Multi-Scale Contour Extraction Based on Natural Image Statistics

Francisco J. Estrada and James H. Elder
Centre for Vision Research, York University
4700 Keele St., Toronto, On.
M3J 1P3, Canada
[paco,elder]@elderlab.yorku.ca

Abstract

Perceptual grouping of the complete boundaries of objects in natural images remains an unsolved problem in computer vision. The computational complexity of the problem and difficulties capturing global constraints limit the performance of current algorithms. In this paper we develop a coarse-to-fine Bayesian algorithm which addresses these constraints. Candidate contours are extracted at a coarse scale and then used to generate spatial priors on the location of possible contours at finer scales. In this way, a rough estimate of the shape of an object is progressively refined. The coarse estimate provides robustness to texture and clutter while the refinement process allows for the extraction of detailed object contours. The grouping algorithm is probabilistic and uses multiple grouping cues derived from natural scene statistics. We present a quantitative evaluation of grouping performance on the Berkeley Segmentation Database, and show that the multi-scale approach outperforms several single-scale contour extraction algorithms.

1. Introduction

We address the problem of computing complete bounding contours of salient objects in a natural image. This problem is of central importance to computer vision, and yet remains largely unsolved. Most current boundary detection algorithms work on data at a single scale (usually the scale that provides the highest amount of detail). These single-scale algorithms can yield useful results on reasonably simple scenes, but natural images with their wealth of detail, texture, clutter, and other artifacts continue to be extremely challenging.

There are many factors that make the problem hard. Here we focus on two:

- *Complexity.* Finding complete bounding contours amounts to finding a specific sequence of n local edge primitives from a much large clutter of N edges. The general problem is of exponential complexity. Algorithms must therefore make strong assumptions that may limit performance, or must limit search to a small portion of the total hypothesis space.
- *Global Constraints.* Many algorithms use Markov assumptions that do not capture global structure. Capturing global constraints can generally be computationally expensive

One reason current algorithms are not more successful is that their single-scale nature is too constraining given the complexity of the grouping task. The abundance of detail on medium- and high-resolution images can cause single-scale grouping algorithms to "miss the forest for the trees". Previous research [6] has demonstrated that prior information about the contents of the image can be used to successfully deal with image complexity. However, such information is generally not available.

In this paper we demonstrate that a multi-scale approach can be used to reduce the complexity of the grouping task in the absence of prior scene knowledge. Our method is based on the observation that coarse object boundaries can be recovered from low-resolution images in which grouping complexity is significantly reduced. Such coarse boundaries can be used to inform and guide the grouping process at finer scales of resolution, culminating in detailed contours at the finest scale. The general procedure is applicable to most existing grouping algorithms provided that a suitable way can be devised to propagate information across scales. Here, we build on the probabilistic grouping framework of Elder et al. [6]. Their algorithm is based on natural scene statistics, and provides us with a sound base upon which to study the multi-scale grouping problem.

We evaluate the algorithm on images from the Berkeley Segmentation Database (BSD), and provide a quantitative comparison between our multi-scale method, the Ratio Contour algorithm of Wang et al. [24], the grouping algorithm of Estrada and Jepson [10], and the single-scale ver-

sion of our framework, which is roughly equivalent to the grouping algorithm of Elder et al. [6].

2. Previous Work

There has been interesting work in applying multi-scale methods to shape modeling. Dudek and Tsotsos [3] introduced a method for 2D and 3D shape representation and recognition based on multi-scale curvature. Dubinskiy and Zhu [4] proposed a coarse-to-fine method for shape encoding and synthesis. In their work, contours are formed with linear combinations of simple shapes; starting with an ellipse, and adding basis shapes progressively to achieve refined reconstructions. More recently, Wang et al. [27] presented a multi-scale image representation based on a sketch pyramid, and show that it is potentially useful for applications such as tracking and super-resolution. The progress made on multi-scale shape representation is encouraging, but it does not address the question of how we get the shapes to be represented out of the image in the first place. That is the main concern of this paper.

Many recent methods for boundary detection treat grouping as a clustering problem. Mahamud et al. [14] define an affinity matrix based on the probabilities of random walks biased by smooth continuation and proximity, and use the spectral properties of this matrix to estimate edge and link saliencies from which closed contours can be extracted. Sarkar and Soundararajan [22] use graph-partitioning techniques to identify subsets of features that are weakly connected to the remaining features in a graph that encodes perceptual grouping principles. Wang et al. [24] present an algorithm for finding contours as maximum-likelihood cycles in a graph. Gdalyahu et al. [11] present a grouping method that finds clusters of features (which for perceptual grouping are short line segments) that often appear together in stochastically generated graph cuts.

