Region Space Analysis

School of Computer Science and Electronics Engineering
Arief Setyanto
Dr. John C Wood, Prof. Mohammed Ghanbary

Outline
- Why segment: we do it
- Salient object isolation: feasible?
- Low level segmentation: practical?
- Metadata from regions: possible?
General Result

(a) Original Video

(b) Simplest Segmentation Result (under segmented)

Introduction

- There are a lot of dots in a digital image
- Pre segmentation reduces required data
- Regions and even objects found using descriptors
- Preservation of boundaries and salient content
- Branches often represent semantic information
- Can be achieved without the need for thresholds
What We do?

- Pre Segmentation
  - watershed
  - SLIC

- Region Merging for Hierarchical Segmentation
  - Colour Based
  - Colour and Motion Direction

- Evolution Analysis
  - Reluctant Merging Detection
  - Region Relative Surround Saliency

- Region Metadata Generation

- Metadata Based Region Query

Object

- Frames/ Images
- Sequence of Frames (Video)
Frame/Image

- A single frame consists of almost millions of **pixels**, for example 720 lines uses 1280 pixels per line (1280 x 720) or 1920x1080.
- Representation – 2D Matrices
- In order to reduce computation, neighbouring pixels which carry similar information groups together to be new unit called region

Video/Sequence of Frames

- Generally immediate consecutive frame share similar information
- In three dimensional representation, video has spatial dimension which are horizontal (x) and vertical (y) axis plus temporal axis (t).
- Every dot picture element in the 3D space so called **voxel**.
- Representation 3D matrices
- Instead of having spatial neighbor's, every voxel has spatial and temporal neighbors.
- According to similarity criteria, the voxels in the 3D space can be grouped and form a volumetric partition or super voxel.
Pre-Segmentation (Spatial)

Original, 288 x 352 pixels

(a) Watershed 2675 regions

(b) SLIC 207 Regions

Region/Super pixel

- Algorithm to produce regions
  - Watershed
  - Mean shift
  - SLIC
- Video analysis using - Region temporal correlation
- Region Descriptor
  - Colour descriptor
  - Colour and neighbourhood descriptor
- Advantage
  - Only need a single frame at every execution
  - Doesn’t need a huge amount of memory
- Disadvantage
  - Need region correlating task
Pre-Segmentation (Spatio Temporal)

Volume/Super voxel of Video

- Method:
  - 3D watershed

- Advantages
  - Doesn’t need to compute region correlation across frame
  - Provide approximation of motion direction for each frame since the pre-segmentation task

- Disadvantage
  - Need a number of frames as an input (usually between 2 cuts)
  - Need a huge number of memory in preprocessing stage
Hierarchical Segmentation

- In the real world there are no single interpretation of a scene.
- One may interpret a face as a single entity while other percept as compound objects, it consist of eyes, nose, lips etc.
- The idea of hierarchical segmentation is keeping the detail information on the lower level while provide generalization on the higher level.
- Our Algorithm record every merging task in a tree. Because every iteration algorithm choose a pair (2 partitions) which have the closest distance (can be colour distance) the result is a binary partition tree.

BPT as Hierarchical Segmentation

Binary Partition Tree

Pre-Segmentation

Partition Labelling

Region/Volume Adjacency Graph (VAG)

Merging

Record Merging History
RAGs for Image

- Region Adjacency Graph

Image Regions

Region Adjacency Graph

Volume Adjacency Graph (VAG)
Merging

- Generally generic segmentation algorithm produce a set of over segmented
- A thousands of tiny partitions obviously mean nothing
- That’s the reason why tiny partitions must be merge in order to obtain the expected object candidates.
- Problem :
  - Which region pair to be merge and when they are merge
  - When the merging must be stopped
- Solution – Identify salient node and propose as salient candidate

Merging

1. Merge the most Similar
2. Issue New Parent Node
3. Update The Volume/Region Adjacency Graph
4. If VAG/RAG is Not Empty
Hierarchical Segmentation on BPT (Image)

Pre-segmentation: Watershed
Initial Nodes: 2675
Total Nodes: 11828
Level: 65

Hierarchical Segmentation on BPT (image)

Pre-segmentation: SLIC
Initial Nodes: 206
Total Nodes: 363
Level: 20
BPT for video (example) Plot Level 2

Simplification

- Using the word ‘simplification’ avoids committing to ‘segmentation’
- A range of tree densities provides a hierarchy for a user to operate in
- Tree densities are controlled by examining the gradients of the graph arcs
- Can be applied to colour size centroid etc. or a combination
Application of Hindsight

- The tree is a documentation of the merging process generated without using thresholds
- An individual path from leaf to root is unique
- The path can be subjected to statistical analysis

The Rule

- On an upward path through the tree a region which is growing consistently is a ‘happy’ region
- If a discontinuity occurs, a reluctance to merge is evident.
- A reluctant event should inhibit further merging
- The same event observed in colour, size and centroid etc. reinforces
BPT Evolution ANALYSIS

- Define all possible path from the lowest leaf toward the root of BPT
  - Path = \{ path1, 1 path2 ..... Pathl \}
  - where l = number of original partition as the result of watershed
- Each path consist of n node starting from the lowest to the root
  - Path_i = \{ Node1, node2 ..... Node_n \} where n is the number of node from certain path (i) from the leaf nodes to the root. n can be vary for individual path
- Observe volume evolution and identify the unhappy merging, which is happen when 2 neighbouring region obliged to be merge while they have big difference. They may belong to different object

Proceed for all possible path

- In our experiment, we choose first, second and last peak from the graph
- First peak result remain over segmented, second more simple and the last peak tend to be under segmented.
- For simplicity reason, all the node under the peak node will be pruned.
Evolution Analysis

- Choose specific leaf nodes from a BPT and form a path, identify the reluctant merge in every merging step.

