Vision Research 50 (2010) 2062–2068

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The contribution of scene context on change detection performance

Eckart Zimmermann *, Fabian Schnier, Markus Lappe

Psychological Institute II and Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, Westf. Wilhelms-Universität, Fliednerstrasse 21, 48149 Münster, Germany

ARTICLE INFO

Article history:
Received 1 July 2009
Received in revised form 4 April 2010

Keywords:
Change detection
Scene context
Natural scenes
Visual representation
Short-term memory

ABSTRACT

The gist of a visual scene is perceived in a fraction of a second but in change detection tasks subjects typically need several seconds to find the changing object in a visual scene. Here, we report influences of scene context on change detection performance. Scene context manipulations consisted of scene inversion, scene jumbling, where the images were cut into 24 pieces and randomly recombined, and scene configuration scrambling, where the arrangement of the objects in the scene was randomized. Reaction times, where significantly lower in images with normal scene context. We conclude that scene context structures scene perception.

1. Introduction

Our visual representation is fragmentary and unstable. The capacity of the visual representation can for instance be tested in change detection experiments, where large changes remain undetected when their motion transient is masked (Grimes, 1996). Performance is poor even in the flicker paradigm, where the two pictures are presented rapidly in alteration, separated by a brief blank (Rensink, O'Regan, & Clark, 1997). Changes go unnoticed although they are potentially visible all the time. However, in real-life situations this almost never poses a problem. Only in laboratory situations one can experience these limits of the visual representation. Therefore, other factors must compensate the inefficiency of the visual short-term memory. It has been claimed that we use the external world as an outside memory (O'Regan, 1992) or we take advantage of an long-term memory representation (Hollingworth & Henderson, 2002; Rensink, O'Regan, & Clark, 2000). In both cases it would be the context of a scene which stands in proxy for the fragmentary visual representation. Images of natural real-world scenes share specific structural elements: for instance the sky is always in the upper part of the image and streets are in the lower part of the image. Scene context is constituted by the global spatial configuration of objects in the scene, e.g. whether it is a traffic scene or a living room, etc. The way objects occur in natural scenes is governed by specific probabilities: the position, the orientation and the covariance with other objects is determined by the context of the scene. Why should scene context be beneficial for change detection? In order to detect the change the observer is faced with two tasks: searching many parts of the image as fast as possible and thereby storing visual information during the blank period. Change detection in the flicker paradigm is essentially a combination of a visual search with a short-term memory task. Scene context influences on visual search have been tested for abstract (Chun, 2000) and for naturalistic scenes (Oliva & Torralba, 2007). For naturalistic scenes one might assume that because visual salient regions attract attention changes in these regions should be detected faster than changes in non-salient regions. Indeed, abrupt onsets of attentional cues facilitate change detection even if these cues are unrelated to the change (Scholl, 2000). However, no difference in reaction times were found for the detection of changing objects in regions with high and low visual saliency (Stirk & Underwood, 2007). High-level image characteristics have a strong influence on reaction times. Changes in central regions of interest are detected faster than changes in marginal regions of interest (Rensink et al., 1997). Scene context influences on short-term memory have been tested by Vidal, Gauchou, Tallon-Baudry, and O'Regan (2005). They asked subjects to decide whether a certain object in the display has changed after the presentation of a brief blank. Depending on the stimulus set size (between 2 and 8 objects) subjects were able to answer correctly. If however a non-target object changed additionally to the target object change detection performance was impaired significantly. The authors claim that in visual short-term memory items are stored in a structural gist with relational information about the whole stimulus configuration. Short-term memory thus seems to depend on the object configuration in the image. Velisavljevic and Elder (2008) directly tested the influence of object configuration in naturalistic scenes on the recognition of scene parts. They presented

* Corresponding author. Address: Psychological Institute II, Westf. Wilhelms-Universität, Fliednerstrasse 21, 48149 Münster, Germany. Fax: +49 251 83 34173.
E-mail address: eckartzi@gmail.com (E. Zimmermann).

0042-6989/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2010.07.019

either normal images or jumbled images. The jumbling consists in dividing the normal images into many segments whose positions were then jumbled. Test images were presented for 70.6 ms, followed by a mask for 517 ms. Then, two probe segments were shown of which one was contained in the test image and the other not. The subject had to decide which of the probe segments was in the image. Recognition performance was higher for images with a coherent global scene structure than for jumbled images. Indeed, the perception of scene context survives a saccadic eye movement or an image jump and helps to compare two versions of an image (Bridgeman & Tseng, 2008).

