An Irregular Pyramid for Multi-scale Analysis of Objects and their Parts*

Martin Drauschke

Department of Photogrammetry, Institute of Geodesy and Geoinformation
University of Bonn, Nussallee 15, 53115 Bonn, Germany
martin.drauschke@uni-bonn.de

Abstract. We present an irregular image pyramid which is derived from multi-scale analysis of segmented watershed regions. Our framework is based on the development of regions in the Gaussian scale-space, which is represented by a region hierarchy graph. Using this structure, we are able to determine geometrically precise borders of our segmented regions using a region focusing. In order to handle the complexity, we select only stable regions and regions resulting from a merging event, which enables us to keep the hierarchical structure of the regions. Using this framework, we are able to detect objects of various scales in an image. Finally, the hierarchical structure is used for describing these detected regions as aggregations of their parts. We investigate the usefulness of the regions for interpreting images showing building facades with parts like windows, balconies or entrances.

1 Introduction

The interpretation of images showing objects with a complex structure is a difficult task, especially if the object's components may repeat or vary a lot in their appearance. As far as human perception is understood today, objects are often recognized by analyzing their compositional structure, cf. [9]. Besides spatial relations between object parts, the hierarchical structure of the components is often helpful for recognizing an object or its parts. E. g. in aerial images of buildings with a resolution of 10 cm per pixel, it is easier to classify dark image parts as windows in the roof, if the building at whole has been recognized before.

Buildings are objects with parts of various scales. Depending on the view point, terrestrial or aerial, the largest visible building parts are its facade or its roof. Mid-scale entities are balconies, dormers or the building's entrance; and small-scale parts are e. g. windows and window panes as window parts. We restrict our focus on such parts, a further division down to the level of bricks or tiles is not of our interest.

Recently, many compositional models have been proposed for the recognition of natural and technical objects. E. g. in [6] a part-based recognition framework

* This work has been done within the project *Ontological scales for automated detection, efficient processing and fast visualization of landscape models* which is funded by the German Research Council (DFG).

is proposed, where the image fragments have been put in a hierarchical order to infer the category of the whole object after having classified its parts. So far, this approach has only been used for finding the category of an object, but it does not analyze the parts individually. This approach has been evaluated on blurred, downsampled building images, cf. [13]. Without resizing the image, the algorithm seems to work inefficiently or even might fail at homogeneous facades or on the repetitive patterns like bricks, because the fragments cannot get grouped together easily. Thus, the approach is not easily applicable to the domain of buildings.

Working on hyperspectral images, a hierarchical segmentation scheme for geospatial objects as buildings has been recently proposed using morphological operations, cf. [1]. Due to the low resolution of the images, the hierarchy can only be used for detecting the object of the largest scale, but not its parts separately.

We work on segmented image regions at different scales, where we derive a region hierarchy from the analysis of the regions. So far, it is purely data-driven, so that the general approach can be used in many domains. A short literature review on multi-scale image analysis is given in sec. 2. Then, we present our own multi-scale approach in sec. 3. For complexity reasons, we need to select regions from the pyramid for further processes. We document this procedure in sec. 4. The validation of our graphical representation is demonstrated in an experiment on building images in sec. 5. Concluding, we summarize our contribution in sec. 6.

2 Multi-scale Image Analysis

Although, the segmentation of images can be discussed in a very general way, we have in our mind the segmentation of images showing man-made scenes. These images usually show objects of various scales. With respect to the building domain, windows, balconies or facades can be such objects. For detecting them, the image must be analyzed at several scales. The two most convenient frameworks for multi-scale region detections are (a) segmentation in scale-space and (b) irregular pyramids. Regarding scale-space techniques, the behaviour of segmentation schemes have been studied, and the watershed segmentation is often favored, even in different domains, cf. e. g. [16], [8], [10] and [3].

We also evaluated the usability of watersheds for segmenting images of buildings. Thereby, our focus was the possibility to segment objects of different scales. In Gaussian scale-space, the smoothing with the circular filter leads to rounded edges and region borders in higher scales. We obtain similar result when using the morphological scale-space as proposed in [12]. Again, the shape of the structural element emerges disturbingly at the higher scales. In the anisotropic diffusion scheme, cf. [17], the region borders of highest contrast are preserved longest, and therefore it can not be used for modeling aggregates of building parts, where the strongest gradient appear at the border between e. g. bright window frames and dark window panes.

