Video Segmentation

Video Data is Complex

- Video data segmentation
  - Cannot use 'Tuple' or 'record'
- Content-based Indexing
  - Key-word search alone does not work well.
- Similarity matching vs Exact match
- Browsing mechanism
  - Exploring or browsing mode to confirm similarity matching

Units of Video Data

- Video
- Scenes
- Shots
- Frames

Manual segmentation is very slow.

Semantic Units

- Shot is a sequence of frames recorded in one camera operation.
- Scene is a sequence of semantically related shots.

Shot Boundary Detection (SBD)

- Pairwise Comparison
- Likelihood Ratio Comparison
- Global Histogram
- Local Histogram
- Average Intensity Difference
- Edge Change Ratio
- Background Tracking

Challenges in Designing SBD

Detecting sharp cuts are now quite good, but detecting gradual transitions is not good.

- Fade-out: Gradually darkens the end of a shot to black
- Fade-in: Gradually lightens the shot from black
- Dissolve: A superimposition of the end of shot A and the beginning of shot B.
- Wipe: Shot B replaces shot A by means of a boundary line moving across the screen. The two shots do not blend.
### Pairwise Comparison

- Compare the corresponding pixels in two consecutive frames and count the number of pixels that have changed over a threshold.
- \( P(k,l) \): Intensity value of the pixel at coordinate \((k,l)\) in frame \(i\).

\[
D^P(k,l) = \begin{cases} 
1 & \text{if } |P(k,l) - P_{i-1}(k,l)| > t \\
0 & \text{otherwise}
\end{cases}
\]

\[
\sum^N_{k=1} \sum^M_{l=1} D^P(k,l) > 100 \quad \Rightarrow \quad \text{Cut !}
\]

### Disadvantages

- A shift of few pixels between two consecutive frames can create a cut (e.g., pan or object movements).
- Use smoothing by replacing each pixel with the mean value of its nearest neighbor.

### Likelihood Ratio

- Partition the frame into a number of blocks.
- For each block in a frame, compute \( m_i \) and \( s_i \).

\[
m_i = \frac{\sum_j P(j)}{N} \quad \text{and} \quad s_i = \sqrt{\frac{\sum_j (P(j) - m_i)^2}{N}}
\]

- Compare corresponding regions of two consecutive frames using the following formula.

\[
\text{Cut !} \quad \text{Increment a count } (c).
\]

**Advantage:** Tolerate slow and small object motion from frame to frame.

### Histogram Comparison (1)

\[
HD_i = \sum_{j=1}^{G} |H_i(j) - H_{i-1}(j)| > t
\]

\[
\text{Cut !} \quad \text{Advantage: Less sensitive to object movements than pairwise comparison.}
\]

### Histogram Comparison (2)

- Problem: Histograms of two consecutive frames can be different even though their colors are perceptually similar.

\[
HD_i = |4-0| + |13-4| + |25-38| = 26
\]

Perceptually, \( SD_i = 0 \)

\[
HD_i = |0-300| + |300-0| = 600
\]

Let \( w \) be between 0 and 1.

If bin 0 is similar to bin 1, \( w \) is near 1, \( HD_i = 2^w \cdot |1-1| \).

If bin 0 is different than bin 1, \( w \) is near 0, \( HD_i = 2^w \cdot |300| \).

Adjust histogram values based on color similarity.
Use a perceptually weighted histogram (PWH)

- Each histogram bin is represented by a color value.
- Compute the color distance between each pixel and the representative color for each bin.
- Select only the x most similar bins.
- Weights inversely proportional to the color distance are added to the count of the x most similar bins.

Example:

Pixel values are 10, 11, and 22.
There are two bins: bin 0 (1-10) and bin 1 (11 and up).

Regular histogram:
10 is assigned to bin 0
11 and 22 are assigned to bin 1.

PWH:
Let's representative color of bin 0 be 5 and representative color of bin 1 be 15.
For 10, \(w(1/(10-5))\) is assigned to bin 0 and bin 1.
For 11, \(w(1/(6-0))\) is assigned to bin 0 and \(w(1/(15-11))\) is assigned to bin 1.

Histogram Comparison (3)

\[ G_{ij} = H_i(j) - H_j(i) \]

\[ SD = \sum_{j} \frac{\left| H_i(j) - H_j(j) \right|^2}{H_i(j)} \]

Note: There are many formula for Chi-square tests.

Histogram and Color Images

1. Compute intensities from the intensities of the three color channels.
   \[ I = 0.299R + 0.587G + 0.114B \] [NTSC standard]
2. Use 3D histogram

Handling flashlights

Assumption: When there are flashlights, no more than half of an entire frame is affected.

- Divide a frame into 16 regions.
- Compute the 16 difference values of corresponding regions.
- A camera break is detected when the sum of eight lowest difference values is greater than a threshold.

