Combining regular decimation and dual graph contraction for hierarchical image segmentation

Fuensanta Torres[a], Rebeca Marfil[a], Yll Haxhimusa[b] and Antonio Bandera[a]

[a]Grupo ISIS, Departamento de Tecnología Electrónica, Universidad de Málaga, Spain
[b]Pattern Recognition and Image Processing, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Austria

fuensantatorres@uma.es, rebeca@uma.es, yll@prip.tuwien.ac.at, ajbandera@uma.es

Abstract The Bounded Irregular Pyramid (BIP) is a hierarchical structure for image representation whose aim is to combine concepts from regular and irregular pyramids. The data structure is a combination of the simplest regular and irregular structures: the $2 \times 2/4$ regular one and the simple graph representation. However, simple graphs only take into account adjacency relationships, being unable to correctly encode the topology of the image. This paper proposes a new version of the BIP, where the regular decimation process is now merged with a stochastic graph decimation strategy. Experiments demonstrate that this new irregular pyramid is able to provide qualitative good segmentation results and to preserve the topology of the input image at higher levels of its hierarchy.

Keyworks image segmentation; regular pyramid; irregular pyramid; dual graph data structure.

1 Introduction

Image understanding aims for analysing an input image into its constituent patterns. Depending on the task, we could be interested on decomposing the image into regions which are homogeneous according to some criteria such as colour or texture [9]. This problem is called image segmentation. The low-level features, like brightness, colour, texture and/or motion attributes, should be used to build sequentially a hierarchy of segmentations [11]. A grouping method should have the following properties [5]: i) to capture perceptually important groupings or regions, which reflect global aspects of the image, ii) to be highly efficient, running in time linear in the number of pixels, and iii) to create hierarchical segmentations [11]. In general, low-level cue image segmentation cannot produce a complete final 'good' segmentation [3].

A pyramid representation consists of a stack of layers, each layer involving a number of processors (nodes). Each processor has information about one part of the image, called its receptive field. The bottom layer (layer 0) corresponds to the level of elements on the sensor (e.g. number of pixels) . The number of processors in the second layer (layer 1) is smaller by a factor of λ - this factor is called a reduction ratio, and it is greater than one. Specifically, one processor (a 'parent') receives input from λ 'children' in the layer below. Processors on higher layers receive only some information about the content of their receptive fields. Pyramids are hierarchical structures which have been widely used to address the image segmentation task [9]. In the pyramid hierarchy, each level is built by computing a set of local operations over the level below. These local operations adapt the hierarchy to the topology of the image, allowing interesting features such as the detection of global features of interest at higher levels of the hierarchy and the reduction of the complexity of the image segmentation task. If each level of the pyramid is encoded as a graph, then the nodes of any level over the base will be linked to the nodes of the level above. That is, the pyramid provides multiple representations with different levels of details (i.e. resolution) for each region of the segmented image. This property can be very interesting for certain tasks:

global relationships of the image can be analysed at higher levels of the hierarchy where the whole image is encoded using a reduced set of nodes; but detailed versions of the regions associated to these nodes are simultaneously available at low levels.

The efficiency of the pyramid in image segmentation is strongly influenced by two characteristics: the data structure used to encode every level of the pyramid (horizontal relations) and the decimation scheme employed to build one level from the level below (vertical relations) [4]. The data structure determines the information which may be encoded at each level of the pyramid (e.g., simple graphs, dual graphs, combinatorial maps...). Thus, it roughly defines the horizontal relations of the entities within the same level of the pyramid. On the contrary, the decimation scheme employed to build the pyramid sets the dynamic properties of the pyramid (height, preservation of details, etc.). It corresponds to the vertical relations between the entities between two subsequent levels of the pyramid. Taking into account these two characteristics, pyramids can be classified as regular or irregular ones. Regular pyramids have a rigid structure where the relationships among nodes of the same level are fixed and the reduction ratio between the number of nodes of two consecutive levels are constant and bigger than one. In these hierarchies, the pyramid structure can be adapted to the image topology by means of the relationships among nodes of consecutive levels. As it was pointed out by Bister et al [2] or Antonisse [1], this is typically inadequate to preserve the topology of the input image at higher levels of the hierarchy. On the other hand, irregular pyramids can adapt their spatial relationships (among nodes of the same level and among nodes of consecutive levels) to the content of the input image. Besides, although the original proposals presented a serious drawback with respect to computational efficiency, this problem has been resolved by the last proposed strategies [4, 8].