Elder and Zucker [7], extract contours as shortest-path cycles in a graph encoding perceptual grouping principles. They introduce a Bayesian grouping framework that accounts for different causes of contour fragmentation. This work was extended by Elder et al. [6] to allow the incorporation of object knowledge. The algorithm is based on natural scene statistics, and incorporates prior knowledge about the objects of interest in the form of the expected photometric appearance of the image around individual object edges. Estrada and Jepson [10] propose a search-based algorithm based on local affinity normalization. The resulting normalized affinities provide robustness to varying degrees of clutter, and result in significantly reduced search complexity.

All of these algorithms work at a single-scale. There are a couple of methods that, although not multi-scale, use multiple levels of inference to extract features. Mohan and Nevatia [17] use a constraint satisfaction network to generate progressively more complex structures. Their method proceeds hierarchically from single edgels to contours, ribbons, and finally surface patches. Sarkar and Boyer [21] propose a Bayesian network model of grouping. In their scheme, progressively more complicated features are formed by the network using voting methods and graph operations supported by perceptual grouping principles. Raman et al. [18] propose a method similar in spirit to the work we describe here. Starting from coarse-scale closed contours of important structures in MRI images, they track and refine the shape and position of each contour through scale-space. Their method is able to generate accurate edge maps while being resistent to noise and clutter.

Saund [20] introduced a fine-to-coarse edge grouping algorithm for extracting larger scale shape primitives from binary (silhouette) images. Liang et al. [13] use a two-scale image-space method to refine the position estimates of arterial walls. Ren and Malik [19] study a multi-scale, Bayesian method for contour completion. The goal of their algorithm is to refine the edge map based upon contour continuity over multiple scales. The algorithm is not designed to extract complete object boundaries, but some improvement in the localization of salient edges is demonstrated.

Several of the boundary detection algorithms mentioned above have been shown to successfully detect simple object boundaries on images from constrained domains. However, on complex natural images current grouping algorithms are easily misled, and ultimately fail. Reliable boundary extraction on complex images remains an unsolved problem. It is also interesting to notice that in the context of object-boundary detection the problem of scale has received relatively little attention. In the remainder of the paper we will show that the use of multi-scale information leads to improved boundary extraction on natural images from the Berkeley Segmentation Database (BSD). We begin by describing the specific objective of our algorithm

3. Training and Testing Image Data

The BSD provides a useful record of how human observers segment natural images into a small number of disjoint regions. However, the objective of our work (and most contour grouping efforts) is not to compute a full segmentation of the image, but to extract complete contours bounding the most salient objects in the scene. In addition to this, many BSD images are mostly composed of irregularly shaped texture regions that do not correspond to individual objects.

In order to be able to use images and ground-truth segmentations from the BSD to train and evaluate our method, we will use only those images from the BSD that contain at least one salient, unoccluded object that is fully contained within image bounds. Two human observers (independently, and including one of the authors) viewed all 200

images from the BSD training set and selected those that satisfied the above criteria. In total, there were 37 training images that both observers agreed satisfy these conditions.

Each image in the BSD has between 4 and 7 associated human segmentations. For each selected image, the observers cooperatively hand-labeled the regions from the corresponding ground-truth segmentations that corresponded to individual objects (which is necessary because a segmentation may split individual objects into several regions). This resulted in 37 sets of ground-truth object boundaries for training. The same process applied to the 100 BSD test images resulted in 20 test images, and 20 sets of ground-truth boundaries for testing.

4. Single-Scale Grouping Algorithm

Our grouping framework is based on the grouping algorithm of Elder et al. [6]. Their method takes as input a set of tangents (line segments) $T = \{t_1, ..., t_N\}$ and produces contours that are guaranteed to be closed and simple (without self-intersections). It is assumed that only a subset $T^o \subset T$ of tangents in the image lie on the boundaries of objects of interest. A contour hypothesis $s = \{t_{\alpha_1}, ..., t_{\alpha_m}\}$ is defined as an injective (non-repeating) sequence of tangents. Only a small subset C of the possible sequences will correspond to actual contours in the image, and only a subset C^0 of these will lie on the boundaries of objects of interest.

Let $I_1^o = \{\alpha_1, ..., \alpha_m\}$ represent the set of tangent indices for the current hypothesis s, and let $I_1^c = \{\{\alpha_1, \alpha_2\}, ..., \{\alpha_{m-1}, \alpha_m\}\}$ represent the set of index pairs for directly successive tangents on the hypothesized contour. Elder et al. approximate the posterior probability of the hypothesis as

$$p(s \in C^0 | D) \;\; \alpha \;\; F(s, D) = \prod_{i \in I_1^o} p_i^o \prod_{\{i,j\} \in I_1^c} p_{ij}^c, \qquad (1)$$

where p_i^o is the posterior probability that tangent t_i lies on an object of interest, and p_{ij}^c is the posterior probability that tangents t_i and t_j are directly successive tangents on a real contour.

