Consciousness and Cognition 43 (2016) 152–166

Contents lists available at ScienceDirect

Consciousness and Cognition

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

The role of edge-based and surface-based information in natural scene categorization: Evidence from behavior and event-related potentials

Qiufang Fu [a,*], Yong-Jin Liu [b], Zoltan Dienes [c], Jianhui Wu [a], Wenfeng Chen [a], Xiaolan Fu [a]

[a] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[b] Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[c] Sackler Centre for Consciousness Science and School of Psychology, University of Sussex, BN1 9QH Brighton, United Kingdom

ARTICLE INFO

Article history:
Received 15 January 2016
Revised 6 June 2016
Accepted 7 June 2016

Keywords:
Natural scene categorization
Edge-based theory
Surface-based theory
SOA
ERPs

ABSTRACT

A fundamental question in vision research is whether visual recognition is determined by edge-based information (e.g., edge, line, and conjunction) or surface-based information (e.g., color, brightness, and texture). To investigate this question, we manipulated the stimulus onset asynchrony (SOA) between the scene and the mask in a backward masking task of natural scene categorization. The behavioral results showed that correct classification was higher for line-drawings than for color photographs when the SOA was 13 ms, but lower when the SOA was longer. The ERP results revealed that most latencies of early components were shorter for the line-drawings than for the color photographs, and the latencies gradually increased with the SOA for the color photographs but not for the line-drawings. The results provide new evidence that edge-based information is the primary determinant of natural scene categorization, receiving priority processing; by contrast, surface information takes longer to facilitate natural scene categorization.

1. Introduction

Humans have a remarkable ability to categorize natural scenes quickly and accurately. The human brain needs only approximately 150 ms to decide whether a color photograph, flashed for 20 ms, contains animals or vehicles (Rousselet, Fabre-Thorpe, & Thorpe, 2002; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001), even with little or no attention applied to the task (Feifei, VanRullen, Koch, & Perona, 2005; Li, VanRullen, Koch, & Perona, 2002; Otsuka & Kawaguchi, 2007; Rousselet et al., 2002). The challenge is to explain how rapid natural scene categorization takes place in the human brain.

A recent fMRI study found that line-drawings generated similar neural activation as color photographs in the parahippocampal place area (PPA) and the retrosplenial cortex (RSC), which suggests that the human visual system uses schematic representations with content that is analogous to simple line-drawings, to encode and process statistical regularities in a scene (Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011). This finding has provided new evidence for an edge-based theory that assumes that edge-based representations are sufficient for object recognition and that surface characteristics such as color,

* Corresponding author at: Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, China.
E-mail address: fuqf@psych.ac.cn (Q. Fu).

http://dx.doi.org/10.1016/j.concog.2016.06.008
1053-8100/

brightness, and texture are less efficient routes for accessing the memorial representation (Biederman, 1987; Biederman & Ju, 1988). Indeed, some studies have found that surface gradients such as color changes had little influence on object classification and identification (e.g., Biederman & Ju, 1988; Cave, Bost, & Cobb, 1996; Joseph & Proffitt, 1996) or even impaired object classification (e.g., Gagnier & Intraub, 2012). For example, Biederman and Ju (1988) demonstrated that the reaction times and error rates were virtually identical for the common objects of color photographs and line-drawings when the images were briefly (50–100 ms) presented. Thus, although scene recognition and object recognition are technically different, the same perceptual processes might be involved. However, due to poor temporal resolution, on the order of one to several seconds (Rossion, Kung, & Tarr, 2004), the above fMRI study cannot discriminate differences in the time course of categorizing color photographs and line-drawings.

The role of surface properties in object or scene recognition remains controversial (e.g., Gagnier & Intraub, 2012; Parron & Washburn, 2010; Wichmann, Sharpe, & Genenfurtner, 2002). In contrast with the edge-based theory, the alternative surface-based theory assumes that surface gradients are central for object recognition and that both contour and surface information provide simultaneous routes for basic-level categorization. This perspective has received support from other studies (e.g., Tanaka, Weiskopf, & Williams, 2001; Wichmann et al., 2002; Wurm, Legge, Isenberg, & Luebker, 1993). For example, color improved object recognition of common food items when there was no time limit on the stimulus presentation (Wurm et al., 1993).

