Similarity-based and perception-based image segmentation — Source link ⬀

Jérémy Huart, Pascal Bertolino

Related papers:

- Pyramid segmentation algorithms revisited

- Unsupervised image segmentation using local homogeneity analysis

- Survey paper on image segmentation using k-means clustering

- Medical Image Segmentation Techniques, a Literature Review, and Some Novel Trends

- A Region Based Segmentation using Pixel Block Fuzzy Similarity

Similarity-based and perception-based image segmentation

Jérémy Huart, Pascal Bertolino

SIMILARITY-BASED AND PERCEPTION-BASED IMAGE SEGMENTATION

Jérémy Huart, Pascal Bertolino

Laboratoire des Images et des Signaux,
BP 46, 38402 Saint Martin d'Hères, France
Pascal.Bertolino@lis.inpg.fr

ABSTRACT

In this paper, we present a segmentation method that provides perceptually relevant partitions without any *a priori* knowledge of the image content: first a local homogeneity analysis detects the image areas that have to be segmented. Then segmentation using a similarity criterion is locally performed. At last, segmented regions are grouped using Gestalt criteria. The whole method is presented in a hierarchical framework.

1. INTRODUCTION

We tackle the problem of generic and automatic object extraction. There are three main segmentation approaches. Firstly, the edge-based methods which mainly contain the active contour algorithms (*e.g.* snake, balloon [1]). Secondly, the region-based methods such as region-growing or split-n-merge algorithms. This category gathers the probabilistic approaches which minimize a global energy function calculated on regions [2]. [1] proposes an original method which mixes energy function and mathematical morphology. And finally, the hybrid methods which merge the two kinds of information, such as the region competition [3]. [4, 5] propose to integrate the previous methods with the scale-space in order to add multiresolution information well adapted to the human visual system.

In order to propose a generic method that is based neither on *a priori* knowledge of the image content nor on any object model, three powerful methods are combined: (1) A **local homogeneity analysis** of the image is performed to initialize a local segmentation in order to prevent any over-segmentation. (2) The irregular graph pyramid realizes this **local segmentation** within the heterogeneous areas of the image. It is a flexible parallel region growing segmentation technique. Using a classical similarity criterion and label propagation, it provides a stack of accurate partitions. (3) The pyramid is used again on the segmentation result as a **perceptual grouping** process according to criteria extracted from the Gestalt theory. These criteria are well adapted to model-free methods because it only takes into account the visual relevance of regions.

2. THE LOCALIZED PYRAMID

The irregular pyramid [6] is a powerful tool that provides hierarchical segmentations with a single process.

In this method an adjacency graph is initialized, where every vertex corresponds to a one pixel region. Using a local (i.e. region independent) algorithm performed on the whole image, similar neighboring regions can merge, yielding a decreasing number of vertices. Regions i and j are similar if for example their average YUV color distance is lower than a threshold: $d(YUV(R_i), YUV(R_j)) < T$. The algorithm iteratively produces a stack of successive graphs (called pyramid) until convergence (no more possible fusions).

Usually, the graph pyramid is initialized with as many vertices as the number of pixels in the image, in order to perform the segmentation of the whole image (figure 1). In a localized pyramid, only a subset of the image pixels are segmented (they are initialized as undefined zones), while the rest of them is arbitrary associated to one (or a few number of) vertex (figure 2) called roots (regions that will belong to the final partition) as the background for instance. During the segmentation process, the pixels of the undefined zones merge together and form regions that stick to one of the roots, thus refining their shape. Localized segmentation is interesting since it provides faster processing times (only a part of the image is processed) and because the risk of segmentation error is spatially well limited.
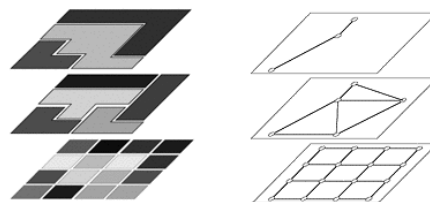

Fig. 1. Example of a graph pyramid built on a 4×4 pixels image: stack of partitions and stack of graphs

Fig. 2. Example of a localized pyramid initialization

3. SIMILARITY BASED SEGMENTATION

To initialize automatically the local pyramid we use a method presented in [7] that performs a local homogeneity analysis on color images. Contrary to their method that computes the homogeneity image (or "H-image") from the RGB components, we use the CIE L*a*b* color space since it provides less false negative discontinuities (see figures 3.b and 4.b). The H-image is a gray-scale image whose pixel values are the H-values representing the image discontinuities according to an homogeneous feature. Three H-images are computed, namely H_L, H_a, H_b, and the final H-image is obtained using the following formula:

$$H = \sqrt{(H_L^2 + H_a^2 + H_b^2)}.$$

The high values of the H-image correspond to possible discontinuities and the low values correspond to homogeneous regions. Then a classification of the H-values provides a binary partition with homogeneous and non homogeneous connected components. The authors of [7] perform a region growing from homogeneous zones used as seeds. On our side, this binary mask initializes the roots and the undefined areas needed for the local segmentation (figure 3.b).

