

Enhancing Boundary Primitives using a Multiscale Quadtree Segmentation

Robert Bergevin and Vincent Bergeron

Laval University, Department of Electrical and Computer Engineering,
Québec, Canada
{bergevin,vbergero}@gel.ulaval.ca

Abstract. A method is proposed to enhance boundary primitives of multi-part objects of unknown specific shape and appearance in natural images. Its input is a strictly over-segmented constant-curvature contour primitive (CCP) map. Each circular arc or straight-line segment primitive from the map has an unknown origin which may be the closed boundary of a multi-part object, the textured or marked region enclosed by that boundary, or the external background region. Five simple criteria are applied in order to weight each contour primitive and eliminate the weakest ones. The criteria are defined on the basis of the superposition of the CCP map on a multiscale quadtree segmentation of the original intensity image. A subjective ground-truth binary map is used to assess the degree to which the final weighted map corresponds to a selective enhancement of the primitives on the object boundary. Experimental results confirm the potential of the method to selectively enhance, in images of variable complexity, actual boundary primitives of natural and man-made multi-part objects of diverse shapes and appearances.

1 Introduction

Delimiting the region occupied by an unknown but interesting object in a static image is both useful and easy for humans. In computer vision, this is still a fundamental problem with no existing general solution. This is particularly notable with complex natural images where objects of interest appear under variations of shape, illumination, surface texture, viewpoint, and background. A novel generic object detection and localization method was proposed recently [1]. Its main assumption is that objects of properly complex shape are of more interest and preferably detected. On that basis, a number of potential object boundaries, defined as ordered groups of primitives from an input map of constant-curvature contour primitives (CCPs), are systematically generated and sorted according to a number of shape grouping criteria and constraints. From a given boundary, the region occupied by the object in the image is easily recovered.

Considering an average input map of 400 CCPs and boundaries of around 30 ordered CCPs, the number of possible boundaries is huge, that is about 10^{86} . Hence, very efficient pruning was required in order to reduce the number of generated boundaries to a manageable number of about a thousand. Still, images

often took up to many minutes to process. Sorting of generated boundaries was quite effective though since the generated boundary most similar to a manual reference was always in the first absolute or relative positions. Quantitative similarities obtained were from 85% to 100%. Besides, high-ranking boundaries were qualitatively very similar to their manual reference (see Figure 1).

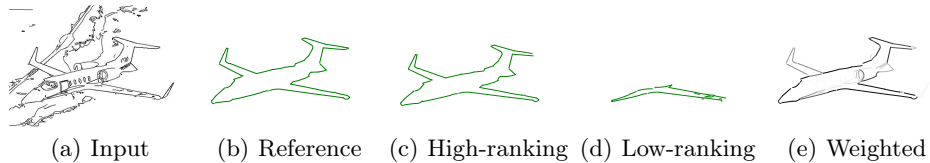


Fig. 1. CCP maps and boundaries. From left to right: binary input map, manual reference boundary, high-ranking boundary, low-ranking boundary, weighted map.

The goal of the method proposed in this paper is to assign a weight to each CCP in the input map that should reflect its potential to end-up on high-ranking boundaries. If properly done, such a process shall enhance CCPs from an object boundary with respect to distractor CCPs in the binary input map. The weighted map obtained could then replace the binary input map in the object detection and localization method. This would likely help reduce both the number of generated boundaries and the time required to generate and rank them. Typically, near 90% of the CCPs of an input map are distractors, either internal texture primitives or external background primitives (see Figure 1(a)). Figure 1(e) presents a weighted CCP map obtained by computing the relative number of times each CCP from the binary input map is used in the thousand or so boundaries generated by the object detection and localization method. The darker a CCP in that weighted map, the more it is used in the generated boundaries. That weighted CCP map could only be computed after all boundaries were generated. In contrast, the method proposed in this paper directly transforms a binary map into a weighted map to be used as input to a more efficient object detection and localization method.