Pyramids are a commonly used representation for scale-space structures, cf. [14]. When working on the regular grid of image blocks, e. g. on pixel-level,

the use of a regular pyramid is supported by many advantages, e. g. access in memory, adjacencies of blocks etc. In contrast to the regular grid, the number of entities rapidly decreases when working on segmented image regions, which also decreases the complexity of many further algorithms. Furthermore, the representation of objects by (aggregated) regions is more precise in the shape of the objects boundary than using rectangular blocks.

In the last years, different pyramid frameworks have been proposed. With respect to image segmentation, we would like to point out the stochastic pyramids, cf. [15], and irregular pyramids as used in [10]. In both approaches, a hierarchy of image regions is obtained by grouping them according to certain condition, e. g. a homogeneity measure. With respect to buildings we often have the problem of finding such conditions, because we want to merge regions of similar appearance on one hand and regions rich in contrast on the other hand. Thus we decided to work on watershed regions in scale-space, and to use this scale-space structure to derive a region hierarchy that forms an irregular pyramid.

3 Construction of the Irregular Pyramid

In this section, we present our multi-scale segmentation framework and the construction of our *region hierarchy graph* (RHG). For receiving more precise region boundaries, we applied an adaptation of the approach of [8].

3.1 Multi-scale Image Segmentation

Many different segmentation algorithms were proposed since the age of digital imagery has started. We decided to derive our segmentation from the watershed boundaries on the image's gradient magnitude. Considering the segmentation of man-made objects, we mostly find strong color edges between different surfaces, and so the borders of the watershed regions are often (nearly) identical with the borders of the objects.

Our approach uses the Gaussian scale-space for obtaining regions in multiple scales. We arranged the discrete scale-space layers logarithmically between $\sigma = 1$ and $\sigma = 16$ with 10 layers in each octave, obtaining 41 layers. For each scale σ, we convolve each image channel with a Gaussian filter and obtain a three-dimensional image space for each channel. Then we compute the combined gradient magnitude of the color images. Since the watershed algorithm is inclined to produce oversegmentation, we suppress many gradient minima by resetting the gradient value at positions where the gradient is below the median of the gradient magnitude. So, those minima are removed which are mostly caused by noise. The mathematical notation of this procedure is described in more detail in [5]. As result of the watershed algorithm, we obtain a complete partitioning of the image, where every image pixel belongs to exactly one region.

3.2 Region Hierarchy Graph

The result of the scale-space watershed procedure is a set of regions R_σ^ν where ν is the index for the identifying label and σ specifies the scale. The area of a

region $|R|$ is the number of its pixels. Since the scale-space layers are ordered in a sequence, we denote neighbored scales by their indices, i. e. σ_i and σ_{i+1}. Our RHG is based on pair wise neighborhoods of scale and we define two regions R^{ν_m,σ_i} and $R^{\nu_n,\sigma_{i+1}}$ of neighbored scales as *adjacent in scale* if their overlap is maxized. Therefore, we determine the number of pixel positions which belong to both regions $|R^{\nu_m}_{\sigma_i} \cap R^{\nu_n}_{\sigma_{i+1}}|$. Concluding, adjacency in scale of two regions of neighbored scales is defined by the mapping

$$R^{\nu_m}_{\sigma_i} \mapsto R^{\nu_n}_{\sigma_{i+1}} \Leftrightarrow |R^{\nu_m}_{\sigma_i} \cap R^{\nu_n}_{\sigma_{i+1}}| > |R^{\nu_m}_{\sigma_i} \cap R^{\nu_k}_{\sigma_{i+1}}| \; \forall \, k \neq n, \tag{1}$$

which defines an ordered binary relation between region, and the mapping symbol $\mapsto$ reflects the development of a region with increasing scale. Observe, no threshold is necessary.

According to [14], there occur four events with region features in scale-space: the *merging* of two or more regions into one, and the *creation*, the *annihilation* or the *split* of a region. Our RHG reflects only two of these events, the creation and the merging. A creation-event is represented by a region of a higher layer that is no target of the mapping-relation, and a merge-event is represented, if two or more regions are mapped to the same region in the next layer. Equ. 1 avoids that a region can disappear, because we always find a region in the next layer. Furthermore, our mapping-relation avoids the occurrence of the split-event, because we always look for the (unique) maximum overlap. Our definition of the region hierarchy leads to a simple RHG, which only consists of trees, where each node (except in the highest scale) has exactly one leaving edge.