The observable cues D can be partitioned into unary object cues D^o, used to compute p_i^o, and binary grouping cues D^c, used to compute p_{ij}^c. The unary cues represent the appearance of the image near individual tangents. The binary cues represent perceptual grouping relationships between pairs of tangents. When the objects of interest are of a specific type (e.g., skin coloured regions), unary cues can provide strong evidence for particular sequences (e.g., hue). In this paper we are interested in computing the boundaries of a broad range of salient objects, so the unary cues are relatively weak. Binary cues such as proximity and good continuation tend to be more stable over object type, although domain specificity can be observed.

The algorithm for computing hypotheses based on these cues is constructive: it generates progressively longer contours by expanding the current set of hypotheses by one tangent each iteration. The expansion is done so as to maximize the posterior probability that the resulting tangent sequence belongs to the set C^o of true contour fragments lying on the boundaries of objects of interest, given the observed data D.

Elder et al. call $F(s, D)$ the foreground term since it depends only on the tangents that are part of the hypothesis s. The posterior probabilities p_i^o and p_{ij}^c are given by

$$p_i^o = \frac{1}{1 + (L_i^o P_i^o)^{-1}} \quad \text{and} \quad p_{ij}^c = \frac{1}{1 + (L_{ij}^c P_{ij}^c)^{-1}},$$

where for a set of m^o unary cues at each tangent (e.g. colour, contrast), and a set m^c of binary cues for each tangent pair (e.g. proximity, smooth continuation),

$$L_i^o = \prod_{k=1}^{m^o} \frac{p(d_i^k | t_i \in T^o)}{p(d_i^k | t_i \notin T^o)} \qquad L_{ij}^c = \prod_{k=1}^{m^c} \frac{p(d_{ij}^k | \{t_i, t_j\} \in C)}{p(d_{ij}^k | \{t_i, t_j\} \notin C)}, \qquad (2)$$

and

$$P_i^o = \frac{p(t_i \in C)}{p(t_i \notin C)} \qquad P_{ij}^c = \frac{p(\{t_i, t_j\} \in C)}{p(\{t_i, t_j\} \notin C)}.$$

Here d_i^k is the observed value of the k^{th} unary cue at tangent t_i and d_{ij}^k is the observed value of the k^{th} binary cue for the pair of tangents $\{t_i, t_j\}$. The binary cues we use are proximity, smooth continuation, and similarity of brightness and contrast between a pair of tangents; with the associated probability distributions (learned from natural scene statistics) reported by Elder and Goldberg in [5].

In [6], unary cues are derived from prior knowledge about the scene. Here we assume no such knowledge. Instead, we will use unary cues to measure the salience of individual tangents and to incorporate the information provided by coarse-scale contours during multi-scale grouping. We describe the use of information provided by coarse scale contours in the next section.

To estimate the salience of individual tangents we use the boundary detector of Martin et al. [16]. Their detector, trained on natural scenes from the Berkeley Segmentation Database [15], uses brightness, colour, and texture to compute the *boundary energy* at each pixel. Martin et al. show that this boundary energy tends to emphasize salient object boundaries while disregarding many types of texture that produce strong responses from typical edge detectors.

To obtain the probability distributions on boundary energy required by our algorithm, we computed the boundary energy for our training dataset. To allow for typical localization errors in the boundary energy map, the boundary energy output was smoothed with a Gaussian kernel and normalized to $[0, 1]$. We then computed likelihood distributions of normalized boundary energy for edges on and off the ground-truth boundaries. The resulting distributions

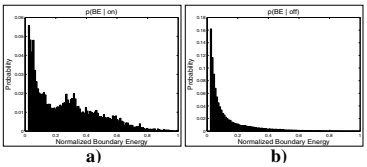

Figure 1. Probability distributions for boundary energy on ground-truth contours (a), and for random edges not part of the ground-truth contours (b).

$p(BE \mid \text{on})$ and $p(BE \mid \text{off})$ (shown in Fig. 1) then become one component of the object cue likelihood ratio L_i^o (2).

The grouping algorithm starts with a set of individual tangents. It keeps a table of the best M possible continuations for each tangent, and iteratively expands its current list of hypotheses as follows:

- For each of the K current hypotheses of length l, generate a set of M expanded contours of length $l+1$ by adding to that hypothesis one of the M tangents in its list of best continuations. This yields $K \cdot M$ new hypotheses of length $l+1$.
- Remove any self-intersecting or duplicate contours.
- Compute the term $F(s_k, D)$ for each of the remaining hypotheses.
- Sort the new set of hypotheses, and keep only the K contours with the highest $F(s_k, D)$.
- Detect and store any closed contours found among the new hypotheses.

