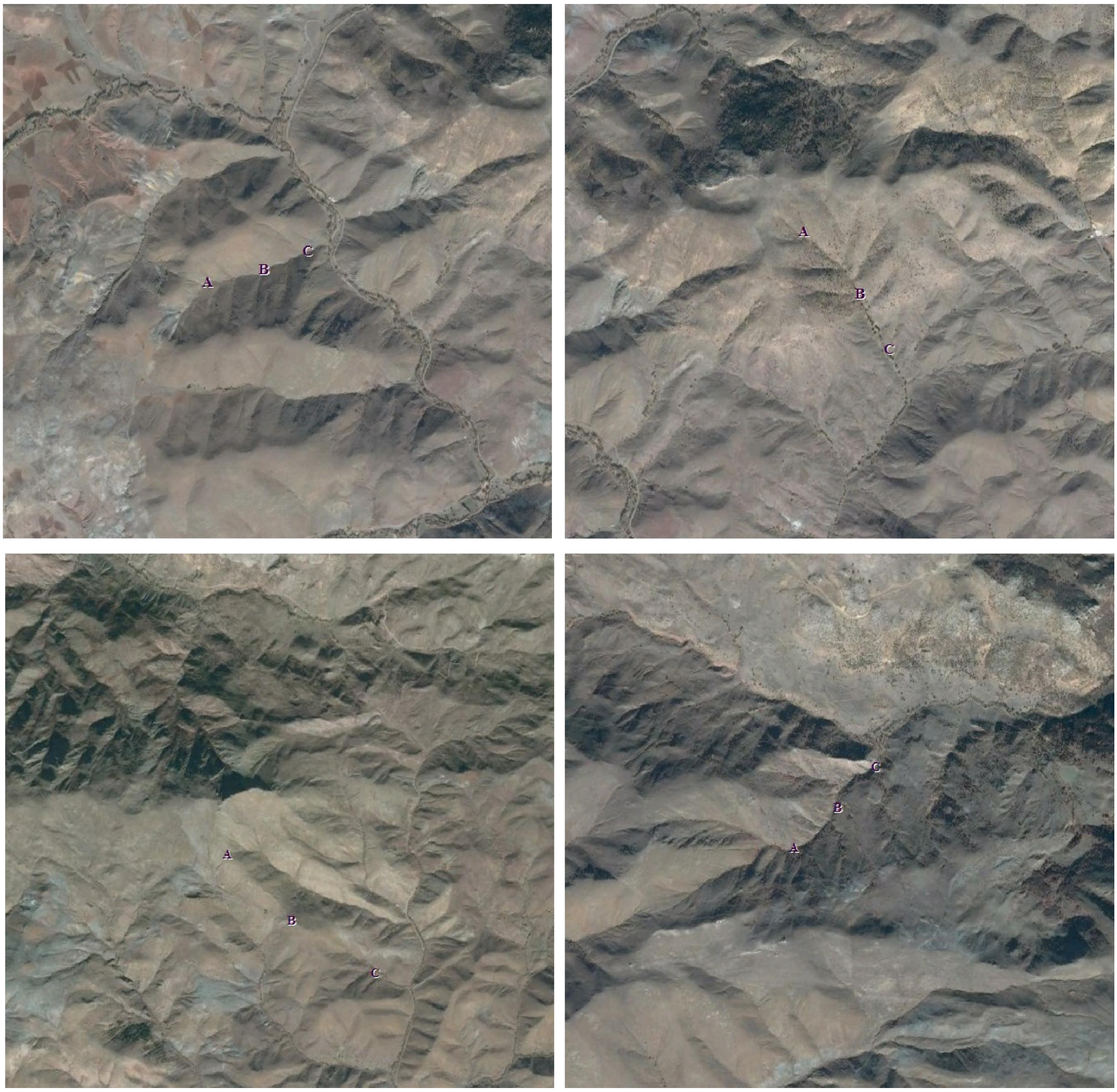

A
B
C

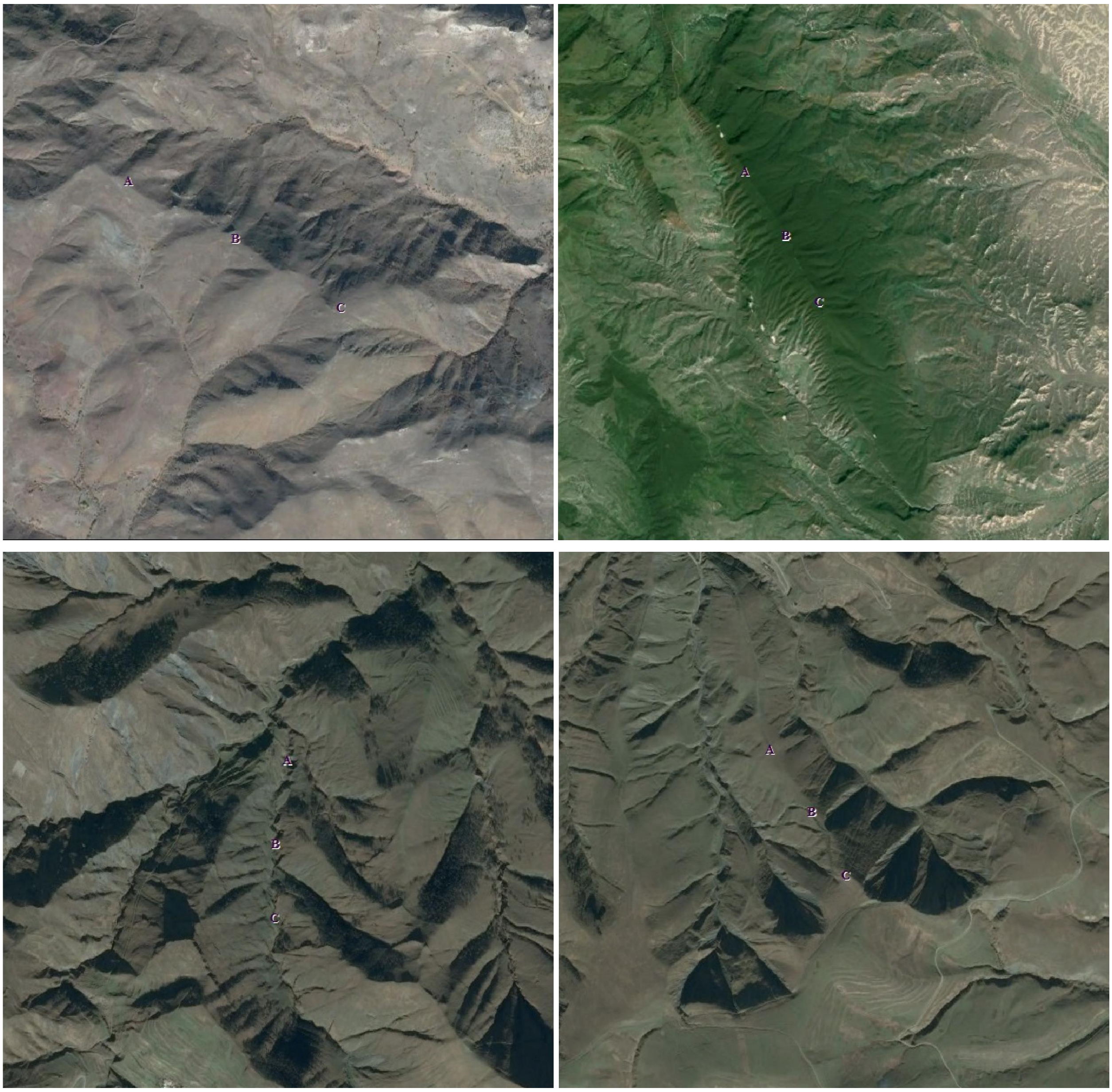

B
C
A


A
B
C

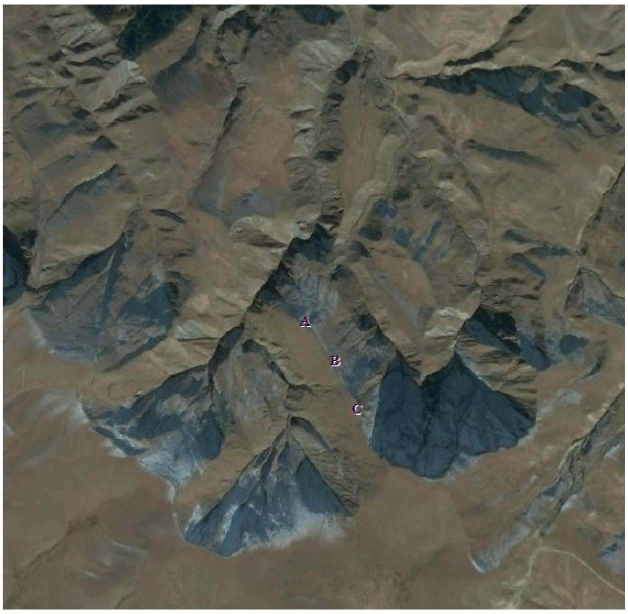

A
B
C

Terrain Type - Middle Rugged




A
B
C
A
B
C
A
B
C
A
B
C

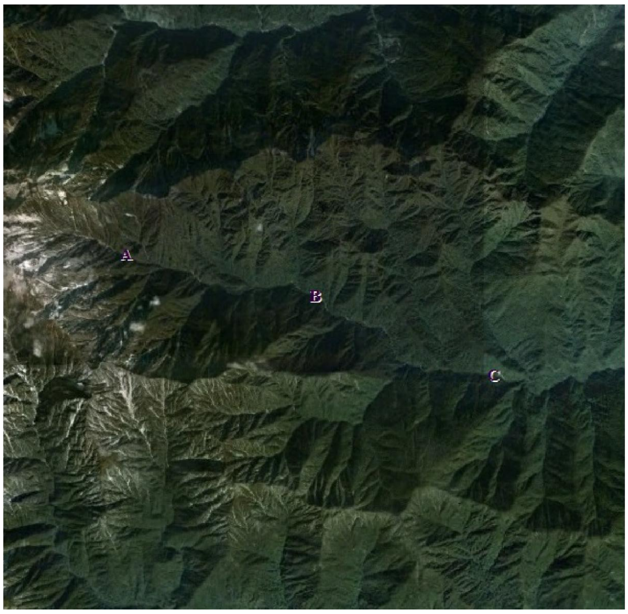

A
B
C


A
B
C

Terrain Type - Highly Rugged


A
B
C


A
B
C


A
B
C


A
B
C


C
B
A


A
B
C

Post-Questionnaire

5_TRE_postquestionnaire

Participant number (to be inserted by the supervisor)

General

Did you find the tasks boring or tiring?

1 low boring/ tiring	2	3 medium boring/ tiring	4	5 high boring/ tiring
○	○	○	○	○

Preference

Now you worked with monoscopic and stereoscopic images. Which one do you prefer?

- ○ Monoscopic images
- ○ Stereoscopic images
- ○ Neither of them

Why?

Anaglyph viewing

Discomfort (Unwohlsein) level with anaglyph viewing

1 low discomfort	2	3 medium discomfort	4	5 high discomfort
○	○	○	○	○

Contradiction

Have you ever noticed a contradiction (Widerspruch) between landform and landcover (snow, vegetation, river)?

- ○ Yes
- ○ No

Contradiction 2

Please explain briefly the observed contradictions (Widersprüche).

How did you answer the question then?

- ○ Based on my perception
- ○ Based on my interpretation
- ○ Can't remember

Did you participate in a similar experiment before (e.g. Julien Biland's study)?

