Commonalities and Differences in Eye Movement Behavior When Exploring Aerial and Terrestrial Scenes

Sebastian Pannasch, Jens R. Helmert, Bruce C. Hansen, Adam M. Larson and Lester C. Loschky

Abstract Eye movements can provide fast and precise insights into ongoing mechanisms of attention and information processing. In free exploration of natural scenes, it has repeatedly been shown that fixation durations increase over time, while saccade amplitudes decrease. This gaze behavior has been explained as a shift from ambient (global) to focal (local) processing as a means to efficiently understand different environments. In the current study, we analyzed eye movement behavior during the inspection of terrestrial and aerial views of real-world scene images. Our results show that the ambient to focal strategy is preserved across both perspectives. However, there are several perspective-related differences: For aerial views, the first fixation duration is prolonged, showing immediate processing difficulties. Furthermore, fixation durations and saccade amplitudes are longer throughout the overall time of scene exploration, showing continued difficulties that affect both processing of information and image scanning strategies. The temporal and spatial scanning of aerial views is also less similar between observers than for terrestrial scenes, suggesting an inability to use normal scanning patterns. The observed differences in eye movement behavior when inspecting terrestrial and aerial views suggest an increased processing effort for visual information that deviates from our everyday experiences.

S. Pannasch (✉) · J. R. Helmert
Department of Psychology, Engineering Psychology and Applied Cognitive Research, Technische Universitaet Dresden, Dresden, Germany
e-mail: sebastian.pannasch@tu-dresden.de

B. C. Hansen
Department of Psychology, Neuroscience Program, Colgate University, Hamilton, NY, USA

A. M. Larson
Department of Psychology, University of Findlay, Findlay, OH, USA

L. C. Loschky
Department of Psychology, Kansas State University, Manhattan, KS, USA

M. Buchroithner et al. (eds.), *Cartography from Pole to Pole*,
Lecture Notes in Geoinformation and Cartography, DOI: 10.1007/978-3-642-32618-9_30,

Keywords Eye movements · Fixation duration · Saccade amplitude · Ambient and focal processing · Aerial scene views · Terrestrial scene views

1 Introduction

Under most circumstances, vision is the dominant sensory modality in humans. During visual perception, information is sampled from the environment via *active vision* (Findlay 1998). Saccades—fast ballistic movements—direct the foveal region of the eyes from one fixation point to another. During saccades, the intake and processing of visual information is largely suppressed and is therefore limited to the periods of fixations, when the eyes are relatively still. This interplay of fixations and saccades is essential, as highest visual acuity is limited to the small foveal region. Eye movement behaviour in many everyday situations, such as reading text or inspecting images, can be described as an alternation between fixations and saccades.

Fixation durations vary a great deal from one fixation to the next. It has been suggested that the length of a fixation is determined by information processing and by eye movement pre-programming (Rayner 1998). Fixation durations typically range from roughly 100 to 500 ms, but can last up to 2–3 s in some cases (Rayner 1998). Similarly, the length of saccades generally varies from between less than 1 to 130° of visual angle (Land 2004). Importantly, Velichkovsky et al. (2005) reported particular relationships in the variation of fixation durations and saccade amplitudes that were related to certain modes of visual processing. Specifically, they found fixations of shorter durations (below 180 ms) are often associated with larger saccades; this combination was termed *ambient processing*, which is assumed to serve the processing of spatial orientation and localization. Moreover, the combination of longer fixations and shorter saccade amplitudes was termed the *focal processing mode*, which is assumed to be concerned with the analysis of object features. The time course of these two processing modes has been investigated under different conditions of free viewing, which has revealed a systematic relationship: during early phases of scene inspection, the ambient mode seems to dominate, giving way to the focal processing mode with increased time (Pannasch et al. 2008). This relationship has been observed across different types of stimuli and various visual tasks. However, the corresponding analyses were relatively coarse (typically consisting of the comparison of gaze behavior in two 2-s time periods, i.e. early [0-2 s] versus late [4-6 s]).

