

Generic Detection of Multi-Part Objects

Jean-François Bernier
Laval University
Computer Vision and Systems Laboratory
Quebec, Canada
jfbernie@gmail.com

Robert Bergevin
Laval University
Computer Vision and Systems Laboratory
Quebec, Canada
bergevin@gel.ulaval.ca

Abstract

A method is proposed to detect multi-part man-made or natural objects in complex images. It consists in first extracting simple curves and straight lines from the edge map. Then, a search tree is expanded by selecting and ordering the segmented primitives on the basis of generic local and global grouping criteria. The set of partial contours provided by the parallel search are combined into more complex forms. Global scores produce a sorted list of potential object silhouettes.

1. Introduction

Multi-part objects are everywhere, from living beings to man-made objects, rigid or deformable, articulated or not. They can be a person, with a head, body, two legs, and two arms, or an airplane, with its nose, body, two wings, and tail. Current work in detection of such objects in images is often too specific and it lacks efficiency and noise tolerance. A new fully deterministic and generic method is proposed whose goal is to come closer to the capacity of humans to detect interest regions in complex images. Multi-part objects are located by orderly selecting contour primitives on their boundary, based on simple grouping criteria.

Powerful techniques were proposed recently to learn local appearance features from exemplar images [1]. Despite impressive results, they are still too limited in terms of viewpoint-invariance and genericity. Other techniques are based on more generic but too simple shape models [4]. Interest points bring information on image contents [9] but are yet to be applied to generic detection. Inter-line affinity is computed in [3]. Proximity, continuity and closing criteria are defined on that basis. The extracted contours are numerous, they may appear anywhere, and they may look anyhow since only local, or simple global [2], criteria are enforced.

2. Basic concept

Figure 2(b) is a constant-curvature contour primitive (CCP) map of Figure 1(a) obtained using a custom segmentation algorithm [6]. Any ordered group of CCPs from the map is a possible detection or solution. Considering an average map of 400 CCPs, the number of possible solutions with 30 CCPs is about 10^{86} . Figure 1(b) is the best solution as interactively selected by a human. It is referred to as SGT, for subjective ground truth. SGT is not known by the algorithm and will only serve in assessing the quality of obtained results.

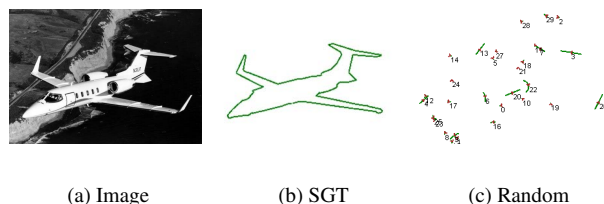


Figure 1. Basic concept

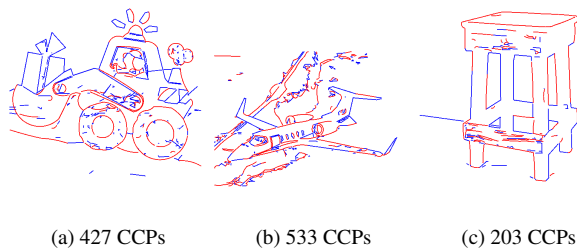


Figure 2. CCP maps

Let us assume that an algorithm provides a scoring function for solutions. FGT, or formal ground truth, is the possible solution with highest score. FGT is usually not known

either as it would require to score a nearly infinite number of solutions. More practically, a subset of the possible solutions is considered and the best one is selected. The latter is called FGTa, an approximation of FGT. FGTa may be the same as FGT, but this can seldom be verified in images of typical complexity. The goal of the proposed method is to generate, in an efficient manner, an FGTa as close as possible to SGT, for a variety of complex images of multi-part objects. Figure 1(c) shows an example of a random solution with 30 CCPs (each CCP has a small arrow and a number). In the proposed method, local grouping criteria help discard such a poor solution early on. Best retained solutions according to global criteria are similar to SGT (see Figure 3).

