Hierarchical Markov Random Fields with Irregular Pyramids for Improving Image Annotation

Annette Morales-González[1], Edel García-Reyes[1], and Luis Enrique Sucar[2]

[1] Advanced Technologies Application Center. 7a # 21812 b/ 218 and 222, Rpto. Siboney, Playa, P.C. 12200, La Habana, Cuba.
{amorales,egarcia}@cenatav.co.cu
[2] Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico
esucar@ccc.inaoep.mx

Abstract. Image segmentation and Automatic Image Annotation (AIA) are two important areas that still impose challenging problems. Addressing both problems simultaneously may improve their results since they are interdependent. In this paper we give a step ahead in that direction considering different segmentation levels simultaneously and possible contextual relations among segments in order to improve the automatic image annotation. We propose to include hierarchical relations among regions of an image in a Markov Random Field (MRF) model for annotation. This relations are obtained from irregular pyramids, which keep parent-child relations among regions through all the levels. Our main contribution is therefore the combination of the irregular pyramid approach with context modeling by means of hierarchical MRFs. Experiments run in a subset of the Corel image collection showed a relevant improvement in the annotation accuracy.

Keywords: automatic image annotation, Markov random fields, irregular pyramids

1 Introduction

Image segmentation and Automatic Image Annotation (AIA) are very important research areas in computer vision due to their relevance for many applications, such as image retrieval and scene understanding. However, the problems that arise in both fields are frequently addressed independently, disregarding the relation between them. Both fields suffer from the well-known semantic gap between low level image features and high level concepts, which is still a struggling point for researchers world-wide. Addressing both problems simultaneously may help closing the semantic gap and solving both more effectively.

Several approaches have been proposed in order to reduce the semantic gap. Probabilistic graphical models are a promising alternative used with the purpose of modeling in a more realistic fashion the context-dependent relations among

data. In particular, Markov Random Fields (MRF) [1] are employed in computer vision due to the possibility of modeling spatial neighborhood relations in images.

MRFs have been used for AIA in several works. In [2], a multiple MRF is proposed, where, instead of building a single MRF, they construct one MRF for each keyword in the vocabulary to capture different semantics among keywords. The proposal of [3] explores dependencies among features and several words. In [4] and [5], the co-occurrence information among words and the probabilities of occurrence of spatial relations between pairs of words respectively are used as interaction potential in an MRF model. While these approaches have focused on exploring dependencies among features and words, and between neighboring pairs of words, we propose to explore hierarchical relations among different segmentation levels of an image. Using hierarchies with MRF models is not a new idea. In [6] they propose a segmentation method using two levels of a hierarchy, where the region classification is performed independently at each level, and later combined within the MRF model. It has been used also for texture segmentation and denoising [7][8], where the hierarchy consists on two or three layers representing different characteristics of the image. The main difference with our proposal is that we will use a hierarchy of image partitions (represented as graphs) at different levels of resolution, and we will construct a MRF at each level that will be fed with the MRF information computed in adjacent levels, for the purpose of improving image annotation.

To obtain a hierarchical representation of images, we use irregular pyramids [9], which are hierarchical structures formed by combinatorial maps. The combinatorial map at each level of the pyramid is equivalent to a Region Adjacency Graph (RAG) [10] relative to a partition of the image. Irregular pyramids provide hierarchical relations among regions found at different levels, and topological relations among regions of the same level.

Our main contribution is therefore the combination of the irregular pyramid segmentation approach with context modeling for improving automatic image annotation. This involves using an additional potential in a MRF model, taking into account the hierarchical information among regions at different levels of the pyramid. A MRF will be modeled for each level of the pyramid, taking into account the initial labels (annotated with a base classifier) and contextual (hierarchical and spatial) relations among image regions. The best configuration of labels for each level will be computed in a bottom-top (taking information from lower levels) and top-bottom (taking information from lower and higher levels at the same time) processes, thus refining the initial annotation. Experiments performed in a subset of the Corel image collection showed that the proposal (called HMRF-Pyr) can clearly improve the annotation results and revealed the advantages of using hierarchical relations.

The remainder of this paper is as follows. Section 2 provides an overview of irregular pyramids. In Section 3 we present basic concepts regarding MRFs. The proposed approach is presented in Section 4, followed by the experimental evaluation in Section 5. Finally we present the conclusions and future work.

