Segmentation Based on Level Combination of Irregular Pyramids

Annette Morales-González, Edel Garcia-Reyes
Advanced Technologies Application Center
7a No. 21812 b/ 218 y 222, Siboney, Playa, P.C. 12200, Havana, Cuba
amorales,egarcia@cenatav.co.cu

Luis Enrique Sucar
National Institute of Astrophysics, Optics and Electronics
Luis Enrique Erro No. 1 Tonantzintla, Puebla 72840, MEXICO
esucar@inaoep.mx

Abstract

In image retrieval applications, if we work with hierarchies of image partitions, it is often necessary to select the proper level of segmentation to be used or whether it should combine more than one level in order to annotate the objects. This decision may be even harder if we do not have a previous knowledge regarding the image. The contributions of this work are a new measure to evaluate the segmentation based on persistence of relevant edges and an algorithm to scan the irregular pyramids combining the better segments in order to build a new graceful image in a perceptive manner.

Keywords: irregular pyramids, image segmentation, hierarchical segmentation

1 Introduction

One important step to perform tasks such as object recognition or content based image annotation and retrieval is to segment images into regions that provide relevant information regarding the objects present in the image. The segmentation process usually produces a set of ”homogeneous” regions regarding low-levels features, which are combined by computing their similarity values. Nevertheless, this homogeneity of low-level cues will not map to the semantics of the image, and the degree of homogeneity of a region is in general quantified by threshold(s) for a given measure [7].

The low-level coherence of brightness, color, texture or motion attributes should be combine sequentially as a hierarchy of partitions [18]. Pyramids are hierarchical structures which have been widely used in segmentation tasks [13]. A pyramid segmentation algorithm describes the contents of the image using multiple representations with decreasing resolution. Each representation is built using some criteria for merging regions from the level below, where the base level of the hierarchy (level 0) is the original image. Some examples of this approach are the regular pyramids, irregular graph pyramids and combinatorial pyramids.

Regular image pyramids construct the hierarchy of partitions by using the neighborhood relationships defined on each image. The reduction window, with fixed size and shape, relates each pixel of the pyramid with a set of pixels defined in the level below. The rigidity of the vertical structure of regular pyramids induces several drawbacks, such as the shift-dependence and scale-dependence problem, and the limited number of regions encoded at a given level of the pyramid [2].

The irregular graph pyramid is a stack of successively reduced graphs (being the base level the high resolution input image) where each graph is built from the graph below by selecting a set of vertices

Luis Enrique Sucar and Hugo Jair Escalante (eds.): AIAR2010: Proceedings of the 1^{st} Automatic Image Annotation and Retrieval Workshop 2010. , volume 1, issue: 1, pp. 1-10

named surviving vertices and mapping each non surviving vertex to a surviving one [9]. Each vertex represents a region in the current partition and each edge represents adjacency between the regions represented by the nodes it connects. The main advantage of using this kind of graphs is that they may contain parallel edges and self-loops, which represent several common boundaries and inclusion relations respectively [3].

A combinatorial pyramid is almost like the irregular graph pyramid, but instead of using graphs for the levels, they use combinatorial maps. This representation allows to define better the inclusion relationship, by specifying which region is inside and which one is outside. It also proves to be more computationally efficient than the irregular graph pyramid representation [4].

Another approach to hierarchical segmentation is the Bounded Irregular Pyramid (BIP), which combines the regular and the irregular approach to construct the levels [14]. This method possesses both approaches advantages, such as the speed construction of the pyramid and the ability to adapt their structure to the data, but inherits the shift-variance problem of the regular pyramids and does not preserve important topological relationships such as inclusion and multiple adjacency.

Although these pyramidal structures provide more information of the image producing several representations at different levels of resolution, if we plan to use one of these hierarchical approaches for image retrieval applications, processing the whole pyramid of segmentations for each image can be very time-consuming. In this case, it would be desirable to select the most suitable level of segmentation to be used (according to some criteria) or whether it should be a combination of several levels, in order to annotate the objects. Performing this selection manually may be easy, but problems emerge when we want to do the same thing automatically. This decision may be even harder if we do not have a previous knowledge regarding the image. For this purpose, we consider that a measure that automatically evaluates image segmentations may be a good indicator of wether a particular segmentation level matches some specific requirements.

