Detection of Unexpected Multi-Part Objects from Segmented Contour Maps

R. Bergevin *, J.-F. Bernier

*Department of Electrical and Computer Engineering, Laval University,
Quebec City, Canada G1K 7P4*

Abstract

A novel method is proposed to detect multi-part objects of unknown specific shape and appearance in natural images. It consists in first extracting a strictly over-segmented map of circular arcs and straight-line segments from an edge map. Each obtained constant-curvature contour primitive has an unknown origin which may be the external boundary of an interesting object, the textured or marked region enclosed by that boundary, or the external background region. The following processing steps identify, in a systematic yet efficient way, which groups of ordered contour primitives form a complete boundary of proper multi-part shape. Multiple detections are ranked with the top boundaries best satisfying a combination of global shape grouping criteria. Experimental results confirm the unique potential of the method to identify, in images of variable complexity, actual boundaries of multi-part objects as diverse as an airplane, a stool, a bicycle, a fish, and a toy truck.

Key words: Multi-part object detection, segmented contour map, grouping constraints, global shape grouping criteria.

1 Introduction

Whether you are sitting looking at a picture or walking in a new place, your attention is often captured by some part of what is in your field of view. Attention is considered a key factor in cognitive science given the limited processing capabilities of humans [1], the quantity of information in the observed scenes, and the variety of perceptual and cognitive tasks and contexts. Attention is sometimes considered a purely reactive behavior where, for instance, a generic

---

* Corresponding author. Tel.: 1-418-656-2131x5173; Fax: 1-418-656-3159.
  Email address: `bergevin@gel.ulaval.ca` (R. Bergevin).

big, moving, red thing leads you to move your eyes in its direction. However, low-level, reactive behaviors are not sufficient to attain the sophistication made possible by higher-level cognitive capabilities. For instance, recognizing a known object would elicit a higher-level attention capture and a specific intelligent reaction.

Apart from these two extreme situations, other types of attention capture may also exist involving various forms and combinations of detection, localization, and recognition of objects at different levels of specificity. For instance, a widely discussed question about human vision is whether detection, localization, and recognition are sequential or parallel processes and, in the former case, what their temporal relationships are. The same questions arise in computer vision. Clearly, it is impossible to recognize a horse and still not know whether a horse is present in the field of view. However, as exemplified by recent research in computer vision, it could be possible to detect that a horse is present in the field of view without being able to localize it precisely. In that case, a specific model of a horse, e.g. a set of, possibly structured, local appearance features, is required a priori. In a different scenario, a four legged animal could also be detected and localized before its specific recognition as a horse in a given pose and posture. That scenario makes more sense from a complexity point of view since the number of known objects is huge and their possible appearances are also numerous. One way toward efficient high-level attention capture is to detect whether an unexpected but interesting object is present in the field of view and if yes, to recognize it more specifically with its pose and posture.

In this paper, a computer vision method for high-level attention capture is proposed on that basis. A major hypothesis made is that no explicit or specific model is available a priori for the shape and appearance of potential objects of interest. Hence, the addressed problem is the detection of unexpected but interesting objects. This differs from problems referred to as object (category) detection, localization, or recognition in the literature, where an explicit (possibly learned) object model of limited genericity is available a priori. One may argue that interest is often highly context-dependent. Humans may be assumed to be generally aware of context but not computer vision systems. Besides, if context is often an important factor, out-of-context vision is not uncommon even for humans. For instance, consider situations such as opening an image book, a photo album or a web image search result window, looking at the first shots of a movie in a theater, zapping television channels, or even awakening in the morning. Out-of-context detection and localization of unexpected or unknown objects by computer vision systems is to become more common as applications get more sophisticated and challenging. If specific context is not known, it means one either has to do without it or recognize it first. The latter is however a problem at least as difficult as object detection and localization itself. In this paper, specific context is assumed to be

unknown and detection and localization of unexpected but interesting objects is sought for.

2   Related Work

Model-based hypothesis-verification methods for object recognition match local image features to a known model in order to find the pose of a specific object [2]. Recent visual object categorization methods apply supervised statistical learning methods in order to build a more generic model for each of a number of known object categories [3–6]. Though local features such as interest points or contour fragments clearly bring useful information about image contents [7], their use in object categorization still consists in identifying features specific to a given class of object appearances. Despite rapid progress, state-of-the-art methods still have important limitations, namely their lack of a precise localization of the detected objects, their limited invariance to viewpoint, their possible confusion between different detectors responding positively to a test image, and their choice of categories which is rarely discussed and appears to be ad-hoc. Scaling of the methods to a large number of categories is another important problem yet to be fully addressed.

As a result, the previous methods may hardly be seen as properly modeling high-level attention capture. In a sense, human attention capture implies that one does not have specific expectations and hence does not look or visually search for a particular object in a given pose but is instead surprised by what is observed. Hence, having a classifier for each type of possibly interesting object appearance and verifying in turn which one is in fact present in an image is unrealistic and inefficient for high-level attention capture. Detecting that an interesting but unexpected object is present in an image implies that it is also localized. Delimiting the region occupied by an interesting object in a static image is both useful and easy for humans. This is sometimes referred to as figure-ground discrimination. In computer vision, the related problem of partitionning an image into *object* and *background* regions is referred to as figure-ground segmentation. This is still a fundamental problem in computer vision with no existing general solution yet. The method proposed in this paper offers a possible solution to that problem for multi-part objects, a large and very generic class of objects.

State-of-the-art segmentation methods in computer vision rely on very generic contextual knowledge, e.g. interesting objects are compact, contrasting, and of at least a certain size, but they fail to provide a satisfying means of detecting and localizing objects of interest in challenging situations. On the other hand, state-of-the-art detection methods typically rely on specific contexts e.g. the object of interest is a horse seen from the side with the head facing

left and whose formal model has specific parameters, in order to provide the best results in proper contexts that is, good combinations of detection and false alarm rates. The method proposed in this paper efficiently detects and precisely localizes, out of the huge number of possibilities, unexpected but interesting objects in a given contour primitive map obtained from a single intensity image. It relies on contextual knowledge more generic than state-of-the-art detection methods but more specific than state-of-the-art segmentation methods. The proposed approach is to use an extended set of grouping criteria and constraints in order to find objects of interesting shapes. Much evidence has been published that perceptual grouping of contour primitives is important in human vision [8,9]. However, as discussed in [10], few results are known concerning high-level grouping for object detection and localization.

In computer vision literature, generic grouping criteria for shapes are typically limited to simple ones, e.g. local continuity and smoothness [11], or global closure [12], [13] and convexity [14]. Elder and Zucker [12] used a shortest-path algorithm to find maximum-likelihood closed contours. Even though no criterion was used to explicitly target multi-part objects, they presented a result where a single contour for the complete shadow of a multi-part wooden doll was extracted. However, very few if any confounding primitives seemed to be present on the object, on its shadow, or in the background (the actual primitive map was not shown). Different partial contours were also extracted on the doll and in between the doll and its shadow. No attempt was made at identifying the most interesting contours. Targeting simpler shapes and high efficiency, Jacobs [14] used a constraint (threshold) on a global salience criterion based on the relative amount of gaps in a convex closed shape. Efficiency resulted mainly from using monotonic constraints, meaning that a subset of primitives could be rejected as soon as it did not satisfy the constraint since adding more primitives could not make the larger subset acceptable. While convexity is properly monotonic, [14] also used the salience constraint as if it were monotonic, even though it is not since adding a long convexity-preserving segment with no gap to a convex group may increase its salience. In fact, relatively few grouping criteria and constraints are proposed in literature and they are mostly monotonic. Unfortunately, this choice limits the possibilities of proposed methods in terms of detecting shapes of interest. A comparative study of [12] and two competing methods was made by Wang et al. [15]. For natural images of animals, the optimal boundary has a simple near-convex shape not representative of the animal shape. The limitation to monotonic criteria and near-convex shapes is also typical of previous saliency-based methods [11]. For instance, saliency networks impose that an optimal curve must be composed of sub-curves that are themselves optimal, a property equivalent to monotonicity and referred to as *extensibility* in [16].

With complex shapes, effective grouping criteria are more likely to be non-monotonic. Non-monotonic grouping refers to the fact that the value of a given

formal criterion is not changing monotonically i.e. either always increasing or always decreasing as the number of grouped (added, merged) primitives increase. For instance, a group of two primitives may get a higher score on a given criterion than another group of two primitives but the same third primitive added to the first group may give a lower score than when added to the second group. This makes it much more difficult to reach even a local optimum simply by following an ascending or descending gradient in the value of a single grouping criterion. In the method proposed in this paper, many such non-monotonic grouping criteria are combined in the objective function which makes typical deterministic and even random discrete optimization techniques most likely ineffective. The effect of grouping non-monotonicity becomes more apparent as the number of elements in each group grows from two to three and more, as is the case in the proposed method. In practice, an important consequence of the non-monotonicity of the grouping criteria is that, as the subsets of primitives considered increase in size, the constraints need to be adapted. For instance, as the shape and interest of a partial object becomes clearer, constraints may appear or disappear or more simply diminish or increase. Previous search-based methods avoided this added complexity by implicitly or explicitly assuming monotonic criteria and constraints. A recent method by Estrada and Jepson [13] extracts salient non-convex contours by developing a search tree, enforcing fixed constraints and rejecting a fixed proportion of less promising partial contours.

In this paper, the hypothesis is made that global, non-monotonic grouping criteria are needed in order to detect and localize unexpected but interesting objects directly from a noisy and cluttered primitive map. A single generic set of grouping criteria is used in the proposed detection method. Our approach is a generalization of [14] where the criteria and constraints are more numerous and non-monotonic. The main contextual knowledge to be relied on in defining the shape-based grouping criteria is that objects of interest are of intermediate complexity. One may argue that intermediate complexity objects are generally interesting since they are complex enough to be discriminated from each other and from distracting structures in complex natural images, and still simple enough to be detected, described, understood, and memorized using limited processing capabilities [17,1]. Here, objects of intermediate shape complexity are called multi-part objects and assumed to be of general interest in out-of-context situations. 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. In a projected view of a 3D multi-part object, evidence for parts lies in the successive concavities and convexities along its external boundary [18]. In practice, as primitive maps are noisy, a single criterion is not sufficient to always localize the most interesting objects. For instance, an object may have the proper number of concavities but be of less interest than a more complete or larger object. A trade-off among a

5

number of criteria is used to define the degree of interest of each possible object boundary.

While many non-monotonic criteria are needed to evaluate the complex boundary of a multi-part object, fewer monotonic criteria could be sufficient to directly detect its parts. Part extraction using contour grouping algorithms have been proposed by a number of authors [19–22]. In practice, as a result of the limited evidence for part connections in images, the performance of part extraction algorithms depends on either assuming very simple part shapes and object structure or having previously determined the region occupied by the interesting object in a cluttered scene [23]. In the former case, generic object detection and localization must precede specific part-based shape analysis and recognition. As an extreme example, sophisticated skeleton-based shape analysis methods always assume binary silhouettes as given or easily obtained. The method proposed in this paper aims at detecting and localizing interesting objects in a generic context, prior to a more specific shape analysis, categorization, or recognition. Basic grouping constraints impose upper or lower constraining or limiting values (thresholds) to basic criteria evaluating partial boundaries while global grouping criteria evaluate complete boundaries. Basic grouping criteria and constraints are used to limit the number of partial boundaries considered while global criteria are used to rank the obtained complete boundaries. Criteria and constraints are detailed in Section 5.

Even though our grouping criteria and constraints are based on shape, it is important to emphasize that our goal is not to directly compare different shapes for classification, recognition, or retrieval purposes. In those cases, binary shape silhouettes are usually the input to be processed and a description or model of each shape is required. This problem is addressed using a formal representation for the shape, e.g. skeletons [24] or codons [25]. A shape matching procedure then computes the similarity of any pair of shapes described in the chosen representation. Here, the goal is different. It is to detect and localize the shapes in images. Furthermore, that detection is not obtained by searching for a shape similar to a given model. Instead, the localization of the shapes likely to capture attention, being of general interest, is sought for. Finally, as the goal is to detect and localize interesting objects with no specific shape model, active contours could be seen as a simpler alternative to grouping primitives from a strictly over-segmented input map. However, to be successful, active contours need to be properly initialized. This is likely to necessitate manual intervention in the context of this problem since shape, position, posture, internal textures and markings, and background structures may vary widely from one image to another. Besides, active contours deformations are likely to be affected by those internal textures and background structures present in natural images.

