Detecting Curved Symmetric Parts using a Deformable Disc Model

Tom Sie Ho Lee
University of Toronto

Sanja Fidler
TTI Chicago

Sven Dickinson
University of Toronto

Abstract

Symmetry is a powerful shape regularity that's been exploited by perceptual grouping researchers in both human and computer vision to recover part structure from an image without a priori knowledge of scene content. Drawing on the concept of a medial axis, defined as the locus of centers of maximal inscribed discs that sweep out a symmetric part, we model part recovery as the search for a sequence of deformable maximal inscribed disc hypotheses generated from a multiscale superpixel segmentation, a framework proposed by [13]. However, we learn affinities between adjacent superpixels in a space that's invariant to bending and tapering along the symmetry axis, enabling us to capture a wider class of symmetric parts. Moreover, we introduce a global cost that perceptually integrates the hypothesis space by combining a pairwise and a higher-level smoothing term, which we minimize globally using dynamic programming. The new framework is demonstrated on two datasets, and is shown to significantly outperform the baseline [13].

1. Introduction

In the formulation of object recognition as object detection, a strong top-down object prior minimizes the need for bottom-up perceptual grouping, *i.e.*, weaker, mid-level priors. However, as object databases grow towards human-like capacities, *i.e.*, tens of thousands of objects, a linear search through a large space of object detectors quickly becomes intractable. It's clear that as object databases grow, the role of perceptual grouping to extract a discriminative, domain-independent indexing structure that can prune a large database down to a small number of promising candidates will increase dramatically.

One of the most powerful indexing structures is a configuration of parts, in which a set of parts (and their relations) belonging to the same object is recovered without any a priori knowledge of scene content, *i.e.*, without the help of an object detector. The bottom-up recovery of a set of generic parts can be traced back to the earliest days of computer vision, and includes Blum's medial axis transform (MAT) [3], Binford's generalized cylinders [2], Pentland's superquadrics [25], and Biederman's geons [1], to name just a few examples. What do all these representations have in common? They're all based on symmetry, a physical regularity in our world that has been exploited by the human visual system to yield a powerful perceptual grouping mechanism – something that the early Gestalt psychologists understood practically a century ago [37] (for a definitive survey on symmetry and its analysis in computer vision, see [16]).

In an attempt to harness the power of a MAT-like representation yet avoid its inapplicability to cluttered scenes, Levinshtein *et al.* [13] introduced a bottom-up approach which first detects symmetric parts and then groups them nonaccidentally to form indexing structures. The key contribution is the modeling of a "deformable" maximal inscribed disc as a superpixel. The image is segmented into superpixels at multiple scales, where each scale yields a graph in which nodes are superpixels. Adjacent superpixels are linked by an edge, to which a learned affinity function assigns a measure of how likely two superpixels represent adjacent maximal discs belonging to the same part. A graph clustering algorithm is then applied to the superpixel graph to yield a set of connected components representing symmetric object parts.

While the framework outperformed previous approaches to symmetric part detection, it suffered from a number of serious limitations, as illustrated in Figure 1. To begin with, the symmetry model was restrictive in assuming that parts had straight axes and constant width. This prevented the correct detection of significantly curved and/or tapered parts, as illustrated in Figure 1(a). Second, superpixel grouping was restricted to a single scale, rather than integrated across multiple scales, preventing the detection of tapered parts whose component superpixels span multiple scales (Figure 1(c)). Finally, the superpixel grouping algorithm did not enforce a notion of good continuation, which led to incorrect detections when faced with ambiguous grouping possibilities, as reflected in the undersegmentation of the two symmetric parts (leaves) into a single part in Figure 1(e).

In this paper, we extend the approach of [13] to over-

come these limitations. Like [13], superpixels are generated at multiple scales and represent hypotheses of maximal inscribed discs. Like [13], adjacent hypotheses are assigned an affinity by a learned affinity function trained on manually detected symmetric parts. And like [13], hypotheses must ultimately be selected and grouped to form symmetric parts. But the model we use to assign the affinities, the nature of the search space of hypotheses which are selected and grouped to become parts, and the grouping algorithm that selects and groups hypotheses are different from [13], and represent the three main contributions of this paper.

In our first contribution, we relax the assumption that a symmetric part is straight with constant width, and extend the model to allow a part to bend and taper. Given two adjacent hypotheses, we fit a deformable ellipse to their union, from which an estimate of bending and tapering can be recovered. We then warp the image of the union, effectively "undoing" the bending and tapering, and yielding an invariant model of a symmetric part whose axis is straight and whose width is constant. By factoring out these deformations, we reduce the variability of the symmetry data used to train the classifier, and allow ourselves to adopt the learning framework of [13] while accommodating much greater within-class variation. In Figure 1(b), we see how this new model can detect curved symmetric parts.

In our second contribution, we relax the assumption that maximal disc hypotheses can only be grouped within a given scale, and extend the grouping process to integrate the hypotheses from multiple scales. In our multiscale approach, we make no assumptions about how the scales relate to each other, nor do we assume that adjacent maximal discs comprising a part must be drawn from adjacent scales. We construct a single, integrated search space of *deformable discs* and assign an affinity to any adjacent or overlapping discs, regardless of which scale they come from. This allows superpixels from different scales to be grouped into the same part, an essential requirement for detecting significantly tapered parts, as illustrated in Figure 1(d).