Several methods have been proposed to evaluate image segmentations. They can broadly be divided into two categories: analytical and empirical methods [5]. Analytical methods directly examine the algorithm by analyzing their properties, whereas empirical methods evaluate the result of the segmentations on given data sets [6]. The empirical approaches can be split up into two main categories: supervised and unsupervised evaluation. The former is based on desirable properties of well segmented images, according to the human visual interpretation while the latter requires a segmentation of reference or a priori knowledge (e.g. number of objects, shape, reference colors, etc.) [6].

Liu and Yang [12] have suggested an unsupervised quality measure based on the homogeneity of region color and limited region size and number. This measure tends to evaluate very noisy segmentations favorably when the average color error of small regions is close to zero. Borssoti et al. [1] have improve this measure in order to penalize the numerous small regions. This two functions have the problem of reaching a minimum value when the only segmented region is the entire image and favor segmentations with a limited number of regions. In [6] the authors propose an improvement to these two measures.

In [17], region-based evaluation methods such as the Hamming Distance, Local Consistency Error (LCE), Bidirectional Consistency Error (BCE) are reviewed. They also revisit boundary-based evaluation methods, such as the Distance Distribution Signatures, Precision-Recall measures and the Earth Mover's distance. This techniques require *apriori* knowledge of the image in order to evaluate the results of the segmentation process.

The contributions of this work are a new measure to evaluate the segmentation levels of the pyramid based on persistence of relevant edges, in order to obtain the levels that "better" depict objects and object's parts, and an algorithm to scan the irregular pyramid combining the best segments in order to build a new graceful image in a perceptive manner.

Section 2 of this paper gives a brief description of the irregular pyramid approach. In Section 3 we describe the proposed method to perform the evaluation of the segmentation levels, and in Section

4 we introduce the algorithm to combine vertices from different levels in order to build a new level. Experimental results are presented in Section 5.

2 Irregular Pyramids Overview

Irregular graph pyramids are formed by a region adjacency graph (RAG) per level. In these graphs $G = (V,E)$ the vertices (V) represent the cells or regions, and the edges (E) represent the neighborhood relations of the regions. The graph content is stored in attributes attached to both vertices and edges (i.e. color, size, gray values of the pixels, a weight measuring the difference between the two end points). The irregular graph pyramid is then a stack of successively reduced graphs (being the base level the high resolution input image). Each graph is built from the graph below by selecting a set of vertices named surviving vertices and mapping each non surviving vertex to a surviving one. Therefore each non-surviving vertex is the child of a surviving one which represents all the non surviving vertices mapped to it and becomes their father [9]. In Figure 1 some of these concepts are illustrated. For further details refer to [10].

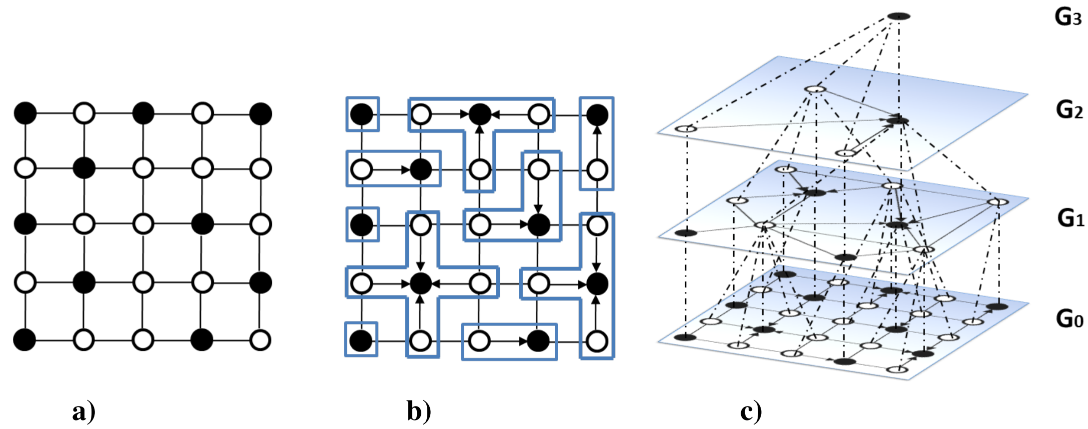

Figure 1: Construction of the irregular pyramid. a) Set of surviving vertices, depicted in black, b) contraction kernels for this set and c) irregular pyramid built using the contraction kernels from b).

