Video sequences: ground truthed data and their quality

B. Fisher, S. Blunsden
School of Informatics
Univ. of Edinburgh

ISSUES:
+ NEW BEHAVE PROJECT GROUND TRUTHED DATA
+ ACCURACY IN CAVIAR VIDEO GROUND TRUTH LABELLING

FUNDING: EC CAVIAR PROJECT IST 2001 37540
EPSRC BEHAVE PROJECT (GR/S98146)
BEHAVE Dataset Overview

Theme: Interacting groups
Group Sizes: 2-5 people
Scenarios: 10

25 FPS, 640x480, 60,000 marked up frames
AVI + JPEGs of frames
Ground plane homography data

Bounding boxes around people (VIPER XML)
VIPER based labeling

2838 page accesses (11/12/07)
BEHAVE Dataset Example (Fight)

Ground truthed data and their quality
BEHAVE Dataset Scenarios

InGroup - in group and not moving much
Approach - Two people or groups approaching
WalkTogether - People walking together
Meet - Two or more people meeting
Split - Two or more people separating
Ignore - Ignoring each other
Chase - One group chasing another
Fight - Two or more groups fighting
RunTogether - The group is running together
Following - A person being followed
Ground Truth

Text file

ID1    ID2    Start   End    Label
[2]    [0,1]   ;60296  ;60349  ;Approach

Group ID1 with person 2 is APPROACHED by Group ID2 with persons 0 & 1 during frames 60296-60349.
EVALUATING GROUND-TRUTHING

WHY? COMPARING PROGRAM RESULTS WITH GROUND TRUTH

WHAT IS TYPICAL QUALITY OF HUMAN LABELED GROUND TRUTH?

DETECTIONS, GEOMETRY, INTERPRETATIONS
CAVIAR GROUND TRUTH LABELING: GEOMETRY

- BOUNDING BOX: ID, CENTRE COORDINATES, WIDTH, HEIGHT
- ORIENTATION OF MAIN AXIS
- SOME: HEADS, HANDS, FEET, SHOULDERS
- LABELLING ONLY IF TARGET MOVED IN SEGMENT
- GROUPS OF INTERACTING INDIVIDUALS

Ground truthed data and their quality
GROUND TRUTH LABELING: BEHAVIOR

FOUR LEVELS OF BEHAVIOUR:

1. INSTANTANEOUS MOVEMENT: INACTIVE, ACTIVE, WALKING, RUNNING

2. INSTANTANEOUS SITUATION: FALLING DOWN, BROWSING, LEAVING OBJECT
3. LONGER TERM CONTEXT: COLLAPSING PERSON, WINDOW-SHOPPING, LEFT SUSPICIOUS OBJECT

4. ROLE: A WALKER, FALLING PERSON, LEFT OBJECT
STATE TRANSITION ACTIVITY

DESCRIPTION

BROWSE CONTEXT: A SEQUENCE OF “MOVE” AND “BROWSE” SITUATIONS
GROUND TRUTH QUALITY
ASSESSMENT

CAVIAR FightOneManDown SEQUENCE

3 LABELERS: PHD STUDENTS
INSTRUCTION: “LABEL ALL MOVING TARGETS”

9-11 INDIVIDUAL TARGETS, 1-2 GROUPS
958 FRAMES: WALKING OR IDLE MAINLY

1 FIGHT
1 DARK SMALL BACKGROUND TARGET
1 HARDLY MOVING FOREGROUND TARGET
CAVIAR AUTOMATIC DETECTION RATE

Individual True Detections vs Overlap

DETECTION: 95+%

Ground truthed data and their quality
POSITION ERROR

Individual Position Error vs Overlap

POSITION < 2 PIXELS

Ground truthed data and their quality
DETECTION TIMING ERROR

Individual Timelag Entry/Exit vs Overlap

ENTER $\leq 2$ FRAMES   EXIT $\leq 1$ FRAME
## SEMANTICS: MOVEMENT LEVEL
### PERCENT CORRECT

Observer 2 & 3 with Observer 1 as true

<table>
<thead>
<tr>
<th>MOVEMENT</th>
<th>INACTV</th>
<th>ACTV</th>
<th>WALK</th>
<th>RUN</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>INACTIVE</td>
<td>85</td>
<td>15</td>
<td>-</td>
<td>-</td>
<td>712</td>
</tr>
<tr>
<td>ACTIVE</td>
<td>8</td>
<td>32</td>
<td>60</td>
<td>0</td>
<td>1480</td>
</tr>
<tr>
<td>WALKING</td>
<td>-</td>
<td>12</td>
<td>87</td>
<td>2</td>
<td>7444</td>
</tr>
<tr>
<td>RUNNING</td>
<td>-</td>
<td>-</td>
<td>20</td>
<td>80</td>
<td>533</td>
</tr>
</tbody>
</table>

