Ariana Familiar
CIS 581
Project 3B Report

Abstract

Humans have the remarkable ability to remember thousands of objects they encounter and generalize over past experiences to recognize new objects. For centuries, philosophers and psychologists have studied how human cognition allows for rapid and automatic object recognition. That being said, most studies have utilized small sets of real-world object stimuli, and examining behavior using such objects is confounded by preexisting knowledge of the objects. Some studies have attempted to use novel objects, however such stimuli are contrived and often very simple in visual properties while real objects are much more complex. In the present project, I developed a process to generate novel objects that are based on real-world objects, which can be used to systematically examine how and under what conditions people learn about visual objects. As a Psychology PhD student, I will be able to use such objects as stimuli in future experiments.

Introduction

Object recognition is a widely studied topic in cognitive psychology and cognitive neuroscience. For centuries, research in this area has attempted to elucidate the conceptual structures and learning mechanisms that allow for such processes. This work has employed stimuli that are either real-world objects or simple novel objects. The use of real-world objects to study object learning and subsequent recognition is useful, but is confounded by prior knowledge of the objects. Moreover, novel objects generated by researchers are contrived and often simple in visual properties, and so may not tap into the same mechanisms as when processing complex real-world objects.

Although there is a rich literature of research on human object recognition, there has been no former work, to my knowledge, that has attempted to create novel object stimuli based on the statistics of real-world objects. My main objective in the present project was to develop novel objects that retain visual (geometric and texture) information about real-world objects without user-definition of this information. These can be used to study and compare recognition abilities in humans. Importantly, these stimuli were generated using shape and color information derived from images of real objects.

Related Work

Some former studies have attempted to generate novel object stimuli, however these are typically simple and contrived by experimenters. A widely used set of novel objects are termed “greebles”, which were originally used in an early neuroimaging study on the brain regions involved in human object perception[1]. While novel, the greebles contain human-like features (e.g. ears), and so using these stimuli is confounded by previous knowledge and similarity to real-world objects. A recent study on object learning used novel stimuli that were 3D and defined by sinusoidal modulations[2]. These stimuli are unique and do not necessarily resemble real-world objects, but are simpler in visual appearance than real-world objects. Additionally, there have been some attempts at generating large databases of novel objects[3], however these items are nonetheless defined by experimenters and are simple and/or resemble real-world objects.

Although focused on computer-based object generation, computer vision work has offered similar results. Research in this area has either aimed to generate new images that combine objects based on separate images of the objects, but retain most of the information in the original images (e.g. virtualized reality), or to generate novel images based on simplified visual properties (e.g. spectral images). Moreover, although focused on creating novel objects that are often complex, computer graphics involves user-generated renderings. Recently, deep-learning approaches have been used to generate novel objects, however, to my knowledge, these have been aimed at generating simple objects (e.g. symbols) or real-world objects[4, 5].

The primary feature that used to generate stimuli in the present work was shape information derived from real-world objects. Former research has established the behavioral relevance and importance of shape as a defining feature of many objects. Shape information is immediately available through visual input and is informative of the functional use of an object and how we may physically interact with it. There is even a specific brain region in the visual system that contains neurons that are selective for shape information[6]. Thus, shape is a relevant feature for object recognition. In this project, I combined the shapes of real-world objects to create novel object shapes that retained naturalistic shape-information.

Implemented Methodology

Research labs have made publicly available large sets of object images for academic purposes. In the current project, I used a set of images of unique objects from a former study[7] (available at: **https://bradylab.ucsd.edu/stimuli.html)**. There are 405 images of objects (256 x 256 pixels), all centered on a white background.

First, images were randomly chosen from the entire set depending on the user definition of how many images to use. Next, these images were convolved with a Gaussian blur kernel (sigma = 1), and canny edge detection was performed. Border detection was then performed on each image to find the global shapes of the objects (using bwboundaries).

Next, two-item morphing was performed until all of the images were used. First, a morph was generated between two randomly chosen images from the subset. Corresponding points were chosen based on the border detection output (100 points per image), such that the points were placed equidistantly along the border of each object. Delaunay triangulation and thin-plate spline warping was used to generate a morph (average between binary versions of the two images). This resulting morph was then morphed with another randomly chosen image from the subset (using the same process described above), until all images in the subset were used (no images were repeated).

