The Role of Image Understanding in Contour Detection

C. Lawrence Zitnick
Microsoft Research, Redmond
larryz@microsoft.com

Devi Parikh
Toyota Technological Institute, Chicago (TTIC)
dparikh@ttic.edu

Abstract

Many cues have been proposed for contour detection or image segmentation. These include low-level image gradients to high-level information such as the identity of the objects in the scene or 3D depth understanding. While state-of-the-art approaches have been incorporating more cues, the relative importance of the cues is unclear. In this paper, we examine the relative importance of low-, mid- and high-level cues to gain a better understanding of their role in detecting object contours in an image. To accomplish this task, we conduct numerous human studies and compare their performance to several popular segmentation and contour detection machine approaches. Our findings suggest that the current state-of-the-art contour detection algorithms perform as well as humans using low-level cues. We also find evidence that the recognition of objects, but not occlusion information, leads to improved human performance. Moreover, when objects are recognized by humans, their contour detection performance increases over current machine algorithms. Finally, mid-level cues appear to offer a larger performance boost than high-level cues such as recognition.

1. Introduction

Segmentation and the related task of contour detection are being leveraged for an increasingly wide variety of computer vision tasks. For instance, segmentation has been used for object recognition [2, 25, 10, 29, 33], optical flow estimation [35], stereo [31], and image compositing [27]. In this paper, we focus on the task of object level contour detection and segmentation.

A fundamental question is the degree to which high-level information is necessary when performing object segmentation [33, 10, 16]. For instance, is object segmentation possible even if the objects in the scene cannot be recognized? How important is the 3D understanding [12, 15] of a scene in segmenting its objects? It is commonly assumed that object boundaries correspond to changes in color or texture [3, 17, 6] (Figure 1(a)). However, boundaries referred to as illusory contours [34] may not be visible (Figure 1(c)), and many color or texture edges correspond to albedo edges and not actual boundaries (Figure 1(b)). Solving these ambiguities may require mid-level information such as contour reasoning [19, 36, 8] or the use of Gestalt laws [14] for perceptual grouping. Finally, it may be necessary to reason at the object level to correctly determine object boundaries [37, 16]. For example, when segmenting a dark horse with white feet the feet are commonly missing when only using low and mid-level cues (Figure 1(d)).

Figure 1: Illustrative examples of (a) an easy to segment object, (b) object with albedo edges, (c) illusory contours and (d) a difficult object to segment without object level knowledge.

Previous works have proposed the use of various cues and their combinations for image segmentation, each with varying amounts of low-, mid- and high-level information. However, the relative importance of these cues is less understood. In this paper, we study the relative importance of these cues to help provide guidance for the future development of contour detection and segmentation algorithms. These cues include low-level information such as color edges, mid-level information related to contours and textures, and high-level information such as object recognition and occlusion reasoning. In this paper, we use the term "mid-level" to refer to non-local gradient, texture and edge information that is commonly used by state-of-the-art contour detection and segmentation approaches. For example, this may include finding long smooth contours [19, 36, 8], or using Gestalt laws [14]. We do not use the term "mid-level" to refer to semantically meaningful cues, such as figure-ground information, object attributes, etc.

We perform our analysis using numerous human studies and machine experiments described in Section 3. Specifically, we address the problem of object boundary detection while varying the amount and type of information available. For instance, we can control the amount of local informa-

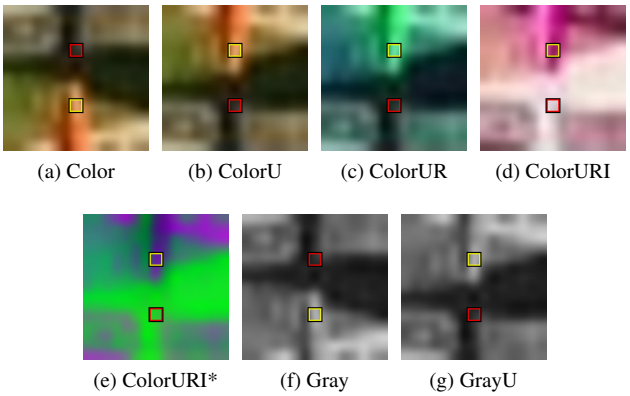
tion by varying the size of the visible patch surrounding the potential object boundary [11]. We may also manipulate the type of information shown by only displaying intensities or by rotating the color channels. For each of these tasks, the difficulty of recognizing the objects and their occlusion relationships varies providing insights into their relative roles. While it is difficult to directly infer causality relationships, as shown in Section 4, correlations between contour detection, low-level information, object recognition and occlusion reasoning can be found.