The present experiment analyzes the time course of viewing behavior in greater detail by showing real-world scenes from different views and under different display conditions: Photos of natural scenes taken either from the terrestrial or aerial perspective were presented either upright or inverted. The research question was whether we would obtain similar gaze patterns to those reported previously when looking at upright terrestrial views. By contrasting the viewing behavior in

this particular case with images of different perspectives (aerial vs. terrestrial) and orientations (upright vs. inverted), we expected to gain further insights about the interplay between ambient and focal processing modes. Particularly, we know from gist recognition studies that compared to upright terrestrial views, both inverted terrestrial scenes and upright (as well as inverted) aerial scenes are much harder to recognize within the time course of a single fixation, and thus appear to require more than a single fixation to reach a high level of gist recognition (Loschky et al. 2010). Thus we predicted differences very early after the image onset; fixation durations should be shortest for terrestrial upright views. Starting with this hypothesis we examined whether the different viewing conditions would influence only the initial gaze behavior or gaze behavior observed throughout a longer time period of scene inspection.

2 Method

2.1 Subjects

Thirty students (19 females) at Colgate University with a mean age of 18.5 years took part in this experiment. One subject was removed from the sample because their performance in image categorization was at chance level. All subjects had normal or corrected-to-normal vision and were given course credit for their time. The study was conducted in conformity with the declaration of Helsinki, and Institutional Review Board-approved written informed consent was obtained.

2.2 Apparatus

Participants were seated in a dimly illuminated, sound-attenuated room. Eye movements were sampled monocularly at 1 kHz using the SR EyeLink 1000 infrared eye tracking system with on-line detection of saccades and fixations and a mean spatial accuracy of better than 0.5°. Saccades were identified by deflections in eye position in excess of 0.1°, with a minimum velocity of 30° s^{-1} and a minimum acceleration of 8000° s^{-2}, maintained for at least 4 ms. Pictures were displayed using an VidiaQuadro NVS 285 dual DVI/VGA graphics card and a CRT display (21-inch Viewsonic, G225fB). Maximum luminance output of the display monitor was 100 cd/m^2, the frame rate was set to 85 Hz, and the resolution was set to 1024 × 768 pixels. The monitor was viewed at a distance of 91.5 cm. Head position was maintained with an SR Research chin and forehead rest.

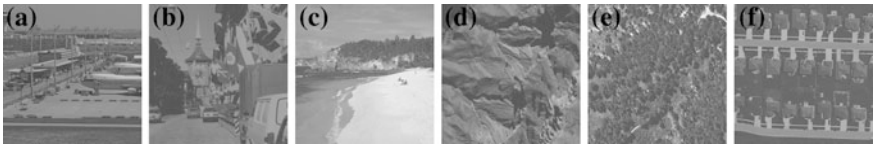

Fig. 1 Examples of the stimuli. Images (**a–c**) represent terrestrial views of the categories airport, city, and coast; images (**d–f**) show aerial views of the categories mountain, forest, and suburbs

2.3 Stimuli

Grayscale images of real-world scenes from six possible categories, viz. airports, cities, coasts, forests, mountains, or suburbs (Loschky et al. 2010), served as stimuli (Fig. 1). Each category was comprised of 60 unique images. Of these, 30 images presented an aerial view (as viewed from a satellite) and the other 30, a terrestrial view. There was no one-to-one correspondence between aerial and terrestrial scene images. Further, each image was presented in an upright as well as inverted manner, resulting in a total of 720 images. In order to use all images of this database, the full set was divided into two groups, each containing 360 images, counterbalanced for categories and views. All images were presented centrally on the screen with a size of 768 × 768 pixels (subtending a visual angle of 18.2° horizontally and vertically).

