Bootstrapped learning of novel objects

Mark J. Brady
Department of Psychology, University of Minnesota, Minneapolis, MN, USA

Daniel Kersten
Department of Psychology, University of Minnesota, Minneapolis, MN, USA

Recognition of familiar objects in cluttered backgrounds is a challenging computational problem. Camouflage provides a particularly striking case, where an object is difficult to detect, recognize, and segment even when in "plain view." Current computational approaches combine low-level features with high-level models to recognize objects. But what if the object is unfamiliar? A novel camouflaged object poses a paradox: A visual system would seem to require a model of an object's shape in order to detect, recognize, and segment it when camouflaged. But, how is the visual system to build such a model of the object without easily segmentable samples? One possibility is that learning to identify and segment is opportunistic in the sense that learning of novel objects takes place only when distinctive clues permit object segmentation from background, such as when target color or motion enables segmentation on single presentations. We tested this idea and discovered that, on the contrary, human observers can learn to identify and segment a novel target shape, even when for any given training image the target object is camouflaged. Further, perfect recognition can be achieved without accurate segmentation. We call the ability to build a shape model from high-ambiguity presentations bootstrapped learning.

Keywords: object recognition, learning, camouflage, segmentation, background, clutter, color, motion, mechanochemical, morphogenesis, novel objects, top down, bottom up

Introduction

A fundamental function of biological vision is to detect and recognize potential food items and predators from naturally cluttered backgrounds. The task can be especially difficult when the form and coloration of the target objects are similar to the background. Camouflage provides a particularly striking example, where natural (or artificial) mechanisms work to disguise an object even when in "plain view." It wasn't until the advent of computer vision research in the 1960s and 1970s that it was realized that most objects, not just those that are camouflaged, blend in with their backgrounds to a surprising extent. In fact, to this day, the segmentation of static figures from cluttered backgrounds has no robust machine vision solution. To the best computer vision system, almost all objects are camouflaged, not just those intending to hide. What we might call the unintentional camouflage of everyday vision is the rule, rather than the exception. Finding object contour boundaries is difficult because image edges can be caused by illumination and material changes, and not just the depth discontinuities defining object boundaries. Further, when an object is seen against a cluttered background, its contour boundaries tend to merge with the contours of background elements. There is an apparent scarcity of local image features that remain invariant over viewpoint, lighting, and background changes. The objective local ambiguity stands in contrast to the speed and accuracy with which humans can identify objects in natural images (Thorpe, Fize, & Marlot, 1996).

It is generally believed that the extraordinary competence of the primate visual system at detecting and recognizing objects involves coupling local image measurements with global knowledge of the shapes and properties of objects and object classes previously seen. Early computational approaches used generic knowledge of surface properties (piece-wise smoothness) to link together contour segments or texture measurements likely to belong to the same surface (Poggio, Torre, & Koch, 1985). Systems relying solely on generic grouping principles, however, tend not to be robust. Greater robustness can be achieved, at the cost of generality, by relying on specific knowledge of familiar objects for both segmentation and for nonsegmented classification (Amit & Geman, 1999; Yuille, 1991). High-level object knowledge is also important for dealing with occlusion and object articulations. However, the importance of high-level object knowledge leaves us with a profound paradox. When object learning begins, there is no model of the object. If ambiguous local image measurements are not sufficient for object recognition, then surely it is not sufficient for model building. We call this the bootstrapped learning problem.

One way out of this paradox is to assume that learning occurs under conditions of low ambiguity. For example, the movement or color of an object may distinguish it from its background. Motion is a well known basis for segmentation (Braddick, 1974; Lamme,

DOI 10:1167/3.6.2 Received December 20, 2002; published July 22, 2003 ISSN 1534-7362 © 2003 ARVO