Using simple graphs (graphs without multiple edges and self-loops) as the levels of the pyramid, the encoding of the spatial structure of the image might not be accurate. The lack of self-loops does not allow to differentiate inclusion from adjacency relationship. The lack of parallel edges prevent from having information regarding multiple common boundaries between two adjacent regions.

To overcome these problems, the dual graph pyramids are introduced. In order to correctly represent the embedding of the graph in the image plane, the dual graph $\overline{G} = (\overline{V}, \overline{E})$ of the RAG is additionally stored at each level. The RAG is also replaced by a RAG+ (enhanced region adjacency graph), which is a RAG that includes non-redundant self-loops or parallel edges [9].

Within the dual graph pyramid framework the reduction process is performed by a set of edge contractions. The edge contraction collapses two adjacent vertices into one vertex and removes the edge. This set is called a Contraction Kernel (CK) [3] (See Figure 1b). A CK is defined on a graph $G = (V,E)$ by a set of surviving vertices S and a set of non surviving edges N such that [9]:

- (V,N) is a spanning forest of G
- Each tree of (V,N) is rooted by a vertex of S

The contraction of the graph reduces the number of vertices while maintaining the connections to other vertices. As a consequence, the decimation of a graph by a CK may induce the creation of some redundant edges. The contraction process must follow two steps [9]:

1. A set of edge contractions on G_K encoded by the CK(S,N) . The dual of the contracted graph G_{K+1} is computed from $\overline{G_K}$ by removing the dual of the edges contained in N
2. The removal of redundant edges encoded by a CK applied on the dual graph. The edge contractions performed in the dual graph has to be followed by edge removals in the initial one in order to preserve the duality between the reduced graphs.

Combinatorial pyramids [3] are introduced in order to properly characterize the inclusion relationship, since using graphs it is not possible to know which region is inside and which one is outside just by having a self-loop. In this case, the edge's orientation around a vertex is needed. A Combinatorial Map (CM) may be understood as a planar graph encoding explicitly the orientation of edges called darts, each dart having its origin at the vertex it is attached to. A CM can be defined as $G = (D, \sigma, \alpha)$, where D is a set of darts (an edge connecting two vertices is composed of two darts $d1$ and $d2$, each dart belonging to only one vertex), α is the reverse permutation which maps $d1$ to $d2$ and $d2$ to $d1$ and σ is the successor permutation which encodes the sequence of darts encountered when turning around a vertex [3]. The dual of a CM is defined by $G = (D, \varphi, \alpha)$ with $\varphi = \sigma \circ \alpha$. The cycles of the permutation φ encode the set of darts encountered when turning around a face of G [3].

A combinatorial pyramid is then a stack of successively reduced combinatorial maps, having the advantages that each CM explicitly encode the orientation of darts around each vertex and the dual is defined on the same set of darts by the permutations φ, therefore, only one data structure has to be encoded and maintained along the pyramid [4].

3 Evaluating the Irregular Pyramid Levels of Segmentation

We use the combinatorial pyramid framework (COMA) [8] to obtain a hierarchy of image partitions, as depicted in Figure 2. In this representation, each level is a combinatorial map [3] and connections between levels are kept among the surviving vertices of each upper level and the vertices that were merged into the surviving one in the level below. This connections can be used to traverse the hierarchy upwards and downwards.

Using the entire pyramid of partitions for other tasks such as object recognition or image retrieval can be very time-consuming. Having this hierarchy of partitions, it is easy for a person to manually select one level or several levels that one might find ”better” segmented according to some criteria. Problems emerge when we want to do the same thing automatically. In this case we confront issues like which level(s) of the pyramid we should use or if we were to select one level, can we be sure that the main object parts are well represented in such partition?

For this reason, we decided to evaluate the levels of the pyramid in order to decide which levels are the best ones (according to the measure defined). We believe that the image edges can be an important criteria to evaluate segmentations results. When a partition does not preserve all relevant edges in the image, it usually means that several regions from different objects or background were merged into one single region, thus loosing very useful information. Moreover, even a partition that segments the object as a whole silhouette may not be the best one, since we are more interested in finding object's parts and its relations, in order to provide discriminative information to the object recognition algorithm.