The following two sections describe the shape grouping primitives and the

6

detection problem. The proposed detection method is then detailed in Section 5. Experimental results are presented in Section 6, which is followed by a final discussion and conclusion.

3 Input Data

Given an image of an unexpected multi-part object, the basic goal is to identify the region occupied by the object in the image. From a raw intensity or color image, state-of-the-art region-based segmentation methods do not always produce satisfying object detection results. Being obtained on the basis of regional photometric statistics, the resulting region map is likely to suffer from both over- and under-segmentation. That is, no single region is to cover the whole object, when its shape and appearance are complex, and some regions are to overlap the object and the image background, when background is complex and object pose is unconstrained. A strictly over-segmented region map would make it possible to recover the interest region solely by grouping connected regions, with no need to either backtrack (completely removing the last added regions) or split (partially removing) regions, as with heuristic search or split-and-merge segmentation. Besides, even if a strictly over-segmentated region map were obtainable, region grouping constraints and criteria needed to efficiently and effectively identify the best interest region on the basis of its shape would be difficult to define.

By comparison, an edge map is trivially strictly over-segmented since each edgel is either on the object boundary or not. However, edgels are both too numerous and local to be a primitive of choice for efficient and effective shape-based grouping. Conversely, discrete image contours obtained by linking neighboring edgels may be small in number but they suffer from two related difficulties. First, grouping constraints and criteria are not likely to be effective unless strong regularities are assumed in the shapes of interest. Secondly, a raw contour map may also suffer from under-segmentation when contrast is poor and image contours overlap boundary and background or boundary and object body (interior). However, this under-segmentation is likely to be more limited since there is no spill out effect typical of region growing.

The best trade-off appears to be a strictly over-segmented map of simple contour primitives. For instance, a small group of simple contour primitives is directly indicative of a too simple shape, which is not the case with a small number of uniform regions or image contours. In particular, constant-curvature contour primitives (CCPs) are amenable to both strict over-segmentation and efficient and effective shape-based grouping for multi-part object detection. Hence, as a preprocessing step to the proposed method, a constant-curvature contour primitive map need to be obtained from the static intensity image.

Extracting CCP primitives directly from a static intensity image is possible using one of many proposed algorithms in the literature, for instance those based on the Hough Transform [26,27]. However, it may be difficult to enforce strict over-segmentation with such direct methods since they typically need to trade off primitive continuity and size. Besides, it may be difficult to properly cover the object boundary without extracting a very large number of overlapping primitives.

Computing straight-line-segment or CCP approximations of open and closed discrete image contours is a better approach which is likely to remove contour under-segmentation. Boundary CCPs are not to overlap the image background or object body unless a perfect accidental alignment of features (continuous in position and curvature) arises between the external boundary and the background or the object body. Many algorithms are available in the literature to segment and approximate contours. Unfortunately, most of them do not properly enforce invariance properties needed in generic contexts, for instance invariance to partial occlusion and scale. The selected algorithm [28] was developed for generic shape-based object recognition and, as such, it explicitly attempts to enforce the needed invariance properties. In that algorithm, each discrete contour is processed independently. A number of straight-line-segment approximations is first built, each starting at one of the discrete contour points. Segments are grown on each side of the starting point, were applicable, with a new segment generated when the appoximation error exceeds a scale-related threshold. A second step attempts to group consecutive straight-line segments into circular arcs along each produced polyline. Again, a new group is started when the approximation error exceeds another scale-related threshold. A final step selects, for each discrete contour, the best CCP approximation among those produced using a combination of statistical criteria. The resulting CCP map for a given image consists in all CCPs in the selected approximations of the different discrete contours. Those contours were obtained using a 8-neighbor edgel linking algorithm. A thinned edgel map was previously obtained using a basic Canny filter. The straight-line segments and circular arcs of the produced CCP map are the input data for the proposed detection method.

In this paper, the algorithm computed the CCP maps using a single one-pixel scale. As a result, the obtained approximations retain the significant shape features along the object boundary, the number of CCPs is manageable, and the CCP maps may be assumed to be strictly over-segmented. A more systematic analysis of the strict over-segmentation of the produced CCP maps would be interesting but it was not attempted here. Given the choice of primitives and the relative complexity of the object shapes, any under-segmentation is to be local and limited in size, not affecting much the boundary growing and evaluation criteria of the proposed method. Figure 2 presents five typical input CCP maps (circular arcs are red and straight-line segments are blue) obtained by

preprocessing the five images in Figure 1. Noise in the input CCP maps arise from low contrast in some areas, small structures, and local contour clutter. The proposed method is meant to be robust to that noise.
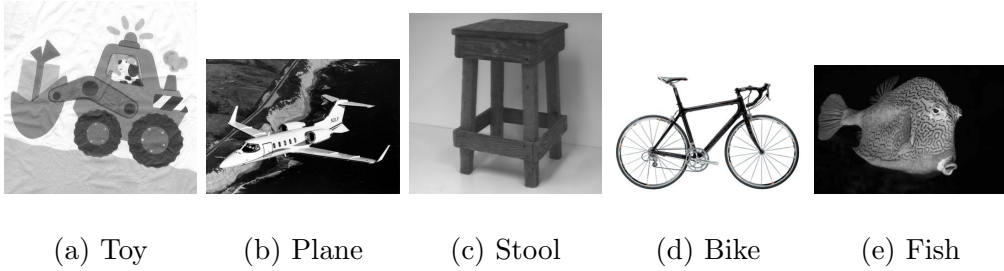

| (a) Toy | (b) Plane | (c) Stool | (d) Bike | (e) Fish |

Fig. 1. Images of multi-part objects. Each dimension is between 256 and 512 pixels.

| (a) 427 | (b) 552 | (c) 203 | (d) 205 | (e) 640 |
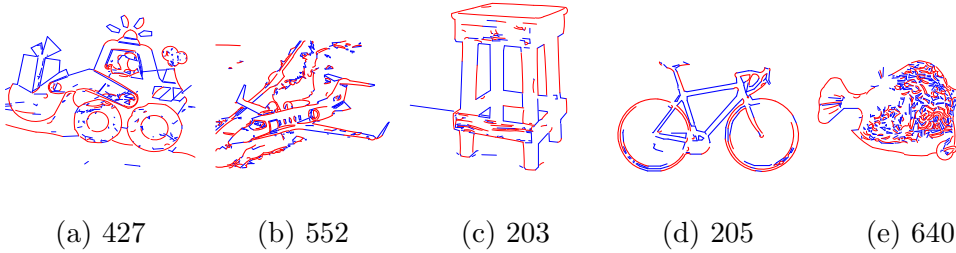
Fig. 2. Input maps. Preprocessing produced the indicated number of CCPs.

4 Detection Problem

Figure 3(b) reproduces the constant-curvature contour primitive (CCP) map of the image in Figure 3(a). Any proposed solution to the multi-part object detection problem has to be an ordered group of CCPs selected from the input map. Such a solution is assumed to represent a complete clockwise or counter-clockwise traversal of the object boundary. From this ordered group of primitives, the region occupied by the object in the image may easily be recovered.

The detection problem thus consists in selecting and ordering the proper subset of primitives from the map. Considering an average map of 400 CCPs and complete boundaries with, say, 30 CCPs, the number of possible complete boundaries is about 10^{86}. In the generic context where a large number of variations in object appearance is possible, as discussed earlier, it is hard to eliminate any of those boundaries a priori. In the proposed method, the detection problem is formalized as a combinatorial optimization problem and the best solution is found according to a multi-criteria cost (or value) function. Given the non-monotonicity of the individual grouping criteria, as discussed

earlier, a random or heuristic search for the best boundary or a small number of good-enough boundaries would not work. Hence, even if practically impossible, an evaluation of every possible boundary is theoretically needed in order to guarantee that the best is identified. In any case, a fundamental requirement is a scoring function able to provide an assessment of the value of any selected boundary from the huge number of possible ones. More importantly, that scoring function has to be consistent with human judgment, a truly difficult requirement.

As discussed earlier, the global shape grouping criteria on the basis of which each boundary is to be scored were not readily available from human or computer vision literature. SAFE (Subjectivity and Formalism Explicitly), a novel methodology [29], has been used to identify a first set of criteria for the present implementation of the detection method. The main idea behind SAFE is to make explicit the role and choices of the system designer (the subjectivity) and to consider them separately from the resulting system (the formalism). A common formalism for grouping algorithms is used, based on the combinatorial optimization of a multi-criteria objective function. The explicit choices to be made are the experimental context (input and output formats, test data, and ground-truth), the conceptual criteria, their formalization and combination, the optimization algorithm, and the metric to objectively compare output with ground-truth. The rationale behind the SAFE methodology is to make it easier to document and compare competing methods at the high level of main design choices. An interactive tool was built to implement the SAFE methodology and it was used to identify the set of conceptual and formal shape grouping criteria used in the detection method proposed in this paper. These criteria are to be discussed in more detail in Section 5.

Figure 3(c) is the best boundary as interactively selected by a human. In SAFE, it is referred to as the subjective ground truth (SGT). SGT is not known by the proposed detection method. It will only serve to assess the quality of its results. Let us assume that a scoring function exists for boundaries. The formal ground truth (FGT), is the boundary with the highest score, out of the huge number of possible boundaries. FGT is usually not known either as it would require that all boundaries be considered and that each of them be scored. More practically, a subset of the possible boundaries is considered and the one with the highest score is selected. The selected optimal boundary is only an approximation of FGT. It is referred to as FGTa. 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 effective FGTa as close as possible to SGT, for natural images of multi-part objects. Figure 3(d) shows an example of a randomly generated boundary with 30 CCPs. Each CCP is displayed with a small arrow and a number since the same set of CCPs in a different order, or with one CCP pointing in the

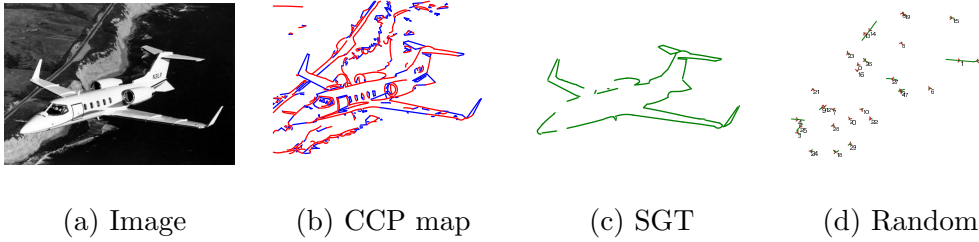(a) Image   (b) CCP map   (c) SGT   (d) Random

Fig. 3. Basic concept. A random boundary is unlikely to be similar to SGT.

opposite direction, represents a different boundary. In the proposed method, grouping constraints applied to partial shapes help discard such a poor boundary early on. Besides, the developed scoring function based on global shape criteria would assign a very low value to that poorly structured boundary. The proposed method represents a first attempt at using the SAFE methodology. As a result, the design choices made could certainly be optimized in various ways. Despite this fact, it is remarkable that the best scoring boundaries produced, including FGTa, are actually quite similar to SGT (see Figure 9).

It is important to realize that grouping, as understood here, is not limited to partitioning the set of primitives e.g. by clustering or labeling approaches. This agrees with the definition in [30] but not with the more restricted definition in [31]. In fact, the previous two generic grouping algorithms from literature are not suitable for our detection problem. In [31], monotonic criteria (in fact, consistency constraints) are explicitly assumed, such as the co-circularity of segments. As for [30], groups are obtained by hierarchical pairing, enforcing the same constraints (called restrictions) and using the same criterion (called similarity) at each level.

5   Detection Method

For each input CCP map, a manageable number of complete boundaries must be produced efficiently. A complete boundary consists of an ordered group of primitives with no self-intersection and a satisfied closure constraint. Each produced group should enclose a shape that scores well with respect to a combination of generic multi-part grouping criteria. Besides, the boundary of a multi-part object actually present in the original image should be among the best-scoring shapes produced. For instance, at most a thousand or so complete boundaries with the ground-truth shape in the first few percentiles is sought for, in a matter of a few minutes or less.