The process terminates when none of the current hypotheses can be expanded, either because there are no suitable continuations, or because the possible expansions lead to self-intersecting shapes. Once the expansion phase has been completed, all closed contours are evaluated and sorted in order of decreasing quality. We estimate the quality of a particular hypothesis with the geometric mean of $F(s_k, D)$. The use of the geometric mean of probabilities along a chain of edges was first proposed as a measure of grouping quality by Mahamud et al. [14]. They noted that the use of the geometric mean enables an algorithm to compare contours of different lengths, removing the bias for short boundaries that occurs when only the product of the probabilities along the contour is considered. For a contour of length γ, the geometric mean is defined as $F(s_k|D)^{1/\gamma}$. In the log domain, this yields

$$\log(F(s_k|D)^{1/\gamma}) = \frac{\sum_{t_i \in s_k} \log(p_i^o) + \sum_{\{t_i,t_j\} \in s_k} \log(p_{ij}^c)}{\gamma}.$$

The algorithm described above is capable of detecting closed contours on natural images at a single scale. However, it can fail to detect complex contours on natural images. The algorithm explores only a small subset of the hypothesis search-space which often does not contain the boundaries of salient objects. In what follows we will demonstrate that a multi-scale approach can be used to successfully guide the algorithm toward better object boundaries.

5. Multi-scale Contour Grouping

We start from the observation that on low-resolution images it becomes possible for a grouping algorithm to explore a much larger portion of the full hypothesis space (or even to perform an exhaustive search). Thus it seems reasonable to use a coarse-to-fine strategy to successively refine contours found at lower resolutions. We expect that at coarse scales the grouping algorithm will find a great variety of closed shapes, at least some of which will correspond to salient object boundaries. These contours can then be refined until detailed boundaries are obtained at the finest scale.

We generate a Gaussian pyramid with W levels for the input image. Separately for each image in the pyramid we detect edges (see [8]), and use a greedy algorithm to fit small line-segments to groups of consecutive edges with similar orientation. This yields W sets of tangents (line segments) that are used as input for the multi-scale algorithm.

The multi-scale grouping proceeds as follows: The single-scale algorithm described above is run on the image at the coarsest level. The best P contours resulting from this step are stored. We then run the grouping algorithm $P+1$ times at the next finer scale. This corresponds to using each of the P spatial priors as an object (unary) cue, plus one additional run without any spatial prior so as to allow the algorithm to capture objects whose contour first becomes discernible at the current scale. The contours output by each of these runs are stored, and after all the runs at the current resolution are complete, the overall best P are selected again to form spatial priors for the next finer scale. This procedure is repeated until contours at the highest resolution are obtained.

To accomplish the above, we need a way to transform coarse-scale contours into a suitable spatial prior to be used at a higher resolution. This problem has two parts. First, we need an appropriate contour representation that can be easily transported across scales. Second, we need to study the relationship between contours at coarse scales and their refined versions at higher resolution, so that we can derive the appropriate statistical distributions to use with our grouping algorithm.

5.1. Fourier Contour Representation

We model contours using their Fourier descriptors. The Fourier representation is attractive because it is relatively inexpensive to compute (using the FFT), can be easily transferred across scales, and can be used directly for object

recognition tasks [25], [1], [12].

Given a list of coordinates (x_k, y_k) for $k = 0, \ldots, N-1$ along an extracted contour, we represent the contour as a curve in the complex plane formed by the points $b_k = x_k + iy_k$. The discrete Fourier transform of the set of points b_k along the contour yields a set of complex coefficients $B(n) = 1/N \sum_{k=0}^{N-1} b(k) e^{-ik2\pi n/N}$ for $n = 0, \ldots, N-1$. These Fourier coefficients can be used to reconstruct the contour to a specified level of detail. Here N is typically on the order of hundreds of sampled points. A good, intuitive description of how Fourier coefficients represent a shape is given in [23].

To determine the appropriate number of coefficients to use in the reconstruction, we studied the Fourier coefficients of the ground-truth boundaries in our training set. We reconstructed the original contours using different numbers of Fourier coefficients, for each of the reconstructions, we computed the distance error between the original curve and the reconstructed one. Our results indicate that typical boundaries can be reconstructed to an average error of less than 1 pixel using as few as 25 Fourier coefficients. This is the number of Fourier coefficients we will use in our grouping algorithm. An additional advantage of the Fourier representation is that it smooths out noise and discretization artifacts, providing a suitable model on which to build a spatial prior. This is particularly useful at coarse scales where discretization artifacts are more pronounced.

5.2. Building the Spatial Prior

For a particular coarse-scale contour, we generate a Fourier model and project a reconstruction of the contour scaled by a factor of 2 onto the image at the next finer scale. We want our grouping algorithm to favour lines whose orientation and spatial location are similar to the location and orientation of the projected contour. To accomplish this, we need to study the typical variations in position and orientation for object contours across scales.