How does the disruption of scene context influence change detection performance? Picture inversion has been used to reduce scene context. (Kelley, Chun, & Chua, 2003; Shore & Klein, 2000). Kelley et al. (2003) used inverted pictures with two changes, one change in the central and one in the marginal region of interest. Subjects should indicate when they detected one of them. Reaction times were not different for inverted compared to upright pictures. In the upright condition subjects more frequently detected the change in the central than in the marginal region of interest. This influence of scene context was reduced in the inverted condition. In another attempt to distort scene context Yokosawa and Mitsumatsu (2003) divided pictures in up to 24 pieces whose positions, where then jumbled. This jumbling did not affect reaction times. Using the one-shot paradigm, where each, the original and the modified picture respectively, is only presented once, Varakin and Levin (2008) found a higher percentage of change detection for normal compared to jumbled images. However, no difference in performance was found between normal and inverted images. The authors suggest that the disruption of surface and object continuity inherent to jumbling is responsible for reduced change detection. We wanted to re-evaluate whether scene context can support our fragmentary visual representation. We chose a change detection task in the flicker paradigm and presented images with natural scene context and scenes with unfamiliar scene context. We hypothesized that familiar scene context facilitates change detection performance and expected differences in reaction times for both images types.

2. Experiment 1

Experiment 1 tested the effect of scene inversion and scene jumbling on change detection performance in the flicker paradigm.

2.1. Apparatus and stimuli

Stimuli were presented on a 22" monitor (20" visible screen diagonal) Iiyama Vision Master Pro 514 with a vertical frequency of 200 Hz at a resolution of 800 × 600 (40° × 30°). All stimuli showed normal real-world scenes consisting of buildings or vehicles (see Fig. 1A). For the estimation of scene interest five observers who did not participate in any of the experiments of this study described each scene after a presentation of 1°s. Regions of central interest were defined as those scene parts that contained objects mentioned by three or more observers (Rensink et al., 1997). Each image was manipulated to obtain a present/absent change of an object inside the central region of interest. The changing object was chosen such that it would not cross any border in the subsequent jumbling of the images in the jumbled condition. The scene jumbling was done by cutting each picture into 24 pieces and recombining these pieces in a random order. The jumbling was done once and the jumbled images were then used in all further experiments. Since the scene jumbling procedure creates spurious edge information along the boundary of the 24 image pieces a grid of black lines was added over the boundary regions. In a control

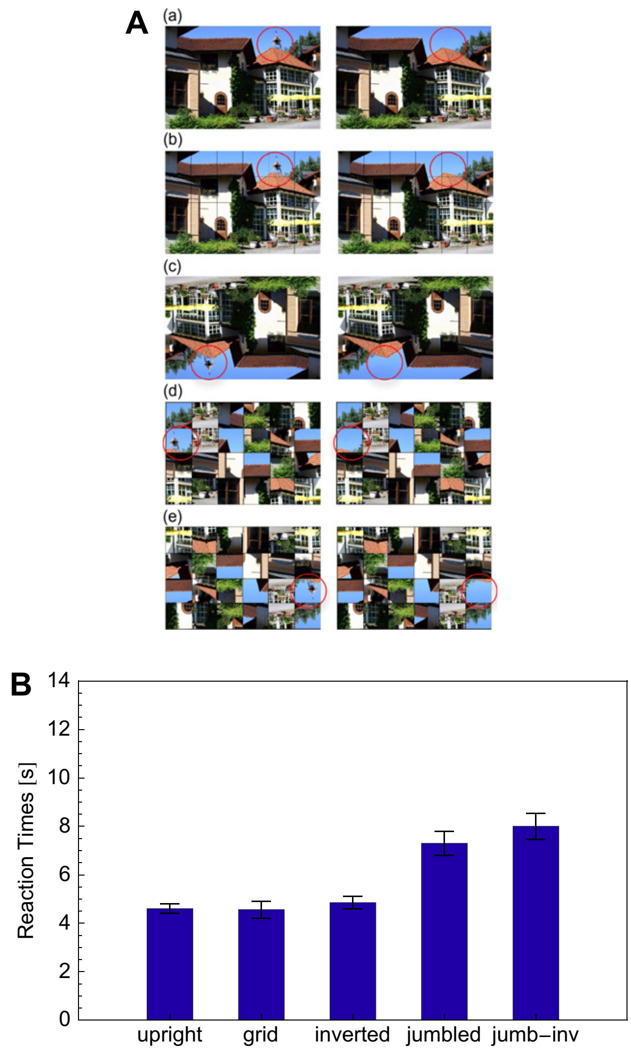

Fig. 1. (A) Example of the stimuli used in Experiment 1. Red circles indicate the changing object. The five conditions are: (a) upright stimuli, (b) upright-grid, (c) inverted stimuli, (d) jumbled stimuli, and (e) jumbled-inverted stimuli. (B) Mean reaction times for all conditions of Experiment 1. Error bars are standard errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