A method with a similar goal was proposed recently [2]. The weight of a given contour primitive was computed using a number of criteria measuring the quality of groups made by pairing the primitive with all other primitives in the map. All but one criteria considered only the geometry or topology of the contour primitive pairs. The last criterion considered the coherence in local appearance of the paired primitives. That is, no criterion considered the appearance of the region enclosed by the boundary primitives. The method proposed in this paper precisely takes this type of complementary criterion into account. Since the method it to operate before object (boundary) detection, the actual object region is unknown and evidence for it must be obtained first.

More specifically, the binary input CCP map, made-up of straight-line segments and circular arcs [3], is superposed on a region map obtained from a low-

level segmentation of the original intensity image. From this superposition, five criteria are applied in order to weight each contour primitive and eliminate the weakest ones. Generic low-level segmentation methods group and select image data according to basic image parameters, irrespective of high-level knowledge as to what constitutes an object of interest [4, 5]. Unfortunately, they tend to produce results that suffer from both under-segmentation and over-segmentation. That is, no single region covers the whole object and some regions overlap the object and the image background. The latter is more problematic since the whole object is no more recoverable by grouping connected regions. By comparison, constant-curvature contour primitives (CCPs) are amenable to strict over-segmentation, which is at the basis of the object detection method in [1]. Raw images are trivially strictly over-segmented since each pixel is either on the object or not. However, pixels are both too numerous and local to be a primitive of choice for the superposition. The selected trade-off in the proposed method is a quadtree segmentation producing uniform regions of different sizes. In the current implementation, uniformity of regions is related to the statistical deviation of the pixel intensities from their average. Other definitions based on color or texture are also possible but they are left for future work. In order to still be robust to noise and detailed texture, the CCP map is superposed on a number of image segmentations obtained at different scales.

The enhancement problem addressed by the proposed method is related to the figure-ground segmentation problem [6, 7]. In both cases, one needs to identify which portions of an image contain important information and which are only distractive to the ultimate processing goal. Whereas generic low-level segmentation methods suffer from both under-segmentation and over-segmentation, specific high-level segmentation methods are more likely to extract significant segments of the image given their specialization to known objects of interest [8–10]. While much progress has been made to high-level methods recently, they still impose specific constraints on the object pose or appearance, which is not in line with our goal. Recent model-based object detection methods are also quite powerful in finding discriminating features of objects [11, 12]. However, they need a training phase specific to each object category. Finally, a number of generic perceptual grouping methods directly attempt to extract closed object boundaries [13–15]. A comparative study of [13] and two competing methods was made by Wang et al. [16]. For natural images of animals, the optimal contour always had a simple near-convex shape not representative of the animal shape. This limitation to near-convex shapes, also typical of previous saliency-based methods [17], is not appropriate for our problem.

The following section describes the proposed method in more details. Experimental results are presented in Section 3. A final section concludes the paper.

2 Proposed Method

As explained earlier, the main innovation of the proposed method is a superposition of the binary input CCP map on a quadtree segmentation of the original

intensity image. Pixels forming quadtree regions are meant to have similar intensities. Five segmentations are obtained, from fine to coarse, using five related thresholds controlling whether a given quadtree region is decomposed into four smaller regions. The finest segmentation, among the first three, with a number of regions significantly higher than the number of regions in the next coarser segmentation is selected as the main quadtree segmentation. The values of five criteria are combined into a weight for each CCP of the binary input map. The values of the first four criteria are computed using a superposition of the CCP with the main quadtree segmentation. The value of the last criterion is based on a superposition with the next two coarser segmentations. Constraints, corresponding to limit values for the criteria, are enforced to directly eliminate some CCPs from the output map.

Five simple and complementary criteria were selected. A short description of each criterion is given next. Thresholds and parameters are determined empirically by maximizing quantitative performance values for a representative set of images.

Local Coherence is the product of the CCP length and the number of quadtree regions it intercepts. The idea is to verify that the CCP is on the border between the region enclosed by the boundary and the external background region. If so, the CCP is to intercept a large number of small quadtree regions. Only the best one-third CCPs are retained.