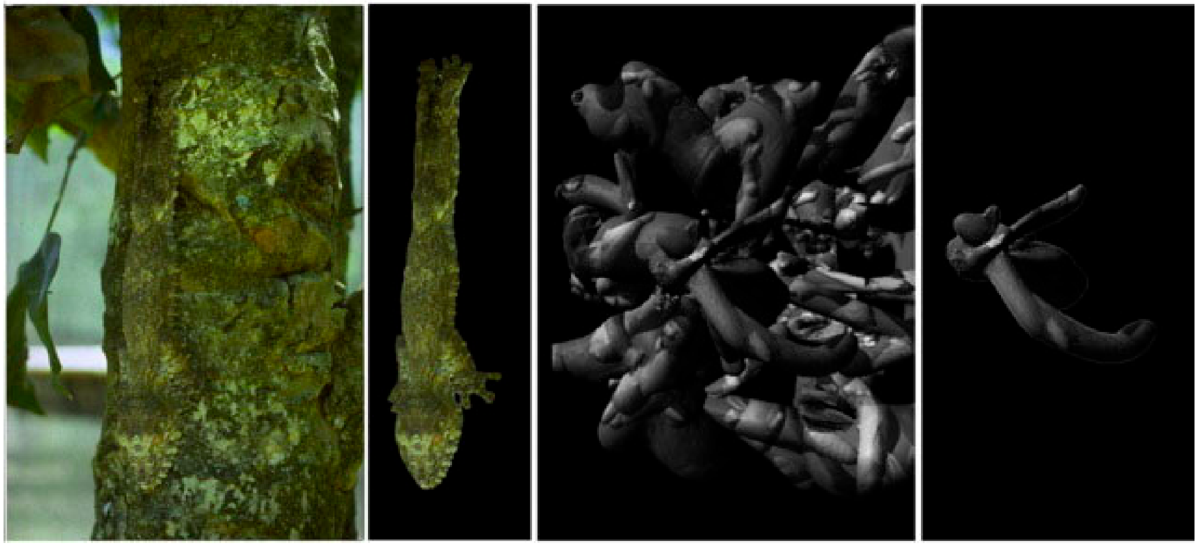

Figure 1. A. A photograph of a flat-tailed gecko, with and without background. B. An artificial morphogenic object, or "digital embryo," also with and without background. Despite the fact that both objects are unoccluded, they cannot be segmented without prior knowledge of the object. Digital embryo scenes mimic aspects of nature's more severe forms of camouflage. Gecko photograph by M. Kramer (http://home.wxs.nl/~mkramer/).

1995), and many animals display vivid color patterns as warnings, or for social recognition (Brown, 1975). Therefore, it may be that object learning is opportunistic, occurring when segmentation is possible, but not under conditions of high ambiguity or camouflage, and it is only after learning that an observer can recognize or segment a static camouflaged object. We tested this hypothesis by training observers on images in which a camouflaged novel object is presented against a cluttered background with (1) motion-defined boundaries, (2) color-defined boundaries, or (3) ambiguous boundaries. The hypothesis predicts that on testing, recognition performance should improve for the low-ambiguity conditions (1 & 2), but not for the ambiguous condition.

Methods (Experiment 1)

Stimuli

To generate unfamiliar camouflaged objects, we specify object, camouflage, and scene models for image variation.

Object model

In order to study how objects are learned in camouflage, it is necessary to have stimuli that retain the generic properties of naturally important surfaces and yet are unfamiliar. Plants and animals are fundamental to our survival; unfortunately, there are no guarantees that any set of plants or animals will be novel to all observers in an experiment. We solved this problem by simulating some aspects of embryological development to grow 3D shapes, rendered using computer graphics (M. J. Brady, 1999). For a demonstration of the growth process, see M. Brady, 1999. Our morphogenic algorithm is mechanochemical (i.e., intracellular forces as well as diffusion of a chemical morphogen are both simulated, such that chemical pattern formation and shape forming movements of cells occur simultaneously). For other examples of mechanochemical morphogenesis, see Ball (1999) as well as Murray (1993). We call our resulting objects digital embryos. (See right panel of Figure 1). Digital embryos appear to be organic forms but do not resemble a familiar class of organism.

Camouflage model

Camouflage occurs when the surface texture and/or shape of an animal or object appears similar to the background (left panel of Figure 1).

Background model. We adopted the extreme form of camouflage in which the target object is set against a background of similar objects, all drawn from the same class, in our case, other digital embryos that also had albedo variations that we describe next (right panel of Figure 1).

Surface texture model. In nature, intra-species albedo variation may seem minor (as in deer) or it may be major (as in zebra and giraffe). Upon close inspection, many of the seemingly minor variations prove to be quite extensive. Within individuals, albedo variation may occur due to mud wallowing, precipitation, sweat, molting, shedding, changing clothes, or wearing makeup. Shadows cast from forest canopies may also be confounded with these albedo variations. In this experiment, we mimic the more rigorous conditions found in the learning of both object classes and individuals by the use of major changes in albedo patterns. The appearance of a given object of fixed shape was varied by painting, or texture mapping,