In our third contribution, we remove the global part symmetry constraint of [13] and relax it with smoothed local symmetry. Moreover, we reformulate the problem from a graph segmentation problem to a sequence finding problem, replacing a simple agglomerative clustering algorithm with an optimal grouping algorithm that captures the perceptual grouping principle of good continuation. Specifically, we adapt the salient curve detection framework of Felzenszwalb & McAllester [11], in which edgels are grouped to form salient, continuous curves, to group superpixels (hypothesized maximal discs) into salient, continuous, symmetric parts. The continuity model provides a much more powerful and flexible global constraint that can help select from among many locally ambiguous groupings, as illustrated in Figure 1(f).
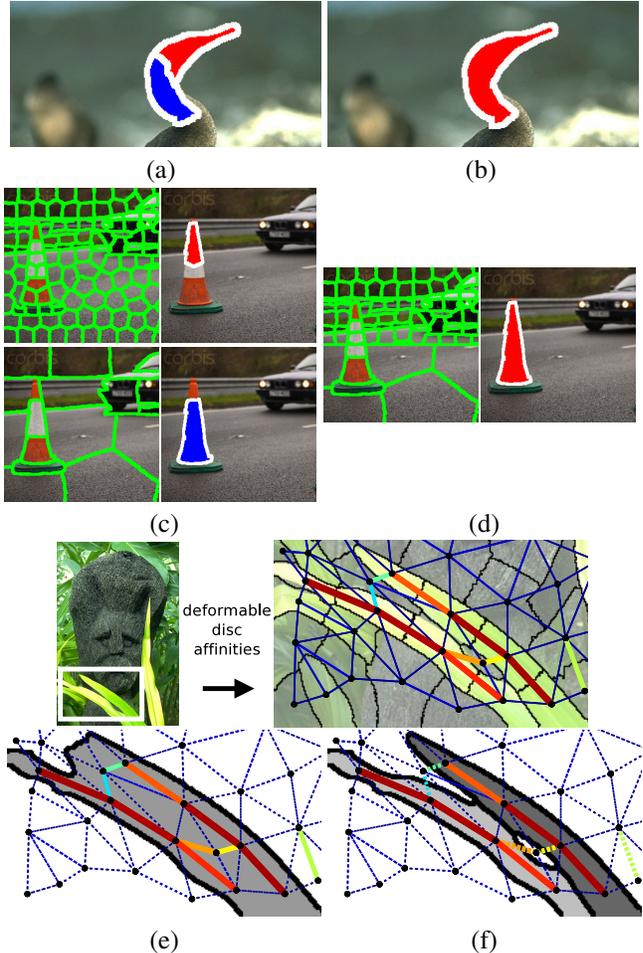
Figure 1. Illustrating the limitations of Levinshtein *et al.* [13] (left column) and demonstrating how our improved framework overcomes them (right column): (a) [13] cannot handle curved parts, and overpartitions them into piecewise straight; (b) we introduce a more powerful model for symmetry that removes this restriction, allowing curved parts to be correctly detected; (c) [13] detects parts by grouping superpixels from the same scale, preventing the detection of parts whose superpixels span multiple scales, *e.g.*, a tapered part; (d) we introduce a multiscale framework that allows superpixels at multiple scales to be grouped into the same part, allowing tapered parts to be detected; (e) [13] adopts a greedy superpixel grouping strategy that can lead to a graph structure which, while exhibiting good symmetry, can undersegment parts; in this case, the two sequences of superpixels representing the two leaves are bridged by a high-affinity path, yielding a tree-structured cluster that fails to distinguish the individual leaves (superpixels are shown as points and affinities in the cluster are colored, with redness proportional to affinity); (f) we adopt a new optimal grouping algorithm, based on the salient curve detection framework in [11], that combines local symmetry and continuity to better choose from among a set of ambiguous groupings; in this case, both symmetric parts (leaves) are correctly detected.

2. Related Work

While Blum's MAT, along with its many descendant representations such as the shock graph [29, 33] and bone graph [19], provided an elegant *decomposition* of a shape into symmetric parts, it assumed that the shape was segmented, an unrealistic assumption when the recognition domain consists of cluttered, occluded scenes. Other classes of approaches have taken a less restrictive approach that first attempts to detect local symmetries, in the form of parts, and then finds nonaccidental groupings of the detected parts to form indexing structures. Example approaches in this domain include the multiscale peak paths of Crowley and Parker [9], the multiscale blobs of Shokoufandeh *et al*. [32], the ridge detectors of Mikolajczyk and Schmid [22], and the multiscale blobs and ridges of Lindeberg and Bretzner [14], and Shokoufandeh *et al*. [31]. Unfortunately, these filter-based approaches yield many false positive and false negative symmetric part detections, and the lack of explicit part boundary extraction makes part attachment detection unreliable.

