The AAAI 2017 Spring Symposium on
Science of Intelligence: Computational Principles of Natural and Artificial Intelligence
Technical Report SS-17-07

Is the Human Visual System Invariant to Translation and Scale?

Yena Han,[†] Gemma Roig,[†‡] Gad Geiger,[†] Tomaso Poggio[†‡]
[†]Center for Brains, Minds and Machines, MIT
[‡]LCSL, Istituto Italiano di Tecnologia at MIT
77 Massachusetts Ave, Cambridge, MA 02139

Abstract

Humans are able to visually recognize objects presented at different scales and positions. It is not clear whether this transformation-invariant recognition is learned by experience, or if the brain computes an invariant representation of objects, allowing invariant recognition even after a single exposure to an object. Numerous behavioral studies with humans provided results on translation and scale invariance, but the stimuli used were mostly familiar objects to the subjects. Also, the inconsistent results on translation invariance in object recognition, which may be due to differences in the nature of the stimuli used in their experiments, make it hard to examine one-shot invariance. We study this one-shot invariance using unfamiliar letters. We analyze the recognition performance in a same-different task presenting the characters on a display for 33ms at different scales and positions. We confront our experimental results with the predictions made by a computational model of the feedforward stream of the human visual cortex. The model is grounded on invariance principles. Also, it characterizes the eccentricity dependency of the receptive fields of the neurons in the primary visual cortex. Our data suggest that the feedforward path of the human visual system computes a representation of objects that are scale invariant. We also observe limited position invariance, and the degree of invariance is maintained if the eccentricity is increased linearly with scale. Those results are consistent with the predictions from the computational model.

Introduction

Human eyes are resource constrained and cannot cover the whole visual field at super-high resolution. But why does resolution decrease linearly with eccentricity? Are there computational reasons or is this just an accident of evolution? We argue that the reason for it is the need to provide invariance to geometric transformations, such as scale and position, to help the basic task of visual recognition.

The question whether the human visual system computes invariant representations to transformations is a fundamental vision problem. Recognition of objects at different scales and positions can take place trivially because of previous experience and memorization of several transformed images of the object. It is however likely that humans can also recognize specific objects seen only once at different positions and scales. There are several studies in the literature reporting inconsistent results on translation-invariant object recognition (Nazir and O'Regan 1990; Dill and Fahle 1998; Dill and Edelman 2001). Such results might be due to the fact that previous works use stimuli of different nature in complexity of the structure of the object, and spatial frequency. Thus, it is not clear from experimental data whether the robustness to recognize objects under different transformations is due to experience, or to the brain computing invariant representations of those objects with respect to geometric transformations.

Limited translation invariance using patterns of randomly distributed dots in a square area was reported by (Nazir and O'Regan 1990); whereas Dill (Dill and Fahle 1998; Dill and Edelman 2001) conducted experiments using animal-like shapes, and found more invariant properties of the human visual system. Yet, when the task was to distinguish animals built from their scrambled parts, their results in (Dill and Edelman 2001) showed that there was no distinction among shapes built from the same scrambled parts. They argued that this might be caused by a pooling operation across space

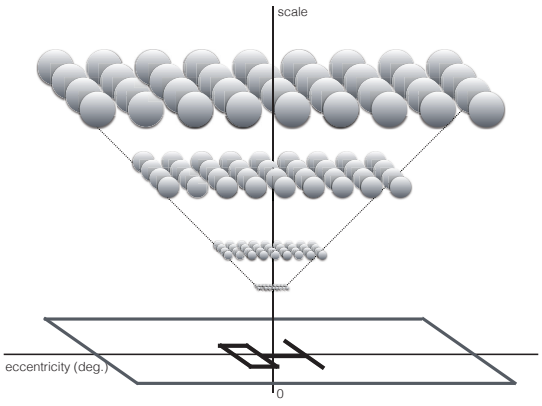

Figure 1: Computational model of the primary visual cortex from (Poggio, Mutch, and Isik 2014). Each circle represents a neuron, and there are the same number of neurons at all scales. Yet, the neurons at a larger scale covers a larger eccentricity than those at a lower scale.

that yields that those animals with the same parts arranged spatially different are indistinguishable.