The proposed strategy is to build the groups in parallel and iteratively, starting with as many groups as there are primitives in the CCP map. Then, for

each existing group, as many new groups are created as there are primitives remaining in the map, adding one of the primitives to each new group. And so on, until each group is complete or rejected. This systematic generation and elimination of ordered subsets of contour primitives from the map is the only way to make sure each possibly interesting multi-part shape is given due consideration.

In order to decide if a group is to be rejected at a given iteration, it is tested with respect to a number of grouping constraints. As discussed earlier, the grouping constraints need to be adaptive, in terms of which ones are used at each iteration and what limiting values they have. At each iteration, all groups failing to satisfy the tested constraints are removed from consideration.

The grouping procedure can be likened to traversing a tree from each starting primitive, adding a primitive at each level and pruning a branch when a group fails to satisfy the constraints. In the tree, each node corresponds to a boundary with one CCP added from the partial boundary associated with its parent node. Each parent node generates as many child nodes as there are CCPs left i.e. CCPs from the input map that are not part of its associated partial boundary. In practice the number of generated nodes may be less since each one has to be validated by two so-called preliminary tests, as described below. Once nodes are generated at a given level of the tree, they are pruned using the grouping constraints. The retained nodes are the parent nodes of the next generation and grouping iteration. Figure 4 illustrates the generation and pruning of nodes at level-1 (left) and level-2 (right) of the tree. An empty partial boundary is associated with the single parent node at level-1. The number of generated nodes corresponds to the number of validated CCPs in the input map. That is, a partial boundary is initiated at each CCP in the map. Pruning removes some of these 1-CCP partial boundaries. As will be shown in Table 1, that first pruning is done on the basis of the length of the CCPs. In the figure, only two 1-CCP partial boundaries are retained. Each is a parent node at level-2 where it will generate a number of 2-CCP partial boundaries that are to be pruned according to an adapted set of grouping constraints, and so on.

As mentioned earlier, besides the need to satisfy the grouping constraints, each CCP to be added to an existing partial boundary must be validated by two preliminary monotonic tests. The first and simplest test consists in verifying that the distance between connecting endpoints is less than a limiting value. That is, the start endpoint of the oriented CCP under test must be within a fixed distance from the finish endpoint of the last CCP of the partial boundary. The limiting value was empirically set at 30 pixels. Nevertheless, gaps larger than 30 pixels may be present in the complete boundaries obtained by merging partial boundaries, as discussed in Section 5.3. The second test is more complex and it is only performed when the first test is successful. It consists in
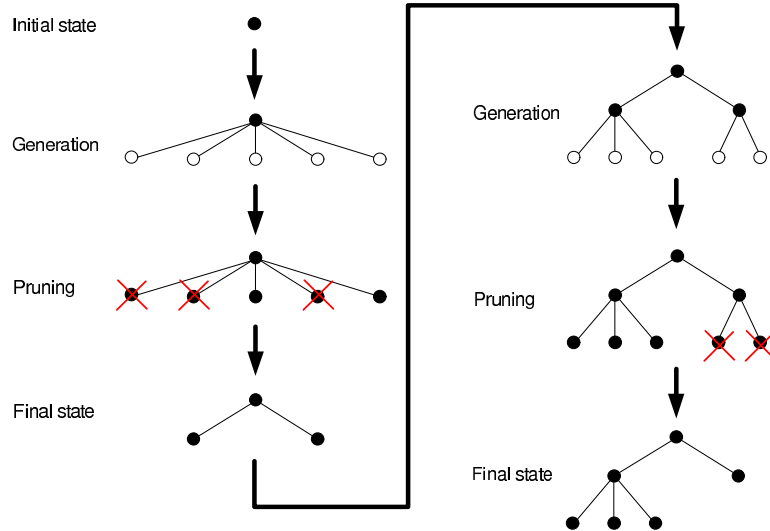
Fig. 4. Node generation and pruning. Grouping constraints controls expansion.

verifying that the CCP under test does not intersect with the partial boundary. This condition eliminates a large number of possible boundaries. Gaps in the partial boundary are filled with virtual straight-line segments to perform the test. In the present implementation, the intersections are recomputed each time a CCP is tested.

Fifteen levels of the tree are developed, each with a generation and pruning stage as described before. A unique CCP from the map is associated to each level on any path from a given root to a leaf node. These 15-node paths are usually partial. That is, the corresponding chain of fifteen CCPs is typically not long enough to totally cover the actual boundary of a multi-part object. As explained in Section 5.3, this first step is followed by one or two merging steps of partial boundaries associated with level-15 nodes. Final scores are computed for the obtained complete boundaries using global shape criteria to be described in Section 5.4. In the reported experiments, CCP maps have from about 100 to more than 600 CCPs. Since boundaries are made of an ordered chain of CCPs selected from the map, each CCP may be added to a partial boundary from either end. Hence, the number of generated nodes at each level is actually twice the number of remaining CCPs. On the other end, very short CCPs are only considered at the end to fill some holes in the produced complete boundaries. In the present implementation, a minimum size of 5 pixels is enforced. For the CCP map in Figure 3(b), the number of retained partial boundaries (parent nodes) at each of the fifteen levels is as follows: 1:131, 2:166, 3:278, 4:301, 5:163, 6:242, 7:623, 8:389, 9:1082, 10:874, 11:690, 12:780, 13:819, 14:937, 15:500. Figure 5 presents a sample of those partial boundaries at levels 7, 10, and 15.

A block diagram of the boundary grouping algorithm appears in Figure 6. Its worst-case complexity corresponds to the generation of all possible ordered
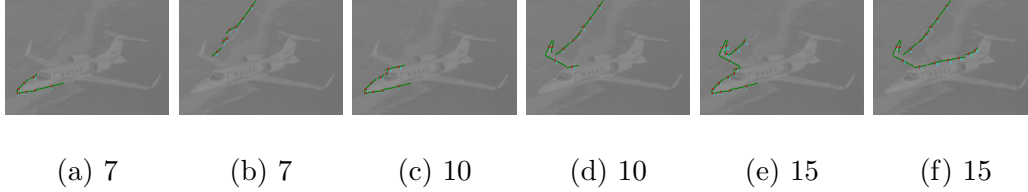
13

Fig. 5. Sample partial boundaries. Level numbers are indicated.

subsets of CCPs from the input map. That is, each generated group at any given iteration satisfies the combined grouping constraint at that iteration. As explained in the previous section, this worst-case complexity is exponential and with a typical number of CCPs in the input map, the number of subsets of a given size typical of actual multi-part object boundaries is unrealistically high. Hence, the combined grouping constraints at the different iterations must be strong enough to end up removing all but a small number of the generated groups. Besides, the few retained complete boundaries should be multi-part shapes and include the actual multi-part object shape present in the image. More details on the different steps of the algorithm appear in the following sections.

5.1 Basic Grouping Criteria and Constraints

Each continuation of a partial boundary consists in adding a given CCP from the map. The added CCP must first be validated according to a distance test and an intersection test, as described earlier. Once the added CCP is validated, the obtained boundary is further tested using a boolean combination of grouping constraints. Each individual constraint corresponds to a limiting value on one of ten basic grouping criteria. Each criterion is associated with a code number. Those numbers are arbitrary. They are used to efficiently encode the boolean expressions defining the combination of constraints at the different levels. The ten basic grouping criteria are: *Continuity* (code 1), *Distance* (code 4), *Number of Parts* (code 8), *Surface Cover* (code 16), *Boundary Size* (code 32), *Contraction* (code 64), *Filled Continuity* (code 128), *Jaggedness* (code 256), *Lack of Cover* (code 512), and *Part Strength* (code 1024). These ten basic criteria were formalized using simple mathematical expressions. Given the number of times a constraint is tested, it is important to have a fast, even if approximate, result. A short description of each criterion is given next.

Continuity is the arithmetic mean of the differences of angles between tangents at the finish endpoint of each CCP and the start endpoint of the following CCP. A zero mean angle is worth 100%. A 180 degrees mean angle is worth 0%. In between, the scale is linear. A 100% continuity boundary has perfectly matching tangents between each pair of consecutive CCPs. Any
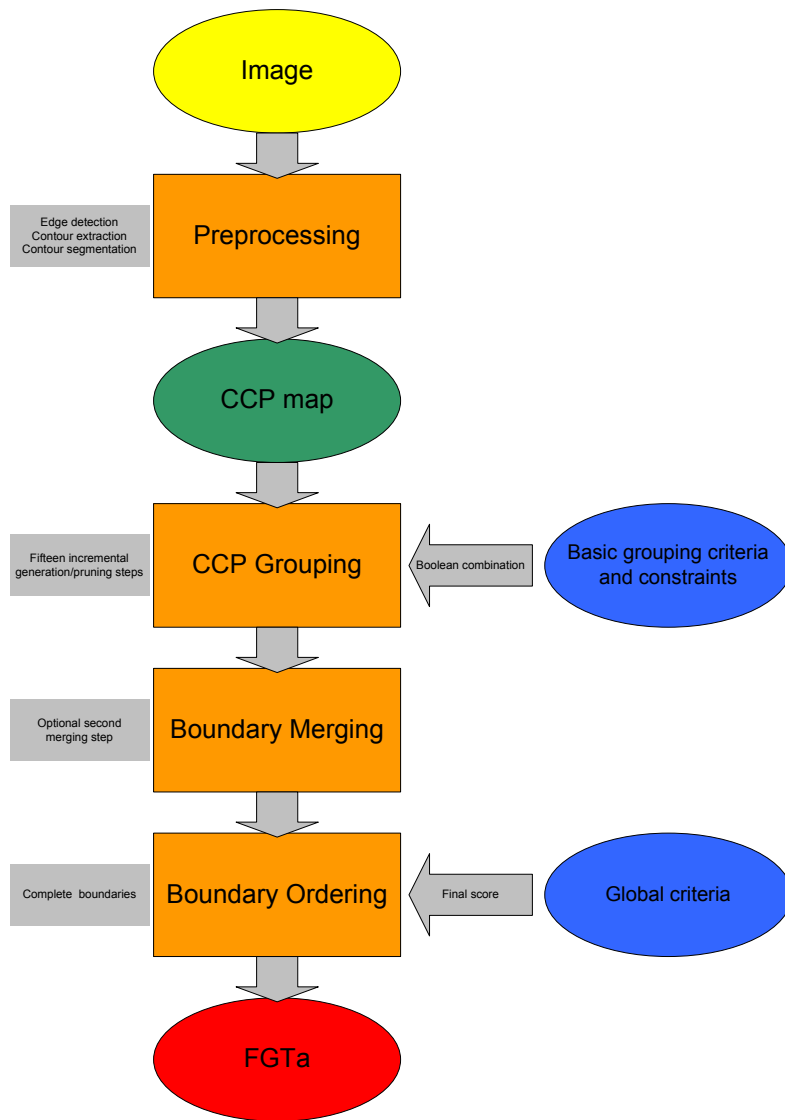
14

Fig. 6. Boundary grouping algorithm. The main output is an ordered set of complete boundaries. A secondary output is a set of open boundaries with 15 to 45 CCPs.

kind of perfectly smooth shape, whatever its complexity in terms of number of parts, would have perfect continuity. This criterion is ideally maximum.

Distance is the Euclidean distance from the finish endpoint of the partial boundary to the start endpoint of the added CCP. It is ideally minimum.

Number of Parts is the number of concavities in the obtained boundary after closing it with a virtual straight-line segment. A number of consecutive segments with small angles between their tangents is replaced by a single virtual segment oriented according to the sum of the angles. Also, gaps between consecutive CCPs of the boundary are filled using virtual straight-line segments. One is added to the resulting number of concavities to take

into consideration the main body of the multi-part object. The ideal 100% value is with four or more parts. With a single part, the value is 0%. Two and three parts have 33% and 67%, respectively.

Surface Cover is the ratio (expressed as a percentage) of the surface enclosed by the virtually closed partial boundary (last primitive is connected with first one using a virtual straight-line segment) to the area of the bounding box of the primitives in the input CCP map. The ideal maximum value is 100%.

Boundary Size is the sum of the total length of the gaps between consecutive CCPs and the total length of the CCPs themselves. Larger values are preferred.

Contraction is the ratio of the length of the virtual closing straight-line segment (see *Number of Parts*) to *Boundary Size*. In the first iterations, the ideal maximum value is 100% in order to go straight away from the starting point. Later on, the ideal minimum value is 0% in order to have a long complete boundary.