A more powerful filter-based approach was recently proposed by Tsogkas and Kokkinos [36], in which integral images are applied to an edge map to efficiently compute 13 features, including a novel spectral symmetry feature, at each pixel at each of 13 scales. Multiple instance learning is used to train a detector that combines these features to yield a probability map which, after nonmaximum suppression, yields a set of skeleton points. The method is computationally intensive yet parallelisable, and the skeleton points still need to be parsed and grouped into parts. But the method shows promise in recovering an approximation to a medial axis transform of an image.

Another class of approaches takes a less holistic approach, and addresses the combinatorial challenge of grouping extracted contours. Example approaches in this domain include Brady and Asada [5], Connell and Brady [8], Ponce [27], Cham and Cipolla [6, 7], Saint-Marc *et al*. [28], Liu *et al*. [15], Ylä-Jääski and Ade [38], and Stahl and Wang [35]. Since these methods are contour-based, they have to deal with the issue of computational complexity of contour grouping, particularly when cluttered scenes contain many extraneous edges. Some require smooth contours or initialization, while others were designed to detect symmetric objects and cannot detect and group the symmetric parts that make up an asymmetric object. A more recent line of methods extract interest point features, such as SIFT [17], and group them across an unknown symmetry axis [18, 12]. While these methods exploit distinctive pairwise correspondences among local features, they critically depend on reliable feature extraction.

A recent approach by Narayanan and Kimia [24] proposes an elegant framework for grouping medial fragments into meaningful groups. Rather than assuming a figure-

(a)   (b)   (c)
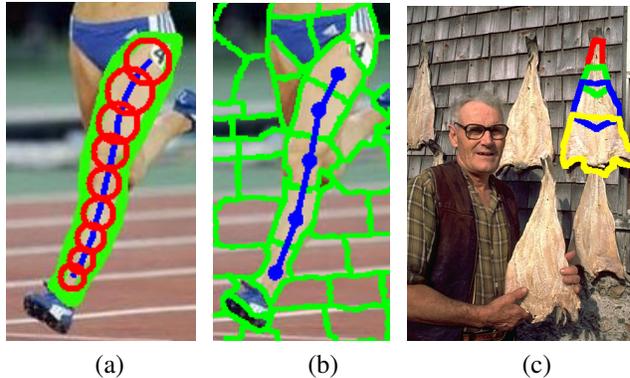
Figure 2. Deformable disc model of an object part: (a) Many maximal discs (red circles) are required to "flesh out" a part in the classical medial axis (from [13]); (b) by allowing maximal discs to deform to the shape of the part, far fewer *deformable dics* are required to define the part – in this case, the deformable discs are modeled as superpixels drawn from the same scale (from [13]); (c) a symmetric part is defined as the union of deformable discs – in this case, the deformable discs making up the tapered part are modeled as superpixels drawn from different scales.

ground segmentation, the approach computes a shock graph over the entire image of a cluttered scene, and then applies a sequence of medial transforms to the medial fragments, maintaining a large space of grouping hypotheses. While the method compares favorably to figure-ground segmentation and fragment generation approaches, the high computational complexity of the approach restricts it to images with no more than 20 contours.

Our approach, extending that of [13], is qualitatively different from both filter-based and contour-based approaches, offering a "region-based" approach which perceptually groups together compact regions (segmented at multiple scales using superpixels) representing deformable maximal discs into symmetric parts. We avoid the low precision that often plagues the filter-based approaches, along with the high complexity that often plagues the contour-based approaches.

3. A Representation for Symmetric Parts

Adopting the framework of [13], we define a symmetric part as a sequence of *deformable discs*, where each deformable disc d is a compact image region (a pixel mask) that roughly corresponds to a maximal inscribed disc. Unlike the classical medial axis transform in which maximal inscribed discs are bitangent to the part's boundary, as shown in Figure 2(a), our deformable discs are not constrained to be circular, and are allowed to deform to the shape of the boundary while maintaining high compactness. As a result, the number of deformable discs required to capture the shape of the part is far less than the number required using maximal discs, as shown in Figure 2(b).

Superpixels, being compact and having the tendency to deform to image boundaries, are ideally suited as a model of a deformable disc. Therefore, a superpixel segmentation can be seen as a set of deformable disc hypotheses. But since we have no a priori knowledge of part scale, and since a tapered part may be captured by deformable discs of different sizes, as shown in Figure 2(c), we generate superpixels (deformable disc hypotheses) at different scales. In contrast to [13], whose framework restricted grouping to superpixels at the same scale and thus could not handle significant taper, we group superpixels from a single hypothesis set that combines superpixels from all scales. We segment each image into 25, 50, 100, and 200 superpixels using a modified version [23] of the normalized cuts algorithm [30].

Given a set $\mathcal{D}$ of deformable disc hypotheses, our goal is to perceptually group deformable discs that belong to the same part. Since the vast majority of superpixels will *not* correspond to true deformable discs, we must manage the complexity of the search space. We adopt a proximity constraint between any two deformable discs and consider grouping together only deformable discs whose underlying superpixels are adjacent or overlapping. We thus capture the set $\mathcal{D}$ in a graph $\mathcal{G}$, whose nodes d_i represent deformable discs and whose edges (d_i, d_j) span pairs of deformable discs whose underlying superpixels are adjacent or overlapping. Pairs of superpixels in which one superpixel is entirely contained by the other are redundant groupings and are not included as edges in $\mathcal{G}$. Each edge is assigned a symmetry-based affinity which, as described in Section 4, reflects the degree to which the pair of deformable discs is believed to belong to the same symmetric part.