Plot the evolution in the BPT

Detected reluctant merging
Simplification Result (carphone) – 1\textsuperscript{st} Peak

Carphone – second peak
Carphone – last peak

Soccer – original
Result – (Soccer) – first peak

Soccer – second peak
Soccer – last peak

Test Result

The table below figure out how far the simplification been done by the algorithm.

<table>
<thead>
<tr>
<th>Video</th>
<th>Size</th>
<th>Pre-Segmentation Result</th>
<th>Entire BPT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nodes</td>
<td>Size</td>
<td>Duration</td>
</tr>
<tr>
<td>carphone</td>
<td>4894</td>
<td>48.2</td>
<td>3.3</td>
</tr>
<tr>
<td>foreman</td>
<td>5673</td>
<td>38.9</td>
<td>3.6</td>
</tr>
<tr>
<td>soccer</td>
<td>7352</td>
<td>28.4</td>
<td>3.1</td>
</tr>
<tr>
<td>stefan</td>
<td>9232</td>
<td>16.1</td>
<td>2.7</td>
</tr>
</tbody>
</table>
Simplification Rate

<table>
<thead>
<tr>
<th>Video</th>
<th>peak</th>
<th>Simplified BPT</th>
<th>Average Per Node</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>salient nodes</td>
<td>level</td>
</tr>
<tr>
<td>carphone</td>
<td></td>
<td>39</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>280</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>third</td>
<td>288</td>
<td>93</td>
</tr>
<tr>
<td>foreman</td>
<td></td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>624</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>third</td>
<td>632</td>
<td>64</td>
</tr>
<tr>
<td>soccer</td>
<td></td>
<td>39</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>570</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>third</td>
<td>580</td>
<td>143</td>
</tr>
<tr>
<td>stefan</td>
<td></td>
<td>39</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>second</td>
<td>906</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>third</td>
<td>908</td>
<td>86</td>
</tr>
</tbody>
</table>

TABLE II: Simplification Result of some Test Video

Metadata

- Image processing domain often involve a complex pixel processing
- One of our work aim to shift image processing area into database processing domain.
- Region as a result of segmentation are translated into textual database records.
Task

- Saliency identification
- Shape Identification for salient region
- Metadata Recording
- Retrieval from Metadata

It can be used either by machine or human to query an information by using “SQL Like Syntax” which is extended with some special keyword in order to perform spatial logic operation such as: next to, on the left of etc.

Region to Textual Metadata

<table>
<thead>
<tr>
<th>RegionID</th>
<th>level</th>
<th>Parent</th>
<th>Left</th>
<th>Right</th>
<th>Shape</th>
<th>Colour/Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>2132</td>
<td>0</td>
<td>0</td>
<td>2128</td>
<td>2131</td>
<td>-</td>
<td>Gray</td>
</tr>
<tr>
<td>2128</td>
<td>1</td>
<td>2132</td>
<td>2122</td>
<td>2086</td>
<td>Face</td>
<td>Silver</td>
</tr>
<tr>
<td>2131</td>
<td>1</td>
<td>2132</td>
<td>2127</td>
<td>2130</td>
<td>-</td>
<td>Blue</td>
</tr>
<tr>
<td>2122</td>
<td>2</td>
<td>2128</td>
<td>1994</td>
<td>2115</td>
<td>Triangle</td>
<td>Silver</td>
</tr>
<tr>
<td>2086</td>
<td>2</td>
<td>2128</td>
<td>2024</td>
<td>2057</td>
<td>-</td>
<td>Yellow</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RegionID</th>
<th>Neighbour</th>
<th>Angle</th>
<th>Position</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>2745</td>
<td>2748</td>
<td>23.1707</td>
<td>Up Right</td>
<td></td>
</tr>
<tr>
<td>2745</td>
<td>2752</td>
<td>58.2401</td>
<td>Up Right</td>
<td></td>
</tr>
<tr>
<td>2746</td>
<td>2748</td>
<td>271.924</td>
<td>Bottom</td>
<td></td>
</tr>
<tr>
<td>2746</td>
<td>2752</td>
<td>32.1977</td>
<td>Up Right</td>
<td></td>
</tr>
</tbody>
</table>
Metadata Content

1. Region Identity
2. Level in the pyramidal tree (BPT)
3. Region statistic (colour, size, centroid)
4. Region relative to its parent and child in the tree
5. Region relative to its neighbour in spatial space
6. Region in temporal domain (how long its alive, how far its moving)
7. Region Shape for object candidate only with certain measure (according to 3 and 4)

Extended SQL for textual metadata
Extended Query Language

- Near to
- Next to
- South Neighbour to
- South East Neighbour to
- East Neighbour to
- ...
- Inside

References

References


THANK YOU Very Much

- I need your feedback to improve this research
- Arief Setyanto
  - e-mail: asetya@essex.ac.uk
- John Charles Woods
  - e-mail: woodjt@essex.ac.uk

University of Essex
Wivenhoe Park
Colchester, CO4 3SQ
Essex, UK