Cost-Sensitive Top-down/Bottom-up Multi-scale Activity Recognition

Mohamed R. Amer\textsuperscript{1}, Dan Xie\textsuperscript{2}, Mingtian Zhao\textsuperscript{2}, Sinisa Todorovic\textsuperscript{1}, and Song-Chun Zhu\textsuperscript{2}

\textsuperscript{1}Oregon State University, Corvallis OR
\textsuperscript{2}University of California, Los Angeles CA

ECCV 2012
Problem – Given

- High-resolution, long video of a large scene
- People engaged in individual actions and group activities
Problem – Goal

Answer WHAT, WHERE, and WHEN queries about individual actions and group activities
Contributions

• Multi-scale activity recognition
  – Jointly addressing activities at different scales

• Cost-Sensitive Inference

• New Dataset
  – High resolution video
  – Allows for digital zoom-in and zoom-out
  – Many co-occurring individual and group activities
Prior Work – Punctual/Repetitive Activities

• Single Actor

Lan et al ICCV11, Rodriguez et al. CVPR08, Kovashka & Grauman CVPR10
Laptev et al. ICCV03, ICCV07, Dollar et al. VS-PETS05, Blank et al. ICCV05 ...

• Single Group

Lan et al PAMI11, Ryoo & Aggarwal ICCV09, Ryoo ICCV11, Choi et al CVPR11
Amer & Todorovic ICCV11, CVPR12 ...
Prior Work – Structured Activities

Gupta et al CVPR09

Ryoo et al ICCV09, Pei et al ICCV11, Brendel et al CVPR11....
Our Approach

• **Unified hierarchical model of:**
  – People and the objects they interact with
  – Individual actions
  – Group activities

• **Cost-sensitive zooming-in/-out for:**
  – Fusing visual cues at different scales
  – Answering: What, where, when
Our Approach – Related Prior Work

Wu & Zhu IJCV11
Model: And-Or Graph

\[ \mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{P}) \]
Model: And-Or Graph

\[ G = (\mathcal{V}, \mathcal{E}, \mathcal{P}) \]

\( \mathcal{V} \): Graph nodes \( (\mathcal{V}_{NT}, \mathcal{V}_T) \)

\( \mathcal{V}_{NT} \): Non-terminal nodes such as A, R, O

\( \mathcal{V}_T \): Terminal nodes such as \( t(A) \), \( t(R) \), \( t(O) \)
Model: And-Or Graph

\[ G = (V, E, P) \]

\( V \): Graph nodes \((V_{NT}, V_T)\)

\( V_{NT} \): Non-terminal nodes such as A, R, O

\( V_T \): Terminal nodes such as t(A), t(R), t(O)

\( E \): Graph edges \((E_{rel}, E_{dec}, E_{switch})\)

\( E_{rel} \): Relation edges

\( E_{dec} \): Decomposition edges

\( E_{switch} \): Switching edges
Model: And-Or Graph

\[ G = (\mathcal{V}, \mathcal{E}, \mathcal{P}) \]

\( \mathcal{V} \) : Graph nodes \((\mathcal{V}_{\text{NT}}, \mathcal{V}_{\text{T}})\)

\( \mathcal{V}_{\text{NT}} \) : Non-terminal nodes such as A, R, O

\( \mathcal{V}_{\text{T}} \) : Terminal nodes such as \(t(A), t(R), t(O)\)

\( \mathcal{E} \) : Graph edges \((\mathcal{E}_{\text{rel}}, \mathcal{E}_{\text{dec}}, \mathcal{E}_{\text{switch}})\)

\( \mathcal{E}_{\text{rel}} \) : Relation edges

\( \mathcal{E}_{\text{dec}} \) : Decomposition edges

\( \mathcal{E}_{\text{switch}} \) : Switching edges

\( \mathcal{P} \) : Probability over all parse graphs
Model: And-Or Graph

\[ W = (K, \{pg_k : k = 1, 2, \ldots, K\}) \]

\[ p(W) = p(K) \prod_{k=1}^{K} p(pg_k) \]

\[ p(pg) = \frac{1}{Z} \exp(-E(pg)) \]

