CAVIAR

IST- 2001- 37540

Context Aware Vision using Image-based Active Recognition

University of Edinburgh — Prof. R. Fisher

Inst. Nat. de Recherce en Informatique et en Automatique — Prof. J. Crowley

Instituto Técnico Superior — Dr. José Santos-Victor

Project active: Oct 1, 2002 - Sept 30, 2005

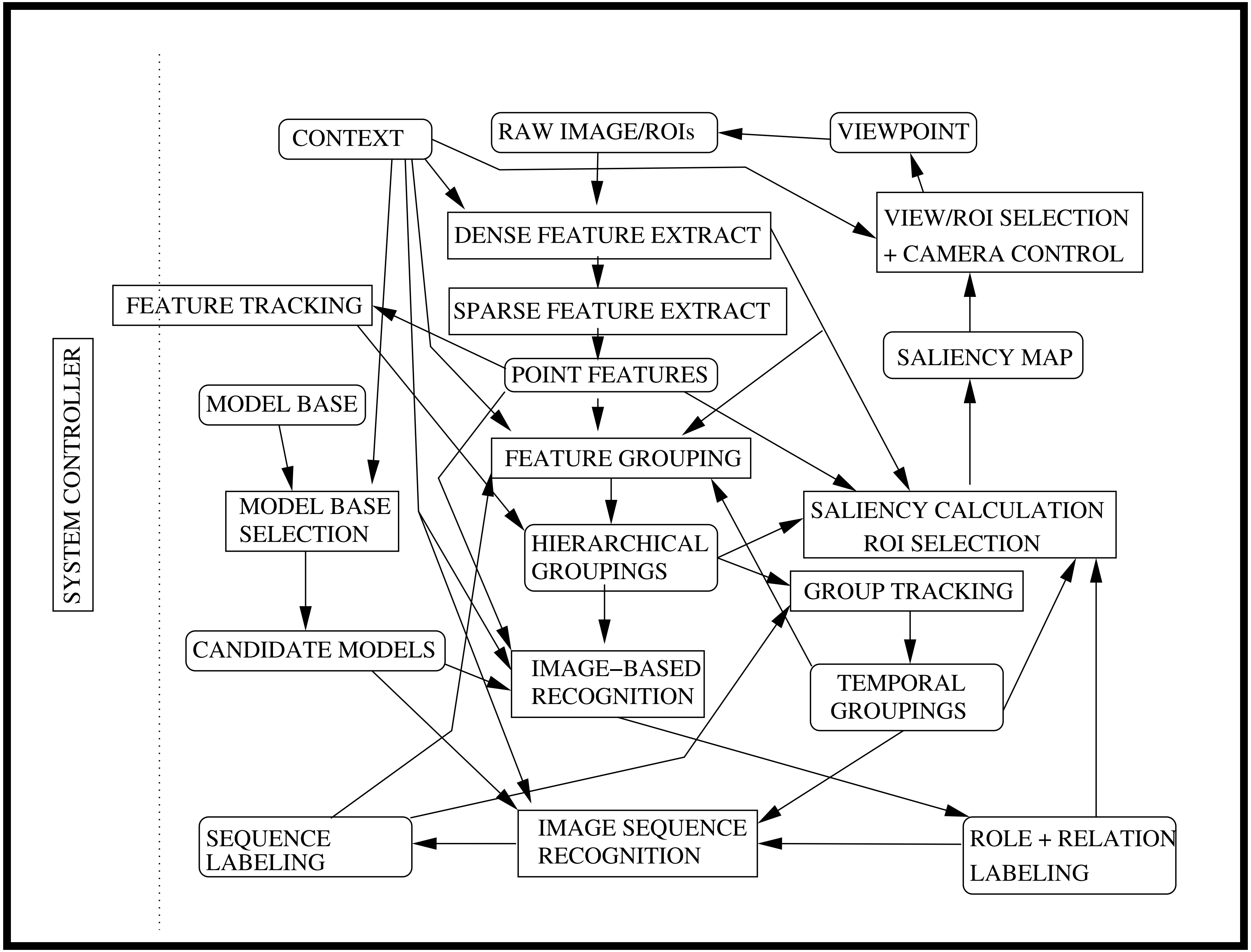
Key scientific question

Can the use of:

- rich local image descriptions,
- selected by a hierarchical visual attention process, and
- guided and processed using task, scene and contextual knowledge

improve image-based recognition processes?

