DOI:10.6129/CJP.202003_62(1).0005
中華心理學刊 民 109，62 卷，1 期，99-128
Chinese Journal of Psychology 2020, Vol.62, No.1, 99-128

Image-based Approach to Face Recognition: Effects of Line-drawn Faces, Pigmentation and Shading, and Spatial Frequency

Jisien Yang*[1], Gary C.-W. Shyi[2,3]
Department of Social Work, National Quemoy University[1]
Department of Psychology and Center for Research in Cognitive Sciences,
National Chung Cheng University[2]
Advanced Institute of Manufacturing with High-tech Innovations,
National Chung Cheng University[3]

The vast number of studies in face recognition have gradually evolved into three major approaches, namely, the image-based, structure-based, and function-based approach. Studies under the category of image-based approach are concerned with image-level face information, such as edge, spatial frequency, pigmentation, shading and shadow. Researchers manipulate these factors and examine the roles they may play in face recognition. In contrast, those ascribed to the structure-based approach are concerned with the effect of featural, configural, and holistic information on face recognition. Researchers adopting this approach consider faces as decomposable objects and explain face processing in terms of structural descriptions. Finally, studies under the third approach investigate face recognition from a functional perspective, where the main issues concern how humans extract biological as well as social information from faces that are pertinent to social interactions, such as identity, gender, age, race, and emotional expression, among others. In this review, we first discuss the differences between these three approaches, and then focus our review in details on the findings accrued within the image-based approach. Our goal is to provide an integrated overview of the image-based approach and emphasize the importance of image-based processing in understanding face recognition. Specifically, our review highlights how basic processing of image information such as border, spatial frequency, pigmentation, may lay the critical foundation for the later processing of the structural and functional aspects of a face.

Keywords: *Face recognition model, image-based research, spatial frequency, shape from shadow*

In the extensive literature on face recognition, three major approaches have gradually evolved, namely the image-based, structure-based, and function-based approaches. In this review, we first discuss the differences between these three approaches, and then focus in detail on the findings of studies using the image-based approach. Studies in this category examine image-level face information, such as the face edge, spatial frequency, pigmentation, shading, and shadow. Researchers have manipulated these factors and examined the roles they may play in face recognition. In contrast, studies using the structure-based approach are concerned with the effects of featural, configural, and holistic information on face recognition. Researchers adopting this approach consider faces as decomposable objects and explain face processing in terms of structural descriptions. Finally, studies taking the third approach investigate face recognition from a functional perspective, focusing on how humans extract from faces biological and social information pertinent to social interaction, such as identity, gender, age, race, and emotional expression.

Bruce and Humphreys (1994) suggested that visual descriptors for face recognition can be divided into edge-based information and surface-based information. "Edge-based information" refers to discontinuities in the image (e.g. edge information), whereas "surface-based information" refers to features such as pigmentation and shadow in the image. Spatial frequency information, which influences both edge-based and surface-based information, is also discussed in this paper.

Edge-based information is obtained from the contours of a face, its features, or other facial characteristics. Line-drawn faces are an extreme example of this type of information, comprising edge information only. Research has shown that edge-based information may be insufficient for face recognition, as indicated by participants' prolonged reaction time and shortened eye gaze when asked to recognize line-drawn faces (Price & Humphreys, 1989; Bruce & Humphreys, 1994; Heuer, 2016). In addition, when photographs of celebrities were turned into line-drawn faces, recognition accuracy decreased dramatically, from 90% to 47% (Davies, Ellis, & Shepherd, 1978). Leder (1996) also found that participants were better at recognizing photographs of configuration-altered faces than line-drawn faces. Together, these results suggest that surface-based information such as pigmentation, shading, and shadow, rather than the mere outline of a face, is critical for face recognition.

In many studies, surface-based information has been divided into two categories: pigmentation and shading. Face recognition ability declines when the pigmentation of a face is reversed, especially when transformed from the positive to the negative. Second, information on a face's shading, the result of lighting, is critical for its three-dimensional construction and hence recognition (Bruce & Langton, 1994; Kemp, Pike, White, & Musselman, 1996).

Philips (1972) discussed three possible reasons why faces are difficult to recognize in photographic negatives (Galper, 1970). First, faces in negatives appear expressionless (Galper & Hochberg, 1971). Second, shadows, which offer a key clue to a face's three-dimensional shape, are difficult to interpret (Bradshaw & Wallace, 1971). Third, human faces are depicted in a wide range of gray tones in negatives.

Liu and Chaudhuri (1998) found that the recognition of internal features in two-tone and multi-tone images suffered more than the recognition of external features when using photographic negatives. They proposed that the processing of face images in photographic positives and negatives differs qualitatively rather than quantitatively.

Hayes, Morrone, and Burr (1986) further examined the influence of spatial frequency on the recognition of photographic positives and negatives. They filtered face images in photographic positives and negatives in five bandwidths, centered at 3.2, 6.4, 12.5, 25, and 50 cycles per face (CPF), with 1.5 octave filters. They found that recognition performance reached a peak at about 20 CPF. In other words, the image in the bandwidth with 20 CPF provided the most useful information for face recognition. In addition, negative images were more difficult to recognize than positive images when components of the images had a low spatial frequency. Negative and positive images whose components had a high spatial frequency were recognized equally well. These results suggest that the difficulty of recognizing photographic negatives results mainly from the processing of low spatial frequency information within the negatives.