- ○ Yes
- ○ No

Page 1

5_TRE_postquestionnaire

In this study we have been testing the terrain reversal effect, a perceptual phenomenon where landforms appear inverted, e.g. we perceive valleys as ridges and vice versa (see image below). Did you notice this phenomenon?

- Yes
- No

Same region 0° and 180° turned

End

Thank you very much for taking the time to participate in this study. If you have any comments, please note them here.

Page 2

Images compilation

OVERVIEW OF THE COMPILATION OF THE IMAGES

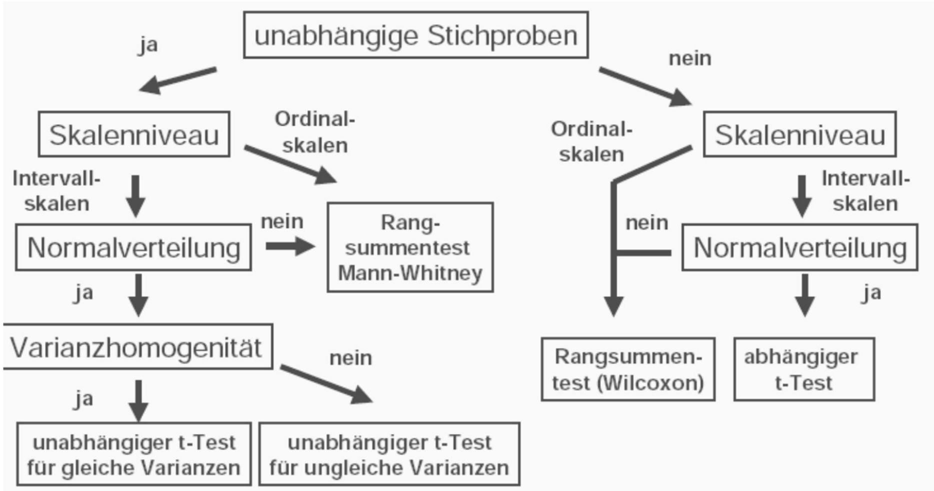
TOTAL	IMAGE ORIENTATION	VISUALIZATION TYPE	TASK TYPE	TERRAIN TYPE
Total → 240 images	North orientated → 120 images	Non-stereo → 120 images	Task Type 1 → 120 images	Subtle change (sc) → 80 images
	180° turned → 120 images	Stereo → 120 images	Task Type 2 → 120 images	Middle rugged(mr) → 80 images
				Highly rugged (hr) → 80 images

TOTAL	IMAGE ORIENTATION	VISUALIZATION TYPE	TASK TYPE	TERRAIN TYPE
A total of 240 images/ participant	120 images north orientated	60 images non-stereo	30 images Task Type 1	10 images sc
				10 images mr
				10 images hr
			30 images Task Type 2	10 images sc
				10 images mr
				10 images hr
		60 images stereo	30 images Task Type 1	10 images sc
				10 images mr
				10 images hr
			30 images Task Type 2	10 images sc
				10 images mr
				10 images hr
	120 images 180° turned	60 images non-stereo	30 images Task Type 1	10 images sc
				10 images mr
				10 images hr
			30 images Task Type 2	10 images sc
				10 images mr
				10 images hr
		60 images stereo	30 images Task Type 1	10 images sc
				10 images mr
				10 images hr
			30 images Task Type 2	10 images sc
				10 images mr
				10 images hr

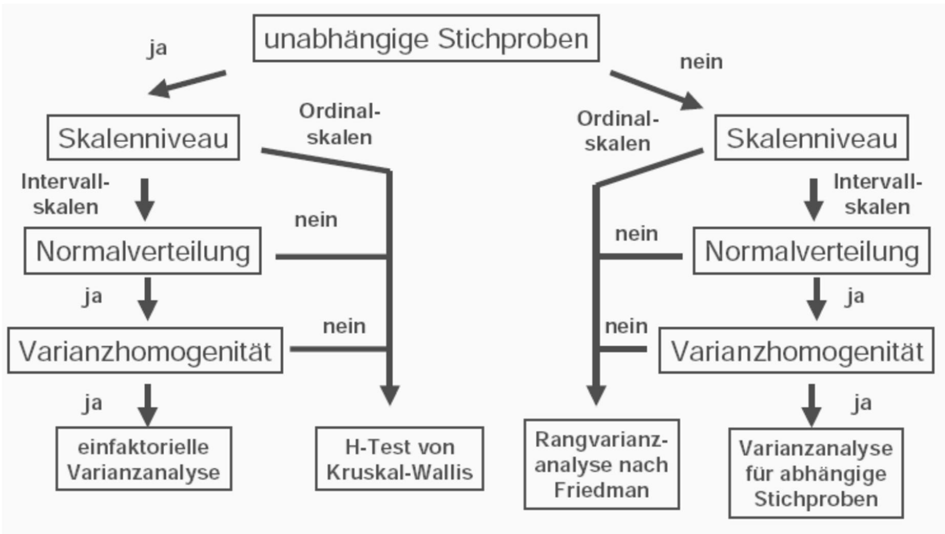
Statistics

DECISION TREES* FOR THE RIGHT STATISTICAL TEST

Decision tree when **two** samples are given:



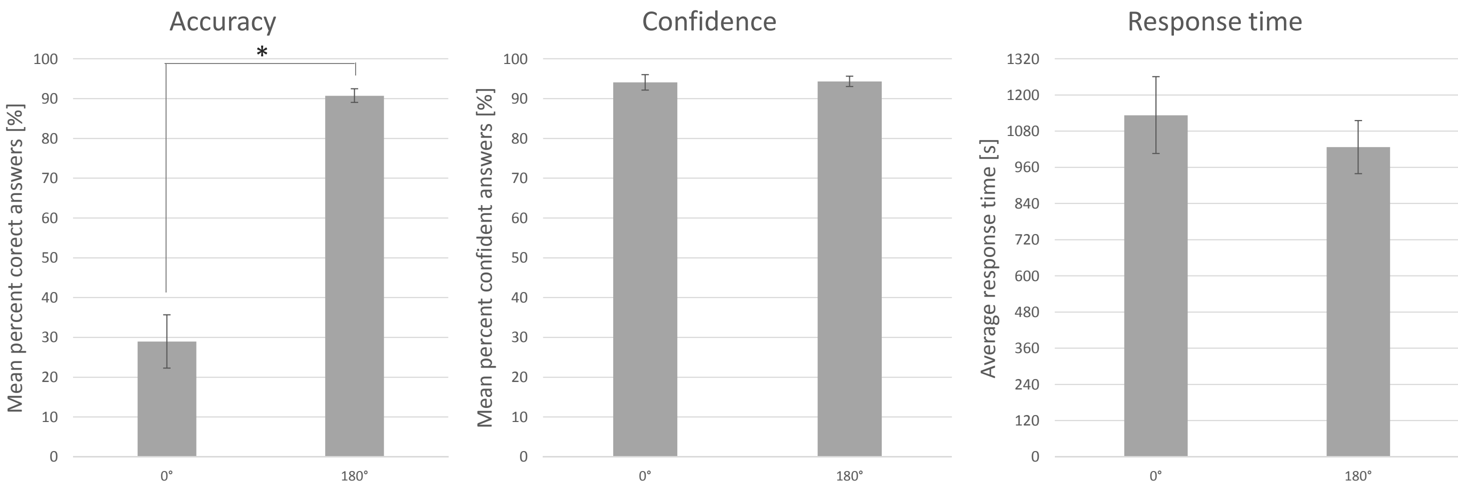
Decision tree when **more than two** samples are given:



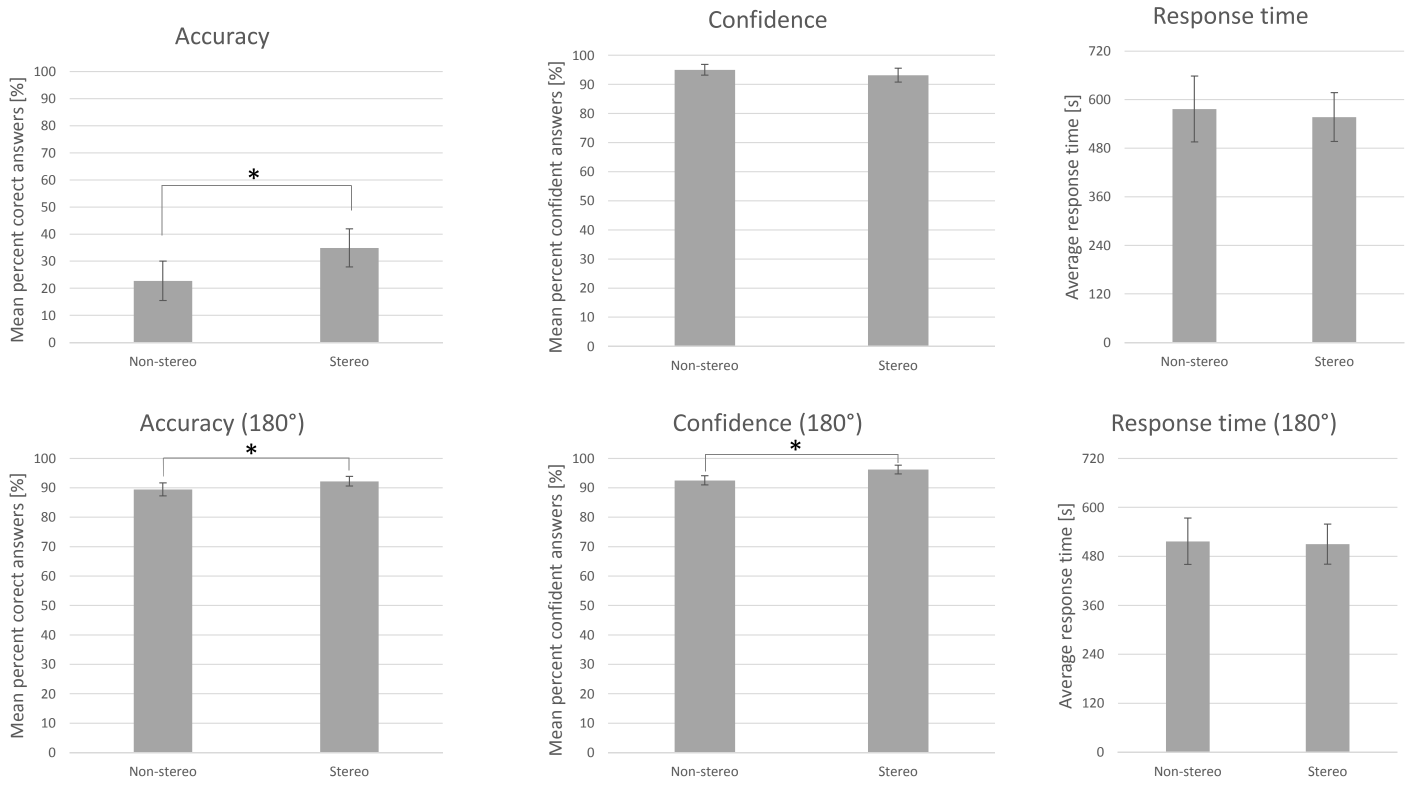
* Illustrations are from the statistical course GEO 246 from the Department of Geography at the University of Zurich