Filled Continuity is the same as the continuity criterion but with virtual straight-line segments filling the gaps between consecutive CCPs.

Jaggedness is the number of obtuse angles along the boundary. Smaller values are preferred. It is complementary to but less important than the continuity criterion.

Lack of Cover is the ratio between the total length of the gaps between consecutive CCPs and *Boundary Size*. Its value is normalized between 0% and 100%. A small value is preferred. This criterion is the complement of the salience fraction used by Jacobs for convex groups [14].

Part Strength is the minimum distance between the final endpoint of the last CCP added to the partial boundary and the other CCPs of the partial boundary. The distances considered are not limited to endpoints. Larger values are preferred in order to limit the occurrence of very thin parts or connections.

It is important that the criteria be generic enough to be effective in the context of the multi-part object detection task. That is, this single set of shape grouping criteria must permit to detect and localize interesting objects of different shapes and appearances. As explained earlier, the criteria have to operate in a context where a large number of intrinsic and extrinsic variations in object appearance are possible. In particular, compared to state-of-the-art methods for detecting expected objects, a greater invariance to changing shape, texture, markings, and viewpoint are needed.

In the present implementation, no attempt was made at systematically optimizing the set of criteria and its formalization. Some criteria are inspired by global shape grouping criteria identified using the SAFE methodology (see Section 5.4). Others are inspired by criteria from the perceptual grouping literature. The main contribution of the proposed method, with respect to those

criteria, is the variable subset used at each level and the adaptive boolean combination and limiting values for the corresponding constraints. These designer choices were made empirically but not systematically. They are detailed next.

5.2 Boolean Combination of Constraints

To be retained as a continuation of the partial boundary, a tested CCP must satisfy a given boolean expression specific to its position on the boundary under construction. As a general rule, constraints combined in the boolean expression are more restrictive in the early levels and more permissive later on. Also, alternative combinations of constraints with different limiting values are possible at a given level.

Table 1 presents the boolean expressions for the fifteen levels of the tree. In each expression, an operand is a constraint represented by a pair of numbers (the criterion code and its associated limiting value). The expressions are written in postfix notation. The three possible combining operators are: AND (*), OR (+), and FAND (X). The first two are standard boolean operators applied to a single pair of constraints. The last one is similar to the first one but it can be applied to an arbitrary number of constraints.

As an example, let us first examine the boolean expression at level 3. In this case, if either of two pairs of constraints is satisfied, the boundary produced by adding the tested CCP to the partial boundary is retained. Each pair imposes a limiting value to the *Continuity* (code 1) and *Distance* (code 4) criteria. That is, either an 80% *Continuity* is combined with a *Distance* of 10 pixels or a 65% *Continuity* is combined with a *Distance* of 2 pixels. A third possibility is a *Distance* of 1 pixel with no constraint on the *Continuity* criterion.

A common pattern is used all the way down the tree. It expresses the trade-off between continuity and distance. Large distance asks for good continuity and smaller distances allow more discontinuities. Notice also the *Boundary Size* criterion (code 32) applied one level out of two in low levels, and more frequently at the end. This way, a chance of survival is left to shorter partial boundaries. By level four, *Lack of Cover* (code 512), *Boundary Size* (code 32), *Part Strength* (code 1024), and *Contraction* (code 64) criteria become meaningful. *Part Strength* is used until the end because every newcomer must be tested for correct positioning. Level six is the first time the *Number of parts* (code 8) criterion is applied. Level ten is quite challenging for groups as all criteria are part of the boolean expression, except *Contraction* (code 64) which is never used after level 4.

Given the complexity of the multi-part shapes and the simplicity of the CCPs,

Table 1
Boolean expressions for the combination of constraints.

Level	Boolean expression
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

it rarely happens that ten CCPs cover the whole boundary. Level 11 to 15
have relaxed constraints so that proper continuation of promising boundaries
is possible. Since the relaxed constraints represent a risk of combinatorial
explosion, an upper limit on the number of retained partial boundaries at
each node is set to 2500. In fact, when this upper limit is passed, the partial
boundaries are ordered and only the best 2000 are used in the next iteration.
A partial score is computed in order to order the different nodes at a given
level. Four of the ten criteria, also members of the global shape criteria to be
described in Section 5.4, are combined with unit weights to produce the score.
They are: *Number of parts* (code 8), *Filled Continuity* (code 128), *Surface
Cover* (code 16), and *Lack of Cover* (code 512).

The boolean expressions in Table 1 were used in all reported experiments.
In the present implementation, neither the expressions nor the parameters
were optimized. Boolean expressions were chosen by hand on the basis of our
understanding of the role and functioning of each formalized criterion in the
context of the problem, irrespective of the specific dataset. Values of parame-
ters in the expressions were adjusted using trial and error on the input maps
of Figure 2. No systematic procedure was used and no systematic optimization
was attempted. Only a small number of parameter values in the expressions
were actually selected by experimenting with this small number of images

(representing less than a third of the test images reported in this paper). The values were also strongly constrained. All were selected from a small number of possible values. A single set of expressions and values appears sufficient as long as image dimensions are within a limited range.

5.3 Boundary Merging

In practice, groups of fifteen CCPs rarely produce a satisfying boundary. However, it was found difficult to efficiently and effectively keep on adding CCPs one at the time with further relaxed constraints. Instead, direct merging of pairs of level-15 nodes is attempted. The rationale is that complex partial boundaries that merge well are likely to belong to the same object boundary. Once again, each potential merge has to be validated using an intersection test and relaxed constraints for a majority of criteria. Boundaries for partly occluded objects may be accepted given the relaxed constraints.

The two merged boundaries result in a longer boundary which corresponds to the second boundary appended to the first one. The ordering of CCPs is kept but CCPs are renumbered for the second boundary, starting at one plus the number of CCPs in the first boundary. When a partial chain of CCPs in the final part of the first boundary is identical to a partial chain of CCPs in the initial part of the second boundary, one of the two partial chains is removed in order to avoid overlap and duplication of CCPs in the merged boundary. In other words, merging a subset of primitives in the final part of a partial boundary to another partial boundary is acceptable (see Figure 7). The minimum number of primitives to be kept from the overlapping boundary is set to 5. Hence, the number of CCPs in a merged boundary is from 20 to 30. In order to limit the complexity of the merging step, only the best 500 level-15 nodes are kept, according to their partial scores. Those 500 nodes correspond to 250 000 candidate pairs for merging.

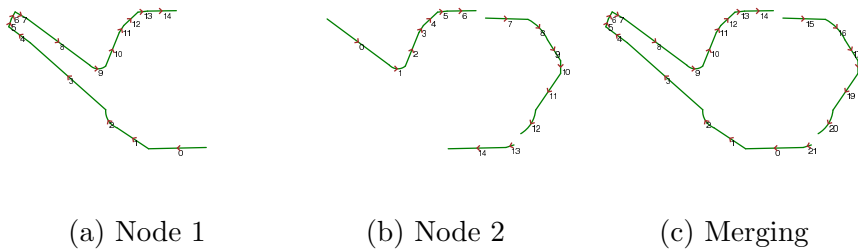(a) Node 1   (b) Node 2   (c) Merging

Fig. 7. Merging. The two partial boundaries overlap.

In our test images, some multi-part object boundaries have more than 30 CCPs. A final merging step further combines merged boundaries that are

not yet complete with level-15 nodes. A merged boundary is complete when *Contraction* is less than or equal to 10% and a virtual closing straight-line segment has no intersections with CCPs of the boundary. Given the same constraint that at least 5 CCPs are kept from the overlapping boundary, the twice-merged boundaries may have from 25 to 45 CCPs.

The second merging step is optional. It will not occur when the score of the best complete boundary is at least as high as the best score in the merged boundaries and the number of CCPs in the best complete boundary is at least 15% of the number of CCPs in the input map. Final scores are now used instead of partial scores. They are computed from the global grouping criteria, as described in the following section. Once boundary merging is over, very short CCPs initially removed are brought back to try filling small gaps in the obtained boundaries. The final score is re-computed if some gaps are filled to produce a refined boundary.

5.4   Global Criteria and Final Scores

Global grouping criteria evaluate each obtained complete boundary in order to find its rank among the retained boundaries. In particular, the best boundary obtained, for a given input CCP map, is our approximation to the formal ground truth (FGTa). Its similarity to the subjective ground truth (SGT) is our measure of success for the method (see Section 6).

A set of ten global grouping criteria was defined and formalized by Randrianarisoa et al. [29]. The first four global criteria were also part of the basic criteria used to rank partial boundaries. They are: *Number of Parts*, *Filled Continuity*, *Surface Cover*, and *Lack of Cover*. The *Lack of Cover* criterion is actually reformulated in order to favor only boundaries with limited gaps covering between 5% and 12% of the boundary. The other six basic grouping criteria are not part of the global grouping criteria as they would be less effective in properly ranking the obtained complete boundaries. They are: *Continuity*, *Distance*, *Boundary Size*, *Contraction*, *Jaggedness*, and *Part Strength*. Six new global grouping criteria are introduced in order to benefit from now having a complete shape. They are: *Cover*, *Visual Balance*, *Non Compactness*, *Gap Distribution*, *Centering*, and *Foreground*. Each criterion is normalized between 0% and 100% and the maximum value is preferred. A short description of the new criteria is given next.

Cover is the ratio between the total length of the CCPs and *Boundary Size*. It is complementary to the original *Lack of Cover* criterion.
Visual Balance is the complement of the normalized distance between the mass center of the CCPs on the boundary and the mass center of the bound-

ing box of those CCPs. The mass center of the CCPs is the mass center of the individual mass centers. A unit weight is given to each CCP, short or long. The normalization factor is half the diagonal of the bounding box.

Non Compactness is the complement of the ratio between 4π times the area and the square of *Boundary Size*. A circle is minimally non compact.

Gap Distribution is computed similarly to *Visual Balance* except that the mass center of the bounding box is replaced by the mass center of the virtual primitives filling the gaps between consecutive primitives on the boundary.

Centering is also computed similarly to *Visual Balance* except that the mass center of the bounding box is replaced by the center of the image or CCP map.

Foreground is a binary criterion that receives a maximum value unless the boundary passes near a side of the image, in which case it receives a minimum value.

The final score of a complete boundary is computed as a weighted sum of the ten global criteria. All but three criteria have unit weights in the sum. *Cover*, *Filled Continuity*, and *Surface Cover* each have a weight of 5. Although it was not found possible to normalize the final score, the range of values obtained is similar for the different images of multi-part objects tested.

6 Experimental Results

Four views on the obtained results are presented. Together, they indicate that the proposed method is remarkably apt at detecting multi-part objects in natural images of variable complexity, on the basis of the developed basic and global criteria and scores.

6.1 Qualitative FGT/SGT Comparison

The best boundary found (best final score boundary, or FGTa) and other retained boundaries need to be compared to the human segmentation (SGT). The comparison is in two parts. The qualitative part consists in displaying the final CCP map for retained boundaries next to the SGT map (with no CCP ordering shown). In this first view on the obtained results, it will be clear that the best retained boundaries, including the FGTa, are quite similar to SGT and much more so than the worst retained boundaries.

Figures 8 to 10 present results for three typical CCP maps. Final scores of displayed boundaries are shown. The retained boundary with the best final score is FGTa. Retained boundaries are ordered using the final score computed

as a weighted sum of the ten global grouping criteria. In each figure, the displayed retained boundaries are the best one (FGTa), the second best, the worst, and the next to worst.

The CCP map of the toy truck (Figure 8) is difficult since the method must face more than one good object. Despite the numerous possibilities, the final FGTa is at the proper complexity and quite similar to the SGT reference. Given the large variety of multi-part shapes, it is difficult to perfectly reproduce the SGT. In fact, the computed final score of SGT is 1663, compared to 1721 for the FGTa. That means, the proposed method considers FGTa as a better multi-part object than SGT. The CCP map of the plane (Figure 9) is probably the most difficult one. It has background clutter and internal textures and markings. As a result, many spurious CCPs create bridges to go around some parts, like the plane's right wing. Hence, some boundaries include only a subset of object parts. However, their final score is lower than the final score of more complete boundaries. The best scoring boundary obtained covers most of the important parts of SGT. In the CCP map of the stool (Figure 10), many internal CCPs, junctions, and holes increase the complexity of the detection task. Fortunately, the proposed method eliminates most partial boundaries rapidly and the FGTa obtained is very similar to SGT.