Cognitive Architecture


SYSTEM CONTROLLER
CONTEXT
RAW IMAGE/ROIs
VIEWPOINT
VIEW/ROI SELECTION
+ CAMERA CONTROL
DENSE FEATURE EXTRACT
SPARSE FEATURE EXTRACT
FEATURE TRACKING
POINT FEATURES
SALIENCY MAP
MODEL BASE
FEATURE GROUPING
MODEL BASE
SELECTION
SALIENCY CALCULATION
ROI SELECTION
HIERARCHICAL
GROUPINGS
GROUP TRACKING
CANDIDATE MODELS
IMAGE−BASED
RECOGNITION
TEMPORAL
GROUPINGS
SEQUENCE
LABELING
IMAGE SEQUENCE
RECOGNITION
ROLE + RELATION
LABELING

Key Achievements

1. **Gabor based interest point detection and local image representations for object recognition** (WP1)
2. **Set of 29 visual features for short term activity recognition** (WP1)
3. **The extension of natural interest points to provide affine invariant "natural interest lines"** (WP1)

4. **Integration of core CAVIAR modules**: INRIA target detection module, UEDIN track formation module, IST dense and sparse feature detection modules, IST short-term activity classification module, UEDIN rule based and statistical behaviour recognition modules and UEDIN statistics module. (WP2)

5. **A software environment for composition of perceptual processes** (WP2)

6. **Methods prioritising target detection and model-based recognition** (WP3)

7. **Accurate (98% correct) short-term activity recognition** (WP4)
8. **HSMM behaviour recognition algorithm** (WP4)
9. **Comparison of rule-based and HSMM behaviour recognition algorithms** (WP4)
10. **Development of a new form of linear auto-associative network for recognition of posture and identity** (WP4)

11. **Methods for automatic learning of probabilistic models for short-term activity** (WP5)
12. **Methods for automatic learning of probabilistic situation models** (WP5)
13. **Target motion detection algorithm comparison** (WP6)
14. **Ground truth dataset extension** (WP6)
15. **Answers to most scientific questions**

Workpackage Review

WP1	Active image acquisition	Santos-Victor
WP2	Control Architectures	Fisher
WP3	Selective Attention	Fisher
WP4	Recognition Processes	Fisher, Santos-Victor
		Crowley
WP5	Model Learning	Crowley
WP6	Performance Evaluation	Fisher
WP7	Dissemination	Fisher
WP8	Management	Fisher

WP1/Q2: Is there a benefit to using Gabor versus Gaussian derivative filters for feature extraction?

Look for eyes in face images

Detection rate:

Gabor filter	Gaussian der	Directional der
47.55%	37.8%	37.8%

Workpackage 2 - Control Architectures

T2.1: Build a baseline architecture for perception control

T2.2: Integrate all of the different visual modalities

T2.3: Implement a central, master-based controller

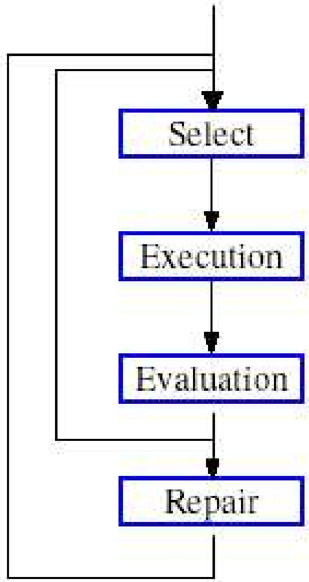
T2.4: Implement auto-regulating supervisory control

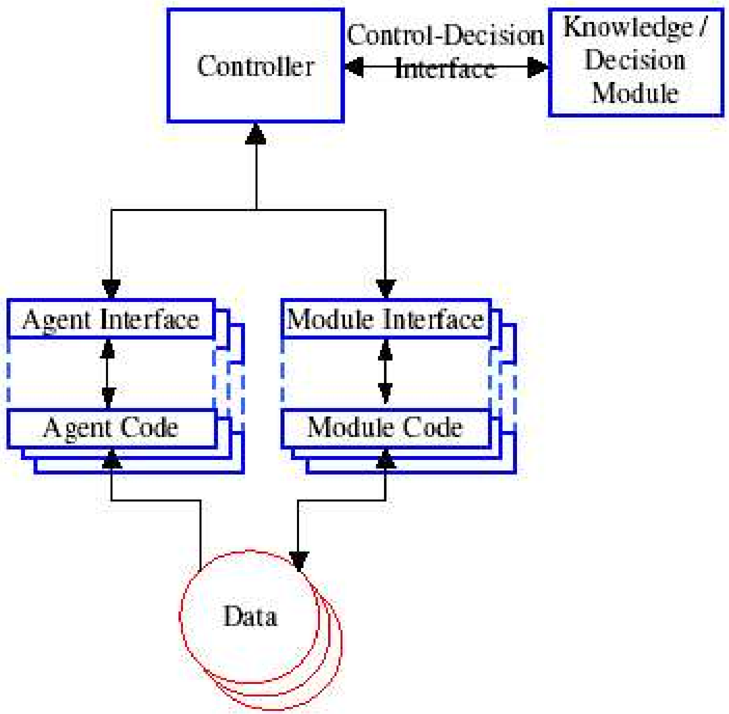
Q8: What control policies optimise performance in video behaviour understanding systems