Note that the relation defined in equ. 1 is asymmetric. When expressing region adjacency with decreasing scale, we take the inverted edges from the RHG. Moreover, the relation is not transitive. Thus, the RHG may contain paths to different regions, if a scale-space layer has been skipped when constructing the RHG. We show a scale-space with three layers and the corresponding RHG in fig. 1.

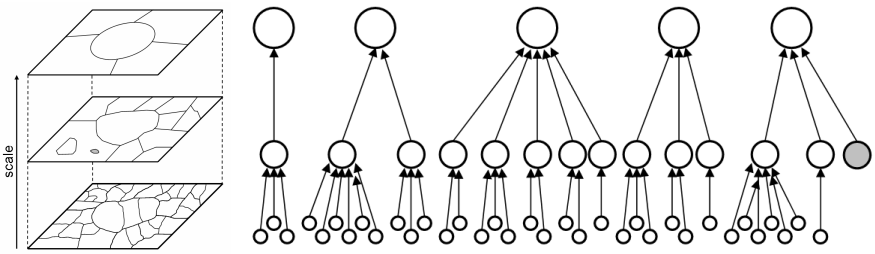

Fig. 1. Segmentation in scale-space and its RHG. Regions from the same scale are ordered horizontally, and the increasing scales are ordered vertically from bottom to top. The edges between the nodes describe the development of the regions over scale. The gray-filled region has been created in the second layer.

3.3 Region Focusing

The Gaussian smoothing leads to blurred edges at larger scales, and corners become rounder and rounder. Therefore, we perform an additional region focusing, which is inspired by [18] and [2]. In [2], the existence of an edge has been recognized in a large scale, but its specific geometric appearance was derived by tracking it to the lowest available scale.

We improve the geometrical precision of our segmented regions by combining information from the RHG with the initial image partition, i. e. the segmentation at the lowest scale $\sigma = 1$. Taking the forest as a directed graph with arcs from higher scale to lower scales we obtain the focused region at a level below a given regions as the union of all regions reachable from the source region. Reaching the initial image partition, we obtain regions $\widehat{R}_{\sigma_i}^{\nu_n}$ by merging all respective regions:

$$\widehat{R}_{\sigma_i}^{\nu_n} = \cup_k R_{\sigma=1}^{\nu_k} \text{ with } \exists \text{ a path from } R_{\sigma=1}^{\nu_k} \text{ to } R_{\sigma_i}^{\nu_n}. \tag{2}$$

In fact, our approach is an adaptation of the segmentation approach in [8]. There, a similar merging strategy for watershed regions has been proposed, where the regions were merged on the basis of their tracked seed points, thus bottom up, whereas our approach ist top down. The procedure in [8] is not suitable to our segmentation scheme, because we have suppressed all minima in the gradient image which are below its median, so we might analyze the development of a huge number of seed point for a single region. Furthermore, our approach with looking for the maximum overlap is also applicable, if a different segmentation is used than watershed regions. We visualize a result of our region focusing in comparison with the original image partition in fig. 2.

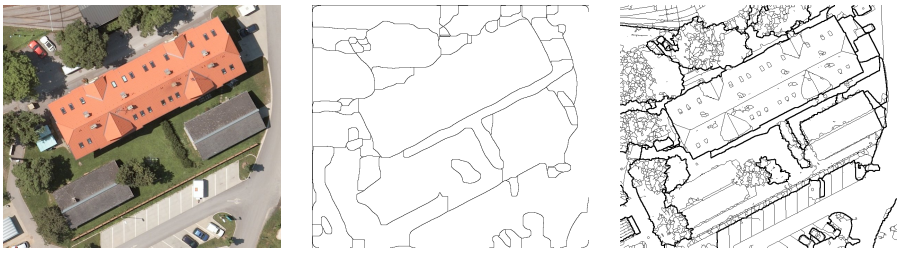

Fig. 2. Image segmentation of an aerial image. Left: RGB image of a suburban scene in Graz, Austria (provided by Vexcel Imaging GmbH). Middle: Original watershed regions in scale $\sigma = 35$. Right: Region focusing with merged regions of scales $\sigma = 12$ (thin) and $\sigma = 35$ (thick). Clearly, both segmentations of scale $\sigma = 35$ are not topologically equivalent, because the newly created or split regions (and their borders) cannot get tracked down to the initial partition by our region focusing.

