Temporal AND-OR Graph
for representation and recognition of Events, Actions, Motions

Song-Chun Zhu, Sinisa Todorovic, and Ales Leonardis

At CVPR, Providence, Rhode Island
June 16, 2012
Goal: Recognize events in daily scenes

For example, an office.

Ref. Pei, Si and Zhu, ICCV2011.
Challenges

1. Events happen over an extended time period
   • Variant time-span
   • Could be interrupted
   • Multiple routes
   • Intention and prediction

2. Actions are hard to recognize
   • Subtle and similar
   • No salient motion/pose at most of the time
   • Contextual objects -- key!!
Overview of our approach

Spatial image parsing

- Background objects
  - Water collector
  - Trash can
  - Water boiler
  - Phone
  - Tea canister

Video

Temporal event parsing

- EVENTS
  - Refill tea
  - Renew tea leaf
  - Refill water
  - make phone call

- Spatial image parsing
- Temporal event parsing

- Arrival
- Bend &
- Leave

- Trash can
- Tea can
- Boiler

- Arrive trashcan
- Bend &
- Leave trashcan
- Interior

- Arrive tea can
- Grab tea can
- Leave tea can

- Arrive boiler
- Bend &
- Leave boiler

- Water collector
- Trash can
- Water boiler
- Phone
- Tea canister

- Arrive trashcan
- Bend &
- Leave trashcan
- Interior

- Arrive tea can
- Grab tea can
- Leave tea can

- Arrive boiler
- Bend &
- Leave boiler

Background objects

Events

Water collector
Trash can
Water boiler
Phone
Tea canister

Spatial image parsing

Temporal event parsing

Events

Refill tea
Renew tea leaf
Refill water

Arrive trashcan
Bend &
Leave trashcan
Arrive tea can
Grab tea can
Leave tea can
Arrive boiler
Bend &
Leave boiler

Water collector
Trash can
Water boiler
Phone
Tea canister

Arrive trashcan
Bend &
Leave trashcan
Arrive tea can
Grab tea can
Leave tea can
Arrive boiler
Bend &
Leave boiler

Overview of our approach

Spatial image parsing

Video

Temporal event parsing

Events

Refill tea
Renew tea leaf
Refill water

Arrive trashcan
Bend &
Leave trashcan
Arrive tea can
Grab tea can
Leave tea can
Arrive boiler
Bend &
Leave boiler

Water collector
Trash can
Water boiler
Phone
Tea canister

Arrive trashcan
Bend &
Leave trashcan
Arrive tea can
Grab tea can
Leave tea can
Arrive boiler
Bend &
Leave boiler

Overview of our approach

Spatial image parsing

Video

Temporal event parsing

Events

Refill tea
Renew tea leaf
Refill water

Arrive trashcan
Bend &
Leave trashcan
Arrive tea can
Grab tea can
Leave tea can
Arrive boiler
Bend &
Leave boiler

Water collector
Trash can
Water boiler
Phone
Tea canister

Arrive trashcan
Bend &
Leave trashcan
Arrive tea can
Grab tea can
Leave tea can
Arrive boiler
Bend &
Leave boiler
Scene parsing
How to define actions and events?

Some neurons in the pre-motor area encode actions

Mirror neurons firing when performing action or seeing other people performing the action

Gallese et al 96, Rizzolatti et a. 96
### Actions = Spatiotemporal relations between body parts and objects in the scene

<table>
<thead>
<tr>
<th>Atomic Actions</th>
<th>Fluent</th>
<th>Foreground Symbols</th>
<th>Background Symbols</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shake Hands (P1, P2)</td>
<td>Near(P1,P2) and Touch (P1.hand, P2.hand)</td>
<td><img src="image1.png" alt="Diagram 1" /></td>
<td><img src="image2.png" alt="Image 2" /></td>
<td><img src="image3.png" alt="Image 3" /></td>
</tr>
<tr>
<td>Use Dispenser (P3)</td>
<td>Bend(P3) and Near(P3, A) and Touch(P3.hand, A)</td>
<td><img src="image4.png" alt="Diagram 2" /></td>
<td><img src="image5.png" alt="Image 5" /></td>
<td><img src="image6.png" alt="Image 6" /></td>
</tr>
<tr>
<td>Pick up Phone (P4)</td>
<td>Touch(P4, B) and On(B)</td>
<td><img src="image7.png" alt="Diagram 3" /></td>
<td><img src="image8.png" alt="Image 8" /></td>
<td><img src="image9.png" alt="Image 9" /></td>
</tr>
</tbody>
</table>

Some of the learned atomic actions by pursuing the co-occurrence of relations.