Our studies support three hypotheses. First, further research on low-level cues may not yield improved segmentation results. Our studies show that humans do not outperform state-of-the-art segmentation and contour detection methods using only small image patches. This supports the earlier findings of Fowlkes [11]. Second, the recognition of objects leads to a notable improvement in contour detection accuracy, while occlusion information is less essential. Assuming there are no confounding factors, our experiments show that the improvements in contour detection accuracy due to larger patch sizes is caused in part by improved recognition of the objects in the patches. Finally, while recognition of objects leads to a significant improvement in contour detection, a larger performance boost is gained from the increase in mid-level information as the analyzed image patches gain in size.

2. Previous work

In this section, we describe previous works using human subjects to study segmentation and contour detection, as well as, various works that have used segmentation-based approaches to object recognition and discovery.

The problem of image segmentation and contour detection are closely related. Rivest and Cavanagh [26] studied various sources of information used by humans for contour localization. Their findings supported the hypothesis that information related to luminance, color, motion and texture are integrated at a common site in the brain. Closely related to our work, Grady *et al*. [13] find that if given a bounding box, high level semantic information is not needed for humans to find a consistent segmentation, and conclude the problem is well-posed. Fowlkes [11] studied the performance of humans at depth boundary detection with varying window sizes. He found machines were roughly equivalent to humans when shown grey-scale patches. A large database of human labeled boundaries was collected and analyzed by Martin *et al*. [17]. Each contour segments a region into foreground and background objects, known as figure-ground labeling. Fowlkes *et al*. [12] study the problem of figure-ground labeling given both local and global information, and find that image luminance does provide additional information over just the knowledge of the depth boundary shape. Peterson [24] hypothesizes that object



(a) Color (b) ColorU (c) ColorUR (d) ColorURI

(e) ColorURI* (f) Gray (g) GrayU

Figure 2: Example patch visualizations from our seven scenarios for our human studies (best viewed in color).

recognition may proceed figure-ground organization in humans. McDermott [18] studied how contour junctions may be detected. The role of motion parallax in segmentation was studied by [38]. Related to segmentation, numerous studies have addressed the problem of selective attention in humans [7]. Recently, the use of human studies has been applied to several computer vision problems to help in understanding the challenges that remain. These include recognizing objects in tiny images [32, 23], the tradeoffs between features, learning algorithm and the amount of training data [21], the roles of local and global information in images [20] and person detectors [22]. In this paper, we conduct human studies to understand the relative importance of low-, mid- and high-level information for contour detection.

Since there are numerous works using segmentation and contours for object recognition, we only reference a few representative examples here. Belongie *et al*. [2] proposed using contours for recognition, while Rabinovich *et al*. [25] used segments as the basis for contextual reasoning between objects in an image. Segmentation approaches to object discovery were proposed by Sivic *et al*. [30] and Lee and Grauman [16]. Shotton *et al*. [29], Ferrari *et al*. [10], and Tu *et al*. [33] all proposed algorithms to jointly segment and recognize objects in an image. In this paper we systematically quantify the dependence between segmentation and high-level tasks such as recognition.

3. Experimental setup

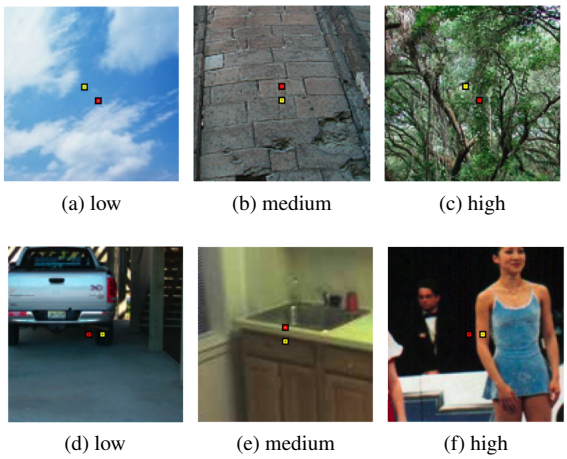
In this paper, our goal is to examine the role of low-level, mid-level and high-level information in object segmentation. Instead of measuring segmentation performance directly, we measure the accuracy of contour detection, i.e., the accuracy of the segment boundaries, as done by previous studies on machine segmentation performance [17, 1]. The detection of contours has the additional advantage in that it

is possible to perform the task using only local information. This allows us to vary the amount of local (low-level) vs. global (high-level) information available during our human studies. However, a set of contours does not necessarily create an image segmentation, since they may not form closed loops. A few missing contours can also join two segments creating a poor segmentation of an image. Despite these drawbacks, measures of contour detection accuracy have been shown to correlate well with segmentation accuracy, and approaches have been proposed to create segmentations from possibly incomplete contours [8, 1].