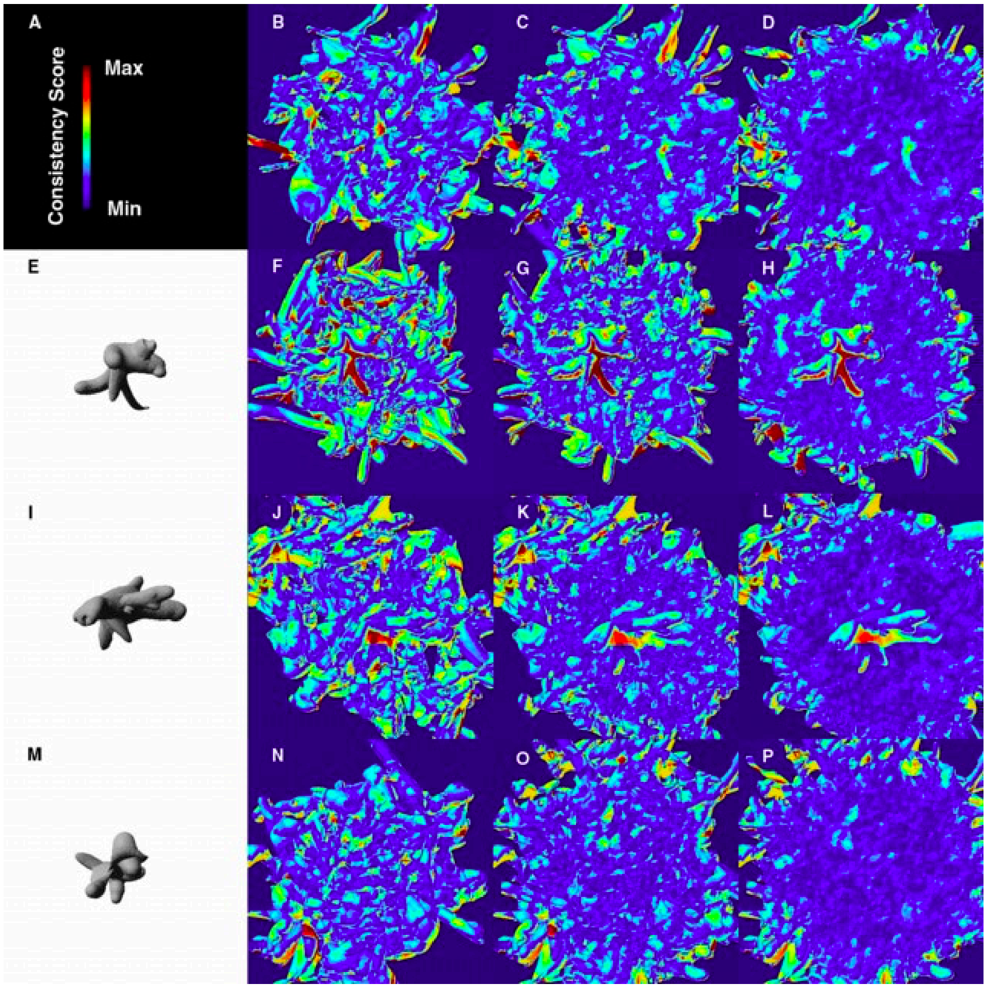

Figure 2. Output of the consistency algorithm. The consistency algorithm takes a set of training images and attempts to find object fragments, which appear repeatedly in the training sequence (see "Appendix B"). The algorithm is challenged because the object images vary due to changing camouflage, translation in 3D space, and minor changes of relative viewpoint induced by the translation. The algorithm attempts to translate a series of training images to align the object of interest in each. In the image resulting from the averaging of the translated training images, background pixels will tend toward a mean value (minimum consistency score), whereas the darker and lighter portions of the object of interest will remain near the extreme pixel values (maximum consistency score). Pseudo-coloration A allows the reader to consider various thresholds for object versus background. B-D. Distance-from-the-mean images from a training set of 3, 7, and 20 images, respectively. The object of interest is shown in E. Some hypothetical fragments may be selected from these images but the uncertainty is high. F-H. Algorithm output for 3, 7, and 20 training images with the same object shown in E, except that the object was not camouflaged during training. The algorithm performs much better, showing the effect of camouflage on image consistency. J-L. Output from 3, 7, and 20 training images containing the object shown in I. Uncertainty is high after 3 training images but decreases with further training. A suitably chosen threshold could segment out a diagnostic fragment for use in object recognition. Other thresholds would produce a mix of false positives and false negatives. Performance was similar on 4 out of 9 object training sets. N-P. Results from the training set for the object shown in M. Uncertainty remains very high. Four out of 9 training sets produced similar results. In all training sets, uncertainty is very high at the onset of training and remains high even after 3 images. One would expect that an observer would not be able to segment the first image of a training set.

different gray-level albedo patterns on the surface. The camouflage was made particularly challenging by using albedo patterns that were themselves images of other digital embryos. One consequence of this manipulation is to introduce albedo edges, internal to the object, that mimic shading seen at self-occluding boundaries or folds (Ben-Shahar, Huggins, & Zucker, 2002). Using albedo patterns, which mimic 3D-surface shading and discontinuities, might seem to be an unrealistically difficult form of camouflage. However, such mimicry appears frequently in nature (Thayer, 1909), a dramatic example of which is found in the moths of the genus Callionima (http://dlp.cs.berkeley.edu/photos/). Also, albedo patterns, which mimic the 3D surfaces of vegetation, are the basis of many popular patterns on modern hunting clothing.

Our camouflage model ensured that no geometric forms other than embryos need to be introduced or their effects explained.

Scene model

We fixed object orientation, distance, and illumination, but varied the background objects, position of the object of interest, and camouflage patterns from trial to trial. In the color training scenes, objects of interest were colored green. In motion training scenes, objects of interest were animated over a parabolic path with quasi-random coefficients. (See “Appendix A” for rendering details). The complete set of stimuli can be viewed and downloaded (Brady & Kersten, 2000).

Characterization of object camouflage

We wanted to characterize algorithmically the extent of albedo and contrast variation, in order to measure the extent to which portions of an object appear repeatedly in a sequence of training images. To do this, we applied a translation invariant consistency algorithm, which detects objects or object parts that appear repeatedly in a series of scenes. The results are shown in Figure 2. Object fragments reappear to varying degrees. However, there is considerable uncertainty whether repeating fragments are from the object or from the background.

Observer Training

We trained six adult observers on scenes of novel objects and then tested their ability to recognize those objects. The training and testing scenes were generated by placing a digital embryo at random in a scene, applying camouflage, then placing other camouflaged embryos to fill in the background (Figure 1, upper right). The camouflage patterns consisted of scenes of novel objects. Every scene had a new set of background objects and new camouflage patterns on the objects.

The goal of training was to imitate a natural object learning scenario, where an animal views an object, such as another animal, the object may disappear for some seconds, then reappear. It may not reappear for some hours or days, when learning may resume. Each time it appears it has a new background and if it is a different instance of the same object class, it may have a new surface pattern. The object may emit a sound or an odor, which helps to identify it. In this experiment, we played a sound with each appearance.