4. Defining a Deformable Disc Affinity

A restricted model of symmetry was used in [13], in which the axis was straight and the width was constant along the axis. An ellipse was fit to the region defined by two adjacent superpixels, defining a scale- and orientation-invariant coordinate system into which a grid was placed; edgels in the vicinity of the region boundary were then populated into the grid. The resulting shape feature, along with a set of appearance-based features computed over the same region, were fed to a classifier that evaluated to the edge affinity between two superpixels. Connected components with high edge affinity yielded symmetric parts.

To handle curvature and taper in symmetric parts, we relax the model of symmetry by replacing the ellipse with a deformable ellipse that accommodates bending and tapering [26]. An overview of the approach is illustrated in Figure 3. A deformable ellipse is fit to the boundary of the region, as shown in Figure 3(a). The axis is allowed to curve and the width to taper along the axis. Figure 3(b) shows edgels in the vicinity of the region boundary to which the model was

(a) (b)

(c) (d)
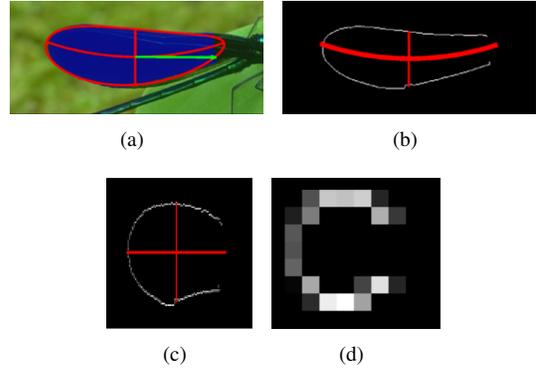
Figure 3. (a) A deformed ellipse fitted to the region shaded in blue. The boundary of the deformed ellipse along with its two axes are shown in red. The boundary edgels shown in (b) are warped into $\mathcal{W}$ resulting in (c), where the medial axis, shown bold, has been straightened and any taper removed. The descriptor formed on $\mathcal{W}$ is shown in (d).

fitted. Next, we "undo" the fitted bending and tapering deformations using the warp W that maps the boundary into the invariant coordinate system $\mathcal{W}$ as shown in Figure 3(c). Finally, boundary edgels are populated into a grid on $\mathcal{W}$ to compute the shape feature of [13], as shown in Figure 3(d).

Details of our model are described as follows. Recall that an ellipse is implicitly defined by a mapping into a space where each mapped point $W(\mathbf{x})$ is constrained to lie on the unit circle, satisfying $W(\mathbf{x}) \cdot W(\mathbf{x}) = 1$. In our deformable ellipse, the warp W is composed of a rigid transformation $R(\mathbf{p}, \theta)$ at position $\mathbf{p}$ and orientation θ, scaling $S(\mathbf{a})$ to major and minor axes lengths $\mathbf{a} = (a_x, a_y)$, and two deformations, namely, circular bending $B(b)$ at radius b and linear tapering $T(t)$ at slope t. We denote the full set of parameters of the warp by $\mathbf{w} = (\mathbf{p}, \theta, \mathbf{a}, b, t)$.

We compute W by fitting the parameters to boundary points $\mathbf{x}_1, \ldots, \mathbf{x}_N$. The problem is framed as least-squares minimization in which we seek a parameter vector $\mathbf{w}$ that locally minimizes the squared errors between the warped boundary points and the unit circle. The sum of squared errors objective,

$$\sum_{i=1}^{N} e(\mathbf{x}_i; \mathbf{w})^2, \tag{1}$$

is defined using the regularized algebraic distance

$$e(\mathbf{x}; \mathbf{w}) = c \cdot (W(\mathbf{x}; \mathbf{w}) \cdot W(\mathbf{x}; \mathbf{w}) - 1), \tag{2}$$

where $c = \sqrt{a_x a_y}$ penalizes highly elongated axes [34]. Deformation parameters are initialized to $b = 0$ and $t = 0$ while the position, orientation, and scale parameters are initialized by fitting a regular ellipse. All parameters are iteratively fit using a nonlinear least squares algorithm, yielding the solution $\hat{\mathbf{w}}$.

4

We compute the shape feature in the invariant coordinate system $\mathcal{W}$ defined by the fitted deformable ellipse. An edge map is computed using the Pb detector [21], thresholding responses at 0.01 to obtain a set of edgel coordinates lying in the vicinity of the region boundary. The edgels are then warped by $W(\cdot; \hat{\mathbf{w}})$ into $\mathcal{W}$, thereby reducing shape variability, in particular straightening the symmetry axis and removing taper. Finally, we compute a normalized spatial histogram of the warped edgels over a 10x10 grid to obtain a 100-dimensional feature. In addition, we compute an appearance feature based on color and texture features computed over the same region. A 27-dimensional feature is obtained from the dissimilarity of RGB and HSV means and variances and color and texture histogram distances.