Intelligent Controller Architecture

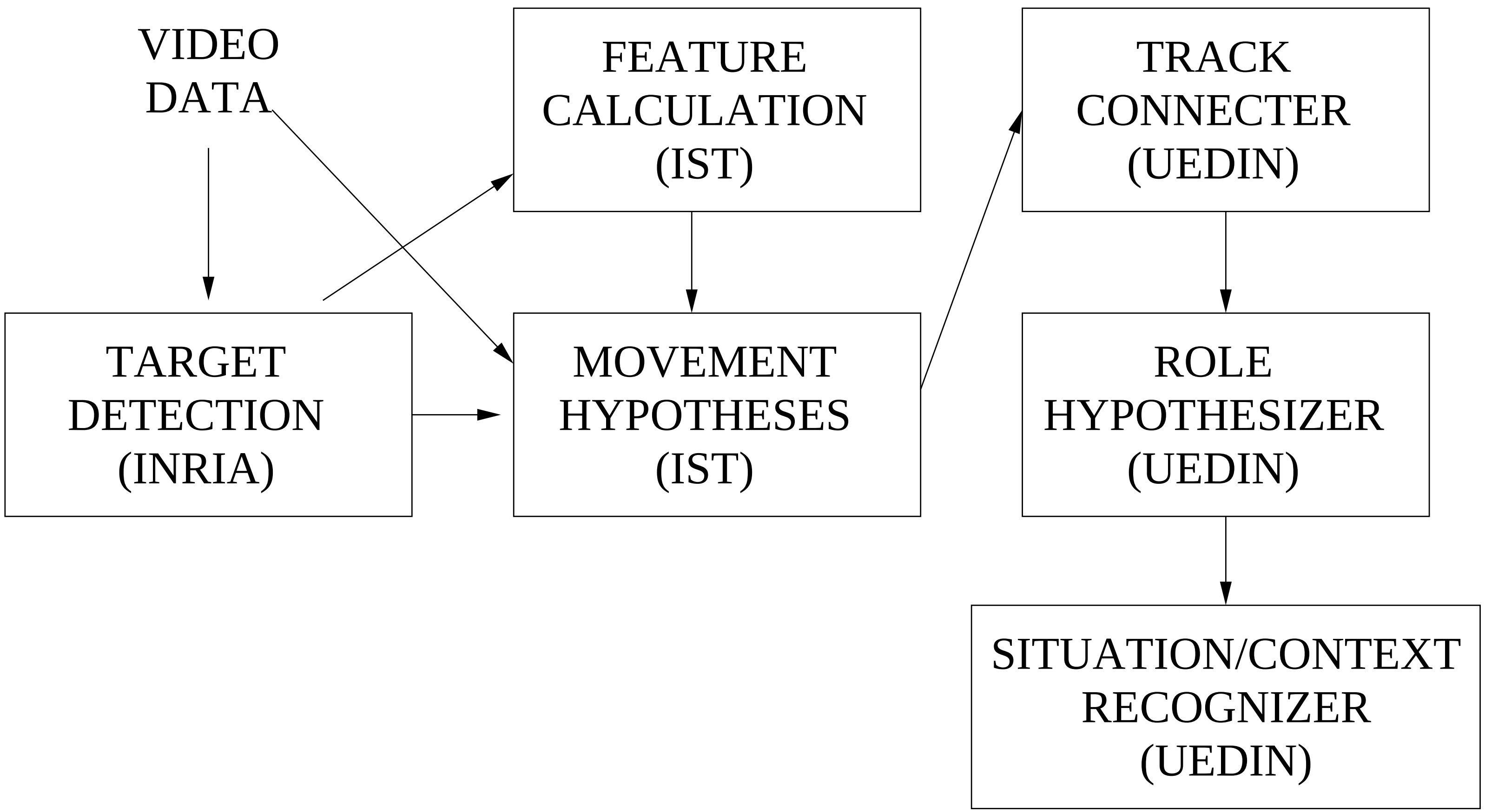
Agents: evaluate parameters, do quick diagnostics

Decision Module: reasoning over next module, parameter changes





Plug-and Play Integrated Modules



Key features

- Data-flow automatically constructed from XML auto-descriptions.
- Rule-based selection of “ready” module to run next
- ImaLab/RAVI environment

Auto-adaptation

- Optimises throughput quality, quantity, rate
- Self-adapting parameter tuning based on auto-descriptive functions and rules
- Global tuning through individual module adjustments
- Uses each module's auto-critic
- Awaiting rule learning & experiments (limited by integration, data collection time)

Q8: What control policies optimise performance in video behaviour understanding systems

Each module has model of quality, quantity, rate vs parameters

Controller uses these to optimise overall performance

Controller constructed

Module model learning constructed

Ready for model learning & experiments

Workpackage 3 - Selective Attention

T3.1: Implement hierarchical object-based salience and saccade process

T3.2: Integrate top-down priming from situation and context priming

T3.3: Generate situation and context alerts

Q3: What attention-focussing strategy for target detection gives most efficient and effective results?