\[ E(pg) = -\sum_l \left[ \sum_{(\vee^l, \wedge^l) \in \mathcal{E}_{\text{switch}}(pg)} \log p(\wedge^l | \vee^l) \right. \]
\[ + \sum_{(\wedge^l, \wedge^{l-}) \in \mathcal{E}_{\text{dec}}(pg)} \log p(X_{\wedge^l} | X_{\wedge^{l-}}) \]
\[ + \sum_{(\wedge_i^{l+}, \wedge_j^{l+}) \in \mathcal{E}_{\text{rel}}(pg)} \log p(X_{\wedge_i^{l+}}, X_{\wedge_j^{l+}}) \]
Inference

\[ W^* = \arg \max_{W \in \Omega} p(W)p(I_A | W) \]

\[ p(W) = p(K) \prod_{k=1}^{K} p(pg_k) \]

\[ p(I_A | W) = q(I_A) \prod_{k=1}^{K} \frac{p(I_{A_{pg_k}} | pg_k)}{q(I_{A_{pg_k}})} \]
Inference

\[ pg^* = \arg \max_{pg \in \Omega(pg)} \left[ \log p(pg) + \log \frac{p(I_{\Lambda pg} | pg)}{q(I_{\Lambda pg})} \right] \]

\[ p(pg) = \frac{1}{Z} \exp(-E(pg)), \quad Z = \sum_{pg} \exp(-E(pg)) \]

\[ E(pg) = -\sum_l \left[ \sum_{(\lor^l, \land^l) \in \mathcal{E}_{\text{switch}(pg)}} \log p(\land^l | \lor^l) \right. \]
\[ + \sum_{(\land^l, \land^l-\lor^l) \in \mathcal{E}_{\text{dec}(pg)}} \log p(X_{\land^l} | X_{\land^l-\lor^l}) \]
\[ + \sum_{(\land^l, \land^l+\lor^l) \in \mathcal{E}_{\text{rel}(pg)}} \log p(X_{\land^l} , X_{\land^l+\lor^l}) \]

\[ \frac{p(I_{\Lambda pg} | pg)}{q(I_{\Lambda pg})} = \sum_{t \in \mathcal{V}_T(pg)} \log \frac{p(I_{\Lambda t} | t)}{q(I_{\Lambda t})} \]
\[
\begin{aligned}
\text{Inference} \\
pg^* &= \arg \max_{pg \in \Omega(pg)} \sum_l \left\{ \log p(\land^l | \lor^l) \\
& \quad + \log \frac{p(t^\land_l | t)}{q(t^\land_l)} \right\} \\
& \quad \text{No zoom} \\
& \quad + \log \frac{p(t^\land_{l-} | t)}{q(t^\land_{l-})} + \log p(X^\land_l | X^\land_{l-}) \quad \text{zoom-out} \\
& \quad + \sum_{i=1}^{N_l} \left[ \log p(X^\land_{i+} | X^\land_i) + \log \frac{p(t^\land_{i+} | t)}{q(t^\land_{i+})} + \sum_{i \neq j} \log p(X^\land_{i+}, X^\land_{j+}) \right] \quad \text{zoom-in}
\end{aligned}
\]
Inference: Structure

\[ p_{g}^{*} = \arg \max_{p_{g} \in \Omega(p_{g})} \sum_{l} \left\{ \log p(\land^{l}|\lor^{l}) \right\} \]

\[ + \log \frac{p(t_{\land^{l}}|t)}{q(t_{\land^{l}})} \]

No zoom

\[ + \log \frac{p(t_{\land^{l-}}|t)}{q(t_{\land^{l-}})} + \log p(X_{\land^{l}}|X_{\land^{l-}}) \]

zoom-out

\[ + p(N^{l}) \sum_{i=1}^{N^{l}} \left[ \log p(X_{\land^{l+}}|X_{\land^{l}}) + \log \frac{p(t_{\land^{i+}}|t)}{q(t_{\land^{i+}})} + \sum_{i \neq j} \log p(X_{\land^{l+}, i}, X_{\land^{l+}, j}) \right] \]

zoom-in

\[ p(\land^{l}|\lor^{l}) \text{ : is the probability of an And node given a parent Or node} \]
Inference: $\alpha$ – Process

$$pg^* = \arg \max_{pg \in \Omega(pg)} \sum_l \left\{ \log p(\land^l | \lor^l) + \log \frac{p(t_\land^l | t)}{q(t_\land^l)} \right\}$$

For No zoom:

$$+ \log \frac{p(t_\land^l | t)}{q(t_\land^l)}$$

For zoom-out:

$$+ \log \frac{p(t_\land^l | t)}{q(t_\land^l)} + \log p(X_{\land^l} | X_{\land^l-})$$

