Taking into account inclusion and adjacency information in morphological hierarchical representations, with application to the extraction of text in natural images and videos

HUỲNH Lê Duy

Under the direction of

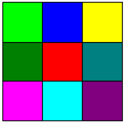
Dr. Yongchao XU and Prof. Thierry GÉRAUD

Sorbonne Université
Laboratoire de Recherche et Développement de l'EPITA

December 13, 2018

Image representation

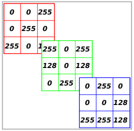
- Pixel-based image representations



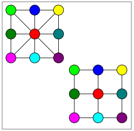
discrete function

$$\begin{aligned} D_I &\rightarrow V \\ p &\mapsto I(p) = v \end{aligned}$$

matrix



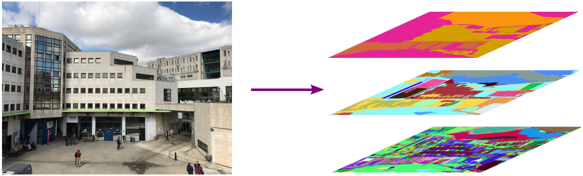
graph



- May not be suitable for higher level of image understanding
- Considering a set of pixels $\Rightarrow$ region-based image representation

Multi-scales nature of images

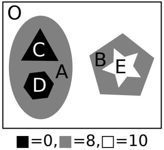
- Features of interest could be found at different scales
  - The building, the sky ⇒ large scales
  - Windows, people ⇒ small scales



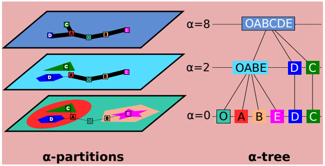
- Many applications may benefit from a region-based multi-scale representation:
  - Computed once
  - Adapted to specific applications afterward
    ⇒ Versatile and efficient
- They could be classified into the **hierarchies of partitions** and the **trees based on threshold decompositions**

Hierarchies of partitions

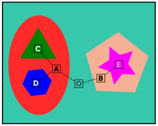
- Hierarchies of partitions:
  - α-tree (hierarchy of quasi-flat zones) [Soille, 2008]
  - Binary Partition Tree [Salembier et al., 2000]
  - Hierarchies of watershed [Najman, 2011]
- **Underlying requirement:** Region model and dissimilarity measure
- **Relationship:** Adjacency of regions



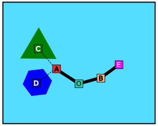
Image



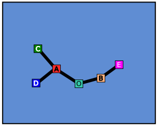
α-tree



α=0 (flat-zone partition)



α=2

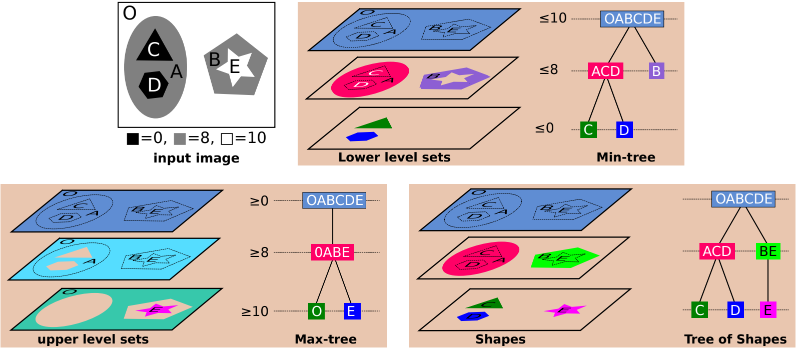


α=8

α-tree: An example of hierarchy of partitions

Trees based on threshold decompositions

- Trees based on threshold decompositions:
  - **Min-tree** [Salembier et al., 1998]: Connected Components (CCs) of lower level sets
  - **Max-tree** [Salembier et al., 1998]: Connected Components (CCs) of upper level sets
  - **Tree of Shapes** [Monasse et al., 2000]: Holles filled CCs of the Min-tree and Max-tree
- **Underlying requirement:** Ordering of the value space
- **Relationship:** Inclusion of level sets

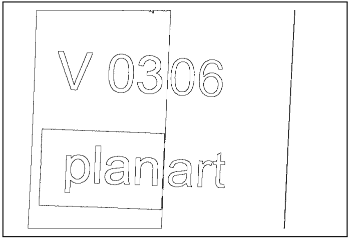


Examples of trees based on the threshold decompositions

The Tree of Shapes (ToS)

- The ToS is a self-dual fusion of the Min-tree and Max-tree.
- The ToS encodes the inclusion of the **image level lines**, which are the **contours of shapes**

Image



Notable level-lines

Relationships between objects of interest

- Object of interest should be represented by nodes on the tree.
- Objects usually do not appear in isolation.
- Especially in the field of text detection and recognition.

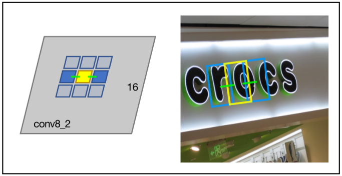
Detecting all these numbers is interesting, but we might also want to know how they are related to each other

- Therefore most classical text detection methods integrate a grouping step to form characters into text strings

Adjacency relationship (1/2)

- “Loose” adjacency: separated by a small number of background pixels
  $\rightarrow$ tell us how objects are organized in the image

3-nearest neighbors graph [Fabrizio et al., 2016]

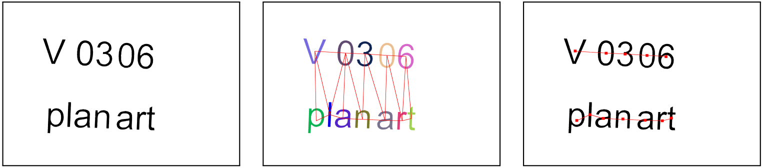


Linking adjacent detections [Shi et al., 2017]

- Most text string formation methods form an adjacency graph of objects of interest
- before applying more spatial assumptions (e.g., distance, alignment) and assumption about similarity of texts to segment that graph.
- However, we may lost important information only considering spatial information.