With the objective of having the best properties of both types of pyramids, the bounded irregular pyramid (BIP) was proposed by Marfil et al [10]. The BIP is a hierarchical structure whose data structure combines the simplest regular and irregular structures: the $2 \times 2/4$ regular one and the simple graph irregular representation. The decimation scheme is also decomposed into two stages, one of them dealing with the regular structure and the other one with the irregular structure. Both decimation processes employ the sequential union-find scheme [4]. The aim is to apply the regular decimation in the homogeneous parts of the image, meanwhile the heterogeneous parts are decimated using the irregular process. Thus, the BIP approximates or even outperforms previously proposed hierarchical segmentation approaches, yet it can be computed much faster [9]. However, it was originally highly affected by the shift variance problem, i.e. it provides an image segmentation which varies when the image is shifted slightly. This problem has been recently solved [12]. However, the new versions of the BIP employs still the simple graph to encode the relationships among nodes of the same level. Simple graphs only take into account adjacency relationships, being unable to distinguish from the graph an adjacency relation from an inclusion relation between two regions. Besides, if there is two non-connected boundaries which will join one region to another one, the simple graph only joins these nodes by one arc. These limitations can be raised if dual graphs are employed because their structure is adapted to the processed data and they correctly encode the topology in 2D. In this paper, the dual graph data structure and the maximal independent edge set (MIES) decimation process proposed by Haxhimusa et al [6] are used to deal with the heterogeneous parts of the image. The use of the dual graph allows to preserve the topology of the image and to correctly encode the relation of adjacency and inclusion between image regions.

The rest of the paper is organised as follows: Section 2 describes the proposed segmentation approach, analysing the decimation processes employed to deal with the regular and irregular parts of the pyramid. Section 3 presents some preliminar experimental results. Finally, conclusions and future work are unfolded at Section 4.

2 Segmentation procedure

The aim of the proposed segmentation approach is to maintain the topological advantages of the dual graph pyramid but reducing its computational load. To do that, this paper propose to use a $2 \times 2/4$ regular decimation in the homogeneous areas of the input image and a dual graph contraction scheme in the rest of the image. Following the BIP strategy, the obtained pyramid can be seen as a combination of regular and irregular structures presenting two types of nodes: nodes belonging to the regular structure (regular nodes) and nodes belonging to the irregular one (irregular nodes). The mixture of both regular and irregular structures generates an irregular configuration which is described as a graph hierarchy in which each level $G_l = (N_l, E_l)$ consists of a set of nodes, N_l, linked by a set of *intra-level edges* E_l. Two nodes $n_i \in N_l$ and $n_j \in N_l$ which are neighbours at level l are linked by an intra-level edge $e_{ij} \in E_l$. These edges define the horizontal relationships of the pyramid.

Each graph G_{l+1} is built from G_l by computing the nodes of N_{l+1} from the nodes of N_l and establishing the *inter-level edges* $E_{l,l+1}$. Therefore, each node n_i of G_{l+1} has associated a set of nodes of G_l, which is called the *reduction window* of n_i. This includes all nodes linked to n_i by an inter-level edge. The node n_i is called *parent* of the nodes in its reduction window, which are called *children*. Two nodes n_i and n_j of N_l are said to be adjacent or *neighbours* at level l, if their corresponding reduction windows w_i and w_j are neighbours at level $l-1$. Two reduction windows $w_i \in N_{l-1}$ and $w_j \in N_{l-1}$ are neighbours if there are at least two nodes $n_r \in w_i$ and $n_s \in w_j$ which are connected by an intra-level edge $e_{r,s} \in E_{l-1}$. The set of nodes in N_l which are neighbours of a node $n_i \in N_l$ is called the *neighbourhood* of n_i. The *receptive field* r_i of a node $n_i \in N_l$ is the set of its children at level 0.

The value of each parent is computed from the set of values of its children using a *reduction function*. Using this general framework, the procedure to build the level G_{l+1} from level G_l consists of three steps:

1. Selection of the nodes N_{l+1} of G_{l+1} among N_l : This selection step is a decimation procedure and selected nodes N_{l+1} are called the surviving nodes.