In the next section we discuss the various machine algorithms used in our experiments, followed by a description of the experimental setup for our human studies.

3.1. Machine experiments

We experiment with six different machine approaches to generating contours, ranging from naive to state-of-the-art methods. Two of these are based on low-level gradient information, while the other four are unsupervised segmentation-based approaches. For the segmentation-based approaches, we use the authors' publicly available implementation. **Gradient:** The first approach is a naive method based on local gradient information. Contours are detected simply by thresholding the magnitude of the gradients at a pixel. **Canny:** The second approach detects contours in the image by running the classical canny edge detector [3]. We used the in-built MATLAB implementation. **Ncut:** This is the segmentation-based approach of [28] that builds a graph on neighboring pixels. The weights of the edges between pixels are determined using the intervening contours cue that depends on the magnitude of the color gradients between pixels. The graph is partitioned using normalized cuts, resulting in a segmentation of the image. **Mean Shift:** This segmentation-based approach uses the mean-shift clustering algorithm for image segmentation [5]. Each pixel is represented by its five dimensional position in color and location. The modes in this five dimensional space that are found using the mean-shift clustering algorithm correspond to a clustering of the image. **FH:** This segmentation-based approach by Felzenszwalb and Huttenlocher [9] is an efficient graph-based approach that defines edge-weights based on not just the gradient across the edge, but relative to gradients observed in a neighborhood. This allows the method to capture perceptually important non-local aspects of the image. It can preserve detail in low-variability image regions while ignoring detail in high-variability regions. **UCM:** This segmentation-based approach uses a hierarchical representation of the image called Ultrametric Contour Maps [1]. It integrates local contour cues along the regions boundaries and surrounding region attributes. Recent results show this approach to achieve state-of-the-art performance [1].



(a) low (b) medium (c) high

(d) low (e) medium (f) high

Figure 3: Examples of low, medium and high gradient patches (largest size) from our Contour patch dataset with (top) no object contour and (bottom) with object contour.

3.2. Human experiments

Our human studies measure the accuracy of our subjects on contour detection, recognition, occlusion (*i.e.* depth boundary) detection and figure-ground labeling tasks on image patches. We perform these tasks under numerous scenarios with varying amounts and types of information available. Each of these scenarios was designed to help separate the influence of different types of low-, mid- and high-level information. For instance the patch size can control the amount of mid and high-level information available. The knowledge of high-level information such as the recognition of objects can be reduced by flipping the patch upside down or by manipulating the color information, with a minimal affect on mid-level non-semantic information.

For each scenario, we showed subjects a patch with small red and yellow squares equidistant from the center, as seen in Figure 2. We asked them several questions related to the red and yellow squares pertaining to contour detection, depth boundary detection, figure-ground assignment and recognition. Specifically, the questions were:

- Do the red and yellow squares lie on the same object, or different objects? The possible answers were: "Same object" or "Different object".
- Is the object under the red square in front of or behind the object under the yellow square? The possible answers were: "Red in front of Yellow", "Yellow in front of Red" and "Neither".
- Which object does the red square belong to? Subjects were to provide a one word free form answer.
- Which object does the yellow square belong to? Again, subjects were to provide a one word free form answer.

.

We conducted these human studies on Amazon's Mechanical Turk. Each subject answered the questions above for 6 patches at the same time. Reasoning about multiple patches for a task before moving on to the next task limits the influence of one task on the other. Getting responses for all tasks pertaining to a patch from the same subject reduces inter-subject variabilities. In our experiments, each patch was assigned to 10 unique subjects. Since the experimental setup and viewing environment cannot be fully controlled using Amazon's Mechanical Turk, these studies are meant to merely serve as a lower bound on human performance.

We presented the patches using seven different scenarios, each with a different visualization to vary the low-level information available (color vs. gray-scale) and the ease with which high-level information can be inferred (manipulating spatial layout and color information). (1) **Color:** A regular RGB patch was shown. (2) **ColorU:** The color patch was flipped upside down before presenting it. (3) **ColorUR:** The RGB channels of the patch were rotated to be BRG, and the patch was flipped upside down before presenting it. (4) **ColorURI:** The RGB channels were rotated and inverted. So the three color channels became 255-B, 255-R and 255-G. The patch was flipped upside down before presenting it. (5) **ColorURI*:** The color channels were rotated, and only one of the channels was inverted. So the three color channels became B, 255-R and G. The patch was flipped upside down before presenting it. (6) **Gray:** An upright patch was presented in gray scale and (7) **GrayU:** A gray scale patch was presented after flipping it upside down. Examples of these visualizations can be seen in Figure 2.

