

LANDFORM PERCEPTION ACCURACY IN SHADED RELIEF MAPS: A REPLICATION STUDY CONFIRMS THAT NNW LIGHTING IS BETTER THAN NW AGAINST THE RELIEF INVERSION EFFECT

A. Çöltekin^{1,*}, V. Rautenbach², S. Coetzee², T. Mokwena²

¹ Department of Geography, University of Zurich, CH-8057, Zurich, Switzerland – arzu@geo.uzh.ch

² Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa – {victoria.rautenbach, serena.coetzee}@up.ac.za, tebogoi renemokwena@gmail.com

Commission IV, WG IV/9

KEY WORDS: Relief inversion effect, Terrain reversal effect, Shaded relief maps, Landform perception, Light direction

ABSTRACT:

Relief inversion effect is a perceptual phenomenon that leads to an inverted perception of convex and concave shapes. This perceptual inversion occurs in scenes where the shading/shadows act as the main depth cue. In visuospatial displays, such as shaded relief maps, the positioning of the shadows in the northern slopes, thus when light source placed broadly in south, mislead the cognitive system based on the 'light from above prior' assumption (Mamassian and Goutcher 2001). Thus, assuming the light must come from above, our mind creates an illusion, and we perceive the landforms incorrectly. To judge the 3D spatial relationships in terrain representations correctly, the relief inversion effect must be avoided. Cartographic convention against this effect is to place the light source at north-west (NW), whereas a recent study demonstrated that north-north-west (NNW), or even north yields more precise results (Biland and Çöltekin, 2016). Since this finding goes against decades of convention, to establish its validity further, we attempted replicating the results with a different sample in South Africa. In this paper, we present our findings, which broadly confirm that the NNW (or also N) is better than NW against the relief inversion effect.

1. INTRODUCTION AND BACKGROUND

The direction at which light illuminates objects may impact the perception of the objects, especially 3D shapes (Ramachandran, 1988). This is linked to the phenomenon called *light from above prior* in which the cognitive system assumes that the light source is above (Mamassian and Goutcher 2001)—possibly for environmental reasons—, and slightly to the left (Sun and Perona, 1998) —possibly explained by dominant right-handedness in people. As a consequence, the human visual system (HVS) is used to processing the visual input in which the shadows are always "below". When shadows are above, the mind still assumes that these surfaces should be below, and creates a stunningly strong illusion in which we perceive convex shapes concave and *vice versa*. This can be demonstrated: If one rotates Figure 1, circles on the left become craters and circles on the right become bumps for majority of the viewers.

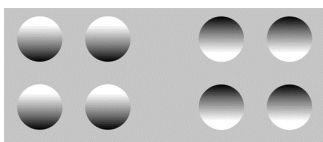


Figure 1. The circles on the left are illuminated from above, and appear as bumps, whereas the circles on the right are illuminated from below and appear as craters. The Figure is recreated from Ramachandran, 1988.

The relief inversion phenomenon expresses itself in visuospatial displays where the main depth cue is shadow. We see relief

inversion (also named 'terrain reversal') in geographic displays as well. For example, we see this effect in orthogonal satellite images that (e.g., Saraf et al., 1996, 2007; Bernabe-Poveda and Çöltekin 2014; Çöltekin and Biland 2018). Various propositions to solve the problem with satellite images do exist (e.g., Bernabé-Poveda, et al., 2011, Zhang et al. 2016). Even more critically, light direction plays an important role for shaded relief maps (SRMs) where 3D landforms are represented through shading. As with other types of displays, using a 'wrong' light direction in SRMs creates the undesired consequence of relief inversion: the 3D landforms appear inverted (Imhof, 1965). A key difference between satellite images and the SRMs is that with the SRMs, we can precisely control the light direction. Therefore, to avoid this effect, classical cartography texts recommend that the light source for SRMs should be placed at 315 degrees azimuth (NW) and 45 degrees elevation based on convention (Kraak and Ormeling 2010; Slocum et al. 2008). However, in a recent study, Biland and Çöltekin (2016) demonstrated that incident light at 337.5 degrees (NNW) results in higher accuracy than the recommended 315 degrees (NW) when participants identify 3D landforms in SRMs.

