Enhancing Contour Primitives by Pairwise Grouping and Relaxation

Robert Bergevin and Alexandre Filiatrault

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

Abstract. A method is proposed to enhance contour primitives of multi-part objects in complex images. It consists in first extracting circular arcs and straight-line segment primitives from the image edge map. The produced binary constant-curvature primitive map includes primitives of interest on the object contour or silhouette, as well as two types of distractors: internal texture segments and external background segments. Each obtained primitive is enhanced using an exhaustive evaluation of a number of pairwise grouping criteria. Finally, an iterative relaxation procedure adjusts the weight of each primitive according to the weights of its best-matching primitives. A subjective ground-truth binary map may be used to assess the degree to which the final weighted map corresponds to a selective enhancement of contour primitives.

1 Introduction

Delimiting the region occupied by an interesting object in a static image is both useful and easy for humans. In computer vision, this is still a fundamental problem with no existing general solution. This is particularly notable with complex natural images where objects of interest appear under variations of shape, illumination, surface texture, viewpoint, and background. Recently, a new generic object detection method was proposed [1]. Its main assumption is that objects of properly complex shape are of more interest and preferably detected. On that basis, a simple set of explicit local and global contour grouping criteria are defined and used to control the expansion of a deterministic search algorithm. Starting from a constant-curvature contour primitive (CCP) map, a number of potential object silhouettes are systematically generated and sorted.

Considering an average map of 400 CCPs and silhouettes of around 30 ordered CCPs, the number of possible silhouettes is huge, that is about 10^{86}. Efficient shape criteria-based pruning reduces the number of computed silhouettes, for images of varying complexity, to less than a thousand. Still, difficult images may take up to many minutes to process. On the other hand, scoring based on global criteria is quite efficient as, in each case, the silhouette most similar to a manual reference ends-up in the first absolute or relative positions. Quantitative similarities obtained are from 85% to 100%. In fact, high-ranking silhouettes are

M. Kamel and A. Campilho (Eds.): ICIAR 2007, LNCS 4633, pp. 222–233, 2007.

(a) Manual (b) High-ranking (c) Low-ranking

Fig. 1. Manual and computed silhouettes

almost always qualitatively very similar to their manual reference, which is a very encouraging result (see Figure 1).

Despite the quality of obtained results, there is place for improvement in terms of reducing both the number of computed silhouettes and the time required to generate and sort them. Typically, more than 90% of the CCPs of an input map are distractors, either internal texture primitives or external background primitives (see Figure 2(b)). The goal of the method proposed in this paper is to assign a weight to each CCP in the input map that should reflect its potential to end-up as a member of a high-ranking silhouette. If properly done, such a process shall enhance CCPs from the silhouette with respect to distractor CCPs of the original binary map.

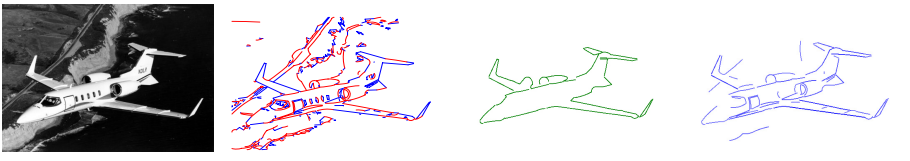
The proposed method falls into what is known as the segmentation problem. The latter consists in identifying which portions of an image contains important information and which are only distractive to the ultimate processing goal. A very large number of studies have been conducted in the past on the segmentation problem. Typically, such studies fall into one of two broad categories. Firstly, generic low-level segmentation methods group and select image data according to basic image parameters, irrespective of high-level knowledge as to what constitutes an object of interest [2,3,4]. While much progress has been made over the years with that approach, the quality of results still strongly depend on the imaging conditions. More precisely, low-level methods applied to complex images tend to produce results that suffer from both under-segmentation and over-segmentation. This makes it quite difficult to usefully exploit their results in subsequent processing stages, even to complete the segmentation process as defined above. In contrast, specific high-level segmentation methods are more likely to extract significant segments of the image given their specialization to known objects of interest [5,6]. While progress has been made to improve high-level methods, they are still too specific to be of general use.

It has proven quite difficult to come up with a generic high-level segmentation method. The main contribution of this paper is a new pre-preprocessing stage whose goal is to improve the efficiency of a promising recently-proposed generic high-level segmentation method [1].

