Vision Research 50 (2010) 2158–2162

Contents lists available at ScienceDirect

Vision Research

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

Time course of the integration of spatial frequency-based information in natural scenes

Ken Kihara *, Yuji Takeda

National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

ARTICLE INFO

Article history:
Received 12 May 2010
Received in revised form 11 August 2010

Keywords:
Natural scene perception
Scene categorization
Spatial frequency
Integration
Backward masking

ABSTRACT

It is known that visual information is processed separately and based on multiple spatial frequencies. Therefore, integration of information is important for categorization of natural scenes. To clarify the time course of visual integration, we examined categorization accuracies for spatially filtered images as a function of image exposure duration. Results indicated that, with image durations of 100-ms, accuracy was superior with spatially integrable images when compared with accuracy levels based upon the probability summation model estimated from accuracies of separately presented low- and high-frequency images. This finding suggests that spatial frequency integration begins earlier than 100-ms after the image onset.

1. Introduction

It is well known that humans can recognize a complex natural scene even when it appears only briefly. For example, we can categorize the image of a natural scene as a target (i.e., it contains animals) among distractor images (i.e., they do not contain animals) when the target appears simultaneously with distractors (Rousselet, Fabre-Thorpe, & Thorpe, 2002; Rousselet, Thorpe, & Fabre-Thorpe, 2004) or when the target appears in a sequence of distractors (Potter, 1975, 1976). This is true even when the exposure duration of these images is relatively brief (approximately 100 ms). It has also been shown that a target object in natural scenes exposed for only 50 ms can be categorized with about 80% accuracy (Grill-Spector & Kanwisher, 2005). In addition, identification accuracy of a target object (e.g., a Priest) within a scene exposed for 80 ms was higher, if the object was contextually congruent with the scene's background (a church) than when it was incongruent with the background (a football field); this suggests that the semantic relationship between object identity and scene gist is rapidly analyzed (Davenport, 2007; Davenport & Potter, 2004). Consistent with behavioral studies, event-related potential (ERP) studies have shown that the ERP component reflecting scene categorization behavior was elicited at about 150 ms after stimulus onset (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Johnson & Olshausen, 2003; Thorpe, Fize, & Marlot, 1996). Taken together, these findings suggest that natural scene images are processed well enough to be correctly categorized even when images appear only for a few hundred milliseconds.

One of the major mechanisms supporting the rapid categorization of natural scenes is parallel processing of information based on multiple spatial frequencies (e.g., Bar et al., 2006; Oliva & Schyns, 1997; Schyns & Oliva, 1994). Our visual system has multiple spatial frequency channels that allow for an initial frequency-specific analysis of a scene (Wilson & Bergen, 1979). High and low frequency-based images provide respectively different content from a scene: higher spatial frequencies contain fine information of image details and/or object boundaries, whereas lower frequencies preserve coarse blobs representing the general framework of object shape and layout. What is important here is that when we view a natural scene; we are consistently aware of a single intact image, not separate multiple images depending on each spatial frequency channel. This indicates that information from these channels must be integrated prior to awareness.

The integration of multiple frequency scales appears to transpire following a coarse-to-fine progression. In this view, coarse-scale information, carried by lower frequency channels, is available earlier than fine-scale information, which is carried by higher spatial frequency channels (Hughes, Nozawa, & Kitterle, 1996; Navon, 1977; Schyns & Oliva, 1994). A schematic content of the scene served by fast processing of low frequency information would facilitate the analysis of high-spatial frequency information, thus contributing to rapid categorization of scenes (Bar, 2004; Henderson & Hollingworth, 1999). However, several studies suggest that the coarse-to-fine manner is flexible (Schyns & Oliva, 1999); that is, under certain circumstances, attentional set might modulate

* Corresponding author. Address: National Institute of Advanced Industrial Science and Technology, Central 6, 1-1-1 Higashi, Tsukuba 305-8566, Japan. Fax: +81 29 861 6732.

E-mail address: ken-kihara@aist.go.jp (K. Kihara).

0042-6989/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2010.08.012

that nature of frequency integration such that it reflects progression from fine-to-coarse channel processing (Morrison & Schyns, 2001; Sowden & Schyns, 2006). In either case, it is plausible that the integration of information supplied by multiple frequency scales is important for scene perception.

Although information provided by different spatial frequency channels should be integrated before we become aware of an intact image, little is known about the time course of such integration. As noted above, 100 ms of viewing images of a natural scene is sufficient to insure correct categorization; it is possible that this time interval is related to course of information integration from multiple frequency channels, although there is no direct evidence for a causal link between the integration of spatial frequencies and rapid scene categorization. However, at present there is no direct evidence revealing time constraints associated with the integration of information supplied by lower and higher frequency channels in scene categorization tasks. To examine the time course of scene integration, we assessed exposure times of scene images required for the integration of low with high frequency information. We expected that integration of low and high frequency information, presented simultaneously, would facilitate categorization performance.