In this paper, we replicate Biland and Çöltekin's (2016) study based on a different set of participants to re-examine the robustness of the findings presented in the mentioned paper. The remainder of the paper is structured as follows: In section 2, we describe the methodology; in section 3 we present the results; and

* Corresponding author

in section 5 we discuss the implication of the results and conclude the paper.

2. METHODS

2.1 Study design

Because ours is a replication effort, we designed the user study based on the work presented by Biland and Çöltekin (2016). This was a within-subject user study, that is, all participants completed all of the tasks. The independent variable in the study was the varying light direction. In this replication study, we worked with five light directions in the NW quadrant at equal intervals. As dependent variables, we measured participants' accuracy (i.e., effectiveness) and their confidence in the accuracy of their responses.

2.2 Materials

Because the goal was to test the robustness of the findings regarding the recommended illumination angle, we used only the images that were produced using a light direction in the NW quadrant; specifically, the SRMs that were rendered with a light direction of 270, 292.5, 315, 337.5 and 0 degrees. These five light directions are only a subset of the SRMs created by Biland and Çöltekin (2016), where they tested also eastern and southern quadrants. The original SRMs were created using ASTER GDEM available from the United States Geological Survey (USGS). Figure 2 demonstrates an example of an SRM used in this study (courtesy of Biland and Çöltekin, 2016).

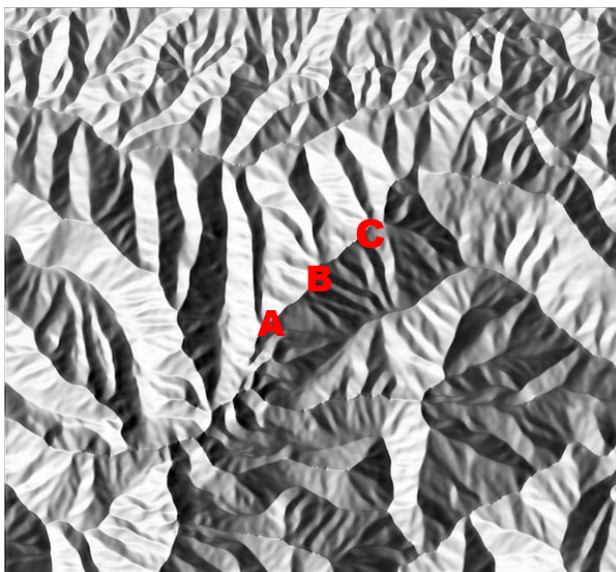


Figure 2. A shaded relief map (SRM) showing a ridge that is illuminated at 270 degrees where the relief inversion (also known as terrain reversal effect) might be present for some viewers. Assuming you see the ABC as a ridge, if you rotate this image, the ridge should turn to a valley (or vice versa). This figure is adapted from Biland and Çöltekin (2016).

2.3 Participants

Eighty-seven university students (44 males, 43 females) ranging from 20 to 29 years old (average age of 22) participated in the study. All the students were pursuing an undergraduate degree at the time of the study. The majority was enrolled for BSc Environmental Science (32.2%), followed by BSc Geography

(19.5%), BSc Geoinformatics (16.1%), BSc Meteorology (16.1%), BSc Geology (8.0%), and other related fields. To establish participants' experience with various types of stimuli they were asked to indicate how often they work or interact with the following for their studies and in their leisure time: fine arts, photography interpretation, 3D (e.g. Google Earth), satellite imagery, and cartography. Participants indicated that they interact (on average) with all of these stimuli often, both for their studies and in their leisure time. Participants' responses on the experience and training levels are shown in Figure 3.

