Vision Research 109 (2015) 99–106

Contents lists available at ScienceDirect

Vision Research

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

The shape of the face template: Geometric distortions of faces and their detection in natural scenes

Kaewmart Pongakkasira *, Markus Bindemann

School of Psychology, University of Kent, Canterbury, UK

ARTICLE INFO

Article history:
Received 30 September 2014
Received in revised form 27 January 2015
Available online 26 February 2015

Keywords:
Face detection
Face template
Natural scenes
Geometric distortions

ABSTRACT

Human face detection might be driven by skin-coloured face-shaped templates. To explore this idea, this study compared the detection of faces for which the natural height-to-width ratios were preserved with distorted faces that were stretched vertically or horizontally. The impact of stretching on detection performance was not obvious when faces were equated to their unstretched counterparts in terms of their height or width dimension (Experiment 1). However, stretching impaired detection when the original and distorted faces were matched for their surface area (Experiment 2), and this was found with both vertically and horizontally stretched faces (Experiment 3). This effect was evident in accuracy, response times, and also observers' eye movements to faces. These findings demonstrate that height-to-width ratios are an important component of the cognitive template for face detection. The results also highlight important differences between face detection and face recognition.

1. Introduction

Human face detection is the process by which observers find faces within the visual environment (see, e.g., Lewis & Edmonds, 2005; Lewis & Ellis, 2003; Tsao & Livingstone, 2008). This process appears to be distinct from subsequent categorization tasks (Bindemann & Lewis, 2013). However, in contrast to other tasks with faces, such as identification (see, e.g., Bruce & Young, 1986; Burton, Bruce, & Johnston, 1990; Burton et al., 2005) and matching (e.g., Burton, White, & McNeill, 2010; Clutterbuck & Johnston, 2002; Johnston & Bindemann, 2013), emotion recognition (e.g., Calder et al., 2001; Calder & Young, 2005), or gaze perception (e.g., Bayliss, Pellegrino, & Tripper, 2004; Driver et al., 1999; Jenkins, 2007), face detection has been studied comparatively little in Psychology. This is surprising considering that detection is an important first step for all other tasks with faces.

The available evidence suggests that face detection is automatic (Lewis & Edmonds, 2003, 2005) and very rapid (Crouzet, Kirchner, & Thorpe, 2010; Fletcher-Watson et al., 2008). This indicates that this process might rely on a "quick and dirty" processing strategy that utilizes salient visual cues to locate likely face candidates (Crouzet & Thorpe, 2011). One possibility for such a strategy could be based on a simple skin-coloured face-shaped template. This idea is based on the finding that skin-colour tones facilitate detection, but only when this is tied to the general shape of a head. Face detection is impaired, for example, when faces are rendered entirely in greyscale or unnatural colours, or when skin-colour tones are preserved in only part of a face (Bindemann & Burton, 2009). Detection performance declines also when the general shape of a face is disrupted by image scrambling (Hershler & Hochstein, 2005). In contrast, face detection appears to be unaffected by some dramatic transformations, such as the removal of the internal facial features (i.e., the eyes, nose, and mouth), provided that general face-shape and colour information is retained (Hershler & Hochstein, 2005).

Viewed together, these studies suggest that face detection might be underpinned by skin-coloured, face-shaped templates. Beyond these findings, however, the nature of such a template remains largely unexplored. One aspect, for example, that has been preserved in all previous studies in this field is the height-to-width ratio of faces. Considering the impoverished nature of facial stimuli that allow for detection to proceed unhindered (e.g., Bindemann & Burton, 2009; Hershler & Hochstein, 2005), such natural aspect ratios might be particularly important for detection. However, while this idea seems plausible, an interesting discrepancy exists that might also undermine this notion. In tasks that require the *identification* of faces, substantial geometric distortions, which dramatically disrupt the typical height-to-width aspect ratios of faces, do not appear to affect performance. For example, even when faces are stretched vertically to 150% (Bindemann et al., 2008) or 200%

* Corresponding author at: Kaewmart Pongakkasira, School of Psychology, University of Kent, CT2 7NP, UK. Fax: +44 (0) 1227 827030.