The white connected components are the roots. The pixels of the black components will have to merge together and/or with a neighboring root according to the similarity criterion [8].

Compared to the result provided by the classical pyramid (fig. 3.c), the result obtained with the local segmentation and the homogeneity mask (fig. 3.d) is much suitable: there are less regions but with the same precision. Notice that in both cases, the same threshold was used.

4. PERCEPTION BASED GROUPING

4.1. Pyramidal region grouping

After the local similarity segmentation process, grouping regions aims at simplifying the content of the obtained partition such as figure 4 c. Two constraints must be respected for an efficient grouping process: first, only the best groupings must locally be retained. That means all combinations of groupings must be tested (among two, three, four, ... neighbors). Secondly all the groupings must be spread on the image so that no part of the image is advantaged.

For managing the grouping process, the irregular pyramid is used as well for three main reasons: first its graph struc-
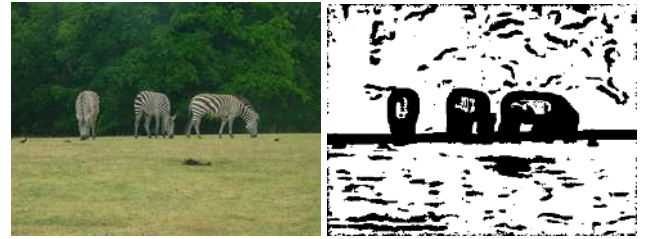
(a) Original image

(b) RGB homogeneity mask

(c) Segmentation of the whole image with a classical irregular pyramid

(d) Segmentation of the heterogenous areas with a localized pyramid

Fig. 3. regular *vs.* localized segmentation

ture is well adapted to the parallel (i.e. independent) manipulation of regions. Secondly the criteria for region merging are easily interchangeable. At last iterating a grouping process is simply done by building new levels. The initial graph of this pyramid is the last graph of the similarity pyramid. Indeed, the local pyramid is extended with additional levels induced by the region grouping.

The visual relevance measures concern either a single region or a grouping of several neighboring regions. In their paper, the authors of [2] group only pair-wise regions. Contrary to their work, in our method, any number of regions can simultaneously merge in one grouping. It supplies more choice in the grouping strategy and more adaptivity to the image content.

4.2. Grouping features

The chosen features are derived from the Gestalt theory [9] which does not use any object model. The human vision performs domain-independent grouping (called Gestalts) mainly based on five properties: proximity, similarity, closure, continuity and symmetry [10].

From these properties, energies can be calculated for any region or group of regions, the aim being to find the lowest energy groupings representing a high perceptual relevance. The cost of a grouping is composed by several energy functions proposed by [2].

E_{fusion} can be seen as the cost of the fusion operation. It is based on the mean difference of Lab components and on

the junctions continuity.

E_{region} is the energy of a region resulting from the fusion. It can be seen as the opportunity of the potential grouping (the better the opportunity, the lower the energy). It is based on its compactness, convexity and area.

A low energy value indicates a strong visual interest. On the contrary a high value indicates an undesirable region or grouping. The energy function of a region resulting from a grouping is given by $E = E_{fusion} + E_{region}$. The aim is to realize the grouping which ensures locally the lowest energy.

4.3. Grouping selection

4.3.1. Selection of the best local groupings

Let v_c be a vertex, $c \in [\![1, N]\!]$ and n_c the number of its neighbors. All the groupings including v_c and its neighboring combinations are considered. The number of combinations is given by:

$$C = \sum_{j=1}^{n_c} C_{n_c}^j$$

$C_{n_c}^j$ being the number of combinations of j neighbors among n_c. For each of these groupings, E_{fusion} and E_{region} are computed.

Let g_c being the grouping including v_c with the lowest energy $E(g_c)$. g_c is a potential grouping if: (1) g_c locally improves the partition energy, (2) $E(g_c)$ indicates a high relevance grouping. Otherwise g_c is not retained.