For zoom-in:

$$+ p(N^l) \sum_{i=1}^{N^l} \left[ \log p(X_{\land^l} | X_{\land^l}) + \log \frac{p(t_{\land^l+i+} | t)}{q(t_{\land^l+i+})} + \sum_{i \neq j} \log p(X_{\land^l+i+}, X_{\land^l+j+}) \right]$$

$p(N^l)$: is an exponential prior over the number of children
Inference: $\beta$ – Process

$$pg^* = \arg \max_{pg \in \Omega(pg)} \sum_l \left\{ \log p(\bigwedge^l | \bigvee^l) + \log \frac{p(t_{\bigwedge^l} | t)}{q(t_{\bigwedge^l})} \right\}$$

No zoom

$$+ \log \frac{p(t_{\bigwedge^l} | t)}{q(t_{\bigwedge^l})} + \log p(X_{\bigwedge^l} | X_{\bigwedge^{l-}})$$

zoom-out

$$+ p(N^l) \sum_{i=1}^{N^l} \left[ \log p(X_{\bigwedge^l_i} | X_{\bigwedge^l}) + \log \frac{p(t_{\bigwedge^l+i} | t)}{q(t_{\bigwedge^l+i})} + \sum_{i \neq j} \log p(X_{\bigwedge^l_i}, X_{\bigwedge^l_j}) \right]$$

zoom-in

$p(N^l)$: is an exponential prior over the number of children

$p(X_{\bigwedge^l_i}, X_{\bigwedge^l_j})$: is the $\beta$-process, the probability of binding two children
Inference: $\gamma$ - Process

\[ pg^* = \arg \max_{pg \in \Omega(pg)} \sum_{l} \\left\{ \log p(\land^l | \lor^l) \right\} \]

\[ + \log \frac{p(t_{\land^l} | t)}{q(t_{\land^l})} \]

No zoom

\[ + \log \frac{p(t_{\land^l} | t)}{q(t_{\land^l})} + \log p(X_{\land^l} | X_{\land^l-}) \]

zoom-out

\[ + p(N^l) \sum_{i=1}^{N^l} \left[ \log p(X_{\land^l} | X_{\land^l}) + \log \frac{p(t_{\land^l+} | t)}{q(t_{\land^l+})} + \sum_{i \neq j} \log p(X_{\land^l}, X_{\land^l+}) \right] \]

zoom-in

$p(N^l)$ : is an exponential prior over the number of children

$p(X_{\land^l} | X_{\land^l-}), p(X_{\land^l+} | X_{\land^l})$ : are the $\gamma$-processes, a child's likelihood given its parent
Inference – $\alpha, \beta, \gamma$ Processes

• $\alpha$: running a detector of the activity

• $\beta$: bottom-up binding of parts of the activity

• $\gamma$: top-down prediction of parts from the activity
α – Process

• Group Activities:
  – Space-Time Volume (STV)

• Primitive Actions:
  – Motion (STIP-HOG)/Appearance (KLT)

• Objects:
  – Deformable Part-based Model (DPM)
β, γ – Process

• β and γ processes are modeled as Gaussian distributions over location, scale and orientation.
\[ \beta - \text{Process: } p(X_{\land i}^{t+}, X_{\land j}^{t+}) = N(X_{\land i}^{t+} - X_{\land j}^{t+}; \mu_{\beta i}, \Sigma_{\beta i}) \]
\( \gamma - \text{Process} \)

\[ p(X_{\wedge i} | X_{\wedge l}) = N(X_{\wedge i} - X_{\wedge l}; \mu_{\gamma l}, \Sigma_{\gamma l}) \]
Cost-Sensitive Inference

• Reinforcement Learning based Inference
  – Explore/Exploit strategy
  – Q-Learning to learn the optimal moves
Explore/Exploit

# of detectors left = 7

$p(pg^{(t)}) = 0$
Explore/Exploit

Q Table for $A_1$

<table>
<thead>
<tr>
<th>(Exploit) $\alpha_1+$</th>
<th>(Exploit) $\alpha_2+$</th>
<th>(Explore) $\alpha_3+$</th>
<th>(Explore) $\alpha_4+$</th>
</tr>
</thead>
</table>

