

# Document

HUYNH Le Duy

19 novembre 2015

## 1 The objective

The objective of our method is to obtain a self-dual hierarchical representation of the input images. At first, we looking for a self-dual segmentation. The advantage is that we don't have to process the image twice to obtain all objects. The self-dual operation will process both the bright-object-on-dark-background and dark-object-on-bright-background the same way. Then, an hierarchical structure of regions extracted, which present the relation of inclusion, is recorded.

## 2 Well-composedness

Our approach involves the notion of well-composedness. It has been introduced to avoid the connectivity paradox which comes from the fact that the Jordan theorem is not fulfilled in discrete topology if we use only one type of connectivity. The Jordan theorem states that a simple closed curve divides the space into two disconnected regions. However, in a set of point using 4-connectivity, a simple curve may result in more than two disconnected regions, while in the 8-connectivity case, a simple curve exist but does not divide the plan. This paradox is solved by using a couple of connectivities, one for the background and the other for the object which require an assumption about the gray level of the background and the object. This choice is better avoid since we want to obtain a self-dual segmentation.

Well-composedness image enjoys useful topological and geometric properties, especially the fact that there is only one connectivity relationship between points of the image. Defined by Latecki et al. : a set is weekly well-composed if any 8-component is a 4-component ; a set is well-composed if both that set and its complement are weekly well-composed. An image is not a priori well-composed, there are 2 approaches to obtain a well-composed images : by modify value of pixel in image or by a well-composed interpolation.

## 3 Well-composed interpolation

Proved in [Links between the morphological tree of shapes and well-composed gray-level images], if the interpolation is calculated by median, a self-dual plain map is obtained. The initial image is not deteriorated in utilising interpolation with subdivision.

TABLE 1 – Well-composed interpolation

a	$(a + b)/2$	b
$(a + c)/2$	$median(a, b, c, d)$	$(b + d)/2$
c	$(c + d)/2$	d

## 4 Original approach

Our approach contain three steps : construction of a hierarchical structure represent the image, pruning of that tree and text grouping under condition. First, we use a self-dual segmentation of the input image. During the segmentation, a hierarchical structure contains inclusion relation of components is also recorded. Then, properties of components will be checked to eliminate regions which are not probable to be text. These regions will be merge to their upper region. Finally, the pruned tree will then be used to group remain components into text candidates. Only components sharing same background (i.e brother nodes) will be grouped together.

### 4.1 Segmentation and construction of the tree :

The segmentation is done by using the zero-crossing of the morphological laplacian to mark boundary of regions. The morphological laplacian is obtain from the .However, the morphological laplacian is not a priori well-composed, therefore we can describe it with one type of connectivity and it will suffer the connectivity paradox or a couple of connectivities could be use but an arbitrate choice has to be made and the process will not be self-dual. Neither choice is appropriate. Segmentation works on the well-composed interpolation of the morphological laplacian allows the use of only one connectivity type without suffer the connectivity paradox. This process will label each connected component with a number. A hierarchical structure based on inclusion relation will also be recorded, called tree of shape of laplacian :

- A node is a connected region having same sign
- The root node is either positive or negative, defined by the median of pixels on the bother.
- A node included in another node belongs to that node.
- Null pixel included in a region belongs to that region, null pixel on the boundary of two node belongs the upper nodes.

### 4.2 Prune the tree :

Region will be verified by these condition to eliminate those which are less probable to be text. We keep those which satisfy these condition :

- Average Gradient magnitude of point on the contour  $> 30$ .
- Bounding box height and width  $> 5$  pixels.
- $0.1 < \frac{\text{Bounding box height}}{\text{bounding box width}} < 5$ .
- $\frac{\text{Region area with all children}}{\text{bounding box area}} > 0.1$

- Region area  $> 30$

For some properties will be collect during the segmentation and construction of tree of shape of laplacian (TOSOL), we can decide to keep these region or not at that time. Properties retrieved after the segmentation have finished such as the third and the fourth will be used to further prune the TOSOL. Nodes does not pass will be marked so it will be transparent to the grouping phase.