We examined our ground-truth training contours at the finest and next-to-finest scale. We generated Fourier descriptors for the lower resolution contours, and projected them (after appropriate scaling) onto the images at the finest scale. We then measured the deviation in position and orientation between the scaled Fourier contours and the ground-truth contours at the fine scale. From these statistics, we estimated four probability distributions: $p(dist|on)$ which characterizes the distance between ground-truth tangents and the scaled contour; $p(ang|on)$ which characterizes the difference in orientation between ground-truth tangents and the model; $p(dist|off)$ which corresponds to the distance between tangents not on the object boundary and the Fourier model; and $p(ang|off)$ that models the difference in orientation between tangents not on the object boundary and the scaled contour. These distributions are shown in

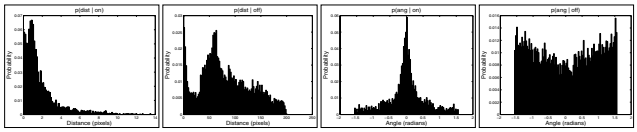

Figure 2. Distributions for a) $p(dist|on)$, b) $p(dist|off)$, c) $p(ang|on)$, and d) $p(and|off)$.

Fig. 2. The likelihood ratios $p(dist|on)/p(dist|off)$, and $p(ang|on)/p(ang|off)$ become additional object cues in the term L_i^o from Eq.(2). These distributions indicate that contour tangents at fine scale show good agreement in position and orientation with regard to the scaled coarse contours, whereas tangents not belonging to the fine-scale contours show no orientation agreement, and are situated farther away from the scaled coarse contours.

Together, the distance and angle cues form a spatial prior on the expected location and orientation of contour tangents given a particular coarse-scale hypothesis. The effect of this spatial prior is that finer-scale contours favour boundaries that correspond to refined versions of the coarse-scale hypothesis. Since the coarse-scale images are unlikely to contain much detail, coarse contours usually capture the rough shape of large objects. This gives the multi-scale grouping algorithm increased robustness against the typical problems introduced by the abundance of clutter and texture at higher resolutions.

Figure 3 shows this coarse-to-fine process. For each level of the Gaussian pyramid we show the spatial prior at that scale derived from the contour extracted at the previous scale, corresponding to the ratio $(p(dist|on) * p(ang|on))/(p(dist|off) * p(ang|off))$; the best contour extracted by the multi-scale procedure; and the best contour extracted by the single-scale grouping method (i.e. without the spatial prior) at that scale. Whereas the multi-scale procedure successfully recovers the boundary of the object, the single-scale procedure becomes confused by the complexity of the line-set at higher resolutions.

6. Experimental Set-Up

There has been a significant lack of research with regard to the proper evaluation of grouping algorithms. Many of the current methods are demonstrated on a handful of images, with only visual comparison as a means of demonstrating their capabilities. Exceptions include the work of Borra and Sarkar [2] who carry out a performance evaluation of three grouping methods in the context of aerial object recognition, and the work of Williams and Thornber [29] who present an evaluation of different affinity measures used in grouping. More recently, Wang et al. [26] presented an interesting comparison of the grouping algorithms in [7], [14], [24].

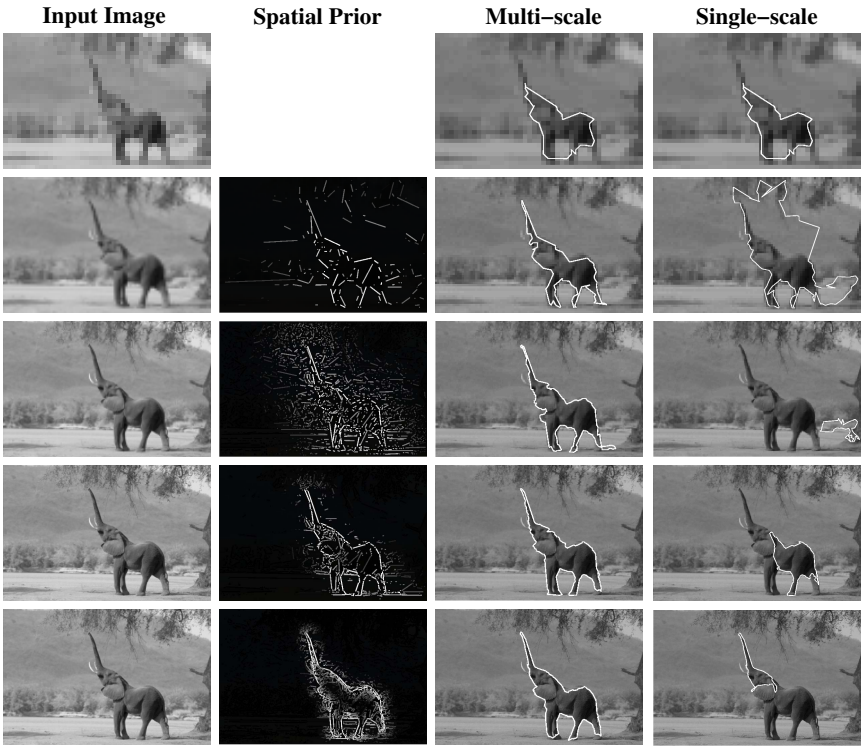

Figure 3. A complete run of the multi-scale algorithm. From left to right, the columns show the original input image (coarsest scale at the top, finest scale at the bottom); the spatial prior generated from the contour at the previous scale, where brightness is proportional to the estimated probability that a given segment is on a the object's boundary; the contour extracted by the multi-scale algorithm; and the contour extracted by the single-scale algorithm (i.e. without using the spatial prior). The image at the highest resolution is 737×486 pixels in size.