Adjacency relationship(2/2)

- Using only adjcency and spatial relationships could lead to misunderstanding.



Binarization of the input image   Adjacency graph   Horizontal adjacency graph

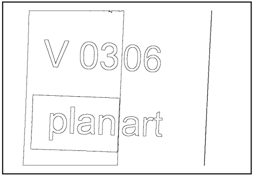
- Sometimes we could spot a problem immediately e.g.,"**planart**" is not an English word.
- Sometimes it is difficult to determine the error.
  $\rightarrow$ Is there anything wrong with "V0306"?

Background/Object relationship

- One important information is not used: “Background/object” relationship, which also carries semantic information
- Problem: How can we deduce the “Background/object” relationship between image regions?

Input image

Text strings we would make, knowing some words are in different backgrounds.



Contour of some notable regions

- Observation: The contour of object is always included in the contour of its background.
- This inclusion relationship happens to be encoded in the ToS.

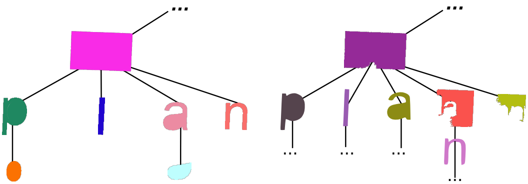
Related objects may appear far on ToS

- The ToS encodes all the level-lines, yet not all of them are interesting.
- Images may be affected by uneven illumination, noise (compression, scene complexity)

Parts of image

Ideal

What we may obtain with ToS



Our approach

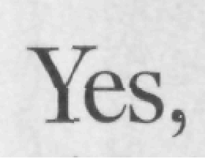
- **Motivation**:
  - Using background/object relationship would add important information to the task of grouping related objects, which would benefit text detection.
  - The background/object relationship could be deduced from the ToS.
- **Problem**: We do not know the distance between related objects on ToS.
- **Approaches**:
  - A simplified ToS on which related objects are siblings.
  - Build a more complex graph with inclusion and other spatial relationship and analyze it in a hiearchical way.

Plan

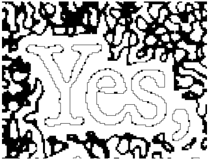
Morphological Laplacian

The morphological Laplace operator:

- $\Delta_{\square} = \delta_{\square} + \varepsilon_{\square} - 2id$
- **simple**, and provides **closed contour**.
- Robust to uneven illumination

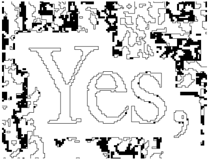


u



$LoG_{17}(u) = 0$

$\Delta_{\square}(u)$



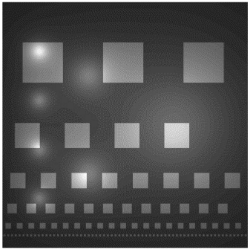
$\Delta_{\square}(u) = 0$

Morphological Laplacian is **less affected by noise** and **does not modify contour position**. . .

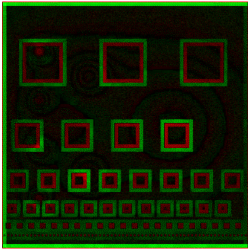
Morphological Laplacian

The Morphological Laplace operator:

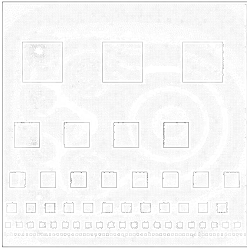
- $\Delta_{\square} = \delta_{\square} + \varepsilon_{\square} - 2id$
- **simple**, and provides **closed contour**.
- robust to uneven illumination



u



$\Delta_{\square}(u)$



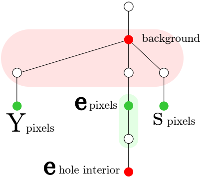
$\Delta_{\square}(u) = 0$

. . . and has high performance in **uneven illumination**

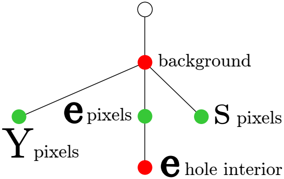
Tree of Shapes of Laplacian sign

- The Tree of Shape of Laplacian signs (ToSoL) encodes the **inclusion of Morphological Laplacian 0-crossings**.
- Tree structure:
  - Nodes: Conected Components (CCs) that have the same signs.
  - Root: infinite background whose sign is defined by median of elements on borders.
  - Zero CCs: merged with upper node.

$\Delta_{\square}(u)$.



$\mathfrak{S}(\mathrm{sign}(\ \Delta^{\mathrm{wc}}_{\square}(u)\ ))$.



Simplified $\mathfrak{S}$, called ToSoL.

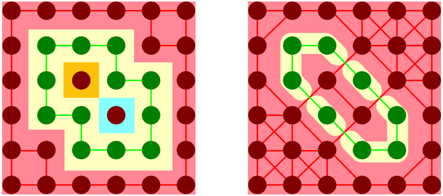
■ ■ □: positive, negative, and zeroes of Δ

About connectivity paradox

- ToSoL is a set of Connected Components
  $\rightarrow$ need to specify the connectivity
  $\rightarrow$ *Jordan Separation Theorem* may not hold [Rosenfeld et al., 1966].

A simple closed curve may

divide the plane into 3 CCs | not divide the plane at all

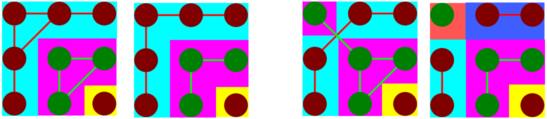


- Unwanted Solution: One connectivity for background, one for objects
  $\rightarrow$ We cannot assume which Laplacian sign is background or object without breaking self-duality.