2 Irregular Pyramids Overview

An irregular graph pyramid is a stack of successively reduced graphs (being the base level the high resolution input image). In these graphs $G = (V, E)$ the vertices (V) represent cells or regions, and the edges (E) represent neighborhood relations of the regions. When we build an irregular pyramid from an image, each level represents a partition of the pixel set into cells, i.e. connected subsets of pixels. On the base level (level 0) of the pyramid, the cells represent single pixels and the neighborhood of the cells is defined by the 4-connectivity of pixels. A cell on level k (parent) is a union of neighboring cells on level $k-1$ (children) [11]. Each graph is built from the graph below by selecting a set of surviving vertices and mapping each non surviving vertex to a surviving one. In this way, each surviving vertex represents all the non surviving vertices mapped to it and becomes their father [11]. This parent-child relations may be iterated down to the base level and the set of descendants of one vertex in the base level is named its receptive field (RF). Some of these concepts are illustrated in Figure 1.

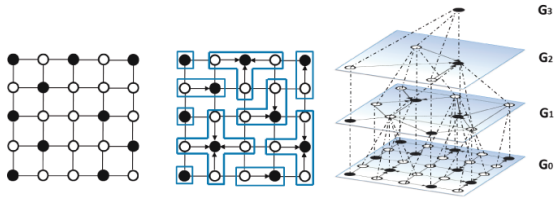

Fig. 1. Construction of the irregular pyramid. (a) Set of surviving vertices, depicted in black, (b) contraction kernels for this set and (c) irregular pyramid built using the contraction kernels from b

Within the irregular 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) [10]. The contraction of the graph reduces the number of vertices while maintaining the connections to other vertices.

This pyramid is able to represent several topological relations between regions. Besides the classical adjacency relationship encoded by the Region Adjacency Graph (RAG), each graph may also contain parallel edges and self-loops, representing several common boundaries and inclusion relations respectively [10].

Combinatorial pyramids [10] are introduced in order to properly characterize the inclusion relationship, which cannot be fully represented using graphs. In this case, the edges 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 moving around a vertex [10]. A combinatorial pyramid is then a stack of successively reduced combinatorial maps, having the advantages that each CM explicitly encodes the orientation of darts around each vertex

3 Markov Random Fields

Intuitively, Markov Random Fields [1] are undirected graphical models that combine information from a set of observations, and interaction information obtained by the relation with neighbors. Formally, we can say that $Y = \{Y_1, Y_2, ..., Y_n\}$ is called a random field, being Y_i random variables on a set of sites S, that can take values y_i in a set of labels L. This can be depicted as an undirected graph, where each vertex i represents the random variable Y_i and the edges represent direct dependence relations between variables. Henceforth, the terms "vertex" and "variable" shall be used interchangeably.

A Markov Random Field is a random field that obeys the Markov property $P(y_i|y_{i-1}, y_{i-2}, ..., y_1) = P(y_i|N(y_i))$, where $N(y_i)$ is the set of neighbors of y_i. This means that given its neighbor set $N(y_i)$, a vertex i is independent of all other vertices in the graph. The most probable configuration of labels Y^* for a MRF is the one that maximizes the joint probability $P(y)$. This joint probability is modeled by some restrictions represented by local probabilities, also known as potentials. The potentials can be interpreted as constrains that penalizes or favors certain configurations of Y. The joint probability is expressed as Eq. 1

$$P(y) = \frac{1}{Z} * \exp^{-U_p(y)} \tag{1}$$

where Z is the partition function or normalizing constant and $U_p(y)$ is the energy function. $U_p(y)$ is computed using the aforementioned potentials (Eq. 2).

$$U_p(y) = V_O(y) + \lambda \sum_I V_I(y, y') \tag{2}$$

$V_O(y)$ stands for the association (or unary) potential, which represents information coming from the observations. $V_I(y, y')$ is the interaction (or pairwise) potential and models the information obtained from neighboring vertices (y, y'). λ is a constant introduced to weight the relevance of the restrictions imposed by the potential functions. The Maximum A Posteriori (MAP) optimal configuration Y^* is obtained by minimizing the value of $U_p(y)$. Common methods for achieving this optimal configuration are the Iterated Conditional Modes (ICM), Simulated Annealing and Loopy Belief Propagation (LBP), among others [12].

Although higher order potentials could be used in the model (making y_i dependent on a number O of variables), we prefer to use only unary and pairwise potentials, since higher order potentials largely increase the computational cost for finding the optimal configuration.

