Causal Video Object Segmentation From Persistence of Occlusions

Brian Taylor, Vasiliy Karasev, Stefano Soatto

Presented by Sehyun Joo

All images from paper
Contents

• Problem Statement
• Problem Analysis
• Approach
• Results
• Comparison with previous work
Last presentation:
Object segmentation

Graph Structure for Object Proposal Selection: Binary Edges

SegTrack Dataset

Region within the red boundary is the object region
Problem Statement

Video object segmentation using occlusion
  • With some depth information of occlusion

input: video
Challenges

(A.B) -> <-(C,D,E)  \hspace{1cm}  (A.B) -> (C,D,E) <- (E)

Modeling assumptions :
\begin{itemize}
  \item Based on surrounding differentiations.
\end{itemize}
The motion of two objects generate occlusions and disocclusions.
• $C(x) = 0$ : background
• $C(x) = 1, 2, ...$ : foreground layers with depth value
Background: Detachable Object Detection

$\mathbf{c}_t = \arg\min_{c_t: c_t \geq 0} \int_D g_t(x) |\nabla c_t(x)| dx$

s.t. $c_t(y^c) - c_t(y) \geq 1 \ \forall (y^c, y) \in O_t.$

- Proposes some detachable objects from images.

Once an object, always an object
Once an object, always an object

\[ c_t(x) \geq 1 \ \forall x \in F, \quad F = \{ c_{t-1}(w_t^{t-1}(x)) \geq 1 \} \]

The set of all pixels ever labeled as foreground
Once an object, always an object

\[ \int_D \kappa_t(x) \max(0, 1 - c_t(x)) \, dx \]

 Gets bigger if region is in F

*Foreground prior
Occlusion cue aggregation

\[ \sum_{i=1}^{N} \lambda_i \max \left( 0, 1 - c_t(y_i^c) - c_t(y_i) \right), \]

\[ \lambda_{t,i} = m_t(y_i^c) \lambda_{t-1,i} + 1 \{ c_{t-1}(w_{t-1}^{t-1}(y_i^c)) \geq c_{t-1}(w_{t-1}^{t-1}(y_i)) \}. \]

Weight function: decreased with large motion

Aggregating all the occlusion objects in the cue
Overall model

$$
c_t = \arg \min_{c_t \geq 0} \int_D g_t(x) |\nabla c_t(x)| dx + \tau \int_D c_t(x) dx \\
+ \int_D \kappa_t(x) \max (0, 1 - c_t(x)) dx \\
+ \sum_{i=1}^N \lambda_i \max (0, 1 - c_t(y_i^c) - c_t(y_i)), \\
(y_i^c, y_i) \in \tilde{O}_t
$$

Persistent layer boundary weights  
Layer regularization

Foreground prior

Aggregation of occlusion cue
Results
Comparison: segmentation

new colors correspond to new objects
Comparison:
foreground & background
Comparison: multilable
Conclusion

• We can segment objects in ‘video’ using the feature of occlusion.

• Still, if the image doesn’t contain enough movements, it’s very easy to fail.

• Segmenting many object simultaneously is also very hard problem.
Thank you!