School of Engineering and Computer Science The Hebrew University of Jerusalem

ARIEL TAL ERAN WEISSENSTERN

Supervisor: GAL ELIDAN

MOVIE WEAVER

Automatic Movie Editor

1. Introduction

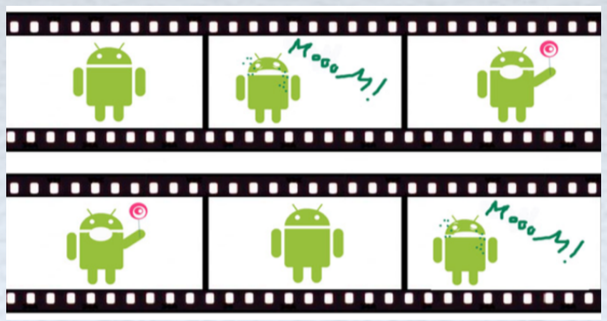
The Problem

- The world is full of video –
- Most video is taken by non-professionals
  - Clips are longer snapshots
  - No time or skill to compose a movie
- An autonomous composition tool is needed

Key Observations

- Notice that context creates meaning
  - Kuleshov's experiment (1920)
- Composing starts with pairs of clips

Context creates meaning



Same clips - Different movies

Challenges

- Hard to define the "best" sequence
- Clip classification
- Selecting a subset of the clips
- Determining the sequence

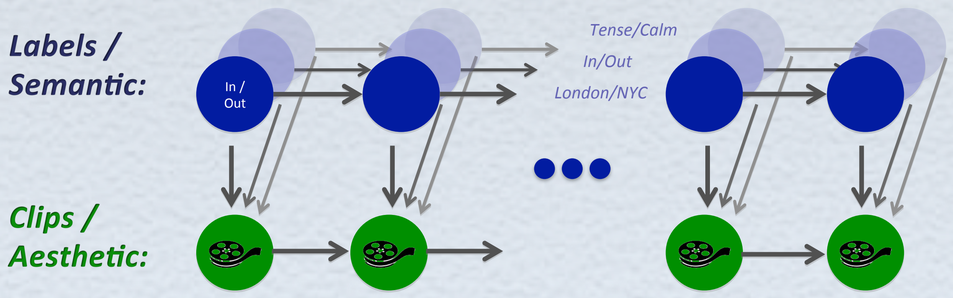
Goals

- Incorporate Cinematic Principles
  - Low-level consistency – Colors, Brightness
  - Rules of the *Invisible Cut*
  - Plot – Actors, Locations
- Computationally Feasible

2. Materials and Methods

1. Solution in a Nutshell

- *Probabilistic Model* over composition of clips
- Define semantic properties of a clip using probabilities
  - Indoor / Outdoor
  - Actors
  - Locations
- Define aesthetic similarity to other clips
- Use inference to find the *most probable clip-sequence*



2. Problem Definition

$$\max_{\mathbb{C},\mathbb{H}^M} P(X_1, H_1^1, ..., H_1^M) \cdot P(X_2, H_2^1, ..., H_2^M | X_1, H_1^1, ..., H_1^M) \cdot \ldots$$

$$\ldots \cdot P(X_K, H_K^1, ..., H_K^M | X_{K-1}, H_{K-1}^1, ..., H_{K-1}^M)$$

$$= \max \sum_{i=2}^{K} \bar{w}_0 \bar{f}_0(X_i, X_{i-1}, H_i) + \sum_{i=2}^{K} \bar{w}_1 \bar{f}_1(H_i^1, H_{i-1}^1) + \ldots + \sum_{i=2}^{k} \bar{w}_M \bar{f}_M(H_i^M, H_{i-1}^M)$$

$$= \max \sum_{i=2}^{K} \bar{w}_0 \bar{f}_0(X_i, X_{i-1}, H_i) + \sum_{m=1}^{M} \left[\sum_{i=2}^{K} \bar{w}_m \bar{f}_m(H_i^m, H_{i-1}^m) \right]$$

- C – The input set of clips
- H – A set of classifiers of size M
- K – Clip position in the final sequence
- Assuming initial probabilities are uniform

3. Encoding Semantic Behavior

Why?

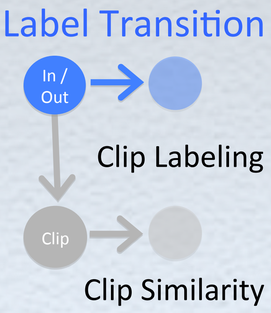
- Humans classify what they see in a semantic level – Indoor vs. Outdoor, Character appearance, Geographical
- Prioritize semantics by their cinematic importance

How?

- Encode transition probabilities between labels
- Indoor-Outdoor transition can encode a stable movie (1), or an hectic movie (2)

(1)
$$\begin{array}{c} \\ in \\ out \end{array} \begin{array}{c} \begin{array}{cc} in & out \end{array} \\ \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix} \end{array}$$

(2)
$$\begin{array}{c} \\ in \\ out \end{array} \begin{array}{c} \begin{array}{cc} in & out \end{array} \\ \begin{pmatrix} 0.1 & 0.9 \\ 0.8 & 0.2 \end{pmatrix} \end{array}$$



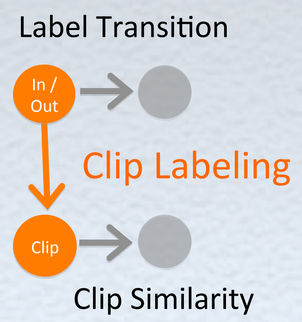
4. Semantic Classification

Why?

- Semantic classification of clips helps us get the feeling of the clip
- Such classifications are easy to percept as humans, and are the building blocks of a professional editor

How?

- We trained classifiers on a tagged set of movies
- Using these classifiers we determine the *affinity* of a clip to each label



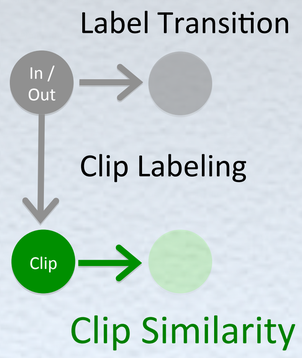
5. Encoding Clip Similarity

Why?

- Encapsulating all the aesthetic properties of a clip – Color Histograms, Visual Symmetry
- Metric to pair-wise clip matching

How?

- Compute the low-level attributes of each clip
- Create a similarity function over pairs of clips



3. Results

(1) Weights aiming at an indoor sequence of clips:

(2) A random sequence

Strengths

- Scalable
  - Semantic classifiers can be added easily – Tempo, Geographical, etc.
  - Fast
- Allows several levels of abstraction - low-level similarity up to a semantic-based narrative
- User preferences are projected as probabilities

Weaknesses

- Tends to repeat selected clip sequences
- The result may not always correspond with the user's preferences

4. Future Work

- Use movies to learn the model parameters
- Learned model quantifies difference between movies – MTV vs. BBC
- Use the model to generate different styled movies –
  - "My trip to India – National Geographic style"
  - "Grandpa's birthday – MTV style"

5. References

- A. Oliva, A. Torralba - "Modeling the shape of the scene"
- S. Eisenstein – "A Dialectic Approach to Film Form", 1949
- K. Murphy - "The Bayes Net Toolbox for Matlab."
- BBC Motion Gallery

6. Acknowledgements

- Eyal Soreq – Dept. Of Screen Based Arts - Bezalel
- Michael Fink – Hebrew University, Google