Since we do not have *apriori* knowledge regarding the image we chose a Canny filter to determine relevant edges, and to use the resulting edges mask as reference to evaluate the segmentation at each

Figure 2: Example segmentation hierarchy obtained using combinatorial pyramids.

level. The Canny edge detector presupposes a notion of continuity by using thresholding with hysteresis for the detection. Before applying the Canny detector, the images are smoothed to reduce the influence of noise. For evaluating each partition of the pyramid we propose the following measures:

$$B_G = \frac{|P \cap R|}{|R|} \tag{1}$$

$$B_B = 1 - \frac{|P \backslash R|}{n} \tag{2}$$

where P is the set of all edge pixels from the partition being evaluated, R is the set of edge pixels in the Canny mask image and n is the total amount of pixels in the image. $|.|$ is the cardinality of set. Measure 1 evaluates how well the partition edges matched those of the Canny mask, and measure 2 evaluates how many border pixels in the partition are not present in the Canny mask. Thus, measure 1 tends to favor over-segmented partitions while measure 2 does the opposite, and penalizes partitions with more edges than those present in the mask, so these measures are combined into a global measure B using two weights W_1 and W_2.

$$B = W_1 * B_G + W_2 * B_B \tag{3}$$

Some sample results of the level evaluation using the B measure can be seen in Figure 3. In this example, the 9th level of the hierarchy obtained the best evaluation. We can see in level 10 that some edges of Lena's face were lost, thus mixing a portion of the face with a background region. Also some edges of the background objects were not kept in this partition. This is why the B value starts to decrease from level 10 onwards.

4 Improving the Segmentation

Beyond evaluating the partition levels, we are proposing to build a new partition, that will be a combination of regions belonging to different levels. This new partition must improve the result of the measure previously proposed.

Figure 3: Example level evaluation of the segmentation hierarchy.

At each level of the irregular pyramid, and for each vertex, a connection is kept to all the vertices in the level below that were merged into it. This is the vertex's CK. Using this information, we can use the best level evaluated by B and for each region, it is possible to search in the upper levels if it was combined with other regions into a bigger one that improves the result of B. Analogously, we can search in the levels below if it is possible to decompose a region into several regions that improve the result of B. This idea is based on the possibility that the best level evaluated by B may contain regions that are still under-segmented fragments of an object (that is better described in a region of an upper level) or it may contain regions that lost edges in the process of merging vertices in levels below, and we should retrieve these lost edges by fragmenting the region again. The selection of the vertices to be combined can be seen in Alg. 1

The process of finding the corresponding vertices in upper/lower levels with the current vertex being analyzed is done by recursively traversing the hierarchy using the connections between each vertex and its CK in the level below. The edges of the new graph can be updated by performing the same operations of contraction and removal defined for the irregular pyramids, but only in the neighborhood of the updated vertices.

We can see in Figure 4 the results of applying this method. It is important to notice that the best segmentation level found for the example images shows the background less segmented than original levels, and some details and prominent edges of objects are recovered.

5 Experimental Results

We ran an experiment using the ETH-80 Image Set database [11] and the B criteria to test if the improved level constructed based on the Canny edge filter mask obtains better score than the rest of original levels in the pyramid. We used the segmentation masks provided by this database as ground truth for the evaluation. For *apples*, in the 98% images, the level constructed outperformed the score of the regular levels when they were compared with the segmentation masks. The percentages for the other categories can be seen in Table 1 and the overall percentage score was 89.8%. It is important to notice in the *cars* category, which obtained the lowest result, that the segmentation mask provided by the database segments the car as a whole, while we are aiming to recover details of the car. Examples of segmented

input : Best level L_K evaluated by B
output: Graph formed by the combination of vertices from different levels of the pyramid
Search upwards;
foreach *vertex* v_K *in* L_K **do**
    find in upper levels the vertex v_{K+n} that merged v_K with its neighborhood $N(v_K)$;
    compute B' for level L_K after replacing v_K and $N(v_K)$ with v_{K+n};
    if $B' > B$ **then**
        the vertex v_{K+n} is kept in level L_K;
        v_K and $N(v_K)$ are removed from L_K;
end
Search downwards;
foreach *vertex* v_K *in* L_K **do**
    find in lower levels the set of vertices $R_{K-m}(v_K)$ that was merged into v_K;
    compute B' for level L_K after replacing v_K with $R_{K-m}(v_K)$;
    if $B' > B$ **then**
        the set of vertices $R_{K-m}(v_K)$ is kept in level L_K;
        v_K is removed from L_K;
end