2 Problem

The input to the proposed method is a constant-curvature contour primitive (CCP) map (see Figure 2(b)) computed from a static intensity image (see Figure 2(a)). The CCP map is computed using Magno, an in-house segmentation

algorithm [7]. The number of CCPs in maps produced by Magno may be from about one hundred primitives to more than one thousand primitives. Figure 2(c) is a binary map of CCPs interactively selected by a human from the input map. It is referred to as SGT, for subjective ground truth. It represents what the person believes is the best subset of CCPs that would produce, if properly ordered, the silhouette of a multi-part object. In this paper, CCPs in the manual references may differ from the ones in [1] since they are selected directly from the input CCP map, without a view of the original image.



(a) Image (b) Input map (c) SGT (d) Output map

Fig. 2. Basic concept. Contour primitives should be selectively enhanced.

SGT is not known by the proposed method. It will only serve to assess the quality of computed solutions. In contrast to [1], computed solutions are in a format different than the manual reference. That is, solutions are weighted CCP maps whereas the reference is a binary CCP map. In that way, the number of possible solutions is actually infinite. On the other hand, CCPs from a weighted map may be sorted according to their weight. The number of possible orders for the input CCPs is finite. Similarly, an infinite number of threshold values could be used to reduce the weigthed map to a binary map (see Figure 2(d)). However, the number of subsets of CCPs from the input binary map is finite.

Let us assume that an algorithm provides a scoring function for computed solutions, either a weighted CCP map, a complete sorted list of CCPs, or simply a subset of input CCPs. FGT, or formal ground truth, is the possible solution with highest score. FGT is usually not known either by the proposed method as it would require to generate all of an infinite or huge finite number of solutions and score each of them. More practically, a subset of the possible solutions is computed and the one with the highest score is selected. The selected optimal solution is only an approximation of FGT. It is referred to as FGTa. In this paper, a single weighted CCP map is computed for each input binary map.

FGTa may be the same as FGT, but this can seldom be verified in images of typical complexity. The goal of the proposed method is to generate, in an efficient manner, an FGTa as similar as possible to SGT, for a variety of complex images of multi-part objects. In this paper, similarity measures proposed apply to thresholded weighted maps, where the threshold is selected according to the number of sorted CCPs kept in the final subset.

3 Proposed Method

There are two main stages in the proposed method. Firstly, a number of pairwise grouping criteria are evaluated for each pair of primitives in order to assign a strength to that pair. For each primitive, the values of its pairs are converted to a single weight. Secondly, the weights of all primitives are adjusted by an iterative relaxation stage according to the weights of their best-matching primitives. Typically, a number of very small primitives are present in the initial CCP map. In order to limit the processing time, the shortest primitives may not be considered. Figure 3 displays resulting maps with 88 random CCPs, the 88 longest CCPs, and the best 88 CCPs according to the pairwise grouping criteria. One may compare them with Figure 2(c). Clearly, length is an important factor but pairwise criteria provide a significant improvement in terms of approaching SGT. For all results presented in this paper, the number of primitives considered is at most four hundred.

(a) Random (b) Length (c) Pairwise

Fig. 3. Selection criteria. Random, length, and pairwise grouping.

3.1 Pairwise Grouping Criteria

A total of six pairwise grouping criteria are used in the proposed method. They were developed on the basis of Gestalt grouping laws [8]. There are two near-primitives grouping criteria: position and angle continuity. Their goal is to measure the degree to which a pair of primitives could be connected on the silhouette of the object. They apply to pairs of primitives distant by at most 10 pixels. Besides, there are two far-primitives grouping criteria: regional coherence and matching. Here, the goal is to measure the degree to which a pair of primitives could symmetrically delimit a part of the object. They apply to pairs of primitives distant by at least 35 pixels. Finally, there are two all-primitives grouping criteria: local intensity and contrast. These last criteria aim at measuring the degree to which a pair of primitives properly delimit a single object or part. They apply to all pairs of primitives. The formalization of the six criteria is described below. Each one is normalized such that values obtained by the pairs are between 0 (criterion is not satisfied at all) and 1 (criterion is perfectly satisfied).

The first four criteria are only applied if the distance between the two primitives of a given pair is in the proper range. Distance calculation is performed according to the types of the two primitives: straight-line segments, circular arcs, or mixed. Computations are quite efficient as they involve a limited number of analytical parameters: positions of endings and center.