4. Results

We now describe our contour patch dataset, and the results of the machine experiments and human studies.

4.1. Contour patch dataset

We build our contour patch dataset from a subset of 185 images in the SUN dataset [4]. The dense and detailed object segmentations in the SUN dataset are appropriate for our study of object boundary detection. We did not use the popular Berkeley Segmentation Dataset [17] since many of the labeled contour boundaries correspond to changes in albedo and not object boundaries. We extract patches from a total of 240 locations across these images. Half of these locations fall on an object boundary as per the SUN ground truth annotations, and the other half do not have an object boundary within a 15x15 pixel neighborhood. To obtain a varied distribution of patch and contour types, as shown in Figure 3, a third of the locations have low-gradients i.e. they have at most a gradient magnitude of 10 in the surrounding 7x7 patch. Another third of the locations have medium-gradients i.e. they have a gradient magnitude between 10 and 30, and none of the pixels in the 7x7 neighborhood have a gradient magnitude higher than 30. The last third of the locations have high-gradients with a gradient magnitude of at least 30 in a 7x7 neighborhood. Since the object boundaries are occasionally not well localized in the SUN dataset, the presence or absence of a contour was verified using Amazon's Mechanical Turk. Of the original 240 patches, 196 were verified and used for our experiments. We extract patches at seven different sizes centered at each of the 196 locations: 7x7, 9x9, 15x15, 25x25, 33x33, 63x63 and 127x127. This results in a total of 1372 patches in our dataset. Each patch is presented using seven different visualizations to 10 subjects, resulting in about 96k responses to each of our four questions. The dataset is available on the authors' website.

4.2. Machine algorithms

We detected contours in the 185 images using the six machine approaches described above. If a contour boundary was detected within a 5x5 window of the central pixel, the patch was considered to have a contour boundary, and otherwise not. This is to account for any small location errors in the SUN dataset, and to mirror the human studies as closely as possible, which had the red and yellow squares separated by 5 pixels. The parameters for the various segmentation algorithms were selected by optimizing their contour detection performance on an independent set of patches extracted from a disjoint set of images from the SUN dataset. The threshold for the gradient-based contour detector was set to 20. We used the default parameters for canny (automatically determined threshold and sigma = 1). We generated 10 segments for every image using normalized cuts. For the mean shift segmentation algorithm, the parameters were SpatialBandwidth = 7, RangeBandwidth = 6.5 and MinimumRegionArea = 200. The parameters for FH were set to sigma = 1, K = 500, min = 200. The parameter for UCM was set to k = 0.1.

The results of the machine tests on our dataset for contour detection can be seen in Figure 6. The local edge based methods, gradient and Canny perform relatively poorly with 58.3% and 63.9% respectively. The segmentation approaches, mean shift (74.2%), FH (74.2%), and UCM (74.7%) all perform roughly the same with normalized cuts (67.0%) doing worse. Clearly, the mid-level information used by the segmentation approaches provides additional accuracy. We provide a comparison between machine and human accuracy in the following section.

4.3. Human Studies

Before we describe our results for the human studies, we describe how human accuracies for the different tasks are computed. The SUN dataset does not supply ground truth for depth boundary detection or figure-ground labeling. For these tasks, we use the majority vote response of

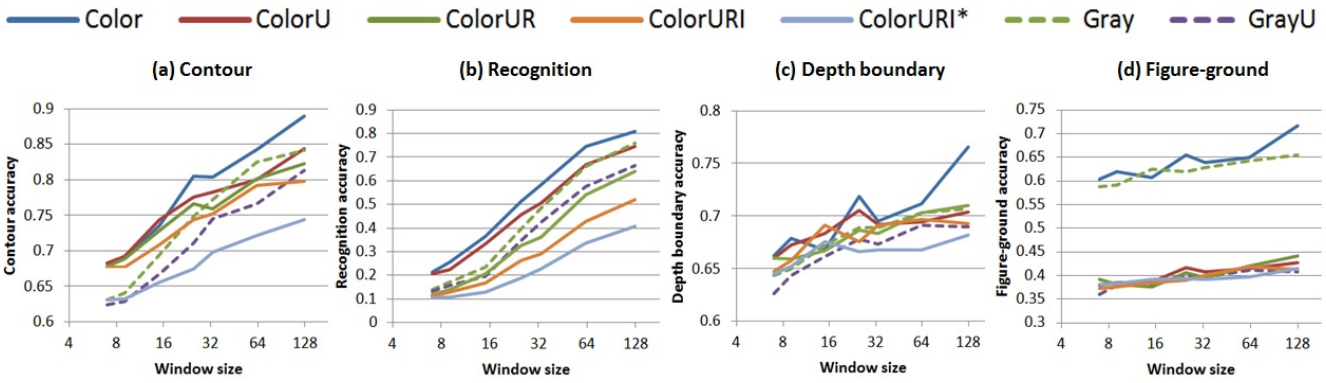


Figure 4: Accuracies of (a) contour detection, (b) recognition, (c) depth boundary detection and (d) figure-ground labeling across window sizes for all scenarios.