3. Parallel search for partial paths

Various existing techniques learn their parameters from training images [5]. Because of the very large number of possible objects and images falling under the scope of our problem, proper formal training is hardly applicable here. It was found more appropriate to define the abstract category of multi-part objects using a limited number of local and global grouping criteria whose definitions were selected and validated on the basis of a new interactive methodology [8]. Details about the selected criteria appear below.

The GraphSearch deterministic algorithm [7] is used to build-up potential pieces of silhouettes. Due to the huge number of possible solutions, most nodes must be removed in order to keep the search under control. This is obtained by rejecting paths with intersections and by applying local grouping criteria. At the end of this procedure, a list of nodes containing paths of fifteen CCPs is produced. This number is typically not enough for the silhouettes of the interest object to be complete. This step is followed by the combination of nodes, as explained later. Scores are computed for the obtained final paths using global criteria.

3.1. Local criteria

A new node must be validated by distance and intersection tests in order to be added to the tree. Some valid nodes are not to be kept either. More tree nodes are removed by the application of a set of local criteria to CCP paths. Failure to satisfy a set of conditions, in the form of a boolean equation, results in pruning of that node. Early removal of a node is more efficient. Hence, conditions are typically very restrictive in the early levels, and more permissive later on.

The ten local grouping criteria are listed in Table 1. Each has a simple formal definition. For instance, the two continuity criteria have a linear scale. A null angular difference between tangents at extreme points of the CCPs worths 100%. A 180 degrees difference worths zero. The arithmetic mean of scores on a path is computed. The number of

parts criterion is computed using the number of concavities. It starts by filtering angles through the silhouette by grouping small angle variations together. Gaps on the boundary are filled. One is added to the result to take into consideration the main body of the multi-part object. Other criteria are also simple to compute.

Table 1. Local criteria encoding

Number	Criteria
1	Continuity
2	Unused
4	Distance
8	Number of parts
16	Surface area
32	Total length
64	Opening
128	Filled continuity
256	Obtuse angles
512	Hole proportion
1024	Early closure

Tree is generated up to level fifteen. After each level generation, the pruning process is executed. Local criteria are expressed in the form of a boolean equation, different for each level of the tree. Table 2 present the equations in postfix format for the fifteen levels of the tree. First number of each pair is the criterion code as displayed in first column of Table 1. Second number is a fixed threshold applied. Operators express logical combination of criteria. Another set of similar equations is used for large images.

From level ten, elimination process changes. All nodes are quite good because they meet tight conditions. The slackened criteria may not be sufficient to reduce combinatorial explosion. To help with that situation, a partial score is computed for each node using a linear combination of individual criteria scores. Then, only the best nodes are kept. This elimination process arises whenever the number of nodes at a given level exceeds 2000. It is then reduced down to 1500.

3.2. Global criteria and scores

Associating a quality score to a node is at the heart of the proposed method. A node is considered as either a completed or under-construction object. The former gets its value from the main score, the latter from the partial score. The main score has been mainly developed by Randrianarisoa et al. [8]. The ten global grouping criteria, with their weights, are as follows: closure (5), visual balance (1), compactness (1), number of parts (1), filled continuity (5), gap distribution (1), object-image position (1), surface area (5), border effect (1), hole proportion (1). Only four ele-

Table 2. Boolean equations

Level	Associated equation
1	32 10
2	1 80 4 10 * 1 65 4 2 * + 4 1 +
3	1 80 4 10 * 1 65 4 4 * + 4 2 +
4	1 75 4 10 * 1 60 4 5 * + 4 2 + 512 10 * 32 50 * 1024 2 * 64 20 *
5	1 75 4 15 * 1 60 4 6 * + 512 5 + 512 12 * 128 55 * 256 2 * 1024 3 *
6	1 70 4 15 * 1 60 4 7 * + 512 5 + 128 55 * 8 1 8 2 + * 32 80 * 512 12 * 1024 4 *
7	1 70 4 15 * 1 60 4 8 * + 1 55 4 5 * + 256 3 * 1024 5 *
8	1 65 4 18 * 1 60 4 10 * + 128 65 * 32 100 * 512 15 * 1024 6 *
9	1 65 4 20 * 1 60 4 12 * + 128 75 * 512 15 * 32 120 * 1024 7 *
10	8 2 1 65 * 128 75 * 16 0.5 * 256 3 * 8 4 1 70 128 75 * 1 65 128 65 * 512 5 * + * 16 1 * 256 4 * + 512 15 * 4 20 * 32 150 * 1024 8 *
11 to 15	4 20 128 70 512 12 X