Each observer was trained on three sets of three objects. One set was shown as moving against its background, one set was shown in color against its background, and one set was shown with no supplemental segmentation clues. We call this last case the ambiguous case.

Movie 1. The first six scenes of a training session, in the ambiguous segmentation case.

Training took place over 4 days. Each training session consisted of showing object A for 10 s, B for 10 s, C for 10 s, and then A, B, C, repeating until each object had appeared 5 times. In each appearance of an object, the camouflage pattern, background, and object location was changed. In order to compare any two scenes of the same object, observers had to hold one scene in memory during a 20-s delay and two intervening learning tasks.

Movie 2. First four scenes of a training session, in the motion segmentation case.

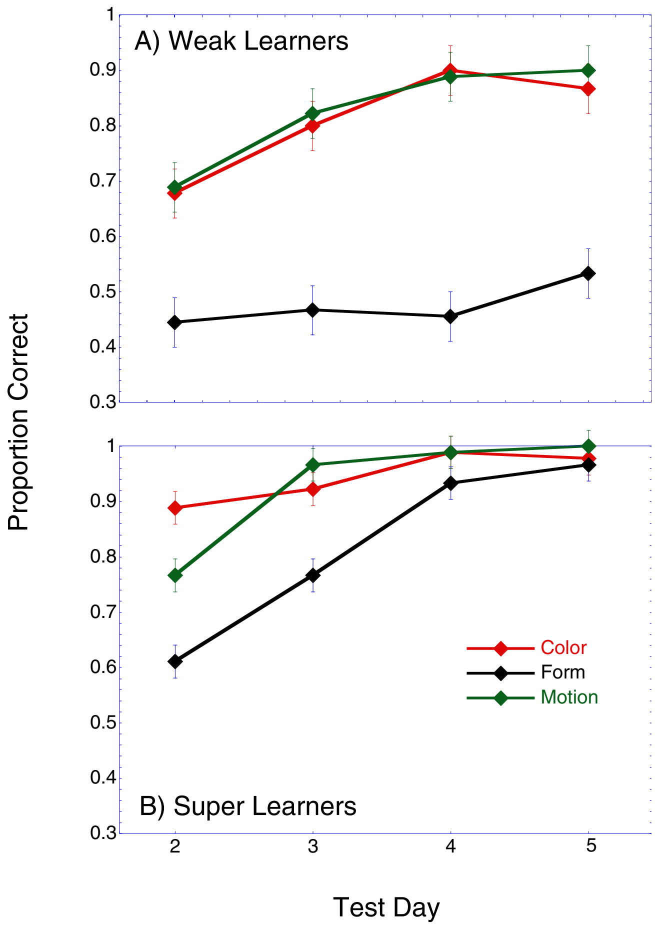

Figure 3. Novel object learning over 4 days of testing. Training occurred on days 1-4. Observers were divided into two groups according to performance on the ambiguous case. A. Weak learners SE, PR, and IB show little or no learning in the ambiguous case, but they were able to exploit situations where segmentation clues were present. Chance performance is .25 for an observer who guesses at all choices with equal probability. However, chance performance is .5 for an observer who always guesses "other" and this is the upper bound of any guessing strategy. Statistics are by analysis of variance (ANOVA) with clue and day as factors. Clue and day effects are significant, F ratio has a *p* value < .0001. Interaction between clue and day is not significant, *p* value = .324. B. "Super observers" AN, WA, and SM provide an existence proof for a powerful bootstrapped learning algorithm. They achieve near perfect performance, even in the ambiguous case. All observers have normal or better acuity and contrast sensitivity. Statistics are by ANOVA. Factors were clue and day. Clue and day effects were significant, with *p* values <.0001. Interaction is also significant, with *p* value < .0001. In an ANOVA comparing groups, supers are significantly better than weak learners, *p* value <.0001.

Testing

Observers were tested on their ability to recognize objects. Test stimuli consisted of a camouflaged object with a background of other camouflaged objects (Figure 1, right). The object of interest was one of the three current training objects or a completely novel object. Each trial was a four alternative forced choice where the choices were "A, B, C, or other." No sounds or segmentation clues were provided and the camouflage patterns and backgrounds changed with each scene. Testing took place on four days. Tests were given 24 hr after each training session and before any new training. Thus, testing was of long term memory.

Results (Experiment 1)

Recall our initial hypothesis that recognition learning cannot occur under conditions of high ambiguity or camouflage, but only when segmentation is possible, and it is only after such opportunistic learning that an observer can recognize or segment a static camouflaged object. Our scenes were constructed so that when segmentation clues were not present, observers could not be certain which elements belong to the object of interest and which belong to background. How do observers perform in learning objects from images with and without segmentation clues?

Figure 3 shows observers' ability to utilize segmentation clues during object learning. However, our initial hypothesis is proven incorrect in general (Figure 3B), because three out of six observers did not depend upon segmentation clues in order to learn new objects. Apparently, they are able to combine information from a number of ambiguous sources to produce a reliable object model. We call this phenomenon bootstrapped learning. Figure 3B shows just how reliable bootstrapped object models can be for some observers.