Algorithm 1: Combining pyramid levels

images from this database can be seen in rows 2 and 3 of Figure 4.

Table 1: Results from the experiment with the ETH-80 Image Set database.

apples	cars	cows	cups	dogs	horses	pears	tomatoes
98 %	**68 %**	**90 %**	**86 %**	**95 %**	**92 %**	**96 %**	**90 %**

We also performed an experiment using the BSDB [15] which evaluates the performance of segmentation techniques based on the comparison of machine detected boundaries with respect to human-marked boundaries using the Precision-Recall framework [16]. Precision is a measure of how much noise is in the output of the detector. Recall is a measure of how much of the ground truth is detected. These two measures are combined into the F-measure, which is the harmonic mean of precision and recall. This benchmark consists of all of the grayscale and color hand-labeled segmentations for 300 images, performed by 30 human subjects. The images are divided into a training set of 200 images, and a test set of 100 images. They present mostly natural scenarios, animals, persons, human-made objects and buildings. The pictures may present one single object or several objects interacting in the same scene.

We built the segmentation hierarchy for the 100 color images in the test set, and computed the new level by combining vertices from different levels. Some examples are presented in the rows 4, 5 and 6 of Figure 4. The pyramids for these images have an average height of 20 levels. We submitted each level obtained for the 100 images to the evaluation process and the results obtained with the F-measure shows that the new constructed level has a higher score than those of the regular levels of the pyramid. This can be seen in Figure 5.

a) b) c) d) e) f)

Figure 4: Examples of the level combination result: a) Original image, b) Image mask resulting from the Canny edge filter, c) - e) The best three levels found according to the *B* measure, f) Improved level found by combining regions of different levels

In the BSDB web page[1], they show a comparison between several segmentation methods using this benchmark. The results regarding the F-measure for color images range between 0.43 and 0.70. The result obtained by the proposed approach (F = 0.49) is not among the best ones. Although we obtained high recall values, the precision had low values, which means that we matched well the ground truth edges, but still have undesired over-segmentation in the new level. It also may have something to do with the fact that the precision/recall measures are not tolerant to refinement, thus it is possible for two segmentations that are mutual refinements of each other to have very low precision and recall scores [17]. Since our approach intends to segment object parts, and usually humans segments objects as a whole silhouette, our result will most likely be a refinement of the segmentation done by humans.

[1]`http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/`

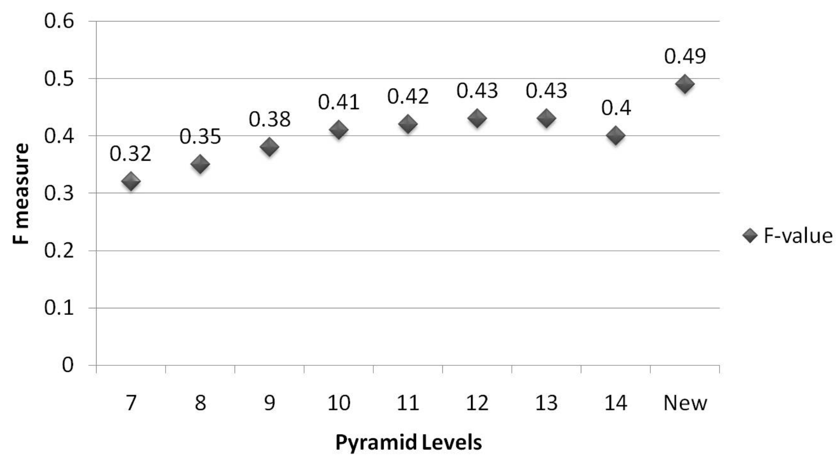

Figure 5: Comparison between regular levels of the pyramid and the new level constructed.