Also, very little research has been done analyzing scale invariance for object recognition in one-shot learning scenario. In (Biederman and Cooper 1992; Furmanski and Engel 2000) experiments were conducted evaluating the recognition performance with familiar objects at different scales. Their results are inconclusive regarding invariant representations in the brain, because high recognition performance of familiar objects of different sizes may be due to experience, *i.e.* the objects have been seen before at different scales by the human subjects. Other literature related to scale invariance, is about determining frequency bands and channels in the primary visual cortex (Chung, Legge, and Tjan 2002; Majaj et al. 2002). These literature indeed suggest a high degree of scale invariance of letters in the human visual system. Yet, no experimental results have been reported studying the recognition rate of unfamiliar objects at different scales.

The simple application of recent mathematics on invariance suggests that the geometry of the retina and of the cortex reflects the first step of computing representations of the image that are invariant to changes in scale and position. This is captured in the computational model of the primary visual cortex presented in (Poggio, Mutch, and Isik 2014). This computational model based on invariant representation principles (Anselmi et al. 2016), accounts the eccentricity dependency of the receptive fields of the neurons. As displayed in Fig. 1, the visual window is modeled with a truncated inverted pyramid model (Poggio, Mutch, and Isik 2014). In the model, receptive fields of different sizes and eccentricities leads to simultaneous invariant representation to scale and translation. The size of the smallest receptive field is a linear function of eccentricity, and the number of receptive fields at each scale is the same. Then, pooling over responses from different receptive fields gives invariance within the range of the inverted pyramid.

We tested the prediction in (Poggio, Mutch, and Isik 2014) by measuring with psychophysical experiments about invariance in object recognition. We used unknown letters to the subjects, which are Korean letters, to study invariance to scale and transition in the human visual system.

Methods

Stimuli and setup To examine the range of invariance through one-shot learning, we used Korean letters that were unfamiliar to the human subjects who participated in our study. Stimuli set of 24 Korean letters was used for scale invariance experiments and 27 letters were used for position invariance experiments. Black letters were presented on a white background. Stimuli were presented on a 60 Hz Dell U2412M monitor at a distance of 1.26 m using the Psychophysics Toobox (Brainard 1997) for MATLAB (Guide 1998). The program was run by a Windows computer. Subjects were seated with a chin rest.

Participants Between six and seven subjects participated in each experiment. Subjects had normal or corrected-to-normal vision.

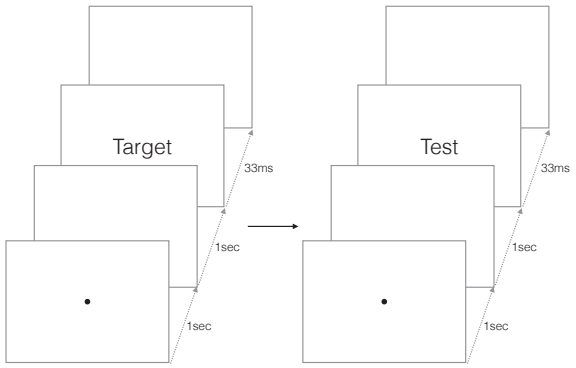

Figure 2: Stimuli sequence.

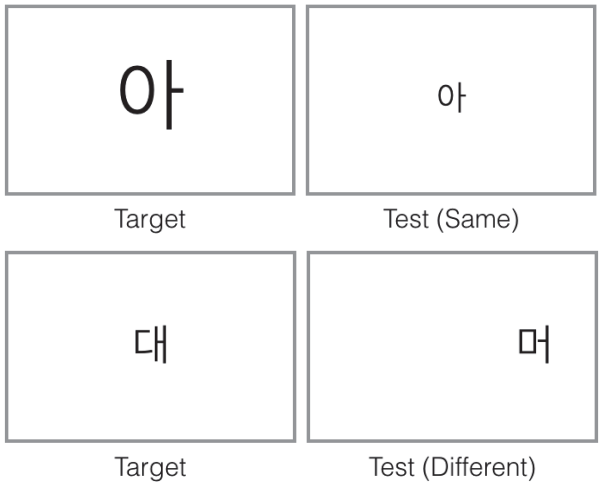

Figure 3: An example set of stimuli used to test scale invariance (top) and position invariance (bottom).