Actions = Spatiotemporal relations between body parts and objects in the scene

<table>
<thead>
<tr>
<th>Unary Relations</th>
<th>Binary Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sign of person</strong></td>
<td><strong>Binary Relations</strong></td>
</tr>
<tr>
<td>Stand(P1)</td>
<td>Fluent (A,B)</td>
</tr>
<tr>
<td>Stretch(P1)</td>
<td>Touch (A,B)</td>
</tr>
<tr>
<td>Bend (P1)</td>
<td>Near (A,B)</td>
</tr>
<tr>
<td>Sit (P2)</td>
<td>Oclude (A,B)</td>
</tr>
<tr>
<td>Status of objects</td>
<td>In(A,B)</td>
</tr>
<tr>
<td>On (phone)</td>
<td><strong>Examples</strong></td>
</tr>
<tr>
<td>Off (phone)</td>
<td></td>
</tr>
<tr>
<td>On(screen)</td>
<td></td>
</tr>
<tr>
<td>Off(screen)</td>
<td></td>
</tr>
</tbody>
</table>

High-order relations: - E.g., surrounded by
Event as temporal And-Or-Graph

And-node
Or-node
Leaf-node

Refill tea

Prepare the cup

Renew tea leaf

Refill water

Dump old tea

Dump water

Null

Arrive trashcan
Bend down
Leave trashcan
Arrive collector
Bend down
Leave collector

Get close to tea can

Grab tea can

Leave tea-can

Arrive boiler
Bend down
Leave boiler
Event as temporal And-Or-Graph

- **And-node**
- **Or-node**
- **Leaf-node**

Make phone call

- Get close to phone
- Pick up receiver
- Talking over phone
- Put down receiver

- Get close to phone
- Pick up phone
- Stand still
- Walk around
- Put down receiver

10
Formulation

\[ p(g) = \frac{1}{Z} \exp\{\text{score}(g)\} \]

\[ \text{score}(g) = \sum_{v_t \in T(g)} \lambda_{v_t} \alpha(v_t) + \sum_{v \in V_0(g)} \lambda_v \omega(v) + \sum_{(i,j) \in E(g)} \lambda_{ij} r_{ij}(v_i, v_j) \]
Formulation

\[ p(g) = \frac{1}{Z} \exp\{\text{score}(g)\} \]

**Grammar**

\[ \sum_{v \in V_0(g)} \lambda_v \omega(v) \]

**Temporal Relations**

\[ \sum_{(i,j) \in E(g)} \lambda_{ij} r_{ij}(v_i, v_j) \]

**Data term**

\[ \text{score}(g) = \sum_{v_t \in T(g)} \lambda_{v_t} \alpha(v_t) \]

\[ \alpha(v_t) = \sum_{i \in F} \beta_i, h_i(v_t) - \text{dist}(P_{\text{person}}, P_{\text{obj}}) \]
Parsed process (Earley Parser [Earley 1970])

- Refill tea
  - Empty cup
    - Dump old tea
      - Arrive trashcan
        - Bend down
      - Leave trashcan
    - Arrive collector
      - Bend down
    - Leave collector
  - Null
    - Renew tea leaf
      - Get close to tea can
      - Grab tea can
      - Leave tea-can
    - Arrive boiler
      - Bend down
    - Leave boiler

- Refill water
  - Dump water
    - Renew tea leaf
    - Leave collector
  - Refill water

And-node
Or-node
Leaf-node
Refill tea

Empty cup

Renew tea leaf

Refill water

Dump old tea

Dump water

Null

Get close to tea can

Grab tea can

Leave tea-can

Arrive boiler

Bend down

Leave boiler

Arrive trashcan

Bend down

Leave trashcan

Arrive collector

Bend down

Leave collector

Null

Grab tea can

Leaf-node

And-node

Or-node

Parsing process (Earley Parser [Earley 1970])