2. Inter-level edges definition: Each node of G_l is linked to its parent node in G_{l+1}. This step defines a partition of N_l.

3. Intra-level edges definition: The set of edges E_{l+1} is obtained by defining the spatial relationships between the nodes N_{l+1}.

In the proposed irregular structure, the regular part is built using a union-find procedure [4] and the irregular part uses the dual graph contraction scheme proposed by Haxhimusa and Kropatsch in [6]. Therefore, the Algorithm 1 shows the process to build the graph pyramid. Next Sections describe the regular decimation process implemented following the union-find strategy and the decimation process employed to deal with the irregular part of the pyramid.

Algorithm 1 Process to build the proposed structure

```
1: k = 0, Input: graph map G_0
2: repeat
3:    Regular decimation on the regular nodes of G_k
4:    Find the edges of the irregular part of G_k with the smallest weight and build the equivalent
      contraction kernel (ECK)
5:    Dual graph contraction on the edge contraction kernel
6:    k = k + 1
7: until G_k = G_{k-1}
end
```

2.1 Regular decimation process

Although regular pyramids can be explained as a graph hierarchy, it is more usual to represent them as a hierarchy of image arrays due to their rigid structure. Therefore, in the regular part of this proposed pyramid, each regular node is represented by (i, j, l), where l represents the level and (i, j) are the x- and y-coordinate within the level. In each of these arrays two nodes are neighbours if they are placed in adjacent positions of the array in an 8-neighbourhood.

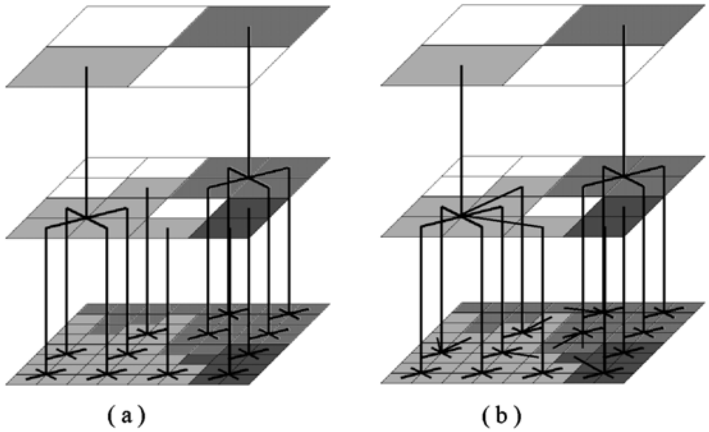

Figure 1: a) Regular nodes and their inter-level edges after the generation step, and b) regular nodes and their inter-level edges after the parent search step.

The first step to build the $2\times2/4$ structure is a 4 to 1 decimation procedure. In order to perform this decimation, each regular node has associated two parameters: homogeneity $Hom(i,j,l)$ and parent link $Parent(i,j,l)$. Regular nodes have $Hom(i,j,l) = 0$ or $Hom(i,j,l) = 1$. $Hom(i,j,l)$ of a regular node is set to 1 if the four nodes immediately underneath are similar according to some criteria and their homogeneity values are equal to 1. Otherwise, it is set to 0. If the node (i,j,l) is a node of the regular structure with $Hom(i,j,l) = 1$, then the parent link of the four cells immediately underneath (sons) is set to (i,j). It indicates the position of the parent of a regular node in its upper level. A regular node without parent has its parent link set to a NULL value. Parent links represent the inter-level edges of the regular part of the proposed pyramid.

All the regular nodes presenting a homogeneity value equal to 1 form the regular structure. Regular nodes with an homogeneity value equal to 0 are removed from the structure. Fig. 1a shows the regular part of the BIP data structure after being built. White nodes are the non-homogeneous ones. In this example the used similarity criteria is the colour distance. Two nodes are similar if they have similar colour. The base level of the structure is formed by the pixels of the 8x8 original image. The 4 to 1 decimation procedure generates a 4×4 level and a subsequent 2×2 level.