condition the same grid was also layed over the normal, not-jumbled images to control for any effects of this grid. In summary, the images were presented in two experiment series with three conditions each: series 1: upright, inverted, and jumbled and series 2: upright, upright-grid, and jumbled-inverted.

2.2. Procedure

The subject was seated 57 cm in front of the screen with the head stabilized by a chin rest. The room was dimly illuminated. Prior to the experiment each subject saw a change detection demonstration to become familiar with the task. The subject started the first trial by a mouse click. The original (target present) and the modified image (target absent) were presented in a loop for 240 ms each with a grey blank of 80 ms in between. The task was to detect the change as fast as possible and to press the space bar on the keyboard immediately. Then the mouse pointer appeared and the subject had to indicate the position, where the change took place. All trials in which the position of the mouse click had a smaller distance than 2.5° to the center of the changing object were counted as correct. The proportion of trials in which the distance was larger than 2.5° was below 2.5% in all experiments. When the change was not detected after 1 min the next trial started automatically. These trials were counted as having a reaction time of 1 min. Images were presented in the conditions upright, inverted and jumbled. The three conditions were counterbalanced across subjects, such that the pictures that were upright for one group were inverted for the second and jumbled for the third. Each subject saw 72 images randomly presented, 24 of each of the three conditions. If scene context supports change detection performance reaction times should be lower in the upright compared to the jumbled condition. Scene inversion however should not increase reaction times since the object relation remain consistent in this condition.

2.3. Subjects

Twenty-one subjects (8 males, 13 females; mean age: 28 years) with normal or corrected-to-normal vision participated. All subjects were students of the Department of Psychology and gave informed consent.

2.4. Results and discussion

For data analysis, first the median reaction time for each subject was calculated. Then, the mean over subjects was taken in each condition. Fig. 1B shows the resulting mean reaction times. The mean reaction time for the upright condition was calculated from the combined data from both experiment series. A repeated measures oneway ANOVA revealed significant differences in the reaction times ($F(2, 60) = 30.27$; $p = 0.001$) of experiment series 1. Post-tests showed that there were significantly higher reaction times in the jumbled condition compared to the upright ($p = 0.001$) and the inverted condition ($p = 0.001$). For experiment series 2 a oneway repeated measures ANOVA confirmed significant differences in the reaction times ($F(2, 60) = 33.14$; $p = 0.001$). Post-tests showed that there were significantly higher reaction times in the jumbled-inverted condition compared to the grid ($p = 0.001$) and the inverted condition ($p = 0.001$). The accuracy of change detection did not differ significantly across conditions of experiment series 1 (oneway repeated measures ANOVA, $F(2, 60) = 0.52$, ($p = 0.623$)). Mean error was 2.47% of images in the upright condition, 1.87% of images in the inverted condition and 2.48% in the jumbled condition. Again, no significant differences in accuracy were found in experiment series 2 (oneway repeated measures ANOVA, $F(2, 60) = 5{,}75$ ($p = 0.281$)). Mean error was 2.26% of images in the upright condition, 0.41% of images in the grid condition and 2.36% of images in the jumbled-inverted condition.

The reaction times in the inverted condition are not different from those of the upright condition. This result is similar to the study of Kelley et al. (2003). However the difference in reaction times between the jumbled and the upright condition is different in contrast to Yokosawa and Mitsumatsu (2003). They obtained reaction times of ca 28 s in the upright as well as in the jumbled condition. This is unusual because reaction times for upright pictures with a changing object in the central region of interest normally range between 6 and 8 s (Rensink et al., 1997). Our reaction times in the upright condition were 4.5 s. The long reaction times of Yokosawa and Mitsumatsu (2003) suggest that their changes were overall much more difficult to detect than ours and those of Rensink et al. (1997). The statistical power of their method therefore was too small to find any effect of scene context on reaction times. A further reason for the long reaction times could be the longer blank duration of 250 ms which Yokosawa and Mitsumatsu (2003) used. However, although change detection performance increases with longer blank time this increase is rather small (Rensink, O'Regan, & Clark, 2000) and cannot fully explain the long reaction times of Yokosawa and Mitsumatsu (2003).