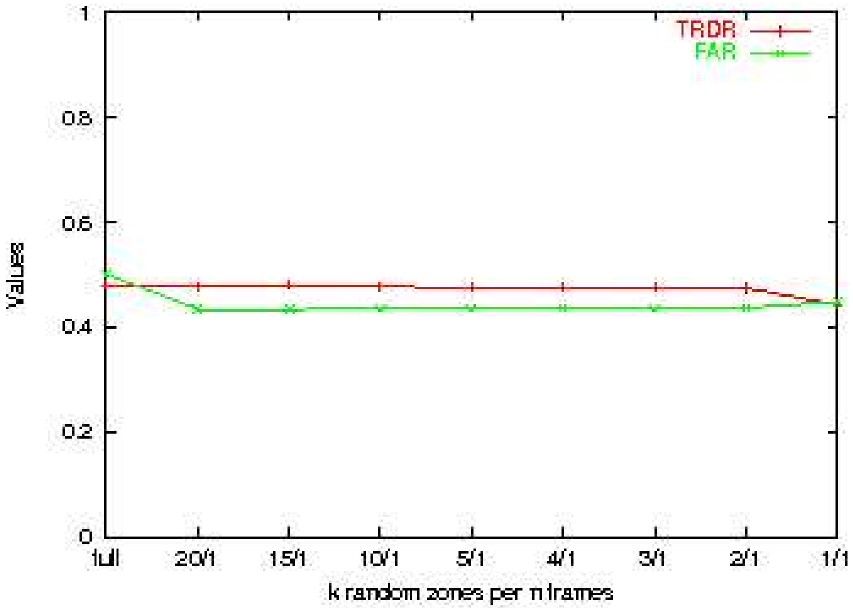
Q5: The value of domain knowledge to focus invocation of context models

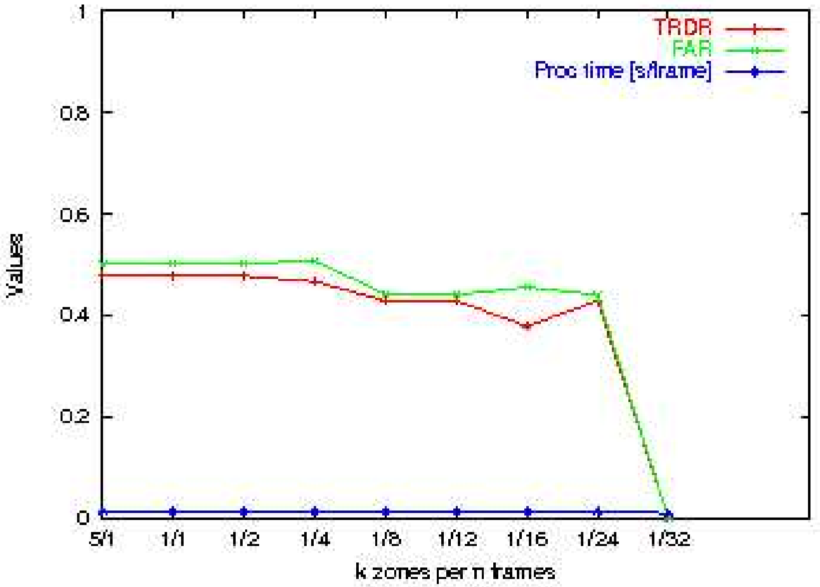
Progress

- INRIA: detection regions: subsampling in space and time gives similar detection rates but is faster
- UEDIN: situation/context model priming using Bayesian network ranks correct model hypothesis near to top for each target.

Q3: What attention-focussing strategy for target detection gives most efficient and effective results?

Tracker detection regions placed: randomly, spatial subsample at extry/exit zones, temporal subsample, complete





Random bad, 15 small detection zones spatially subsampling full detection zone subsample temporally 1 frame in 4. 78 > 82 fps

Q5: The value of domain knowledge to focus invocation of context models

Bayesian Model Network

Fuses target and model interestingness

8 context models

Use previous frame context: avg. rank = 0.09 out of 7

No previous frame context: avg. rank = 1.2 out of 7

Workpackage 4 - Recognition Processes

T4.1 Define and implement visual processes for observing properties and relations

T4.2 Define representations for object, situation and context

T4.3 Develop process for object & behaviour recognition under variations in viewpoint

T4.4 Develop process for diagnosing and recovering from errors

Q1: What are good features for video-based human behaviour understanding?
Q4: The effect of the temporal window length for describing activity
Q6: Rule-based versus probabilistic behaviour recognition
Q7: The relative importance of feature data [ie. p(situation|data)] versus model relations [ie. context graph]
Q9: What is a good balance between data-driven processing and model-driven processing?

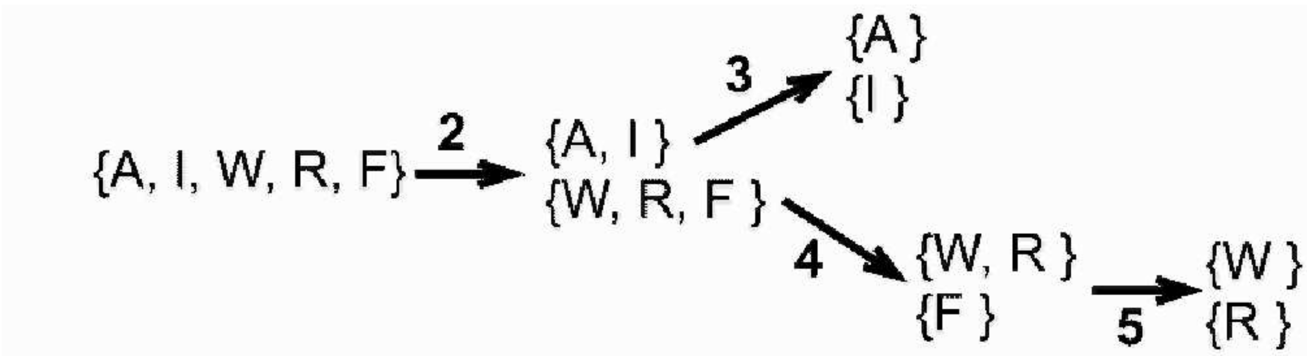
IST+INRIA Contributions (to follow)

- IST short-term activity classifier, feature selection methods
- INRIA natural interest lines
- INRIA linear auto-associative memory object recogniser

Q1: What are good features for video-based human behaviour understanding?