4 Proposed Approach

The definition of MRFs (see Section 3) and most of its applications in images, deal with two basic relations: the relation of a region feature (observation) with a label, and the relation of neighboring labels. We are proposing to include in this framework a parent-child relationship, driven by the notion that in a hierarchical representation, the children regions may have a trustworthy vote regarding its parent's classification and viceversa.

We propose to build a MRF for each level of the pyramid, starting from bottom to top, and at each level l, the information regarding the best configuration of labels ${Y^*}_{l-1}$ obtained in level $l-1$ is used as additional information to compute the current level's label configuration ${Y^*}_l$. When the top level is reached, the same process is repeated from top to bottom, now using ${Y^*}_{l-1}$ and ${Y^*}_{l+1}$ to compute ${Y^*}_l$. The MRF for each level will have the same structure of the underlying RAG in the irregular pyramid.

The Markovian neighborhood $N(y_i^l)$ of label y_i^l can be split into two neighborhoods: the spatial neighborhood and the hierarchical neighborhood. The spatial neighborhood of label y_i^l corresponding to vertex i, is composed by the labels assigned to all the vertices adjacent to i in the RAG of level l. The hierarchical neighborhood is formed by all the labels assigned to the Contraction Kernel of vertex i in level $l-1$ and by its parent's label in level $l+1$.

We propose to compute the energy function as depicted in Eq. 3.

$$U_p(y_l) = \lambda_O V_O(y_i^l) + \lambda_I \sum_i V_I(y_i^l, y_j^l) + \lambda_H \left(\sum_{ch} V_{Ch}(y_i^l, y_k^{l-1}) + V_P(y_i^l, y_m^{l+1}) \right) \tag{3}$$

This is an extension of Eq. 2, introducing $V_{Ch}(y_i^l, y_k^{l-1})$ as a hierarchical potential that models the relation of label y_i^l (assigned to vertex i of level l of the pyramid) with its child label y_k^{l-1}, and $V_P(y_i^l, y_m^{l+1})$ that models the relation of y_i^l with its parent label y_m^{l+1}. The label y_k^{l-1} was assigned to vertex k in level $l-1$ of the image pyramid by the MRF computed for this level. Vertex k belongs to the CK (See Section 2) of vertex i ($k \in CK(i)$). The observation, interaction and hierarchical potentials are weighted by λ_O, λ_I and λ_H respectively, and $\lambda_O + \lambda_I + \lambda_H = 1$.

Having stated the energy function, we defined each potential as follows. The association potential $V_O(y_i^l)$ is defined as in [4]. We use as base annotation system the k-nearest neighbors (KNN) classifier. In order to rank candidate labels for a given region we use the distance of the test instance to the top k-nearest neighbors as relevance weight. As presented in [4], relevance weighting is obtained using Eq. 4.

$$P^R({y_i}_j^l) = \frac{d_j(x_i^l)}{\sum_{n=1}^{k} d_n(x_i^l)}, \quad y_i^l \in Y, x_i^l \in X \tag{4}$$

where $d_j(x_i^l)$ is the Euclidean distance in the attribute space of observation x_j (corresponding to the j-nearest neighbor) to x_i^l (the test instance), being Y the set of labels and X the set of observations.

The association potential is then expressed as in Eq. 5,

$$V_O(y_i^l) = \frac{1}{P^R(y_i^l)} \tag{5}$$

the interaction potential is defined in Eq. 6,

$$V_I(y_i^l, y_j^l) = \begin{cases} 0 & \text{if } y_i^l = y_j^l \\ 1 & \text{if } y_i^l \neq y_j^l \end{cases} \tag{6}$$

the potential related with the children information is defined by Eq. 7 and the potential related with the parents information is presented in Eq. 8

$$V_{Ch}(y_i^l, y_k^{l-1}) = \begin{cases} 0 \text{ if } y_i^l = y_k^{l-1} \\ 1 \text{ if } y_i^l \neq y_k^{l-1} \end{cases} \tag{7}$$

$$V_P(y_i^l, y_m^{l+1}) = \begin{cases} 0 \text{ if } y_i^l = y_m^{l+1} \\ 1 \text{ if } y_i^l \neq y_m^{l+1} \end{cases} \tag{8}$$

where $k \in CK(i)$ and $i \in CK(m)$.

The interaction potential penalizes neighbors with different labels with respect to the current vertex, while hierarchical potentials punish children or parents having different labels than the current vertex. In order to obtain the optimal configuration Y^*, we used the ICM algorithm, which is efficient and, although usually criticized for converging to local minimums, for this case the results were very similar to other more complex methods. This might indicate that for the current problem, ICM is actually converging to the global minimum.