Arrive trashcan

Bend down

Leave trashcan

Arrive collector

Bend down

Leave collector

NULL

Grab tea can

Leaf-node

And-node

Or-node

Leaf-node

Earley 1970

14
Refill tea

Empty cup

Renew tea leaf

Refill water

Dump old tea

Arrive trashcan, Bend down, Leave trashcan

Arrive collector, Bend down, Leave collector

Dump water

Get close to tea can

Grab tea can

Leaf-node

Renew tea leaf

Leaf-node

Refill water

Arrive boiler, Bend down, Leave boiler

Null

Leaf-node

Parsing process (Earley Parser [Earley 1970])
Parsing process (Earley Parser [Earley 1970])

- And-node
- Or-node
- Leaf-node

- Refill tea
- Empty cup
- Renew tea leaf
- Refill water
- Dump old tea
- Dump water
- Null
- Arrive trashcan
- Bend down
- Leave trashcan
- Arrive collector
- Bend down
- Leave collector
- Get close to tea can
- Grab tea can
- Leave tea-can
- Arrive boiler
- Bend down
- Leave boiler
Refill tea

Empty cup

Renew tea leaf

Null

Refill water

Dump old tea

Arrive trashcan

Bend down

Leave trashcan

Arrive collector

Bend down

Leave collector

Dump water

Get close to tea can

Grab tea can

Leave tea-can

Null

Renew tea leaf

Arrive boiler

Bend down

Leave boiler

Parsing process (Earley Parser [Earley 1970])
Refill tea

Empty cup

Dump old tea

Dump water

Null

Renew tea leaf

Get close to tea can

Grab tea can

Leave tea-can

Arrive boiler

Bend down

Leave boiler

Arrive trashcan

Bend down

Leave trashcan

Arrive collector

Bend down

Leave collector

Parsing process (Earley Parser [Earley 1970])
Parsing process (Earley Parser [Earley 1970])

And-node
Or-node
Leaf-node

Refill tea
Renew tea leaf
Refill water

Dump old tea
Dump water
Null

Arrive trashcan → Bend down → Leave trashcan
Arrive collector → Bend down → Leave collector
NULL

Get close to tea can → Grab tea can → Leave tea-can
Arrive boiler → Bend down → Leave boiler
Refill tea
Empty cup
Renew tea leaf
Refill water

Dump old tea
Arrive trashcan
Bend down
Leave trashcan

Null
Get close to tea can
Grab tea can
Leave tea-can

Arrive boiler
Bend down
Leave boiler

Parsing process (Earley Parser [Earley 1970])
Parsing process (Earley Parser [Earley 1970])

- Arrive trashcan
- Bend down
- Leave trashcan
- Arrive collector
- Bend down
- Leave collector
- Null
- Get close to tea can
- Grab tea can
- Leave tea-can
- Arrive boiler
- Bend down
- Leave boiler
- Refill tea
- Renew tea leaf
- Refill water
Parsng process (Earley Parser [Earley 1970])

And-node
Or-node
Leaf-node

Refill tea
Empty cup
Renew tea leaf
Null
Refill water

Dump old tea
Arrive trashcan
Bend down
Leave trashcan

Dump water
Arrive collector
Bend down
Leave collector

Get close to tea can
Grab tea can
Leave tea-can

Null
Grab tea can
Arrive boiler
Bend down
Leave boiler

And-node
Or-node
Leaf-node
Parsing process (Earley Parser [Earley 1970])

- And-node
- Or-node
- Leaf-node

Activities:
- Arrive trashcan
- Bend down
- Leave trashcan
- Arrive collector
- Bend down
- Leave collector
- Null
- Get close to tea can
- Grab tea can
- Leave tea-can
- Arrive boiler
- Bend down
- Leave boiler
- Refill tea
- Renew tea leaf
- Dump old tea
- Dump water
- Renew tea leaf
- Refill water
Parsing process (Earley Parser [Earley 1970])

And-node
Or-node
Leaf-node

Dump old tea
Arrive trashcan → Bend down → Leave trashcan

Dump water
Arrive collector → Bend down → Leave collector

Null
Get close to tea can

Renew tea leaf
Grab tea can → Leave tea-can

Refill tea

Refill water
Arrive boiler → Bend down → Leave boiler
Parsing process (Earley Parser [Earley 1970])