We use a logistic regressor to combine shape and appearance features computed over a region r into $\sigma(r) \in [0, 1]$. The logistic regressor combines the shape output $\sigma_s(r)$, obtained by an SVM with RBF kernel, and the appearance output $\sigma_a(r)$, obtained by a logistic regressor with quadratic kernel. We train on a set of images containing symmetric parts that are annotated with pixel masks. Positive masks are adjacent deformable discs spanning the width of a part, and negative masks oversegment across the width, span over the part boundary, or undersegment the part. Details on part annotation and examples are provided in Section 6.

5. Finding Sequences of Deformable Discs

Given a graph $\mathcal{G}$ capturing deformable disc hypotheses, in which edge affinities reflect the degree to which adjacent deformable discs are believed to belong to the same symmetric part, the final step is to find sequences of deformable discs representing symmetric parts. In [13], a greedy agglomerative clustering algorithm based on [10] was adopted. At each iteration a candidate node was added to a cluster provided that the union of the node and cluster satisfied a *global* symmetry constraint.

The strategy suffered from two serious limitations. First, while the global symmetry constraint offered a useful abstraction mechanism, it was enforced by fitting an ellipse to the region corresponding to the entire cluster, thus restricting detectable parts to straight axes. The second limitation concerned the grouping algorithm, which was designed for graph clustering rather than for sequence clustering. A lack of a notion of continuation allowed clusters to branch into tree structures, leading to undersegmentation.

We overcome both limitations by framing the problem as a search for *sequences* over $\mathcal{G}$. For any sequence $P = (d_0, \ldots, d_n)$ of adjacent discs, the underlying graph edges correspond to binary terms $\{s(d_{i-1}, d_i)\}_{i=1}^n$. We define ternary terms $\{t(d_{i-1}, d_i, d_{i+1})\}_{i=1}^{n-1}$ to cover slightly larger subsequences over which a notion of smoothed local symmetry can be computed. This is in contrast to the restrictive global scope over which symmetry was computed in

(a)

(b)
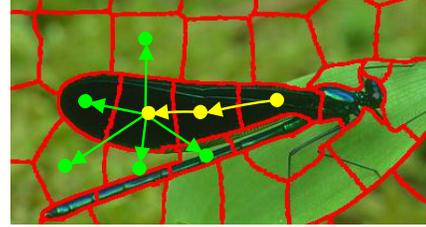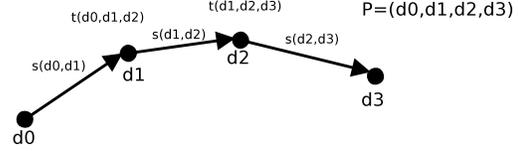
Figure 4. (a) Grouping is formulated as finding a sequence P of adjacent deformable disc hypotheses. (b) In the dynamic programming algorithm, a candidate sequence (yellow) is dequeued and possible extending discs (green) are considered.

[13]. Moreover, as explained below, good continuation is enforced by extending a candidate sequence only from its *last* disc, thereby preventing branching from its sides.

We find optimal sequences $P_1, P_2, \ldots$ over $\mathcal{G}$ using the global cost defined below in Eq. 3. We have adapted the method of Felzenszwalb & McAllester [11], which grouped sequences of edgels into salient curves, to our setting in order to group sequences of deformable discs into salient symmetric parts. Accordingly, we sum over binary and ternary terms and normalize by the length n:

$$\text{cost}(P) = \frac{A}{n} + \frac{\sum_{i=1}^n s(d_{i-1}, d_i) + \sum_{i=1}^{n-1} t(d_{i-1}, d_i, d_{i+1})}{n}.$$

(3)

The cost includes a term A/n that favors longer sequences, where we have set $A = 0.1$. The binary and ternary terms are set to $1 - \sigma(r)$, where the region r is defined by the set of discs given to the respective term. Note that we have assumed an equal contribution to the total cost from each disc regardless of its size.

The cost function in Eq. 3 is minimized by the dynamic programming algorithm in [11], which pursues a best-first search strategy to find the best sequence. Initially, a priority queue, Q, contains all possible (candidate) sequences of unit length. As shown in Figure 4(a), each edge (d_i, d_{i+1}) is directed such that a sequence of edges terminating at d_{i+1} can be extended with an edge starting at d_{i+1}. At each iteration, the most promising sequence P^* is dequeued from Q, and new candidate sequences are proposed by extending the end of P^* with adjacent discs, as shown in Figure 4(b). If an extended sequence ending at an edge s improves the cost of another sequence ending at s, it is enqueued back onto Q. Multiple sequences are found by successively min-
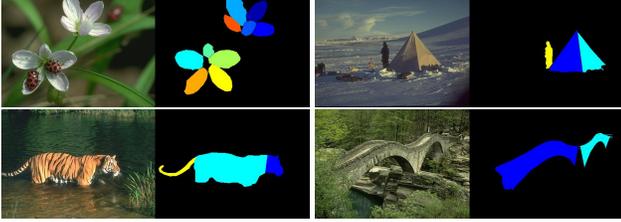
Figure 5. Example images with manually annotated symmetric parts.

imizing Eq. 3 over remaining disc hypotheses, and stopping when a maximum cost is reached.