E-mail address: kp281@kent.ac.uk (K. Pongakkasira).

http://dx.doi.org/10.1016/j.visres.2015.02.008
0042-6989/© 2015 Elsevier Ltd. All rights reserved.

(Hole et al., 2002) of their actual size, while the original horizontal dimensions are maintained, the speed and accuracy of recognition is unaffected. This suggests also that face perception can be remarkably insensitive to manipulations that grossly distort stimulus shape.

In this study, we therefore wish to explore how face detection is affected by such geometric distortions, to further investigate the nature of the template that might be used for this process. For this purpose, observers were asked to locate faces in images of natural scenes in a paradigm that is adopted from previous studies (Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Burton & Bindemann, 2009). In contrast to these studies, faces were either presented with their original aspect ratios intact or these ratios were manipulated. The aim here was to examine whether this would affect the efficacy with which faces can be detected, by recording observers' eye movements and response times to faces. If so, this would suggest that these aspect ratios are an important dimension of a face detection template.

2. Experiment 1

The first experiment examined how vertical stimulus distortions affect face detection. In this experiment, observers searched natural visual scenes for frontal views of faces, which were either presented in their original aspect ratio or were stretched vertically to increase the height-to-width ratio. Two different stretch conditions were used. In these, either the original height of the face stimuli was preserved but the width was compressed by half, or the original face width was preserved but height was increased to double. These two conditions therefore provide identical height to width ratios (of 2:1), but one is comparable to the original face stimuli by retaining their height, whereas the other retains their width. If detection operates on a face-template that is sensitive to the height–width ratio of faces, then such geometric distortions should impair detection. As a result, observers should be slower to fixate these stretched faces in visual scenes and to make appropriate detection responses.

2.1. Method

2.1.1. Participants

Twenty-seven undergraduate students (8 male, 19 female) from the University of Kent, with mean age of 19.7 years (SD = 2.2), participated in this experiment for course credit. All reported normal or corrected-to-normal vision.

2.1.2. Stimuli

The stimuli were adopted from previous detection studies (Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Burton & Bindemann, 2009) and consisted of 24-bit RGB photographs of 120 indoor scenes, which were taken inside houses, apartments and office buildings. These scene images measured 1000 (W) $\times$ 750 (H) pixels at a resolution of 72 pixels/inch (subtending a visual angle of 30.5° $\times$ 23.8° at a viewing distance of 60 cm). For each scene, four versions were prepared which were identical in all aspects, except for the following differences. Three of these versions contained a photograph of a frontal face. The faces shown in these scenes were of twenty unfamiliar models (ten male, ten female) of white Caucasian origin. To ensure that the face locations were unpredictable throughout the experiment, the scenes were divided into an invisible 3 $\times$ 2 grid of six equally-sized rectangles. Across the stimulus set, the faces were equally likely to appear in any of these regions.

Apart from these commonalities, the three versions of these face-present scenes differed in terms of the aspect ratio of the faces. In the *original* face condition, the height-to-width ratios of all faces were preserved. However, the size of the faces was varied across scenes, ranging from 36 (H) $\times$ 27 (W) pixels (1.2° $\times$ 0.9° of VA) for the smallest face photograph to 139 $\times$ 115 pixels (4.7° $\times$ 3.9°) for the largest face image (mean face image dimensions, 58.7 $\times$ 47.2 pixels (2.0° $\times$ 1.6°); SD, 19.4 $\times$ 16.2 pixels (0.7° $\times$ 0.5°)). This was done to ensure that participants could not adopt a simple search strategy based on the size of the faces (see Bindemann & Burton, 2009). The height-to-width ratio of these faces was also calculated. Height was measured as the maximum

Fig. 1. Example stimuli for Experiment 1, depicting a scene without face (top left), and faces in the original (top right), horizontally compressed (bottom left), and vertically stretched condition (bottom right). (For interpretation to colours in this figure, the reader is referred to the web version of this paper.)