### 4.3 Text grouping

After pruning the tree, remain nodes will be grouped to provide text candidate. 2 nodes will be grouped together if :

- They have the same parent.
- Their heights are not different than 2 times.
- Distance from a node to another is smaller than the maximum of its height and width.
- Horizontal alignment.

## 5 Our implementation

**The segmentation and construction of the tree** is implement by a font propagation of labels. **Tree pruning** is divided in two parts, during the segmentation of image (by using properties can be obtain during the labeling process) and after the segmentation (using properties obtained after). The original image will be convert to gray level in function of their luminal. The morphological laplacian will be computed. A morphological gradient will also be calculated to provide information to prune the tree. A 1 pixel border will also be add with the value equal the median of all point on the border to act like the root node. Their resolution will be double with a local well-composed interpolation. For each connected region having same sign, a label is propagated. As defined, all connected zeros will be labeled. Because the laplacian is well-composed, only 4-connectivity is used. Following the propagation of each label, there will be unlabeled regions, which have different sign, left behind. Before an unlabeled regions is labeled, by following points on the outer contour, we will obtains the average gradient magnitude on the contour and the bounding box of that region. The outer contour is retrieved by collecting pixels having unlabelled pixels on its right, until forming a closed contour. The starting point is in the left of the top left most point of an unlabelled regions (which is the first point that is reached when the image is read from left to right, top to bottom). If properties of this region pass the condition, new label is used to mark that region. Otherwise, the label of surround regions will be used. A record shows inclusion relationship of labelled will be keep. In the end, we obtains a labelled image and a table of parent.

In the subsequent stage, the labelled image is further simplifier using properties obtain after the labelling process. Nodes do not pass criteria will be transparent to the grouping process.

For the **text grouping process**, in our implementation, the first and second criteria are strictly followed. The third and last one are implemented by using 3 searching line in the horizontal direction (in the middle, 10% of height from the top and bottom) to look for brother nodes on the left and right. Two nodes will be considered if one of their searching lines reaches the others.

## 6 Modified approach

Because the well-composed interpolation cost, it also double the resolution and make later steps more expensive. We will label the laplacian implicitly. The precedent interpolation is local, because interpolated points depend only on its nearest neighbours. It is not the only self-dual well-composed interpolation. We will chose an approach that is familiar with the nature of our labelling process so it can be process implicitly without having to store the whole image. The most important result of the well-composed interpolation is the decision for the two critical configuration : which value is connected and which one is separated. The precedent approach use the local value of these point, the modified approach uses the information of inclusion : (The pixels will be separated if they are is surrounded by another colour), all interpolated point will have the colour of the surrounded pixels (in which surrounded pixels refers the pixels which will be labelled earlier in the labelling process). This approach is still :

- **Self-dual** : The complement operation does not affect the inclusion relation.
- **Well-composed** : No criteria configuration left or created after the interpolated (prove ?).  
(Because all in-between points have the same value, these points and any exist point cannot create a criteria configuration)

This approach disjoints the inside region at any criteria configuration found, which is not always the case when using old approach. It allow us to treat the original image implicitly as it has been interpolate. Focusing on the original point only, result of this interpolation on image can be described as the outside region uses 8-connectivities and the inside uses 4-connectivities. In its turn, the inside region may include other region so its must be described by 8-connectivities. Actually, only the border of the inside region needs to be described by 4-connectivities.

## 7 Implementation

In reality, results of the new well-composed interpolation allows us to treat the original image implicitly without having to process with the interpolated one : image will be described using 8-connectivities except points on a region's border with its container regions.

In this approach, we do not have to interpolate, only a 1-pixel-wide border needed to be add to the original laplacian to define the root node. Most of the modification is in the labelling process : Point on the border will be keep trace because they are 4-connected. The propagation of labels is done with 8-connectivity except at marked point. During the propagation, neighbours which have different sign will be marked as they are on the inner contour of other regions. To obtain the inner contour of regions (to obtain average gradient magnitude and bounding box), we only need to follow an 4 connected closed curve of marked point.

Processing time reduce significantly because we do not have to calculate the interpolated point and other process (labeling, grouping) work on image at almost original size.