**OVERALL: 78%**

**SOME QUANTITY OF ACTIVITY AMBIGUITY**
## SEMANTICS: ROLE PERCENT CORRECT

<table>
<thead>
<tr>
<th>ROLES</th>
<th>BROWSER</th>
<th>WALKER</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BROWSER</td>
<td>48</td>
<td>52</td>
<td>248</td>
</tr>
<tr>
<td>WALKER</td>
<td>1</td>
<td>99</td>
<td>9921</td>
</tr>
</tbody>
</table>

**OVERALL: 97%**

**WALKER/BROWSER LABELLING INSTRUCTION AMBIGUITY**
## SEMANTICS: SITUATION PERCENT CORRECT

<table>
<thead>
<tr>
<th>SITUATION</th>
<th>MOVING</th>
<th>INACTIVE</th>
<th>BROWSING</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOVING</td>
<td>94</td>
<td>6</td>
<td>0</td>
<td>8440</td>
</tr>
<tr>
<td>INACTIVE</td>
<td>30</td>
<td>61</td>
<td>8</td>
<td>1481</td>
</tr>
<tr>
<td>BROWSING</td>
<td>2</td>
<td>50</td>
<td>48</td>
<td>248</td>
</tr>
</tbody>
</table>

OVERALL: 88%

INACTIVE/BROWSING LABELLING INSTRUCTION AMBIGUITY
## SEMANTICS: CONTEXT PERCENT CORRECT

<table>
<thead>
<tr>
<th>CONTEXT</th>
<th>BRWS</th>
<th>IMMOB</th>
<th>WALK</th>
<th>FALL</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BROWSING</td>
<td>47</td>
<td>52</td>
<td>0</td>
<td>-</td>
<td>1166</td>
</tr>
<tr>
<td>IMMObILE</td>
<td>85</td>
<td>-</td>
<td>15</td>
<td>-</td>
<td>657</td>
</tr>
<tr>
<td>WALKING</td>
<td>-</td>
<td>2</td>
<td>98</td>
<td>-</td>
<td>4868</td>
</tr>
<tr>
<td>FALLING</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>100</td>
<td>3478</td>
</tr>
</tbody>
</table>

**OVERALL: 87%**

**BROWSING/IMMOBILE LABELLING INSTRUCTION AMBIGUITY**

**IMMOBILE/WALKING DATA AMBIGUITY**
ASSESSING SEMANTICS IGNORING AMBIGUITY

CONFLATE: IDLE/BROWSE, WALK/RUN,
ACTIVE/INACTIVE
ALLOW ± 20 FRAMES FOR TIMING ERRORS

\[ d_{i,j,p,t} : \text{NUMBER OF FRAMES WHERE OBSERVER } i \text{ AND } j \text{ DIFFER FOR TARGET } p \text{ AND PROPERTY } t \]

\[ n_{i,p,t} : \text{NUMBER OF TRACKED FRAMES FOR OBSERVER } i, \text{ TARGET } p \text{ AND PROPERTY } t \]

INCONSISTENCY : \[ c_{i,j,t} = \frac{\sum_p (d_{i,j,p,t} + d_{j,i,p,t})}{\sum_p (n_{i,p,t} + n_{j,p,t})} \]
## INCONSISTENCY RESULTS

<table>
<thead>
<tr>
<th>OBSERVERS</th>
<th>MOVE’T</th>
<th>ROLE</th>
<th>SIT’N</th>
<th>CONTEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &amp; 2</td>
<td>0.068</td>
<td>0.030</td>
<td>0.068</td>
<td>0.026</td>
</tr>
<tr>
<td>1 &amp; 3</td>
<td>0.018</td>
<td>0.028</td>
<td>0.018</td>
<td>0</td>
</tr>
<tr>
<td>2 &amp; 3</td>
<td>0.107</td>
<td>0</td>
<td>0.107</td>
<td>0.027</td>
</tr>
</tbody>
</table>

~10% DIFFERENCE FOR INSTANTANEOUS INTERPRETATIONS
~3% DIFFERENCE FOR LONGTERM INTERPRETATIONS
CONCLUSIONS

EVEN WITH MULTIPLE MARKERS AND REVIEW, ERRORS IS:

- GEOMETRIC QUANTITIES: 1-2 PIXELS
- TEMPORAL QUANTITIES: 1 SECOND
- DETECTIONS MISSED: ≤ 5%
- FALSE ALARMS: 0% (IE. ONLY INTERPRETATION AMBIGUITIES)
- SEMANTICS LABELLING: 5-10% (AFTER INTERPRETATION AMBIGUITIES)

BOUNDS FOR COMPARISON OF PROGRAM RESULTS TO GROUND TRUTH