For near primitives, position continuity is computed as follows:

$$PC(i,j) = e^{-\frac{x}{K}} \tag{1}$$

where x is the closest distance between a pair of endings from the two primitives i and j. An experimental value $K = 10$ was selected. Similarly, angle continuity is computed as follows:

$$AC(i,j) = 1 - e^{-\frac{|\tan \Delta\angle|}{10}} \tag{2}$$

where $\Delta\angle$ is half the angle difference in degrees betweeen the tangents at the closest endings. It is computed such that close-by parallel primitives obtain a very low value.

Once PC and AC are computed, they are merged into a near-primitives value (NPV). The merging function is defined in such a way that a single good value out of the two could lead to a good value for the pair:

$$NPV = PC^{K_1} \times AC^{K_2} \tag{3}$$

where the constants $K_1 = \frac{1}{2}$ and $K_2 = \frac{1}{5}$ were determined experimentally.

Figure 4 confirms the complementary nature of the above two criteria. The shown maps are thresholded to keep only the best 88 primitives.

(a) PC (b) AC (c) NPV

Fig. 4. Near-primitives criteria. Each criterion is combined with length.

Regional coherence, the first of two far-primitives criteria, must take into account that a multi-part object may have both convex and concave parts. Besides, size and aspect-ratio of parts should be balanced with respect to image size. An integration of those different aspects resulted into the following formula:

$$RC(i,j) = max(TanOver(i,j), TanOver(i,mj), TanOver(mi,mj)) \times f(x) \tag{4}$$

where

$$f(x) = LogSig((x - k_1)/\delta_1) - LogSig((x - k_2)/\delta_2) \tag{5}$$

and

$x = RadOver(i, j)$ measures the radial overlap of the two primitives.
k_1, k_2 are the limits of an interval for which f(x) is above 0.5. They determine how the distance between the two primitives affect RC. In this paper, $k_1 = 0.2$ and $k_2 = 1.0$.
δ_1, δ_2 control the speed at which RC declines when the primitives are either too close or too far. In this paper, $\delta_1 = 0.06$ and $\delta_2 = 0.1$.
$TanOver(i, j)$ measures the tangential overlap of the two primitives.
mi, mj are mirror primitives of i and j, respectively.
$LogSig()$ is a standard function LogSig(x) $= 1/(1+e^x)$

In order to compute RC for the three types of pairs, each straight-line segment is temporarily transformed into a circular arc of small curvature. Originally, the RC criterion was developed for a multi-scale CPP map. Hence, each straight-line segment could be transformed into a circular arc according to its corresponding scale parameter. In this paper, a single small extraction scale is assumed for all primitives. For this reason, the performance of RC is not as good as expected with a pair of straight-line segments. This explains the creation of the complementary regional matching (RM) criterion.

RM compares the shape delimited by the two primitives joigned by their corresponding extremities to a square shape. In practice, it first averages the projection of each primitive onto the other as follows:

$$AP(i, j) = (\sqrt{Projection(i, j)} + \sqrt{Projection(j, i)})/2 \tag{6}$$

RM is then obtained by scaling that value to take into account the relative size of the primitives, the position of their endings, as well as the ratio between their lengths and distance:

$$RM(i, j) = AP(i, j)/max((P_1, \frac{1}{P_1}) \tag{7}$$

where

$$P_1 = min(length(i), length(j))/max(distance(i, j), 10) \tag{8}$$

Again, once RC and RM are computed, they are merged into a far-primitives value (FPV). The merging function is now:

$$FPV = K_1 \times RC + K_2 \times RM \tag{9}$$

where the constants $K_1 = 2$ and $K_2 = 5$ were determined experimentally.

Figure 5 shows that RC, though a powerful pairwise criterion, is not globally sufficient in itself.

The all-primitives criteria do not actually limit themselves to the binary CCP map. More precisely, the local image intensities on each side of a primitive are combined into two values used in the final two criteria: a measure of average contrast for the primitive and a measure of average intensity on each side. The intensities are sampled at a distance of four pixels from the primitive, to take into account the uncertainty in its position and shape.