3. Experiment 2

In Experiment 1 the changing objects were always placed in regions of central interests, which were determined for the upright pictures. It has been demonstrated that change detection is superior for changes in central regions of interest (Rensink et al., 1997). The fast deployment of attention to gist-related parts of a scene (Torralba, Oliva, Castelhano, & Henderson, 2006) might explain this difference. Jumbled scenes do not contain a gist comparable to natural scenes. In Experiment 2 we tested whether the change detection advantage remains when changes are placed in regions of marginal interest. Here we used the same pictures, but now we positioned the changing objects outside the central region of interest. The image material was the same as in Experiment 1 with the exception that the changes were now applied to parts of the image that were not in the central region of interest.

3.1. Methods

Twenty-one subjects (10 males, 11 females; mean age: 28 years) with normal or corrected-to-normal vision participated. All subjects gave informed consent. The procedure was the same as in Experiment 1. The upright with grid, inverted, and jumbled conditions were used.

3.2. Results and discussion

Mean reaction times are shown in Fig. 2. Oneway repeated measures ANOVA revealed significant differences in the reaction times ($F(2, 60) = 4.32$; $p = 0.048$). Bonferroni post-tests showed that there were significantly higher reaction times in the jumbled condition compared to the upright condition ($p = 0.039$). Change detection accuracy did not differ significantly across conditions (oneway repeated measures ANOVA, $F(2, 60) = 0.39$; $p = 0.18$). Mean error was 1.98% of images in the grid condition, 0.84% in the inverted condition and 1.3% in the jumbled condition. All reaction times were higher in images with changes in marginal regions of interest. Although the difference between the normal and jumbled condition is smaller than in Experiment 1 change detection performance is superior even when changes are placed in marginal regions of interest.

4. Experiment 3

The jumbling in Experiment 1 and Experiment 2 cut across the objects of the images and disrupted surface continuities. Therefore,

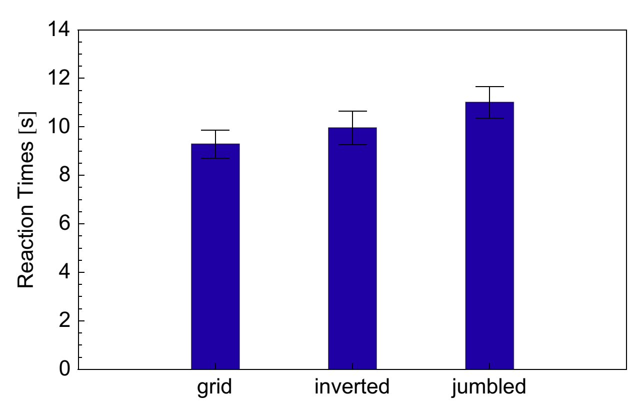

Fig. 2. Mean reaction times for all conditions of Experiment 2. Error bars are standard errors.

the jumbled images contained more surface fragments than the upright and inverted images. The segmentation of the image could thus take more time, which makes the search for the target object harder. The higher reaction times in the jumbled and the jumbled-inverted condition then would not be related to the distortion of the spatial object configuration but rather to the different image layout. In Experiment 3 we tested this hypothesis with abstract art images (from painters Kandinsky, Miro and Delaunay) which did not resemble real-world scenes (see Fig. 3A). We chose these images since they have a context which differs fundamentally from that of natural scenes. The stimuli were selected to consist of many abstract objects which were not semantically arranged. In these images, the scene was an accumulation of geometric figures, like circles, rectangles or triangles. We reasoned that jumbling would induce surface and object disruption, much as in the natural scenes of Experiment 1 or 2. Therefore, if this disruption is responsible for longer search times, search duration should be higher in the jumbled than in the normal condition. If on the other hand the spatial object configuration is important, search times should be similar in both conditions since neither condition shows much natural context.

4.1. Methods

Eighteen subjects (12 males, 6 females; mean age: 28 years) with normal or corrected-to-normal vision participated. All subjects gave informed consent. The stimuli consisted in abstract art images. A total of 48 images was displayed in two conditions: grid and jumbled, with 24 images in each condition.

4.2. Results and discussion

Mean reaction times are shown in Fig. 3B. A paired t-test revealed no significant difference between the conditions grid and jumbled ($p = 0.56$). The abstract art images were chosen to contain geometric objects without any scene context. Accuracy of change detection was 2% of images in the grid condition and 1.4% in the jumbled condition and did not differ significantly from each other (paired t-test, $p = 0.13$). If segmentation or disruption of object continuity would be the explanation for the higher reaction times in the jumbled condition of Experiment 1 and Experiment 2, one would expect a difference in reaction times also in Experiment 3. This was not found.

