IOWA STATE UNIVERSITY
Human-Computer Interaction Program & Department of Psychology

Object Perception, Attention, & Memory 16th Annual Meeting, Chicago, Illinois
November 13, 2008

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

Dense Regions of View-Invariant Features Promote Object Recognition

Introduction

Although saliency (Parkhurst, Law & Niebur, 2002) and invariance (as determined by Scale Invariant Feature Transform – SIFT; Lowe, 1999) are highly correlated, we found that invariance more accurately predicted overt attentional selection, as measured by eye movements, than did salience (Still, Dark & Parkhurst, 2007).

To investigate whether invariant features that are important for computer vision really contribute to human object recognition, we created two photograph fragments for each photo. One condition contained 50% of the pixels associated with the densest regions of invariant features as defined by the adapted SIFT algorithm (see Figure 2). The other condition, the complement, contained the remaining 50% of the pixels. We found that objects were more easily identified when the fragments contained more invariant features (Wolff, Still, Parkhurst & Dark, 2007).

In the current experiment, we manipulated the percentage of pixels displayed with the displayed regions selected either randomly or in terms of density of invariant features. If invariant features are critical for object recognition, naming accuracy should be higher for the invariant conditions compared to the random conditions. In addition, the benefit for invariant regions should decrease as number of pixels is increased.

Method

Stimuli were a set of 64 nameable photographs of complex objects from the Amsterdam Library of Object Images (Geusebroek, Burghouts and Smeulders, 2005). Eight fragments were constructed for each object. The fragments varied in whether pixel regions were chosen randomly or in terms of density of invariant features (see Figure 1); they also varied in the proportion of pixels chosen (20%, 40%, 60%, 80%; See Figure 3).

Figure 1: Creating an Invariant Features Pre-attentional Map

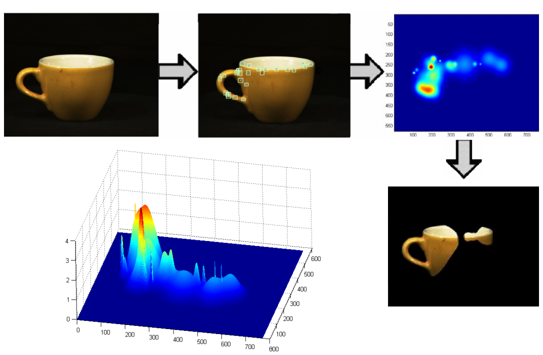
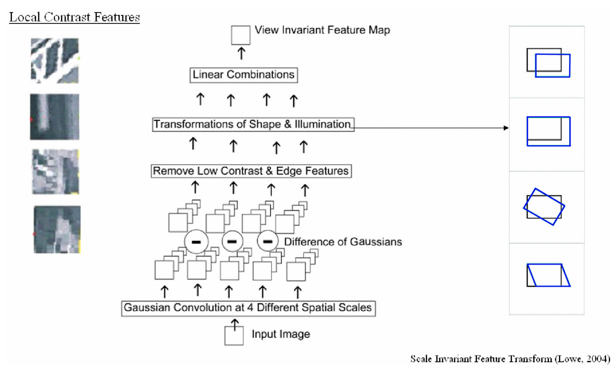


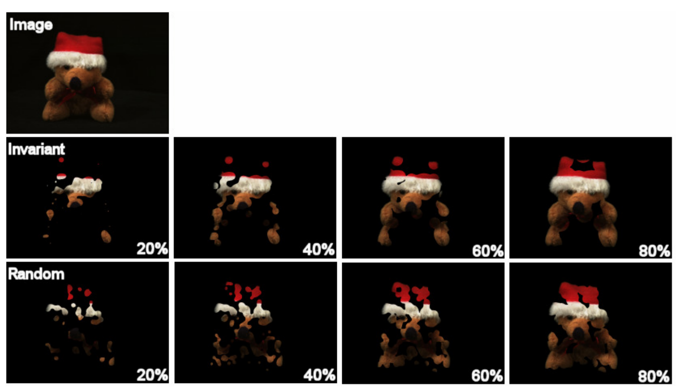
Figure 2: Schematic of our Adaptation of the SIFT Algorithm



Procedure

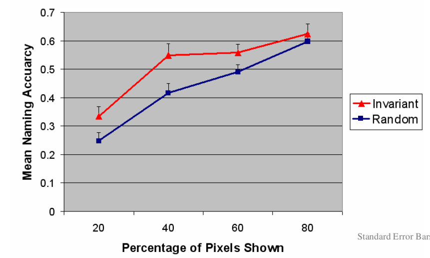
Participants (N = 32) viewed each fragmented object for 1,000 msec and then typed the object name, guessing when unsure. Fragment type (invariant or random) and proportion of pixels were manipulated within-subjects with each participant viewing 8 images for each of the 8 conditions. Counterbalancing insured that each participant saw only one version of an object and that each object was presented in all conditions.

Figure 3: Example Stimuli



Results

Figure 4: Participant Object Recognition Naming Accuracy



Invariant fragments (i.e., those constructed to capture invariant features) were more likely to be named correctly than random fragments $F(1, 31) = 24.88$, $MSE = .016$, $p < .001$. As the proportion of fragments increased, naming accuracy increased, $F(3, 29) = 43.91$, $MSE = .027$, $p < .001$. We had predicted that the benefit to invariant fragments would decrease as the proportion of pixels shown increased. Although the predicted interaction was not significant, one-tailed comparisons showed that accuracy in the invariant condition was higher than in the random condition with 20% pixels, $t(31) = 2.10$, $SEM = .04$, $p = .02$, and with 40% pixels, $t(31) = 2.44$, $SEM = .05$, $p = .01$. The difference was marginally significant with 60% pixels, $t(31) = 1.66$, $SEM = .04$, $p = .05$, but not significant with 80% pixels, $t < 1$, $p = .26$.

Discussion

Regions containing invariant features appear to be important for human object perception, confirming our previous findings (Wolff et al., 2007). The same type of invariance that is important for computer vision (Lowe, 1999) was important for human object recognition.

It is possible that dense regions of invariant features are extracted by a pre-attentive, low-level system and that these regions provide information for programming eye movements (Rensink, 2000). This guidance of attention to invariant regions may play an early role in object perception. Taken together, our research supports the hypothesis that the visual system is biased to select visual features likely to be important for object recognition.

References

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.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91-110.

Lowe, D. G. (1999, September). Object recognition from local scale-invariant features. Paper presented at the *International Conference on Computer Vision*, Corfu, Greece, 1150-1157.

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*, 107-123.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17-42.

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

Wolff, T., Still, J. D., Parkhurst, D. J. & Dark, V. J. (2007, May). *Invariant Features Detected with Computer Vision Allow Better Human Object Recognition in Photographs*. Poster presented at the meeting of the Midwestern Psychological Association, Chicago, IL.