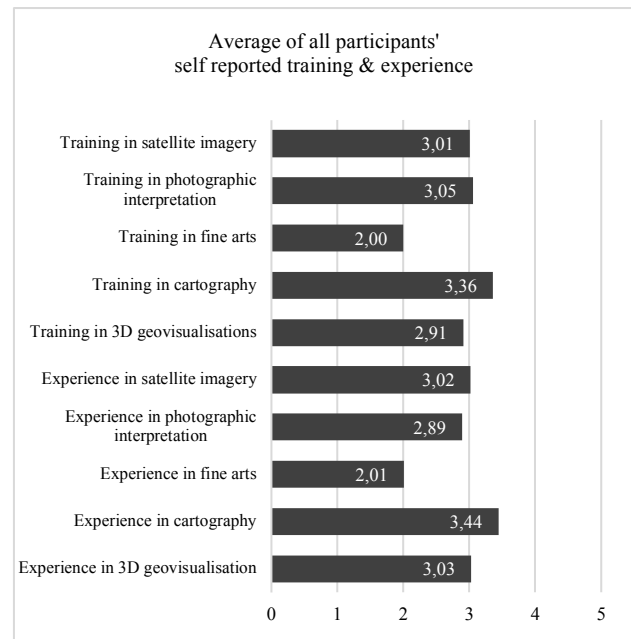


Figure 3. Participants' self-reported training and experience in relevant domains (averaged for all participants). In the horizontal axis, 0 means no training/no experience, and 5 means highly trained/very experienced.

2.4 Procedure

The study was conducted in a classroom. We consider the classroom experiment 'controlled', as all participants were exposed to identical conditions: The computers, the display sizes, keyboards as well as the bandwidth were identical for all participants. Before participants started, the instructor explained the process, what would be expected of them, and asked participants not to communicate with each other during the session (thus, each participant solved the questions alone). Two assistants were also present to aid the participants with any technical issues. Participants were informed that they had an opportunity to ask any questions before starting with the study. At this point, the experiment began. Participants were first presented with various questions to capture demographic information (e.g. gender, age and level of education). First a set of demographic questions were presented using a software called Qualtrics (<https://www.qualtrics.com/uk/>). Then, the SRMs and the main tasks were presented in randomized order, again using Qualtrics. The study consisted of 40 questions (20 SRMs featured ridges, and 20 valleys). Using the SRMs, participants were asked to identify a specific landform (either a valley or a ridge) with each SRM. The question was posed using a five-point Likert scale (i.e., clearly a valley, maybe a valley, not sure, maybe a ridge, and clearly a ridge), as it was in the original study. Using a Likert scale allows measuring ambiguity of the perceived landform, but also participants' confidence in their response is inferred from it (i.e., 1 and 5 are high confidence, 2 and 4 are

lower than 1 and 5, whereas 3 is neutral). After participants responded to the questions, and completed the tasks, we thanked them and finished the session. The participation was voluntary, and we offered no compensation to the participants.

3. RESULTS

The 40 tasks were completed by eighty-seven participants, thus we collected a total of 3480 responses. An accuracy score in percent was calculated for each participant (maximum number of correct responses per participant would be 40). The accuracy score is negatively skewed (-1.594) and has a kurtosis of 2.805. The skewness of the accuracy score per light direction varied: 270 degrees (-0.101), 292.5 degrees (-1.075), 315 degrees (-2.014), 337.5 degrees (-2.580) and 0 degrees (-2.764). The strong negative skewness indicates that the participants generally performed well, but the accuracy was particularly high when landform was illuminated by a (virtual) light source at 337.5 degrees and 0 degrees. We deduced confidence data from the

Likert scale. If a participant selected *clearly a ridge/valley* they received a 2 towards their confidence score, 1 for *maybe a ridge/valley* and 0 for *not sure*. This resulted to a confidence score out of 80. The confidence score was also negatively skewed (-1.037) with a kurtosis of 0.091.

3.1. Participants' performance and confidence per light direction

Figure 4 provides an overview of the relationship between the participants' accuracy scores and their average confidence per SRM. Even though overall we see a strong correlation (*Spearman's rho* = 0.278, *p* = .000 2-sided) between the average accuracy score and the average confidence. Figure 4 illustrates that the participants were not necessarily aware which light directions allow them to identify the landforms correctly, or unambiguously (confidence is high even when accuracy rates are low).