2. Experiment 1

In Experiment 1, we investigated the categorization accuracy of intact, low-pass, high-pass, and low/high-pass filtered images of natural scenes that were presented for various durations. Intact scenes received no spatial frequency filtering. The low-pass and high-pass images isolated, respectively, coarse (low frequency) and fine (high frequency) spatial properties of these imaged scenes. The low/high-pass images contain information from both low- and high-pass frequency filters, but they lacked information from an intermediate frequency-band.

For the present purpose the critical comparison was between the categorization accuracies of low/high-pass images (experimentally recorded) and estimated accuracies based upon combining accuracy levels from low and high-pass filtered images. The latter were based on the assumption that low- and high-pass information independently contribute to scene categorization. Specifically, it is assumed that observers can categorize a scene correctly when either low- or high-pass information was processed well enough for categorization. Because the integration of information from multiple frequency channels plays an important role in scene perception, we expected that accuracy of categorization in the low/high-pass condition might be higher than the estimated accuracy computed from accuracy levels of low and high-pass filtered images if low and high frequency information in the former condition were integrated into a unified image. On the contrary, low/high-pass images would not necessarily produce superior categorization (relative to the estimated accuracy) if the low- and high-pass information were not integrated.

To control for availability of frequency information to the integration process, we manipulated exposure duration of images using backward-masking techniques that can terminate processing of visual stimuli. In this way, we compared the performance in the low/high-pass condition with the estimated accuracy in each exposure duration. If low and high-pass information were integrated within the same time window (i.e., between stimulus onset mask), then performance levels in the low/high-pass condition should be superior to the estimated accuracy.

2.1. Methods

2.1.1. Participants

Twenty-six adults (13 male and 13 female, range 19–24 years) from the subject pool at the National Institute of Advanced Industrial Science and Technology (AIST) participated in this experiment. All participants received payment for their participation. All had self-reported normal or corrected-to-normal vision. Written informed consent was obtained from all participants. This experiment was approved by the Committee of Ethics, AIST.

2.1.2. Stimuli and design

We manipulated two factors in a within-participants design: the frequency conditions of the test image (intact, low-pass, high-pass, and low/high-pass) and duration of the test image (33-, 100-, and 250-ms). We selected 1440 test images from commercially available picture libraries. The resolution of the images was 320×240 pixels. Two categories of images were defined by the presence versus absence of vehicles: half the pictures contained vehicle(s) and half did not. All test stimuli were converted into gray-scale images. Each of the two picture sets (vehicle, non-vehicle) was further divided into 12 subsets of 60 images each. The mean power spectra of each octave band in spatial frequency (i.e., 0–2, 2–4, 4–8, 8–16, 16–32, 32–64, 64–128, and more than 128 spatial cycles/image) were approximately equal across 12 subsets (Peli, 1990). A preliminary experiment with participants who did not participate in Experiments 1 and 2 confirmed that vehicle detection (i.e., hit rate in the vehicle subsets and the correct rejection rate in the non-vehicle subsets) was approximately equal across all 12 subsets.

The 12 subsets within each of two stimulus categories (vehicle, non-vehicle) were randomly assigned to four frequency conditions (i.e., intact, low-pass, high-pass, and low/high-pass) for each participant. In the intact condition, test images were not filtered. In the low-pass condition, the test images were filtered in Fourier space, using a fourth-order Butterworth filter, set to filter low frequencies (<16 cycle/image; viewed as <3.33 cycle/degree). In the high-pass condition, test images were filtered with a fourth-order Butterworth high-pass filter (> 24 cycle/image; viewed as > 5 cycle/degree). In the low/high-pass condition, test images were created by averaging gray-levels of the low- and high-pass images of an identical scene. Examples of images of four frequency conditions appear in Fig. 1.

Three levels of test image exposure duration were employed (33-, 100-, and 250-ms). In each frequency condition, images in the vehicle and non-vehicle categories were randomly assigned (in equal numbers) to one of the three exposure conditions. This resulted in 60 vehicle images and 60 non-vehicle images. All stimuli were presented on a 17-in color CRT monitor with a 60 Hz refresh rate controlled by MATLAB with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The viewing distance was 57 cm: all stimuli subtended 3.6° of visual angle vertically and 4.8° horizontally.

2.1.3. Procedures

A trial began with the display of a fixation cross which remained exposed until the spacebar was pressed. Next, a test image was presented at the center of the display, followed by a 1000-ms mask image. The display duration of a test image was varied according to the duration condition. The mask image was created from the preceding test image, using randomization of the phase of the Fourier spectrum without changing the amplitude spectrum in each trial. After the presentation of the mask image, participants were instructed to categorize the test image as either a vehicle or a non-vehicle image by pressing one of two corresponding response keys. Participants responded at their own pace and were encouraged to guess if they were uncertain whether a target was presented or not.

A previous study has indicated that stimulus visibility may be affected by the spatial frequency of the stimulus presented on the immediately preceding trial, when spatial frequency of stimuli