Interestingly, Laws and Hunter (2006) did not find a significant difference in the accuracy between the objects in color photographs and line-drawings with a 20-ms presentation of each image, which is consistent with the findings of Biederman and Ju (1988), but a marginally significant advantage for color photographs over line-drawings was found ($p = 0.07$) with a 1000-ms presentation of each image, which is principally consistent with the findings of Wurm et al. (1993). A comparison of the above studies also reveals that most of the studies in support of the edge-based theory limited the presentation times or processing duration to a very short time, while there was no time limit or a long processing time in the study that supported the surface-based theory. Thus, we predict that the stimulus presentation or processing duration could modulate the role of the surface information in scene perception. Specifically, if the processing duration is long enough, then the surface information should facilitate the recognition; however, if the processing duration is extremely short, then surface information could even impair recognition performance if edge-based information is thereby harder to extract. If the latter occurs, then the result will provide new evidence for edge-based information receiving priority processing.

The purpose of the present study was to address this issue by adopting event-related potentials (ERPs) in a backward masking task of categorizing natural scenes. To manipulate the processing duration, a backward masking paradigm was adopted in the present ERP study, in which the stimulus duration was constant but the stimulus onset asynchrony (SOA) between the image and mask was varied. Backward masking is useful in investigating the time course of information processing in the visual system in that it allows processing to be interrupted at different times (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Hansen & Loschky, 2013; Holcomb & Grainger, 2006; Kovács, Vogels, & Orban, 1995; Loschky, Hansen, Sethi, & Pydimarri, 2010; Loschky et al., 2007; Macknik & Livingstone, 1998; Rieger, Braun, Bülthoff, & Gegenfurtner, 2005; Rolls, Tovée, & Panzeri, 1999; VanRullen & Koch, 2003). Usually, when the SOA becomes longer, the behavioral performance and neural activation recorded by fMRI increase and so does the ERP differential activity, roughly between 150 and 250 ms on the targets and distracters (Bacon-Macé et al., 2005; Holcomb & Grainger, 2006). Because accuracy increases significantly with SOAs below 44 ms (i.e., 6.25, 12.50, 18.75, 25, 31.25, 43.75 ms, see Bacon-Macé et al., 2005), the SOA was set at 13, 27, 40 and 213 ms in the present study. Moreover, to explore the role of edge-based and surface-based information in natural scene categorization, we adopted color photographs and line-drawings as stimuli because the color photographs include both edge-based information (e.g., edge, line, and conjunction) and surface properties (e.g., color, brightness, and texture), whereas line-drawings include only edge-based information, as established in a previous study (Walther et al., 2011).

Previous ERP studies have shown that natural scene categorization involves two stages: a perception stage that extracts information about different features of the visual input and a decision stage that evaluates the relevance of the information in making a decision (Bacon-Macé et al., 2005; VanRullen & Thorpe, 2001). Early ERP components such as P1 and N1 are associated with feature detection or integration (Hillyard & Münte, 1984) and are sensitive to elemental features of stimuli (e.g., Holcomb & Grainger, 2006; Itier, Latinus, & Taylor, 2006). Given that color photographs involve both edge-based and surface-based information while line-drawings include only edge-based information, differences in the components at the perceptual stage could be elicited by color photographs versus line-drawings. Specifically, if surface-based information is processed simultaneously with edge-based information such that both facilitate categorization, then the latencies of the early components should be at least as fast for the color photographs as for the line-drawings. Conversely, if the latencies of early components are faster for the line-drawings than the color photographs, then it indicates the importance of edge-based information, with surface information analyzed as a secondary route for visual cognition which does not facilitate early on. Indeed, Walther et al. (2011) found that there was only a low correlation between the neural activity that was generated by color photographs and the neural activity that was generated by line-drawings in early visual areas, which suggests that the feature analysis in early visual processing differs between color photographs and line-drawings. The later ERP components, such as N2 and P3, are related to decision making (Folstein & Van Petten, 2008; Nieuwenhuis, Aston-Jones, & Cohen, 2005). If only an edge-based representation is sufficient in the decision-making process, then there will be no difference in the pattern of later components of color photographs and line-drawings, as suggested by Walther et al. Conversely, if