Detection of Manipulation Action Consequences (MAC)

Yezhou Yang, Cornelia Fermüller, Yiannis Aloimonos
Computer Vision Lab, University of Maryland, College Park

**Why Consequences?**
A fundamental concept in understanding manipulation actions, are the action consequences. There is a small set of fundamental primitive consequences that provides a systematic high-level categorization of manipulation actions.

**Fundamental Types**
Six fundamental categories are: DIVIDE, ASSEMBLE, CREATE, CONSUME, TRANSFER, and DEFORM. Each of these categories is represented by Visual Semantic Graph (VSG):

1. **DIVIDE**: one object breaks into two objects, or two attached objects break the attachment;
2. **ASSEMBLE**: two objects merge into one object, or two objects build an attachment between them;
3. **TRANSFER**: an object is moved from one location to another location;
4. **DEFORM**: an object has an appearance change;
5. **CREATE**: an object is brought to, or emerges in visual space;
6. **CONSUME**: an object disappears from visual space.

**Continuous Object Monitoring**
A new tool for continuously monitoring objects under manipulation has been developed. The approach combines an Attention mechanism, Foreground-background segmentation and Stochastic tracking. The first MAC dataset (MAC 1.0) was compiled, and experiments that evaluate the detection of consequences were conducted.

**Weighted Graph Cut**
A weighted graph-cut method:
- The unary terms are defined on the tracked points on the basis of their color.
- Every sample is weighted by comparing with the fixation point.
- The binary terms are defined on all points on the basis of edge information.

\[ Q(f) = \sum_{x \in X} U_x(l_x) + \lambda \sum_{(x,y) \in \Omega} V_{x,y}(l_x, l_y). \]

**Active Tracking**
An active monitoring process:
- Tracking the object to obtain temporal correspondence.
- Segmenting the object to obtain its topological structure and appearance model for the VSG.

**Consequence Detection**
Experiment 1: Detecting the various consequences. A consequence is detected using the VSG change.

**Video Classification**
Experiment 2: Video classification based on different consequences.

**More Results**

**References**

**Acknowledgements**