

Considerations Regarding the Minimum Spanning Tree Pyramid Segmentation Method^{*}

(Why does it always find the lady?)

Adrian Ion, Walter G. Kropatsch, and Yll Haxhimusa

Vienna University of Technology, Pattern Recognition and Image Processing Group,
Favoritenstr. 9/1832, A-1040 Vienna, Austria
{ion,krw,yll}@prip.tuwien.ac.at

Abstract. The minimum spanning tree pyramid is a hierarchical image segmentation method. We study its properties and the regions it produces. We show the similarity with the watershed transform and present the method in a domain in which this is easy to understand. For this, a short overview of both methods is given. Catchment basins are contracted before their neighbouring local maximas. Smooth regions surrounded by borders with maximal local variation are selected. The maximum respectively minimum variation on the border of a region is larger than the maximum respectively minimum variation inside the region.

1 Introduction

Image segmentation is the process of partitioning the image into salient parts, i.e. partitioning the image into regions, such that each region is homogeneous with respect to some criteria such as greyvalue, colour, or texture. A segmentation method should have the following [1, 2]: create a hierarchy, capture perceptually important groupings, and run in linear time.

The presented work is motivated by the desire to further understand and improve the results of one such segmentation method, the minimum spanning tree pyramid (MST Pyramid) [3], and better fit it to the necessities of higher level processing [4]. During the past years, we have had the chance to use and test different implementations of the MST Pyramid method and even though random selection mechanisms are used [5] and in most of the cases the neighbourhood graph of an image does not have a unique minimum spanning tree, the most important entities, like the lady in Fig. 1, could always be found in the produced results.

After benchmarking the method using human made segmentations [6] the necessity for a more analytical approach has risen, for which the results are presented here. While looking in detail at the properties of the method, a certain similarity with the watershed transform [7] has also been observed and is included in this discussion.

^{*} This paper was supported by the Austrian Science Fund under grants FSP-S9103-N04 and P18716-N13.

This paper is organised as follows: Sections 2 and 3 contain a short description of the MST Pyramid and the Watershed transform, Section 4 presents the results of our study, Section 5 contains the outlook, and we end with the conclusions in Section 6.

2 The MST Pyramid Segmentation

Initially developed in the dual graph contraction and dual irregular pyramid framework and recently adapted to 2D combinatorial maps and combinatorial map pyramids [8], the MST Pyramid method [3] takes as input a weighted neighbourhood graph (NG) and produces a hierarchy of partitions by using the minimum spanning tree (MST) algorithm by Borůvka [9] and region internal/external contrast concepts [1].

To apply this method for image segmentation, the input NG is obtained by associating a vertex to each pixel and connecting two neighbouring vertices by an edge weighted with the distance of the two pixel values in some featurespace (we have experimented with difference in greyscale and RGB colour). Internal contrast of a region is defined as the biggest weight of the edges of its MST. External contrast between two neighbouring regions is defined as the smallest weight of the edges connecting vertices from the two regions. Algorithm 1 shows a description of the MST Pyramid method, and Fig. 1 shows some results. Step 4. of the Algorithm is called the contrast step. More details can be found in [10].

Algorithm 1 – MST Pyramid segmentation

Input: Attributed neighbourhood graph G_0 .

- 1: $k = 0$
- 2: **repeat**
- 3: $ME_k = \bigcup$ smallest edge around each vertex of G_k
- 4: $CE_k =$ edges from ME_k connecting two regions having larger internal contrast than the external contrast between them {contrast test step}
- 5: $G_{k+1} = (G_k$ with the edges from CE_k contracted)
- 6: $k = k + 1$
- 7: **until** $G_k = G_{k-1}$

Output: An attributed neighbourhood graph at each level of the pyramid (G_0, G_1, \dots, G_k) .