2.4 Procedure

Subjects were informed that the purpose of the study was to investigate eye movement patterns in the perception of natural scenes and were asked to study the images in order to categorize each scene after the presentation. Subjects were randomly assigned to one of two image groups. To limit the total duration of the experiment, participants of both groups inspected only 240 scenes (40 from each category, 20 aerial and 20 terrestrial views, shown either upright ore inverted), selected from each group's image database (360 scenes). This procedure made sure that within each group the overlap of images between any two subjects was at least 66 % but there was no overlap between the groups. Each image was shown for 6500 ms, followed by a screen showing the six category labels arranged in a 2 × 3 grid. Category labels were randomly shuffled on a trial-by-trial basis, to avoid influences of responding to a preferred grid location (e.g., top left) across trials. Participants had to categorize the previous image by clicking on the appropriate category, and then the next trial started. The full experiment took 60 min in total to complete. An initial 9-point calibration and validation was performed before the start of each block, and calibration was checked prior to every 20th trial in the experiment.

2.5 Data Analysis

Raw eye movement data were preprocessed by removing fixations around eye blinks and outside the presentation screen, resulting in the inclusion of a total of 136, 406 fixations and saccades (86 % of the original data). Because of the positive skew in the distributions of fixation durations and saccade amplitudes, median rather than mean values were used as a measure of central tendency. For statistical testing the respective median values were subjected to repeated measures analyses of variance (ANOVA). All ANOVA factors were within-subjects, unless otherwise stated in the text. *Eta*-squared values are reported as estimates of the effect size (Levine and Hullett 2002).

3 Results and Discussion

3.1 Behavioral Measures

First we examined the percentage of correct responses for the four different types of scene presentations. The correct responses of all subjects were analyzed in a 2 (view: aerial, terrestrial) × 2 (rotation: upright, inverted) repeated measures ANOVA. We obtained a significant main effect for view, $F(1,28) = 43.43$, $p < 0.001$, $\eta^2 = 0.942$, in which viewers were more accurate for terrestrial than for aerial scene views (97.2 vs. 90.6), consistent with Loschky et al. (2010). However, there was no main effect for rotation, $F(1,28) = 1.35$, $p = 0.255$. This was due to the fact that there was a significant view × rotation interaction, $F(1,28) = 6.77$, $p = 0.015$, $\eta^2 = 0.052$, in which viewers showed slightly better classification performance for upright than inverted terrestrial images (97.7 vs. 96.7 %) whereas for aerial views the opposite was found (89.6 vs. 91.6 %). The lack of effect for image inversion on terrestrial scene categorization accuracy was likely due to the long stimulus presentation times of 6500 ms.

3.2 Eye Movement Measures

Our objective was to examine eye movement behavior during the free visual exploration of real-world scenes. Therefore, we systematically varied the scene viewpoint (aerial vs. terrestrial) and rotation (inverted vs. upright). In particular, we were interested in possible influences of these factors on fixation durations and saccade amplitudes at specific points in the time course of scene inspection. Furthermore, we investigated if scene viewpoint and rotation influence scene exploration behavior over time by comparing the eye movement scan paths as a function of these factors.

For the first analysis, fixation durations and saccade amplitudes were entered into two 2 (view: aerial vs. terrestrial) × 2 (rotation: inverted vs. upright) repeated measures ANOVAs, respectively. For fixation durations we observed significant differences for view, $F(1,28) = 5.33$, $p = 0.029$, $\eta^2 = 0.336$, and for rotation, $F(1,28) = 13.4$, $p = 0.001$, $\eta^2 = 0.4$. Fixation durations were longer for aerial views (329 vs. 320 ms) and longer if scenes were shown inverted (331 vs. 317 ms). There was also a significant interaction of view × rotation, $F(1,28) = 15$, $p < 0.001$, $\eta^2 = 0.264$, which was based on shorter fixation durations for terrestrial views if shown upright (309 vs. 326 ms), whereas no such difference was found for aerial views (both 331 ms). For saccade amplitudes we found significant differences for the influence of view, $F(1,28) = 22.21$, $p < 0.001$, $\eta^2 = 0.97$, but not for rotation, $F(1,28) = 1.14$, $p = 0.295$. Longer saccades were made during the exploration of aerial images (4.8 vs. 4.5 deg). No significant interaction was found, $F < 1$. Thus, the overall comparisons of fixation durations and saccade amplitudes as a function of view and scene orientation suggest that there are differences in gaze behavior depending on the type of scene people are looking at.