Kemp, Pike, White, and Musselman (1996) explored the effect of negatives on face recognition from a different perspective, focusing on the roles of shape-from-shading and pigmentation. They manipulated hue and luminance separately and inferred that the negative effect was due primarily to the shape-from-shading source. However, Russell, Sinha, Biederman, and Nederhouser (2006) provided yet another point of view. Specifically, they asked participants to match faces differing only in shape or only in pigmentation. They found that the pigmentation was the only factor that impaired performance. Therefore, they claimed that the perception of pigmentation, not shape, is selectively disrupted by negation, and that this disruption is the main source of the effect of negatives on recognition. Together, these results suggest that both shading and pigmentation are critical to face recognition and causes of the negativity effect.

Studies that have manipulated spatial frequency and examined its effects have focused on three main areas, namely low-pass filtering, high-pass filtering, and band-pass filtering. In low-pass filtering, image information above a preset spatial frequency is filtered out, leaving only information below the preset spatial frequency; this results in a coarser image. In high-pass filtering, information below the preset spatial frequency is filtered out and that above the preset spatial frequency remains, yielding a finer image. In interval-pass or band-pass filtering, the information above and below a preset interval between two spatial frequencies is filtered out and the information within the interval remains (Campbell & Robson, 1968; Bachmann, 1991; Hayes, Morrone, & Burr, 1986).

Neuroscientific research has found that cells may respond selectively to different spatial frequencies in the cortex; they are thus called spatial frequency analyzers. These neurons show different turning curves when responding to information on different spatial frequencies (Maffei & Fiorentini, 1973; De Valois, De Valos, Yund, 1979; Robson, Tolhurst, Freeman, & Ohzawa, 1988; De Valois & De Valos, 1980; De Valois, Albrecht, & Jacobs, 1982; Graham & Nachmias, 1971). Lehmkule, Kratz, Mangel, and Sherman (1980) also found that different types of cells in the lateral geniculate nucleus deal with high and low spatial frequencies, respectively, in separate visual pathways.

No consensus has yet been reached on the relative speed and priority of processing of high and low spatial frequency information. The “coarse-to-fine” hypothesis contends that low spatial frequency information is processed earlier and faster than high spatial frequency information (Parker, 1980; Parker & Salzen, 1977; Vassilev & Mitov, 1976; Parker & Costen, 1999; Parker, Lishman, & Hughes, 1992, 1996; Schyns & Gosselin, 2003; Parker & Dutch, 1987; Schyns & Oliva, 1994). Schyns and Oliva (1994), for example, used four kinds of stimuli, namely full spectrum images, low-pass filtered images, high-pass filtered images, and hybrid spatial frequencies, to examine the roles of coarse and fine information in scene processing. Their findings supported the hypothesis of coarse-to-fine processing. The authors argued that recognition occurs on both coarse and fine spatial scales. However, the coarse spatial scale may be activated first to quickly achieve a rough estimate of the input, followed by the activation of the fine scale to refine the information.

Parker, Lishma, and Hughes (1996) proposed an entirely different viewpoint. They found that pre-cuing via a fine scale (i.e., high spatial frequency) is more conducive to obtaining a later correct target and more disruptive of irrelevant targets than pre-cuing via a coarse scale (i.e., low spatial frequency). The results of their experiments favored the notion that higher processing priority is given to high than to low spatial frequency information.

Instead of debating the processing priority of high versus low spatial frequency, Schyns and Oliva (1999) examined the hypothesis of flexible scale usage proposed by Schyns and Gosselin (2003). This hypothesis suggests that the brain adjusts the processing priority given to different spatial frequencies based on the demand of the task. Schyns and Oliva (1999) found that in a task requiring participants to judge the expressiveness of the face, processing was biased toward high spatial frequency information, whereas in a task involving the categorization of expressions, processing was biased toward low spatial frequency information. Importantly, they found that the priority of processing and spatial frequency bias was not fixed but could be dynamically changed according to task demand. When a spatial frequency bias—either high or low-was induced, participants continued to favor that specific spatial frequency channel when judging gender in subsequent images. As a result, they sometimes perceived the same face to be of different genders. It seems that we are flexible in applying specific spatial frequency channels that suit our needs, and that this bias can continue to influence our later perceptions.

Many studies have also found that various aspects of face processing rely on specific spatial frequencies, such as band passes of 5.5-10.5 CPF (Fiorentini, Maffe, & Sandini, 1983), 8-16 CPF (Costen, Parker, & Craw, 1994), 17.6 CPF (Tieger & Ganz, 1979), and 8-13 CPF, two octaves wide (Nasanen, 1999). For example, in numerous studies of gender judgment, the participants relied primarily on low spatial frequency information

(Schyns, Bonna, & Gosselin, 2002; Goffaux, Jemel, Jacques, Rossion, & Schyns, 2003; Deruelle & Fagot, 2005). In addition, fearful expressions coded as low spatial frequency led to strong activation in the amygdala, but fearful expression coded as high spatial frequency did not (Vuilleumier, Armony, Driver, & Dola, 2003). Finally, Goffaux, Hault, Michel, Vuong, and Rossion (2005) found that high spatial frequency information was more useful than low spatial frequency information when processing feature information in faces. However, holistic processing primarily relies on low spatial frequencies (Goffaux & Rossion, 2006).

To summarize, this article provides a comprehensive and integrated review of the image-based approach and emphasizes the important role played by image-based processing in understanding face recognition. Most importantly, our review shows that the basic processing of image information, such as edges, spatial frequency, and pigmentation, may lay an important foundation for the later processing of the structural and functional aspects of a face.