Once the regular structure is generated using the 4 to 1 decimation procedure, there are some regular orphan nodes (regular nodes without parent). From each of these nodes, a search is made for a node (i_1,j_1,l) in its 8-neighbourhood $\xi_{(i,j,l)}$ which satisfies the following conditions:

- $Hom(i_1,j_1,l) = 1$
- $Parent(i_1,j_1,l) = (i_p,j_p,l+1)$
- $d((i,j,l),(i_1,j_1,l)) < T$ (1)
- $d((i,j,l),(i_1,j_1,l)) \leq d((i,j,l),(i_k,j_k,l)) \quad \forall (i_k,j_k,l) \in \xi_{(i,j,l)}$

being $d(n_i,n_j)$ a similarity measurement between the nodes n_i and n_j and T a similarity threshold. (i,j,l) is linked to $(i_p,j_p,l+1)$ (*parent search step*). For example, in Fig. 1b, there are four orphan nodes at level 1, but only for two of them a suitable parent node is found that satisfies (1).

Specifically, the proposed approach employs the similarity criterion proposed in [6], which is based on the internal and external contrast measurements. This pairwise group merge criterion will be briefly explained in Section 2.2.

2.2 Dual graph contraction

A simple graph is a non-weighted and undirected graph containing no self-loops. In this hierarchy, the graph edges E_l represent adjacency relationships among pyramidal nodes of the level l. Simple graphs encode the adjacency between two nodes by only one edge, although their receptive fields may share several boundary segments. Therefore, a graph edge may thus encode a non-connected set of boundaries between the associated receptive fields. Moreover, the lack of self-loops in simple graphs does not allow to differentiate an adjacency relationship between two receptive fields from an inclusion relationship. In a dual graph pyramid, a level consists of a dual pair $(G_l, \tilde{G}_l)$ of planar graphs G_l and $\tilde{G}_l$. If level l defines a partition of the image into a connected subsets of pixels, then the nodes of G_l are the representatives of these subsets and the edges of G_l represent their neighborhood relationships. The edges of $\tilde{G}_l$ represent the boundaries of these connected subsets in level l and the nodes of $\tilde{G}_l$ define meeting points of boundary segments of $\tilde{G}_l$.

Within the dual graph pyramid framework, the set of edges that define the adjacency relationships among nodes of the level $l+1$ is generated in two steps. First, the set of edges that connects each non-surviving node to its parent is contracted using a contraction kernel. A contraction kernel of a level l is the set of surviving nodes of l and the edges that connect each non-surviving node with its parent. The edge contraction operation collapses two adjacent nodes into one node, removing the edge between them. This operation may create redundant edges such as empty self-loops or double edges. The removal of these redundant edges constitutes the second step of the creation of the set of edges E_{l+1}. These redundant edges are characterized in the dual of the graph and removed by a set of edge removal kernels [7]. The key idea of the dual graphs is that a contraction in a graph implies a removal in its dual, and vice versa, in order to maintain the duality between the newly generated graphs. Thus, the generation of the edges in level $l + 1$ can be resumed as follows:

1. Contraction of edges in G_l which connect non-surviving nodes with their parents. Removal of their corresponding edges in $\tilde{G}_l$. Fig. 2b shows the reduction performed by the contraction kernel in Fig. 2a.

2. Contraction of redundant edges in $\tilde{G}_l$ and removal of their corresponding edges in G_l. In Fig. 2c, the dual node a has a face defined by nodes **A** and **B**. The boundary between the regions defined by these nodes is artificially split by this dual node. Then, the two dual edges incident to this dual node (e'_1 and e'_2) can be contracted. The contraction of these dual edges has to be followed by the removal of one associated edge (e_1 or e_2) in order to maintain the duality between both graphs. In the same way, the dual node b encodes an adjacency relationship between two nodes contracted in the same node. This relationship can be removed by eliminating this direct self-loop and contracting the associated dual edge.

Using such a reduction scheme each edge in the reduced graph corresponds to one boundary between two regions. Moreover, inclusion relationships may be differentiated from adjacency ones in the dual graph.