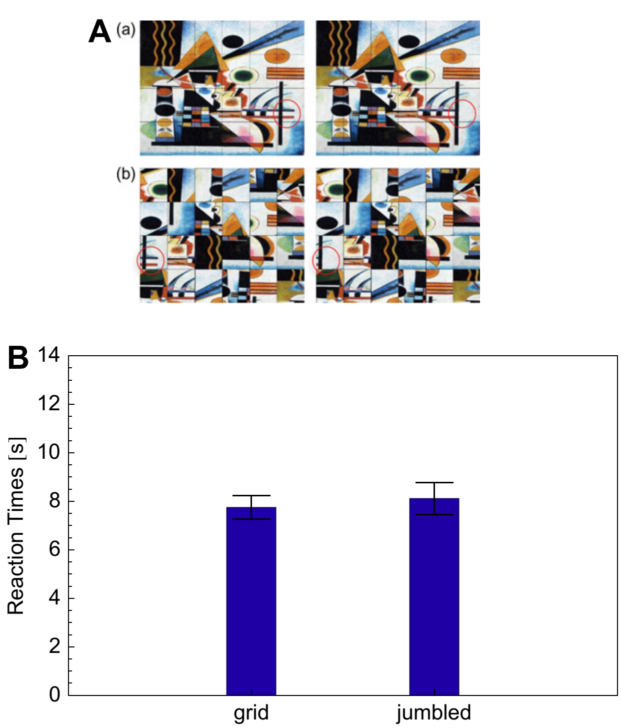

Fig. 3. (A) Example of the stimuli used in Experiment 3: (a) grid stimuli and (b) jumbled stimuli. Red circles indicate the changing object. (B) Mean reaction times for all conditions of Experiment 3. Error bars are standard errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Experiment 4

Experiment 3 showed that surface disruption by jumbling has no detrimental effects on change detection in scenes without natural scene context. Experiment 4 uses scenes with natural scene context and scenes with disturbed object configurations without disrupting surfaces. This disturbance was done by scrambling the spatial arrangement of the objects in the scene (see Fig. 4A). The stimuli consisted of images of furnished rooms which were constructed from the google sketchup library (http://sketchup.google.com). They were chosen in such a way, that one room contained at least 15 objects. This was done to ensure that the task would not become too easy. Rooms for example were kitchens, living rooms, bed rooms, etc. which contained typical furniture as objects. The disturbance of the contextual semantic arrangement of the furniture in the room was done by scrambling the positions and orientations of the furniture. For example, a chair could be placed at the ceiling and a ceiling lamp could be placed on the floor. Scrambling was done by reordering the furniture in such