ments are kept in the partial score, with unit weights: number of parts, filled continuity, surface area, hole proportion.

4. Combination of level 15 nodes

The combination procedure aims at producing complete silhouettes from partial paths. Only the best 500 best nodes are kept, according to partial score. The minimum number of primitives to add to a node is set to 5 in experiments. Thus, the number of CCPs in combinations ranges from 20 to 30. Before accepting a tested combination, it must be validated by intersection checks and further application of local criteria. In order to deal with both complete and incomplete objects, threshold values are permissive. There are many objects with a silhouette of more than 30 CCPs. An optional step takes combinations of level fifteen and combines them again with level fifteen nodes. This way, solutions may contain from 25 to 45 CCPs, which is now enough for typical test images. After combination and potential recombination, short CCPs initially removed are used to complete small gaps on the silhouette of the object.

5. Results

In Figure 3, the main score of each solution appears in parentheses. This is an interesting case with background clutter and internal textures and markings. Many CCPs create bridges to go around some parts, like the plane's right

wing. Pruning those CCPs helps converging to correct answers faster.

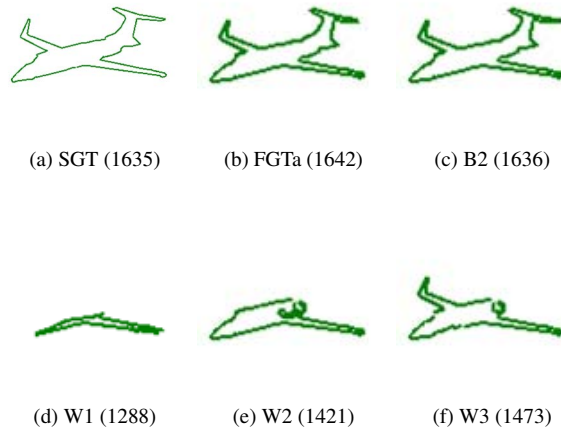


Figure 3. SGT/FGT: 2 best, 3 worst

Table 3 show the position of the most similar solution computed by the method for a variety of images. The total number of retained solutions is in parentheses in the first column.

Table 3. Position of the most similar solutions

Image (#End-answers)	Position	Similarity
Juice (53)	3 (96%)	93%
Airplane (449)	2 (100%)	89%
Water can (2)	1 (100%)	100%
Angel fish (35)	3 (94%)	99%
Stool (427)	13 (97%)	92%
Toy truck (1277)	49 (96%)	87%
Fish (157)	1 (100%)	85%
Bike (857)	21 (98%)	90%
Hand (7)	3 (71%)	97%
Man (158)	1 (100%)	93%

CCP-Usage tells how much each CCP is used in the generated solutions. In CCP-Usage maps, the darker the CCP, the more it is used at a given step. Figure 4 shows CCP-Usage maps for the airplane. Background and texture noise rapidly disappear through levels of the tree and combinations.

Average main score, similarity, precision, and recall are computed for a sample of one hundred solutions (the blue points in the graphs) at each algorithmic step. For main score, the solutions are: a third from the best scores, a third from the middle ones, and the last third from the worst ones. For the other three graphs, the best solutions are retained.