Contrast is the difference between the average intensities on the two sides of the CCP. On each side, the average intensity of the quadtree regions intercepted by a band parallel to the CCP is used. A CCP with a too small contrast is eliminated.

Salience is the difference between the average size of the quadtree regions intercepted by the CCP and the average size of the quadtree regions on either side. On each side, the average size of the quadtree regions intercepted by a band parallel to the CCP is used. A CCP with a too small salience is eliminated.

Global Coherence is a binary criterion that obtains a true value if and only if the CCP intercepts a computed main region. Computation of the main region is explained in Section 2.4.

Scale Coherence is a ternary criterion that obtains a true or doubly true value when the CCP is among the best 20% CCPs in one or two of the additional coarser segmentations, respectively. For each coarser segmentation, the CCPs are sorted according to their local coherence, as long as they are not eliminated due to their poor contrast or salience.

The weight of a CCP is its local coherence, doubled when it is globally coherent and further multiplied by 1.5 when it is scale coherent or 2.25 when it is doubly scale coherent. Eliminated CCPs have zero weight. For example, the binary input map in Figure 4 has 552 CCPs. Local coherence eliminates 370 CCPs (two-thirds), contrast further eliminates 49 CCPs, and salience eliminates 20 CCPs. The number of CCPs in the output map is thus 113, about 20% of the

original. Prior to evaluating the five criteria for each CCP in the input map, a number of computations are performed as described in the following sections.

2.1 Quadtree Segmentation

Since images have arbitrary dimensions, quadtree regions are rectangular instead of square. Otherwise, the classical quadtree segmentation algorithm is used, starting with the complete image and recursively dividing it into four regions until either the standard deviation of the pixel intensities in a region is smaller than a given threshold or a region has less than four pixels. Let us define $\Delta_1 = 20\% * S_{total}$ et $\Delta_2 = 10\% * S_{total}$, where S_{total} is the standard deviation of the pixel intensities for the whole image. The five segmentations are obtained using the following thresholds: $S_{total} - \Delta_1$, $S_{total} - \Delta_2$, S_{total} , $S_{total} + \Delta_2$, and $S_{total} + \Delta_1$. The pixel intensities are obtained from the three channels of the color image using $Y = (0.299 * R + 0.587G + 0.114 * B)$. Quadtree segmentations at five different scales are presented in Figure 2.

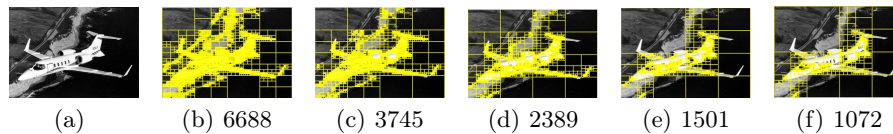


Fig. 2. From left to right: the original intensity image and five quadtree segmentations. The number of rectangular regions is indicated for each segmentation.

2.2 Region Intersections

Each of the five criteria needs a list of the quadtree regions intercepted by a given CCP. The intersections between each CCP and the quadtree regions are computed only once and stored in a look-up table. In order to compute the intersections, each quadtree region is represented by four straight-line segments, its four sides. Intersections are computed between a CCP, a straight-line segment or a circular arc, and each side. There are three possible cases: the CCP is fully contained in the region, the CCP intersects one or more sides, or the CCP is outside the region. The first two cases correspond to a CCP intercepting the quadtree region.

2.3 Bands of Parallel CCPs

On each side of a CCP, two parallel CCPs form a band to be used in computing the contrast and salience criteria. The distance between the CCPs is 5 pixels. When the original CCP is a circular arc, the CCPs in the band have different radii.

2.4 Main Region

The main region is a coarse object segmentation obtained using a region growing process on the main quadtree segmentation. In the present implementation, the main region is simply grown by recursively adding connected regions, starting from a small region seed. An empirical limit of 30 pixels is imposed on the area of an added region. The seed region has the smallest area in the quadtree. Once the first region is grown, another seed region is selected from the remaining regions, as far as its area is smaller than 30 pixels. A second region is grown from the new seed, and so on until all remaining regions are larger than 30 pixels. The largest grown region is retained as the main region. On most tested images, the main region covers the object at least in part. Future implementations of the proposed method should improve the main region extraction.

