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

BEHAVE Dataset Overview

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

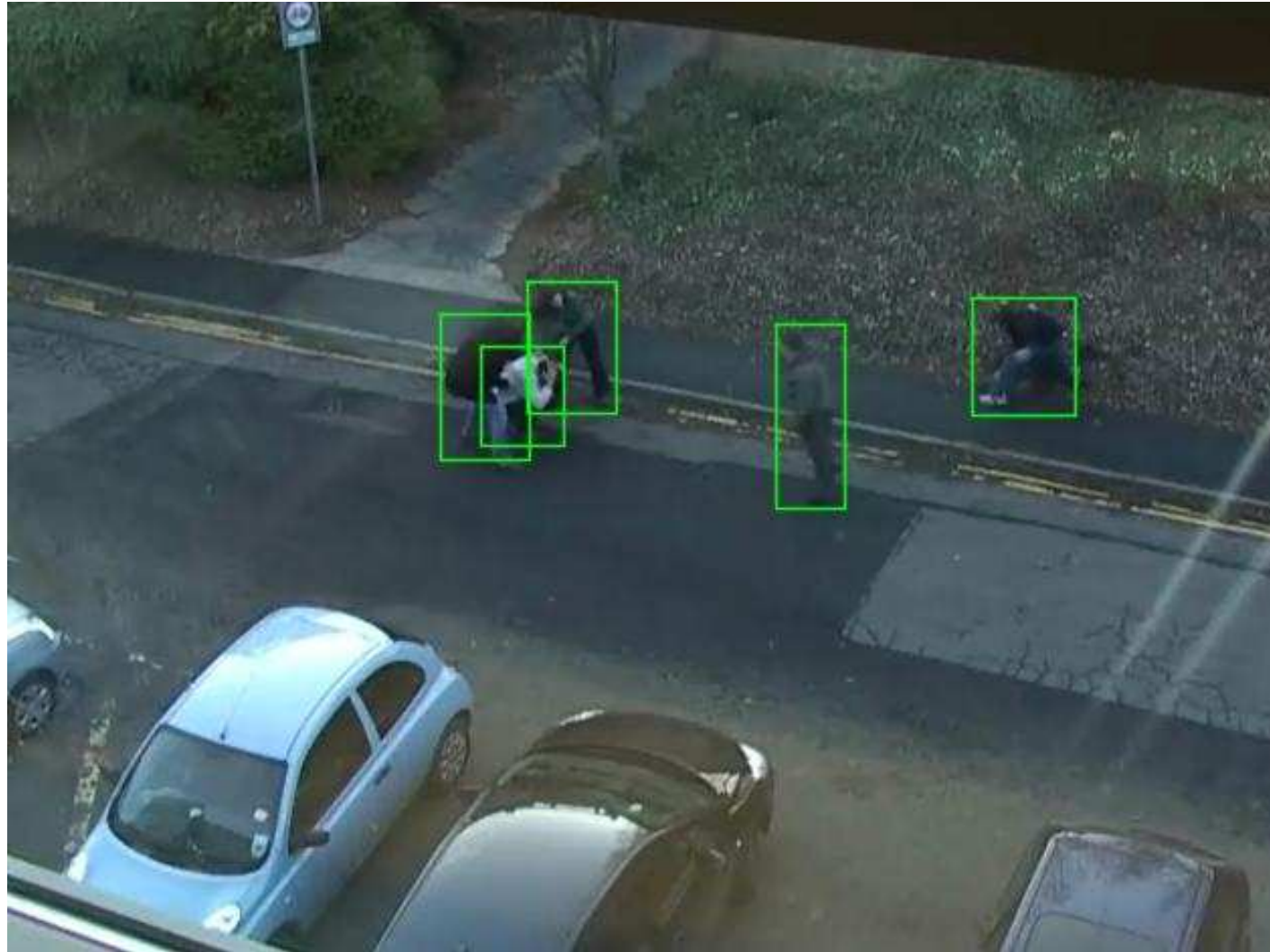25 FPS, 640x480, 60,000 marked up frames
AVI + JPEGs of frames
Ground plane homography data