# of detectors left = 7

$p(p_{g(t)}) = 0$
Explore/Exploit

\[ \alpha_1^+ + \alpha_2^+ + \alpha_3^+ + \alpha_4^+ \]

Q Table for A

\begin{tabular}{|l|}
\hline
(Exploit) \( \alpha_1^+ \) \\
(Explore) \( \alpha_2^+ \) \\
(Explore) \( \alpha_3^+ \) \\
(Explore) \( \alpha_4^+ \) \\
\hline
\end{tabular}

# of detectors left = 6
\( p(pg^{(t+1)}) = 0.2 \)
\( p(pg^{(t)}) = 0 \)
Explore/Exploit

Q Table for $\alpha_1$

(Exploit) $\alpha_5^{-}$
(Explore) $\alpha_6^{+}$
(Explore) $\alpha_7^{-}$
(Explore) $\alpha_8^{-}$
(Explore) $\alpha_9^{+}$

# of detectors left = 6
$p(pg^{(t)}) = 0.2$
Explore/Exploit

Q Table for $\alpha_1$

- (Exploit) $\alpha_5^-$
- (Explore) $\alpha_6^+$
- (Explore) $\alpha_7^-$
- (Explore) $\alpha_8^-$
- (Explore) $\alpha_9^+$

# of detectors left = 5

$p(p_{g^{(t+1)}})=0.2$

$p(p_{g^{(t)}})=0.2$
Explore/Exploit

# of detectors left = 5
\[ p(pg(t)) = 0.2 \]

Q Table for \( \alpha_1 \)

- (Exploit) \( \alpha_5^- \)
- (Explore) \( \alpha_6^+ \)
- (Explore) \( \alpha_7^- \)
- (Explore) \( \alpha_8^- \)
- (Explore) \( \alpha_9^+ \)
Explore/Exploit

# of detectors left = 4
\[ p(pg_{(t+1)}) = 0.4 \]
\[ p(pg_{(t)}) = 0.2 \]
Explore/Exploit

\[ \alpha_{12} - \alpha_{7} - \alpha_{11} + \alpha_{10} + \alpha_{1} + A_{1} \]

Q Table for \( \alpha_7 \)

<table>
<thead>
<tr>
<th></th>
<th>(Exploit) ( \alpha_{10}^+ )</th>
<th>(Explore) ( \alpha_{11}^+ )</th>
<th>(Explore) ( \alpha_{12}^- )</th>
</tr>
</thead>
</table>

\# of detectors left = 4
\[ p(pg^{(t)}) = 0.4 \]
Explore/Exploit

\[
\begin{align*}
\alpha_{12}^- & \quad \alpha_{11}^+ \quad \alpha_{1}^+ \\
\alpha_{7}^- & \quad A_1 \\
\end{align*}
\]

\# of detectors left = 3
\[p(pg^{(t+1)})=0.4\]
\[p(pg^{(t)})=0.4\]

Q Table for \(\alpha_7\)
- (Exploit) \(\alpha_{10}^+\)
- (Explore) \(\alpha_{11}^+\)
- (Explore) \(\alpha_{12}^-\)
Explore/Exploit

Q Table for $\alpha_7$
- (Exploit) $\alpha_{10}^+$
- (Explore) $\alpha_{11}^+$
- (Explore) $\alpha_{12}^-$

# of detectors left = 3
$p(pg^{(t)})=0.4$
Explore/Exploit

# of detectors left = 2
$p(pg^{(t+1)}) = 0.4$
$p(pg^{(t)}) = 0.4$
Explore/Exploit

Q Table for $\alpha_7$

| (Exploit) $\alpha_{10}+$ |
| (Explore) $\alpha_{11}+$ |
| (Explore) $\alpha_{12}-$ |

$\alpha_{12}-$ 
$\alpha_7-$ 
$\alpha_7+$ 
$\alpha_{10}+$ 
$\alpha_{11}+$ 
$\alpha_{12}+$

$\# \text{ of detectors left} = 2$

$p(pg^{(t)})=0.4$
Explore/Exploit

Q Table for $\alpha_7$

- (Exploit) $\alpha_{10}+$
- (Explore) $\alpha_{11}+$
- (Explore) $\alpha_{12}-$

$\alpha_{12}-$ $\rightarrow$ $\alpha_{7}-$ $\rightarrow$ $\alpha_1+$ $\rightarrow$ $A_1$