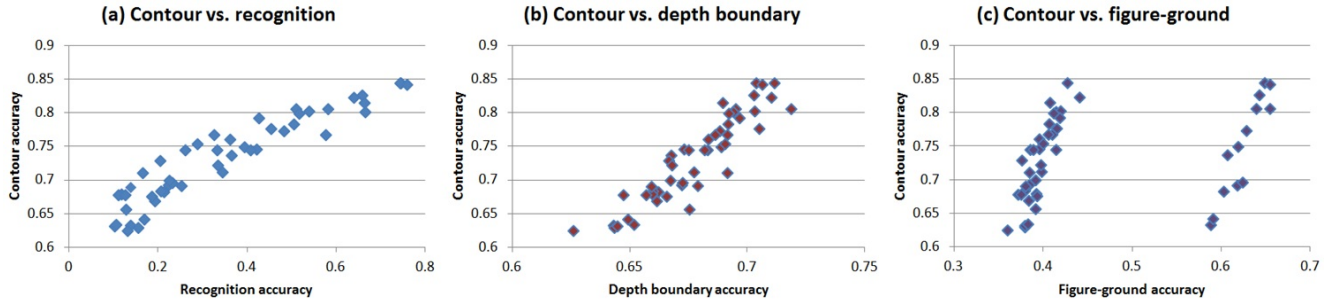


Figure 5: Scatter plots of contour detection accuracies vs. (a) recognition accuracies (corr = 0.944), (b) boundary detection accuracies (corr = 0.914) and (c) figure-ground labelings (corr = 0.378) for all scenarios and window sizes.

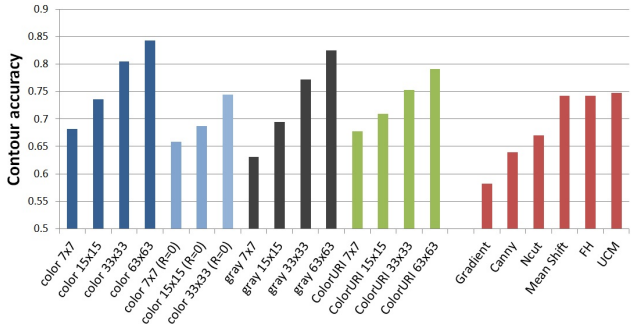


Figure 6: Graph showing the accuracies of various machine approaches (red) compared to human accuracies on the Color scenario (dark blue), the Color scenario when an object in the patch was not recognized (light blue), Gray scenario (gray), and ColorURI scenario (green).

our subjects at the largest patch size (127x127) using the Color visualization as ground truth. For the depth boundary detection task, we ignore the polarity of the response (red in front of yellow or yellow in front of red), and simply check if the subject correctly detected the presence of a depth boundary or not, i.e., did he choose "neither" or not. To measure recognition accuracy, we generated a ground truth dataset of object labels by gathering 30 responses in addition to our original 10 responses to name objects under the red and yellow squares using (127x127) patches with the Color visualization. The freeform answers provided were compared to the answers provided by the original subjects. If corresponding words were found, ignoring capitalization and punctation, the objects were said to be recognized. If at least one object was recognized correctly in the patch, it was labeled as recognized.

We now analyze the results of our human studies. We begin by providing an overview of the results and analyze the accuracies given low-level and mid-level cues. Next, we address the dependencies between contour detection and image understanding, i.e., recognition, depth boundaries (occlusion) and figure-ground labels. Finally, we compare humans accuracies to machines.

An overview of our human studies results can be seen in Figures 4 and 5. Figure 4 plots the contour detection accuracy, recognition accuracy, depth boundary detection accuracy, and figure-ground labeling accuracy vs. the patch size for all scenarios. In each case, the accuracies increase with the patch size visible to the subjects. Figure 5 shows three scatter plots of contour detection accuracy vs. recognition accuracy, depth boundary detection accuracy, and figure-ground labeling accuracy. Each data point corresponds to a scenario and window size. A strong correlation is visible

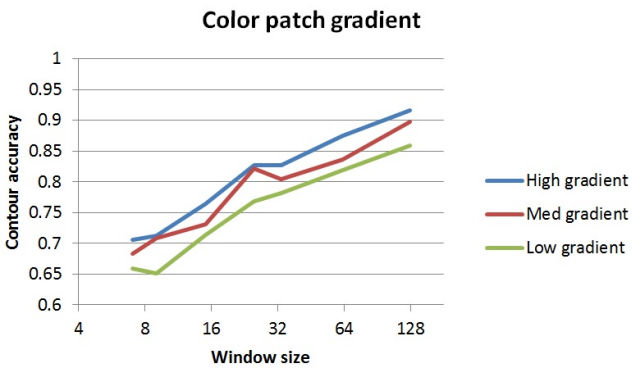


Figure 7: Plot of contour detection accuracies across patch sizes for the human study Color scenario with high, medium and low gradient patches.