And-node
Or-node
Leaf-node

Refill tea
Empty cup
Renew tea leaf
Refill water

Dump old tea
Arrive trashcan
Bend down
Leave trashcan

Dump water
Arrive collector
Bend down
Leave collector

Null
Get close to tea can
Grab tea can

Renew tea leaf
Leave tea-can

Refill water
Arrive boiler
Bend down
Leave boiler
Parsing process (Earley Parser [Earley 1970])

And-node
Or-node
Leaf-node

Refill tea
Empty cup
Renew tea leaf
Refill water

Dump old tea
Dump water
Null

Arrive trashcan
Bend down
Leave trashcan
Arrive collector
Bend down
Leave collector
NULL
Get close to tea can
Grab tea can
Leave tea-can
Arrive boiler
Bend down
Leave boiler
Parsing process (Earley Parser [Earley 1970])

And-node

Or-node

Leaf-node

Refill tea

Empty cup

Renew tea leaf

Refill water

Dump old tea

Dump water

Null

Get close to tea can

Grab tea can

Leave tea-can

Arrive trashcan

Bend down

Leave trashcan

Arrive collector

Bend down

Leave collector

NULL

Arrive boiler

Bend down

Leave boiler
Parsing: A modified Earley parser [Earley 1970]
Intention and prediction

Time 1

Refill tea

Make phone call

![Diagram showing intention and prediction with two branches: Refill tea and Make phone call. Each branch has nodes representing actions, with arrows indicating intention and prediction.]
Intention and prediction

Time 2

Refill tea

Make phone call

Grab tea can
Pick up receiver

intention
prediction

Refill tea Make phone call

Grab tea can Pick up receiver

30
Handle event interruption

- And-Node
- Or-Node
- Leaf-Node

- First Partial parse tree of take water
- Parse tree of take a phone
- Second Partial parse tree of take water

Observed Data

31
Demo
Comparison with human prediction

![Graph comparing goal inference with judge points]

**Mean of participants**

1. Refill tea
2. Phone call

**Algorithm**

1. Refill tea
2. Phone call


related work: [Baker, Saxe and Tenenbaum 2009]
Computation complexity of parsing

- Initially the number of interpretations above a threshold grows rapidly over time.

- At certain decisive moments, i.e. when informative actions are observed, large number of unlikely interpretation drops below the threshold and hence is pruned.

Pickup phone  Reach water boiler
Weakly Supervised Learning of Temporal AND-OR Graph

Stochastic activity has a random number of:

- actors,
- activity parts,
- spatiotemporal configurations
Examples: Activities with Stochastic Structure
Temporal AND-OR Graph

- AND nodes = Particular space-time configurations
- OR nodes = Alternative configurations
- Terminal nodes = BoWs
Temporal AND-OR Graph

\[ S(C) = 0.5(0.4x_1 P_1 + 0.2\overline{x}_1(1 - P_1)) + \cdots \]

Posterior:

\[ P(X|C) = \frac{S(C)}{S_{X=1}} \]

OR nodes:

\[ S_i(C) = \sum_{j \in i^+} w_{ij} S_j(C) \]

AND nodes:

\[ S_k(C) = \prod_{l \in k^+} S_l(C) \]
Learning – Variational EM

1. Learn AND-OR graph structure parameters – $W$

2. Learn Counting Grid parameters – $\mathbf{\pi}$

\[
V = \sum_t \left[ \sum_b Q_b \log \left( \frac{(w_{ib1}x_b^t - w_{ib2}\overline{x}_b^t)}{Q_b} \right) + \sum_b Q_b \sum_z (c_{bz}^t + \theta_z - 1) \log \left( \sum_{u \in H_b} \pi_{uz} \right) \right],
\]

\[
Q_b \propto \exp \left[ \sum_{t,z} (w_{ib1}x_b^t - w_{ib2}\overline{x}_b^t)(c_{bz}^t + \theta_z - 1) \log \left( \sum_{u \in H_b} \pi_{uz} \right) \right].
\]
Bottom up/Top Down Most Probable Explanation

MPE: $\hat{a} = \arg\max_{a \in A} \hat{S}(C; \alpha)$
Results –Volleyball Dataset
Results – Volleyball Dataset