

The Redundancy Pyramid and its Application to Image Segmentation ¹⁾

Allan Hanbury, Jocelyn Marchadier and Walter G. Kropatsch

Pattern Recognition and Image Processing Group, Vienna University of Technology

Favoritenstraße 9/1832, A-1040 Vienna, Austria

e-mail: {hanbury, jm, krw}@prip.tuwien.ac.at, <http://www.prip.tuwien.ac.at/>

Abstract:

Irregular pyramids organise several topological partitions which can be deduced from a partition (called the base of the pyramid) by successive unions of regions. In this paper, we introduce the redundancy pyramid structure. This structure accounts for redundant topological structures present in several topological partitions. We apply it to the problem of segmentation fusion, where we show that using the redundancy of the boundaries of several partitions leads to useful segmentations. These are evaluated using a publically available segmentation benchmark.

1 Introduction

Image segmentation is an important component of many machine vision applications. In general, segmentation techniques aim to partition an image into connected regions having homogeneous properties. Properties taken into account vary according to a specific technique from constant or smoothly varying intensity, to colour and complex textures. More generally, these techniques can be classified based on three different aspects:

- The **model** used for segmenting, the required aspect or interpretation of the “homogeneity” stated above. It is generally explicitly stated in the methods.
- The **optimized criterion**, which is in general implicit to the different methods. It is related to the model used, and has a great influence on the quality of the results.
- The **algorithm** which is employed to compute the segmentation. Even when the optimized criterion is clearly stated, the algorithms for computing segmentations are often sub-optimal compared to the optimized criterion.

A large variety of segmentation methods can be found in the literature [8]. Of more interest to us are the segmentation methods which try to retrieve regions using different models (e.g.

¹⁾This work was supported by the Austrian Science Foundation (FWF) under grants P14445-MAT and P14662-INF and the EU NoE MUSCLE (FP6-507752).

in [7]). The idea of using a combination of different segmentations to obtain the best segmentation of an image has been suggested by Cho and Meer [1]. However, they make use of small differences resulting from random processes in the construction of a Region Adjacency Graph (RAG) pyramid to generate their segmentations. To our knowledge, no general principle for combining segmentations retrieved using different techniques has been presented.

How can it be done? In this contribution, we emphasize the *redundancy* principle, a general principle which is widely used in robust estimation, and more generally, but implicitly, in robust computer vision techniques. In (robust) estimation, redundancy is defined as the difference between the number of parameters of a functional model, and the number of equations [2]. When the redundancy increases, the computed model is not only more precise, but also more reliable. In general, if r is the redundancy of a functional model, then $r - 1$ gross errors (i.e. aberrant measurements introduced in the model) can be detected [2]. In stereo matching, it is well known that the most reliable procedures are those that use more than 3 images. This is because spurious matches can be filtered using the *redundant* images. Even if some noisy primitives have been detected on one image, they will be left aside during the computation, as random noise usually does not *repeat* on consecutive images.

The last statement is also true for segmentations of a single image. Consider that an image has been segmented by different techniques. Usually the noise introduced by a particular technique will not be repeated on the other segmentations. Taking the redundancy of the segmentations into account can then filter the noise inherent to a particular segmentation, and produce a segmentation which is more reliable. In Section 2, we propose taking the redundancy of two dimensional partitions into account by a redundancy pyramid. The use of the redundancy pyramid framework in the context of segmentation is explored in Section 3, where we also evaluate the results using a publically available segmentation benchmark.

2 The boundary redundancy pyramid

We want to account for the redundant sub-parts of a number of topological partitions (or maps) of the plane. We suppose that those partitions have been aligned beforehand. This is true when they are obtained from different segmentations of the same image or from segmentations of different images observed by a stationary camera under different illumination conditions.

A partition P_i of the plane is defined as a set of closed connected regions $r_{i,j} \in P_i$ partitioning the plane. The intersection of two regions r and r' is the region $r \cap r'$ which contains only the boundary common to the two regions (it is not empty if the regions are adjacent as the regions are closed sets). We say that two regions r and r' are adjacent if $r \cap r'$ is reduced to connected paths. r and r' are disjoint if $r \cap r' = \emptyset$. r and r' are overlapping if $r \cap r'$ is neither empty nor reduced to paths. The union of two adjacent regions r and r' is the region $r \cup r'$.