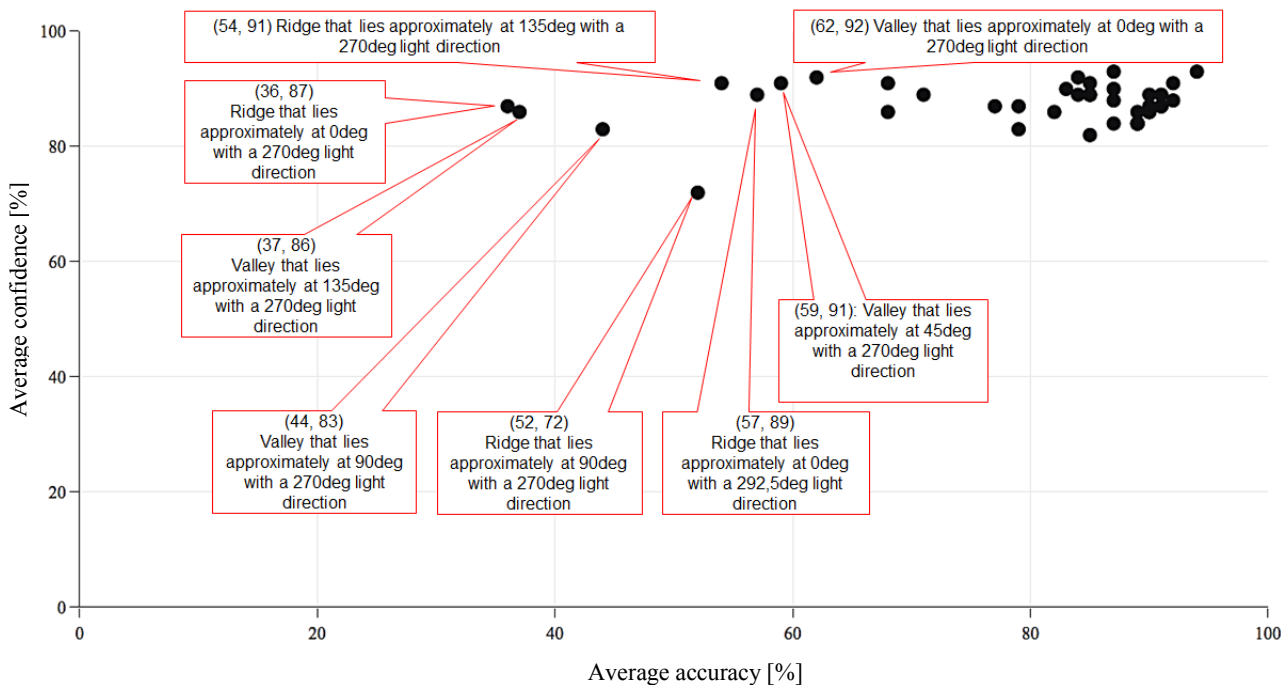


Figure 4. The average accuracy scores of the participants compared to the average confidence score (per question). We annotated the plot to demonstrate *low score – high confidence* pairings. A clear pattern emerges from this clustering: even when the accuracy rates are low, participants are highly confident. Specifically, when the light source is placed at 270°, participants feel confident (~over 70%) even though their accuracy in landform identification is low (~under 60%).

Next, to examine the so-called convexity bias (Hill and Bruce, 1994) when the input is ambiguous, people are more likely to say 'ridge' because of this bias. While Biland and Çöltekin (2016) did not observe a global convexity bias, in a follow-up study by Çöltekin and Biland (2018) a global convexity bias was observed. As mentioned earlier, our data is identical to the first study (Biland and Çöltekin, 2016), thus, we expected to reproduce the 'no effect' in this case. We then examined if participants' overall accuracy scores differ when asked to identify either a valley or ridge. Our analysis revealed that the average accuracy scores were similar, and for both valley and ridge questions, the distribution is negatively skewed (-1.668 for ridges and -1.451 for valleys). As can be seen in Figure 5, descriptive statistics suggest that people are slightly more likely to identify ridges

than valleys, however, this difference is not statistically significant: $t(86) = -0.08046$, $p = 0.726$. This result thus confirms the study we are attempting to replicate (Biland and Çöltekin, 2016). A closer comparative examination to the Çöltekin and Biland (2018) study in which a convexity bias was observed, would be an interesting follow-up.