Experiment procedure The performance of learning unfamiliar letters was measured in a same-different task. The subjects first fixates to a black dot at the center of the screen. For each trial, after 1 sec the fixation dot disappears and the target letter is presented for 33 ms, followed by a white screen for 1 sec. Then, the test letter is shown for 33 ms, again followed by a white screen for 1 sec. Finally, the question of the task appears, in which the subject is asked if the target and test letters displayed previously were the same or different. In Fig. 2 we show the sequence of letter presentations for one trial. The letters used were different for every trial. The presentation time was limited to 33ms to avoid eye movements. Otherwise, the subjects would view the letters the entire stimuli in their fovea, regardless of the size and position of letters.

Experimental design To test scale invariance, both target and test letters were presented at the center, and the sizes of letters were varied. In the first scale invariance experiment, invariant recognition of 30 min and 1 deg letters were tested. Specifically, the combinations of the target and test letter sizes were (30 min, 30 min), (30 min, 1 deg), and (1 deg, 30 min). Similarly, in the second scale invariance experiment, (1 deg, 5 deg), (1 deg, 5 deg), and (5 deg, 1 deg) pairs were used.

Position invariant recognition was evaluated by changing the position of test letters while keeping the letter size the same as the target letters. The test settings can be divided into two cases: learning in the foveal vision, where target letters were presented at the center and test letters at the periphery; and learning in the peripheral vision, where target letters were presented at the periphery and test letters appeared at the same position, at the center, or at the periphery on the opposite side. 30 min, 1 deg, and 2 deg letters were used for the position invariance experiments.

In order to study the linear relation of letter size with the range of invariance to position, the positions of letters at the periphery were determined linearly with the letter size with a factor of 2. Therefore, under the peripheral presentation condition, 30 min, 1 deg, and 2 deg letters were presented at 1 deg, 2 deg, and 4 deg, respectively. For comparison, 2 deg letters were also tested at 5 deg, which is beyond the range of the linear increase.

Analysis Analyses of variance (ANOVAs) were computed to analyze the differences of mean performance under different presentation conditions. For a comparison between two performance levels acquired from the same group of subjects, repeated measures ANOVAs were applied.

Results

Scale invariance The first hypothesis tested was invariant recognition of 30 min and 1 deg letters. The three conditions compared in Fig. 4 are when the size of target and test letters are (30 min, 30 min), (30 min, 1 deg), and (1 deg, 30 min). Mean performance rates under all three conditions were higher than 0.8, which is well above chance (0.5). Although the results from applying repeated measures ANOVA to the data reveal that the difference of performance across the conditions were statistically significant ($F(2, 10) = 4.3, p < 0.05$), since the overall performance was over 0.8, the drop in performance was not manifest. Moreover, pairwise comparisons showed that the difference between the presentation combinations (30 min, 1 deg) and (1 deg, 30 min) was significant ($F(1, 5) = 34.1, p < 0.003$), while there was no statistically significant difference between (1 deg, 30 min) and (30 min, 30 min) or (1 deg, 1 deg) ($F(1, 5) = 0.28, p > 0.61$, $F(1, 11) = 0.08, p > 0.78$, respectively). This implies that even though the sizes used in the experiment were the same, whether the letter size increased or decreased has an effect on performance.