3 The Watershed Segmentation

A well known method used for segmentation but not only, the watershed transform has its origins in mathematical morphology. An intuitive way to view it is that of a landscape (topographic surface) being flooded by water (rain), and the watersheds being the lines which separate the different domains of attraction of

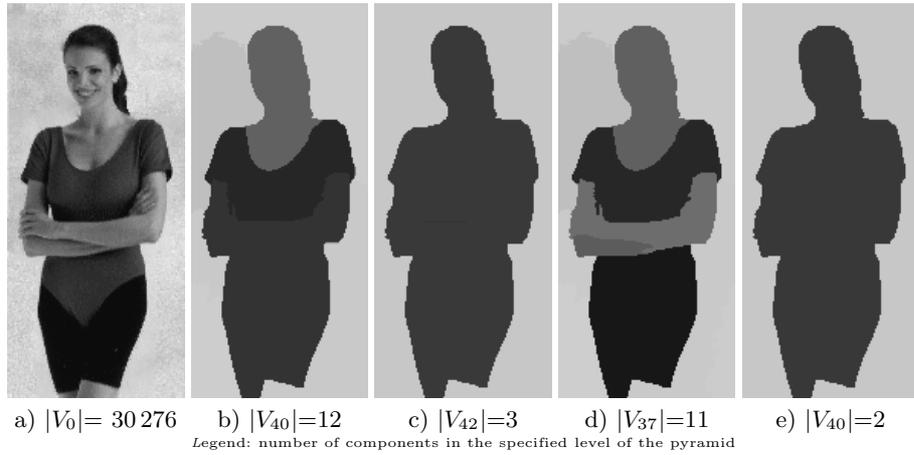


Fig. 1. Levels of the MST Pyramid segmentation of the image of a Woman: with (b,c) and without (d,e) the contrast test step.

rain over the relief [11]. Another way to imagine it, is to think of the landscape with holes made in the local minima, being immersed in water. Starting at these holes (local minima), catchment basins fill with water and the watersheds are the dams build in the places where two such catchment basins would meet to stop them from merging.

As mentioned above, the method can be applied to any topographic surface, and in the case of segmentation, it is most often applied to the gradient image of the image to process. The resulting catchment basins define the segments of the image. For a survey of existing methods that can be used to obtain the watershed transform and a detailed description see [7]. Fig. 2 shows an example result.

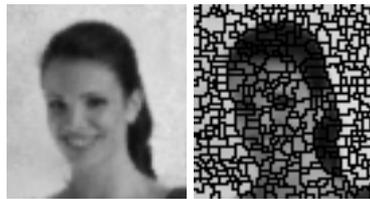


Fig. 2. Watershed segmentation of the image of a Woman's head.

4 Understanding Global Properties of the MST Pyramid

Local decisions taken when merging regions make the MST Pyramid method well suited for parallel processing. On the other side, having global information makes estimating, characterising, and influencing the results much easier.

After doing experiments, we have noticed that the majority of edges filtered by step 4 (contrast test) of Algorithm 1 pass the test, and that removing the filter and just contracting all the proposed edges does not significantly change the results in most of the levels of the pyramid (a discussion of this, follows at the end of Section 4.2). Because of this, we have simplified the model for the current study and removed the contrast test (step 4 in Algorithm 1) from it.

4.1 Case Study - a 1D Image

Let I be a 1D image, defined as $I(p), p = 1, \dots, m$. For a certain p , $I(p)$ identifies the pixel at position p in the image, $I(p_1)$ and $I(p_2)$ are neighbours if $|p_1 - p_2| = 1$. The neighbourhood graph $NG=(V, E)$ of such an image is a chain of vertices $v \in V$ (one for each pixel in the original image), with the vertices associated to each two neighbouring pixels joined by an edge $e \in E$, and its minimum spanning tree is the graph itself (See Fig. 3a,b).

The edge graph $EG=(VE, EE)$ of a graph is a graph where each vertex $ve \in VE$ represents an edge in the original graph (in our case NG), and two vertices are joined by an edge $ee \in EE$ if their corresponding edges in the original graph share a common vertex. (The EG will be used to show the similarity with the watershed segmentation).

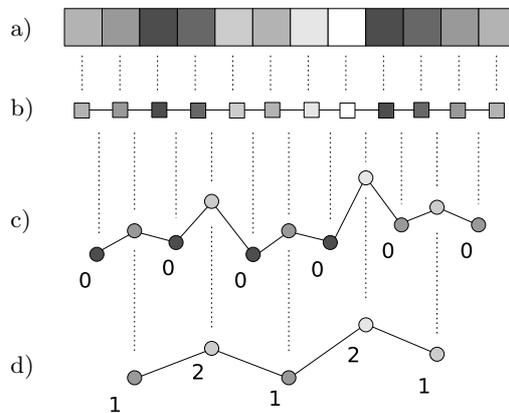


Fig. 3. MST based contraction of a 1D image: a) Image; b) associated NG ; c) associated EG with survival levels specified (higher vertex position means larger weight); d) associated EG of second pyramid level, with survival levels specified.