Low-level activity classification, as in {**A**ctive, **I**nactive, **W**alking, **R**unning, **F**ighting}

Bayesian hierarchical classifier, GMM 29 features selected

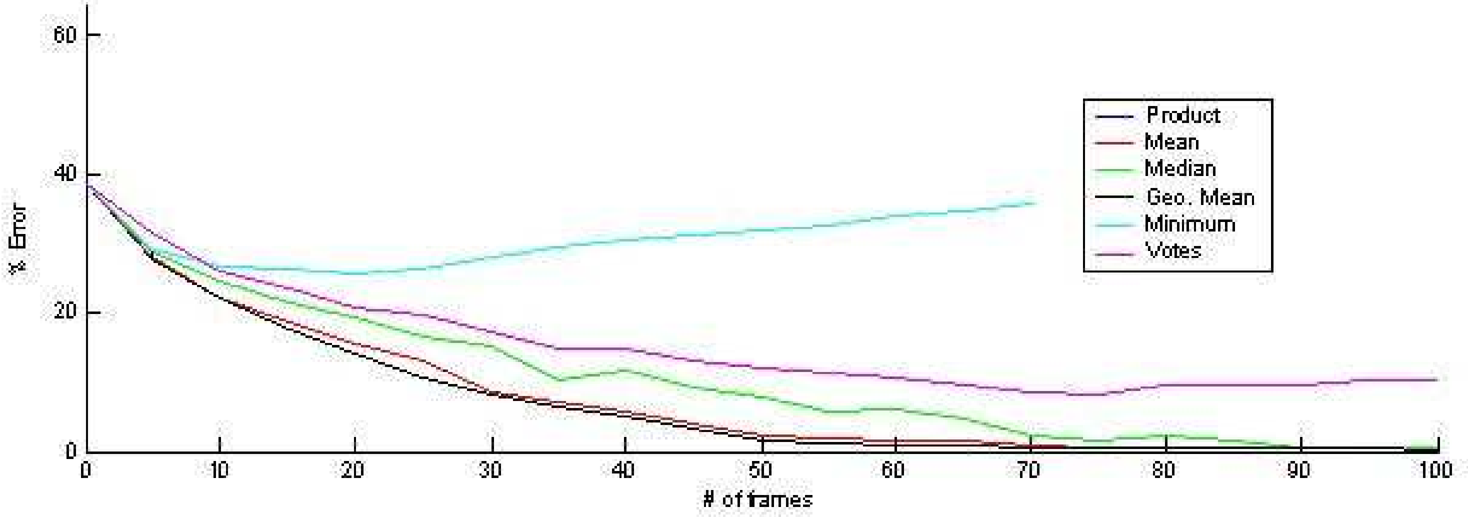


98.2% correct

Q4: The effect of the temporal window length for describing short-term activities

5 activity classes

Error rates for different data fusion methods:

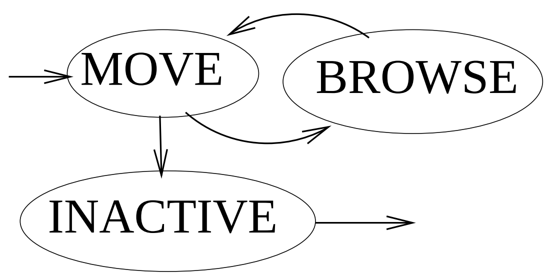


Used 50 frames

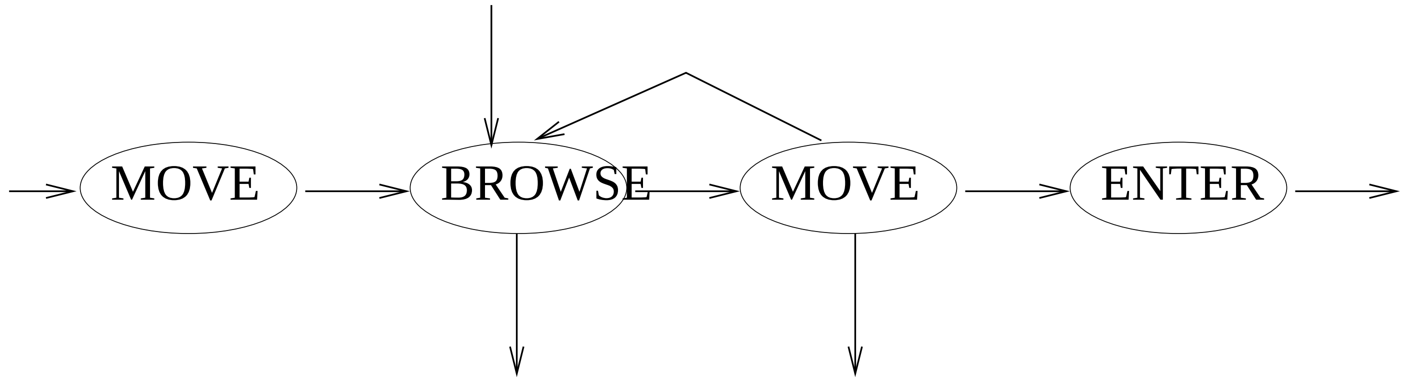
Also need to consider time lag and multiple activities

