A Sketch-Based Approach To Video Retrieval Using Qualitative Features

Koustav Ghosal and Anoop Namboodiri

Center for Visual Information Technology
IIIT Hyderabad, India
Content Based Video Retrieval

Different Form Of Queries:

Text: Example Sketch

Text Based Query:

Video Search from Youtube

“Coke Studio Pakistan”
“Markov random Field Daphne Koller”
“Rohit Sharma’s Double Century”

Metadata: Tags, Comments etc.
Content Based Video Retrieval

Different Form Of Queries:
- Text
- Example
- Sketch

Example Based Query:
Video Google

Input:
Image or Objects of interest

Output:
Relevant Shots or Key Frames

1. Sivic et al, ICCV 2003
Content Based Video Retrieval

Different Form Of Queries:
- Text
- Example
- Sketch

Sketch vs Text/Example based

“A diving video where the diver does a somersault before going down … ”

“A pool shot where the player strikes from below and the white ball moves north-east, then north west then south-west and finally strikes another ball. The second ball drops inside the pocket … ”

Example is not always available.
Sketch vs Text/Example based
Challenges

Background Extraction
- Camera Motion
- Occlusion
- Illumination
- Background Noise
- and many more…

Perceptual Variability

Users perceive the same motion differently
Qualitative vs Quantitative Feature Space

**Approximations** instead of absolute numeric values to remove perceptual variability

**Qualitative Spatio-Temporal Features**
Features that capture “how” (qualitative) objects moved rather than “how much” (quantitative)
• The Problem Statement: Modelling the query (user sketch) and original trajectories using “qualitative” features instead of “quantitative” features to remove perceptual variability.

• Assumption: Trajectories are available for a set of videos.
Aspects of Motion

- **Shape**: Linear and non-linear sub-trajectories.
- **Direction**: Similar in shape but have different directions.
- **Scale**: Different scales of same motion.

S-A-I-O
S-C-I-O
S-C-O-I
**Shape : Circle Based Features**

\[
J = \min_{x_0, y_0, r} \sum_{i=1}^{n} x_i^2 + y_i^2 - 2x_0x_i - 2y_0y_i + x_0^2 + y_0^2 + r^2
\]

\[
S = (x_\mu, y_\mu, r, m, s)
\]

\[
(x_\mu, y_\mu), r = \text{Center, Radius}
\]

\[
m = \text{Slope}
\]

\[
s = \text{Normalized Length}
\]

**K- Means , Bag-Of-Motion**

**Trajectory = Histogram of motion-segments = Loss of temporal Information !!**

1. Sivaprasad et al, ICDAR 2013
Direction And Scale

\[ \text{Trajectory} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \]  
\[ \alpha_k = \sin(\theta_k) \quad \ldots (1) \]

\[ \text{Distance from mean} \]

\[ \text{Trajectory} = (d_1, d_2, \ldots, d_n) \]  
\[ d_k = \quad \text{Distance from mean} \quad \ldots (2) \]
Summary of Features

1. Bag-Of-Motion

2. Ordered Motion Segments $(s_1, s_2, \ldots, s_m)$ where $s_k = (x_\mu, y_\mu, r, m, s)$

3. Change of Direction $(\alpha_1, \alpha_2, \ldots, \alpha_n)$

4. Change of Scale $(d_1, d_2, \ldots, d_n)$

Both For User-Sketches and Videos
Dataset

- 100 Synthetic videos and 100 pool videos

- Data was collected from 50 different users
- Interval of 6 hours between video and sketch
Cascade
Results: An example query

BOM
Sequence of BOM
Sequence of Directions
Sequence of Scales
Results

Precision - Recall Top K Accuracy

Pool

Synthetic

Precision - Recall

Top K Accuracy
Results

Mean Reciprocal rank

\[ MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i} \]

Pool

Synthetic
Summary

• Strengths
  ▪ **Unconstrained Query** : No initial frame
  ▪ **Qualitative Features** : Robust to user-level variations
  ▪ **A novel retrieval strategy** : Cumulative Scoring Mechanism

• Limitations
  ▪ **Dependency** : A strong trajectory extraction algorithm
  ▪ **Not generic** : Doesn’t work for videos where motion is not the most salient feature
Future Work

• Finding new features to expand the scope of this representation to more generic videos

• Using additional information about object color and shape to refine the search

• Generating consistent tracks from videos with camera motion and large perspective variations
Thank you