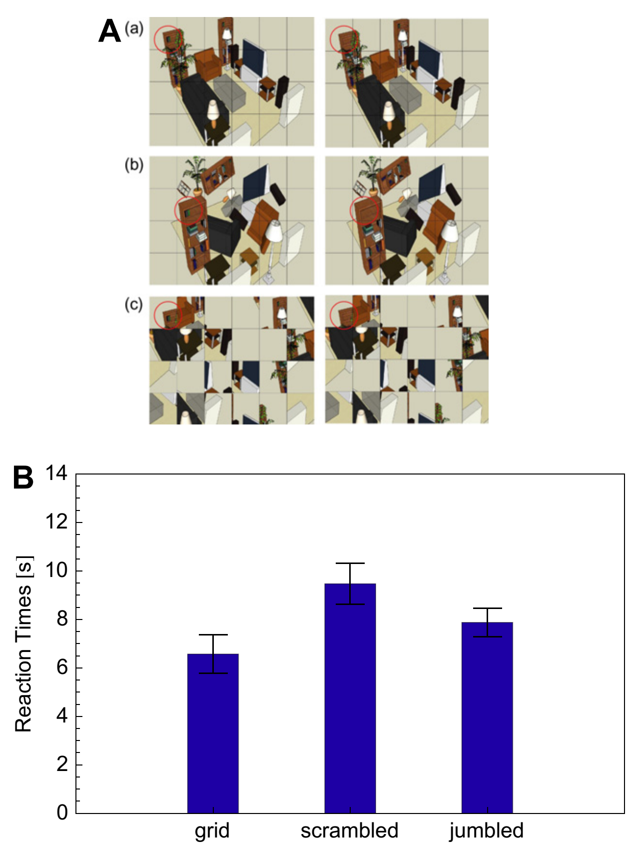

Fig. 4. (A) Example of the stimuli used in Experiment 4: (a) upright stimuli, (b) scrambled stimuli, and (c) jumbled stimuli. Red circles indicate the changing object. (B) Mean reaction times for all conditions of Experiment 4. Error bars are standard errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a way that object size and segmentation would be left intact. Therefore the only aspect, which was altered was the configuration of the objects in the room. Image jumbling served as a second control condition besides the normal arrangement. A grid was used in all of the three conditions. If a spatial object configuration that conforms to the context of a scene is important for quick change detection, reaction times in the scrambled conditions should be higher than in the normal condition. If surface continuity is important for quick change detection, reaction times in the scrambled conditions should be lower than in the jumbled condition.

5.1. Methods

Twenty-seven subjects (10 males, 11 females; mean age: 26 years) with normal or corrected-to-normal vision participated. All subjects gave informed consent. The stimuli consisted of graphic images. The images were displayed in three conditions: normal with a grid, configuration-scrambled and jumbled in 24 pieces. A total of 30 images was presented, 10 in each condition.

5.2. Results and discussion

Mean reaction times are shown in Fig. 4B. Oneway repeated measures ANOVA revealed significant differences in the reaction times ($F(2, 78) = 1.78$; $p = 0.001$). Bonferroni post-tests showed that there were significantly higher reaction times in the configuration-scrambled compared to the grid condition ($p = 0.001$) and the configuration-scrambled condition compared to the jumbled condition ($p = 0.003$). Change detection accuracy was similar in all conditions. A oneway repeated measures ANOVA did not reveal any significant differences ($F(2, 78) = 1.34$, $p = 0.169$). Mean error was 1.32% of images in the grid condition, 0.82% in the jumbled condition and 1.89% in the scrambled condition. The results of Experiment 4 demonstrate that the distortion of the spatial configuration of objects in the scene deteriorates change detection performance. Reaction times in the configuration-scrambled condition were higher than in the other conditions. Since the scrambling procedure does not modify the semantic aspect of scene context these images evoke specific expectations about the locations of objects in the scene. However, the configuration of objects in the scene is completely distorted such that expectations about the object locations would misguide the search for the target objects. This could explain why reaction times in the configuration-scrambled condition were even higher than in the jumbled condition. We thus conclude that scene context guides the search for the target object in the change detection flicker task.

6. Experiment 5

Our jumbling and scrambling methods used in Experiments 1–4 increased the amount of visual clutter in the images. This might explain differences in reaction times since visual clutter is known to influence reaction times in visual search tasks (Rosenholtz, Li, & Nakano, 2007), global search efficiency and eye movement behavior (Henderson, Chanceaux, & Smith, 2009). Rosenholtz et al. (2007) defined visual clutter in three respects. First, as feature congestion, which is the local variability in specific image features. Second, as sub-band entropy, which is a measures of how efficiently an image can be encoded. Efficiency will be higher the more redundant an image is. Third, as edge density, which reflects the number of edges in an image. Our jumbling and scrambling methods used in Experiments 1–4 modify all three aspects of visual clutter. In Experiment 5 we controlled the amount of visual clutter. The stimuli used in Experiment 5 are self-drawn pictures showing normal scenes (see Fig. 5A). The images consisted of 3 × 3 pieces. First,