2.5 Algorithm

The complete primitive enhancement algorithm consists in the following steps:

1. compute the five quadtree segmentations from the intensity image.
2. select the main segmentation.
3. compute the local coherence, the contrast, and the salience of each CCP.
4. eliminate the CCPs with smaller local coherence.
5. eliminate the CCPs with too small contrast.
6. eliminate the CCPs with too small salience.
7. compute the main region.
8. compute the global coherence and the scale coherence of each retained CCP.
9. compute the weight of each retained CCP in the output map.

3 Experimental Results

Ten images of variable complexity were tested. They are displayed in Figure 3. Images **c**, **f**, and **g** are considered simple since they have limited background structure and texture. Images **a**, **e**, **h**, **i**, and **j** are considered complex since they have normal texture and background. Finally, images **b** and **d** are considered difficult mainly because of the poor contrast between the object and the background in some areas. Experiments used a fixed set of parameter values.

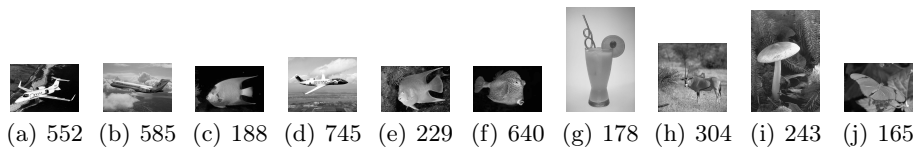


Fig. 3. Ten tested images. The number of CCPs in the input map is indicated.

Two types of qualitative results are presented. A detailed result for a complex image is presented first, showing maps at different intermediate steps. Then, a final result is presented for one sample image in each category of simple, complex, and difficult images. In this case, the binary input CCP map, the manual reference, and the thresholded output map are shown. Finally, a table of quantitative performance values is presented for the ten images.

3.1 Qualitative Results

A detailed result is presented in Figure 4. Each criterion is applied in turn and the binary map showing the retained CCPs is displayed. For criteria 4 and 5, maps before and after the application of the criterion are thresholded to the same number of CCPs in order to show the ranking differences. Sample results for simple, complex, and difficult images are shown in Figure 5.

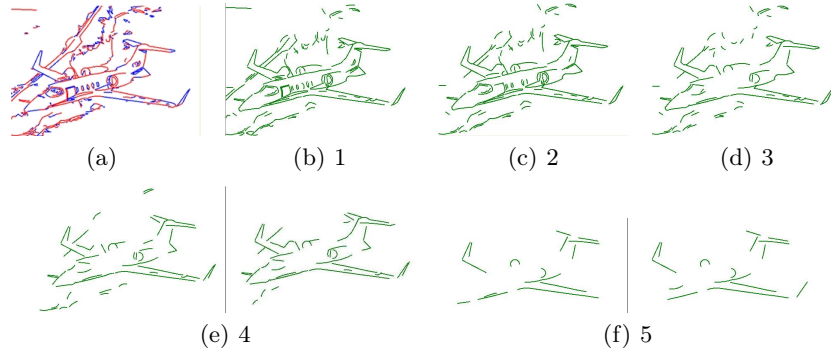


Fig. 4. Sequence of binary maps obtained by applying the five criteria. For criteria 4 and 5, input and output maps are thresholded at 60 CCPs and 20 CCPs, respectively. Circular arcs are red and straight-line segments are blue in the binary input map.

3.2 Quantitative Results

Table 1 presents the precision (PRE) and recall (REC) values for thresholded output maps. The number of CCPs in the input map (CCP), the reference map (SGT), and the thresholded output map (OUT) are indicated for each image. The precision and recall values are computed using a unit weight for each CCP, irrespective of its size and actual computed weight. The equal-error-rate (EER) and area-under-curve (AUC) values are computed from a recall versus precision curve (RPC), where the number of retained CCPs in the output map varies from one to the number obtained after applying the first three criteria. When CCPs included in the reference map are eliminated, the maximum recall value is less than 100%. For this reason, the RPC curve and the associated performance values are more challenging than the usual in object detection or image retrieval.