Since we use the RHG for performing the region focusing, the RHG nearly remains unchanged. We only delete all newly created regions from all scale-space layers above the initial partition. Hence, the respective nodes and edges must be removed from the RHG. Furthermore, all regions must be removed which only

develop from these newly created regions. The updated RHG of the example in fig. 1 will contain all white nodes and the their connecting edges.

4 Selection of Regions from Irregular Pyramid

Up to this point, we only described the construction of our irregular image pyramid, but we have not mentioned its complexity. On relatively small images with a size of about 400×600 pixels, the ground layer of our irregular pyramid often contains 1500 or more regions, and their number decrease down to 10 to 30 in the highest layer. Assuming that the number of regions in a layer decreases with a constant velocity, the complete pyramid contains over 30.000 regions. Since most of these regions do not represent objects of interest, a selection of regions seems to be helpful to reduce the complexity of further processes.

The integration of knowledge about the scene could later be done in this step, e. g. one could choose regions with a major axis which leads in direction of the most dominant vanishing points, or one could choose regions which represent a repetitive pattern in the image, so that it might correspond to a window in the image. But nevertheless, the search for such reasonable regions in the whole pyramid is still a task with a very high complexity.

We have tested our algorithms by segmenting images showing man-made scenes, preferably buildings. These objects mostly have clearly visible borders, so that the according edges can be detected in several layers of the pyramids. Therefore, we focus on stable regions in our irregular pyramid and defined a stability measure $\varsigma_{m,i}$ for a region $R_{\sigma_i}^{\nu_m}$ to the adjacent region in the next scale-space level $i+1$ by

$$\varsigma_{m,i} = \frac{R_{\sigma_i}^{\nu_m} \cap R_{\sigma_{i+1}}^{\nu_n}}{R_{\sigma_i}^{\nu_m} \cup R_{\sigma_{i+1}}^{\nu_n}}, \tag{3}$$

where region $R_{\sigma_{i+1}}^{\nu_n}$ is adjacent in scale to $R_{\sigma_i}^{\nu_m}$ and, therefore, both regions are connected by an edge in the RHG. Then we define the stability measure $\bar{\varsigma}$ of a scale range with d scale-space levels by

$$\overline{\varsigma_{m,i}} = \max_{k=0..d} \left\{ \min_{j=i-d+k..i+k} \varsigma_{m',j} \right\}, \tag{4}$$

where m' corresponds to the region of layer j that is connected to $R_{\sigma_i}^{\nu_m}$ by a path. We call all regions with $\overline{\varsigma_{m,i}} > t$ stable, where t is a threshold, e. g. $t = 0.75$.

If we find a stable region in our pyramid, than we will find at least $d-1$ additional regions with a similar shape. All these regions can be represented by the same region. This is the first step, when we reduce our pyramid. The stable regions are not necessarily adjacent in scale to other stable regions. In fact, this happens seldom. We are able to keep the information of the RHG, if we arrange the stable regions in a hierarchical order and include the merging events, where paths from two or more previously stable regions reach the same region of the pyramid. Due to the limited space, we cannot go more into detail here, we present a sketch of our method in fig. 3. Its result is a *tree of stable regions* (TSR), where we inserted an additional root-node for describing the complete scene.

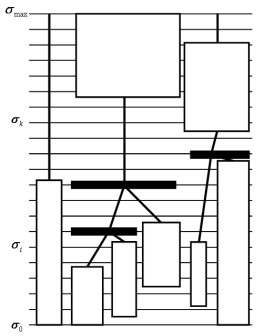

Fig. 3. Tree of stable regions: the layers of the pyramid are arranged in a vertical order (going upwards), each rectangle represents a node in the TSR, the white ones correspond to stable regions, the black ones the merging events from the RHG. The horizontal extensions of the rectangles show their spatial state, and the vertical extension corresponds with the range of stability. The idea of this figure is taken from the interval tree and its representation as rectangular tessellation in [18].

5 Experiments

Our approach is very general, because we used only two assumptions for generating the TSR: the color-homogeneity of the objects and the color-heterogeneity between them, and that the objects of interest are stable in scale-space or are merged stable regions. Now, we want to present some results of our experiments. Therefore, we analyzed the TSR of 123 facade images from six German cities: Berlin, Bonn, Hamburg, Heidelberg, Karlsruhe and Munich, see fig. 4. These buildings have a sufficient large variety with respect to their size, the architectural style and the imaging conditions.