In this proposal, the process to build the dual graph in each level is based on the Hierarchy of Partitions (BoruSeg) algorithm [6]. Let G_k be the graph on level k of the pyramid. Every node $n_i \in G_k$ is a **connected component** $CC(n_i)$ of the partition P_k. The attributes of every edge $e \in G_k$, $attr_e(n_i, n_j)$, is defined in this work as the difference of RGB colour between the end nodes n_i and n_j of edge e. The goal is to find partitions $P_k := \{CC(n_1); CC(n_2); \cdots, CC(n_n)\}$ in k^{th} level of the pyramid such that these elements satisfy certain properties. Neighboring nodes (partitions) are compared pairwise to check for similarities [5]. A pairwise group merge criterion $Comp(CC(n_i), CC(n_j))$ is defined to decide whether there is a boundary between two partitions $CC(n_i), CC(n_j) \in P_k$, i.e. the difference along the boundary of two components is defined relative to a measure of differences of components' internal differences. Note that this defines an intuitive notion of contrast: a contrasted region containing two connected components whose inner differences are less than differences within this two regions. $K_{0,k}(n_i)$ [7] is the equivalent contraction kernel (ECK) of a node $n_i \in G_k$. The **internal contrast** of the $CC(n_i) \in P_k$ is the **largest**

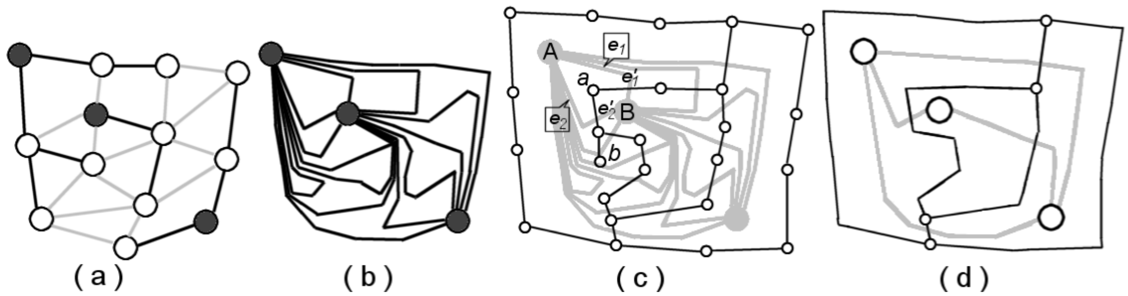

Figure 2: Contraction and removal kernels: a) contraction kernel composed of three nodes (surviving nodes are marked in black); b) reduction performed by the equivalent contraction kernel in a); c) redundant edges characterization; and d) dual graph pair $(G, \tilde{G})$ after dual decimation step.

dissimilarity of component $CC(n_i)$, i.e. the largest edge weight of the $K_{0,k}(n_i)$ of node n_i. Let $n_i, n_j \in V_k$ be the end nodes of an edge $e \in E_k$. The **external contrast** between two components $CC(u_i), CC(u_j) \in P_k$ is the **smallest dissimilarity** between component $CC(n_i)$ and $CC(n_j)$, i.e. the smallest edge weight connecting $N_{0,k}(n_i)$ and $N_{0,k}(n_j)$ of nodes $n_i \in CC(n_i)$ and $n_j \in CC(n_j)$. Formal definition of internal and external contrast can be found in [6]. One can use the internal and the external contrast to define a merging criterion, which check if there is a border or not between to components.

The algorithm to build the hierarchy of partitions (segmentations) is as follows. The idea is to collect the smallest weighted edges (minimum spanning tree principle) e that could be part of the ECK and then to check if the edge weight ($attr_e(e)$) is smaller than the internal contrast of both of the components (internal and the external contrast defined over the ECK of end nodes of e). If the internal contrast of the two components is larger than the external contrast, represented by the weight $attr_e(e)$ of the connecting edge, then the these components are merged. After that all the edges that are going to be contracted form the contraction kernels $K_{k,k+1}$, which is then used to create the graph $G_{k+1} = C[G_k, N_{k,k+1}]$ by using the DGC algorithm [7]. The DGC algorithm [7] contracts edges and creates 'super' nodes with parent-children relations between nodes in subsequent levels of the pyramid, whereas Borůvka's minimum spanning tree principle is used to find edges with the smallest weight and create horizontal relation between nodes in the same level. The algorithm iterates until there are no more edges to contract. Each level of the pyramid is represented by a region adjacency graph (RAG) which is created in an agglomerative way by topology-preserving edge contraction.