- Solution: Make the Laplacian map well-composed [Latecki et al., 1995].

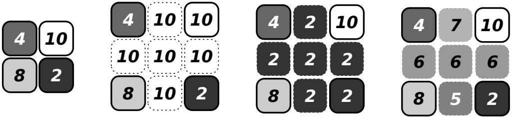
Making an image well-composed

- In a well-composed image, any 8-CCs are also 4-CCs.



Well-composed   Not well-composed

- Well-composedness could be obtained
  - Local well-composed interpolation → Median-based interpolation is a solution for 2D self-duality [Géraud et al., 2015].



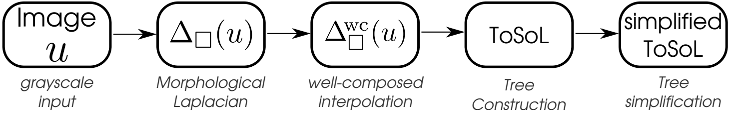
u $I_{max}(u)$ $I_{min}(u)$ $I_{median}(u)$

Example of well-composed interpolation

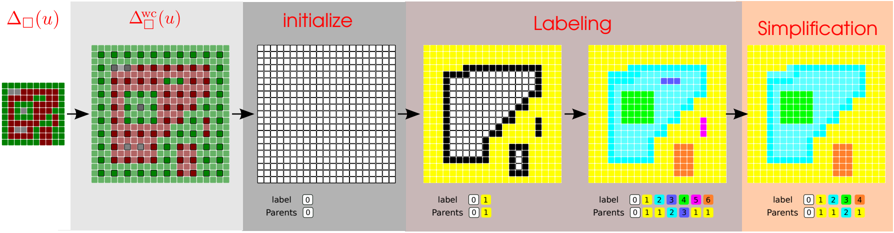
  - Non-local well-composed interpolation [Boutry et al., 2015]

Tree construction

- The ToSoL is constructed top-down.



ToSoL construction



From a Laplacian map to a ToSoL

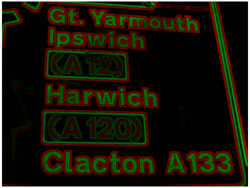
Optimization

- More optimization are detailed in manuscript:
  - The emulatable well-composed interpolation
  - The tree simplification step

  $\rightarrow$ Our ToSoL is computed and simplified with linear time complexity.

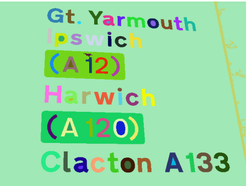
How does the ToSoL look like (1)

Input



Morphological Laplacian

Labeling Map of ToSoL (646 nodes)



Labeling Map of simplified ToSoL (67 nodes)

How does the ToSoL look like (2)

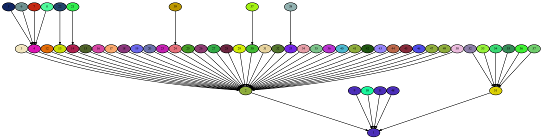
Input

Laplacian

Even for low contrast images . . .

How does the ToSoL look like (3)

The labeling map

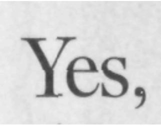


The tree structure

. . . we still obtain a clear representation of objects of interest

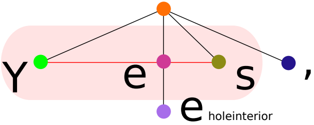
ToSoL for Text in natural images

- Our contributions are mainly in **localization** and **extraction**.
- Proposed solution:
  - Build a simplified ToSoL that represents only notable regions.
  - Build an adjacency graph that links only similar and horrizontally aligned sibling nodes.
  - Return large groups as the output of our methods.



Image

ToSoL labling map



Extract large horrizontally adjacent siblings

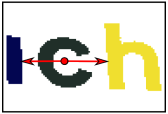
Objectives

We want to extract text characters that are

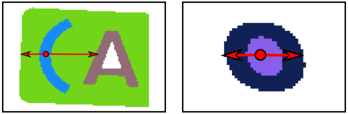
- Recognizable $\rightarrow$ we filter the ToSoL with these criteria:
  - Sufficiently contrasted: $G_{avg} > 30$ (Average Gradient Magnitude)
  - Sufficiently large: $h > 5px$ and $w > 5px$ (Height and Width)
  - Not too irregular: • $0.1 < \frac{h}{w} < 10$ (Height over Width)
- Semantically related $\rightarrow$ we build a Graph of Shapes with these criteria:
  - In the same background : sibling nodes
  - Roughly horizontally aligned: achieved by a simple spatial search.
  - Not too far apart: by setting the maximum searching distance
  - Similar (e.g. height): $SH(c_i, c_j) = \frac{min(h_{c_i}, h_{c_j})}{max(h_{c_i}, h_{c_j})} < 0.5$

An adjacency graph w.r.t inclusion (1)

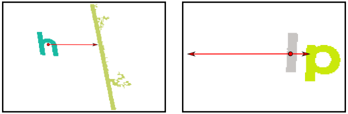
- A simple adjacency graph: A node of ToSoL is connected to at most two other siblings



A region has at most 2 neighbors, one on each side



Searching stops when stepping out of parent. . .

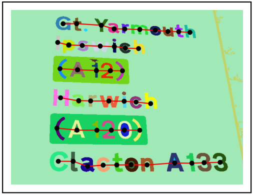


. . . or does not find a similar region before reaching maximum distance

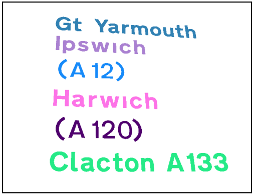
An adjacency graph w.r.t inclusion (2)

- Each large CC is a text candidate of our methods

Input



Adjacency graph w.r.t inclusion



Extration of large CCs

Results

- Localization of text elements (its bounding boxes).
- Segmentation of text from background.
- Detections are grouped based on semantic relationship.