5 Experiments

In order to validate our proposal, we ran experiments on a subset of the Corel image collection. Specifically, we used the CorelA subset developed by [13]. This dataset contains 205 natural scene images split into two subsets with 137 images for training and 68 images for testing. All images have been segmented using normalized cuts [14] and they have been manually annotated with 22 classes.

The irregular pyramids computed for these images have an average of 20 levels. For these experiments we tested all the levels ranging from level 6 to 16, in order to avoid extreme oversegmentations or undersegmentations. Following the idea of [15], we used as visual features for each vertex (region) of each graph, the quantization of the RGB values in 16 bins per channel, yielding a 48-dimensional color histogram, and a local binary pattern (LBP) histogram to characterize texture in the region.

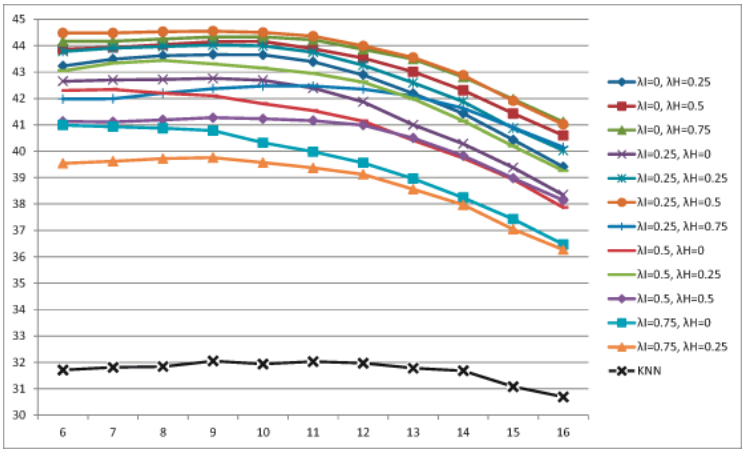

Fig. 2. Accuracy results in the CorelA dataset using different parameter combinations.

The experiments were performed using as base annotation system a KNN classifier as implemented in [4]. We used this base classifier in order to be fair in the comparison, sticking to the conditions imposed by [5]. Nevertheless, we believe that, by using a more sophisticated classifier, our results can be improved. Since we used a different segmentation approach than the one used by [5] to label image regions, we measured the accuracy of the annotation at pixel level. Results can be seen in Table 1. The first three rows show the average accuracy over 10 runs obtained using the algorithms for each level of the pyramid. We are comparing the base classifier KNN with our hierarchical MRF approach (HMRF-Pyr) and the traditional MRF approach (without hierarchical relations). In row 4 we can see the relative improvement of HMRF-Pyr with respect to the KNN base classifier and row 5 shows the relative improvement of HMRF-Pyr over MRF. We tested several combinations of λ_O, λ_I and λ_H, having better results with 0.25, 0.25 and 0.5 respectively. This results can be analyzed in Fig. 2. Horizontal axis show the pyramid levels and vertical axis depict annotation accuracy (in %). For clarity, in Fig. 3 we can see an extract of Fig. 2, only showing the best results using the hierarchical information, the best results when this information is not used ($\lambda_H = 0$), and the base classifier KNN results.

In Fig. 4, some segmentation and annotation results can be seen for images of the CorelA set. From these results we can see that the HMRF-Pyr approach involving hierarchical and neighborhood relations improved the annotation accuracy with respect to the base classifier and with respect to the traditional MRF approach. This is consistent with the initial assumption that the annotation of children regions have influence in its father classification and viceversa.

Comparing our approach with other methods that were tested on this dataset, we can see that the highest annotation accuracy obtained by [5] is of 45.64%, while our best result is 44.6% (See Table 2). These results are similar (1 point difference), however some test conditions (segmentation method and low level

Table 1. Results obtained in the CorelA subset for each level of the pyramid

	Pyramid levels										
Algorithm	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
KNN	31.7%	31.8%	31.8%	32.0%	31.9%	32.0%	31.9%	31.7%	31.6%	31.1%	30.7%
MRF	42.2%	42.3%	42.2%	42.1%	41.8%	41.6%	41.2%	40.3%	39.7%	38.9%	37.9%
HMRF-Pyr	**44.5**%	**44.5**%	**44.5**%	**44.6**%	**44.5**%	44.4%	44.0%	43.5%	42.8%	41.9%	41.0%
Imp. HMRF-Pyr/KNN	12.8%	12.7%	12.7%	12.5%	12.6%	12.3%	12.0%	11.8%	11.2%	10.8%	10.3%
Imp. HMRF-Pyr/MRF	2.2%	2.1%	2.3%	2.5%	2.7%	2.8%	2.9%	3.1%	3.1%	3.0%	3.2%