Statistics

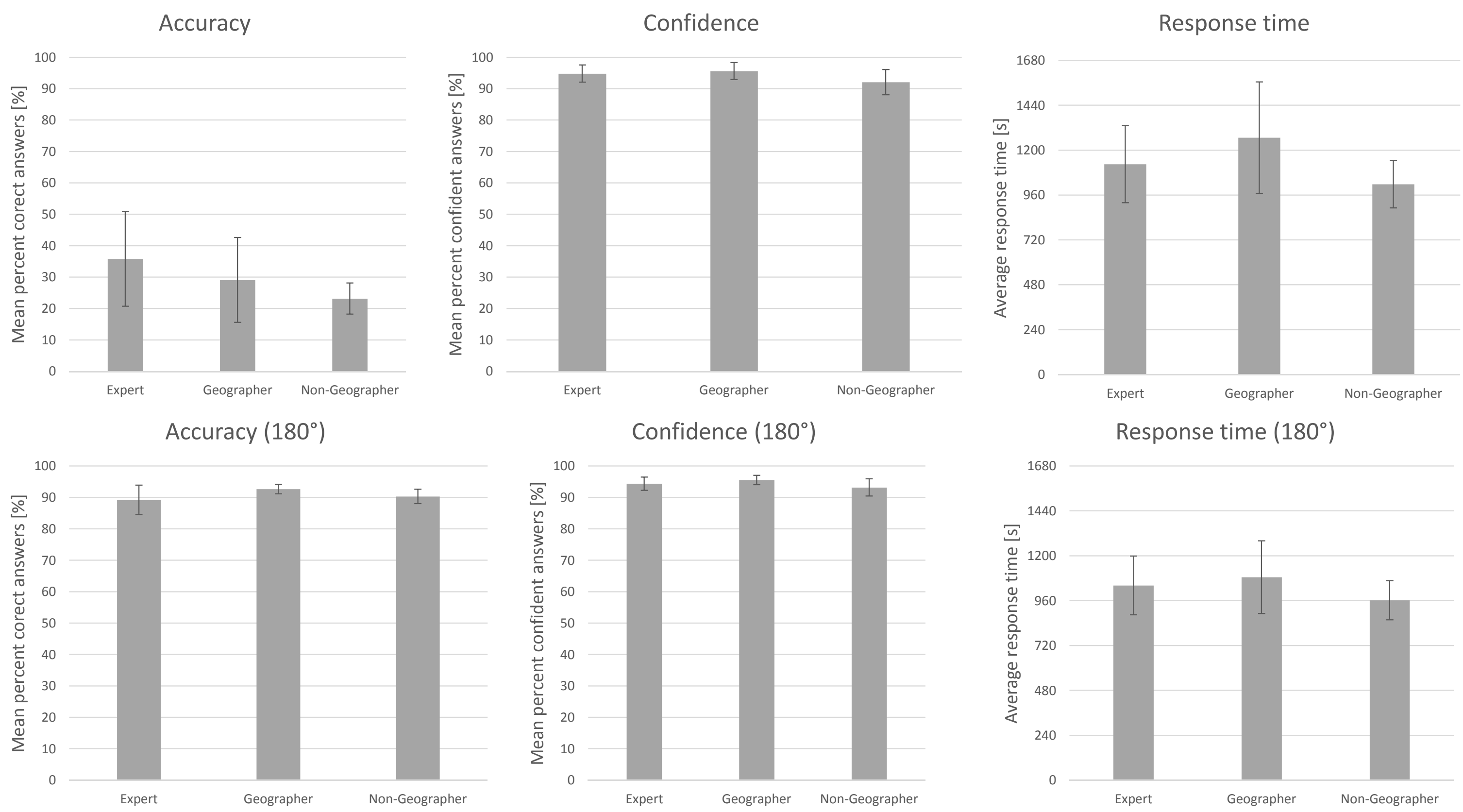
Image orientation	120 images for both image orientations	Accuracy, confidence rate, response time



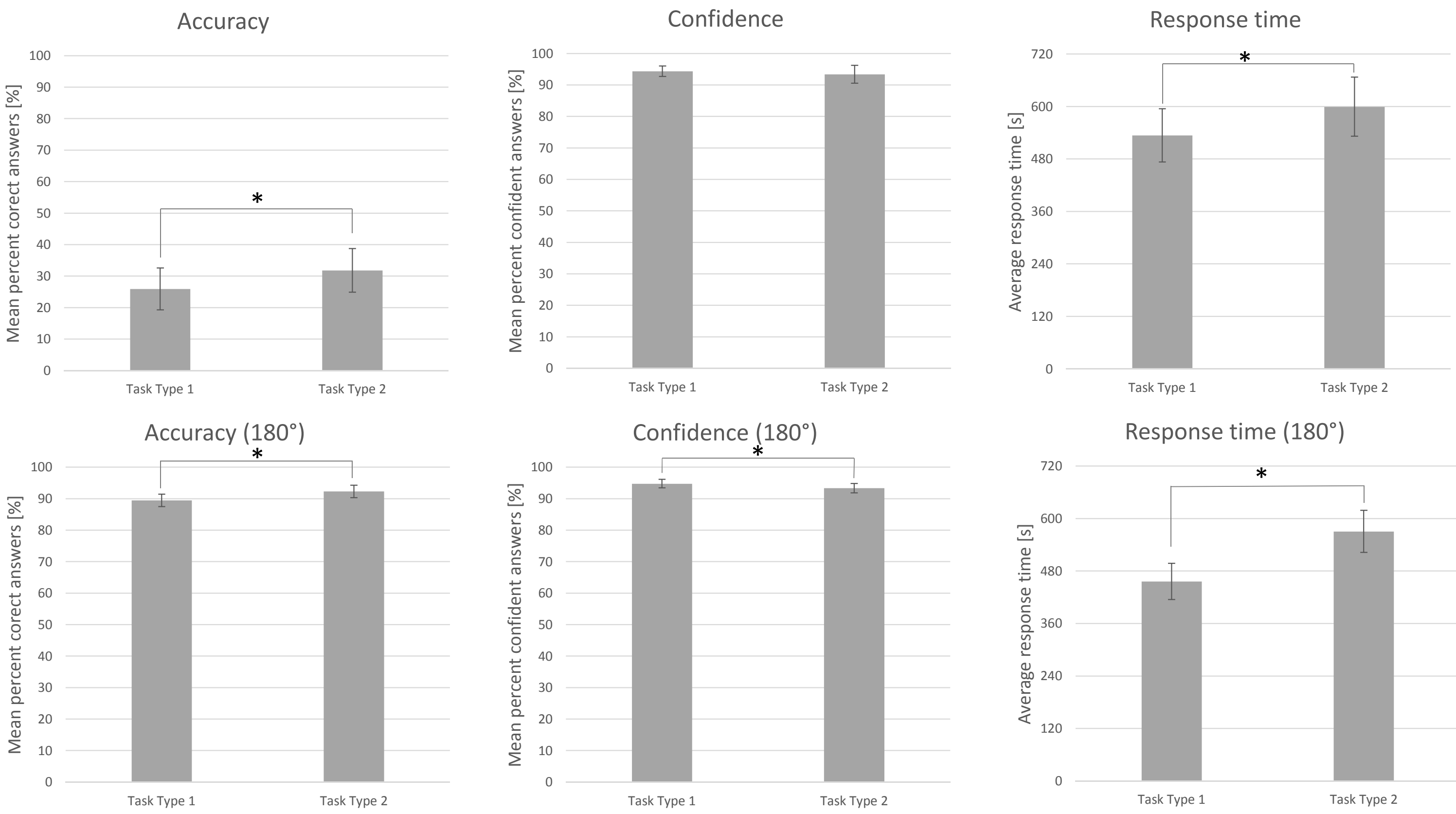
Visualization type (0° & 180°)	120 images for non-stereo and 120 images for stereo part	Accuracy, confidence rate, response time