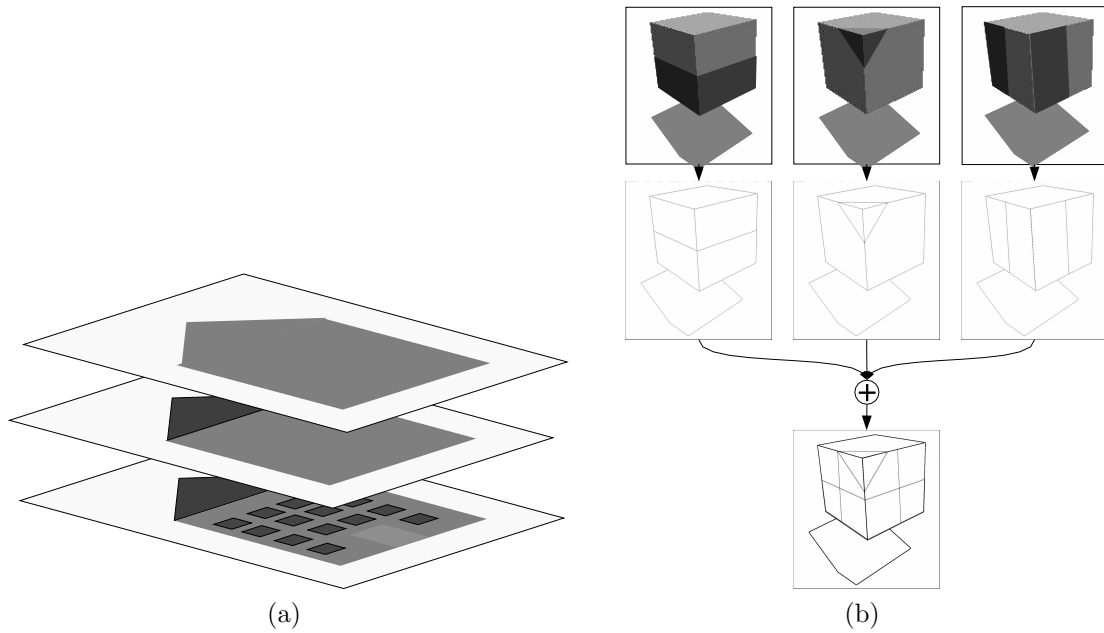


Figure 1: (a) The two top partitions depicted on this figure can be obtained from the first one by merging regions. The second partition represents the facade and roof of the schematic house, and the third one represents the house. (b) Three different cubes have a redundant structure as shown in the stacked redundancy pyramid.

Let us consider a set $\mathcal{P} = \{P_1, \dots, P_n\}$ of n partitions. We want to account for the redundancy of substructures of P_1, \dots, P_n . More precisely, we want to study the redundancy on the set of regions and arbitrary unions of regions. The main idea is to produce the intersection of the n partitions. The intersection is a partition whose regions are exactly composed of sub-regions. Intuitively, trying to merge regions in single partitions in order to see redundant unions of regions on all the partitions can be very complicated. By observing that this analysis can be performed simply by counting redundant edges of regions, we can deduce a simple structure which constructs redundant unions. This structure is based on irregular pyramids [6], which enable the representation of hierarchies of bidimensional partitions of images, like the one depicted in Figure 1a. Each partition at an upper level is deduced from a lower level by merging adjacent regions, merging adjacent edges which are in between exactly two regions, and removing dangling edges which are contained in a region. The boundary redundancy pyramid is simply a pyramid whose first level is the intersection of the partitions used, while the upper levels are constructed by erasing the edges that are redundant in less than l partitions, where l is the constructed level. A simple example is depicted in Figure 1b. It shows three cubes with different colours (in the upper part), similar to the example studied by Keselman et al. [5]. The partition into regions can be seen below. Finally, the redundancy pyramid is in the lower part of the figure, where edges have been colored according to their redundancy. The dark edge, which corresponds to the generic shape of the cube, has redundancy 3. The other edges have redundancies of 1, as they appear only in a single image.

3 Application to segmentation

We apply the redundancy approach to segmentation in two ways. Firstly, we add together a number of different gradients of an image. Secondly, we construct the redundancy pyramid by calculating a series of watershed segmentations of the combined gradient.

3.1 Combining segmentations

The first use of the redundancy principle is in combining a number of gradients of an image, each calculated based on different criteria. We make use of ten different gradients, which are described in more detail in the next section. We simply add the gradients together to form a single combined gradient image G . Each individual gradient image is stretched to fill the full dynamic range of an 8-bit image before being added to the total, thereby ensuring that each gradient image has a similar amount of influence on the result. The combined gradient image for the building image of Figure 2 is shown in Figure 3d.