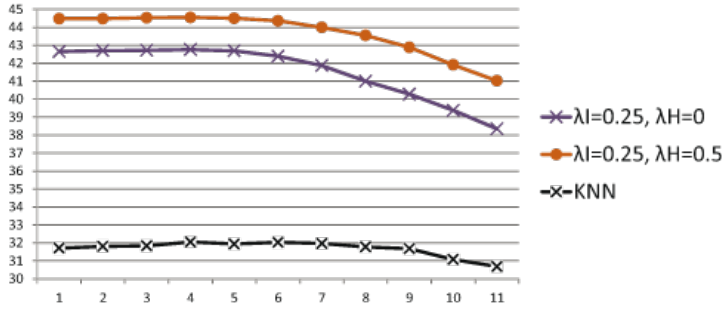

Fig. 3. Extract of Fig. 2, for clarifying the improvement obtained when using the hierarchical potential ($\lambda_H \neq 0$), with respect to not using it ($\lambda_H = 0$).

features) are different and it is more relevant the gain with respect to the base classifier (last column of Table 2). For them, KNN scored 36.8%, while we obtained 32%. The best improvement for [5] over the base classifier was 8.82%, while our relative improvement is 12.8%. In our opinion, this relative improvement shows the importance of the hierarchical information in problems with context-dependent information, and the relevance of combining segmentation and annotation simultaneously. It can be seen in Fig. 4 that segmentation can be enhanced with annotation, simply by joining regions with the same label.

6 Conclusions

In this paper we proposed an approach that combines irregular pyramid segmentation with image annotation based on Markov Random Fields. MRFs allow to take into account contextual relations when performing the annotation, and we proposed an enhance to this model by using the hierarchical relations among regions of different levels of the irregular pyramid. As experimental results showed, hierarchical information actually provides relevant information to the annotation process, which combined with neighborhood information, can represent an important improvement with respect to the base classifier.

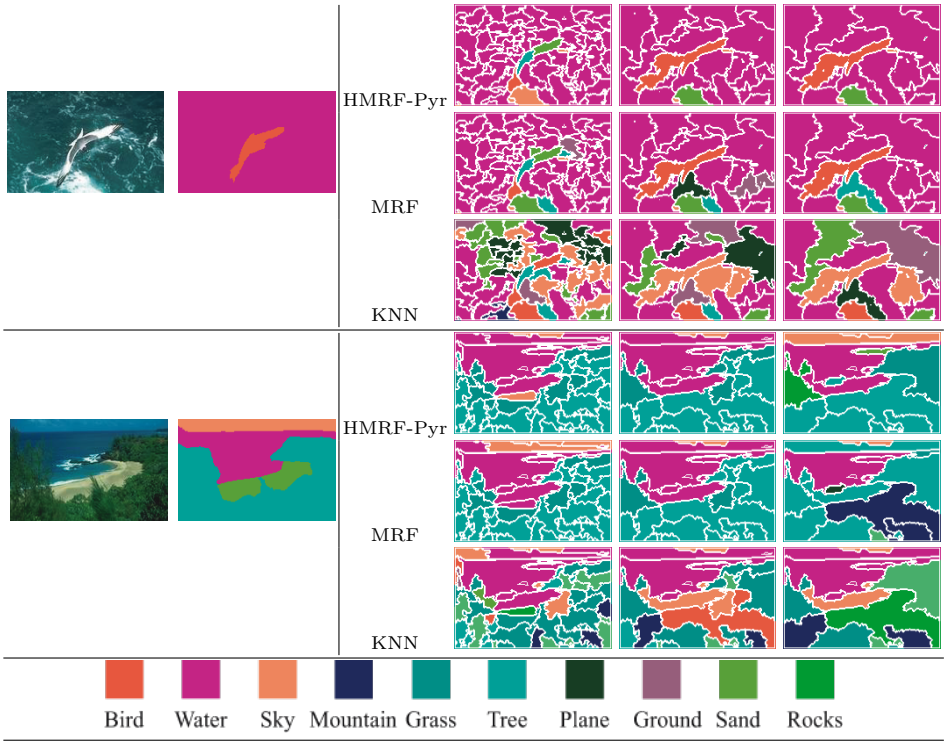

Fig. 4. Examples of Corel image annotation results. First column shows original images with its ground truth annotation mask in second column. Columns 4 to 6 show different pyramid levels. For each image, the first row of levels shows the annotation result with HMRF-Pyr, the second row shows the MRF annotation result and the third one shows the annotation result achieved with KNN. At the bottom we can find a color legend to understand the annotation results. (Best seen in color)

In future work, we have the intention of elaborating a multilevel segmentation/annotation algorithm where the structure of higher levels is modified by the lower levels information.