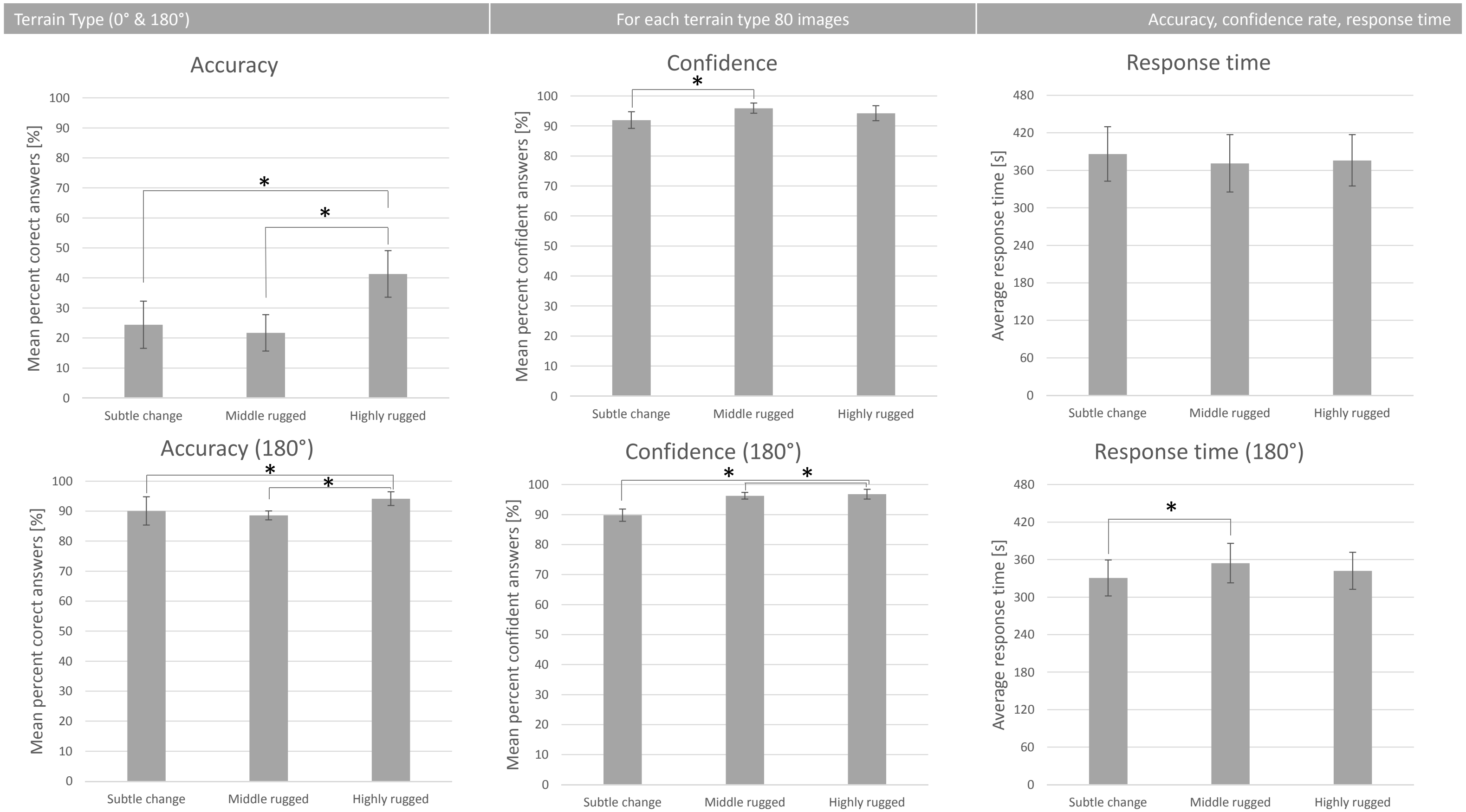


Level of expertise (0° & 180°) | 240 images / participant | Accuracy, confidence rate, response time