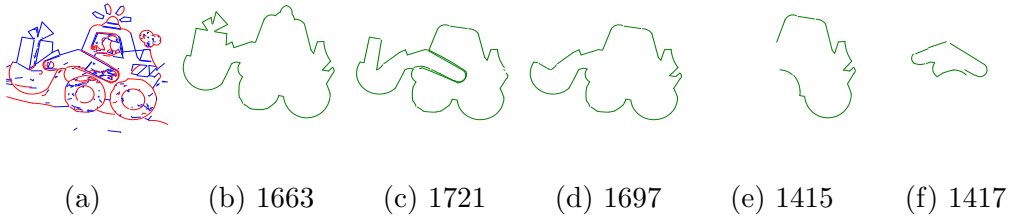(a)   (b) 1663   (c) 1721   (d) 1697   (e) 1415   (f) 1417

Fig. 8. Toy. From left to right: CCP map, ground-truth boundary (SGT), best retained boundary (FGTa), second best, worst retained boundary, next to worst.

(a)   (b) 1582   (c) 1642   (d) 1636   (e) 1288   (f) 1410
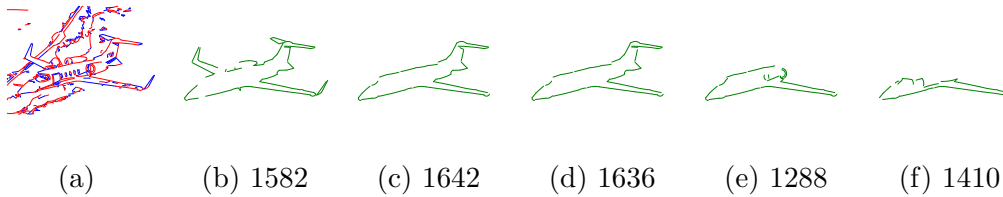
Fig. 9. Plane. From left to right: CCP map, ground-truth boundary (SGT), best retained boundary (FGTa), second best, worst retained boundary, next to worst.

The previous three examples demonstrate qualitatively the effectiveness of the proposed method in detecting unexpected objects of interest of proper complexity in natural images, delimiting precisely the region occupied by the object in the image.
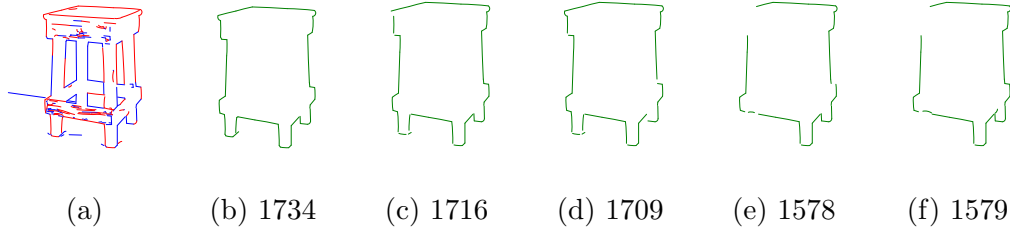
| (a) | (b) 1734 | (c) 1716 | (d) 1709 | (e) 1578 | (f) 1579 |

Fig. 10. Stool. From left to right: CCP map, ground-truth boundary (SGT), best retained boundary (FGTa), second best, worst retained boundary, next to worst.

6.2 Quantitative FGT/SGT Comparison

The quantitative part of the FGT/SGT comparison consists in formally expressing the differences and similarities in terms of the CCPs in each map and their ordering. A new informative measure is defined to compare retained boundaries to SGT. *Similarity* is the ratio between the total length of consecutive CCPs in the two boundaries that are common and in the same order and the total length of the CCPs in the longest of the two boundaries. To compute the similarity, common subsets of consecutive CCPs that are in the same order in the two boundaries are first obtained. For instance, the CCPs at position 4,5,6, and 7 in boundary 1 may be the same as CCPs at position 11,12,13, and 14 in boundary 2. The similarity is the sum of the lengths of those four CCPs divided by the sum of the lengths of all CCPs for the longest boundary. In practice, a similarity above 50% indicates notable similitudes with some large differences, while a similarity above 70% usually means all differences are minor. A similarity above 90% is obtained for boundaries almost identical to SGT that is, with mostly meaningless differences. In the following section, two further standard quantitative measures are to be used to compare partial and final retained boundaries to SGT: *Precision* and *Recall*.

The quantitative FGT/SGT comparison differs from state-of-the-art methods for detecting expected objects where features voting for a true positive detection are rarely validated as actually member of the object of interest. That is, neither the completeness of the feature set, in terms of covering the whole set of ground-truth features, nor its exactness, in terms of consisting only of ground-truth features, are estimated. Hence, the type of comparison performed here is more demanding and the results obtained have to be judged in that respect.

In order to obtain this second view on the results, the set of retained boundaries is first ordered according to the final scores. The similarity measure introduced earlier is then computed for each retained boundary. Finally, the boundary most similar to SGT is identified in the ordered set and its rank is obtained.
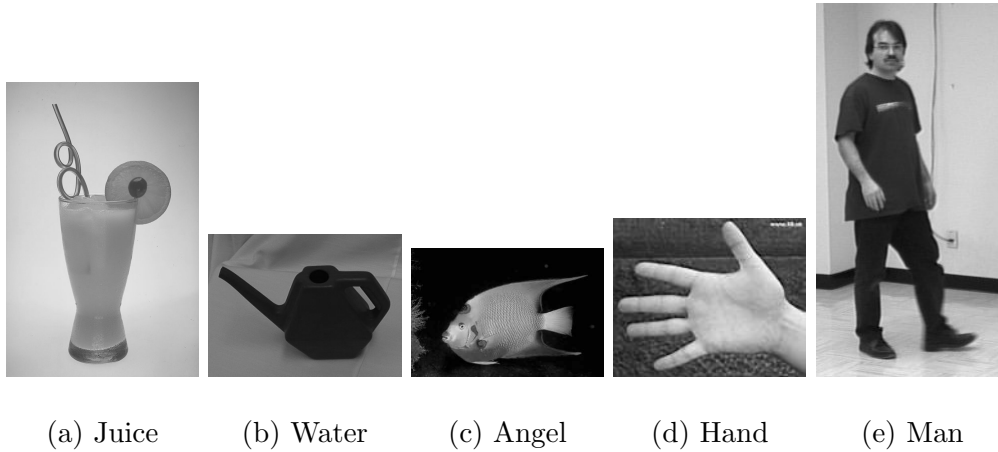
(a) Juice   (b) Water   (c) Angel   (d) Hand   (e) Man

Fig. 11. Images of multi-part objects. Each dimension is between 124 and 741 pixels.

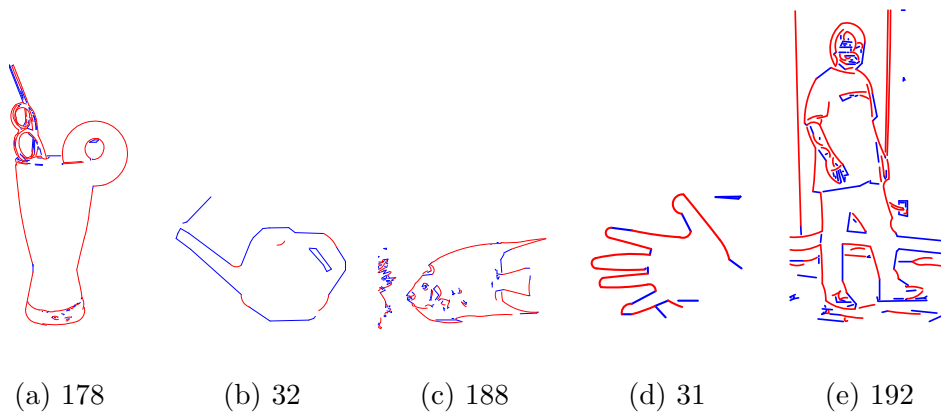(a) 178   (b) 32   (c) 188   (d) 31   (e) 192

Fig. 12. Input maps. Preprocessing produced the indicated number of CCPs.

Table 2 shows the obtained rank and the corresponding similarity of the retained boundary most similar to SGT, for ten images of multi-part objects. The first five images and their CCP maps appeared in Figures 1 and 2. The last five images and their CCP maps appear in Figures 11 and 12. Rank is an integer between one (best retained boundary according to final score) and the number of retained boundaries (third column). The latter is always much smaller than the huge combinatorial number of possible boundaries, as discussed in Section 4. Rank is also expressed as a grade (100% being ideal) using $Grade = 100(1 - (Rank - 1)/number)$ rounded to the nearest integer. The results indicate that the most similar boundary is always close (either in absolute or relative terms) to the best boundary according to the final score. Running times are discussed in Section 8.

As expected, the number of retained boundaries is much less than the huge number of possible boundaries. Even the worst retained boundaries (according to their final scores) contain good partial boundaries that respected the

24

constraints at each of the fifteen levels and at the merging steps. Hence, the rejected boundaries are usually much worse than the retained ones. For instance, a boundary such as the one in Figure 3(d) is always rejected. Moreover, the final score of retained boundaries is a good indication of their formal similarity to SGT, which means their ranking is effective and FGTa is close to SGT.

Table 2
Statistics for the retained boundary most similar to SGT.

CCP map	Number	Rank	Grade (%)	Similarity	Time (sec)
Toy	1259	27	98	0.87	386
Plane	236	2	100	0.68	227
Stool	393	1	100	0.98	307
Bike	857	1	100	0.81	375
Fish	157	1	100	0.85	256
Juice	53	3	96	0.93	82
Water	2	1	100	1.00	25
Angel	35	3	94	0.99	71
Hand	7	3	71	0.97	62
Man	158	1	100	0.93	217

As indicated by the previous results, objects with a part structure visible at a proper scale successfully produce an FGTa that is actually quite close to SGT and among the best ranked retained boundaries. According to the similarity values obtained, only minor or meaningless differences are expected. In practice, the Fish, the Toy, and the Bike images have some visible differences between their most similar boundary and their SGT. The Plane has the worst similarity in the previous ten test cases. Its input map is probably the most difficult one. Its SGT contains a number of large gaps. More importantly, it has a total of 61 CCPs which is too many given the maximum of two merging steps in the current implementation (retained boundaries are limited to a maximum of 45 CCPs). For this reason, many CCPs from SGT are not in the achieved FGTa which results in a low recall (see below) at 59%. Despite these current limitations, the achieved FGTa does not overlap the background or the internal texture. Hence, most CCPs from the retained FGTa are in SGT. In fact, FGTa has a precision (see below) of 94%.

The final two views on obtained results are presented next. They track performance measures step-by-step as processing unfolds.

A third view on the obtained results consists in performance graphs displaying the distribution of four performance measures at the different algorithmic steps. Four types of graphs are produced: the final score (Figure 13(a)), the similarity (Figure 13(b)), the precision (Figure 14(a)), and the recall (Figure 14(b)). Formal final score and similarity measures were defined earlier. Formal precision and recall are defined as follows. *Precision* is computed as the number of CCPs in the boundary that are present in the SGT, divided by the total number of CCPs in the boundary. It can be seen as the degree of purity of a boundary. *Recall* is computed as the number of CCPs in the boundary that are present in the SGT, divided by the number of CCPs in the SGT. It corresponds to the proportion of CCPs from SGT that are successfully found.

In each graph, algorithmic steps 1 to 15 correspond to the levels of the tree, 16 is the first merging, 17 is the optional merging, 18 is the cleaned complete boundaries list (only the few best of very similar boundaries are kept), and 19 is the refined complete boundaries, where very short primitives may be added in order to fill small gaps. At each step, a point is drawn in each graph according to the value of its performance measure (final score, similarity, precision, or recall) for each boundary in a sample set of one hundred boundaries. For the final score graph, the sample boundaries are obtained in the following way: a third from the best scores, a third from the middle ones, and the last third from the worst scores. For the precision and recall graphs, the samples correspond to the best one hundred boundaries according to the final score. Finally, for the similarity graph, the best one hundred boundaries according to the similarity are selected. The asterisk is the average value of the sample boundaries.