Graphical Context Models

Drop Dead Context

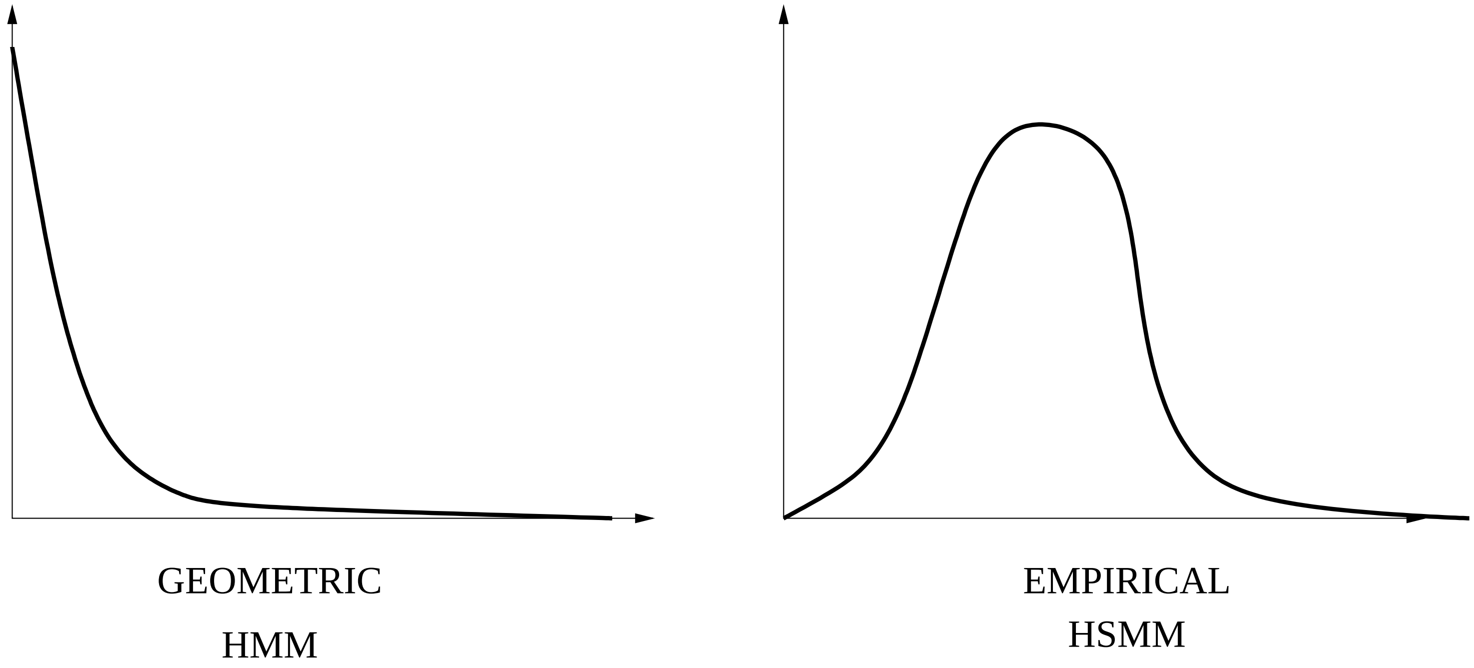


Window-shop Context



Linear Time HSMM Recognizer

Replace HMM's implicit geometric time-in-state distribution by learned distribution → Hidden Semi-Markovian Model



- Previous vision algorithms are $\mathrm{O}(T^2)$
- Adapted a gene sequence matching algorithm for $\mathrm{O}(T)$ recognition
- Allows continuous operation
- Uses $p(activity \mid data)$ from IST and $p(role \mid data)$ from UEDIN

Q6: Rule-based versus probabilistic behaviour recognition

Rule based parser: hand constructed rules with learned parameters

Probabilistic HSMM: with learned distributions

Test over ground truth tracking, activities, roles

	Situations	Contexts
Rule based	70%	57%
HSMM	74%	65%

Q7: The relative importance of feature data [ie. p(situation|data)] versus model relations [ie. context graph] for behaviour recognition

---

Baseline HSMM recogniser on GT tracking, GT roles, GT* activity

1. **Randomly corrupted activity labels in a block**
2. **Randomly corrupted role labels in a block**
3. **Randomly corrupted role labels randomly**
4. **HMM (use a weaker model)**

Default Situation Recognition in 1000 frames (rounded)