Bounding boxes around people (VIPER XML)
VIPER based labeling

2838 page accesses (11/12/07)

BEHAVE Dataset Example (Fight)

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 TRUTH LABELING: BEHAVIOR

FOUR LEVELS OF BEHAVIOUR:

1. INSTANTANEOUS MOVEMENT: INACTIVE, ACTIVE, WALKING, RUNNING
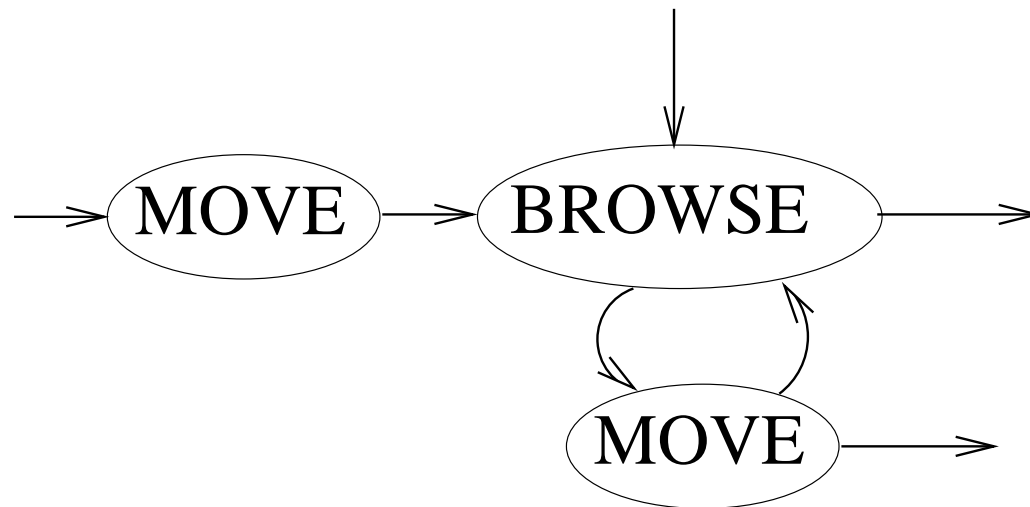
2. INSTANTANEOUS SITUATION: FALLING DOWN, BROWSING, LEAVING OBJECT