The construction of the redundancy pyramid converts the combined gradient image into a stack of region boundaries. The pyramid is built as the superposition of a series of watersheds starting from different minima calculated on the combined gradient image. The minima are located using the h -minima operator [11], which filters out minima whose depth is less than h . The input of the algorithm is the combined gradient image G , and it proceeds as follows:

1. set an accumulator A of the same size as G to zero.
2. for h from 1 to *upperlimit* in steps of *stepsize*
 - (a) Calculate the h -minima of G for the current value of h .
 - (b) Calculate the watershed of G using the minima as the seeds where the flooding process starts.
 - (c) Extract the segment boundaries from the resultant watershed segmentation to form a binary image B .
 - (d) Add B to the accumulator A .

The accumulator now contains a count of the number of times each segment border appears. This is basically the construction of a hierarchy based on the depth of minima in the gradient image. As h becomes larger, the shallower minima fuse with the deeper minima, eliminating some segment boundaries. Note that the watershed boundaries at different levels are extremely stable, as the only effect of increasing the value of h is to fuse some minima, which in the watershed segmentation removes only the boundary between these two minima without changing the shape of the remaining boundaries. The upper limit for h (*upperlimit* in the algorithm) can be fixed or chosen adaptively. For the results presented, we used a fixed value of 751. The value of *stepsize* changes the fineness of the pyramid levels, and hence also the total number of levels. We use a value of 50 for *stepsize*. The result of this redundancy pyramid

algorithm applied to the combined gradient of the building image is shown in Figure 3e. This algorithm is similar to the one based on dynamics proposed by Najman and Schmitt [10].

3.2 Gradients

We make use of ten gradients, labelled from $G0$ to $G9$. $G0$ is a Deriche edge detector applied to the luminance channel of the colour image, with $\sigma = 1.5$. $G1$ is a combined hue and saturation gradient, calculated with the circular-centred morphological gradient [4] using a Euclidean distance within the hue-saturation disc. $G2$ – $G6$ are the five colour invariant gradients proposed by Geusebroek et al. [3]. These are invariant to various subsets of viewing direction, surface orientation, highlights, illumination direction, illumination intensity and illumination colour. For $G7$ – $G9$ we use the boundaries obtained by the CG, BG+TG and BG+CG+TG methods introduced by Martin et al. [9]. These are obtained by training a classifier to detect features on an image corresponding to object boundaries. The classifier makes use of features based on brightness and colour gradients (BG and CG respectively) or texture gradients (TG) or combinations of these. It is trained on human segmentations of 200 images, with each image having been segmented by five or more humans. In the resultant images, each pixel in the image is assigned a confidence level based on the certainty that it is part of an object boundary or not. As such, these are not strictly gradient images. They have the disadvantage that the object contours found are not necessarily continuous. It is therefore not obvious how one should convert from the contour representation to a region representation.

3.3 Segmentation evaluation

We made use of a set of 300 images with ground truth segmentations and an evaluation protocol available as the Berkeley Segmentation Dataset and Benchmark¹⁾ [9]. Each ground truth segmentation has been done by at least five humans. For the exploratory work described in this paper, seven images from this database were used, with the results of three of these being discussed in more detail. These three images and the corresponding human segmentations are shown in Figure 2. The evaluation method proposed in the Berkeley benchmark is based on contours. Regions are not taken into account at all. The benchmark produces a precision-recall curve of the segmentation of each image. A series of n thresholds are applied to the computer result, extracting all pixels which have probabilities above the threshold. Exact correspondances between the pixels in each of these binary images and the ground truth are found by solving a minimum cost bipartite assignment problem [9]. The *precision* is then the fraction of detections that are true positives (i.e. correspond to points in the ground truth) rather than false positives, while *recall* is the fraction of true positives that are detected rather than missed. A plot of the precision and recall values at each threshold is finally produced.

¹⁾<http://www.cs.berkeley.edu/projects/vision/grouping/segbench/>

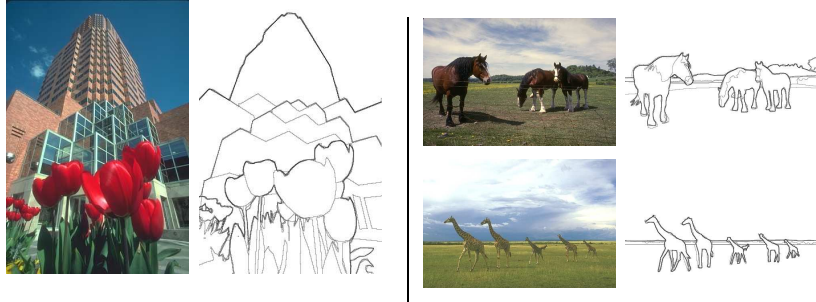
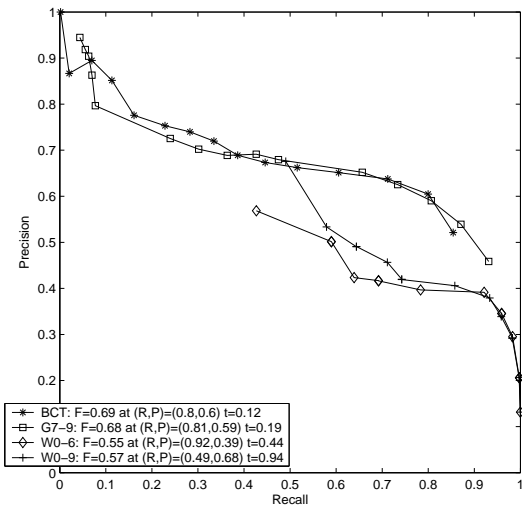