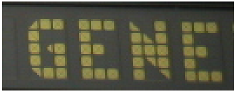
Input | ToSoL labling map | Detection

Method	Recall	Precision	F-score	Consistency
SWT [Epshtein et al., 2010]	46.42%	88.61%	60.92%	50.50%
ER [Neumann et al., 2016]	61.31%	89.20%	62.92%	72.67%
TMMS [Fabrizio et al., 2016]	**78.46%**	75.22%	**76.80%**	79.13%
Our	63.62%	**93.36%**	**75.65%**	**84.98%**

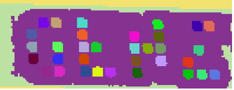
Text segmentation comparison on ICDAR 2013 dataset (*ICDAR Robust Reading Competition: Focused scene text*).

Motivation

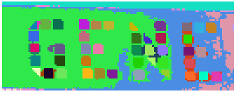
- ToSoL is too simplified → we may prefer ToS to retrieve image details



Input



ToSoL



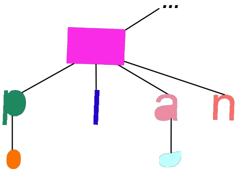
MToS [Carlinet et al., 2015]

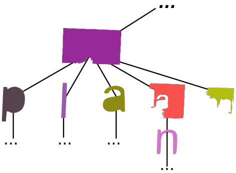
- Some related objects are far from each other on the ToS's structure.

Parts of image

Ideal

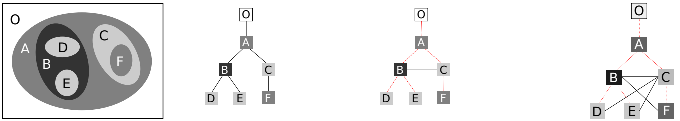
What we may obtain with ToS





Spatial alignment graph w.r.t inclusion

- A spatial alignment graph with respect to inclusion
- A node of ToS is connected with all regions **lying horizontally w.r.t. it**, that is **neither its decendant nor its ancestor**

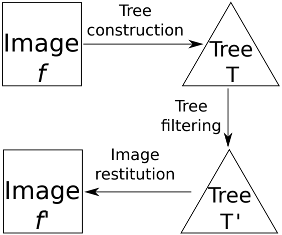


Input ToS Spatial adjacency graph w.r.t inclusion Spatial alignment graph w.r.t inclusion

- This graph is analyzed by shape-based morphology [Xu et al., 2016]

Tree-based implementation of connected operator

- One popular implementation of connected operators.

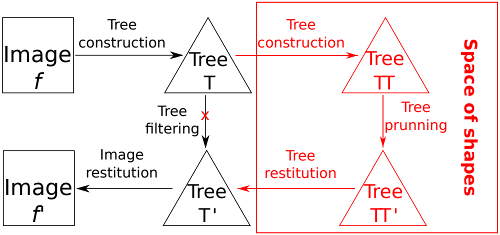


Tree-based connected filter

- Tree filtering depends on attributes of nodes which could be:
  - **Increasing** → only **pruning strategies**, i.e. removing some subtrees.
  - **Non-increasing**
    → **Pruning** strategies (Min,Max,Viterbi) based on **multiple nodes**.
    → **Non-pruning** strategies (Direct,Subtractive) based on **individual node**.

Shape-spaces Morphology [Xu et al., 2016]

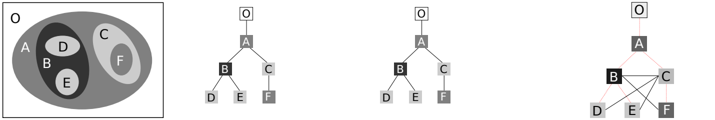
- Non-pruning approach but based on non-local attributes.
- A tree = a graph (i.e. a space of shapes) → use a second tree for analysis.



Workflow of connected filter in tree-based shape space

Our extension of the shape-spaces morphology framework (1)

- Shape-space topology is strictly defined by the parent children relationship of the first tree
  $\rightarrow$ However it could be any space represented by the graph of shapes. e.g., the spatial alignment graph



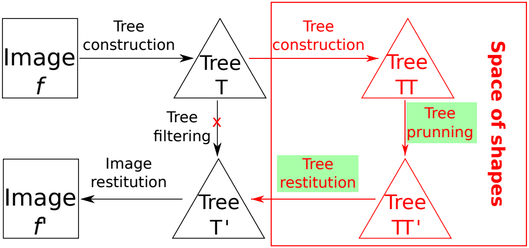
Input — ToS — Tree-based shape-space — Proposed shape-space

Our extension of the shape-spaces morphology framework (2)

- The method of tree restitution only allows the pruning of second tree.

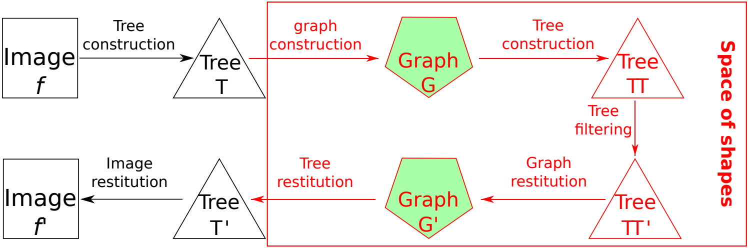
$\rightarrow$ We propose a new tree restitution process that:

- is consitent with how the image is reconstructed from the first tree.
- allows any filtering strategy on the second tree.