The algorithm keeps track of only the best sequence ending at each edge for all possible lengths. It is possible for a symmetric part to have a long spiral axis, however we do not expect the length of a sequence to be on the order of the number of superpixels. We have used a maximum length of 20 in our implementation. We refer the reader to [11] for details of the algorithm.

6. Results

Following the criteria in [13] for symmetric part detection, we evaluate the ability of the algorithm to find *detection masks* corresponding to object parts in a cluttered scene. To obtain a variety of objects appearing in a range of scales in background clutter, we have selected images from the Berkeley Segmentation Database (BSDS) [20] for evaluation. We denote as BSDS-Parts a set of 36 images which are annotated with ground-truth masks corresponding to the symmetric parts of prominent objects (*e.g.*, duck, horse, deer, snake, boat, dome, amphitheater); examples of annotated images are shown in Figure 5. We also evaluate on the Weizmann Horse Database [4], denoted as WHD, using 61 annotated images as was done in [13]. Note that while each image prominently features one or more horses, we are interested in detecting object parts.

Our quantitative evaluation is summarized in Figure 6, which includes a baseline comparison with Levinshtein *et al.* [13] on BSDS-Parts in Figure 6(a) and WHD in Figure 6(b). The results indicate a significant improvement over the baseline. For each method, we obtain a precision-recall curve varying a threshold over the costs of detected parts. A detection mask m_{det} is counted as a hit if its overlap with the ground-truth mask m_{gt} is greater than 0.4, where overlap is measured by intersection-over-union (IoU) $|m_{det} \cap m_{gt}|/|m_{det} \cup m_{gt}|$. Note that low precision is partly due to the lack of annotations on many background objects in both datasets.

We have run our optimization method using the baseline features to isolate the effect of the grouping algorithm. To minimize our cost (Eq. 3), baseline features are com-

puted only at binary and ternary scope. The results for both datasets ("baseline + sequences" in Figure 6), show that the algorithm is responsible for a significant part of the improvement. We also evaluate the contribution of computing symmetry over a 3-disc subsequence as compared to a 2-disc subsequence ("ours w/o smoothing" in Figure 6), and find that the larger scope is beneficial.

Figure 7 highlights specific strengths and weaknesses of our approach. Example (a) demonstrates the successful recovery of a snake along with a second symmetric part representing its shadow (left side of snake). The image is overlaid with detection masks, over which the linear structure of each part is indicated by green line segments that join adjacent deformable discs. Examples (b) and (c) show the detected parts of a plane and a bird. Most of the plane parts are correctly segmented, with the left wing overpartitioned due to the engine, while the bird's parts are correctly segmented; configurations of such detected parts can provide a powerful index into a database of part-based shape models. In example (d), many of the symmetric parts (both straight and curved) comprising the boat are correctly recovered. Examples (e) and (h) show symmetric parts detected in a variety of scenes of intermediate complexity.

Examples (f) and (g) illustrate limitations of our approach. The wing and abdomen of the fly are successfully recovered, however the occluded leaf was not. In (g), low contrast along segments of the snake cause overpartitioning of the part. One could imagine a higher-level grouping module that could group the symmetric parts produced by our framework according to principles of collinearity/co-curvilinearity. We also note that part recall depends on the ability of the superpixel segmentation algorithm to yield deformable maximal discs that make up a symmetric part. Finally, we note that we trained our classifier on 20 WHD images to produce all of our detection results, even on images from BSDS. Note that with our agnostic approach to class labels, we have demonstrated that 1) symmetry is a powerful shape regularity that's ubiquitous in nature; and 2) the symmetry features learned from horse parts can reasonably generalize to a much broader class of symmetric parts that include curvature and taper.

7. Conclusions

Symmetry is a powerful regularity in our world that projects to a powerful regularity in the image. In the absence of an object prior, symmetry is a powerful cue for detecting parts whose configurations, in turn, can help manage search in a large-scale recognition task. The symmetric part detection framework of Levinshtein *et al.* [13] draws on the power of the medial axis while avoiding its pitfalls. However, it suffers from some serious limitations that limit its ability to detect more general classes of symmetric parts. We have addressed these limitations by introducing a num-
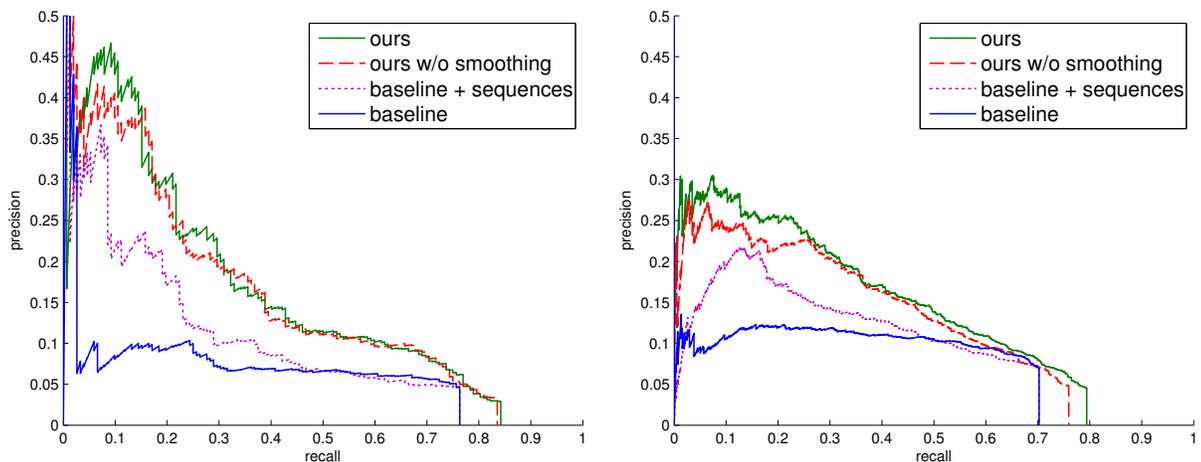
(a) Results on BSDS-Parts.