Methods (Experiment 2)

The surprising result of Experiment 1 was that there exist observers who could learn to recognize and segment objects in the ambiguous training condition. In order to further explore the generality of this finding, we did a second experiment with just the ambiguous training. Further, we want to empirically establish that, prior to training, human observers could not segment the objects with any degree of certainty. Experiment 2 quantified the observers' ability to segment scenes before and after training.

Stimuli

The same stimuli as in Experiment 1 were used, except that the color and motion cases were omitted.

Subject training

Six observers were trained as in Experiment 1, except that there was no training on colored or motion scenes.

Testing

Observers were tested in their ability to recognize and segment objects. Recognition testing was the same as in Experiment 1.

For segmentation testing observers were asked to trace object contours in test scenes. Prior to any training, observers were asked to trace three novel objects in test scenes. These objects were not part of an individual subject's subsequent training or testing set, but were used by the other subjects as part of the balanced experimental design to control for object-specific effects. After 4 days of training, observers were asked to trace the three training objects of interest in novel scenes and were asked to trace three still novel objects in test scenes. Tracing errors were of two types, missed contour segments and extraneous tracing. Tracings were scored by combining tracing path lengths as follows:

$$\text{Trace Score} = (\text{Trace Correct}) / \{(\text{Trace Correct}) + (\text{Trace Missing}) + (\text{Trace Extraneous})\}$$

At Trace Score = .5, the amount of correct tracing equals the amount of tracing error.

Results (Experiment 2)

Example tracings are shown in Figure 4. Prior to training, observers are uncertain of what is the object and what is the background. They may miss the object entirely (Figure 4A) or, because the object of interest is in the foreground, they may trace part of the true boundary. However, part of the background (Figure 4B and 4C) is included as well. After training, both segmentation and recognition performance improve significantly (Figure 5).

Discussion

Over several days, and with relatively few image presentations (20 scenes per object in Experiments 1 and 2), observers are able to learn to recognize and segment camouflaged objects. Further, they were able to do this with images that were highly ambiguous as shown in observers' inability to initially segment any given training image. Evidence of segmentation uncertainty comes not only from the objective consistency score and the subjective tracing tasks, but also from the recognition task. Recognition of objects after ambiguous training is near chance, even after a day of training. If observers had a high confidence in object feature selection or segmentation, they are unable to exploit this during recognition. Given that Experiment 1 shows observers'

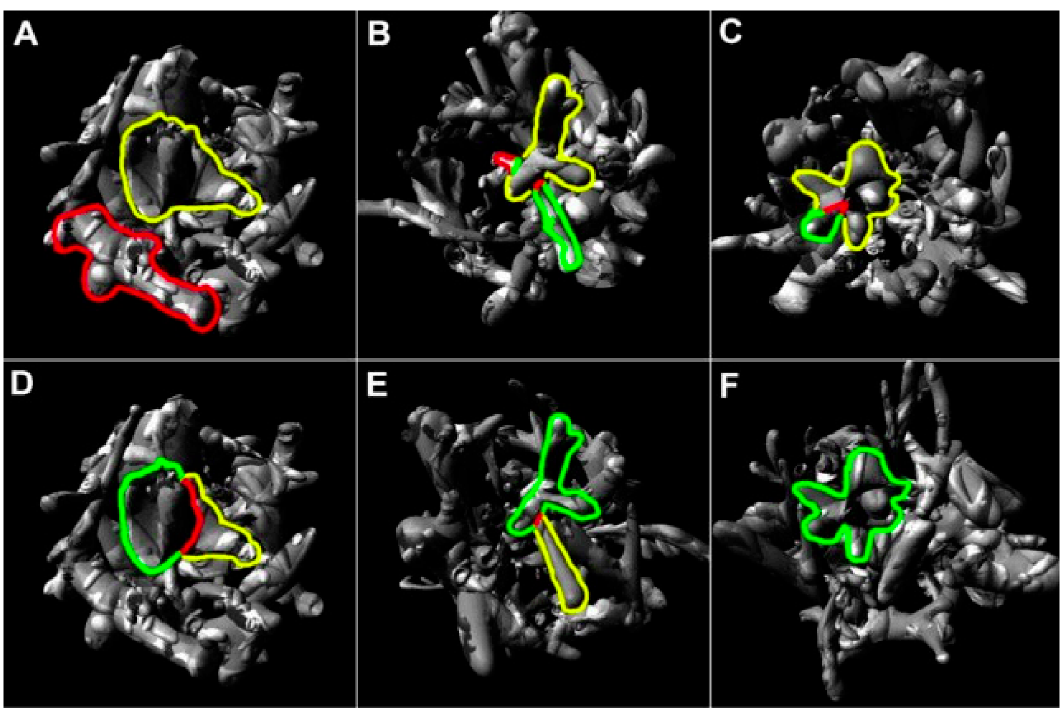

Figure 4. Tracing examples. Green and red segments were originally drawn by the observer and color coded by the experimenter. Green segments are correct and red are extraneous. Yellow segments were missing in the observer's trace and added by the experimenter. A. Observer YH's trace of object D on day 1, prior to any training. Score is 0.0. B. Observer YH's tracing of object A on day 1. Score is .21. C. FF's tracing of object 1 on day 1. Score is .25. D. BK's tracing of D on day 5 after 4 days training on the object. Score is .41. E. KH's tracing of A on day 10 after 4 days training on the object. Score is .60. F. BK's tracing of 1 on day 10 after 4 days training on the object. Score is 1.0.

ability to utilize segmentation information when available, an *inability* to utilize segmentation information would be even more surprising than bootstrapped learning.