Figure 8: Contour accuracies conditioned on correctly recognizing an object and correctly detecting the presence of a depth boundary across window sizes averaged over all scenarios. (blue, R=1) an object is correctly recognized, (red, R=0) not recognized, (solid, D=1) depth boundary correctly labeled, (dotted, D=0) depth boundary incorrectly labeled.

between the accuracy of contour detection and both recognition accuracy and depth boundary detection accuracy.

In Figure 7, we show the contour detection accuracy for the patches with low, medium an high gradients for the Color scenario. Accuracy typically varies by 5% between patches with low and high contrast. Other scenarios show similar differences.

Low-level We can study the performance of contour detection using low-level cues by analyzing the accuracies using small windows in Figure 4(a). The results can be clearly split into two groups. The scenarios in which the relative color information is kept, Color, ColorU, ColorUR and ColorURI, all have higher accuracies than the scenarios without color, Gray and GrayU, and ColorURI*, which does not maintain relative color information. It is interesting to note that low-level contour detection appears to be invariant to color rotation, and the patch being flipped upside down.

Mid-level We refer to non-local and non-semantic contour and texture information as "mid-level information." We do not use the term "mid-level" to refer to figure-ground or object attribute information. In Figure 4, the contour detection accuracies increase with window size, but it is unclear whether this increase is due to more mid-level cues being visible or to better image understanding using high-level knowledge. To separate the effect of high-level knowledge from mid-level cues, we plot in Figure 8 the contour detection accuracies conditioned upon whether at least one object was recognized correctly (R=1) or not (R=0). We also conditioned on whether the presence of a depth boundary was correctly labeled (D=1) or not (D=0). The results are averaged across all scenarios. In each of the four possible cases, the accuracies increase substantially with window size by about 15%. For example, even if an object was not recognized (R=0) and the depth boundary was not correctly labeled (D=0), contour detection accuracies still increased with window size from 64% to 78%. If the increase in accuracy with window size in Figure 4 were solely do to recognizing objects or detecting depth boundaries we would expect the curves in Figure 8 to be flat. This provides strong evidence that mid-level cues are important for object segmentation, assuming there are no other significant sources of high-level information beyond recognition and depth understanding that led to the observed increase.

Depth boundaries In Figure 5(b), the knowledge of a depth boundary within the patch appears strongly correlated (corr = 0.914) to contour detection accuracy. However, in Figure 8 when the contour detection accuracy is conditioned on labeling the presence of a depth boundary correctly (solid vs. dotted lines) there is only a negligible difference. Specifically, the contour detection accuracy is approximately the same regardless of whether the subject correctly labels the presence of a depth boundary (solid green line) or not (dotted green line). Hence, the correlation observed in Figure 5(b) may be primarily due to both depth boundaries and contours being easier to detect with larger patch sizes. In Figure 5, the change in depth boundary detection accuracy (65% to 72%) across patches sizes and scenarios is also quite small with respect to the changes in contour detection accuracy (65% to 85%.) These observations make it unlikely that the understanding of depth had a significant impact on the accuracy of the detected contours in our studies.

Figure-ground Unlike depth boundary detection, figure-ground knowledge does not appear to be strongly correlated (corr = 0.378) to contour detection accuracy as shown in Figure 5(c). The use of patches flipped upside down in some scenarios and right side up in others, resulted in two distinct groupings in Figure 5(c). In upside down patches, the shading information available might lead to the wrong figure-ground interpretation, since humans typically assume a scene is lit from above. However, this misinterpretation does not appear to adversely affect contour detection per-

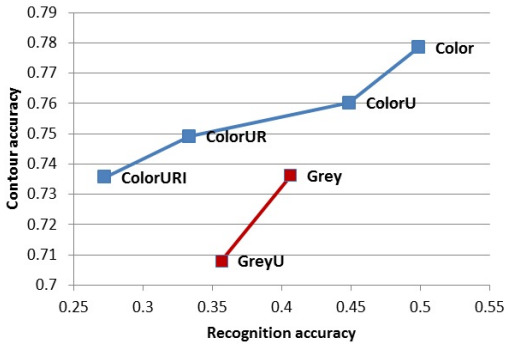


Figure 9: Trends in contour detection accuracy across scenarios. The scenarios in which the patches are easier to recognize achieve the highest contour detection accuracies.