(a) RC (b) RM

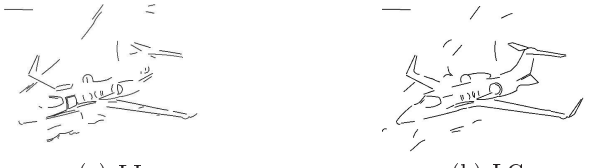
Fig. 5. Far-primitives criteria. Each criterion is combined with length.

The local intensity (LI) criterion finds the best matching average intensities among the two sides of each primitive of the pair. The two remaining sides provide a second match. The two matches are normalized and then combined using weights selected experimentally:

$$LI(i,j) = 0.6 * BestMatch(i,j) + 0.4 * OtherMatch(i,j) \tag{10}$$

The local contrast (LC) criterion is the minimum average contrast of the two primitives. Once LI and LC are computed, they are merged into a all-primitives value (APV). The merging function is like the one for FPV.

Figure 6 shows that LC is quite powerful when the object contrast is strong, as is the case here.



(a) LI (b) LC

Fig. 6. All-primitives criteria. Each criterion is combined with length.

At this point, one or two values among NPV, FPV and APV are obtained for each pair according to the distance between the two primitives. The pair score is the sum of the obtained values weighted by the square-root of the length of each primitive. The last step of this first stage consists in converting the scores of pairs into a weight for each primitive in the map. This is simply obtained by summing the square of the computed score for all pairs containing the given primitive. The resulting weight is normalized using the highest weight among the primitives. Figure 3 shows the best 88 CCPs according to the resulting weights of all CCPs in the map.

3.2 Relaxation Procedure

At each iteration, the weight of a primitive is adjusted according to the weights of its best-matching primitives, one in the near-distance range and one in the far-distance range. This way, a large number of average primitives can not improve

the weight of another primitive. Besides, adjustments are performed in parallel. Hence, no bias is introduced as a result of using a specific adjustment order. The adjustment function is as follows:

$$CCP_{i,j} = K_1 \times CCP_{i,j-1} + (1 - K_1) \times (CCPm + CCPn) \tag{11}$$

where i is the adjusted CCP weight, j is the iteration number, $CCPm$ is the best-matching primitive using near-primitives criteria, and $CCPn$ is the best-matching primitive using far-primitives criteria.

A key question is how much relative influence supporting primitives have at each iteration. This is controlled by parameter K_1. The value $K_1 = 0.9$ was determined experimentally.

As mentioned earlier, the goal of the proposed method is to generate, in an efficient manner, an FGTa as similar as possible to SGT, for a variety of complex images of multi-part objects. In this paper, similarity measures proposed apply to thresholded weighted maps, where the threshold is selected according to the number of sorted CCPs kept in the final subset. Precision and recall are two standard measures in the literature. Precision is high when the number of false alarms is low. That is, the primitives kept mostly belong to SGT. Formally, precision is the number of SGT primitives in the thresholded map (FGTa) over the total number of primitives in that map. Recall is high when the number of false negatives is low. That is, most SGT primitives are kept. Formally, recall is the number of SGT primitives in the thresholded map over the total number of primitives in SGT.

Figure 7 displays thresholded 88-CCP maps obtained after the first, second, fifth, and tenth iterations. The first iteration is where the effect is more obvious. Compared to Figure 3, more than half erroneous background CCPs were removed. Besides, a number of contour CCPs were added. Recall and precision are improved from 61% to 70% and from 47% to 53%, respectively. Afterwards, some improvements are added at the expense of a few introduced errors.

(a) Iteration 1 (b) Iteration 2 (c) Iteration 5 (d) Iteration 10

Fig. 7. Relaxation progress. Results after 1, 2, 5, and 10 iterations.

A global evaluation score may also be defined as the area under the Recall versus iteration number curve. Typically, that global score increases slowly after the first few iterations to converge towards an asymptotic value at or before the tenth iteration.

4 Experimental Results

A number of experiments were conducted using a fixed set of parameter values. Results are presented for simple images (limited background and texture), complex images (normal texture and background) and difficult images (with respect to the pairwise grouping criteria). In all cases, the color image, the complete input binary CCP map, the SGT manual reference, and the thresholded output map are shown. Let us recall that the image intensity is used in two criteria.

4.1 Simple Images

Figure 8 displays a 178-CCP input map whose manual SGT reference has 43 CCPs. The shown thresholded output map has 33 CCPs. Precision and recall are 88% and 67%, respectively.