Limitations of

Pairwise Comparison,
Likelihood ratio,
Histogram Comparison

- Fast moving objects or large moving objects
- Flashing lights
- Flickering objects such as computer monitors.
**Average Intensity Difference**

- Compute the normalized sum of pixel intensity values for each color plane \([R,G,B]\) of frame \(f\), with a size of \(M \times N\).

\[
S_f = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(f(i,j))}{M \times N}
\]

- Evaluate the inter-frame difference between frames \(i-1\), \(i\) and \(i+1\).

\[
D_{ij}^{i+1} = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |I(f(i,j)) - I(g(i,j))|}{M \times N}
\]

- Compute the global cut measure.

\[
D_{cut} = \sum_{p(x,y)} D_{cut}^{i+1}
\]

- \(D_{cut} > \text{threshold} \) \(\Rightarrow\) Cut!

**Edge Change Ratio** (Zabihi'95)

**Assumption:** New edges appear far from old edges and old edges disappear in location far from the new edges.

Use edge detection techniques to find local discontinuities in some image attributes such as intensity or color.

These discontinuities are of interest due to the fact that they are likely to occur at the boundaries of objects.

**Edge Change Ratio (Cont'd)**

- Use Canny edge detection to detect edges in a video frame.
- Compute ECR as follows:

\[
ECR = \frac{\text{Number of entering and exiting edge pixels}}{\text{Total number of edge pixels}}
\]

- ECR > threshold \(\Rightarrow\) Cut!

- Perform better than color histogram:
  \(\rho_e > \rho_w\)
  \(\rho_{in} > \rho_{out}\)
  \(\rho_{cut} > \rho_{in} > \rho_{out}\)

**Shot Detection by Tracking Background** (Oh'00)

- Frames consist of Object(s) and Background.
- While objects change, background does little.

**Fixed Background & Object Areas**

- Cover horizontal camera movement
- Cover vertical camera movement

**How Does FBA Track Camera Movement**
Facilitate Comparison  
(Transform FBA to TBA)

Signature & Sign

• "Gaussian Pyramid" was used to reduce an image to a smaller size.
• We adapted it to reduce TBA and FOA to their "signature" and "sign".

Dimensions

Examples (n = 166)

max = 166
min = 16 
\( \frac{n}{6} \) + 3

Background Tracking (BGT)

Max. number of continuous matches is the score.
How much they share the common background.

Shot Detection Procedure

A Shot Detected by BGT

Edge Change Ratio and Color Histogram incorrectly make six cuts.
**Test Set**

- **BGT scheme**
- **CHD : Color Histogram Difference**
- **ECR : Edge Change Ratio**

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Duration</th>
<th>Status</th>
<th>Method</th>
<th>Shots Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>BGT</td>
<td>Fisherman's Cove</td>
<td>14 votes</td>
<td>3686</td>
<td>1300</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Chicago (Dense/Lines)</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Little Venice</td>
<td>15 votes</td>
<td>1073</td>
<td>130</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Local Park</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>My Little Summer</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Mount (Inland)</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Northwest Beach Park</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
</tbody>
</table>

**Performance Metrics**

- **Recall**: The ratio of the number of shots detected correctly over the actual number of shots.
- **Precision**: The ratio of the number of shots detected correctly over the total number of shots detected.

**BGT is Better**

<table>
<thead>
<tr>
<th>Video</th>
<th># Shots</th>
<th>BGT</th>
<th>CHD</th>
<th>ECR</th>
<th>Shots Correct</th>
<th>Duration</th>
<th>Status</th>
<th>Method</th>
<th>Shots Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fisherman's Cove</td>
<td>14</td>
<td>3686</td>
<td>1300</td>
<td>95</td>
<td>95</td>
<td>14 votes</td>
<td>3686</td>
<td>1300</td>
<td>95</td>
</tr>
<tr>
<td>Chicago (Dense/Lines)</td>
<td>15</td>
<td>107</td>
<td>150</td>
<td>85</td>
<td>85</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
<tr>
<td>Little Venice</td>
<td>15</td>
<td>1073</td>
<td>130</td>
<td>85</td>
<td>85</td>
<td>15 votes</td>
<td>1073</td>
<td>130</td>
<td>85</td>
</tr>
<tr>
<td>Local Park</td>
<td>15</td>
<td>107</td>
<td>150</td>
<td>85</td>
<td>85</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
<tr>
<td>My Little Summer</td>
<td>15</td>
<td>107</td>
<td>150</td>
<td>85</td>
<td>85</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
<tr>
<td>Mount (Inland)</td>
<td>15</td>
<td>107</td>
<td>150</td>
<td>85</td>
<td>85</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
<tr>
<td>Northwest Beach Park</td>
<td>15</td>
<td>107</td>
<td>150</td>
<td>85</td>
<td>85</td>
<td>15 votes</td>
<td>107</td>
<td>150</td>
<td>85</td>
</tr>
</tbody>
</table>