3 Experimental results

Fig. 3 illustrates the results obtained from the segmentation of one image using the proposed approach, the original BIP [9] and the hierarchy of partitions (HP) [6]. In all cases, we have chosen as segmented image the lower level of the hierarchy where the leaf is represented as an unique connected component. It must be noted that in the case of the proposed approach this level is the level 12 meanwhile in the case of the HP this level is the level 20. As it is shown in Fig. 4.a, the proposed approach provides higher reduction factors than the HP at low levels of the hierarchy. At these levels, these factors are very similar to the ones provided by the original BIP. Besides, the total height of the hierarchical representation provided by the proposed approach for this image is also lower than the total height of the hierarchy obtained using the HP, specifically these total heights are 12 and 22, respectively. It must be noted that, while in the case of the HP the level for a 'good' segmentation of the leaf is lower than the uppest level of the structure, in the case of the proposed approach both levels are the same. Similar remarks can be concluded when other images are employed for comparison purposes. Finally, Fig. 4.b represents, at logarithmic

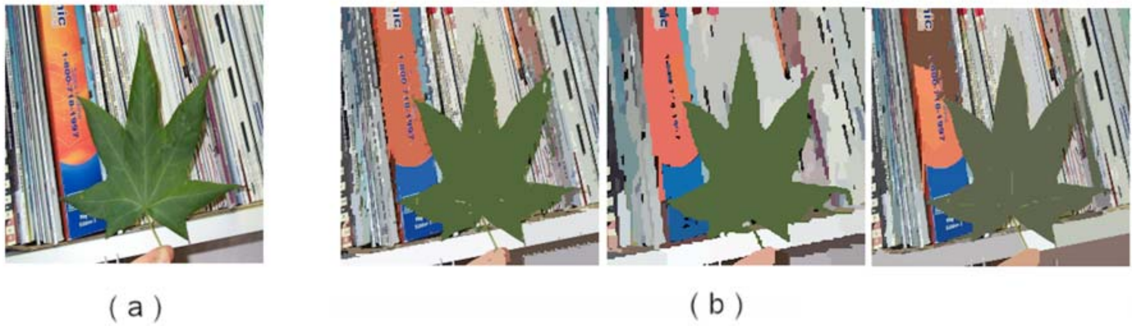

Figure 3: Segmentation example: a) Original image; and b-d) segmentation regions obtained using the proposed approach, the BIP and the HP, respectively.

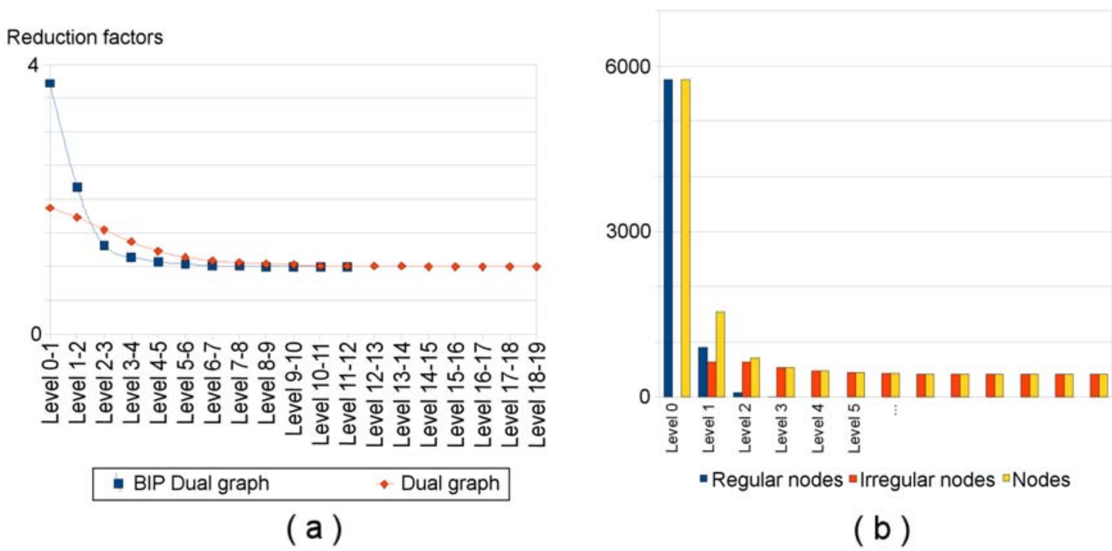

Figure 4: Segmentation statistics associated to the segmentation example shown in Fig. 3: a) Reduction factors; and b) number of regular and irregular nodes with respect to the total number of nodes per level (see text for details).

scale, the total number of nodes of each hierarchy level. It can be noted how the percentage of irregular nodes, with respect to the total number of nodes, increases for higher levels of the pyramid. In fact, at higher levels, the pyramid is only composed by irregular nodes.