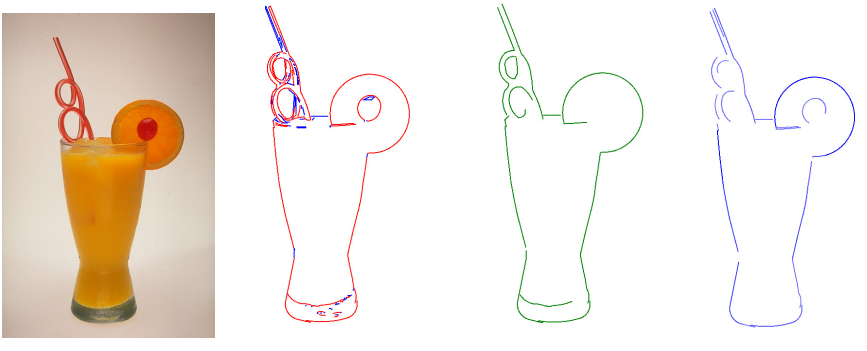


Fig. 8. Final result for a simple image

Similarly, Figure 9 displays a 188-CCP input map whose manual SGT reference has 32 CCPs. The shown thresholded output map has 29 CCPs. Precision and recall are 90% and 72%, respectively. Given the limitation of the method to six simple pairwise grouping criteria, the obtained results are very satisfying, both qualitatively and quantitatively. As can be expected, adding more CCPs in the output map improves the recall at the expense of reducing the precision. For instance, the addition of another 10 ordered CCPs results in a recall of 86% with a precision of 79%, which is still quite good.

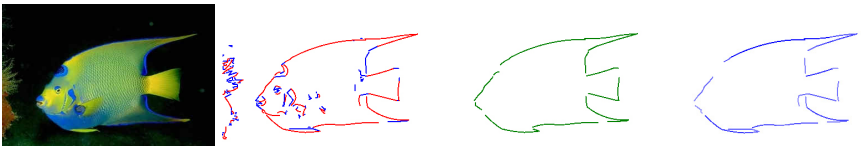


Fig. 9. Final result for another simple image

4.2 Complex Image

Results for a typical complex image were shown in Figure 2(d). The number of CCPs in the complete map is 552. Only the longest 400 were used as input by the method. The manual reference has 79 CCPs. The shown thresholded output map has 88 CCPs. This number was obtained using a fixed ratio. This ratio was found optimal with respect to the global evaluation score over a number of test images. As mentioned earlier, the actual output is the sorted list of weighted CCPS. For instance, although an important CCP is missing at the front of the plane in the thresholded map, it is actually only at position 90 in the sorted list. With 88 CCPs, the computed precision and recall are 55% and 72%, respectively.

4.3 Difficult Images

Figure 10 displays a detailed result on a difficult image. The number of CCPs in the complete map is 745, which is quite high considering that only 400 are used as input. Many short CCPs are either eliminated from the input or they provide noisy measures for the grouping criteria. Compared to a manual SGT reference of 48 CCPs, the thresholded output map has 78 CCPs. The resulting precision and recall are 37% and 60%, respectively.

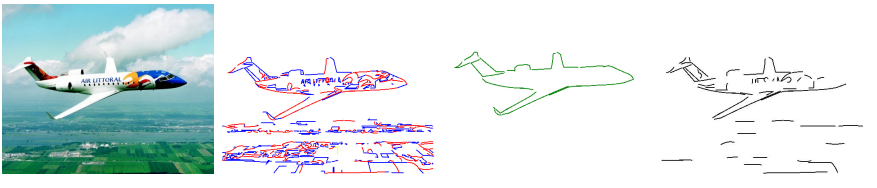


Fig. 10. Final result for a difficult image

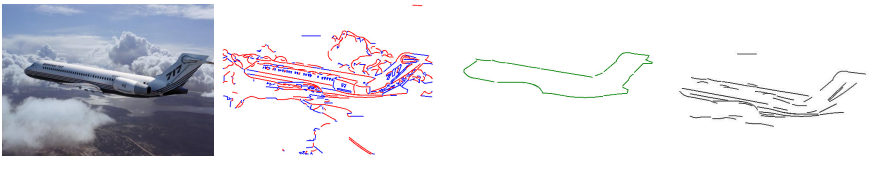


Fig. 11. Final result for another difficult image

Figure 11 presents another type of difficulty. That is, the contrast is limited between the object and the background. For this reason, many texture and background CCPs are present in the complete map. Besides, the internal texture has a structure similar to the object. Despite these difficulties most of the important contour CCPs are kept. For a thresholded map of 54 CCPs (compared to a manual SGT reference of 34 CCPs), the obtained precision and recall are 33% and 53%, respectively. Unfortunately, a number of parallel CCPs from the