(b) Results on WHD.

Figure 6. Our approach significantly improves the baseline of Levinshtein *et al*. [13] on BSDS-Parts and WHD.

(a)   (b)   (c)   (d)

(e)   (f)   (g)   (h)
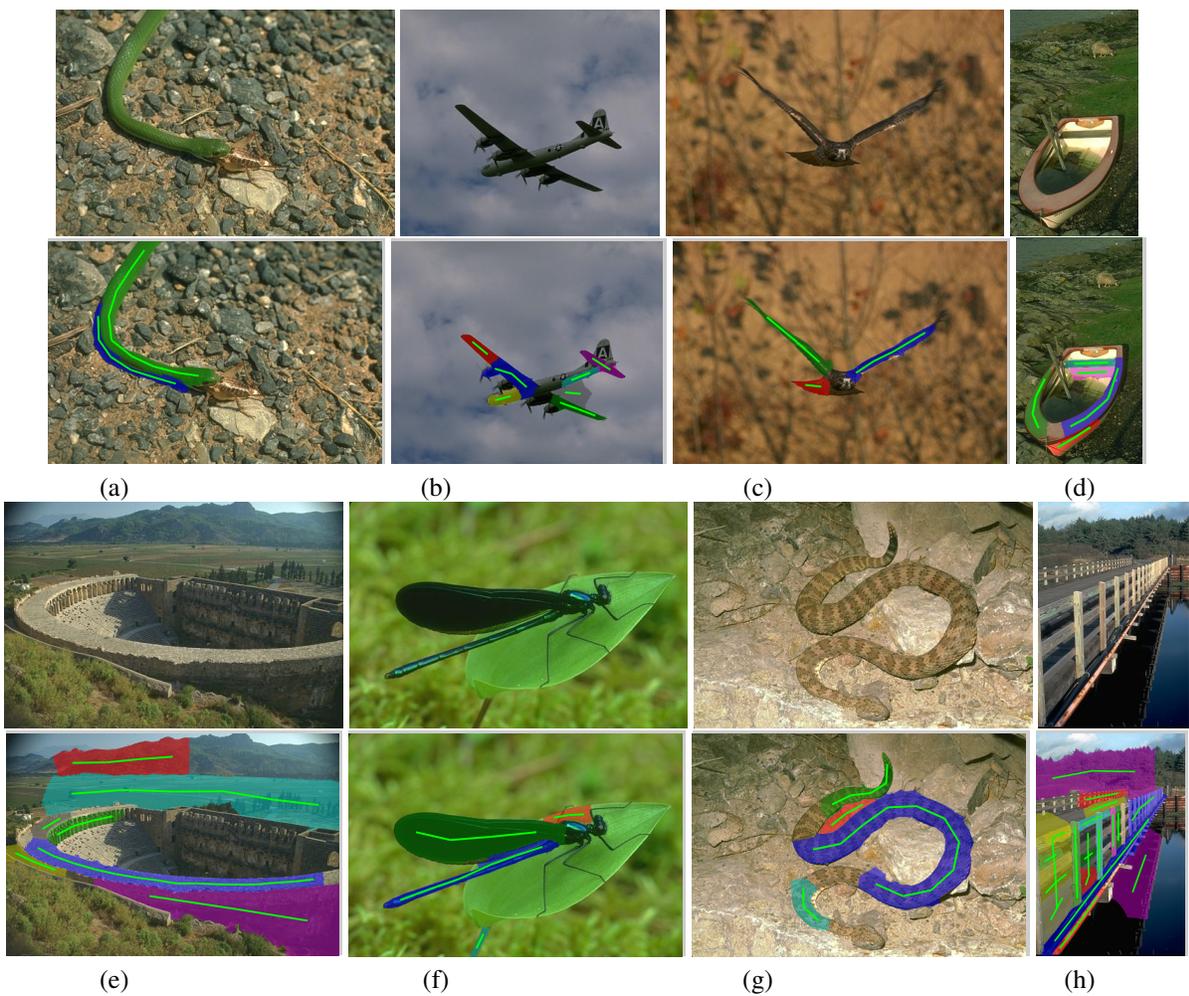
Figure 7. Symmetric object parts detected by our approach.

ber of extensions to [13], including both a richer yet more flexible model for symmetry, a multiscale framework, and an optimal grouping strategy based on the regularity of good continuation. The resulting framework significantly outperforms that of [13], offering an improved perceptual grouping framework for recovering symmetric parts without a priori knowledge of scene content. In future work, we will address the problem of part grouping to yield part configurations whose relational information offers the discriminative power to prune a large database down to a small number of promising candidates.