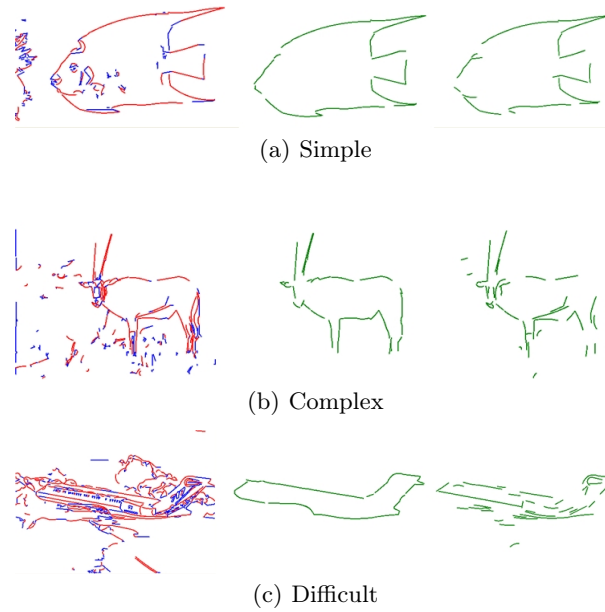


Fig. 5. Input, reference, and thresholded output maps for three sample images.

Image	CCP	SGT	OUT	PRE	REC	EER	AUC
a	552	50	60	73	60	65	66
b	585	33	45	31	40	32	17
c	188	32	27	80	84	84	78
d	745	45	55	38	40	41	32
e	229	33	32	84	77	84	71
f	640	44	44	72	72	72	63
g	178	30	28	70	80	70	69
h	304	41	43	46	45	45	41
i	243	32	29	62	52	55	45
j	165	29	30	40	35	31	20

Table 1. Quantitative performance values. The first three columns indicate the number of CCPs in the input, reference, and thresholded output maps, respectively. The last four columns indicate the performance values expressed as percentages. Precision and recall are obtained from thresholded output maps. Equal-error-rate and area-under-curve are obtained from a recall versus precision curve. Higher values are better.

The computing time for the complete algorithm, excluding the generation of the binary input map, varies depending on the input map size and the image contents. For the ten tested images, the range is from about half a minute to a minute and a half with a Visual.NET C# implementation running on an IBM ThinkPad with a 2.0 GHz Intel Pentium M processor and 1GB of RAM.

4 Conclusion

The goal was to assign a weight to each CCP in a binary input map according to its potential to be on an object boundary of proper complexity. Typically, near 90% of the CCPs of an input map are distractors, either internal texture primitives or external background primitives. Given that objects in tested images have unknown shape and appearance, only generic elimination and weighting criteria could be proposed for CCPs. The weight of each retained CCP was computed by taking into account the coherence between a quadtree segmentation of the original image and the input map. Five simple criteria used to combine region and contour segmentations are the main contribution of the method.

Results obtained on images of variable complexity are significant. Even for image **b**, which obtained the worst quantitative performance values, the retained CCPs with highest final weights provide a useful starting point for an object detection algorithm, as one may conclude by comparing the maps in Figure 5(c). In general, quality of results is as expected by the intuitive categorization of the images as simple, complex, or difficult. Excellent results are obtained for all simple images. Results for most complex images are at least good. Image **j** did not perform well but this might partly be an artifact of the evaluation process. Indeed, most of the longest boundary primitives are retained with a high weight, along with a number of small distractors. A thresholded output map provides a clear qualitative improvement with respect to the input map. Finally, performance values for difficult images were low but the resulting weighted maps are nevertheless an improvement, as just mentioned.