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
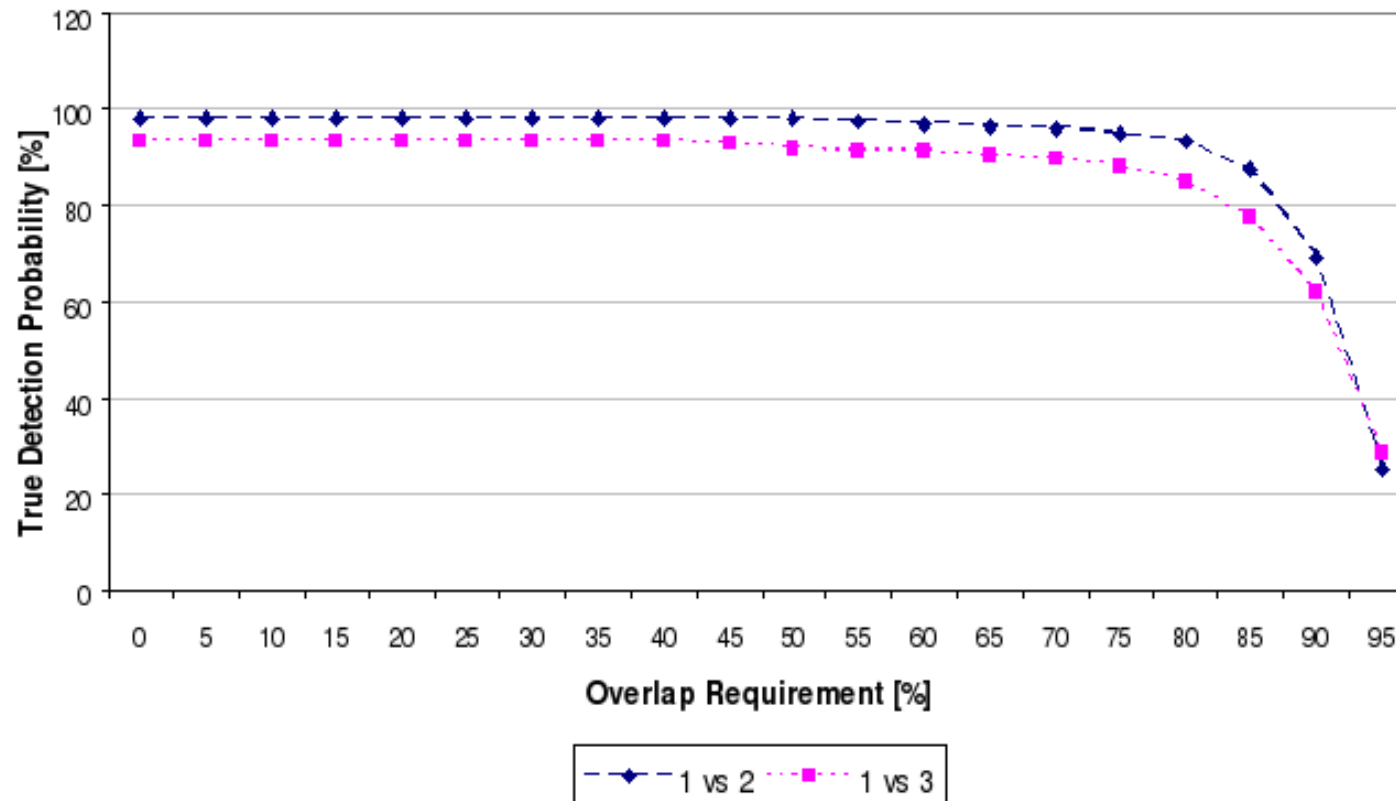958 FRAMES: WALKING OR IDLE MAINLY