The results from the second set of experiments, testing invariant recognition of 1 deg and 5 deg letters, were similar to those from the first setting. Performance rates, overall, were lower than those from the first set of experiments. Nevertheless, correct rates of discriminating letters of the same or different sizes were again significantly higher than chance, and the difference of performance across three presentation combinations was statistically non-significant ($F(2, 8) = 0.45, p > 0.65$).

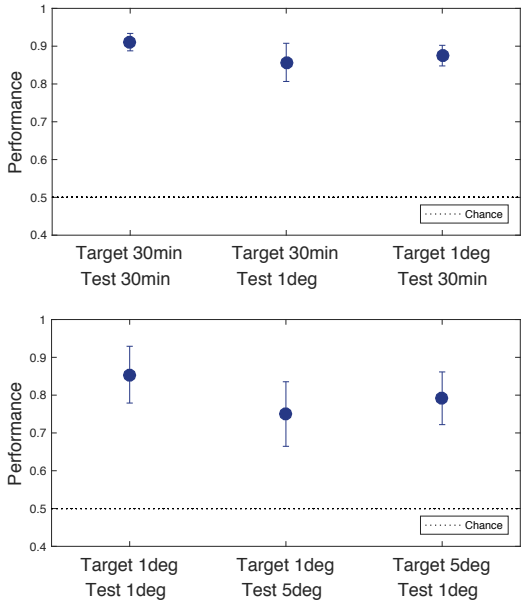

Figure 4: Scale invariance experimental results.

Translation invariance The results of translation experiments, using three different letter sizes, are shown in Fig. 5 (a) and (b). Recognition performances of foveal and peripheral learning, defined by whether the target letters are learned in the foveal or peripheral visual field, are plotted respectively. The results indicate that recognition of unknown letters is not completely invariant to translation. In foveal learning, when letters are 1 deg or 2 deg, performance significantly decreased as test stimuli are shifted to the periphery ($F(1, 5) = 6, p < 0.06$; $F(1, 4) = 53.82, p < 0.002$). However, it is hard to interpret whether this is due to limited invariance or lower visual acuity at the periphery, since the difference between the conditions (0 deg, D deg) and (D deg, D deg) was not significant when letters are 1 deg ($F(1, 5) = 0.014, p > 0.9$) while it was significant in the case of 2 deg letters ($F(1, 4) = 12.56, p < 0.03$). Also, when an outlier, subject 1, in the experiment using 2 deg letters was excluded from the analysis, the difference in performance was no longer significant ($F(1, 3) = 3.52, p > 0.15$).

In peripheral learning, the position of target and test letters are (D deg, D deg), (D deg, 0 deg) or (D deg, opposite D deg). The results of invariant recognition, which are from the last two conditions, were compared with the baseline condition, *i.e.* (D deg, D deg). Overall, the performance of (D deg, 0 deg) condition was slightly lower than (D deg, D deg) condition, but the difference was not statistically significant for