Notice that in the experiments, for computational cost, the maximum number of neighbors per combination is set to 5 or 6, that is $C \leq 31$ or $C \leq 63$.

4.3.2. Selection of the best global groupings

A set G of potential groupings is now defined. The potential groupings that will be really performed are selected in G by increasing order of energies. When a grouping g_s is selected, any grouping in G which intersects with g_s becomes prohibited. Then, the true merging within each selected grouping can be done. Another level of the pyramid is created to obtain the new partition. This selection ensures the best groupings spread on the whole image.

The grouping process is iterated until the number of vertices remains constant in 2 successive levels. These different pyramid levels constitute the multiresolution aspect of the method.

5. RESULTS

Different results obtained with our method are shown in figures 4, 5 and 6. For complex images whose dimension is around 300×300 pixels, the local segmentation generally provides a partition of 100-200 regions. From this partition, the grouping stage gives a partition of less than 20 regions.

In general the similarity-based stage converges in less than 100 levels and the perception-based stage takes only 10-15 additional levels.

In figure 4.d a partition of 13 regions is obtained which defines with a good accuracy the animals. In figure 5.d the method gives only 2 regions whose one defining the animal.

Figure 6 illustrates the fact that increasing the number of neighbors per grouping can help to make a better choice. But it also increases the processing time and the number of regions in the partitions is roughly the same.

The multiresolution aspect of the pyramid constitutes a great advantage because when in the upper levels of the pyramid, relevant objects are unfortunately lost, the user can easily go down through the levels to find them back.
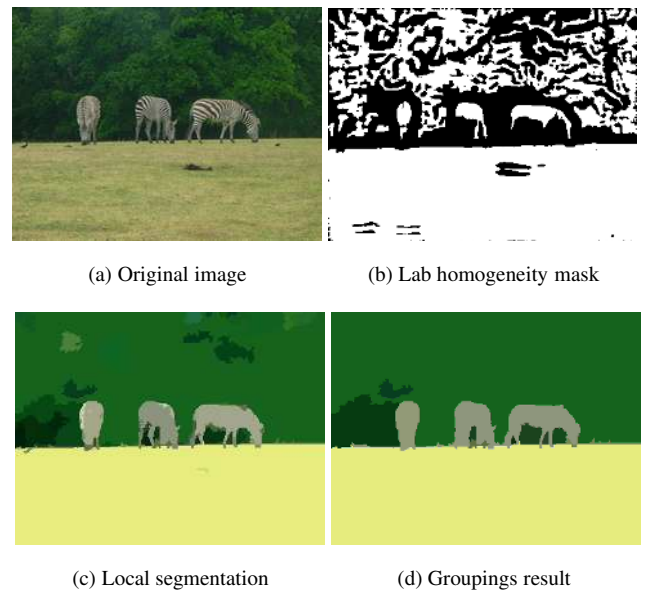

(a) Original image   (b) Lab homogeneity mask

(c) Local segmentation   (d) Groupings result

Fig. 4. Different levels of the segmentation with several relevant objects

6. CONCLUSION

We have presented a generic method combining an accurate localized segmentation that takes into account the color information of the image, with a grouping process that merges regions to produce objects that are perceptually relevant. The final result is a stack of partitions with very few objects. This method is specially suitable either to describe the structure of an object or for supervised segmentation.