6 Conclusions and Future Work

In this work we presented a new method to evaluate the image segmentation levels obtained using irregular pyramids, since in many applications, it is desirable to work with the partitions which better represent the objects of the image, and not with the entire pyramid. This is the case of the object recognition task, where the matching between two entire pyramids has a high computational cost. Having this in mind, we proposed a measure that evaluates the segmentations based on its automatically detected edges and we also presented an approach that creates a new segmentation, which improves the initial pyramid partitions by using this criteria as reference.

In general, the new segmentation obtained combining different levels from the irregular pyramid is more graceful in a perceptive manner, preserving relevant edges and subregions. Of course, the results are dependent on the edges detected originally and the results showed that we still have over-segmentation in the new level. In future work, we plan to introduce some measure from the gestalt principles in order to improve the results. Furthermore, we will study the weights W_1 and W_2 in order to find a good trade off between under and over segmentation.

References

[1] M. Borsotti, P. Campadelli, and R. Schettini. Quantitative evaluation of color image segmentation results. *Pattern Recogn. Lett.*, 19(8):741–747, 1998.

[2] L. Brun and W. Kropatsch. Introduction to combinatorial pyramids. pages 108–128, 2001.

[3] L. Brun and W. Kropatsch. Contains and inside relationships within combinatorial pyramids. *Pattern Recogn.*, 39(4):515–526, 2006.

[4] Luc Brun and Walter Kropatsch. Contraction kernels and combinatorial maps. *Pattern Recogn. Lett.*, 24(8):1051–1057, 2003.

[5] D. P. Dogra, A. K. Majumdar, and S. Sural. Evaluation of segmentation techniques using region size and boundary information. In *PReMI '09: Proceedings of the 3rd International Conference on Pattern Recognition and Machine Intelligence*, pages 285–290, Berlin, Heidelberg, 2009. Springer-Verlag.

[6] Estelle Glory, Vannary Meas-Yedid, Christian Pinset, Jean-Christophe Olivo-Marin, and Georges Stamon. A quantitative criterion to evaluate color segmentations application to cytological images. In *ACIVS*, pages 227–234, 2005.

[7] Yll Haxhimusa, Adrian Ion, and Walter G. Kropatsch. Comparing hierarchies of segmentations: Humans, normalized cut, and minimum spanning tree. In Frank Lenzen, Otmar Scherzer, and Markus Vincze, editors, *Proceedings of 30th OEAGM Workshop*, pages 95–103, Obergurgl, Austria, 2006. OCG.

[8] T. Illetschko, A. Ion, Y. Haxhimusa, and W. G. Kropatsch. Effective programming of combinatorial maps using coma - a c++ framework for combinatorial maps. Technical Report PRIP-TR-106, PRIP, TU Wien, 2006.

[9] Walter G. Kropatsch, Yll Haxhimusa, and Pascal Lienhardt. *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, chapter 13. Hiearchies relating Topology and Geometry. Lecture Notes in Computer Science. Springer, Berlin Heidelberg, Dagstuhl, September 2004.

[10] Walter G. Kropatsch, Yll Haxhimusa, Zygmunt Pizlo, and Georg Langs. Vision pyramids that do not grow too high. *Pattern Recognition Letters*, 26(3):319–337, 2005.

[11] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, pages 409–415, 2003.

[12] J. Liu and Y.H. Yang. Multiresolution color image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:689–700, 1994.

[13] R. Marfil, L. Molina-Tanco, A. Bandera, J. A. Rodríguez, and F. Sandoval. Pyramid segmentation algorithms revisited. *Pattern Recogn.*, 39(8):1430–1451, 2006.

[14] R. Marfil and F. Sandoval. Energy-based perceptual segmentation using an irregular pyramid. In *IWANN '09: Proceedings of the 10th International Work-Conference on Artificial Neural Networks*, pages 424–431, Berlin, Heidelberg, 2009. Springer-Verlag.

[15] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[16] David R. Martin, Charless C. Fowlkes, and Jitendra Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5):530–549, 2004.

[17] Fernando C. Monteiro and Aurélio C. Campilho. Performance evaluation of image segmentation. In *ICIAR (1)*, pages 248–259, 2006.

[18] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2000.