Iowa State University
Department of Psychology

Midwestern Psychological Association's Annual Meeting, Chicago, Illinois
May 3, 2007

Trevor Wolff, Jeremiah D. Still, Derrick J. Parkhurst & Veronica J. Dark

Invariant Features from Computer Vision Allow Better Human Object Recognition in Color Photographs

Introduction

Biederman (1987) suggested that shape-related non-accidental features (e.g., vertices and parallel lines) are important for human object recognition. He demonstrated that object naming accuracy of fragmented line drawings was higher when the fragments contained such features.

We explored whether a similar effect would be found using photographs of real objects. To determine non-accidental features we adopted a computer vision algorithm developed by David Lowe (1999) referred to as the Scale Invariant Feature Transform (SIFT). This algorithm identifies regions in photographs that are likely to be invariant to orientation, rotation, and spatial frequency.

We investigated whether the presence of these invariant features would increase object recognition in humans presented with fragments of photographs of objects that included natural characteristics such as color, texture, and shading.

Method

The SIFT algorithm was applied to 150 photographs of nameable objects downloaded from the Amsterdam Object Library (Geusebroek, Burghouts, & Smeulders, 2005). The invariant features detected by the SIFT algorithm were transformed into a pre-attention map (see Figure 1). The "hot spots" within the pre-attention map reflect areas of densely grouped invariant features (see Figure 2). We then used the maps to create our stimuli (see Figure 3).

Figure 1: *Creating Invariant Features Pre-attentional Map*

Procedure

From each whole picture, we constructed two types of picture fragments: the invariant fragment and its complement (see Figure 3). Each fragment contained 50% of the pixels comprising the object. The invariant fragments contained the 50% of pixels representing the highest density of invariant features. The complement fragment contained the remaining pixels.

Figure 2: *Peaks in Pre-attentional Map Reflect Regions Dense in Invariant Features*

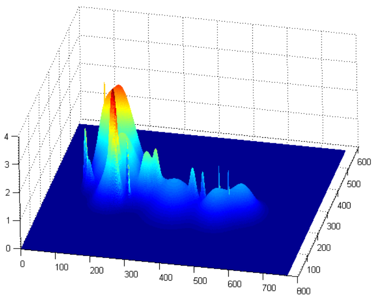
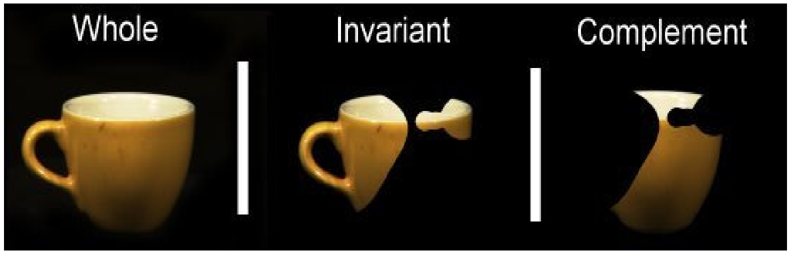


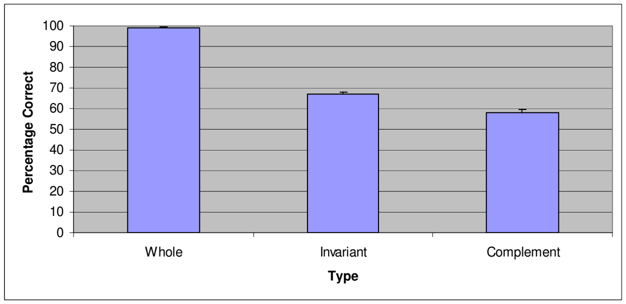
Figure 3: *Example Stimuli*



Participants (N = 21) viewed 50 whole pictures, 50 invariant fragments, and 50 complements for 2 seconds each. Immediately after viewing an image, participants typed the name of the object they had just seen.

We predicted higher naming accuracy for the invariant fragments compared to their complements.

Results



Identification was virtually perfect for the whole pictures, indicating that the objects were namable. A matched-pairs t-test showed more accurate object naming for fragments containing more invariant features, $t(20) = 4.79$, SE = .018, $p < .01$.

Discussion

These results indicate that the SIFT algorithm, which was designed for computer vision, is detecting features that are useful to the human object recognition system. Participants were better able to name the fragments containing more of the invariant features identified by SIFT.

In another experiment, we found that these invariant features attract attention as reflected in eye fixations. Further, we showed that SIFT invariant features better accounted for fixations than an account based on feature saliency (Still, Dark, & Parkhurst, 2007).

These experiments support the hypothesis that dense regions of invariant features play an important role for object recognition.

References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.

David G. Lowe, Object recognition from local scale-invariant features, *International Conference on Computer Vision*, Corfu, Greece (September 1999), pp. 1150- 1157.

Geusebroek, J., Burghouts, G., J., & Smeulders, A., W. M. (2005). The Amsterdam library of object images. *International Journal of Computer Vision*, 61(1), 103-112.

Still, J. D., Dark, V. J., & Parkhurst, D. J. (2007). *Viewpoint Invariant Object Features Attract Overt Visual Attention*. Poster presented at the meeting of the Vision Sciences Society, Sarasota, FL.

[For more information e-mail: jeremiah@iastate.edu]