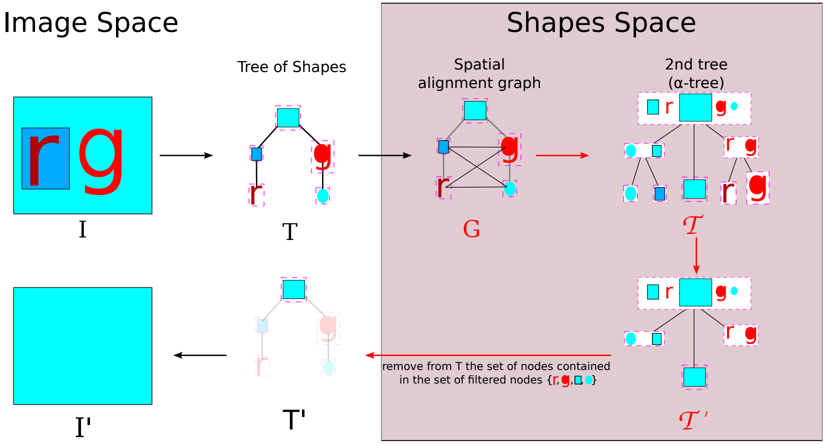


Workflow of connected filter in tree-based shape space

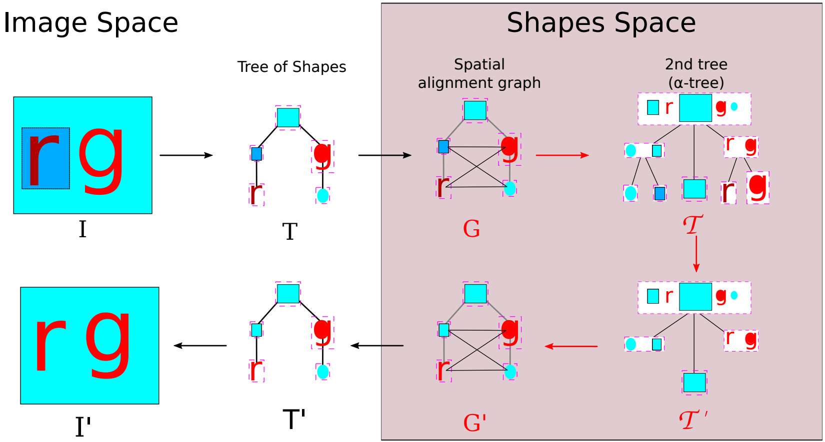
The proposed workflow



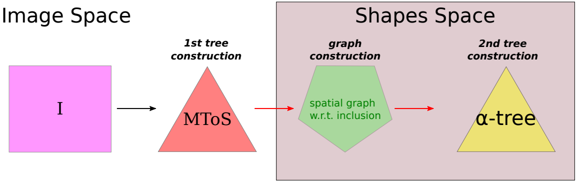
Original approach in new shape-space



New reconstruction approach in new Shape-space

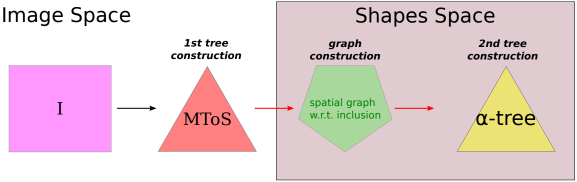


Choice of hierarchies (1)



- First tree: A **simplified Multivariate ToS** [Carlinet et al., 2015]
  - An extension of ToS to Multivariate images (in our case RGB) without imposing a total order on the vector space.
  - Simplified by minimizing the Mumford-Shah cartoon model constrained by the tree topology [Xu et al., 2013]

Choice of hierarchies (2)



- Second tree: an α-tree.
  - We want to group similar, nearby regions $\rightarrow$ hierarchies of segmentation.
  - We want more robustness to uneven illumination $\rightarrow$ a single linkage structure like α-tree.
  - Edges are weighted with dissimilarity in height, colors, distance on trees and spatial distance. . .
  - Simple, could be computed effectively [Najman et al., 2013].
- Text groups are expected to be represented by nodes on the α-tree.

Quality of the second tree

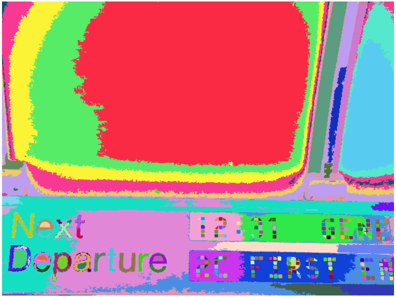
- (A) Best fitting subset: Set of ToS nodes that maximize IoU (Intersection over Union) with Ground Truth.
- (B) Best fitting node: A node on α-tree that has highest IoU with Ground Truth.
- Quality q: IoU(B) over IoU(A)

Comparasion of different dissimilarity functions on *ICDAR Robust Reading Competition: Focused scene text* dataset.

Dissimilarity function	Average Quality q
DS_{height}	79,38%
$L2_{RGB}$	84,16%
ΔE (L*a*b*)	**87,74%**

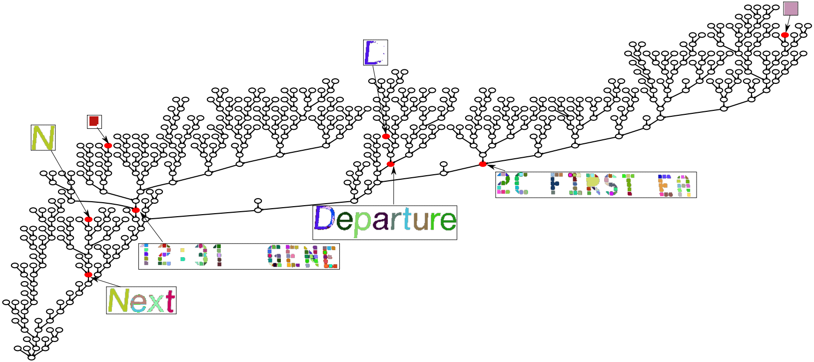
How does it look like at each step? (1)