Various techniques were proposed recently for generic segmentation and object detection with standard image datasets, e.g. the Berkeley Segmentation Dataset and the PASCAL Visual Object Classes, used for comparison. Given that our method is only a preprocessing step and that it considers a very generic object category with no supervised training, it is more appropriate to compare it to grouping methods. Unfortunately, no standard image dataset was proposed for them. Some of our test images are from the Berkeley Dataset. Typically, salient boundaries extracted by state-of-the-art methods are of simpler shapes.

A number of changes could be made to the proposed method in order to improve its performance. A study of the relative importance of the five criteria could result in a different algorithm with, for instance, a different way to combine the criteria, the addition of new criteria, or the removal of the less useful ones. As mentioned earlier, an improved segmentation algorithm adding color or texture uniformity could also improve the results, especially for the difficult images. Similarly, regions of arbitrary shapes could be better adapted to other types of criteria. Finally, a fusion of the weighted output map with the map obtained by a complementary perceptual grouping method [2] would likely improve the respective performances of the two methods.

5 Acknowledgment

This work is supported by a NSERC discovery grant.

References

1. Bernier J.-F., Bergevin, R.: Generic Detection of Multi-Part Objects. In: Proc. of the 18th International Conference on Pattern Recognition (ICPR2006), Hong Kong, China, 2006.
2. R. Bergevin, A. Filiatrault, Enhancing Contour Primitives by Pairwise Grouping and Relaxation, in: Proc. of the 4th International Conference on Image Analysis and Recognition (ICIAR2007), Montréal, Canada, 2007.
3. Mokhtari, M., Bergevin, R.: Generic Multi-Scale Segmentation and Curve Approximation Method, in: Proc. of the Third International Conference on Scale-Space and Morphology in Computer Vision. Vancouver, Canada, 2001, pp. 000–007.
4. D.R. Martin, C.C. Fowlkes, Jitendra Malik: Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26** (2004) 530–549.
5. Lau, H.F., Levine, M.D.: Finding a small number of regions in an image using low-level features. *Pattern Recognition*, **35** (2002) 2323–2339.
6. Héroult, L., Horaud, R.: Figure-Ground Discrimination: A Combinatorial Optimization Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15** (1993) 899–914.
7. Grigorescu, C., Petkov, N., Westenberg, M.A.: Contour and boundary detection improved by surround suppression of texture edges. *Image and Vision Computing*, **22** (2004) 609–622.
8. Yu, S.X., Shi, J.: Object-Specific Figure-Ground Segregation, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2003). Madison, WI, 2003, pp. 39–45.
9. H. Arora, N. Loeff, D.A. Forsyth, N. Ahuja: Unsupervised Segmentation of Objects using Efficient Learning, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2007), Minneapolis, Minnesota, 2007.
10. Borenstein, E., Ullman, S.: Class-Specific, Top-Down Segmentation, in: Proc. of the 7th European Conference on Computer Vision (ECCV2002). Copenhagen, Denmark, 2002, pp. 109–122.
11. A. Kushal, C. Schmid, and J. Ponce: Flexible object models for category-level 3d object recognition, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2007), Minneapolis, Minnesota, 2007.
12. S. Savarese, L. Fei-Fei: 3D generic object categorization, localization and pose estimation, in: Proc. of the 11th International Conference on Computer Vision (ICCV2007), Rio de Janeiro, Brazil, 2007.
13. J.H. Elder, S.W. Zucker: Computing Contour Closure, in: Proc. of the 4th European Conference on Computer Vision (ECCV1996), 1996, pp. 399–412.
14. D.W. Jacobs: Robust and Efficient Detection of Salient Convex Groups, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18** (1996) 23–37.
15. F.J. Estrada, A.D. Jepson: Perceptual Grouping for Contour Extraction, in: Proc. of the 17th International Conference on Pattern Recognition (ICPR2004), 2004, pp. 32–35.
16. S. Wang, J. Wang, T. Kubota: From Fragments to Salient Closed Boundaries: An In-Depth Study, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2004), Washington, DC, 2004, pp. 291–298.
17. S. Mahamud, L.R. Williams, K.K. Thornber, K. Xu: Segmentation of Multiple Salient Closed Contours from Real Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25** (2003) 433–444.