	SM	SB	SI	SEn	SEx	SErr	Tot	%
SM	139		-			26	165	84
SB		10				10	20	49
SI			13			6	19	66
SEn	-			2		2	4	42
SEx					2	4	6	25
TOT							215	77

Default Context Recognition, in 1000 frames (rounded)

	CW	CB	CI	CEn	CEx	CWi	CErr	Tot	%
CW	66	1	4	-			2	74	89
CB	-	15					10	26	57
CI	-	-	23				7	31	75
CEn	1			15		-	7	24	66
CEx	-				21		14	36	60
CR							1	2	24
CWi		9				2	4	16	14
TOT								211	69

Experiment Summary

Experiment	Situation %	Context %
Default	77%	69%
Randomly block corrupt 20% of GT* short-term activities	47%	48%
Randomly block corrupt 20% of GT roles	11%	10%
Randomly individual corrupt 20% of GT roles	6%	6%
HMM classifier	80%	71%

Q9: What is a good balance between data-driven processing and model-driven processing?

Probabilistic Role Estimation, kernal based classifier

Classes of top hypotheses:

	RB	RW	REn	REx	Tot	%
RB	14636	7199	798	858	23491	62
RW	27399	140789	5557	12801	186546	75
REn	125	1475	818	1420	3838	21
REx	1194	1331	1701	2077	6303	32
Total					220178	72

Workpackage 6 - Performance Evaluation

T6.1: Set-up an automated performance characterisation process

T6.2: Design and collect performance characterisation datasets

T6.3: Keep current the "best version" performance results

Nightly Statistics

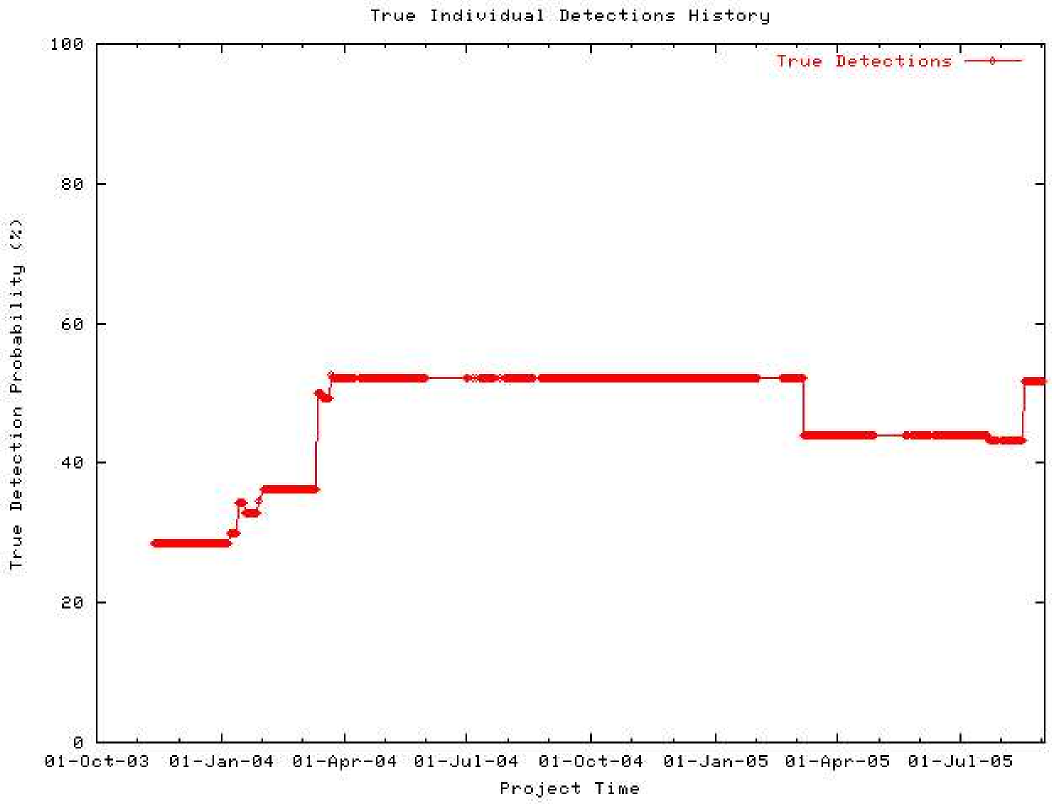
Statistics Pages for Caviar Architecture Performance