Input



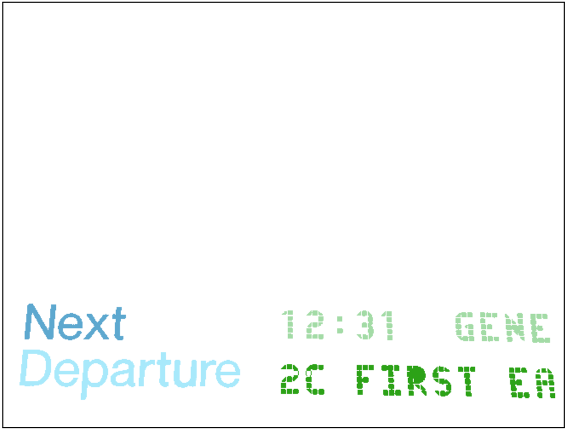
Labeling Map of MToS

How does it look like at each step? (2)



α-tree (565 nodes, indexed by node's height)

How does it look like at each step? (3)



Some interesting nodes

Some results

Good Segmentation | "Broken" objects | Connected objects | Obscurity

Image

Ground Truth

ToS labeling map

Best fitting subset

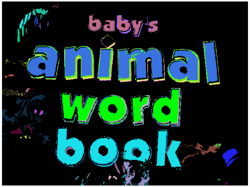
Best fitting node ($CIE\Delta E$)

Selection of candidates

Objective: select a small set of notable nodes as text candidates for false possitive removal and recognition steps.

Proposed method:

- An (ω)-CC approach [Soille, 2008].
- Select nodes that are largely different from their parent.



(1.8)-CCs

top 1.25% nodes with highest IR_{area} (83 nodes)

Interactive segmentation

- Segmentation with marker [Salembier et al., 2000]:
  - Input: image and a background marker (B) and an object marker (O).
  - Output: Largest image regions that intersect (B) but not (O).

Input and markers Results

Motivation

- **Motivation**:
  - Background/object relationship would add important information to the task of grouping related objects, that would benefit text detection.
  - The background/object relationship could be deduced from the ToS.
- **Problem**: We do not know the distance between related objects on ToS.

Conclusion

- Tree of Shapes of Laplacian Signs.
  - A new Tree of Shapes based on Morphological Laplacian.
  - Linear time algorithms.
  - Related nodes are siblings in tree structure.
  - Application in text detection and segmentation.
- Extension of shape-space morphology.
  - Show that the framework could be extended for any shape-space.
  - A new reconstruction process that allows flexible manipulation of the second tree.
- Spatial alignment graph w.r.t inclusion
  - That allows manipulation of both inclusion and spatial alignment.
  - Could be analyzed with shape-space morphology for grouping of unconnected components.
- Implementation of these algorithms in C++ (with Olena and Pylene).

Perspective

- Tree of Shapes of Laplacian Signs.
  - ToSoL for multivariate image.
- Extension of shape-space morphology.
  - Explore other applicative aspects of this work, e.g. filtering/image simplification
- Spatial alignment graph w.r.t inclusion
  - Choice of the first tree.
  - Construction of second tree as a multivariate tree.
- Application in text detection and recognition.
  - Integrate our approach in a text detection pipeline to get an end-to-end evaluation.

Thanks for your attention

Publications

Conference papers

- Lê Duy Huỳnh, Yongchao Xu, and Thierry Géraud (2016). “Morphology-Based Hierarchical Representation with Application to Text Segmentation in Natural Images”. In: Proceedings of the 23st International Conference on Pattern Recognition (ICPR). Cancún, México
- Lê Duy Huỳnh, Yongchao Xu, and Thierry Géraud (2017). “Morphological Hierarchical Image Decomposition Based on Laplacian 0-Crossings”. In: Proceedings of the 13th International Symposium on Mathematical Morphology (ISMM)

References I

Boutry, Nicolas, Thierry Géraud, and Laurent Najman (2015). "How to Make nD Functions Digitally Well-Composed in a Self-dual Way". In: Mathematical Morphology and Its Applications to Signal and Image Processing. Ed. by Jón Atli Benediktsson et al. Vol. 9082. Cham: Springer International Publishing, pp. 561–572 (cit. on p. 18).

Calarasanu, Stefania, Jonathan Fabrizio, and Severine Dubuisson (2016). "What is a good evaluation protocol for text localization systems? Concerns, arguments, comparisons and solutions". In: Image and Vision Computing 46, pp. 1–17 (cit. on p. 59).

Carlinet, Edwin and Thierry Geraud (2015). "MToS: A Tree of Shapes for Multivariate Images". In: IEEE Transactions on Image Processing 24.12, pp. 5330–5342 (cit. on pp. 30, 39).

Epshtein, Boris, Eyal Ofek, and Yonatan Wexler (2010). "Detecting text in natural scenes with stroke width transform". In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, pp. 2963–2970 (cit. on pp. 28, 59).

Fabrizio, Jonathan et al. (2016). "TextCatcher: a method to detect curved and challenging text in natural scenes". In: International Journal on Document Analysis and Recognition (IJDAR) 19.2, pp. 99–117 (cit. on pp. 8, 28, 59).

Garcia, C. and X. Apostolidis (2000). "Text detection and segmentation in complex color images". In: Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 4. IEEE, pp. 2326–2329 (cit. on p. 56).

Géraud, Thierry, Edwin Carlinet, and S. Crozet (2015). "Self-Duality and Discrete Topology". In: International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing. Vol. 9082. LNCS. Springer, pp. 573–584 (cit. on p. 18).

Jaderberg, Max et al. (2014). "Reading Text in the Wild with Convolutional Neural Networks". In: arXiv:1412.1842 [cs] (cit. on p. 56).