$\alpha_{10}+$ $\rightarrow$ $\alpha_{11}+$ $\rightarrow$ $\alpha_1+$

$\alpha_{10}+$ $\rightarrow$ $\alpha_{11}+$ $\rightarrow$ $\alpha_1+$

# of detectors left = 1

$p(p_{g_{(t+1)}}) = 0.5$

$p(p_{g_{(t)}}) = 0.4$
Explore/Exploit

Q Table for $\alpha_{10}$

<table>
<thead>
<tr>
<th>Action</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{13}+$</td>
<td>(Exploit)</td>
</tr>
<tr>
<td>$\alpha_{14}+$</td>
<td>(Explore)</td>
</tr>
<tr>
<td>$\alpha_{15}+$</td>
<td>(Explore)</td>
</tr>
<tr>
<td>$\alpha_{16}+$</td>
<td>(Explore)</td>
</tr>
</tbody>
</table>

$\alpha_{10} + \alpha_{13} + \alpha_{14} + \alpha_{15} + \alpha_{16}$

$\# \text{ of detectors left} = 1$

$p(p_g(t)) = 0.5$
Explore/Exploit

Q Table for $\alpha_{10}$

<table>
<thead>
<tr>
<th>(Exploit) $\alpha_{13}$+</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Explore) $\alpha_{14}$+</td>
</tr>
<tr>
<td>(Explore) $\alpha_{15}$+</td>
</tr>
<tr>
<td>(Explore) $\alpha_{16}$+</td>
</tr>
</tbody>
</table>

# of detectors left = 0
$p(pg^{(t+1)})=0.6$
$p(pg^{(t)})=0.5$
Explore/Exploit

\[ p(pg^*) = 0.6 \]
Q-Learning

• States: $\mathcal{S} = \{s\}$
  - Query
  - Current node in the And-Or graph

• Moves: $\mathcal{M} = \{m\}$
  - Run detectors applicable to the current state

• Reward: $\mathcal{R}$
  - Reward the move that increments the log posterior
    $$R_t(s, m; q) = \frac{1}{1 + \exp\left(-\left(\log p(\mathbf{g}_t | \mathcal{M}) - \log p(\mathbf{g}_t | \mathcal{M} \cup \{m\})\right)\right)}$$

• Transitions: Deterministic simulator.
Varying the Explore/Exploit trade-off
Varying the Number of Detectors Run

E² strategy for answering the query about Walking

Ground Truth

Precision and recall

Number of detectors run
New Dataset

• Footage: 106min
• Frame Rate: 30 fps
• Resolution: 2560x1920 pixels
• Annotations:
  – Group (activities, formation)
  – Individual (actions, poses, facing direction)
  – Objects
Domain Knowledge

• 6 Group Activities:
  – Walking together, Queuing, Campus tour, ...

• 10 Individual Actions:
  – Walking, Sitting, Riding a bike, ...

• 17 Objects:
  – Food truck, Vending machine, Bike, Backpack, ...
## Available Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Resolution</th>
<th>Object</th>
<th>Individual</th>
<th>Group</th>
<th>Background</th>
<th>Instances</th>
<th>Poses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Dataset</td>
<td>2560x1920</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Cluttered</td>
<td>7+</td>
<td>Yes</td>
</tr>
<tr>
<td>VIRAT Ground</td>
<td>1920x1080</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Cluttered</td>
<td>4-</td>
<td>No</td>
</tr>
<tr>
<td>CompCollective</td>
<td>1440x960</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Cluttered</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>Collective</td>
<td>720x480</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Cluttered</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>UT-Interaction</td>
<td>720x480</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Clear</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>KTH</td>
<td>160x120</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Clear</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>Weizmann</td>
<td>180x144</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Clear</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>UCF Youtube</td>
<td>240x500</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Cluttered</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>UCF 50</td>
<td>240x500</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Cluttered</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>Olympic Sports</td>
<td>360x450</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Cluttered</td>
<td>1</td>
<td>No</td>
</tr>
</tbody>
</table>
Queries Example
All Parse Graphs for Group Queries
All Parse Graphs for Individual Queries
Results – Courtyard Dataset