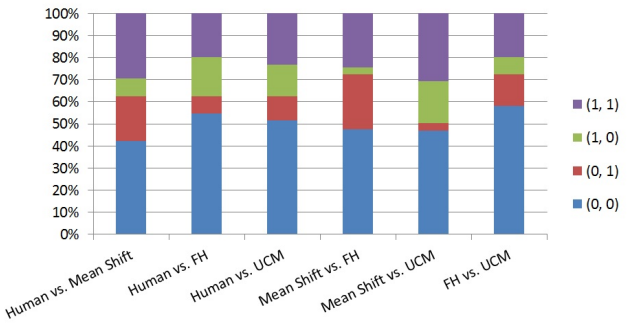


Figure 10: Illustration of the confusion matrices for various human and machine experiments for contour labeling: (purple) both label with contour, (blue) neither label with contour, (green, red) the labelings disagree.

formance.

Recognition The role of recognition in contour detection accuracy is quite interesting. Figure 5(a) shows a strong correlation (corr = 0.944) between recognition accuracy and contour detection accuracy. Unlike the results conditioned on correctly labeling the depth boundary, Figure 8 shows a consistent increase in contour detection accuracy (red vs. blue line) when the subjects correctly recognize at least one object in the patch across window sizes. From these figures it is unclear whether their is causal relationship or just a correlation between recognition and contour detection. That is, does the knowledge of the objects in the patch help to determine whether a contour is present? Or is it just that patches in which objects can be recognized are also easy to segment? To test the causality argument we can examine the average results across various scenarios. For instance, in Figure 9 we plot the average recognition and contour detection accuracy for four color patch scenarios (Color, ColorU, ColorUR, and ColorURI) shown by the blue line, and two gray patch scenarios (Gray, GrayU) shown by the red line. In each of these cases, it can be argued that the low-level and mid-level information remains constant since the patches are just being flipped or the colors rotated or inverted. What does change is the ease in which the objects in the patch are recognized. Assuming there are not other confounding variables, the positive sloping lines in Figure 9 strongly support a causal relationship between recognition and contour detection. Figures 8 and 9 suggest that recognition of an object in the image patch can increase contour detection accuracy from 4% to 6%.

Human vs. machine Figure 6 shows the contour detection performance of various machine algorithms vs. human performance on various scenarios. The segmentation-based machine algorithms perform quite well with respect to humans, achieving similar performance to humans with 15×15 color patches. Human detection using smaller patches is similar to using Canny edge detection. If we examine patches in which the human subjects did not recognize objects (R=0) or in which the patches are hard to recognize (ColorURI), machine performance is similar to human performance on 33×33 patches. However, humans outperform machines by 5% on the Color scenario in which recognition rates are much higher. We hypothesize the improved performance over machines is due to this increase in recognition.

Finally, we examine several confusion matrices for human contour detection performance on 15×15 patches in the Color scenario with Mean Shift, FH and UCM machine algorithms, in Figure 10. The labeling agreement between humans and Mean Shift is slightly less than FH and UCM. UCM and FH produce the most similar labels.

5. Discussion

The experiments show several interesting trends. For instance, even after accounting for the knowledge of depth boundaries and objects, an increase of approximately 15% can be seen as the patch size increases. This implies a larger performance boost is gained from mid-level non-semantic information than the recognition of objects, which typically shows a 5% gain in performance. However, it is possible that other types of high-level information, which are not modeled in our studies are partially responsible for this increase.

Our studies did not find a strong relationship between depth understanding and contour detection. However, humans viewing real-world scenes have access to depth information directly from stereo imagery and motion parallax. If these cues were also given to our subjects, depth information would probably play a much larger role in contour detection.

When comparing the human studies to machine results, the state-of-the-art algorithms [5, 9, 1] do surprisingly well. For instance, machine accuracies are nearly identical to humans with 33x33 patches when an object isn't recognized.

Since humans have access to significant low- and mid-level information in a 33x33 patch, it is worth asking to what degree segmentation performance can improve using only low and mid-level information. Perhaps the recognition of objects is necessary to significantly improve segmentation performance. Should segmentation and recognition be performed jointly [10, 33, 29]?

In conclusion, this paper presents numerous human and machine studies on image segmentation. We find evidence that machines perform as well as humans using low-level information. Mid-level information appears to provide a larger boost in contour detection accuracy than the recognition of objects. Finally, we hypothesize the recognition of objects, but not depth boundary detection is necessary to achieve human level performance.