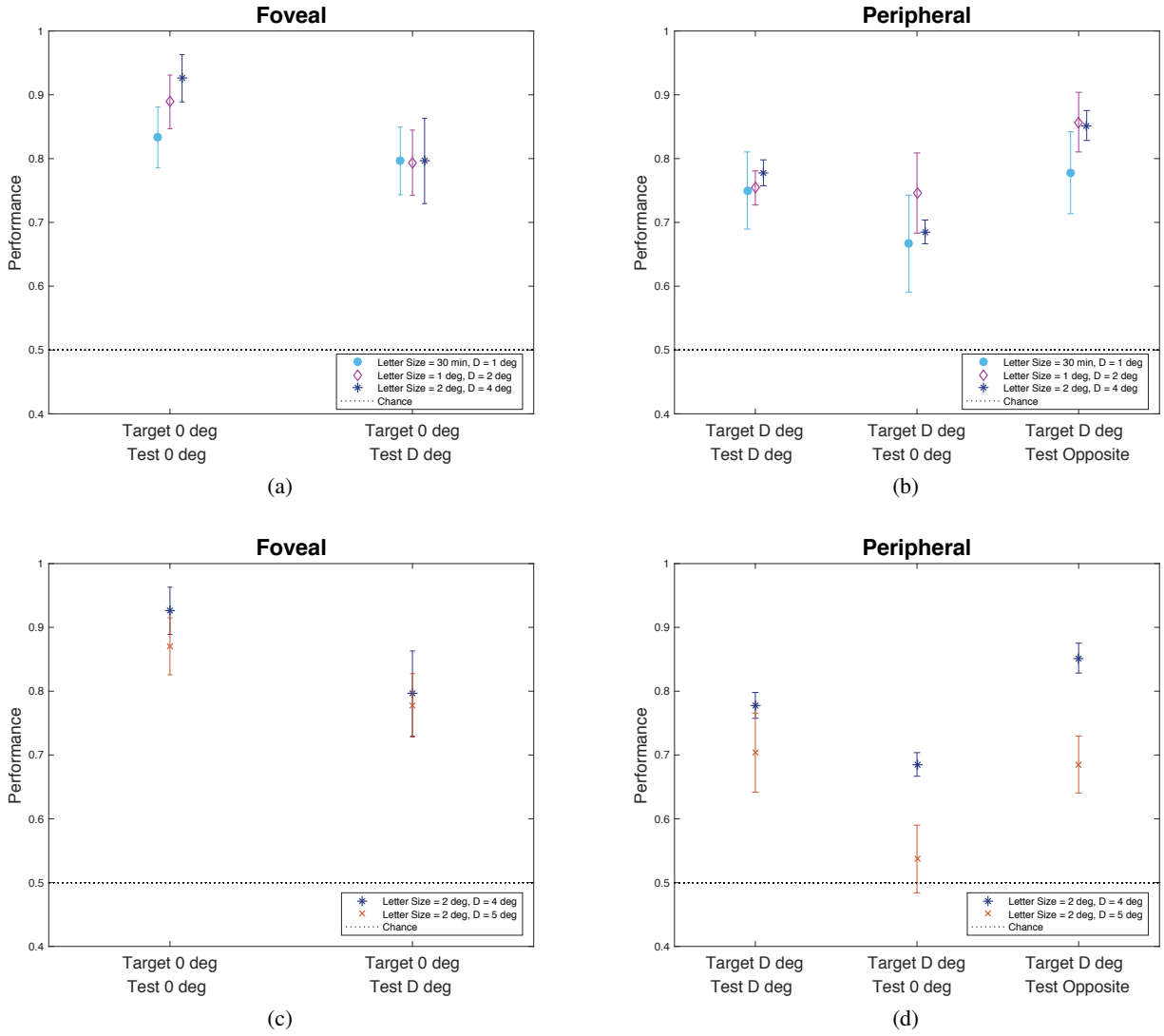

Figure 5: Translation invariance experimental results. (a) and (b) are comparisons of performance when changing the letter size linearly with the eccentricity with a factor of 2. (c) and (d) show results from increasing the eccentricity of letters more than linearly with the letter size.

all three letter sizes ($F(1,5) = 2.25, p > 0.19$, reporting the lowest F and p value among three). Surprisingly, performance increased slightly under (D deg, opposite D deg) condition. The effect was statistically significant when letters were 2 deg ($F(1,4) = 5.28, p < 0.09$), but not for 30 min and 1 deg letters. Since some stimuli were symmetric, this performance enhancement may originate from recognition of symmetric stimuli on the opposite site being easier than information transfer to other shifts in position. Nonetheless, the answer to the question whether viewing letters on the opposite side indeed had benefits over recognizing at the same position as where the letters were first learned requires further study.

Scale-translation invariance relation Recall that the distance of the letters presented at the periphery was increased linearly with the letter size to test the hypothesis of the linear relation between scale and translation invariance. Our results suggest that the degree of invariance is preserved when the letter size is increased linearly with translation. With the linear increase in scale and translation, the conditions testing invariant recognition *i.e.* target and test letters at (0 deg, D deg), (D deg, 0 deg) or (D deg, opposite D deg) did not result in a significant change in performance. Corresponding F and p values computed from one-way ANOVA were $F(2,16) = 0, F(2,16) = 0.52, F(2,16) = 0.84$, and 1.0, 0.60, and 0.45.