### Query about group activities

<table>
<thead>
<tr>
<th>$E^2$ strategy</th>
<th>Standing-in-line</th>
<th>Guided-tour</th>
<th>Discussing</th>
<th>Sitting</th>
<th>Walking</th>
<th>Waiting</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B = 1$, Precision</td>
<td>62.2%</td>
<td>63.7%</td>
<td>68.1%</td>
<td>65.3%</td>
<td>69.4%</td>
<td>61.2%</td>
<td>5s</td>
</tr>
<tr>
<td>$B = 1$, FP</td>
<td>7.2%</td>
<td>2.3%</td>
<td>9.8%</td>
<td>12.6%</td>
<td>8.1%</td>
<td>10.4%</td>
<td>5s</td>
</tr>
<tr>
<td>$B = 15$, Precision</td>
<td>65.4%</td>
<td>66.1%</td>
<td>69.0%</td>
<td>68.7%</td>
<td>70.3%</td>
<td>66.5%</td>
<td>75s</td>
</tr>
<tr>
<td>$B = 15$, FP</td>
<td>10.1%</td>
<td>4.7%</td>
<td>11.1%</td>
<td>11.1%</td>
<td>8.7%</td>
<td>10.9%</td>
<td>75s</td>
</tr>
<tr>
<td>$B = \infty$, Precision</td>
<td>68.0%</td>
<td>70.2%</td>
<td>75.1%</td>
<td>71.4%</td>
<td>78.6%</td>
<td>72.6%</td>
<td>230s</td>
</tr>
<tr>
<td>$B = \infty$, FP</td>
<td>13.6%</td>
<td>10.3%</td>
<td>17.1%</td>
<td>13.7%</td>
<td>10.1%</td>
<td>12.2%</td>
<td>230s</td>
</tr>
</tbody>
</table>

### Query about primitive actions

<table>
<thead>
<tr>
<th>$E^2$ strategy</th>
<th>Walk</th>
<th>Wait</th>
<th>Talk</th>
<th>Drive Car</th>
<th>Ride S-board</th>
<th>Ride Scooter</th>
<th>Ride Bike</th>
<th>Read</th>
<th>Eat</th>
<th>Sit</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B = 1$, Precision</td>
<td>63.3%</td>
<td>61.2%</td>
<td>58.4%</td>
<td>65.8%</td>
<td>63.5%</td>
<td>60.1%</td>
<td>56.8%</td>
<td>55.3%</td>
<td>60.9%</td>
<td>54.3%</td>
<td>10s</td>
</tr>
<tr>
<td>$B = 1$, FP</td>
<td>12.1%</td>
<td>16.2%</td>
<td>11.4%</td>
<td>3.4%</td>
<td>10.2%</td>
<td>11.6%</td>
<td>6.2%</td>
<td>8.2%</td>
<td>2.2%</td>
<td>5.3%</td>
<td>10s</td>
</tr>
<tr>
<td>$B = 15$, Precision</td>
<td>67.6%</td>
<td>63.4%</td>
<td>62.3%</td>
<td>67.2%</td>
<td>67.1%</td>
<td>65.9%</td>
<td>59.3%</td>
<td>61.2%</td>
<td>66.3%</td>
<td>59.2%</td>
<td>150s</td>
</tr>
<tr>
<td>$B = 15$, FP</td>
<td>14.2%</td>
<td>17.1%</td>
<td>15.1%</td>
<td>7.1%</td>
<td>13.8%</td>
<td>13.2%</td>
<td>9.3%</td>
<td>10.3%</td>
<td>4.3%</td>
<td>7.1%</td>
<td>150s</td>
</tr>
<tr>
<td>$B = \infty$, Precision</td>
<td>69.1%</td>
<td>67.7%</td>
<td>69.6%</td>
<td>70.2%</td>
<td>71.3%</td>
<td>68.4%</td>
<td>61.4%</td>
<td>67.3%</td>
<td>71.3%</td>
<td>64.2%</td>
<td>330s</td>
</tr>
<tr>
<td>$B = \infty$, FP</td>
<td>18.7%</td>
<td>20.2%</td>
<td>17.9%</td>
<td>9.7%</td>
<td>17.1%</td>
<td>16.3%</td>
<td>12.3%</td>
<td>12.1%</td>
<td>7.7%</td>
<td>9.0%</td>
<td>330s</td>
</tr>
</tbody>
</table>
Conclusion

• New problem of Multi-scale activity recognition.

• Efficient formulation using And-Or graphs

• Cost-sensitive inference using RL

• New dataset
ACKNOWLEDGMENTS

MSEE FA 8650-11-1-7149

MURI N00014-10-1-0933
Questions