Object Extraction for Video Indexing

Jianping Fan
Department of Computer Science
University of North Carolina at Charlotte
Charlotte, NC 28223
jfanie@uncc.edu
http://www.cs.uncc.edu/~jfanie
1. Why we need video objects?

Necessity is the mother of invention!

Query: show me the videos include president Bush?

All the videos in database should be indexed via objects

For building the database to support this service, we should extract video objects for database indexing
2. What’s the video object?

We, *human being*, may not know the answer exactly.
2. What’s the video object?

Who is who?

You may know them, but your grandma may not!
3. How to detect video object?

a. Feature-based image segmentation

3. How to detect video object?

Edge Detection:

(a) Difference Calculation

\[
\begin{array}{ccc}
1 & 2 & 1 \\
0 & 0 & 0 \\
-1 & -2 & -1
\end{array}
\quad
\begin{array}{ccc}
1 & 0 & -1 \\
2 & 0 & -2 \\
1 & 0 & -1
\end{array}
\quad
\begin{array}{ccc}
2 & 1 & 0 \\
1 & 0 & -1 \\
0 & -1 & -2
\end{array}
\quad
\begin{array}{ccc}
0 & 1 & 2 \\
-1 & 0 & 1 \\
-2 & -1 & 0
\end{array}
\]

Horizontal edge, Vertical edge, Northeastern, Southwestern

\[
HOE(x, y) = |I(x - 1, y - 1) + 2I(x, y - 1) + I(x + 1, y - 1) - I(x - 1, y + 1) - 2I(x, y + 1) - I(x + 1, y + 1)|
\]
3. How to detect video object?

Edge Detection:

(b) Decision

(1) Local Maximum

\[ MOE(x, y) = \max\{HOE(x, y), VOE(x, y), NOE(x, y), SOE(x, y)\} \]

(2) Binary Classification

\[ Y_E(x, y) = \begin{cases} 1, \text{edge \_ pixel}, \ MOE(x, y) \geq T \\ 0, \text{non \_ edge \_ pixel}, \ MOE(x, y) < T \end{cases} \]
3. How to detect video object?
3. How to detect video object?
3. How to detect video object?

b. Similarity-based region growing

3. How to detect video object?

Similarity-based region growing

(a) Neighboring Similarity Calculation

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$(x-1, y-1)$</td>
<td>$(x, y-1)$</td>
<td>$(x+1, y-1)$</td>
</tr>
<tr>
<td>$(x-1, y)$</td>
<td>$(x, y)$</td>
<td>$(x+1, y)$</td>
</tr>
<tr>
<td>$(x-1, y+1)$</td>
<td>$(x, y+1)$</td>
<td>$(x+1, y+1)$</td>
</tr>
</tbody>
</table>
3. How to detect video object?

Similarity-based region growing

(b) Decision

(1) Similarity Distance

\[ D[(x, y), (x-1, y-1)] = |I(x, y) - I(x-1, y-1)| \]

(2) Binary Classification

\[
S(x, y) = \begin{cases} 
1, & \text{similar}, \quad D[(x, y), (x-1, y-1)] < T \\
0, & \text{dissimilar}, \quad D[(x, y), (x-1, y-1)] \geq T 
\end{cases}
\]
3. How to detect video object?
3. How to detect video object?
3. How to detect video object?

Computers can detect homogeneous regions or edges based on calculable features.

*How we can let computer know the semantic objects?*

Face detection
Object Seed for Human Object: Faces

Resources:

a. Human faces always have different colors from other objects: skin color.

b. Regions for human face have special outline shape: Ellipse

c. The components for human faces: eyes, noses, mouth (they have different color from skin color and they are detected as holes in face region) should have special relationships.
Skin Color Map Generation

a. Select dataset for human faces
Skin Color Map Generation

b. Mark the skin area for human face
Skin Color Map Generation

b. Mark the skin area for human face
Skin Color Map Generation

c. Generate the probability for each color can be taken as skin

$$P_{\text{skin}}(u,v) = \frac{q(u,v)}{Q(u,v)}$$
Automatic Image Segmentation

Edge detection → Region growing
Automatic Image Segmentation

Edge detection

Region growing
Skin Color Region Determination

a. Test each pixel in the same homogeneous region

\[ SCL(x, y) = P_{skin}(u(x, y), v(x, y)) \]

b. Test average skin-like for a homogeneous region

\[ \beta_i = \frac{\sum_{(x, y) \in R_i} SCL(x, y)}{\text{Total number of Pixels}} \]
Skin Color Region Determination

c. Decision making

\[
\begin{aligned}
\beta_i &> T \quad \text{skin region} \\
\beta_i &< T \quad \text{non skin region}
\end{aligned}
\]
a. Aspect Ratio Filter

\[
\text{Aspect ratio} = \frac{|X_{\text{right}} - X_{\text{left}}|}{|Y_{\text{down}} - Y_{\text{top}}|}
\]
Face Filters

b. Density Ratio Filter

\[ Density\_ratio = \frac{N\_\text{face}\_\text{pixel}}{|Y_{down} - Y_{top}| \parallel X_{right} - X_{left}|} \]
c. Shape Filter: Ellipse like?

\[
\frac{(x - x_c)^2}{R_1} + \frac{(y - y_c)^2}{R_2} = R
\]
d. Relationship filters

\[ D_{nose\_left\_eye} = D_{nose\_right\_eye} \]

\[ D_{mouth\_left\_eye} = D_{mouth\_right\_eye} \]
3. How to detect video object?