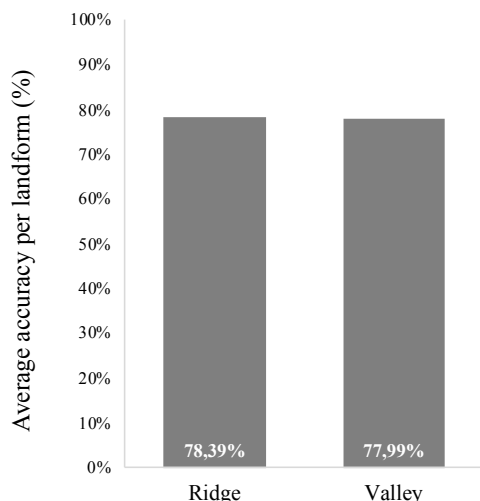


Figure 5. The average accuracy scores for each landform type for all participants presented in percentage. Note that here we did not round the numbers for the percent values presented in the bars (in white ink), if we round them, they would be both at 78%, which demonstrates the values were nearly identical.

3.2. Participants performance and confidence per light direction interval

As mentioned earlier, participants were asked to identify landforms illuminated with light from five distinct directions. Figure 6 provides an overview of the participants' average accuracy and confidence scores for each of light direction (270, 292.5, 315, 337.5 and 0 degrees).

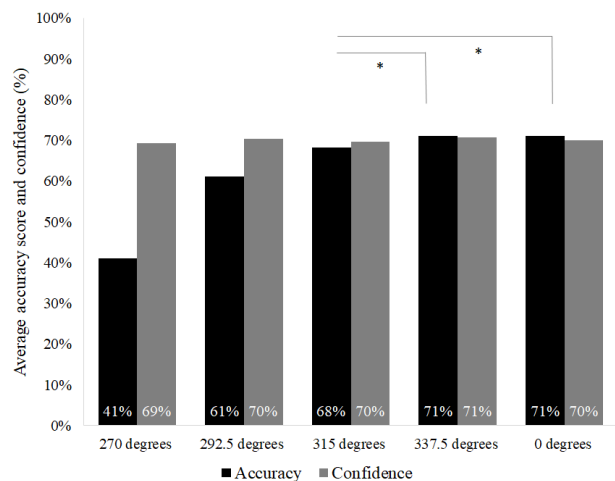


Figure 6. The average accuracy scores for each light direction for all participants. * $p < .05$. Note that here we only present the inferential statistics for the accuracy with NW (315°) against NNW (337.5°) and N (0°), because the other two light directions (270° and 292.5°) clearly impair landform detection accuracy and are not desirable.

We see in Figure 6 that the accuracy gets better as we move to northern illumination angles, overall confirming the previous studies. Furthermore, we analyzed the pairwise differences in participants' accuracy between the conventional 315 degrees and the two other northern illumination angles using t-tests. Similar to the original study's findings, we see that NNW (337.5) and N (0) are better than the conventional NW: $t(86) = -0.28736$, $p = 0.018$ for NNW and $t(86) = -0.28736$, $p = 0.023$ for N.

Different to the original study; we see that the N and NNW yield nearly identical results.

3.3. Exploratory group differences in performance and confidence based on experience and gender

The participants were classified as either *experienced* or *inexperienced* based on their self-evaluation. If a participant rated themselves as *good* or *excellent* (i.e. a 4 or higher) on average on the measures reported in Figure 3, the participant was classified as experienced. This resulted in an imbalanced sample, i.e., we had six experienced participants (3 males and 3 females) and 81 inexperienced participants (41 males and 40 females). Because of this severe imbalance in the sample, we do not report inferential analysis. Nonetheless, as a preliminary indication of possible hypothesis building to explore further later (i.e., for the sake of exploratory analyses), below we report the difference between the two groups. Based on the descriptive statistics, it appears that the inexperienced participants 'outperform' the experienced participants (Figure 7). This is an expected result, as experts tend to 'second guess' when the elements in the scene appear odd (e.g., based on geomorphological reasoning and terrain interpretation).