References

[1] I. Biederman. Human image understanding: Recent research and a theory. *CVGIP*, 32:29–73, 1985.

[2] T. O. Binford. Visual perception by computer. In *ICSC*, 1971.

[3] H. Blum. A Transformation for Extracting New Descriptors of Shape. In W. Wathen-Dunn, editor, *Models for the Perception of Speech and Visual Form*, pages 362–380. MIT Press, Cambridge, 1967.

[4] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In *ECCV*, pages 109–124, 2002.

[5] M. Brady and H. Asada. Smoothed local symmetries and their implementation. *IJRR*, 3(3):36–61, 1984.

[6] T.-J. Cham and R. Cipolla. Symmetry detection through local skewed symmetries. *IVC*, 13(5):439–450, 1995.

[7] T.-J. Cham and R. Cipolla. Geometric saliency of curve correspondences and grouping of symmetric contours. In *ECCV*, pages 385–398, 1996.

[8] J. H. Connell and M. Brady. Generating and generalizing models of visual objects. *AI*, 31(2):159–183, 1987.

[9] J. Crowley and A. Parker. A representation for shape based on peaks and ridges in the difference of low-pass transform. *PAMI*, 6(2):156–169, 1984.

[10] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, 2004.

[11] P. F. Felzenszwalb and D. McAllester. A min-cover approach for finding salient curves. In *WPOCV*, 2006.

[12] S. Lee and Y. Liu. Curved glide-reflection symmetry detection. *PAMI*, 34(2):266–278, 2012.

[13] A. Levinshtein, C. Sminchisescu, and S. Dickinson. Multi-scale symmetric part detection and grouping. In *ICCV*, 2009.

[14] T. Lindeberg and L. Bretzner. Real-time scale selection in hybrid multi-scale representations. In *Scale-Space*, volume 2695 of *Springer LNCS*, pages 148–163, 2003.

[15] T. Liu, D. Geiger, and A. Yuille. Segmenting by seeking the symmetry axis. In *ICPR*, volume 2, pages 994–998, Aug 1998.

[16] Y. Liu, H. Hel-Or, C. S. Kaplan, and L. V. Gool. Computational symmetry in computer vision and computer graphics: A survey. *Foundations and Trends in Computer Graphics and Vision*, 5(2):1–195, 2010.

[17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[18] G. Loy and J. Eklundh. Detecting symmetry and symmetric constellations of features. In *ECCV*, 2006.

[19] D. Macrini, K. Siddiqi, and S. Dickinson. From skeletons to bone graphs: Medial abstraction for object recognition. *CVPR*, 2008.

[20] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, pages 416–423, 2001.

[21] 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.

[22] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *ECCV*, pages 128–142, 2002.

[23] G. Mori, X. Ren, A. A. Efros, and J. Malik. Recovering human body configurations: Combining segmentation and recognition. In *CVPR*, pages 326–333, 2004.

[24] M. Narayanan and B. Kimia. Bottom-up perceptual organization of images into object part hypotheses. In *ECCV*, 2012.

[25] A. Pentland. Perceptual organization and the representation of natural form. *AI*, 28:293–331, 1986.

[26] M. Pilu, A. W. Fitzgibbon, and R. Fisher. Training pdms on models: the case of deformable superellipses. *PRL*, 20(5):463–474, 1999.

[27] J. Ponce. On characterizing ribbons and finding skewed symmetries. *CVGIP*, 52(3):328–340, 1990.

[28] P. Saint-Marc, H. Rom, and G. Medioni. B-spline contour representation and symmetry detection. *PAMI*, 15(11):1191–1197, 1993.

[29] T. Sebastian, P. Klein, and B. Kimia. Recognition of shapes by editing their shock graphs. *PAMI*, 26(5):550–571, 2004.

[30] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 22(8):888–905, 2000.

[31] A. Shokoufandeh, L. Bretzner, D. M. F. Demirci, C. Jönsson, and S. Dickinson. The representation and matching of categorical shape. *CVIU*, 103(2):139–154, 2006.

[32] A. Shokoufandeh, I. Marsic, and S. Dickinson. View-based object recognition using saliency maps. *IVC*, 17(5-6):445–460, 1999.

[33] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Z. Y. Shock graphs and shape matching. *IJCV*, 35:13–32, 1999.

[34] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *PAMI*, 12(2):131–147, 1990.

[35] J. Stahl and S. Wang. Globally optimal grouping for symmetric closed boundaries by combining boundary and region information. *PAMI*, 30(3):395–411, 2008.

[36] S. Tsogkas and I. Kokkinos. Learning-based symmetry detection in natural images. In *ECCV*, 2012.

[37] M. Wertheimer. Laws of organization in perceptual forms. In W. Ellis, editor, *Source Book of Gestalt Psychology*. 1938.

[38] A. Ylä-Jääski and F. Ade. Grouping symmetrical structures for object segmentation and description. *CVIU*, 63(3):399–417, 1996.