Previous psychophysical studies of unsupervised novel shape learning have tended to focus on scenes without background clutter, with components that are segmentable without learning. Fiser and Aslin (2001)

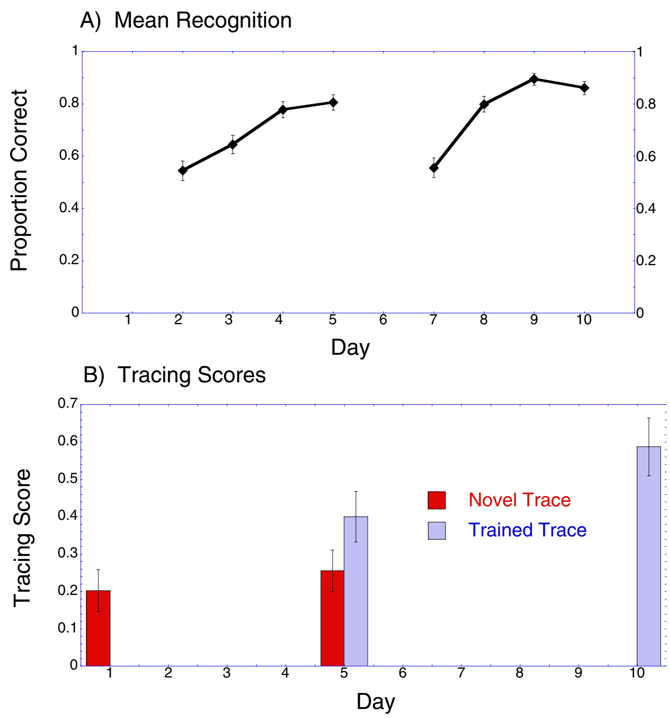

Figure 5. Recognition (top) and tracing (bottom) of camouflaged objects with background and without segmentations clues. Observers start with a new set of objects on day 6. Observers are significantly better at recognizing and tracing familiar objects. A two factor ANOVA of the tracing data, with novelty and subject factors, shows significant effects for novelty (p =.00041), subject (p =.00042), and insignificant interaction (p =.64). An ANOVA was performed on data averaging the two 1-week blocks. The difference between day 1 novel tracing and day 5 novel tracing was not significant in a t test (p =.29). The difference between day 5 trained tracing and day 10 trained tracing was significant in a t test (p =.02) but may be due to either a generalized task training effect or an object effect, because object effects are not controlled for within a single week. For examples of these tracing score values, see Figure 4. Observers learn to recognize objects in spite of an initial inability to segment them. In 5 of 12 day-1 tracings, super observers had a 0.0 tracing score. Segmentation ability evolves in parallel with recognition ability, as expected in the case of bootstrapped learning. There were 4 super learners in Experiment 2, who learned to perfect or near perfect performance, and 1 weak learner, whose performance did not improve by a measurable amount. A sixth observer went from being a weak learner in week 1 to being a super learner in week 2. Data is averaged over super and weak groups. Novel scene tracing at day 5 is a control for general task learning.

showed that human observers can learn shape-composites based on probabilistic co-occurences of potential parts. Bootstrapped learning is distinguished from other types of object learning in that observers learn from scenes where the classification of edge segments into boundary segments and background segments is uncertain.

Computer models of novel object learning with background clutter are rare. However, Weber, Welling, and Perona (2002) have developed an algorithm that can learn uncamouflaged objects in cluttered scenes. This algorithm does not necessarily compute edges, whereas our human observers do in the segmentation tracing task. Shams, Brady, and Schaal (2001) have developed a somewhat similar algorithm, which has shown some ability to learn uncamouflaged digital embryos in background. However, it is unlikely to be able to cope with the degree of camouflage found in the present experiment. Both of these algorithms make use of a constellation of features approach, in which a set of features is collected, along with their relative positions.

We believe that bootstrapped learning for our observers may have been accomplished by a process such as the following:

1. The first image containing object A (image A1) is presented.
2. Features or object parts are extracted from image A1, by a method such as that described by Malik et al., (2001) and Tu and Zhu (2002a, 2002b), and stored in a working memory buffer. Such features must be more subtle than templates of object images or templates of object image parts. (See algorithmic characterization of scenes above, especially Figure 2M-2P.)
3. Working memory content persists while two or more unrelated images (B1, C1, etc.) are processed over a period of 20 s or more.
4. Image A2 is presented.
5. Features are extracted from A2.
6. The intersection of the two feature sets is tested for preservation of relations between features within each image.
7. The resulting subset of features is bound together and stored in long-term working model memory as an evolving model. This long-term working model memory can persist for at least 24 hr.
8. Steps 1-7 are repeated for images A2, A3, etc., except that an evolving high-level model is now available to help segment relevant features via a top down mechanism.

Any algorithm used by the bootstrapping observers must depend upon two memories: the working memory buffer and the long-term working model memory.