5.1 Manual Annotations

The ground truth of our experiments on facade images are hand-labeled annotations[1]. On one side, the annotation contains the polygonal borders of all interesting objects that are visible in the scene. On the other side, part-of relationships have also been inserted in the annotations. An extract of the facade ontology is shown in fig. 5.

5.2 Results

We investigate the coherence between our automatically segmented image regions taken from the TSR and the manual annotations. Our experiment consists of two tasks. First, we document the detection rate of the annotated objects, and secondly, we test, if the hierarchical structure of the TSR reflects the aggregation structure of the annotated objects.

[1] The images and their annotations were provided by the project *eTraining for interpreting images of man-made scenes* which is funded by the European Union. The labeling of the data has been realized by more than ten people in two different research groups. To avoid inconsistencies within the labeled data, there was defined an ontology for facade images with a list of objects that must be annotated and their part-of relationships. A publication of the data is in preparation. Please visit www.ipb.uni-bonn.de/etrims for further information.

Fig. 4. Left: Facade images from Berlin, Bonn, Hamburg, Heidelberg, Karlsruhe and Munich (f. l. t. r.), showing the variety of our data set. Right: Two levels from the irregular pyramid of the Hamburg image.

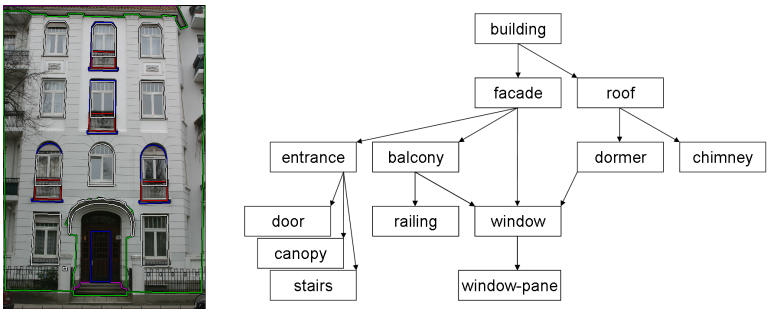

Fig. 5. Left: Facade image from Hamburg with manually annotated objects. Right: Major classes and their part-of relationships from the defined building-scene ontology.

In the 1st test, we perform a similar evaluation as it is done in the PASCAL challenge, cf. [7]. There, it is sufficient enough to map an automatically segmented region to the ground truth region, if the quotient of the intersection and the union of both regions is bigger than 0.5. So, we compute this quotient for each region in the TSR with respect to all annotated objects. Then, the maximum quotient is taken for determining the class label of the segmented region. If the ratio is above the threshold, then we call the object *detectable*. Otherwise, we also look for *partial detectability*, i. e. if the segmented region is completely included by an annotation. This partial detectability is relevant, e. g. if the object is occluded by a car or by a tree. Furthermore, we do not expect to detect complete facades, but our segmentation scheme could be used for analysis of image extracts, i. e. the roof part or around balconies.

Regarding the 2nd experiment, our interest is, if the TSR reflects the class hierarchy. This would be the case, if e. g. a *window*-region includes *window pane*-

regions, i. e. they both are connected by a path upwards in the TSR. So, we only focus on those annotated objects, which were (a) detectable or partially detectable and (b) annotated as an aggregate. In this case, the annotation includes a list of parts of this object. Then, we determine, whether we find other regions in the TSR, which are (a) also at least partially detectable and (b) are connected to the first region by a path upwards in the TSR. Then the upper region can get described as an *aggregate* containing at least the lower one. Additionally, we also check, whether not at least one but all parts of the aggregated object have been found, i. e. if the list of detectable parts is *complete*. Our results are shown in tab. 1.

class	objects	det.	part.	summed	aggregates	aggreg. compl.	aggreg.
all	9201	58%	26%	84%	2303	48%	13%
balcony	285	31%	62%	93%	243	53%	13%
entrance	72	47%	38%	85%	57	26%	11%
facade	191	49%	46%	95%	172	74%	13%
roof	178	46%	46%	92%	89	51%	13%
window	2491	56%	33%	89%	1369	46%	12%
window pane	2765	68%	8%	76%	0	-	-

Table 1. Results on detectabilty of building parts: 84% of the annotated objects have a corresponding region in the TSR or are partially detectable. The columns are explained in the surrounding text.