4 Conclusions and Future work

This paper presents a novel approach for image segmentation which extends the hybrid decimation scheme employed inside the BIP framework. As in the previous version of the BIP, the regular decimation process encodes the homogeneous parts of the image in a set of connected components, each one of them will be finally linked to an irregular node. The level at which this irregular node is located will depend on the size and shape of its associated receptive field and, at this level, this node will be correctly related to its neighbour irregular nodes. The main novelty of this proposal is that, instead of a simple graph, it now uses a dual graph to encode the irregular part of the pyramid. This will allow to preserve the topology of the original image at upper levels of the hierarchy.

The main drawback of the proposed approach is the use of a sequentially conducted regular decimation process. Future work will be focused on managing the hierarchy using a mixture of regular and irregular decimation processes which could run in a parallel way. Other research line could be focused on studying the possibility of combining the fast original BIP with other topology-preserving strategy in a different way. The BIP could be used at low levels to divide the image in a set of superpixels, which will be then employed by a second topology-preserving stage conducted at higher levels of the hierarchy.

Acknowledgments

This work has been partially granted by the Spanish Junta de Andalucía under project P07-TIC-03106 and by the Ministerio de Ciencia e Innovación (MICINN) and FEDER funds under projects no. TIN2008-06196 and AT2009-0026.

References

[1] H.J. Antonisse. Image segmentation in pyramids. *Comput. Graphics Image Process.*, 19:367–383, 1982.

[2] M. Bister, J. Cornelis, and A. Rosenfeld. A critical view of pyramid segmentation algorithms. *Pattern Recognition Letters*, 11:605–617, 1990.

[3] S. Borra and S. Sarkar. A framework for performance characterization of intermediate-level grouping modules. *Pattern Recognition and Image Analysis*, 19(11):1306–1312, 1997.

[4] L. Brun and W. G. Kropatsch. Construction of combinatorial pyramids. In M. Vento E. Hancock, editor, *Graph Based Representations in Pattern Recognition, LNCS,*, volume 2726, pages 1–12. Springer, Berlin, 2003.

[5] P.F. Felzenszwalb and D.P. Huttenlocher. Image Segmentation Using Local Variation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 98–104, June 1998.

[6] Y. Haxhimusa and W.G. Kropatsch. Segmentation graph hierarchies. In A. L. N. Fred et al., editor, *Joint IAPR Int. Workshops, SSPR2004 and SPR2004*, volume 3138 of *LNCS*, pages 343–351. Springer, 2004.

[7] W. Kropatsch. Building irregular pyramids by dual graph contraction. *IEEE Proc. Vision, Image and Signal Processing*, 142(6):366–374, 1995.

[8] W. Kropatsch, Y. Haxhimusa, Z. Pizlo, and G. Langs. Vision pyramis that do not grow too high. *Pattern Recognition Letters*, 26(3):319–337, 2005.

[9] R. Marfil, L. Molina-Tanco, A. Bandera, J.A. Rodriguez, and F. Sandoval. Pyramid segmentation algorithms revisited. *Pattern Recognition*, 39:1430–1451, 2006.

[10] R. Marfil, J.A. Rodriguez, A. Bandera, and F. Sandoval. Bounded irregular pyramid: a new structure for color image segmentation. *Pattern Recognition*, 37(3):623–626, 2004.

[11] J. Shi and J. Malik. Normalized Cuts and Image Segmentation. In *Proceedings IEEE Conference Computer Vision and Pattern*, pages 731–737, 1997.

[12] R. Vazquez-Martin, R. Marfil, P. Nuñez, A. Bandera, and F. Sandoval. A novel approach for salient image regions detection and description. *Pattern Recognition Letters*, 30:1464–1476, 2009.