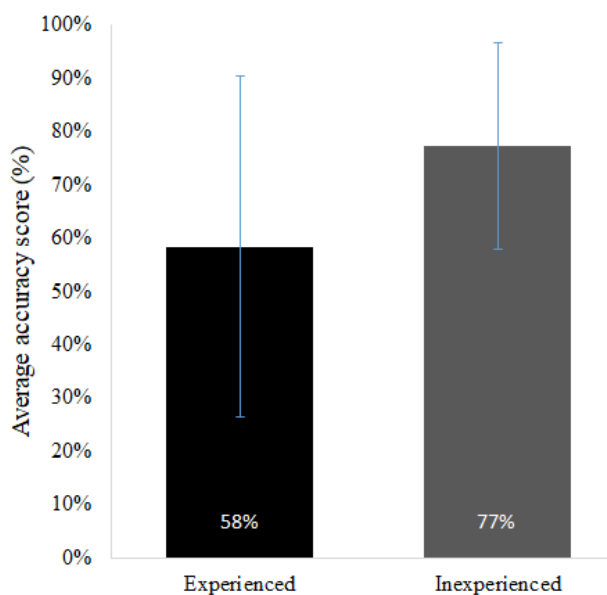


Figure 7. Average accuracy scores (in percentage) of the experienced ($n=6$) and inexperienced ($n=81$) participants for each level. Error bars: \pm SD.

As a part of the exploration, we also checked if there are performance differences based on participants' gender (more precisely, biological sex). Descriptive statistics suggest that female participants (on average) achieve slightly higher scores (m: 74%, SD: 23%; f: 78%, SD: 19%), but men display (ever so slightly) more confidence (m: 86%, SD: 13.5%, f: 85%, SD: 12.6%). The confidence analysis based on gender in Biland and Çöltekin (2016) study yielded a nuanced gender difference that interacted with expertise: inexperienced men displayed overconfidence while the other groups (experienced men, and all women) displayed similar levels of confidence. In this study we cannot revisit this finding first because of the large difference in

the sample size for experienced vs. inexperienced participants, but also because there may be a number of other confounding variables. In other words, the gender analysis for accuracy is most likely contaminated by previous experience, education and professional exposure levels. Similarly confidence can be affected by levels of expertise and exposure to similar situations (e.g., performance anxiety could affect performance as well as confidence negatively, or working with terrain and feeling 'at home' might affect both performance and confidence negatively). A thorough examination of all these factors are left as a future study.

4. DISCUSSION AND CONCLUSIONS

This study was conducted as a replication effort, not because we doubted the design or findings of the original study, but because of three distinct reasons: 1) the implications of changing a convention is large and it should be examined more than once, 2) sampling the participants from a different continent (especially because how we experience light might be different in southern and northern hemispheres), different country, different linguistic and educational characteristics to arrive at similar results would make the findings more plausible, and 3) there is an emerging conversation on how scientific studies do not hold up against replication efforts, and thus replicating an experiment is, in principle, a good practice. Having these three concerns in mind, we believe our replication effort broadly 'paid off', that is, even though the sample was entirely different, and the participants worked only with a subset of the original SRMs, overall tendencies are similar in the experiment conducted in South Africa and in Switzerland. Overall we can say that the original finding was robust and is replicated with another sample.

Even though we interpreted the results somewhat unconventionally in the previous section (Results) to some degree, couple of findings are possibly important to further interpret. First, the fact that participants overall perform quite well is to be expected. In the original study, the entire spectrum of lighting directions were sampled (also eastern and southern light directions, see Biland and Çöltekin, 2016), in this study, because we were interested in whether one should or should not challenge the cartographic convention of using NW (315 degrees north-west) illumination direction, we used this light direction and four others in the north-western quadrant which are reasonable to compare. As in Biland and Çöltekin (2016), we observed that NNW, and also N, are better than NW. In fact, in this experiment, we see that N is quite the same as NW, which might suggest a different 'left bias' (Sun and Perona, 1998). We will examine and compare left bias in northern and southern hemisphere samples in a follow-up comparative study.

As by-products of having measured the landform perception accuracy using equal amounts of valleys and ridges in the SRM set for counterbalancing purposes, we were able to check for what is known as the convexity bias, which was also examined in Biland and Çöltekin's (2016) earlier experiment, and in a follow up study by the same team (Çöltekin and Biland, 2018). The results were replicated as expected: Our findings are similar to Biland and Çöltekin's (2016), that is, we do not observe a convexity bias in this experiment.