Note: the automatically segmented regions were only compared with the labeled data, no classification step has been done so far. We have presented first classification results on the regions from the Gaussian scale-space in [4], where we classified segmented regions as e. g. windows with a recognition rate of 80% using an Adaboost approach. With geometrically more precise image regions, we expect to obtain even better results. Furthermore, the detected regions can be inserted as hypotheses to a high-level image interpretation system as it has been demonstrated in [11]. It uses initial detectors and scene interpretations of mid-level systems to infer an image interpretation by means of artificial intelligence, where new hypotheses must be verified by new image evidence.

A similar experiment on aerial images showing buildings in the suburbs of Graz, Austria, is in preparation. There, we expect even better results, because the roof parts only contain relatively small parts which often merge with the roof in our observed scale range.

6 Conclusion and Outlook

We presented a purely data-driven image segmentation framework for a multi-scale image analysis, where regions of different size are observable in different scales. A defined region hierarchy graph enables us to obtain geometrically significantly more precise region boundary than we obtain by only working in the Gaussian scale-space. Furthermore, the graph can be used for detecting structures of aggregates. So, far we only compared the segmented regions to the annotated ground truth and did not present a classifier for the regions.

In next steps, we will insert more knowledge about our domain, e. g. the regions can be reshaped using detected edges. Then, the merging of region does not only depend on the observations in scale-space, but also on the not-occurrence of an edge. Therefore, we need a projection of the detected edges to the borders of the detected image regions in the lowest layer. Another way would be a multiple-view image analysis, where 3D-information has been derived from a stereo pair of images.

Our region hierarchy graph can further be used as the structure of a Bayesian network, where each node is a stochastic variable on the set of classes. The part-of relations between the regions are analogously taken to model the dependencies between these stochastic variables. This will enable a simultaneous classification of all regions taking the partonomy into account.

References

1. H. G. Akçay and S. Aksoy. Automatic detection of geospatial objects using multiple hierarchical segmentations. *Geoscience & Remote Sensing*, 46(7):2097–2111, 2008.
2. F. Bergholm. Edge focusing. *PAMI*, 9(6):726–741, 1987.
3. L. Brun, M. Mokhtari, and F. Meyer. Hierarchical watersheds within the combinatorial pyramid framework. In *12th ICDGCI*, LNCS 3429, pages 34–44, 2005.
4. M. Drauschke and W. Förstner. Selecting appropriate features for detecting buildings and building parts. In *Proc. 21st ISPRS Congress*, IAPRS 37 (B3b-2), pages 447–452, 2008.
5. M. Drauschke, H.-F. Schuster, and W. Förstner. Detectability of buildings in aerial images over scale space. In *PCV06*, IAPRS 36 (3), pages 7–12, 2006.
6. B. Epshtein and S. Ullman. Feature hierarchies for object classification. In *Proc. 10th ICCV*, pages 220–227, 2005.
7. Mark Everingham and John Winn. The pascal visual object classes challenge 2008 (voc2008) development kit. online publication, 2008.
8. J. M. Gauch. Image segmentation and analysis via multiscale gradient watershed hierarchies. *Image Processing*, 8(1):69–79, 1999.
9. E.B. Goldstein. *Sensation and Perception (in German translation by M. Ritter).* Wadsworth, 6th edition, 2002.
10. L. Guigues, H. Le Men, and J.-P. Cocquerez. The hierarchy of the cocoons of a graph and its application to image segmentation. *Pattern Recognition Letters*, 24(8):1059–1066, 2003.
11. J. Hartz and B. Neumann. Learning a knowledge base of ontological concepts for high-level scene interpretation. In *Proc. ICMLA*, pages 436–443, 2007.
12. R. Harvey, J. A. Bangham, and A. Bosson. Scale-space filters and their robustness. In *Scale-Space*, LNCS 1252, pages 341–344, 1997.
13. I. Lifschitz. Image interpretation using bottom-up top-down cycle on fragment trees. Master's thesis, Weizmann Institute of Science, 2005.
14. T. Lindeberg. *Scale space theory in computer vision.* Kluwer Academic, 1994.
15. P. Meer. Stochastic image pyramids. *CVGIP*, 45:269–294, 1989.
16. O.F. Olsen and M. Nielsen. Multiscale gradient magnitude watershed segmentation. In *Proc. 9th ICIAP*, LNCS 1310, pages 9–13, 1997.
17. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffussion. *PAMI*, 12(7):629–639, 1990.
18. A. Witkin. Scale-space filtering. In *Proc. 8th IJCAI*, pages 1019–1022, 1983.