Latecki, L. J., U. Eckhardt, and A. Rosenfeld (1995). "Well-Composed Sets". In: Computer Vision and Image Understanding 61.1, pp. 70–83 (cit. on p. 17).

Li, Hui and Chunhua Shen (2016). "Reading Car License Plates Using Deep Convolutional Neural Networks and LSTMs". In: arXiv:1601.05610 [cs] (cit. on p. 56).

Matas, J. et al. (2004). "Robust Wide-Baseline Stereo from Maximally Stable Extremal Regions". In: Image and Vision Computing 22.10, pp. 761–767 (cit. on p. 59).

Monasse, P. and F. Guichard (2000). "Fast computation of a contrast-invariant image representation". In: IEEE Transactions on Image Processing 9.5, pp. 860–872 (cit. on p. 5).

Najman, Laurent (2011). "On the Equivalence Between Hierarchical Segmentations and Ultrametric Watersheds". In: Journal of Mathematical Imaging and Vision 40.3, pp. 231–247 (cit. on p. 4).

Najman, Laurent, Jean Cousty, and Benjamin Perret (2013). "Playing with Kruskal: Algorithms for Morphological Trees in Edge-Weighted Graphs". In: International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing. Vol. 7883. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 135–146 (cit. on p. 40).

References II

Neumann, L. and J. Matas (2016). "Real-Time Lexicon-Free Scene Text Localization and Recognition". In: IEEE Transactions on Pattern Analysis and Machine Intelligence 38.9, pp. 1872–1885 (cit. on pp. 28, 59).

Rosenfeld, Azriel and John L. Pfaltz (1966). "Sequential Operations in Digital Picture Processing". In: J. ACM 13.4, pp. 471–494 (cit. on p. 17).

Salembier, P. and L. Garrido (2000). "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval". In: IEEE Transactions on Image Processing 9.4, pp. 561–576 (cit. on pp. 4, 47, 57, 58, 60).

Salembier, P., A. Oliveras, and L. Garrido (1998). "Antiextensive connected operators for image and sequence processing". In: IEEE Transactions on Image Processing 7.4, pp. 555–570 (cit. on p. 5).

Shi, Baoguang, Xiang Bai, and Serge Belongie (2017). "Detecting Oriented Text in Natural Images by Linking Segments". In: arXiv:1703.06520 [cs]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3482–3490 (cit. on p. 8).

Soille, P. (2008). "Constrained connectivity for hierarchical image partitioning and simplification". In: IEEE Transactions on Pattern Analysis and Machine Intelligence 30.7, pp. 1132–1145 (cit. on pp. 4, 46).

Xu, Yongchao, Thierry Géraud, and Laurent Najman (2013). "Salient level lines selection using the Mumford-Shah functional". In: Image Processing (ICIP), 2013 20th IEEE International Conference on. IEEE, pp. 1227–1231 (cit. on p. 39).

— (2016). "Connected Filtering on Tree-Based Shape-Spaces". In: IEEE Transactions on Pattern Analysis and Machine Intelligence 38.6, pp. 1126–1140 (cit. on pp. 31, 33).

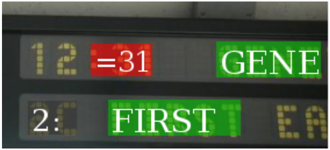
Yin, Xu-Cheng et al. (2014). "Robust Text Detection in Natural Scene Images". In: IEEE Transactions on Pattern Analysis and Machine Intelligence 36.5, pp. 970–983 (cit. on p. 56).

Deep vs Ours: an observation

The detection could critically affect the performance of text detection and recognition methods [Garcia et al., 2000]

Input

Adelaide_ConvLSTMs



Deep2Text II-2

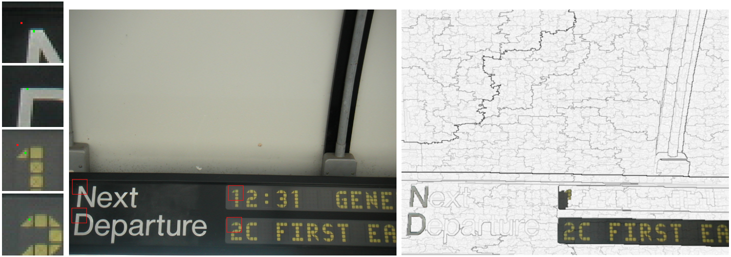
Our detection + Tesseract

- Adelaide_ConvLSTMs [Li et al., 2016] and Deep2Text II-2 [Yin et al., 2014] [Jaderberg et al., 2014] have public results on Robust Reading Competition website (`http://rrc.cvc.uab.es/?ch=2&com=evaluation&task=4`)

Interactive segmentation

- Segmentation with marker [Salembier et al., 2000]:
  - Input: image and a background marker (B) and an object marker (O).
  - Output: Largest image group that intersect (O) but not (B).

Apply on a hierarchies of minimum spanning forests [Cousty.2011.ismm]



Input and 1x1 markers

Results

OCR by Tesseract:

```
Empty page!!
Empty page!!
```

Interactive segmentation

- Segmentation with marker [Salembier et al., 2000]:
  - Input: image and a background marker (B) and an object marker (O).
  - Output: Largest image group that intersect (O) but not (B).