Lastly, the texture of the morphed shape was defined as follows. A texture sample was generated by selecting a central region of each image (52 x 52 pixels), and cross-dissolving the samples across images (equal proportion of each image). Texture synthesis was then used to expand this sample to a larger image (256 x 256 pixels)[8]. The resulting texture was applied to the morphed shape, thus creating a final stimulus output.

This process was repeated as many times as necessary in order to output the number of generated stimuli defined by the user. There was a constraint that no image from the set was processed more than once across stimuli.

Comparison with proposed methodology

The overall approach of the stimulus generation process follows my proposed methodology exactly. The only major change I had to make was in the border detection process. Originally, I thought it would suffice to use the edge detection results and 'erode' the center region of the object, ultimately leaving just the overall border of the object. However, because the edge detection process resulted in holes in the edges of the objects, this led to sub-optimal border results and thus sub-optimal morphing results.

I tried to fix this using multiple different approaches, none of which led to better results across images. Some of these were: (1) used very liberal thresholds in the edge linking phase, (2) used the built-in Matlab canny edge detector, (3) added an eroded binary mask of the original image to the edge detection results, and (4) included median filtering of the results of (3) to reduce noise.

Ultimately, I used the built-in Matlab function bwboundaries for border detection, which outputs pixel coordinates of the exterior boundaries of objects. While this does not result in perfect border detection across all images (e.g. holes in light-colored regions, Figure 1), it resulted in better morphing and more reasonable final results (Fig. 2), especially across the variance of objects in the set.

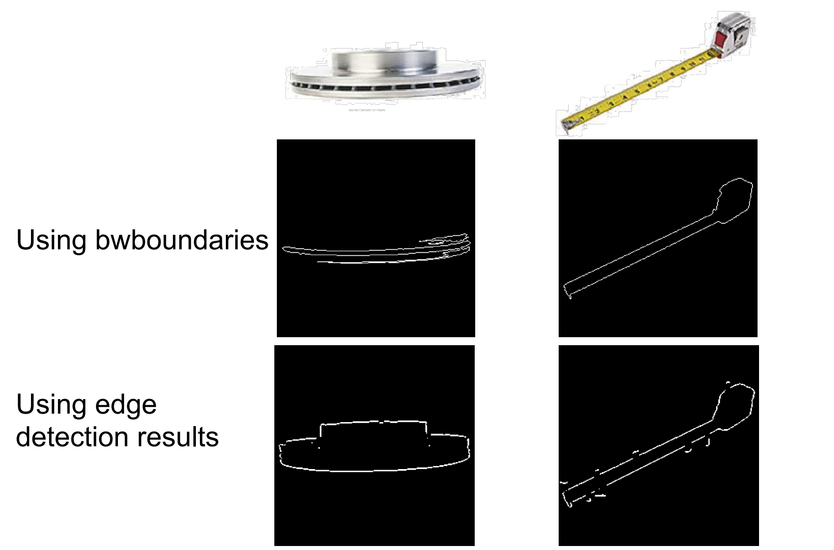

Figure 1 Examples of border detection results

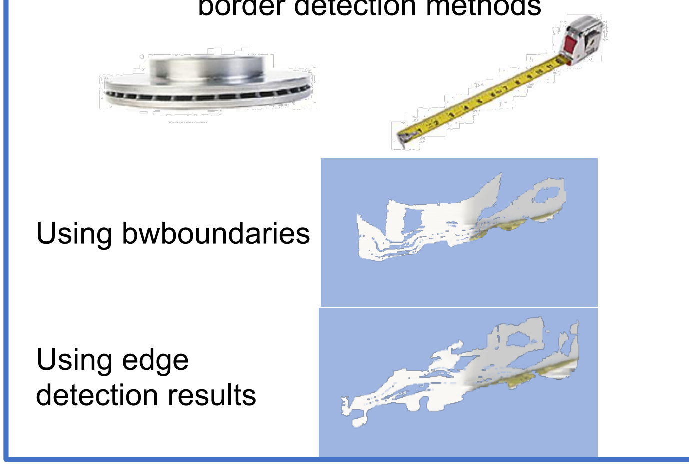

Figure 2 Examples of results using different border detection methods

Results

I generated 5 sets of novel stimuli. One set consisted of 2 stimuli generated from 4 object images, one of 3 stimuli from 3 images, one of 2 from 6, and one of 4 from 6 (Fig. 3). This was done in order to show how the results changed as a function of the number of images being combined, and to show the variance across resulting stimuli within an image subset.