In the second analysis, we focused on particular periods of scene inspection to explore in detail the nature of the global differences reported above. First, we were interested in the earliest influences of the stimulus material, namely possible differences in gaze behavior immediately following the scene onset, which can be related to the processing of gist (i.e., the understanding of the overall meaning of a scene, see e.g., Oliva and Torralba 2006; Greene and Oliva 2009; Loschky and Larson 2010). Thus, we took a closer look at the fixations that started before the scene onset began, and ended after the scene onset. Those fixations were divided into the time before and after the scene onset. We predicted similar fixation durations for the time before the scene onset but expected differences for the fixation durations after the scene onset, if the content of a scene has an early influence on gaze behavior. The 2 (view: aerial vs. terrestrial) × 2 (rotation: inverted vs. upright) repeated measures ANOVA on the fixation times before the image onset revealed no differences, $F < 1$, while the same analysis on the remaining fixation time after the image onset demonstrated significant differences for view, $F(1,28) = 52.7$, $p < 0.001$, $\eta^2 = 0.9$, but not for rotation, $F(1,28) = 1.97$, $p = 0.171$. No significant interaction was found, $F(1,28) = 3.01$, $p = 0.093$. Thus, the remaining fixation time after the scene onset was larger for aerial than terrestrial scenes (239 vs. 214 ms). Finally, we analyzed the saccade amplitudes following the picture onset but found no differences with regard to the scene viewpoint, all $F < 1$.

In our final analysis, we were interested in the eye movement behavior (i.e., fixation durations and saccade amplitudes) over the time course of scene exploration. For this analysis, we divided the 6.5 s image presentation into three time periods: 0–2 s, >2–4.5 s, and >4.5–6.5 s. The first time period was from immediately following the scene onset to 2.0 s after onset. During this period, we assumed a more global scanning strategy (Antes 1974) which has been described in terms of ambient processing (Unema et al. 2005; Pannasch et al. 2008).

Consistent with previous studies, this interval included only eye movements up to 2 s after image onset (for this analysis we excluded the eye fixation spanning the scene onset and started with the first fixation after the image onset). The second time period was between these initial 2 s of scene exploration and up to 4.5 s, during which time we assumed there would be a transition from ambient to focal processing. The third time period was for the final 2 s of scene exploration, during which time we predicted a larger proportion of focal processing being more related to the extraction of details and particular object features.

Fixation Durations: We conducted a 3 (time period: first, second, third) × 2 (view: terrestrial, aerial) × 2 (rotation: upright, inverted) repeated measures ANOVA. For fixation durations we obtained significant main effects for time period, $F(2,56) = 46.9$, $p < 0.001$, $\eta^2 = 0.824$, view, $F(1,28) = 12$, $p = 0.001$, $\eta^2 = 0.082$, and rotation, $F(1,28) = 16.5$, $p = 0.005$, $\eta^2 = 0.045$. Furthermore, we observed a significant interaction for view × rotation, $F(2,56) = 12.1$, $p = 0.019$. No further interactions were found. Regarding the main effect for time period, post hoc testing revealed a significant increase in fixation durations across the time periods, from the first to the third (254 vs. 269 vs. 288 ms), all $p < 0.001$. Fixation durations were shorter for upright terrestrial scenes (260 ms) than for inverted terrestrial (272 ms) or either rotation of aerial views (both 274 ms; see Fig. 2, left panel).

Saccade Amplitudes: We conducted the same 3 (time period: first, second, third) × 2 (view: terrestrial, aerial) × 2 (rotation: upright, inverted) repeated measures ANOVA for saccade amplitudes and obtained significant main effects for time period, $F(2,56) = 69.4$, $p < 0.001$, $\eta^2 = 0.854$, and view, $F(1,28) = 13.5$, $p < 0.001$, $\eta^2 = 0.124$, but not for rotation, $F < 1$. No further interactions were found. The main effect for time period was due to the significant decrease of