internal texture could not be eliminated using only the pairwise criteria. Apart from limiting the potential efficiency gains of following processing stages, these spurious segments are not likely to affect the results obtained e.g. by a robust object detection method, such as the one presented in [1].

4.4 More Results

Figure 12 presents a number of additional input-output CCP map pairs.

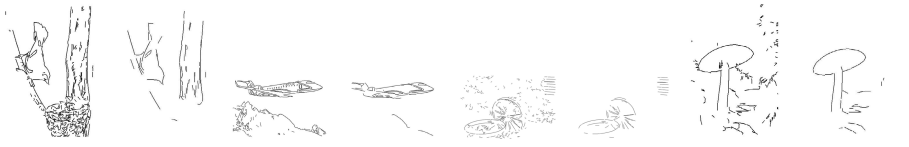


Fig. 12. Input-output pairs. Output maps are thresholded.

5 Conclusion

The proposed method was tested on a number of images of varying complexity and difficulty. It was shown that each developed grouping criterion is helpful but incomplete in itself. Besides, it was found that a small number of iterations of a simple relaxation procedure improves the quality of the obtained maps. Final thresholded maps were qualitatively and quantitatively compared to a subjective ground-truth binary map representing the object silhouette.

Many existing grouping techniques have goals and results that are not as precise as here. That is, regions of interest, sometimes with fuzzy borders, are more likely to be found than a well-defined and complete object silhouette. For instance, the "'large groups"' of Sarkar and Soundararajan [9] are typically under-segmented, mixing contour, texture, and background, even when the main object is centered and unoccluded. Given the challenging and fundamental goal faced by the proposed method, results appear to be quite encouraging.

The proposed method could be improved in a number of ways. A better formalization of the identified criteria might give rise to more efficient and effective computations. The relaxation stage could as well be modified in order to take into account a larger number of supportive primitives. More significant changes are also possible. For instance, criteria could be defined on triples of primitives instead of pairs. Finally, evaluation measures based on the complete weighted output map need to be developed. A weighted manual reference map showing the relative importance of contour primitives could be used for that matter.

Acknowledgment

This work is supported by a NSERC discovery grant.

References

1. Bernier, J.-F., Bergevin, R.: Generic Detection of Multi-Part Objects. In: ICPR06. Proc. of the 18th International Conference on Pattern Recognition, Hong Kong, China (2006)
2. Grigorescu, C., Petkov, N., Westenberg, M.A.: Contour and boundary detection improved by surround suppression of texture edges. Image and Vision Computing 22, 609–622 (2004)
3. Lau, H.F., Levine, M.D.: Finding a small number of regions in an image using low-level features. Pattern Recognition 35, 2323–2339 (2002)
4. Malik, J., Belongie, S., Leung, T., Shi, J.: Contour and Texture Analysis for Image Segmentation. Int. J. Comput. Vision 43, 7–27 (2001)
5. Yu, S.X., Shi, J.: Object-Specific Figure-Ground Segregation. In: CVPR2003. Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition, Madison, WI, pp. 39–45 (2003)
6. Borenstein, E., Ullman, S.: Class-Specific, Top-Down Segmentation. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 109–122. Springer, Heidelberg (2002)
7. Mokhtari, M., Bergevin, R.: Generic Multi-Scale Segmentation and Curve Approximation Method. In: Kerckhove, M. (ed.) Scale-Space 2001. LNCS, vol. 2106, pp. 227–235. Springer, Heidelberg (2001)
8. Desolneux, A., Moisan, L., Morel, J.-M.: Gestalt Theory and Computer Vision. In: Carsetti, A. (ed.) Seeing, Thinking and Knowing, pp. 71–101. Kluwer Academic Publishers, Dordrecht (2004)
9. Sarkar, S., Soundararajan, P.: Supervised Learning of Large Perceptual Organization: Graph Spectral Partitioning and Learning Automata. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 504–525 (2000)