In the rest of the section, the numbering of vertices and edges in both the NG and the EG is done depending on the position of the associated element in the image, i.e. in the NG, v_i is associated to $p(i)$ and e_i is the edge connecting v_i with v_{i+1} , and in the EG, ve_i is associated to e_i respectively to the edge connecting v_i with v_{i+1} (See Fig. 3b,c).

A step in the MST Pyramid: We recall that edges in the NG are attributed with the difference in some featurespace of the two neighbouring pixels' values. The same value is used to attribute their associated vertices in the EG.

When searching for edges to be contracted, the MST Pyramid method selects the smallest edge connecting one vertex in the NG with it's neighbours. In the case of unequal values this results in a unique solution.

Because in our case, in one selection step any edge $e_i, i = 1, \dots, m - 1$ connecting two vertices v_i and v_{i+1} is part of two such tests, we conclude that e_i is selected if $e_i < e_{i+1}$ or $e_i < e_{i-1}$ or one of its bounding vertices is a leaf. Which, in it's associated EG, is equivalent to ve_i is not a local maximum or ve_i is a leaf i.e. $\neg(ve_i > \max(ve_{i+1}, ve_{i-1})) \vee (i \in \{1, m - 1\})$. This means, that in one such step only local maxima survive (See Fig. 3b,c).

What happens further in the MST Pyramid? The selected edges are contracted i.e. the new NG contains only the surviving (non-selected) edges and each group of vertices connected by the selected (non-surviving) edges are merged into one single vertex. In the EG this is equivalent with removing all the selected vertices and connecting each two surviving vertices if they were connected by a path of non-surviving vertices. (See Fig. 3c,d). The whole process of selection-contraction is repeated until no more contraction is possible.

Characterising the regions: The initial aim of the present study was to try to characterise the regions produced by the method i.e. given an image and a connected region in it (a cut), to be able to say what properties (internal/external) must the edges inside, outside, and on the region-border have, such that the region is produced as one segment in one of the levels of the hierarchy. In the case of our 1D image, this is reduced to: given the image and 2 edges, how can we best characterise the region between the two edges?

Recall that in one MST Pyramid step, from the level below only local maxima survive, which is equivalent to applying the watershed transform, on the gradient image of our 1D image (See Table 1).

Each local maximum that survives to a certain level k , defines in each level $l_i, i = 1, \dots, k$, on each side, an attraction area. (See Fig.4) These attraction areas contain only values smaller than that of the local maximum and depend on it and its neighbours (up to the local maxima that survived to level l_i). The higher we go in the pyramid, the larger these attraction areas become, and two such neighbouring regions, defined by 2 neighbouring local maxima, define a catchment basin which will be merged in the next step. (See Fig.4)

	Borůvka MST Pyramid	Watershed segmentation
Domain	edge graph / derivative along edge in the NG	gradient image / derivative in each pixel
Method	local maxima survive	

Table 1. Domain similarity of the MST Pyramid and the Watershed segmentation

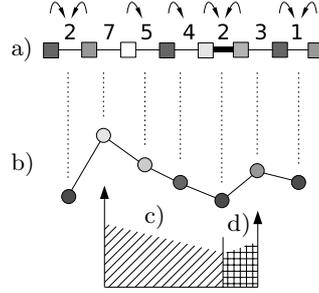


Fig. 4. Attraction areas in a catchment basin: a) NG, each vertex chooses it's smallest edge; b) associated EG; c) attraction area of edge 7 from the NG; d) attraction area of edge 2 from the NG.

Let $ve_i(k)$ and $ve_j(k)$ be the two neighbouring local maxima that define the two attraction areas $aa_1 = \{ve_{i+1}(k), ve_{i+2}(k), \dots, ve_q(k)\}$ and $aa_2 = \{ve_{q+1}(k), ve_{q+2}(k), \dots, ve_{j-1}(k)\}$, with $i < q < j$ denoting vertex indices in our chain-like edge graph and k denoting a level in our pyramid. From the above, we get:

$$\begin{aligned} \max(ve_i(k), ve_j(k)) &> \max(ve_{i+1}(k), \dots, ve_{j-1}(k)) \\ \min(ve_i(k), ve_j(k)) &> \min(ve_{i+1}(k), \dots, ve_{j-1}(k)) \end{aligned}$$