1 FIGHT
1 DARK SMALL BACKGROUND TARGET
1 HARDLY MOVING FOREGROUND TARGET

CAVIAR AUTOMATIC DETECTION RATE

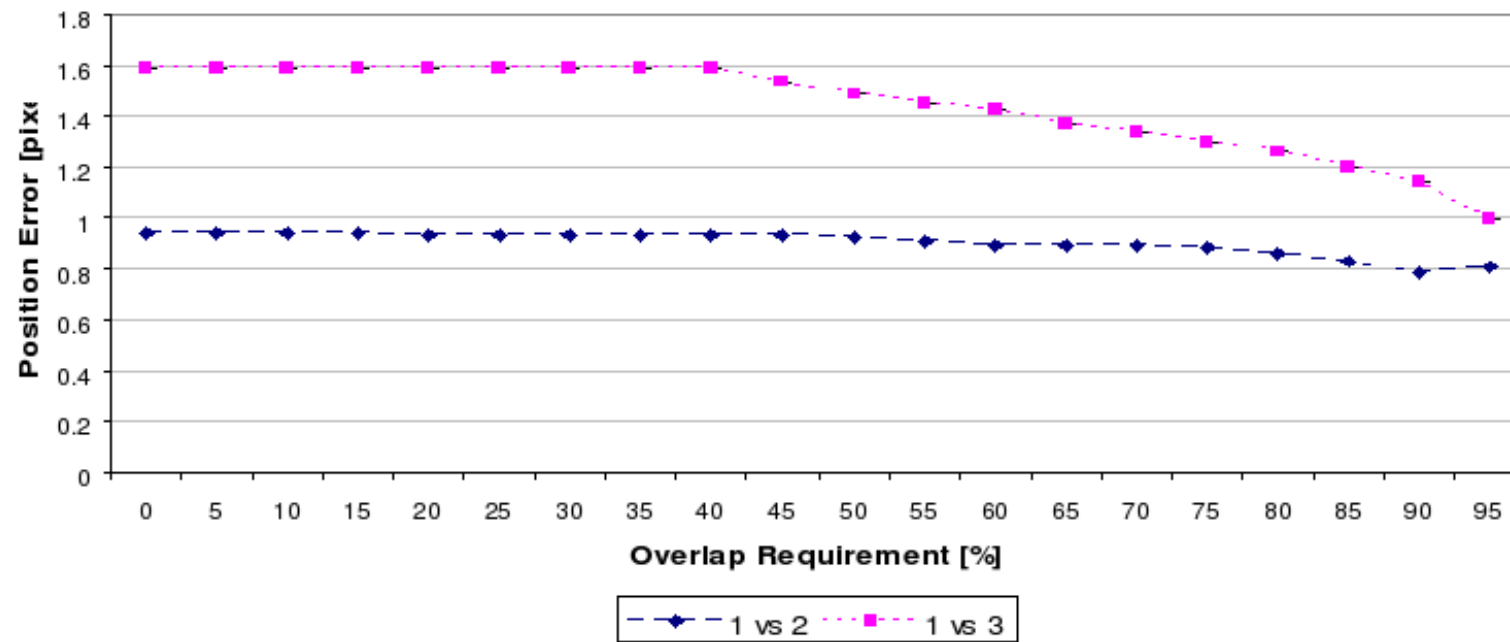**Individual True Detections vs Overlap**
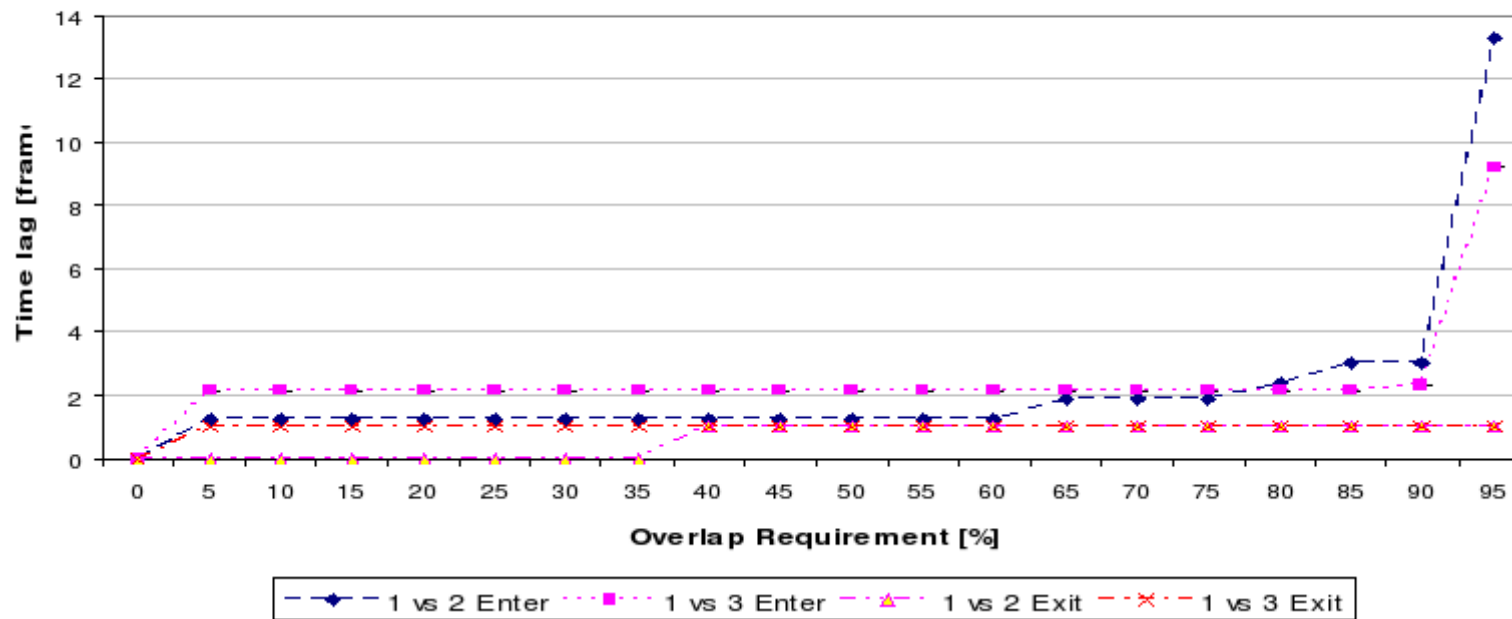