Here, we evaluate the performance of our algorithm using the set of ground-truth boundaries generated from the test images of the BSD. We provide a comparative analysis of the quality of the contours extracted by our multi-scale method (MS), the single-scale version of our algorithm (SS), and two alternate grouping methods: The Ratio Contour algorithm of Wang et al. [24] (RC), and the grouping method of Estrada and Jepson [10] (EJ)[1].

We used half-scale BSD images 241×162 pixels in size. This is because in our current implementation detected-contour quality degrades significantly at higher resolutions (see the discussion below). For the multi-scale algorithm we build a 4-scale Gaussian pyramid, which results in images 31×21 pixels in size at the coarsest scale. All images were processed with the same edge detector [8], however; the edge maps produced by this method are typically very complex. Our algorithm deals with this by using the boundary energy of Martin et al. [16]. To obtain a fair comparison, we pre-filtered the edges used as input to RC and EJ by selecting only the edges for which a blurred version of the boundary energy term is above some small threshold (determined experimentally on training images). The blurring of the boundary energy was the same used for estimating the boundary energy distributions for our algorithm. On training images this filtering step greatly enhanced the results obtained both with RC and with EJ.

From the edge data, RC generates its own input features by fitting spline curves to small sequences of edges. EJ uses the same line segments that were input to our algorithm. Since each algorithm has a different measure of saliency for extracted contours, we disregard the ranking of the shapes produced by the algorithms during the comparison (otherwise, it would be difficult to ascertain whether the observed variations are due to differences search completeness, or to the different ranking measures used by each method). Instead, we allowed each algorithm to generate 20 contours for each of the test images using its own quality measure. From these 20 boundaries, we selected for comparison the one with the smallest error with regard to the ground-truth.

We use an error measure based on the pixel-by-pixel agreement between the regions enclosed by the ground-truth contours and the computed boundaries

$$Error = \frac{A_{GT} - A_{GT \cap CT}}{A_{GT}} + \frac{A_{CT} - A_{GT \cap CT}}{A_{CT}} \qquad (3)$$

where A_{GT} is the area of the region enclosed by the ground-truth boundary, A_{CT} is the area enclosed by the

[1]The authors would like to acknowledge the kindness of Dr. Song Wang in providing us with his implementation of RC.

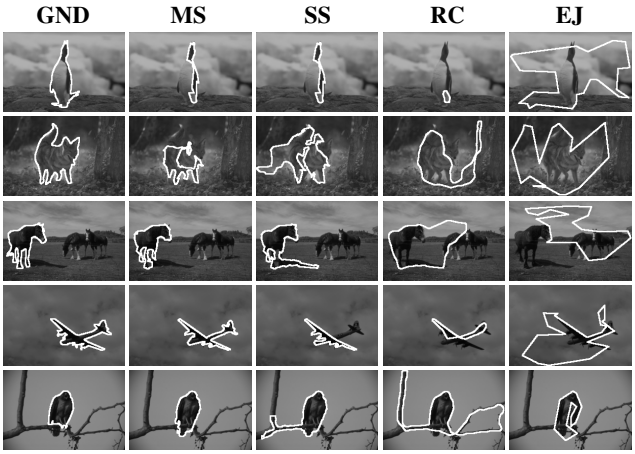

Figure 4. Ground-truth boundaries and best contours extracted by each algorithm on 20 of our test images.

computed contour, and $A_{GT \cap CT}$ is the area of the intersection between the two. The first term of the error measure penalizes a contour for not capturing the entire shape defined by the ground truth. The second term penalizes the contour for extending into the background. Error values are in $[0, 2]$ and will be zero when the contours are identical.

7. Results and Quantitative Evaluation

Fig 4 shows several images from our test set, along with the best contours extracted by each of the algorithms. The multi-scale algorithm recovers boundaries that more accurately describe the shape of the main object in each of the images. Fig 5a shows the mean error for each of the algorithms over the complete testing set. The results indicate that the multi-scale method outperforms the competing algorithms. The error values show high correlation between the algorithms (e.g. the $R2$ correlation between the error values for MS and RC is .37), so the absolute error may not be very informative. To verify that the difference is meaningful, we computed the relative errors of SS, RC, and EJ taking the multi-scale error as a baseline (this is just the difference of the corresponding errors on each image). Fig. 5b shows the mean relative errors of SS, RC, and EJ with regard to the multi-scale algorithm. Here, positive scores indicate that the corresponding algorithm has larger error on average compared to the multi-scale method. A standard t-test on the relative error measurements indicates that the observed differences are statistically significant with 95% confidence. It is clear from these results that the multi-scale algorithm is significantly superior to the single-scale version of the method, and that in general, it outperforms other single-scale grouping algorithms. It is worth pointing out that the test images constitute difficult perceptual organization problems, so the enhanced robustness that results from the use of multi-scale information is an important result. It