References

1. F. Spitzer, *Random Fields and Interacting Particle Systems: Notes on Lectures Given at the 1971 MAA Summer Seminar, Williamstown - Mass*, 1971.
2. Y. Xiang, X. Zhou, T. Chua, and C. Ngo, "A revisit of generative model for automatic image annotation using markov random fields," in *Proceedings of CVPR 2009*, 2009, pp. 1153–1160.
3. A. Llorente, R. Manmatha, and S. Rüger, "Image retrieval using markov random fields and global image features," in *Proceedings of the ACM International Conference on Image and Video Retrieval*. 2010, CIVR '10, pp. 243–250, ACM.

Table 2. Comparison with other methods in the CorelA subset. Second column shows the accuracy of each algorithm. For those who use a base classifier (KNN), the accuracy of this classifier is shown in third column. Fourth column shows the relative improvement achieved by the algorithms over the base classifier.

Algorithm	Accuracy	KNN accuracy	Imp. from KNN
gML1 [13]	35.7%	-	-
gML1o [13]	36.2%	-	-
gMAP1 [13]	35.7%	-	-
gMAP1MRF [13]	35.7%	-	-
MRFs AREK [5]	**45.6**%	**36.8**%	8.8%
HMRF-Pyr	44.6%	32.0%	**12.8**%

4. H. J. Escalante, M. Montes, and E. Sucar, "Word co-occurrence and markov random fields for improving automatic image annotation," in *Proceedings of the 18th British Machine Vision Conference (BMVC-2007)*, September 2007.
5. C. Hernández-Gracidas and L. E. Sucar, "Markov random fields and spatial information to improve automatic image annotation," in *PSIVT*, 2007, pp. 879–892.
6. M. Keuper, T. Schmidt, M. Rodriguez-Franco, W. Schamel, T. Brox, H. Burkhardt, and O. Ronneberger, "Hierarchical markov random fields for mast cell segmentation in electron microscopic recordings," in *Proceedings of the 8th IEEE International Symposium on Biomedical Imaging, ISBI 2011*, 2011, pp. 973–978.
7. D. H. Kim, I. D. Yun, and S. U. Lee, "New mrf parameter estimation technique for texture image segmentation using hierarchical gmrf model based on random spatial interaction and mean field theory," in *Proceedings of ICPR 2006 - Volume 02*, Washington, DC, USA, 2006, ICPR '06, pp. 365–368, IEEE Computer Society.
8. Y. Cao, Y. Luo, and S. Yang, "Image denoising based on hierarchical markov random field," *Pattern Recogn. Lett.*, vol. 32, no. 2, pp. 368–374, Jan. 2011.
9. L. Brun and W. Kropatsch, "Introduction to combinatorial pyramids," *Digital and image geometry: advanced lectures*, pp. 108–128, 2001.
10. L. Brun and W. Kropatsch, "Contains and inside relationships within combinatorial pyramids," *Pattern Recogn.*, vol. 39, no. 4, pp. 515–526, 2006.
11. Y. Haxhimusa and W. Kropatsch, "Segmentation Graph Hierarchies," in *Proceedings of Joint International Workshops on Structural, Syntactic, and Statistical Pattern Recognition S+SSPR 2004*, vol. LNCS 3138.
12. R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, June 2008.
13. P. Carbonetto, "Unsupervised statistical models for general object recognition," Tech. Rep., The Faculty of Graduate Studies, Department of Computer Science, The University of British Columbia, West Mall Vancouver, BC Canada, 2003.
14. J. Shi and J. Malik, "Normalized cuts and image segmentation," in *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, Washington, DC, USA, 1997, CVPR '97, pp. 731–, IEEE Computer Society.
15. A. Morales-González and E. García-Reyes, "Simple object recognition based on spatial relations and visual features represented using irregular pyramids," *Multimedia Tools and Applications*, pp. 1–23, 2011, 10.1007/s11042-011-0938-3.