Figures 13(a) and 13(b) display the final score and similarity graphs for the toy truck. Final scores, from steps 2 to 15, are from about 600 to more than 1600 that is, a range of around 1000. This confirms the high diversity of partial boundaries considered at the different steps. Merging is seen to be effective since it rapidly concentrates the scores to the top values. Similarity grows pretty linearly with step number, with an important discontinuity at the first merging step. One may also notice that the range of the final scores is reduced at step 16 as merging adds new criteria. Average final score and similarity show a clear improvement at that step too. Otherwise, the trends are as expected with a regular increase in performance as the algorithm progresses and better partial boundaries are expanded.

Figures 14(a) and 14(b) display the precision and recall graphs for the toy truck. The precision behavior is intimately linked to the criteria used. In the first levels, precision takes only a few values due to the small number of possi-
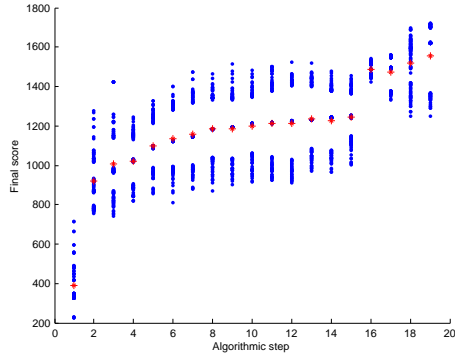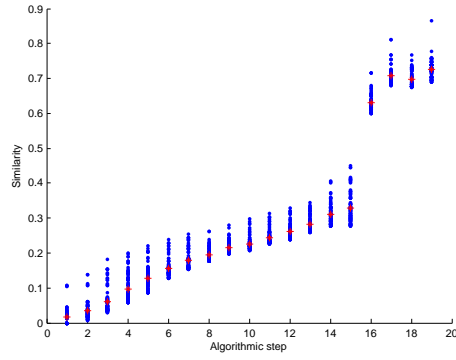
(a) Final score   (b) Similarity

Fig. 13. Final score and similarity graphs for the toy truck. Asterisk is average value.

bilities. This creates graphs with a typically rising tendency from level 1 to 10 where severe constraints force the selection of CCPs on the actual boundary of the plane. Those constraints are relaxed thereafter in order to find missing pieces of the boundary. That permits spurious CCPs to enter boundaries, explaining the temporary drop in precision between levels 11 and 15. This is necessary in order not to miss good complete boundaries. Recall also rises linearly during the boundaries expansion. At each step, the recall score tops at the number of CCPs in a boundary, divided by the number of CCPs in the SGT. So, the 100% mark is attained only at the end. Precision, however, can top 100% right at the beginning, and even drop afterwards, if the number of CCPs in SGT is low enough. For instance, an early good boundary may be spoiled by adding spurious CCPs.
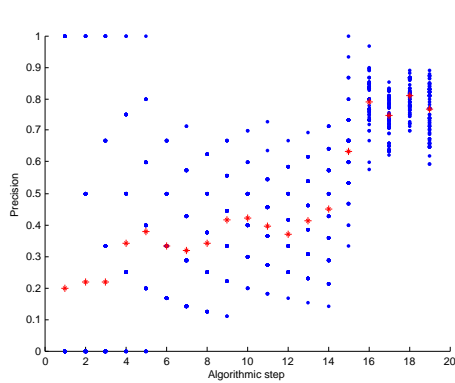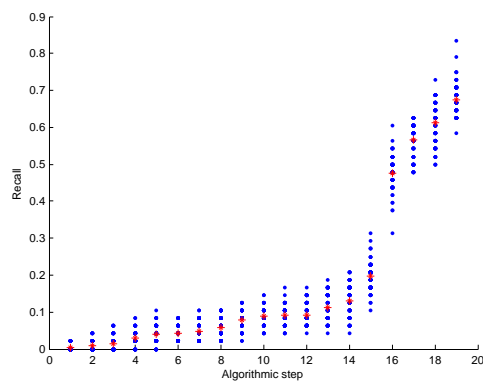
(a) Precision   (b) Recall

Fig. 14. Precision and recall graphs for the toy truck. Asterisk is average value.

27

As the fourth and final view on the obtained results, normalized maps are produced showing the relative number of times each CCP from the input map appears in boundaries at different algorithmic steps. Although not a performance indicator as such, it allows one to precisely and globally track when some background and texture features disappear or when some partial boundaries become stronger.

Figure 15 shows the normalized CCP maps for the plane. Background and texture noise rapidly disappear through levels of the tree, except in specific areas like the background road, due to its continuity and its very good connection with the plane. Still, merging successfully cut through this difficult case. Interestingly, the final normalized CCP map is a quite good probabilistic map for boundary CCPs, except for the plane's right wing.

(a) Level 2   (b) Level 3   (c) Level 10

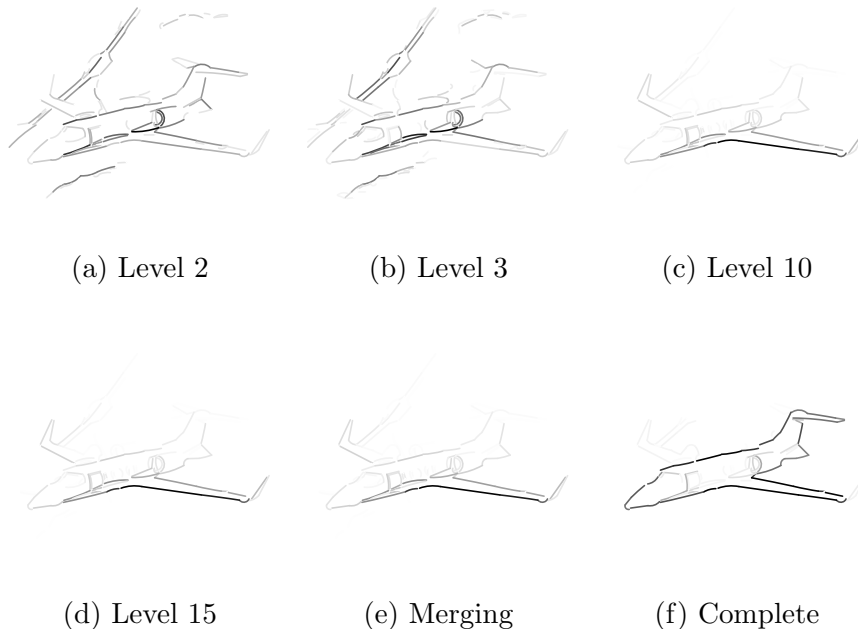(d) Level 15   (e) Merging   (f) Complete

Fig. 15. Normalized CCP maps. The darker a CCP, the more it is used in boundaries.

7   Discussion

The intended scope of the proposed method is mainly non-occluded macroscopic objects in natural settings, as could be observed by the human eye or a robot camera from proper viewpoints. Proper or adequate viewpoints ensure that images of multi-part objects are in the scope of the method. In images

obtained from proper viewpoints, the size of the object is large enough to produce a sufficiently precise input CCP map after preprocessing. Besides, the multipart structure of the object is visible. That is, the viewpoint is not such that the object looks much simpler that it is in reality. For instance, a top view of a lamp may not be indicative of the object's structure. An example of a proper viewpoint of a thin object is Figure 2e. Proper viewpoints are usually natural viewpoints. For most 3D objects, a very small proportion of viewpoints are inadequate.

As implemented, the method does not favor objects with very thin parts (e.g. characters) or high-frequency variations in the boundary contrast (e.g. a high-resolution view of tree with its leaves, a high-complexity shape with a very large number of parts). Nevertheless, even a very simple shape (e.g. a ball) could end-up being the most interesting in a given image if no other more interesting shape is present.

A better assessment of the scope of the proposed method may be obtained by looking at the eight additional input maps, ground-truths, and best retained boundaries presented in Figures 16, 17, and 18. For all eight images, the best retained boundary is quite good and it represents a significant achievement with respect to state-of-the-art methods (see below). The lamp, the butterfly, the mushroom, and the statue have near perfect precision and recall. In each case, only a few short ground-truth CCPs are not extracted. The best similarities obtained are above 96% for boundaries at ranks 1 to 3. The best similarities for the octopus, the cat, the compass, and the hat are from 72% (cat) to 84% (compass), all at rank 1 expect the octopus which is at rank 42 (out of 915 retained boundaries). The octopus has a perfect precision but a 60% recall due to small object structures with many CCPs. The cat has a 89% recall but a 67% precision due to internal structure in the obtained boundary. The compass has a 91% recall with a 80% precision. Finally, the hat has a 82% precision with a 69% recall. In this last case, the ground-truth is more ambiguous since the boy, the hat, and the bucket are different objects but they touch each other to form a complex structure.
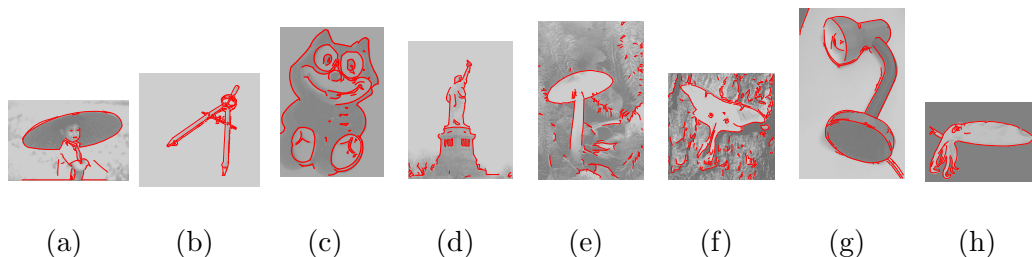

| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) |

Fig. 16. Additional input CCP maps. CCPs are superposed on each input image

As shown, tested images varied in complexity, some having less background clutter and internal textures and markings than others. Given that all possi-
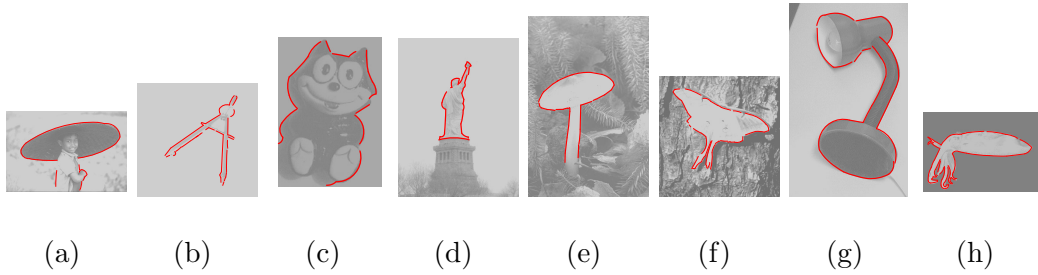
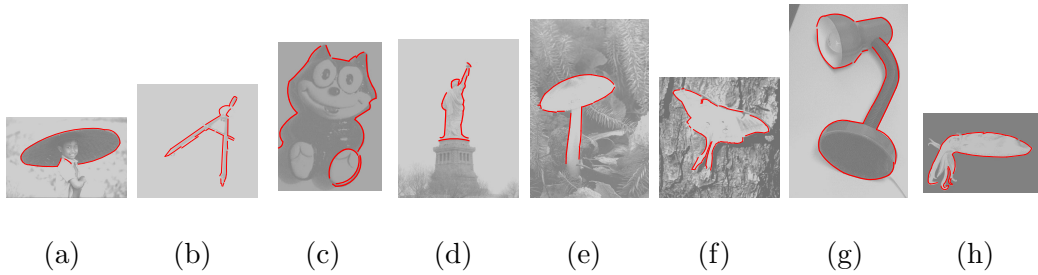Fig. 17. Ground-truths. SGT boundary is superposed on each input image

Fig. 18. Results. FGTa boundary is superposed on each input image

ble boundaries are considered, additional distractor CCPs in the background and inside the object could have no effect, apart from possibly longer running times. Similarly, images with more than one multi-part object in the foreground could produce boundaries for both objects in the first ranks, as long as each one respects global criteria just as well as the other. More likely, one object is to score higher than the other. A sequential detection of the foreground objects, removing CCPs of an object after its detection, is likely to succeed in such a case. This conjecture needs to be systematically tested in future work.

In the current implementation, the main limitation is that no complete boundaries are found in some CCPs maps that actually contain an interesting object. The problem is usually caused by a very irregular and/or disconnected boundary in the CCP map. Such a boundary arises, for instance, when an object has either very fine part structure, e.g. a flower or a tree, or fine texture connected to the boundary, e.g. an exotic butterfly, a tiger, a bear, or a zebra. Extraction of coarser scale edge and CCP maps could offer some help towards solving the figure-ground problem in such cases. However, automatic scale selection might not be easily obtained. Examples of problematic CCP maps are presented in Figure 19. The level at which no partial boundary satisfies the grouping constraints is indicated. A future improvement to the method could consists in relaxing the constraints when this situation arises.