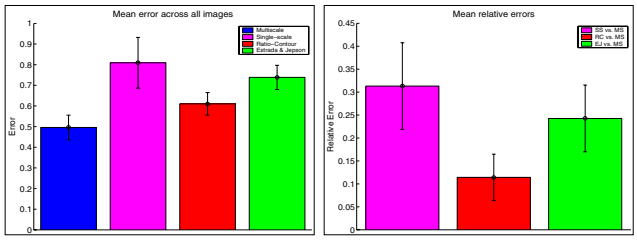

Figure 5. Error data for the grouping algorithms on our test image set. a) Mean absolute error for each algorithm. b) Mean relative error of alternate algorithm with regard to multi-scale.

should be noted that the multi-scale method described here must solve $((M-1) \times (P+1)) + 1$ individual grouping problems for an M-level Gaussian pyramid. Run-time for our current (Matlab) implementations is between 45 min. and 1 hr. for the multi-scale method, and between 3 and 5 min. for the single-scale version. RC and EJ (both of which are binary packages) take less than 1 min. to complete.

8. Discussion

Even with the use of multi-scale information, our algorithm can miss object boundaries in complex images. In particular, we have noticed degraded performance at full resolution, this may be due to a number of factors. The algorithm has no prior model for shapes, it can be led into extracting complicated boundaries winding their way through regions rich in structured texture (which, incidentally, are usually associated with high boundary energy). In addition to this, we have noticed that the geometric mean we use to estimate the quality of a contour is probably not the optimal measure on natural images. In general, the boundaries that resulted in the smallest error with regard to the ground-truth were ranked outside of the top 5 contours. Given that our algorithm chooses the best contours at each scale to refine at higher resolutions, the use of a more robust quality measure should improve upon the results presented above. It seems likely that such a measure would take into account not only the shape of the boundary and the relationship between features along the contour, but also the appearance of image regions inside and outside the hypothesized contour. The development of an appropriate measure of contour quality that yields good results on natural images is one of our current research topics.

Another interesting direction for future research is the addition of a prior global shape model based on natural scene statistics. Such a model could be used not only to guide the formation of boundaries, but also to evaluate the extracted contours. The main difficulty here is in developing a shape model that can be used to give both local predictions for suitable grouping directions, and global estimates of the quality of an extracted shape. Finally, current

research by one of the authors [9] shows that the use of appearance information from regions near a growing contour results in significantly better contour detection at a particular (single) scale.

9. Conclusion and Future Work

We have presented a Bayesian, coarse-to-fine framework for boundary detection. Our algorithm exploits reduced search complexity in low-resolution images to generate a set of coarse contour hypotheses, these boundaries are later refined until detailed object contours are obtained at the finest scale. Since the coarse contours capture general shape information, the procedure is less likely to be misled by the large amounts of texture and detail that exist at higher resolutions. We have shown that the proposed algorithm is able to detect complex object boundaries in images where a single-scale grouping algorithm would fail. We evaluated our algorithm on images from the BSD, and out results indicate that the multi-scale procedure yields, on average, better boundaries than alternate grouping algorithms. This is a promising result, and it is our hope that it will encourage further research in multi-scale boundary detection.

References

[1] Arbter, K., Snyder, W. E., Burkhardt, H., and Hirzinger, G., "Application of Affine-Invariant Fourier Descriptors to Recognition of 3-D Objects," *PAMI*, Vol. 12, No. 7, pp. 640-647, 1990. 5

[2] Borra, S., and Sarkar, S., "A Framework for Performance Characterization of Intermediate-Level Grouping Modules," *PAMI*, Vol. 19, No. 11, pp. 1306-1312, 1997. 5

[3] Dudek, G., and Tsotsos, J. "Shape Representation and Recognition from Multiscale Curvature," *CVIU*, Vo. 68, No. 2, 1997. 2

[4] Dubinskiy, A., and Zhu, S. C., "A Multi-scale Generative model for Animate Shapes and Parts," *ICCV*, pp. 249-256, 2003. 2

[5] Elder, J. H., and Goldberg, R. M., "Ecological Statistics of Gestalt Laws for the Perceptual Organization of Contours," *Journal of Vision*, Vol. 2, pp. 324-353, 2002. 3

[6] Elder, J. H., Krupnik, A., and Johnston, L., "Contour Grouping with Prior Models," *PAMI*, Vol. 25, No. 6, pp. 661-674, 2003. 1, 2, 3