Figure 2: The images discussed (on the left of each pair) and their corresponding human segmentations. Segment boundaries drawn in by more than one human appear darker. The image ID numbers in the database are 86000, 197017 and 253055 respectively, although we refer to them as the *building*, *horse* and *giraffe* images.

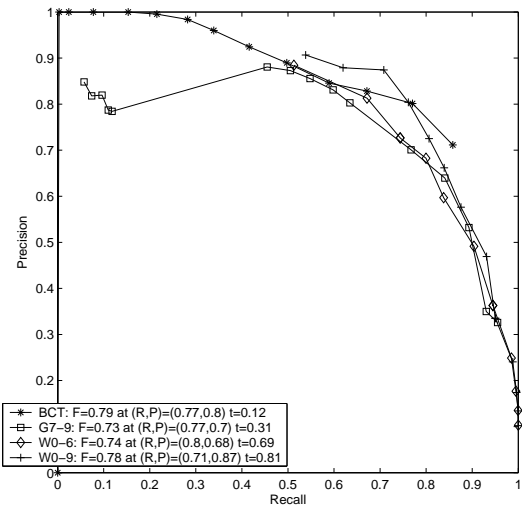
The whole curve is summarised by a single F -value, which, for a set of n precision-recall values (P_i, R_i) , is calculated as $F = \max \{2P_iR_i / (R_i + P_i) \mid i = 1, \dots, n\}$

We evaluated three different segmentation schemes. The precision-recall curves for the three example images are shown in Figure 3a–c. For each curve, 15 equally-spaced thresholds of the segmentation image were computed. The segmentation schemes are labelled G7-9, W0-6 and W0-9. The notation $Gx-y$ indicates the sum of all the gradients between and including Gx and Gy . The notation $Wx-y$ indicates the redundancy pyramid built on the $Gx-y$ image. W0-6 therefore contains a redundancy pyramid built on the sum of the seven true gradients, i.e. excluding the Martin et al. boundaries. W0-9 contains the redundancy pyramid built on the sum of all the gradients. Note that it is impossible to apply the watershed operator to the G7-9 image as the contours in this image are not closed. One therefore finds very few minima in this image, with each one spanning many regions. This once again highlights the difficulty of converting this boundary representation to a region representation. We therefore apply the benchmark directly to the G7-9 image. In each precision-recall graph, the best result of the benchmark applied to the Martin et al. BCT, BT and C images independently is shown for comparison. For the images chosen, this happens to always be the BCT image.

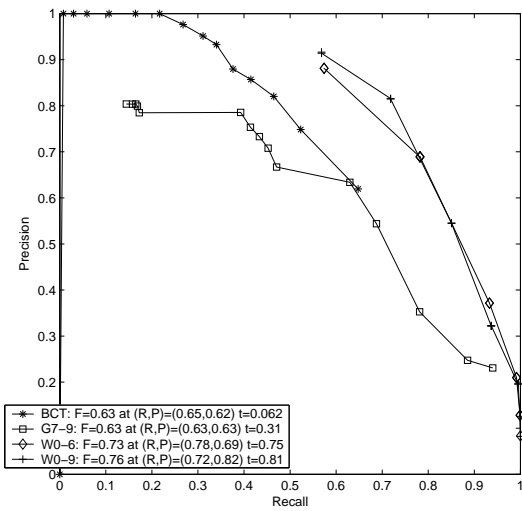
The redundancy pyramid results ranged from worse to significantly better than the Martin et al. BCT method. We have chosen one example illustrating a significant improvement, one illustrating a similar result and one illustrating a poorer result. For the giraffe image, the addition of the extra gradients resulted in a better detection of the boundaries of the bottom part of the giraffes, improving the F -score from 0.63 for the Martin et al. BCT boundaries to 0.76 for W0-9. For the horse image, the results of the redundancy pyramid are similar to those of the BCT approach. An analysis of the results of the building image segmentation provides interesting insights. The W0-9 curve begins at a point on the BCT curve, but then descends towards the right at a faster rate than this curve. The reason, as can be seen in Figure 3e, is that our segmentation also includes the building details, such as the windows, albeit at a lower level of the redundancy pyramid. The presence of these extra unmatched