An interesting method bearing similarities with our approach was proposed by Saund in the domain of sketches and drawings [32]. As here, image con-

30

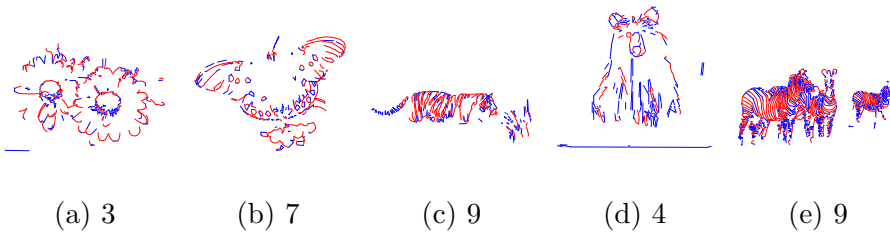(a) 3   (b) 7   (c) 9   (d) 4   (e) 9

Fig. 19. Problematic input CCP maps. The maximum level reached is indicated.

tours are initially segmented into fragments. However, the raw fragments are kept whereas the proposed method segments and approximates contours using CCPs primitives, as a result of the necessity to address the under-segmentation problem. In [32], basic grouping criteria are used to control path expansion the same way as done here. Three criteria are used (continuity, smoothness, and turning) which are somewhat specific to the problem domain and may not apply to multi-part objects. More importantly, those criteria express conflicting trade-offs, as in the proposed method, but instead of being combined to find the best contours of a given type (multi-part shapes in our case) they are used independently to find salient contours of different types (e.g. small shapes or smooth shapes). A larger number of basic criteria are defined in the proposed method and they are combined adaptively at each iteration of the tree expansion. Another major difference between the two methods is in the global evaluation criteria. In [32], compact or loosely convex shapes are favored whereas the proposed method prefers non-compact shapes. Both methods have a closure criterion but the proposed method has eight additional global criteria. Also, whereas backtracking best-first search is used to follow a single optimal path in [32], the proposed method follows all paths simultaneously with iterative pruning and no backtracking. Another difference between the two methods is that pre-computation of local groups e.g. possible junctions in a given neighborhood of a contour primitive is used in [32] to eliminate some paths a priori whereas the proposed method follows all paths with no a priori elimination.

A state-of-the-art boundary extraction method was recently defined, as an improvement to the best method among the three compared in [15]. Basically, an enclosed area criterion was introduced to better favor large objects [33]. This state-of-the-art graph-based optimization method uses a simple objective function (ratio of continuity to area) with a no-self-intersection constraint. Its output is the single best closed boundary. Multiple boundaries may also be obtained from a given image by removing the segments from the input map after each boundary extraction. This is in contrast with the proposed method where retained boundaries may share CCPs and additional boundaries may

31

overlap the best one. Figure 20 presents results obtained using this state-of-the-art method on original test cases used in this paper (default parameters using the authors' software available online). The obtained boundaries are either incomplete or they overlap the object and background significantly. The proposed method is more robust in that at least one of the retained boundaries is to be close to SGT. Moreover, that boundary is to be ranked high. Finally, each retained boundary is evaluated and its relative goodness is quantified by the proposed method. Figure 21 presents comparative results obtained by the two methods on test cases from [33]. Again, compared to the state-of-the-art method, the proposed method similarly or better captures the objects in the tested images.
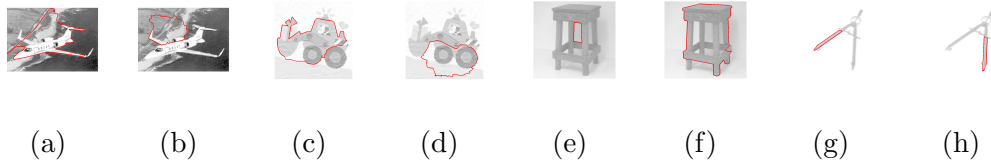

| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) |

Fig. 20. State-of-the-art results. Best two boundaries obtained by a state-of-the-art method on original test cases.

| (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) | (i) |

Fig. 21. Comparative results. Best boundaries obtained on state-of-the-art test cases. a-f: state-of-the-art-method, g-i: proposed method.

The main fundamental difference between the state-of-the-art method and the proposed method lies in their choice of grouping criteria. The state-of-the-art method is based on a single small set of simple grouping criteria whereas the proposed method has two different larger sets of grouping criteria, one for growing the boundaries and one for evaluating them. Apart from this fundamental difference, a number of reasons may explain their relative performance on specific images. For instance, preprocessing in the state-of-the-art method includes a line extraction step that allows filling gaps to merge different image contours. No such gap-filling is presently permitted at the preprocessing stage in the proposed method. This may allow the state-of-the-art method to end-up with a smaller number of segments after preprocessing a noisy edge map. A too large number of CCPs on the boundary may affect the performance of the proposed method given its current limitation to two merging steps. This was demonstrated with the plane image for which the best retained boundary did not include a part of the object. Most importantly, the proposed method produces solutions with limited overlap with the background and the internal

texture, as was shown with the plane image. In cases where the best obtained boundary is either incomplete or of poor shape, as indicated by the final score, the proposed method could also attempt further merging steps. In Figure 22, the best two boundaries obtained with the state-of-the-art method are compared with an incomplete boundary obtained in the current implementation. The latter could be further merged with a partial boundary obtained at level 15 to produce a complete boundary.

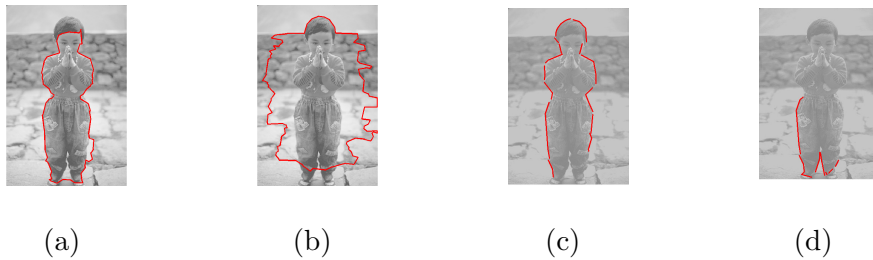(a)   (b)   (c)   (d)

Fig. 22. Large number of boundary CCPs. a-b: best two boundaries obtained by a state-of-the-art method, c: an incomplete boundary obtained using current implementation of proposed method, d: a complementary partial boundary obtained at level 15.

Another more fundamental but related reason for the varying relative performance is that the state-of-the-art method adds virtual segments between endpoints of real segments, allowing to fill even larger gaps between extracted segments. Again this may favor the state-of-the-art method when parts of the object boundary have poor contrast. However, virtual segments could make the number of primitives much too large given practical time and memory limitations. Hence, the state-of-the-art method is restricted to only add virtual segments of limited length. The choice of a proper threshold value is difficult. In some cases, a boundary gap is properly filled but, in other cases, the obtained boundary significantly overlaps the external background or the internal texture and markings. A third reason explaining some differences in performance is that the current implementation of the proposed method does not take image border into account. That is, an object touching the image border may not be properly extracted given its incomplete, open boundary. A virtual segment joining endpoints of two CCPs touching the image border would help the proposed method in such a case. However, this could again be at the expense of producing some partly under-segmented boundaries that include background structures.

7.1 Selection of Parameters

A direct consequence of the addressed problem and the selected approach is the large number of design decisions and parameters in the proposed method. For instance, if each basic grouping criterion could be used irrespective of the size of the partial boundary, any subset of grouping criteria could be used at each incremental grouping step. If 15 steps are used with 10 grouping criteria, there are 2^{150} possible combinations. At each step, a boolean expression is used to logically combine chosen criteria (a criterion may be used more than once in a given expression). Some of the grouping criteria do not make sense when the size of the partial boundary is too small or too large. At each grouping step, a choice was made on which criteria to use. No experimentation was made with the expressions themselves. As far as the number of retained boundaries is manageable, within acceptable time and memory limits, and the best boundaries (those on the actual object boundary) are among the retained boundaries, the expression is acceptable. In future work, we intend to explore the space of retained boundaries at each step in more detail. Even for a single test case, a lot of data has to be looked at and the proper methodology to do the analysis still has to be figured out. This could lead to an improvement of the expressions with even better final results for the proposed method.

Given the need to have a strictly over-segmented input map of primitives and the complexity of the shapes of interest, a relatively large number of primitives are needed to cover the complete object boundaries. The number of incremental grouping steps used in the method before merging may be seen as a parameter of the algorithm. By looking at typical object and image sizes, the shape complexity of objects of interest, and the description scale in the CCP map, a value of 15 was selected to obtain a complete boundary after a limited number of merging steps (one or two in the current implementation). A smaller value would make it difficult to efficiently prune out bad solutions. A larger value would increase the processing time with no improvement in the quality of the results.

With each boolean expression, there are parameters (thresholds) setting acceptable limits for each chosen criterion, actually transforming the criterion into a constraint. Most of the parameters in the proposed method are from those boolean expressions. Their values in the current implementation were adjusted using trial and error on the input maps of Figure 2. However, no systematic procedure was used and no systematic optimization was attempted. More precisely, only a small number of parameter values in the Boolean expressions were selected by experimenting with a small number of images (less than a third of the tested images). The values were also strongly constrained, that is selected from a small number of reasonable integer values.

In the proposed method, pruning parameters determine how many boundaries are kept after some of the processing steps. These were chosen on the basis of computational (time and memory) limits. Merging steps further include a maximum overlap size parameter for which a single reasonable value was chosen. Besides, all criteria are used with relaxed constraint values. We did not experiment with different values. In cases were no complete boundary is found, the constraints could be further relaxed. Most importantly, weighting parameters are needed to combine the global grouping criteria into a single final score. Again, no experimentation was done to optimize those parameters. The present weights are most probably not the best possible ones for the reported test cases.

Finally, the three preprocessing stages needed to get contour primitives are based on simple algorithms and their number of parameters may hardly be reduced. The quality of the input CCP map will vary according to the quality of the image (noise, contrast, blur, etc.). The most important aspect is the strict over-segmentation of the input CCP maps which was confirmed by observation. The precision of the map is not that important. That is, a slight variation in position, orientation, or length of the primitives should affect similarly all possible boundaries and it should not affect their pruning and ranking. A more thorough investigation of these aspects is also planned as future work.

8   Conclusion

In this paper, a hypothesis was made that global, non-monotonic grouping criteria are needed in order to detect and localize unexpected but interesting multi-part objects directly from a noisy and cluttered primitive map. A single generic set of shape-based grouping criteria were used in the proposed detection method. The main contextual knowledge relied on in defining the grouping criteria was that objects of interest are of intermediate complexity. Objects of intermediate shape complexity were called multi-part objects.

The main contributions of the proposed method are: i) the identification and formalization of a proper set of global shape criteria and weights that qualify and quantify the interest of a complete object shape, ii) the identification and formalization of a proper set of basic shape criteria and weights that qualify and quantify the potential interest of a partial object shape, and iii) the empirical identification of adaptive limiting values (constraints) on a subset of the basic shape criteria that enable an early rejection of unpromising shapes. Significant qualitative and quantitative results were obtained under realistic practical constraints for a problem with very high fundamental complexity. This represents the main accomplishment of the proposed method.

Interesting objects of different shapes and appearances were detected and localized, limiting misses and false alarms, in the generic context where a large number of variations in object appearance are possible. These variations may be of both intrinsic (shape, size, aspect ratio, reflectance, texture, markings) and extrinsic (viewpoint, partial occlusion, illumination, light dispersion, scene embedding, projection, lens distortion, sensor noise) origins. The actual improvement made by the proposed method to state-of-the-art detection methods is the greater invariance to changing shape, texture, markings, and viewpoint. Compared to state-of-the-art segmentation methods, high-level attention capture by interesting objects of properly complex shape is now attainable. The proposed approach consists in preprocessing input images to obtain a strictly over-segmented input CCP map which is then processed by grouping CCP primitives to obtain a single, complete object boundary (no over-segmentation) that does not overlap with the internal texture and markings or the external background (no under-segmentation). The obtained results confirm that, compared to state-of-the-art boundary extraction methods, best retained boundaries in the proposed method are more complete and they overlap less with the background and the internal texture and markings.