The stimuli in all sets were very novel, in that they did not resemble a particular item in the subset of images used to produce them, or a particular object in general. The shapes and textures took on a combination of parts from the input images. The resulting textures were very interesting, as they were novel but included the color statistics of the input images.

Figure 3 Results

Compared versus expected results

The overall difference in the compared and expected results was the shape of the resulting stimuli. I was not anticipating as much noise in the border detection and morphing processes, which resulted in noise in the outputted shape of the resulting morphs. This noise was primarily introduced by images of objects that were light-colored or included transparent or very light regions, and objects that included very thin parts (Fig. 4). Moreover, there was a significant amount of noise introduced by the original images, as the borders around the objects were not perfect – there were some pixels that extended outside of the object that should have been background pixels but instead were not the same values as the background (Fig. 4).

In these cases, it was difficult to define the border using the normal operation, and so the morphing was influenced by sub-optimal borders (because the corresponding points were not properly covering the border of the object). In developing these operations, I spent the most amount of time attempting to find a process of border detection and morphing that worked across the many different types of objects, but I could not find a method that worked for the difficult images (see Fig. 4 for examples). However, some additional pre-processing of the images (e.g. using very saturated versions of the images or better cropping of the objects) may help to alleviate these issues in the future.

Furthermore, I originally assumed that it would take many iterations of morphing different images in order to generate shapes that appeared novel (e.g. 50 images). In fact, it took very few iterations to generate stimuli that appeared novel with the implemented methodology (e.g. 2-4).

Figure 4 Examples of difficult images

Future work

In future work, the inputted images and some of the steps in the stimulus generation process can be modified to produce better results. For instance, a pre-processing step of over-saturating the object images may benefit the edge and border detection processes, specifically for objects with unclear borders (e.g. very thin objects or transparent/light-colored objects).

Furthermore, the texture definition process could benefit from improving the texture sampling step. For instance, choosing a central region of the image may incorporate some background pixels for thin objects, so it may be better to include object-only pixels (e.g. the central 100 pixels within the object). This would require images in which the object is clearly defined from the background (e.g. better cropping). Also, a method of choosing only the most informative, or most defined, texture pixels may benefit the final texture output.

Moreover, future versions could implement 3D object information to produce novel 3D stimuli that can be used for testing. Also, the process could be used to generate stimuli based on real-world categories (e.g. artificial vs. natural), which would generate novel objects within such classes for testing categorization in humans and computers.

In sum, this work provides a crucial first step to utilizing natural object information to generate novel stimuli for future experimentation. The present process was developed to generate novel object stimuli based on real-world shape and texture information. These stimuli can be used in tests of object recognition in humans and computers, and can be used to compare performance between the two.

References

1. Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: a framework for studying expert object recognition processes. *Vision research*, *38*(15), 2401-2428.
2. Mattar, M. G., Thompson-Schill, S. L., & Bassett, D. S. (2017). The network architecture of value learning. *Network neuroscience*.
3. Horst, J. S., & Hout, M. C. (*in press*). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods.* Retrieved [Nov. 2, 2017] from: http://michaelhout.com/?page_id=759.
4. Sinha, A., Unmesh, A., Huang, Q., & Ramani, K. (2017). SurfNet: Generating 3D shape surfaces using deep residual networks. *arXiv preprint arXiv:1703.04079*.
5. Lehman, J., Risi, S., & Clune, J. (2016). Creative generation of 3D objects with deep learning and innovation engines. In *Proceedings of the 7th International Conference on Computational Creativity*.
6. Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*(5534), 1506-1509.
7. Brady, T. F., Konkle, T., Alvarez, G. A. and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, USA*, 105(38), 14325-14329.
8. bgheneti, Implementation of the non-parametric sampling based Efros Leung texture synthesis algorithm in Matlab, (2015), GitHub repository, https://github.com/bgheneti/textureSynth