for any level k with $ve_i(k)$ and $ve_j(k)$ being local maxima in level k (they survive to level $k + 1$) and $ve_{i+1}(k), \dots, ve_{j-1}(k)$ not being local maxima. If we recursively follow the previous we get that:

$$\begin{aligned} \max(ve_i(k_1), ve_j(k_1)) &> \max(ve_{i+1}(k_2), \dots, ve_{j-1}(k_2)) \\ \min(ve_i(k_1), ve_j(k_1)) &> \min(ve_{i+1}(k_2), \dots, ve_{j-1}(k_2)) \end{aligned}$$

for $k_1 > k_2$, i.e. the biggest of the values surrounding a certain region in level k_1 is larger then biggest of all the values from any level $k_2 < k_1$ below. The previous holds for the smallest also. So, the maximum edge weight on the border of any region in any level, is larger than the maximum edge weight inside, i.e. maximum variation on the border of a region is larger then maximum variation inside the region, and the same holds for the minimum.

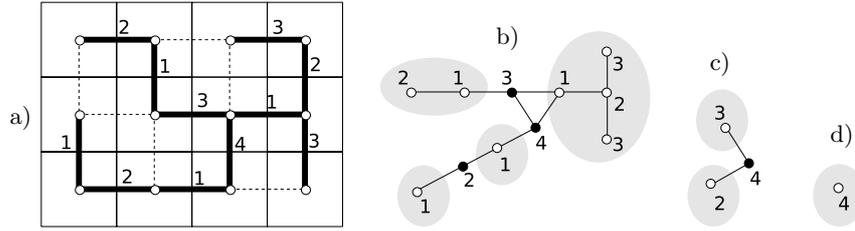


Fig. 5. MST based contraction, 2D case: a) image (thin continuous line) with associated NG (dashed line), its MST (thick line) with its edge weights; b) EG of the MST_NG (vertices of the same component are white and in the same grey ellipse, local maxima i.e. surviving vertices are black); c) EG - second level; d) EG - third level.

4.2 The 2D Case

To continue in the same line of ideas, we present the MST Pyramid edge selection mechanism for a 2D image, in a domain in which the presented similarities with the watershed method remain valid. For this, we do not focus on finding the minimum spanning tree (MST) itself, but on the way the MST Pyramid selects edges for contraction and thus constructs the MST by creating increasingly bigger parts from smaller ones.

For a given 2D Image and its associated NG. We determine the edge graph (EG) of the MST of the NG (MST_NG). Each vertex from the EG is attributed with the weight of its associated edge from the MST_NG. (See Fig. 5a,b)

According to Algorithm 1, in one MST Pyramid selection step each vertex from the NG selects the smallest edge around it (which is guaranteed to be on the MST of the NG). In the context of the EG of the MST_NG of the image, this can be described in a watershed like manner as follows:

1. **Initial configuration:** all the vertices in the EG have no labels and their attribute is the weight of the their corresponding edges in the NG;
2. From minimum to maximum **progressively threshold** the values in the vertices and each unlabelled vertex with a value below the threshold:
 - gets a unique numeric label, if no neighbours are numerically labelled yet (we found a new catchment basin/local minimum),
 - gets the unique numeric label of its labelled neighbours, if no 2 neighbours have different numeric labels (belong to different watersheds) and at least one is numerically labelled (watershed increases),
 - is labelled as “don’t contract/survive”, if none of the previous apply

At the end of each such step, the vertices labelled with the same numeric value are joined and they define connected regions (their corresponding edges in the MST_NG are contracted). A new EG is obtained by keeping the vertices with no numeric labels, the edges connecting them, and additionally connecting any 2 such vertices if in the labelled graph from the current step, they could be connected by paths made only of numerically labelled vertices. (See Fig. 5b,c,d)

As in the case of the 1D image, in each step local maxima from the previous level survive, and the properties observed in the 1D image case study remain valid for the MST-NG of a 2D image. Let r be a connected region in a 2D image. Let $\text{NG}=(V, E)$ be its associated neighbourhood graph. Let $E_c \subset E$ be the cut-edges connecting the vertices $V_r \subset V$, associated to the pixels of r , to the rest of NR ($V \setminus V_r$). Also let $E_r = \{(v_i, v_j) \in E \mid v_i, v_j \in V_r\}$, $\text{MST-NG}=(V_{mst}, E_{mst})$ the MST of NG, $E_{cmst} = E_c \cap E_{mst}$, and $E_{rmst} = E_r \cap E_{mst}$. If r is a region produced by the MST Pyramid (without the contrast step) then:

$$\begin{aligned} \max(E_{rmst}) &< \max(E_{cmst}), \\ \min(E_{rmst}) &< \min(E_{cmst}), \\ R_{mst} = (V_r, E_{rmst}) &\text{ is a connected graph.} \end{aligned}$$