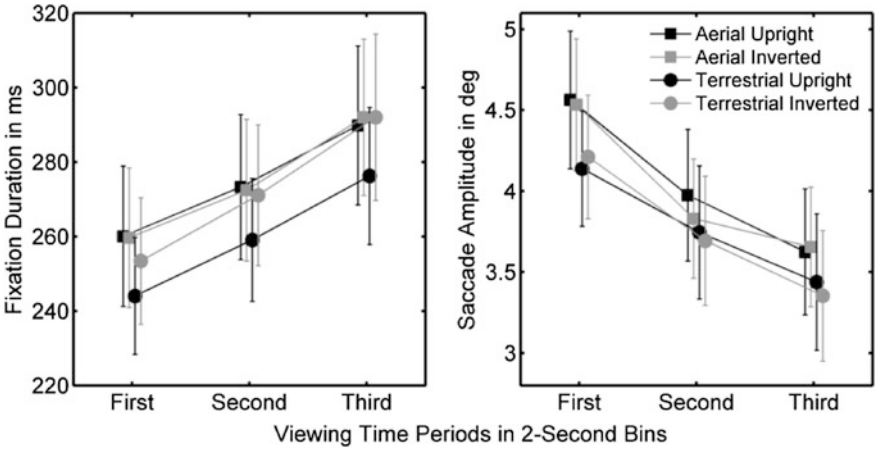

Fig. 2 Mean fixation durations (*left*) and mean saccade amplitudes (*right*) for view and rotation in the respective viewing time periods. The error bars represent 95 % confidence intervals

saccade length from the first to the last time period (4.36 vs. 3.81 vs. 3.52°), all $p < 0.001$. The second main effect was the result of longer saccades for aerial views (4.03 vs. 3.76°; see Fig. 2, right panel).

Finally, we examined the similarity of the fixation strategies for each view and rotation. Therefore, we used the ScanMatch algorithm (Cristino et al. 2010) which is based on the Needleman–Wunsch global sequence alignment algorithm (Needleman and Wunsch 1970) and creates letter sequences retaining fixation location, time and order information for each individual scene inspection. Pairs of sequences are then compared to find the optimal alignment between a pair. The similarity between two sequences is expressed by a normalized ScanMatch score, with similarity magnitude indicated by ScanMatch score distance from 0. The obtained similarity scores were averaged for each image and applied to a 2 (view: aerial vs. terrestrial) × 2 (rotation: inverted vs. upright) repeated measures ANOVA. A significant main effect was obtained for view, $F(1,179) = 26.1$, $p < 0.001$, $\eta^2 = 0.909$, but not for rotation, $F < 1$. No further interaction was found, $F(1,179) = 2.2$, $p = 0.14$. The main effect for view was due to viewers showing less similar fixation strategies for aerial views (0.548 vs. 0.569).

4 Conclusion

Our results revealed three major findings. First, fixation durations and saccadic amplitudes differed according to the respective view and orientation of the scenes. Fixations were longest for aerial views (both upright and inverted) and shortest for upright terrestrial views, suggesting greater difficulty in processing aerial than terrestrial scene images. Saccade amplitudes were larger for aerial than for terrestrial views, whether upright or inverted. Based on this observation, it would therefore seem that identifying aerial views requires the inspection of a larger proportion of image space. However, since we did not analyze the fixation locations in the image, the larger amplitudes could also result from repeated refixations of peripheral image regions. Thus, future work should incorporate the examination of fixation locations according to different fixation sequences within aerial versus terrestrial views (see, e.g., Çöltekin et al. 2010).

Second, the general gaze patterns over the time course of scene inspection followed earlier observations, namely there was a transition from early ambient to later focal processing, indicated by the increase in fixation durations and decrease in saccade amplitudes. However, fixation durations and saccade amplitudes are longer if the explored scene deviates from our everyday experience, namely terrestrial upright scenes. The longer fixation durations most likely indicate increased difficulty in processing and interpreting the visual information in the scene while the longer saccade amplitudes suggest intensified search for useful information in the scene. Future work should apply more advanced analytical methods, such as gaze map matching (Kiefer and Giannopoulos 2012) in order to provide further insight into these issues.