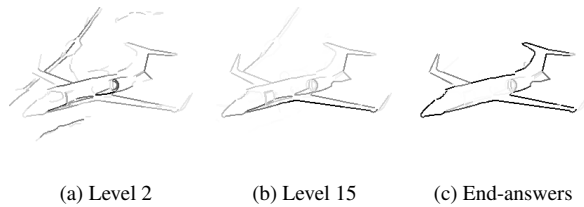


Figure 4. CCP-Usage maps

Figures 5 and 6 show airplane graphs. Similar graphs are obtained with the other images. They are meant to show the importance of each algorithmic step and the need for a good sampling of solutions.

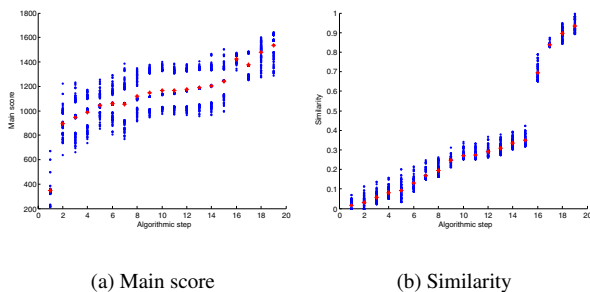


Figure 5. Main score and similarity graphs

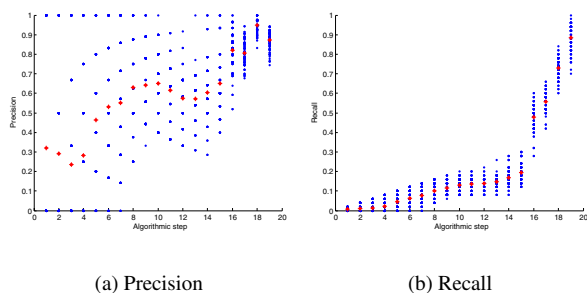


Figure 6. Precision and recall graphs

6. Conclusions

A simple set of explicit local and global grouping criteria are combined to detect multi-part objects in complex images. A deterministic generic detection method based on parallel search tree expansion and pruning was developed and applied to a variety of noisy contour primitive maps. Input images show significant amounts of internal textures,

markings, and background structure. The method is able to target the main subject of an image as long as it corresponds to a multi-part object of the proper complexity. From the obtained object silhouette, it is straightforward to extract the corresponding region. On a Pentium 4 2.0 GHz with 1024 MB of RAM, computation times range from 25 seconds for a simple image like the water can to 6.5 minutes for a complex image like the toy truck. These numbers may be improved by various optimizations including parallel implementation.

7. Acknowledgment

This work is supported by an NSERC discovery grant.

References

- [1] G. Dorko and C. Schmid. Selection of scale-invariant parts for object recognition. In *Proc. of the 9th International Conference on Computer Vision (ICCV'03)*, pages 634–639, 2003.
- [2] J. Elder and S. Zucker. Computing contour closure. In *Proc. 4th European Conference on Computer Vision*, pages 399–412, 1996.
- [3] F. J. Estrada and A. D. Jepson. Perceptual grouping for contour extraction. In *Proc. 17th International Conference on Pattern Recognition*, pages 32–35, 2004.
- [4] S. Mahamud, L. R. Williams, K. K. Thornber, and K. Xu. Segmentation of multiple salient closed contours from real images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):433–444, April 2003.
- [5] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):530–549, May 2004.
- [6] M. Mokhtari and R. Bergevin. Generic multi-scale segmentation and curve approximation method. In *LNCS 2106: Scale-Space and Morphology in Computer Vision, Third International Conference*, pages 227–235, 2001.
- [7] K. Nilsson. *Principles of Artificial Intelligence*. Tioga, Palo Alto, CA, 1980.
- [8] V. Randrianarisoa, J.-F. Bernier, and R. Bergevin. Detection of multi-part objects by top-down perceptual grouping. In *Proc. 2nd Canadian Conference on Computer and Robot Vision*, pages 536–543, 2005.
- [9] D. Walther, U. Rutishauser, C. Koch, and P. Perona. Selective visual attention enables learning and recognition of multiple objects in cluttered scenes. *Computer Vision and Image Understanding*, pages 41–63, 2005.