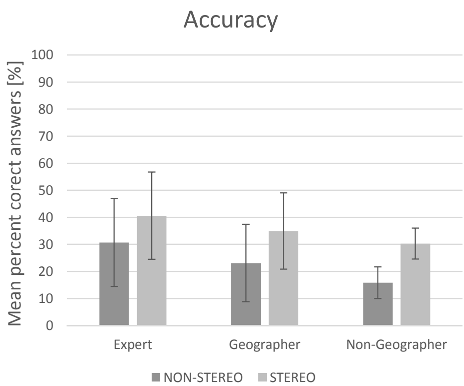


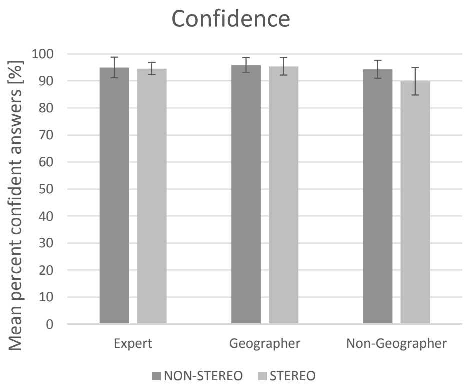
Task Type (0° & 180°)	120 images for Task Type 1 and 120 images for Task Type 2	Accuracy, confidence rate, response time

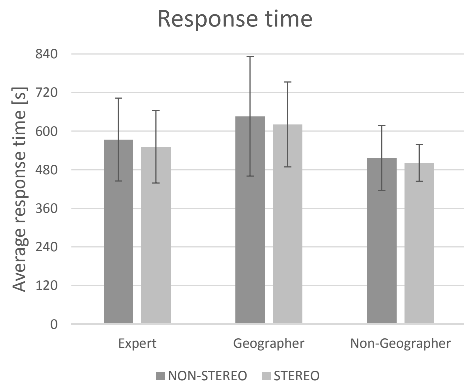


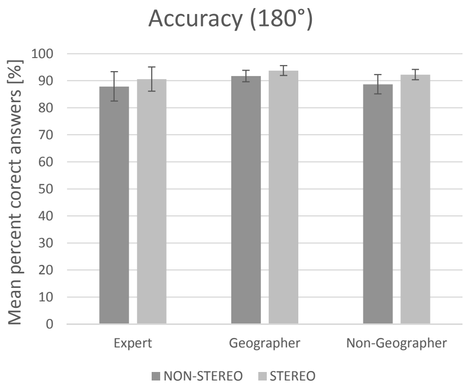

Terrain Type (0° & 180°)
For each terrain type 80 images
Accuracy, confidence rate, response time
Accuracy
Mean percent corect answers [%]
Subtle change
Middle rugged
Highly rugged
Confidence
Mean percent confident answers [%]
Response time
Average response time [s]
Accuracy (180°)
Confidence (180°)
Response time (180°)
*

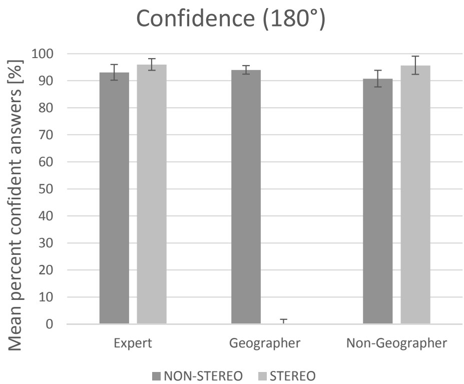
Visualization type & Level of expertise (0°) | Accuracy, confidence rate, response time