Fig. 5 (c) and (d), by contrast, depict the results when the linear relation between scale and translation was not ful-

filled. The position of 2 deg letters presented at the periphery was increased from 4 deg to 5 deg. These results indicate that the increase in distance significantly affected peripheral invariant recognition, but had no effect on the foveal case. Among the invariant recognition test conditions, (0 deg, D deg) did not show a significant drop in performance between $D = 4$ and $D = 5$ ($F(1, 10) = 0.05$, $p = 0.83$). However, performance reduced significantly for the two peripheral invariant recognition cases, (D deg, 0 deg) and (D deg, opposite D deg) ($F(1, 10) = 6.96$, $p < 0.03$; $F(1, 10) = 10.95$, $p < 0.008$).

Discussion

We conducted psychophysics experiments with humans to test scale and position invariance in object recognition. We used rapid categorization in a same-different task with a one-shot learning paradigm. We used Korean letters that were unknown to the human subjects. This allows to distinguish whether recognition is due to experience or if the brain computes invariant representations of the objects.

In our experimental data we observed almost perfect scale invariance in recognition. The high recognition rates were maintained even when the letters were tested at a different scale than they were initially learned. This suggests that the feedforward path of the primary visual cortex computes an object representation that is invariant to scale.

We observed limited translation invariance. On one hand, learning the Korean letters at the center of fixation and testing at the periphery, was less invariant than learning the letters at the periphery and testing at the center. However, the performance of learning at the center of fixation with different letter sizes and testing at the periphery—with eccentricity depending on the letter size—were not significantly different among the different letter sizes. Remarkably, when learning the letters at some eccentricity and testing on the opposite eccentricity with respect to the fixation point, performance was generally even higher than the trivial case of learning and testing at the same eccentricity.

We confronted our experimental results with predictions from the computational model of the feedforward path of the primary visual cortex by (Poggio, Mutch, and Isik 2014). Our results on invariance match the predictions by this model, in which a preferred invariance to scale over transition is captured by the distribution of neurons in a truncated inverted pyramid organization, which responses are pooled to obtain an invariant representation.

Acknowledgments

This material is based upon work supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216.

References

Anselmi, F.; Leibo, J. Z.; Rosasco, L.; Mutch, J.; Tacchetti, A.; and Poggio, T. 2016. Unsupervised learning of invariant representations. *Theoretical Computer Science* 633:112–121.

Biederman, I., and Cooper, E. E. 1992. Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance* 18(1):121.

Brainard, D. H. 1997. The psychophysics toolbox. *Spatial vision* 10:433–436.

Chung, S. T.; Legge, G. E.; and Tjan, B. S. 2002. Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision research* 42(18):2137–2152.

Dill, M., and Edelman, S. 2001. Imperfect invariance to object translation in the discrimination of complex shapes. *Perception* 30(6):707–724.

Dill, M., and Fahle, M. 1998. Limited translation invariance of human visual pattern recognition. *Perception and Psychophysics* 60(1):65–81.

Furmanski, C. S., and Engel, S. A. 2000. Perceptual learning in object recognition: Object specificity and size invariance. *Vision research* 40(5):473–484.

Guide, M. U. 1998. The mathworks. *Inc., Natick, MA* 5:333.

Majaj, N. J.; Pelli, D. G.; Kurshan, P.; and Palomares, M. 2002. The role of spatial frequency channels in letter identification. *Vision research* 42(9):1165–1184.

Nazir, T. A., and O'Regan, J. K. 1990. Some results on translation invariance in the human visual system. *Spatial vision* 5(2):81–100.

Poggio, T.; Mutch, J.; and Isik, L. 2014. Computational role of eccentricity dependent cortical magnification. *CBMM Memo 017. arXiv preprint arXiv:1406.1770*.