Apply on alpha-tree of Spatial alignment graph

Input and 1x1 markers

Results

OCR by Tesseract:

Next 12:31 GENE
Departure 2: FIRST EH

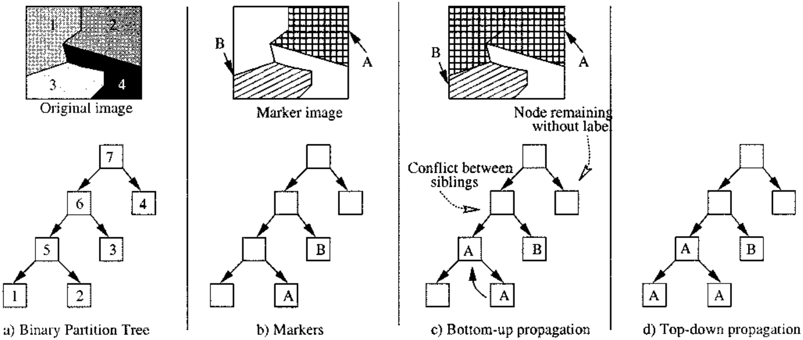
About the text segmentation results

- Stroke Width Transform implementation come from `https://sites.google.com/site/roboticssaurav/strokewidthnokia`
- Text Detection based on Extremal Regions [Matas et al., 2004]; [Neumann et al., 2016], implemented on OpenCV.
- Toggle Mapping Morphological Segmentation [Fabrizio et al., 2016].
- Classical pixel based evaluation [Calarasanu et al., 2016]
- Consistency value measuring how much ground-truth text components are split into several pieces.

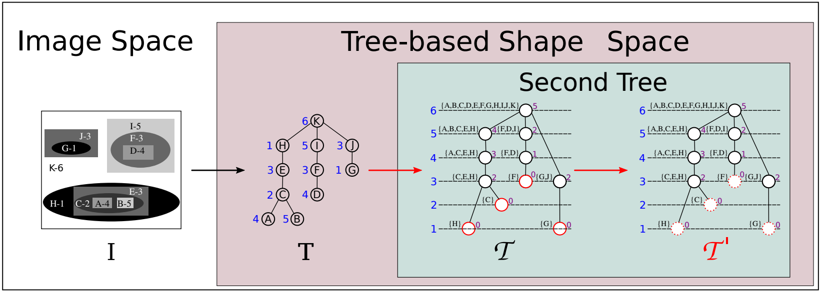
Method	Recall	Precision	F-score	Consistency
SWT [Epshtein et al., 2010]	46.42%	88.61%	60.92%	50.50%
ER [Neumann et al., 2016]	61.31%	89.20%	62.92%	72.67%
TMMS [Fabrizio et al., 2016]	**78.46%**	75.22%	**76.80%**	79.13%
Our	63.62%	**93.36%**	**75.65%**	**84.98%**

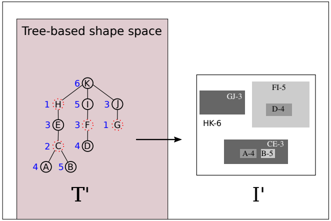
Detail of the marker-based interactive segmentation

- Input: image and a background marker (B) and an object marker (O).
- Output: Largest image group that intersect (O) but not (B).
- The marker could be provided in the shapes space (some shapes was marked) or in image space (shapes whose proper pixels was marked are consider marked).
- The method is described as "marker and propagation" Salembier et al., 2000
  - Both label (B and O) are propagate bottom-up.
  - Propagate top-down from nodes that do not have label conflict.

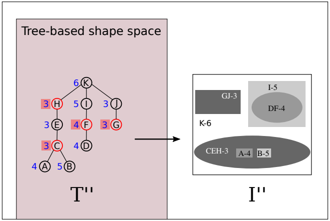


Tree-based shape space original vs proposed tree restitution





Original tree restitution



Proposed tree restitution

Optimization ToSoL construction

- More optimization are detailed in manuscript:
  - The well-composed interpolation
    - Interpolation leads to larger image
    - Local interpolation require computation of actual value

    $\rightarrow$ We define a new non-local self-dual well-composed interpolation that could be easily emulated.
  - The tree simplification step
    - Interesting feature could be extracted from the object contours.

    $\rightarrow$ We allow the filtering during construction of ToSoL by a simple labeling decision.

$\rightarrow$ Our ToSoL is computed and simplified with linear time complexity.

Emulation of well-composed interpolation

- Motivation:
  - Interpolation $\rightarrow$ 4x number of pixels, 3/4 will be removed at the end.
  - Local interpolation compute the actual value, we just ned the sign.

$\rightarrow$ maybe we could emulate the interpolation?

- We could define a nonlocal interpolation with:

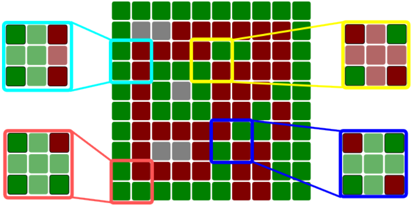
$$f(a_1, a_2 ..., a_n) = \begin{cases} sign(a_1) & sign(a_1) = sign(a_2) ... = sign(a_n) \\ \chi & \exists a_h, a_k; sign(a_h) \neq sign(a_k) \end{cases}$$

where χ is the sign of outside CC

$\longrightarrow$ f is self-dual and in-between.
insert interpolation results

Emulate the emulatable well-composed interpolation

- This interpolation:
  - Follows the same scheme as our ToSL construction,
  - Its topological behavior is deterministic,
  - It could be easily emulated.



- Observe the connectivity of primary pixels:
  - Critical configuration only appears at borders of regions
  - At critical configuration:
    - The outside sign are connected (equ. using C8)
    - The inside sign are disconnected (equ. using C4)
  - At other configuration, C8 and C4 are still equivalent
- Implement by marked which pixels are in the contour and apply corresponding connectivities.

Tree simplification during tree construction.

- We already know the countour of new regions before we label it
- Some interesting properties could be deduced from contour:
  - Its contrast vs its parent by mean gradient magnitude it contour.
  - Its size by the length of its contour, height, width.

  $\rightarrow$ why not filtering the tree during its construction?
- Implement by evaluating the regions before deciding which label to assigned

ToSoL construction