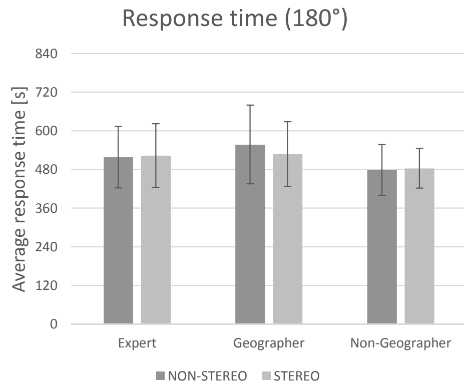




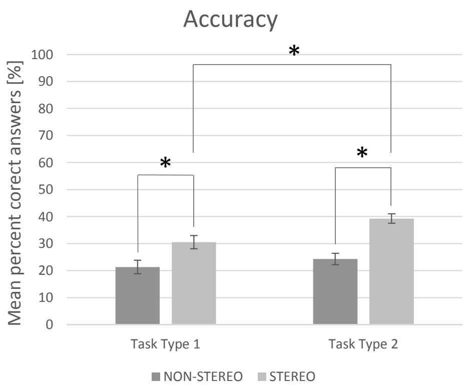


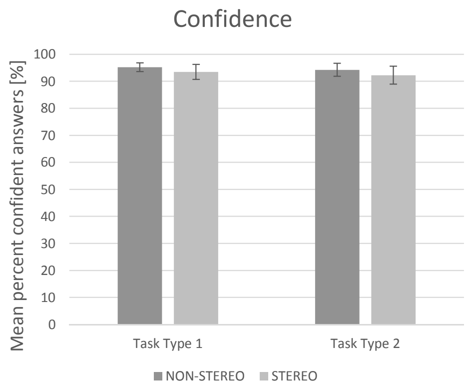


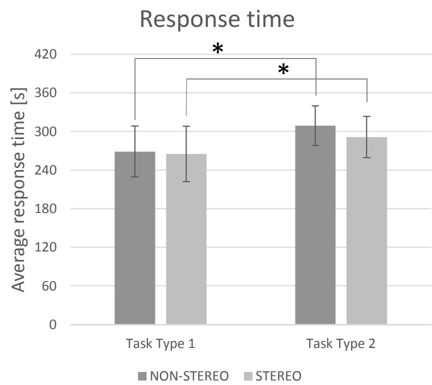


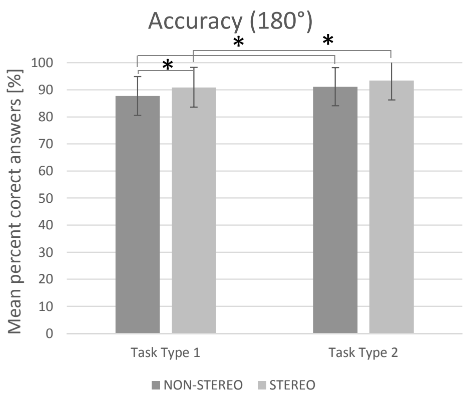


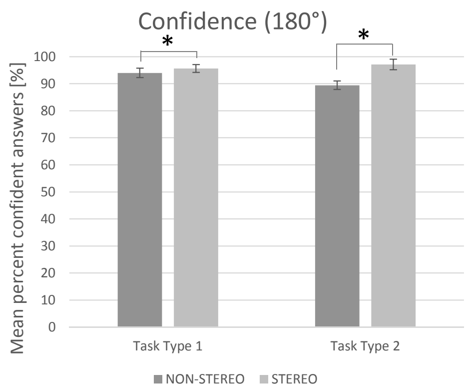
Visualization type & Task Type (0°) | | Accuracy, confidence rate, response time

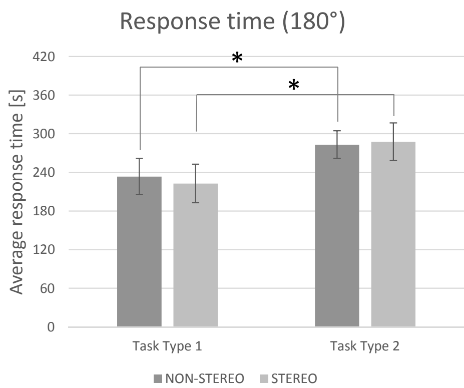




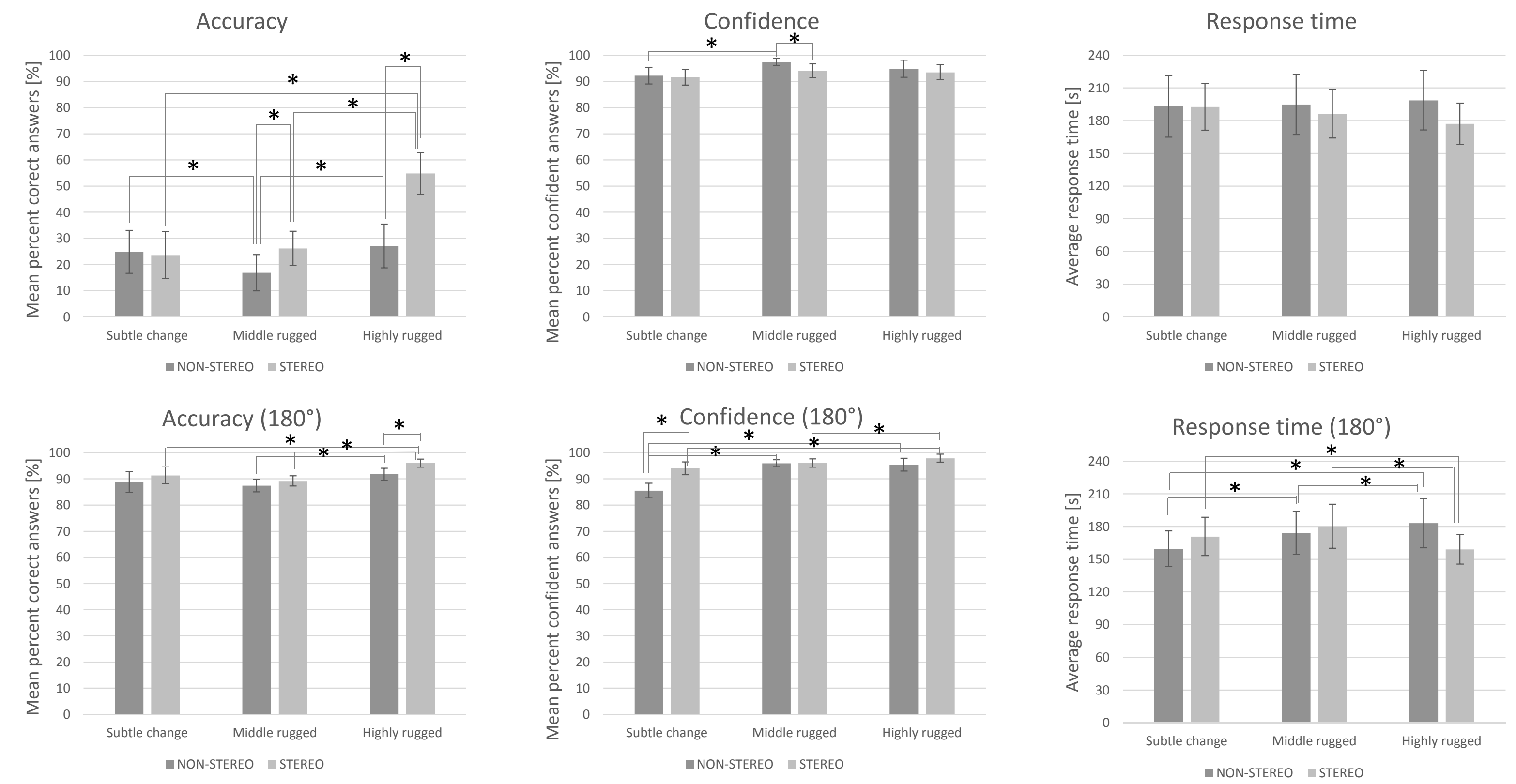








Visualization type & Terrain Type (0° & 180°) | Accuracy, confidence rate, response time



Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Martina Meyer
Zurich, September 30th 2015