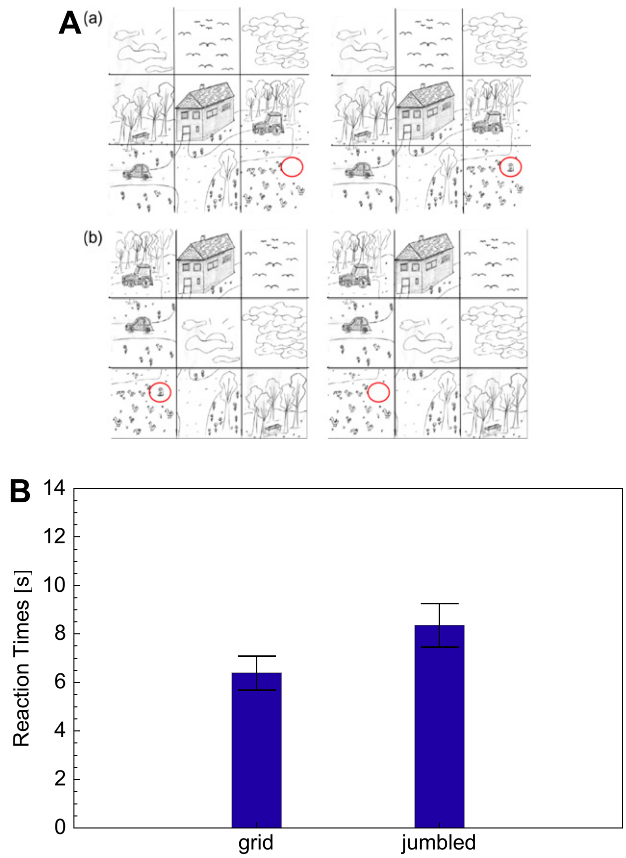

Fig. 5. (A) Example of the stimuli used in Experiment 5: (a) upright stimuli, (b) scrambled stimuli, and (c) jumbled stimuli. Red circles indicate the changing object. (B) Mean reaction times for all conditions of Experiment 5. Error bars are standard errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the objects in the images were arranged in such a way that they lay within the borders of one image piece. The jumbling process therefore could not cut through any of the objects. The number of edges remains the same in the normal and the jumbled condition. Second, the amount of visual information was evenly distributed over the image such that jumbling the image pieces would not add clutter to the image. The images were cut into nine pieces and recombined in a random order.

6.1. Methods

Eighteen subjects (10 males, 8 females; mean age 25 years) with normal or corrected-to-normal vision participated. All subjects gave informed consent. The images were presented in two conditions: normal and jumbled. A total of 14 images was presented, seven in each condition.

6.2. Results and discussion

Mean reaction times are shown in Fig. 5B. A paired t-test revealed a significant difference between the normal and the jumbled condition ($p = 0.005$). No significant differences in change detection accuracy have been found, (paired t-test, $p = 0.3$). Mean error was 1.5% of images in the normal condition and 1.2% of images in the jumbled condition. Even when the amount of visual clutter is kept constant change detection performance is reduced in images without scene context. The effect of visual scene context on change detection performance thus cannot be explained by differences in visual clutter that is contained in the images. The results from Experiment 5 also provide further evidence that image segmentation might not influence change detection. The jumbled images used here did not contain additional edges or boundaries. Nevertheless similar reaction time differences between context and no-context images were obtained.

7. General discussion

Our results demonstrate a clear contribution of natural scene context to change detection performance. Subjects were faster in detecting changes when the images contained scene context. This is in contrast with the study of Yokosawa and Mitsumatsu (2003), who did not find an effect of jumbling. The changes in the images they used have been very hard to detect, since they obtained reaction times of 28 s even in their normal condition. This makes the search for the target object very difficult and may cancel out the influence of scene context. Since we used larger changes the reaction times in our normal condition of Experiment 1 were in the range of 4 s, which is a usual result in change detection tasks in the flicker paradigm (Rensink et al., 1997). Therefore, we were able to demonstrate that scene context indeed facilitates change detection. The influence of semantic scene context is highlighted by the finding that changes, which are placed in central regions of interest will be detected faster than changes which are placed in regions of marginal interest (Rensink et al., 1997). With changes in the marginal regions of interest we observed a mean reaction time of 9 s in normal images. One could argue that because of the disruption of global scene context, jumbled images do not provide central regions of interest. But nevertheless, when the changes were placed in marginal regions of interest there was also a significant difference between upright and jumbled images. The contribution of scene context to change detection thus is not exhausted by the guidance of attention to the central regions of interest. Reaction times did not differ between upright and inverted images. Scene inversion increases the difficulty for recognizing the global scene context but does not modify the contextual relations between objects. Jumbling introduces erroneous edges to the image and the number of surface segments is increased. Any image segmentation process would be harder. Varakin and Levin (2008) found an effect of scene jumbling in the one-shot paradigm, where the original and the modified image are presented just once. They argue that the disruption of the surfaces and layout, which constitute scene structure could be responsible for the lower performance in the jumbled condition. We directly tested this hypothesis in Experiment 3. We presented abstract art images, which did not contain natural scene context as defined above. No significant difference in reaction times was obtained when these images were presented in a normal and a jumbled condition. Therefore, the disruption of surfaces itself does not affect change detection performance. Although the images used in Experiment 3 contain a context, namely that of abstract art images, our participants profited only from natural scene context. We assume that the expectations about the objects in natural scenes which structured the search for the target object could not be applied in the abstract art images. It would be interesting to see whether expertise with abstract art influences performance.