A major challenge for machine vision research is to create a system that can learn to recognize objects from example images. But, which way should one proceed? The target system might need to be opportunistic and subsist on a diet of well-segmented scenes. These scenes would have to be prepared manually and therefore require a great deal of human labor. Alternatively, the opportunistic system might be placed in a rich environment and simply wait for well-segmented images to occur. This could take a considerable amount of time and require its own form of automation. Fortunately, we have been able to demonstrate the existence of a learning algorithm, which does not depend on special opportunities. Instead, it proceeds directly to learn objects given only the most ambiguous, yet commonly occurring, images of those objects. By defining what information is required by an object learner, machine vision researchers are able to pursue one avenue to machine object learning rather than several.

Conclusions

Natural scenes tend to be highly cluttered, which presents a challenge to observers learning to see new objects. Yet there are opportunities when segmentation may be easier, such as when color or motion segmentation clues are present. Our study began with the hypothesis that learners of novel objects would necessarily rely on such opportunities to overcome segmentation problems, especially when dealing with severely camouflaged objects. We found, on the contrary, that there exist two routes to object recognition. One is opportunistic whereas the other relies on bootstrapped learning.

Appendix A. Method Details

Six observers participated in each experiment. All were 20/20 or better on the standard Lighthouse test at 4 m and in a modified Lighthouse at the experimental viewing distance of 61 cm.

Images were 18-cm square and viewed on an iMac computer at 72-dpi resolution.

Images were generated using the Inventor library on a Silicon Graphics computer. The position of objects of interest were randomized at an (x,y) location with mean $(x,y,z) = (0,0,0)$. The (x,y) position of the object of interest varied ±.375. In each scene, 21 background objects had (x,y) coordinates ±1.5 and z coordinates from -1.8 to -9. The projection was perspective with the virtual camera at (0,0,5). Units are arbitrary Inventor units.

Lighting was directional with fixed direction vector (1, -1, -1). Objects were rendered with Phong shading using Inventor parameters diffuse color = (.8, .8, .8), specular color = (1, 1, 1), ambient color = (.5, .5, .5), and shininess = 1.

Texture wrapping uses Inventor's default method. First, the bounding box for each object is computed. Next, the texture image is projected onto each side of the box and then onto the object's polygons.

Each work week (5 days) observers were trained on day 1, tested for recognition and then trained on days 2-4, and they were tested for recognition on day 5. This was repeated for each segmentation clue type in Experiment 1.

In Experiment 2, observers performed (a) a novel tracing task on day 1, week 1; (b) a novel and a familiar object tracing task on day 5, week 1; and (c) a familiar tracing task on day 5, week 2. During each week, the observers used a different set of three training objects. Object sets were permuted evenly among observers to control for object difficulty and ordering effects.

Appendix B. Consistency Algorithm

This program measures the consistency with which object regions in a series of n images appear, relative to random background fragments. In the interest of modeling the iterative nature of human learning, the algorithm collects information one image at a time and consolidates it with a current assessment of image consistency.

The algorithm first takes the pixel-wise log of each image. It then computes a sequence of composite images. Composite pixel intensity value $c(k,i,j)$ at (i,j) in the k^{th} composite is defined recursively as

$$c(k,i,j) = \frac{(k-1)c(k-1,i,j) + p(k,i+tx(k),j+ty(k))}{k} \quad ,k = 1 \ldots n$$

where $p(k,i,j)$ is a pixel intensity value in the k^{th} training image and $(tx(k),ty(k))$ is a translation of the k^{th} training image. $c(1,i,j)$ is simply $p(1,i,j)$. $(tx(k),ty(k))$ is chosen so as to minimize

$$\sum_{i,j} \left| c(k-1,i,j) - p(k,i+tx(k),j+ty(k)) \right|$$

The final translated mean image M is simply the n^{th} composite, where n is the number of training images. Let the mean within M be $\mu(M)$ Background pixels will tend toward this mean. Therefore, any pixels, which tend toward the extremes are likely to be on object fragments. To visualize this, compute a new image

$$E = \text{abs}(M - \mu(M))$$

A threshold may be applied to select candidate object fragments, or as we have done in this paper, the image may be pseudo-colored to portray numerous possible thresholds at once.

Appendix C. Growing Digital Embryos

Digital embryos are generated using simulated hormonal diffusion, simulated physical forces, and polygon fission. These operations are applied repeatedly to an evolving polyhedron. Any polyhedron can be used as a starting shape. In the current application, a regular icosahedron was used.

Two loops operate concurrently. One loop controls the lifecycles of morphogen secreting cells and morhogen diffusion among cells. A second loop controls cell division and simulates the physical dynamics of the cells. Cells are represented by vertices in the polyhedron geometry.

The morphogen secretor lifecycle and loop is simple. A fixed number of vertices (3 in the current experiment) are maintained as morphogen secretors. Each generator is assigned a finite lifespan at random. At the end of a particular generator's lifespan, it is replaced by another generator somewhere else on the surface of the embryo. The location is determined at random.

Active morphogen secretors retain a fixed high morphogen concentration, which diffuses to adjacent calls or vertices. Morphogen flows between vertices, which are connected by edges. The flow rate is proportional to the difference in the concentrations of the adjacent vertices. There is also a constant leakage of morphogen, out of each cell, into the surrounding fluid. The result of these effects is that the morphogen concentration in any nonsecreting cell i, with neighbors j is

$$C_{i,t+1} = (1-L)\left(C_{i,t} + \frac{R\sum_j \left(C_{j,t} - C_{i,t}\right)}{n} \right)$$

at time t+1. R is a diffusion rate and n is the number of vertices connected to vertex i.