DETECTION: 95+%

POSITION ERROR

Individual Position Error vs Overlap

POSITION < 2 PIXELS

DETECTION TIMING ERROR

Individual Timelag Entry/Exit vs Overlap

ENTER ≤ 2 FRAMES   EXIT ≤ 1 FRAME

SEMANTICS: MOVEMENT LEVEL PERCENT CORRECT

Observer 2 & 3 with Observer 1 as true

MOVEMENT	INACTV	ACTV	WALK	RUN	TOTAL
INACTIVE	85	15	-	-	712
ACTIVE	8	32	60	0	1480
WALKING	-	12	87	2	7444
RUNNING	-	-	20	80	533

OVERALL: 78%

SOME QUANTITY OF ACTIVITY AMBIGUITY

SEMANTICS: ROLE PERCENT CORRECT

ROLES	BROWSER	WALKER	TOTAL
BROWSER	48	52	248
WALKER	1	99	9921

OVERALL: 97%

WALKER/BROWSER LABELLING INSTRUCTION
AMBIGUITY

SEMANTICS: SITUATION PERCENT CORRECT

SITUATION	MOVING	INACTIVE	BROWSING	TOTAL
MOVING	94	6	0	8440
INACTIVE	30	61	8	1481
BROWSING	2	50	48	248

OVERALL: 88%

INACTIVE/BROWSING LABELLING INSTRUCTION
AMBIGUITY

SEMANTICS: CONTEXT PERCENT CORRECT

CONTEXT	BRWS	IMMOB	WALK	FALL	TOTAL
BROWSING	47	52	0	-	1166
IMMOBILE	85	-	15	-	657
WALKING	-	2	98	-	4868
FALLING	-	-	0	100	3478

OVERALL: 87%

BROWSING/IMMOBILE LABELLING INSTRUCTION
AMBIGUITY
IMMOBILE/WALKING DATA AMBIGUITY

ASSESSING SEMANTICS IGNORING AMBIGUITY

CONFLATE: IDLE/BROWSE, WALK/RUN,
    ACTIVE/INACTIVE

ALLOW $\pm$ 20 FRAMES FOR TIMING ERRORS

$d_{i,j,p,t}$: NUMBER OF FRAMES WHERE OBSERVER i AND j
DIFFER FOR TARGET p AND PROPERTY t

$n_{i,p,t}$: NUMBER OF TRACKED FRAMES FOR OBSERVER i,
TARGET p AND PROPERTY t

$$\text{INCONSISTENCY}: \quad c_{i,j,t} = \frac{\sum_p (d_{i,j,p,t} + d_{j,i,p,t})}{\sum_p (n_{i,p,t} + n_{j,p,t})}$$

INCONSISTENCY RESULTS

OBSERVERS	MOVE'T	ROLE	SIT'N	CONTEXT
1 & 2	0.068	0.030	0.068	0.026
1 & 3	0.018	0.028	0.018	0
2 & 3	0.107	0	0.107	0.027

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

CONCLUSIONS

EVEN WITH MULTIPLE MARKERS AND REVIEW, ERROR IS:

- GEOMETRIC QUANTITIES: 1-2 PIXELS

- TEMPORAL QUANTITIES: 1 SECOND

- DETECTIONS MISSED: $\leq 5\%$

- FALSE ALARMS: 0% (IE. ONLY INTERPRETATION AMBIGUITIES)

- SEMANTICS LABELLING: 5-10% (AFTER INTERPRETATION AMBIGUITIES)

BOUNDS FOR COMPARISON OF PROGRAM RESULTS TO GROUND TRUTH