In Experiment 4 we tested whether the higher change detection performance in the upright condition relies on the semantic contextual arrangement of objects in the scene. Here we used images in which the scene context was disrupted by modifying the spatial configuration of the objects. With this method no extra surfaces were created. This condition produced a significant increase in reaction times. This finding is consistent with Hollingworth (2007) who found that scrambling the locations of objects in the test image reduces change detection accuracy. Color also influences the facilitating effect of scene context on change detection performance. Nijboer, Kanai, de Haan, and van der Smagt (2008) tested change detection performance in color and grayscale natural scene images. Reaction times were shorter in grayscale than in color images when the images contained scene context. However, when the images did not contain natural scene context no differences in reaction times between color and grayscale images were observed.

Images with natural scene context induce expectations about the configuration of objects in the scene, which facilitate the search for a specific object in that scene. In the change detection task scene context can support the search for the target object at different stages: the perceptual stage, visual short-term memory and long-term memory. At the perceptual level scene context could help to see more details in one glance. In the change detection task it is essential to compare the original and the modified version of the images. Since in the flicker paradigm each version is only seen for a short duration (in our experiments 240 ms), one is forced to recognize many objects in one glance to be efficient. A lot of studies demonstrate that scene context supports the recognition of individual objects (Bar, 2004). Scene context thus may contribute to change detection performance through a broader object representation. At the memory level scene context could support visual-short-term memory. More objects would thus be remembered over the blank period. First, a configurational scene representation helps to remember objects (Hollingworth, 2007; Jiang, Olson, & Chun, 2000; Velisavljevic & Elder, 2008).

Second, a long-term memory representation of the scene context can guide visual search. Rensink et al. (2000) proposed that two systems were used to guide attention: a medium-term memory to provide geometric context (layout) for guidance, and a long-term memory to provide semantic context (scene schema) for guidance. Indeed, a long-term memory representation of scene context has been shown to influence reaction times (Hollingworth, 2009). We conclude that a context representation is used for the visual search scenes and compensates our fragmentary visual representation in active scene perception.

Acknowledgments

We thank Kerstin Rautenberg for the design of the stimuli of Experiment 5.

References

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617–629.

Bridgeman, B., & Tseng, P. (2008). Change blindness by substituting one natural image with another. *Journal of Vision, 8*(8), 1108a.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4*(5), 170–178.

Grimes, J. (1996). On the failure to detect changes in scenes across saccade. In K. Akins (Ed.). *Perception, vancouver studies in cognitive science* (Vol. 5, pp. 89–110). New York: Oxford university press.

Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search: Evidence from search efficiency and eye movements. *Journal of Vision, 9*(1), 32.1–32.8.

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance, 33*(1), 31–47.

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition, 17*, 273–291.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 113–136.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 683–702.

Kelley, T. A., Chun, M. M., & Chua, K.-P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *Journal of Vision, 3*(1), 1–5.

Nijboer, T. C. W., Kanai, R., de Haan, E. H. F., & van der Smagt, M. J. (2008). Recognising the forest, but not the trees: An effect of colour on scene perception and recognition. *Consciousness and Cognition, 17*(3), 741–752.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*(12), 520–527.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*(3), 461–488.

Rensink, R. A., O'Regan, J., & Clark, J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368–373.

Rensink, R. A., O'Regan, J., & Clark, J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition, 7*, 127–145.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2), 17.1–17.22.

Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition, 7*(1–3), 377–396.

Shore, D. I., & Klein, R. M. (2000). The effects of scene inversion on change-blindness. *Journal of General Psychology, 127*, 27–44.

Stirk, J., & Underwood, G. (2007). Low-level visual saliency does not predict change detection in natural scenes. *Journal of Vision, 7*, 1–10.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*(4), 766–786.

Varakin, D. A., & Levin, D. T. (2008). Scene structure enhances change detection. *Quarterly Journal of Experimental Psychology (Colchester), 61*(4), 543–551.

Velisavljevic, L., & Elder, J. H. (2008). Visual short-term memory for natural scenes: Effects of eccentricity. *Journal of Vision, 8*(4), 28.1–28.17.

Vidal, J. R., Gauchou, H. L., Tallon-Baudry, C., & O'Regan, J. K. (2005). Relational information in visual short-term memory: The structural gist. *Journal of Vision, 5*(3), 244–256.

Yokosawa, K., & Mitsumatsu, H. (2003). Does disruption of a scene impair change detection? *Journal of Vision, 3*(1), 41–48.