**Object Model: Knowledge Of Object Type**

Put your knowledge in the programming!

4. How to track object among frames?

a. Change Detection (Motion Detection)

Noise + change induced by object motion
4. How to track object among frames?

b. Reduce the influence from noise

4. How to track object among frames?

c. Reduce the influence from uncovered background

4. How to track object among frames?

d. Motion estimation

However, tracking is a traditional problem for computer vision.
Object Extraction: Summary

Edge Detection  Image Segmentation

Integration

Object Seed Detection

Semantic Object Generation

Contour-Based Temporal Tracking

Temporal Tracking

Image Input
Segmentation: Edges & Regions

Face Detection

Region Aggregation

Object Extraction
5. Semi-automatic video object extraction

a. Why we need this technique?

- It is hard for current computer vision techniques to obtain semantic objects automatically.

- On the other hand, it is easy for human beings to find the semantic object from the complex scene according to their knowledge.

- It is a reasonable solution: human users define the semantic objects in the video via a human-computer interaction procedure, and then the computer tracks the objects among frames automatically.
5. Semi-automatic video object extraction
5. Semi-automatic video object extraction
5. Semi-automatic video object extraction

- Active Contour Tracking

\[
E(S) = \int_{0}^{1} \left( E_{\text{int}}(S(u)) + E_{\text{img}}(S(u)) + E_{\text{ext}}(S(u)) \right) du
\]

- \( S(u) \): representation of contour
- \( E_{\text{int}} \): energy of contour to control its smoothness
- \( E_{\text{img}} \): energy of contour to control image-based similarity such as color-based or edge-based similarity
- \( E_{\text{ext}} \): energy of contour to control external similarity such as human perception
5. Semi-automatic video object extraction

\[ E_{\text{internal}} = \alpha(s) \left| \frac{dv}{ds} \right|^2 + \beta(s) \left| \frac{dv^2}{ds^2} \right|^2 \]  

\[ u(s) = (x(s), y(s), z(s)) \]  

\[ E_{\text{image}} = \omega_{\text{line}} E_{\text{line}} + \omega_{\text{edge}} E_{\text{edge}} + \omega_{\text{term}} E_{\text{term}} \]  

\[ E_{\text{line}} = f(x, y) \]  

\[ E_{\text{edge}} = |\nabla f(x, y)|^2 \]
5. Semi-automatic video object extraction

How to do it?
5. Semi-automatic video object extraction
5. Semi-automatic video object extraction

b. How to do it?
5. Semi-automatic video object extraction

b. How to do it?
5. Semi-automatic video object extraction

b. How to do it?
5. Semi-automatic video object extraction
How to Track Moving Objects: Part-Based Approach
Application of Techniques
Application of Techniques
6. Video Object Representation

a. Color histogram

![Graph showing color histograms for (0,0,0) and (110,112,80)].

- 3-D Column 1
- R
- G
- B
6. Video Object Representation

b. Texture: directional edge histogram
6. Video Object Representation

c. Motion: trajectory of moving object

\[(x, y, z, t)\]

Position + time
6. Video Object Representation

d. Object shape:

(1) boundary-based approach
6. Video Object Representation

(2) Density-based approach: Rectangular box of object
Density of object pixels
Feature-Based Video Content Representation

- **Color**
  - HSV color histogram,
  - dominant color, ...

- **Texture**
  - Edge histogram,
  - Tamura, ....

- **Shape**
  - Rectangular box,
  - moments, .....

- **Motion**
  - Trajectory, motion
  - histogram, ...

- **Other features**
Advantages for Object-Based Video Content Representation
Which Objects We Should Care?

- Domain-Dependent
- Application-Dependent
- User-Dependent
- Visual Attention Models
Video Attention Models

Salient Region Detection
Video Attention Models

Salient Region Detection
Video Attention Models

Bottom-up Attention Models

Gabor Pyramid + Orientation Filters

Subtract low-res (3-4 octaves) from higher res

Normalize \((0..1)\)

map \(* (1 - \max_{ave})^2\)

add maps

Average Maps

Inhibition + Excitation
Video Attention Models

Salient Region Detection
Video Attention Models

Salient Region Detection
Video Attention Models
VideoQ at Columbia University

classes

objects

regions

features: Color, texture, shape

region 1

Object 1

Class

Object n

Region m

Region i

classification

grouping

segmentation
How to answer query by example in VideoQ?

- Color
- Texture
- Motion
- Shape
- Size

Tc, Tt, Tm, Th, Ts

objects

VideoQ at Columbia University
Automatic Video Annotation via Faces

CNN Headline News
Name-It System in CMU

- Detect Face from News Videos
- Detect Keywords from Closed Captions
Name-It System in CMU
7. Object-based Video Access

Query: show the videos in the database include this object

(find the papers include this key word)

Obtain the features from your query example

Compare it with that of objects in database

Query results presentation and streaming
7. Object-based Video Access
7. Object-Based Video Access
(b)
8. Who is working on this topic?

1. Philips research center
2. IBM research center
3. Microsoft research at Redmond
4. NEC research institute
5. Xerox Palo Alto research center
6. Sarnoff Corp.
7. Thompson
8. INRIA, France
9. ATR, Japan
9. Requirement

If you are asked to do video object extraction, the basic requirements are:

a. Video processing techniques:
   segmentation, motion estimation, ...

b. Color space & human perception

c. Video coding standards:
   MPEG-1, MPEG-2, MPEG-4, MPEG-7

d. Skills on C++