Finally, we analyzed additional parameters to obtain a better understanding of viewers' ongoing processing. For example, consistent with our predictions based on gist recognition for aerial scenes, we found longer first fixation durations (i.e., the remaining time of a fixation immediately after the picture onset) for both types of aerial views. This suggests that the difficulty in interpreting aerial scene images begins on the very first eye fixation on the image, which then influences further eye movements and visual processing while looking at the image. Furthermore, we determined the similarity of scan paths for each image across participants. Comparing the similarity indices for the different conditions revealed higher similarity between participants viewing terrestrial views compared to the exploration of aerial views. This may be due to viewers being better able to utilize normal viewing routines while looking at more familiar types of scenes, namely those taken from terrestrial views.

To summarize, our results reveal interesting differences in viewers' gaze patterns when inspecting the same information from our environment but presented from aerial versus terrestrial perspectives. In the case of aerial views, we found in various gaze parameters that this type of scene view has a clear influence on the balance of ambient and focal processing. While the general time course of ambient to focal processing remains stable, the proportion of focal processing is reduced and a greater dominance of the ambient mode is observed. The present study therefore provides insight into the general processing mechanisms involved in viewing complex imagery from different viewpoints. Future work is needed in order to assess how these basic mechanisms change as a function of expertise (e.g. Lloyd et al. 2002; Ooms et al. 2011).

References

Antes JR (1974) The time course of picture viewing. J Exp Psychol 103:62–70

Çöltekin A, Fabrikant SI, Lacayo M (2010) Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. Int J Geogr Inf Sci 24:1559–1575

Cristino F, Mathot S, Theeuwes J, Gilchrist ID (2010) ScanMatch: a novel method for comparing fixation sequences. Behav Res Methods 42:692–700

Findlay JM (1998) Active vision: visual activity in everyday life. Curr Biol 8:R640–R642

Greene MR, Oliva A (2009) The briefest of glances: the time course of natural scene understanding (Research Article). Psychol Sci 20:464–472

Kiefer P, Giannopoulos I (2012) Gaze map matching: mapping eye tracking data to geographic vector features. In: Proceedings of the 20th international conference on advances in geographic information systems, ACM, Redondo Beach, California, pp 359–368

Land MF (2004) The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. Exp Brain Res 159:151–160

Levine TR, Hullett C (2002) Eta-square, partial eta-square, and misreporting of effect size in communication research. Hum Commun Res 28:612–625

Lloyd R, Hodgson ME, Stokes A (2002) Visual categorization with aerial photographs. Ann Assoc Am Geogr 92:241–266

Loschky LC, Larson AM (2010) The natural/man-made distinction is made before basic-level distinctions in scene gist processing. Vis Cogn 18:513–536

Loschky LC, Ellis K, Sears T, Ringer R, Davis J (2010) Broadening the horizons of scene gist recognition: aerial and ground-based views. J Vis 10:1238

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Oliva A, Torralba A (2006) Building the gist of a scene: the role of global image features in recognition. Prog Brain Res 155:23–36

Ooms K, De Maeyer P, Fack V (2011) Can experts interpret a map's content more efficiently? In: Proceedings of the 25th international cartographic conference (ICC 2011), Paris, France, 3–8 July 2011

Pannasch S, Helmert JR, Roth K, Herbold A-K, Walter H (2008) Visual fixation durations and saccadic amplitudes: Shifting relationship in a variety of conditions. J Eye Mov Res 2(4):1–19

Rayner K (1998) Eye movements in reading and information processing: 20 years of research. Psychol Bull 124:372–422

Unema PJA, Pannasch S, Joos M, Velichkovsky BM (2005) Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration. Vis Cogn 12:473–494

Velichkovsky BM, Joos M, Helmert JR, Pannasch S (2005) Two visual systems and their eye movements: evidence from static and dynamic scene perception. In: Bara BG, Barsalou L, Bucciarelli M (eds) Proceedings of the 27th conference of the cognitive science society. Lawrence Erlbaum, Mahwah, pp 2283–2288