The proposed method represents a first attempt at using the SAFE methodology. As a result, the design choices made could certainly be optimized in various ways. In the present implementation, no attempt was made at systematically optimizing the set of criteria and its formalization. Despite this fact, it is remarkable that the best scoring boundaries produced, including FGTa, were actually quite similar to SGT. The main contribution of the proposed method, with respect to those criteria, is the variable subset used at each level and the adaptive boolean combination and limiting values for the corresponding constraints. As explained, these designer choices were made empirically but not systematically.

In addition to those discussed in Section 7, various improvements could be made to the proposed method. For instance, an experimental study of the significance of the various criteria could result in a different combination function or even the removal of the less useful ones. A simpler CCP map with a smaller number of distractors could also be beneficial, in terms of both effectiveness and efficiency. For instance, a value could be attributed to each CCP in the input map and only the best ones selected for multi-part object detection. A method has been developed to selectively enhance CCPs on the basis of pairwise grouping criteria and relaxation [34]. Another possibility is to enhance CCPs on the basis of their compatibility with region maps [35]. Finally, the present merging algorithm is very simple and it could easily be improved to produce additional boundaries when the best retained ones have poor final scores.

Given the a priori definition of an SGT, as required by the SAFE methodology,

it is possible to objectively determine the degree of failure of the method for any tested image (with respect to its SGT). To the best of our knowledge, this is far from being common practice in the computer vision literature, especially in the context of object boundary extraction in the figure-ground segmentation context. In the present implementation, the method sometimes give rise to a particular case of failure where no complete boundary is found despite a number of good partial boundaries in the final levels of the tree. The reason is grouping constraints may be too successful at pruning contours. It would be possible to adaptively relax constraints in the final levels to ensure a fixed number or percentage of partial contours after pruning. This condition would not apply to the first levels of the tree if the algorithm were to avoid producing complete boundaries for images containing no multi-part object.

Running times are indicated in Table 2. On a Pentium IV 2.0 GHz with 1 GB of RAM, they presently range from 25 seconds for a simple image like the water can to 6.5 minutes for a complex image like the toy. The plane required about 4 minutes and the stool a little more. Precomputation of table entries for geometrical relationships between CCP pairs could improve the efficiency of the method. Similarly, earlier rejection of candidate pairs for merging would provide important benefits since a large part of the running time is for boundary merging. Finally, a parallel computation of scores for competing partial boundaries would further improve running times. Memory limits exist with respect to the number of CCPs in the input map. Numbers shown are in the range of acceptable size in the present implementation.

To conclude, it was said earlier that the rationale behind the SAFE methodology is to make it easier to document and compare competing methods at the high level of main design choices. Given that the performance of any computer vision algorithm intimately depends on the input data, another way to compare methods is to use standard image datasets. A number of those were recently proposed for object detection and categorization in the literature, for instance the Weizmann Horse Database, Caltech 101, and the PASCAL Visual Object Classes. As discussed in Section 2, successful detection methods use these datasets in a supervised learning framework where a model in a selected representation is obtained from a statistical training procedure. Such an approach is not appropriate here given the larger number of variations in object appearance that are possible compared to each single category in standard datasets. The detection problem addressed in this paper is more generic, being related to the high-level attention capture problem. Recent segmentation methods proposed in the literature also adopt the statistical learning framework. None are appropriate for our problem since they either limit themselves to local, low-level criteria [36], or impose severe constraints on the object pose or appearance [37–39]. In the literature, there are no learning-based methods for detection or segmentation that target an object category as generic as the multi-part objects. No published detection or segmentation method goes as

far as localizing precisely the boundary of various multi-part objects whose specific shapes and appearances are initially unknown. State-of-the-art salient boundary extraction methods do not yet extract nor rank boundaries of an appropriately complex shape. Hence, the method proposed in this paper is unique with respect to its goal, its approach, and its obtained results.

9   Acknowledgments

References

[1] G.A. Miller, The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, Psychological Review, 1956, Vol. 63, No. 2, pp. 81–97.

[2] J.S. Beis, D.G. Lowe, Learning indexing functions for 3-D model-based object recognition, in: Proc. of AAAI Fall Symposium: Machine Learning in Computer Vision, 1993, pp. 275–280.

[3] G. Dorko, C. Schmid, Selection of Scale-Invariant Parts for Object Recognition, in: Proc. of the 9th International Conference on Computer Vision (ICCV'03), 2003, pp. 634–639.

[4] F. Jurie, C. Schmid, Scale-Invariant Shape Features for Recognition of Object Categories, in: Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR04), Washington, DC, 2004, pp. 90–96.

[5] A. Kushal, C. Schmid, and J. Ponce, Flexible object models for category-level 3d object recognition, in: Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR07), Minneapolis, Minnesota, 2007, pp. 1–8.

[6] S. Savarese, L. Fei-Fei, 3D generic object categorization, localization and pose estimation, in: Proc. of the 11th International Conference on Computer Vision (ICCV'07), Rio de Janeiro, Brazil, 2007, pp. 1–8.

[7] D. Walther, U. Rutishauser, C. Koch, P. Perona, Selective Visual Attention Enables Learning and Recognition of Multiple Objects in Cluttered Scenes, Computer Vision and Image Understanding, Vol. 100, No. 1–2, 2005, pp. 41–63.

[8] K. Koffka, Principles of gestalt psychology, Harcourt, Brace and World, New York (1935).

[9] J.H. Elder, S.W. Zucker, Evidence for Boundary-Specific Grouping, Vision Research, Vol. 38, No. 1, 1998, pp. 143–152.

[10] J. Feldman, The role of objects in perceptual grouping, Acta Psychologica, Vol. 102, No. 2–3, 1999, pp. 137–163.

[11] S. Mahamud, L.R. Williams, K.K. Thornber, K. Xu, Segmentation of Multiple Salient Closed Contours from Real Images, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 25, No. 4, 2003, pp. 433–444.

[12] J.H. Elder, S.W. Zucker, Computing Contour Closure, in: Proc. of the 4th European Conference on Computer Vision (ECCV96), 1996, pp. 399–412.

[13] F.J. Estrada, A.D. Jepson, Perceptual Grouping for Contour Extraction, in: Proc. of the 17th International Conference on Pattern Recognition (ICPR04), 2004, pp. 32–35.

[14] D.W. Jacobs, Robust and Efficient Detection of Salient Convex Groups, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 18, No. 1, 1996, pp. 23–37.

[15] S. Wang, J. Wang, T. Kubota, From Fragments to Salient Closed Boundaries: An In-Depth Study, in: Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR04), Washington, DC, 2004, pp. 291–298.

[16] T.D. Alter, R. Basri, Extracting Salient Curves from Images: An Analysis of the Saliency Network, in: Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR96), San Francisco, CA, 1996, pp. 13–20.

[17] I. Biederman, Recognition-by-Components: A Theory of Human Image Understanding, Psychological Review, 1987, Vol. 94, No. 2, pp. 115–147.

[18] D.D. Hoffman, W.A. Richards, Representing smooth plane curves for recognition: Implications for figure-ground reversal, in: Proc. of the National Conference of the American Association for Artificial Intelligence, Pittsburg, PA, 1982, pp. 5–8.

[19] R. Bergevin, M.D. Levine, Generic object recognition: building and matching coarse descriptions from line drawings, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 15, No. 1, 1993, pp. 19–36.

[20] A. Jacot-Descombes, T. Pun, Asynchronous Perceptual Grouping: From Contours to Relevant 2-D Structures, Computer Vision and Image Understanding, Vo. 66, No. 1, 1997, pp. 1–24.

[21] M. Zerroug, R. Nevatia, Part-based 3D descriptions of complex objects from a single image, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 21, No. 9, 1999, pp. 835–848.

[22] G.-A. Bilodeau, R. Bergevin, Part segmentation of objects in real images, Pattern Recognition, Vol. 35, No. 12, 2002, pp. 2913–2926.

[23] G.-A. Bilodeau, R. Bergevin, Qualitative part-based models in content-based image retrieval, Machine Vision and Applications, Springer, Vol. 18, No. 5, 2007, pp. 275–287.

[24] P.N. Klein, T.B. Sebastian, B.B. Kimia, Shape matching using edit-distance: an implementation, in: Proc. of the Twelfth Annual Symposium on Discrete Algorithms, Washington, DC, 2001, 781–790.

[25] W. A. Richards, D.D. Hoffman, Codon constraints on closed 2D shapes, Computer Vision, Graphics, and Image Processing, Vol. 31, No. 1, 1985, pp. 265–281.

[26] J.P. Princen, J. Illingworth, J.V. Kittler, A Hierarchical Approach to Line Extraction Based on the Hough Transform, Computer Vision, Graphics, and Image Processing, Vol. 52, No. 1, October 1990, pp. 57-77.

[27] H.K. Yuen, J.P. Princen, J. Illingworth, J.V. Kittler, Comparative Study of Hough Transform Methods for Circle Finding, Image and Vision Computing, Vol. 8, No. 1, February 1990, pp. 71-77.

[28] M. Mokhtari, R. Bergevin, Generic Multi-scale Segmentation and Curve Approximation Method, in: Proc. of the Third International Conference on Scale-Space and Morphology in Computer Vision, 2001, pp. 227–235.

[29] V. Randrianarisoa, J.-F. Bernier, R. Bergevin, Detection of Multi-part Objects by Top-down Perceptual Grouping, in: Proc. of the 2nd Canadian Conference on Computer and Robot Vision, 2005, pp. 536–543.

[30] E.A. Engbers, A.W.M. Smeulders, Design considerations for generic grouping in vision, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 25, No. 4, 2003, pp. 445–457.

[31] A. Amir, M. Lindenbaum, A Generic Grouping Algorithm and its Quantitative Analysis IEEE Trans. Pattern Anal. Mach. Intell., Vol. 20, No. 2, 1998, pp. 168–185.

[32] E. Saund, Finding Perceptually Closed Paths in Sketches and Drawings, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 25, No. 4, 2003, pp. 475–491.

[33] J. S. Stahl, S. Wang, Edge Grouping Combining Boundary and Region Information, IEEE Transactions on Image Processing, Vol. 16, No.10, 2007, pp. 2590–2606.

[34] R. Bergevin, A. Filiatrault, Enhancing Contour Primitives by Pairwise Grouping and Relaxation, in: Proc. of the 4th International Conference on Image Analysis and Recognition (ICIAR 2007), Montreal, Canada, 2007, pp. 222–233.

[35] R. Bergevin, V. Bergeron, Enhancing Boundary Primitives using a Multiscale Quadtree Segmentation, in: Proc. of the 4th International Symposium on Visual Computing (ISVC 2008), Las Vegas, Nevada, 2008, pp. 45–54.

[36] D.R. Martin, C.C. Fowlkes, Jitendra Malik, Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues, IEEE Trans. Pattern Anal. Mach. Intell., Vo. 26, No. 5, 2004, pp. 530–549.

[37] T. Cour, J. Shi, Recognizing objects by piecing together the Segmentation Puzzle, in: Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR07), Minneapolis, Minnesota, 2007, pp. 1–8.

[38] S.X. Yu, J. Shi, Object-Specific Figure-Ground Segregation, in: Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR03), Madison, Wisconsin, 2003, Vol. 2, pp. 39–45.

[39] H. Arora, N. Loeff, D.A. Forsyth, N. Ahuja, Unsupervised Segmentation of Objects using Efficient Learning, in: Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR07), Minneapolis, Minnesota, 2007, pp. 1–7.

10 About the Authors

R. BERGEVIN received the Ph.D. degree in Electrical Engineering from McGill University in 1990. He joined the Computer Vision and Systems Laboratory at Laval University in 1991. His research interests are in image analysis and cognitive vision. Dr. Bergevin is a member of the Province of Quebec's Association of Professional Engineers (OIQ) and the IEEE Computer Society. He serves as Associate Editor for the Pattern Recognition journal and Area Editor for the Computer Vision and Image Understanding journal.

J.-F. BERNIER received the M.Sc. degree in Electrical Engineering from Laval University in 2006. He previously received the B.Eng. degree in Computer Engineering from Laval University in 2004. He is currently working as a R&D engineer in a startup company.