Based on the exploratory analyses, we also see that the trends regarding interaction between participant characteristics and their performance as well as confidence measures also confirm the original study's findings. Experience appears to lower both accuracy and confidence, which Biland and Çöltekin (2016)

interpreted as 'overthinking' on the experts' side, and later it was also demonstrated by Çöltekin and Biland (2018) that expertise leads to (most likely implicit) interpretation of the scene, and possibly cognitive processes override perceptual signals. Our findings are in the same direction, although should be interpreted cautiously as the sample sizes of experienced and inexperienced groups are very uneven. Both the expertise and gender-related findings need further scrutinizing for confounding variables in a proper controlled study in the future.

From a human factors perspective, there are a number of problems with 3D displays we must reflect upon (e.g., Çöltekin et al., 2016), and where we rely on shading/shadows as primary depth cues, relief inversion is an interesting cognitive challenge to overcome. While there are still several issues to tackle regarding relief inversion (or terrain reversal) effect, we can now have empirical evidence that the light direction, and thus the positions of the shadows in the scene are critical and must be controlled carefully. In conclusion, we find that the main findings in the original study are replicated, and thus we recommend the light direction to be placed at the NNW rather than NW when creating shaded relief maps.

ACKNOWLEDGEMENTS

We are grateful for Julien Biland for his courtesy in sharing the SRMs that he prepared for his project for our re-use in this experiment. This work was partially supported by the Swiss National Science Foundation (Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung) [grant number 200021_149670] (SNSF project VISDOM).

REFERENCES

- Bernabé Poveda, M.A., Çöltekin, A., 2014. Prevalence of the Terrain Reversal Effect in Satellite Imagery. *International Journal of Digital Earth*, 8, pp. 640–655.
- Bernabé-Poveda, M.A., Sánchez-Ortega, I., Çöltekin, A. 2011. "Techniques for Highlighting Relief on Orthoimagery." *Procedia - Social and Behavioral Sciences* 21: 346–52.
- Biland, J., Çöltekin, A., 2016. An empirical assessment of the impact of the light direction on the relief inversion effect in shaded relief maps: NNW is better than NW. *Cartography and Geographic Information Science*, 44(4), pp. 358–372.
- Çöltekin, A., Biland, J., 2018. Comparing the terrain reversal effect in satellite images and in shaded relief maps: an examination of the effects of color and texture on 3D shape perception from shading. *International Journal of Digital Earth*,
- Çöltekin, A., Lokka, I. E., & Zahner, M. (2016). On the Usability and Usefulness of 3D (Geo)Visualizations - A Focus on Virtual Reality Environments. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Commission*. Prague.
- Hill, H., Bruce, V., 1994. A Comparison between the Hollow-Face and 'Hollow-Potato' Illusions, *Perception*, 23, pp. 1335 – 1337.
- Imhof, E., (1965). Reprinted in 2007. Shading and Shadows. In *Cartographic Relief Presentation*, English language edition edited by H. J. Steward, Redlands, CA: ESRI Press, pp. 159–212.

Kraak, M.J., and Ormeling, F., 2010. *Cartography: Visualization of Spatial Data*. 3rd ed. Harlow, UK: Prentice Hall.

Mamassian, P., and Goutcher, R., 2001. Prior Knowledge on the Illumination Position. *Cognition*, 81(1): B1–9.

Ramachandran, V.S., 1988. Perceiving Shape from Shading. *Scientific American*, 259(2), pp. 76–83.

Saraf, A.K., Das, J.D., Agarwal, B., Sundaram, M., 1996. False topography perception phenomena and its correction. *International Journal of Remote Sensing*, 17(18), pp. 3725–3733.

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. *International Journal of Remote Sensing*, 28(5), pp. 811–821.

Slocum, T.A., McMaster, R.B., Kessler, F.C., Howard, H., 2008. *Thematic Cartography and Geovisualization*. 3rd ed. Upper Saddle River: Prentice Hall.

Sun, J., Perona, P., 1998. Where Is the Sun? *Nature Neuroscience*, 3, pp. 183–84.

Zhang, D., Zhu, W., Chen, X., Zheng, Z., Zhou, X., Jiang, T., 2016. A Correction Technique for False Topographic Perception of Remote-Sensing Images Based on an Inverse Topographic Correction Technique. *International Journal of Digital Earth*, 9(10), pp. 1021–1034