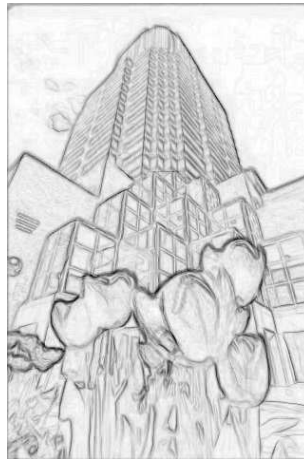
(a) Building



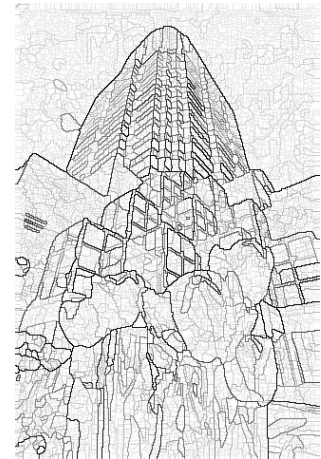
(b) Horse



(c) Giraffe



(d) Gradient G0-9



(e) Redun. pyr. W0-9

Figure 3: (a)–(c) The precision-recall curves for the three example images. (d) Combined gradient G0-9. (e) Redundancy pyramid W0-9 for the building image.

lines in the redundancy pyramid segmentation therefore decreases the precision. This is a good demonstration of the problem of generating a good segmentation ground truth based on segmentations provided by humans, who tend to have a-priori context information available. The Martin et al. methods attempt to overcome this problem by learning the characteristics of boundaries in local neighbourhoods based on the human segmentations. However, they never produce 100% recall, so there are always boundaries missing in the results, even at the lowest threshold value. The redundancy pyramid contains all the contours in all three examples, although they only all appear at low values of precision. This is a strong argument to modify the benchmark to focus less on the presence of parts of contours (the presence of a piece of contour is not very useful in object recognition) by incorporating region information.

4 Conclusion

We have presented the redundancy principle, and proposed a structure, the redundancy pyramid, that can be used to perform the analysis of redundant sub-structure of several topological partitions. We have explored its use in image segmentation. In this application, we make use of the redundancy principle twice: first in the construction of the combined gradient image, and second in building the redundancy pyramid of region boundaries. We have evaluated the segmentations obtained using a publically available benchmarking database.

Our results confirm that it is not simple to improve image segmentations in a general way. Even segmentation by humans is guided by what they intend to do with the segmentation, their purpose. We therefore suggest making use of the redundancy pyramid to develop purpose-guided segmentation. One can envisage, for example, implementing learning of a gradient weighting which best corresponds to the user's purpose (based on interaction with the user). Further work includes the comparison of the segmentation methods with the whole benchmark database, as well as the definition of an addition to the benchmark which can take region subdivision into account.

References

- [1] K. Cho and P. Meer. Image segmentation from consensus information. *Computer Vision and Image Understanding*, 68(1):72–89, 1997.
- [2] W. Förstner. Generic estimation procedures for orientation with minimum redundant information. 2nd Course on Digital Photogrammetry, 1999.
- [3] J.-M. Geusebroek, R. van den Boomgard, A. W. M. Smeulders, and H. Geerts. Color invariance. *IEEE Trans. PAMI*, 23(12):1338–1350, December 2001.
- [4] A. Hanbury and J. Serra. Morphological operators on the unit circle. *IEEE Trans. Image Processing*, 10(12):1842–1850, December 2001.
- [5] Y. Keselman and S. Dickinson. Generic model abstraction from examples. In *Proc. CVPR*, volume 1, pages 856–863, December 2001.
- [6] W. G. Kropatsch. Abstraction Pyramids on Discrete Representations. *Discrete Geometry for Computer Imagery, 10th DGCI*, pages 1–21, Bordeaux, France, 2002.
- [7] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.
- [8] L. Lucchese and S. Mitra. Color image segmentation: A state-of-the-art survey. *Proc. of the Indian National Science Academy (INSA-A)*, 67(2):207–221, 2001.
- [9] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. PAMI*, 26(5):530–549, 2004.
- [10] L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Trans. PAMI*, 18(12):1163–1173, 1996.
- [11] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag, 1999.