The above also explains why most of the edges pass the contrast test (step 4 in Algorithm 1), and why this step does not significantly change the results. The purpose of the contrast step is to ensure that the algorithm produces regions with small variation surrounded by borders with large variation, but this is already achieved in most of the cases by the edge selection mechanism in step 3. Where the results differ significantly is that without using the contrast step, the pyramid always reaches an apex. The results when using the contrast step are better if we are looking for a segmentation that spans just one pyramid level and we have small regions. Here the additional condition stops these small regions to be merged with the surrounding while the rest of the graph is contracted.

Because of the way edges from the NG are attributed (distance in some featurespace), having the MST also gives us an upper bound on the weights of all the other edges. The difference between the values of two neighbouring pixels is less or equal to the sum of the weights of the edges along the path connecting their associated vertices in the MST of the NG.

5 Outlook

The previous study should help in improving the method and using it as a basis for reaching higher level abstraction. We plan to add the slope when calculating the edge weights to prevent “leakage”. Knowing the properties of the method allows us to easily controll it and insert a priory information from e.g. a successfull previous segmentation, or a high level process. Knowing the properties of the regions produced allows us to select a “best segmentation” that spans multiple levels and which can be used by higher level processes that need only one segmentation, or as a start seed for the ones that are able to use hierarchies but use a single segmentation at some instance of time (e.g. object recognition).

6 Conclusion

We have presented a set of properties of the regions produced by the MST Pyramid segmentation method and showed its similarity with the watershed transform of an image. Attraction regions are contracted before their neighbouring

local maxima. Smooth parts of the image surrounded by borders with maximal local variation are selected. Maximum and respectively minimum variation on the border of a region is bigger then the maximum and respectively minimum variation inside the region. Internal/external contrast conditions do not affect too much the lower levels of the pyramid.

References

1. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. *International Journal of Computer Vision* **59** (2004) 167–181
2. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22** (2000) 888–905
3. Haxhimusa, Y., Kropatsch, W.G.: Segmentation graph hierarchies. In: *Proceedings of Joint Workshops on Structural, Syntactic, and Statistical Pattern Recognition S+SSPR*. Volume 3138 of *Lecture Notes in Computer Science.*, Lisbon, Portugal (2004)
4. Keselman, Y., Dickinson, S.J.: Generic model abstraction from examples. *IEEE Trans. Pattern Anal. Mach. Intell.* **27** (2005) 1141–1156
5. Kropatsch, W.G., Haxhimusa, Y., Pizlo, Z., Langs, G.: Vision pyramids that do not grow too high. *Pattern Recognition Letters* **26** (2005) 319–337
6. Haxhimusa, Y., Ion, A., Kropatsch, W.G.: Evaluating graph-based segmentation algorithms. In: *Proceedings of the 18th International Conference on Pattern Recognition*, Hong Kong (2006)
7. Roerdink, J.B.T.M., Meijster, A.: The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta Informaticae* **41** (2000) 187–228
8. Brun, L., Kropatsch, W.G.: Irregular Pyramids with Combinatorial Maps. In: *Proceedings of Joint Workshops on Structural, Syntactic, and Statistical Pattern Recognition S+SSPR*. Volume 1876 of *Lecture Notes in Computer Science.*, Alicante, Spain (2000) 256–265
9. Neštril, J., Miklovà, E., Neštrilová, H.: Otakar Borůvka on minimal spanning tree problem translation of both the 1926 papers, comments, history. *Discrete Mathematics* **233** (2001) 3–36
10. Haxhimusa, Y.: Structurally Optimal Dual Graph Pyramid and its Application in Image Partitioning. PhD thesis, Vienna University of Technology, Faculty of Informatics, Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group (2006)
11. Meyer, F.: Graph based morphological segmentation. In Kropatsch, W.G., Jolion, J.M., eds.: *2nd IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition*. Volume 126., Vienna, Austria, OCG (1999) 51–60