[7] Elder, J. H., and Zucker, S., "Computing Contour Closure," *ECCV*, pp. 399-412. 1996. 2, 5

[8] Elder, J. H., and Zucker, S., "Local Scale Control for Edge Detection and Blur Estimation," *PAMI*, Vol. 20, No. 7, pp. 699-716, 1998. 4, 6

[9] Estrada, F. J., and Jepson, A. D., "Robust Boundary Detection With Adaptive Grouping," *5th IEEE Workshop on Perceptual Organization in Computer Vision*, 2006. 8

[10] Estrada, F., and Jepson, A., "Perceptual Grouping for Contour Extraction," *ICPR*, Vol. 2, pp. 32-35, 2004. 1, 2, 6

[11] Gdalyahu, Y., Weinshall, D., and Werman, M., "Self-Organization in Vision: Stochastic Clustering for Image Segmentation, Perceptual Grouping, and Image Database Organization," *PAMI*, Vol. 23, No. 10, pp. 1053-1074, 2001. 2

[12] Kuthirummal, S., Jawahar, C. V., and Narayanan, P. J., "Fourier Domain Representation of Planar Curves for Recognition in Multiple Views," *Pattern Recognition*, Vol. 27, pp. 739-754, 2004. 5

[13] Liang Q, Wendelhag I, Wikstrand J, Gustavsson T., "A multiscale dynamic programming procedure for boundary detection in ultrasonic artery images," *IEEE Transactions on Medical Imaging*, Vol. 19, No.2, pp. 127-142, 2000. 2

[14] Mahamud, S., Williams, L., Thornber, K., and Xu, K., "Segmentation of Multiple Salient Closed Contours from Real Images," *PAMI*, Vol. 24, No. 4, pp. 433-444, 2003. 2, 4, 5

[15] Martin, D., and Fowlkes, C., "The Berkeley Segmentation Database and Benchmark," Computer Science Department, Berkeley University, `http://www.cs.berkeley.edu/projects/vision/grouping/segbench/` 3

[16] Martin, D., Fowlkes, C., and Malik, J., "Learning to Detect Natural Image Boundaries Using Brightness and Texture," *NIPS*, 2002. 3, 6

[17] Mohan, R., and Nevatia, R., "Perceptual Organization for Scene Segmentation and Description," *PAMI*, Vol. 14, No. 6, pp. 616-635, 1992. 2

[18] Raman, A., Sarkar, S., and Boker, K., "Tissue Boundary Refinement in Magnetic Resonance Images Using Contour-Based Scale Space Matching," *IEEE Transactions on Medical Imaging*, Vol. 10, No. 2, pp. 109-121, 1991. 2

[19] Ren, X., and Malik, J., "A Probabilistic Multi-scale Model for Contour Completion Based on Image Statistics," *ECCV*, pp. 312-327, 2002. 2

[20] Saund, E., "Symbolic Construction of a 2-D Scale-Space Image," *PAMI*, Col. 12, No. 8, 1990. 2

[21] Sarkar, S., and Boyer, K. L., "Integration, Inference, and management of spatial information using Bayesian networks: Perceptual Organization," *PAMI*, Vol. 15, No. 3, pp. 256-274, 1993. 2

[22] Sarkar, S., and Soundararajan, P., "Supervised Learning of Large Perceptual Organization: Graph Spectral Partitioning and Learning Automata," *PAMI*, Vol. 22, No. 5, pp. 504-525, 2000. 2

[23] Staib, L. H., and Duncan, J. S., "Boundary Finding with Parametrically Deformable Models," *PAMI*, Vol. 14, No. 11, pp. 1061-1075, 1992. 5

[24] Wang, S., Kubota, T., Siskind, J. M., and Wang, J. "Salient Closed Boundary Extraction with Ratio Contour," *PAMI*, Vol. 27, No. 4, pp. 546-561, 2005 1, 2, 5, 6

[25] Wang, Y. F., Magee, M. J., and Aggarwal, J. K., "Three-Dimensional Objects Using Silhouettes", *PAMI*, Vol. 6, No. 4, pp. 513-518, 1984. 5

[26] Wang, S., Wang, J., and Kubota, T., "From Fragemtns to Salient Closed Boundaries: An In-Depth Study," *CVPR*, Vol. 1, pp. 291-298, 2004. 5

[27] Wang, Y., Bahrami, S., Zhu, S.C., "Perceptual Scale Space and its Applications," *ICCV*, pp. 58-65, 2005. 2

[28] Williams, L., and Jacobs, D., "Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Salience," *Neural Computation*, Vol. 9, No. 4, pp. 837-858, 1997.

[29] Williams, L., and Thornber, K., "A Comparison of Measures for Detecting Natural Shapes in Cluttered Backgrounds," *IJCV*, Vol. 2/3, No. 34, pp. 81-96, 2000. 5