Acknowledgements: This work was supported in part by NSF IIS-1115719.

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *PAMI*, 33(5):898 –916, 2011. 2, 3, 7

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 24(4), 2002. 1, 2

[3] J. Canny. A computational approach to edge detection. *PAMI*, 8(6), 1986. 1, 3

[4] M. J. Choi, J. Lim, A. Torralba, and A. Willsky. Exploiting hierarchical context on a large database of object categories. In *CVPR*, 2010. 4

[5] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 24(5), 2002. 3, 7

[6] P. Dollár, Z. Tu, and S. Belongie. Supervised learning of edges and object boundaries. In *CVPR*, June 2006. 1

[7] J. Driver and R. S. Frackowiak. Neurobiological measures of human selective attention. *Neuropsychologia*, 39(12):1257 – 1262, 2001. 2

[8] J. Elder and S. Zucker. Computing contour closure. In *ECCV*. 1996. 1, 3

[9] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2), 2004. 3, 7

[10] V. Ferrari, T. Tuytelaars, and L. Van Gool. Simultaneous object recognition and segmentation by image exploration. In *ECCV*. 2004. 1, 2, 8

[11] C. C. Fowlkes. Measuring the ecological validity of grouping and figure-ground cues. *Thesis*, 2005. 2

[12] C. C. Fowlkes, D. R. Martin, and J. Malik. Local figure-ground cues are valid for natural images. *Journal of Vision*, 7(8), 2007. 1, 2

[13] L. Grady, M.-P. Jolly, and A. Seitz. Segmentation from a box. In *ICCV*, 2011. 2

[14] G. W. Hartmann. Principles of gestalt psychology. *Journal of Applied Psychology*, 20:623–628, 1936. 1

[15] D. Hoiem, A. Efros, and M. Hebert. Putting objects in perspective. In *CVPR*, 2006. 1

[16] Y. J. Lee and K. Grauman. Collect-cut: Segmentation with top-down cues discovered in multi-object images. In *CVPR*, 2010. 1, 2

[17] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *PAMI*, 26:530–549, 2004. 1, 2, 4

[18] J. McDermott. Psychophysics with junctions in real images. *Perception*, 33:1101–1127, 2004. 2

[19] P. Parent and S. Zucker. Trace inference, curvature consistency, and curve detection. *PAMI*, 11(8):823 –839, 1989. 1

[20] D. Parikh. Recognizing jumbled images: the role of local and global information in image classification. In *ICCV*, 2011. 2

[21] D. Parikh and C. Zitnick. The role of features, algorithms and data in visual recognition. In *CVPR*, 2010. 2

[22] D. Parikh and C. Zitnick. Finding the weakest link in person detectors. In *CVPR*, 2011. 2

[23] D. Parikh, C. Zitnick, and T. Chen. From appearance to context-based recognition: Dense labeling in small images. In *CVPR*, 2008. 2

[24] M. Peterson. Object recognition processes can and do operate before figure-ground organization. *Current Directions in Psychological Science*, 3, 1994. 2

[25] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. *ICCV*, 2007. 1, 2

[26] J. Rivest and P. Cabanagh. Localizing contours defined by more than one attribute. *Vision Research*, 36(1):53 – 66, 1996. 2

[27] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. In *SIGGRAPH*, 2004. 1

[28] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 22(8), 2000. 3

[29] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *ECCV*. 2006. 1, 2, 8

[30] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman. Discovering objects and their location in images. In *ICCV*, 2005. 2

[31] H. Tao, H. Sawhney, and R. Kumar. A global matching framework for stereo computation. In *ICCV*, 2001. 1

[32] A. Torralba, R. Fergus, and W. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *PAMI*, 30(11), 2008. 2

[33] Z. Tu, X. Chen, A. L. Yuille, and S.-C. Zhu. Image parsing: Unifying segmentation, detection, and recognition. *IJCV*, 63:113–140, 2005. 1, 2, 8

[34] R. von der Heydt, E. Peterhans, and G. Baumgartner. Illusory contours and cortical neuron responses. *Science*, 224(4654), 1984. 1

[35] J. Wang and E. Adelson. Representing moving images with layers. *TIP*, 3(5), 1994. 1

[36] L. R. Williams and D. W. Jacobs. Stochastic completion fields: a neural model of illusory contour shape and salience. *Neural Comput.*, 9:837–858, 1997. 1

[37] J. Winn and N. Jojic. Locus: learning object classes with unsupervised segmentation. In *ICCV*, 2005. 1

[38] A. Yoonessi and C. L. Baker. Contribution of motion parallax to segmentation and depth perception. *Journal of Vision*, 11(9), 2011. 2