The polygon fission operation proceeds as follows: All polygons in the present implementation are triangles. A triangle is marked for fission if the average morphogen concentration of its constituent vertices is above some threshold. The triangle is split into four new triangles as shown in Figure 6. After fission, vertex I is a full-fledged vertex but vertices K and J are not. They cannot be allowed to move as a normal vertex would because it might cause triangles AED and DFC to become quadrangles, and non-planar ones at that. Therefore, vertices K and J remain dependent vertices. What this means, in the case of K, for example, is that K must remain on a line between D and E regardless of what forces act upon it. K will be promoted to a nondependent vertex when AED is split.

Vertices move about in space according to the sum of forces that act upon them. The amount of motion per time increment is proportional to the magnitude of the force, whereas the direction of motion is determined by the total force vector. All vertices in an embryo repel all other vertices according to an inverse square law. At the same time, vertices, which are attached by an edge, are attracted according to Hooke's law.

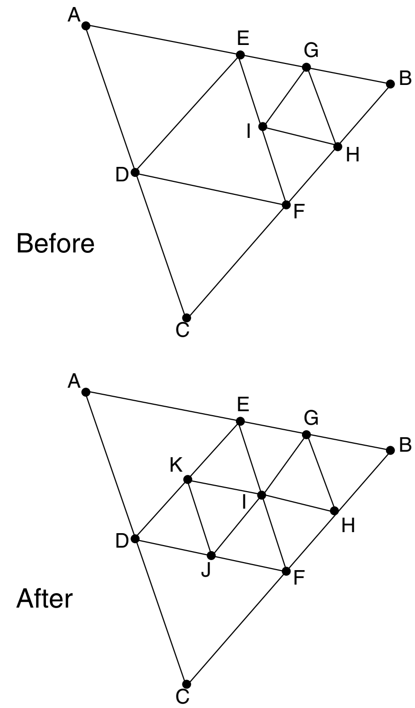

Figure 6. Triangle DEF before and after fission. DEF will eventually be replaced by KEI, IFJ, JDK, and KIJ. However, DEF may persist for awhile as the neighbor of AED and DFC.

Acknowledgments

We thank an anonymous reviewer for suggesting the log transform in the consistency algorithm. This research was supported by National Institutes of Health Grant RO1 EY02857. Commercial relationships: none.

References

Amit, Y., & Geman, D. (1999). A computational model for visual selection. *Neural Computation, 11*(7), 1691-1715.[PubMed]

Ball, P. (1999). *The Self-Made Tapestry* (1st ed.). Oxford: Oxford University Press.

Ben-Shahar, O., Huggins, P. S., & Zucker, S. W. (2002). On computing visual flows with boundaries: The case of shading and edges. In H. H. Bülthoff, S.-W. Lee, T. Poggio, & C. Wallraven (Eds.), *Biologically motivated computer vision, BMCV 2002,* (2525 ed.). Berlin: Springer-Verlag.

Braddick, O. J. (1974). A short-range process in apparent motion. *Vision Research, 14*, 519-527. [PubMed]

Brady, M. (1999). *Growing Digital Embryos*, from http://gandalf.psych.umn.edu/~kersten/kersten-lab/camouflage/degrowqt.html

Brady, M., & Kersten, D. (2000). *The Camouflage Challenge*, from http://gandalf.psych.umn.edu/~kersten/kersten-lab/camouflage/camouflageChallenge.html

Brady, M. J. (1999). *Psychophysical investigations of incomplete forms and forms with background.* Ph. D. Thesis. University of Minnesota, Minneapolis. [Link]

Brown, J. L. (1975). *The Evolution of Behavior*. New York: W. W. Norton.

Fiser, J., & Aslin, R. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological science, 12*(6), 499-504. [PubMed]

Lamme, V. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience, 15*(2), 1605-1615. [PubMed]

Malik, J., Belongie, S., Leung, T., & Shi, J. (2001). Contour and texture analysis for image segmentation. *International Journal of Computer Vision, 43*(1), 7-27.

Murray, J. D. (1993). *Mathematical Biology* (2nd, corrected ed. Vol. 19). Berlin: Springer.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature, 317*, 314-319. [PubMed]

Shams, L., Brady, M. J., & Schaal, S. K. (2001). Graph matching vs. mutual information maximization for object detection. *Neural Networks, 14(3),* 345-354. [PubMed]

Thayer, G. (1909). *Concealing-Coloration in the Animal Kingdom*. New York: Macmillian.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522. [PubMed]

Tu, Z., & Zhu, S. (2002a). Image segmentation by data-driven markov chain monte carlo. *IEEE Tranactions on Pattern Analysis and Machine Intelligence, 24*(5), 657-673.

Tu, Z., & Zhu, S. (2002b, May). *Parsing images into region and curve processes.* Paper presented at the 7th European Conference on Computer Visions, Copenhagen, Denmark.

Weber, M., Welling, M., & Perona, P. (2000). *Unsupervised learning of models for recognition.* Paper presented at the 6th European Conference on Computer Vision, ECCV2000, Dublin, Ireland.

Yuille, A. (1991). Deformable templates for face recognition. *Journal of Cognitive Neuroscience, 3*(1), 59-70.