In the future, we would like to study in detail the interactions of the different energies to be able to learn more about the relevance of each of them according to the local neighborhood. We also aim at processing videos to add the temporal information to this process. It could provide additional merging information or be used to query the stack of
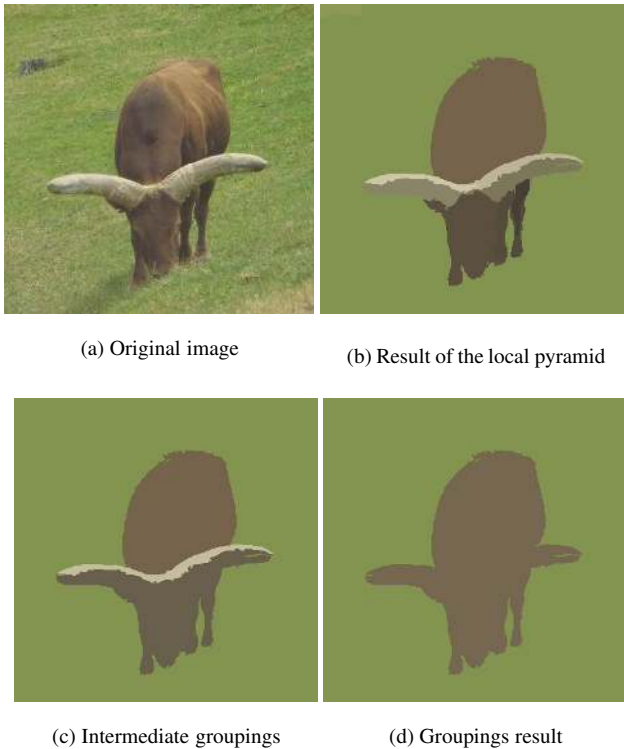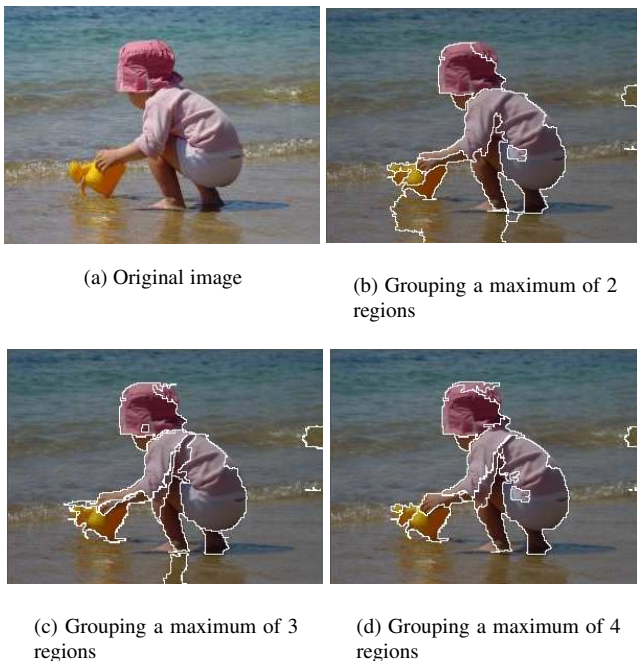
partitions.

7. REFERENCES

[1] Hieu Tat Nguyen, Marcel Worring, and Rein Van Den Boomgaard, "Watersnakes: Energy-driven watershed segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, March 2003, vol. 25 (3).

[2] Jiebo Luo and Cheng en Guo, "Perceptual grouping of segmented regions in color images," in *Pattern Recognition*, April 16 2003, pp. 2781–2792.

[3] S.C. Zhu and A.L. Yuille, "Region competition: Unifying snakes, region growing and bayes/mdl for multiband image segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, September 1996, vol. 18 (9), pp. 884–900.

[4] Ana Petrovic, Oscar Divorra Escoda, and Pierre Vandergheynst, "Multiresolution segmentation of natural images: from linear to nonlinear scale-space representations," in *IEEE Transactions on Image Processing*, august 2004, vol. 13 (8).

[5] Ming Tang and Songde Ma, "General scheme of region competition based on scale space," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, December 2001, vol. 23 (12).

[6] A. Montanvert, P. Meer, and A. Rosenfeld, "Hierarchical image analysis using irregular tessellations," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, April 1991, vol. 13(4), pp. 307–316.

[7] F. Jing, M. Li, H. Zhang, and B. Zhang, "Unsupervised image segmentation using local homogeneity analysis," in *Proc. IEEE International Symposium on Circuits and Systems*, 2003.

[8] J. Huart, G. Foret, and P. Bertolino, "Moving object extraction with a localized pyramid," in *7th International Conference on Pattern Recognition*, Cambridge, UK, august 2004.

[9] M. Wertheimer, "Principles of perceptual organization," in *Readings in Perception*, 1958, pp. 115–135.

[10] N. Zlatoff, B. Tellez, and A. Bazkurt, "Image understanding and scene models: a generic framework integrating domain knowledge and gestalt theory," in *ICIP*, Singapore, October 24-27, 2004.

(a) Original image

(b) Result of the local pyramid

(c) Intermediate groupings

(d) Groupings result

Fig. 5. Different levels of the segmentation with one relevant object

(a) Original image

(b) Grouping a maximum of 2 regions

(c) Grouping a maximum of 3 regions

(d) Grouping a maximum of 4 regions

Fig. 6. Results with different maximum numbers of regions at each local grouping