Today's Statistics Statistics History Caviar Home

Running Results

Individuals

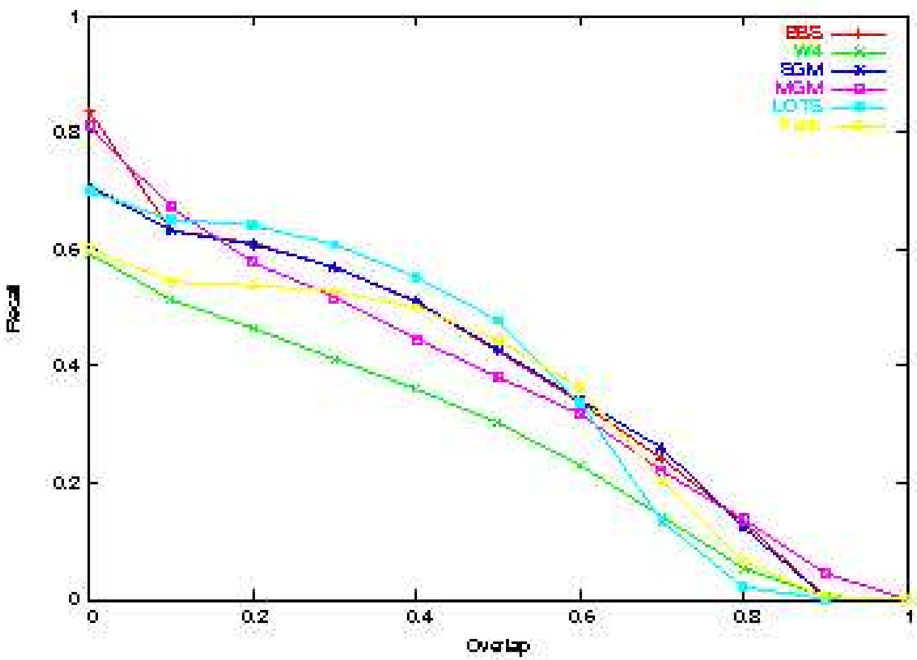
- True Detection Error
- Missed Detection Error
- False Detection Error
- Position Detection Error
- Size Detection Error
- Orientation Detection Error
- Time Lag Entry
- Time Lag Exit
- Dropped Frame Rate
- Drop Count
- Boxes with same Event
- Boxes with same Movement
- Boxes with same Role
- Boxes with same Situation
- Boxes with same Context
- Boxes with same Scenario



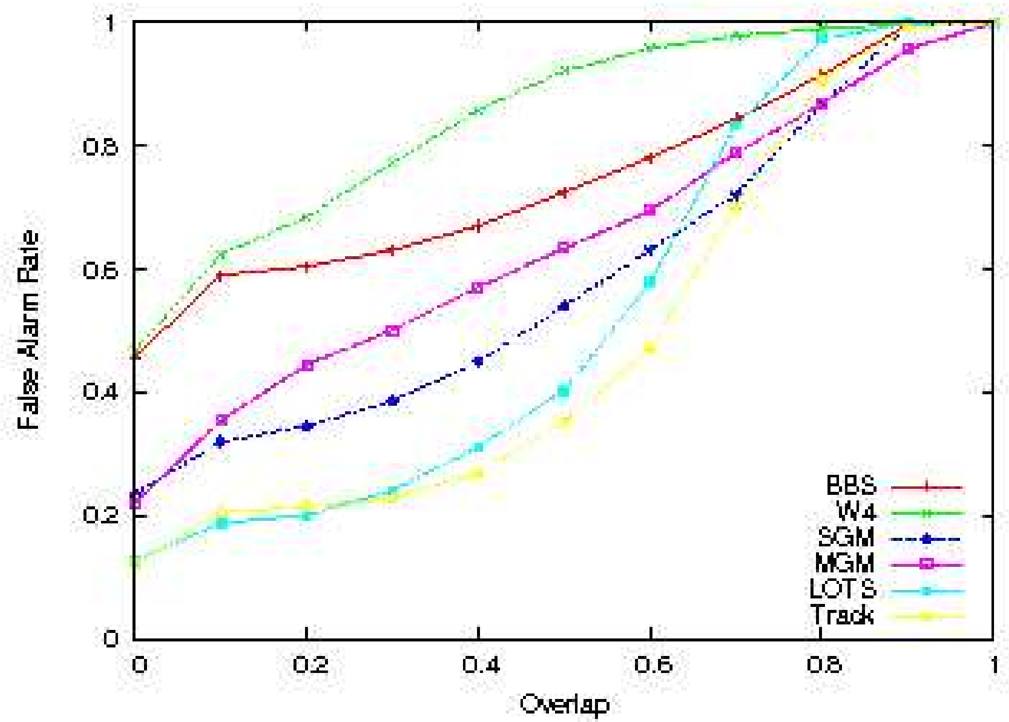
Q10: Performance comparison of target detectors

6 algorithms compared

TRUE DETECT



FALSE ALARM



Comparable detection

LOTS (Boult) has lowest false alarm rate

Unexpectedly low detect rates due to counting of groups

Status

- XML groundtruth labeling consistency improved. 6341 accesses since March 2005.
- Heads, hands, feet, shoulders & gaze added to groundtruth, 19 sequences, 108 people, 52+K frames
- Sequence with 3 labellings used for labelling consistency experiments: good geometric, timing consistency, semantic labelling status confusing (maybe 1% labelling errors + 40+% semantic ontology/timing differences)

Workpackage 7 - Dissemination

- Web site maintained and updated
- HAREM 2005 workshop organised (Human Activity Recognition and Modeling) at BMVC 2005
- 24 new publications (6 Year 1 + 17 year 2)
- Blue Eye Video / Observit relationship continuing

Workpackage 8 - Management

- Management meetings: 5 - 10/04, 1/05, 6/05, 11/05, 11/05
- Scientific meetings: 3 - 10/04, 1/05, 6/05 (1 day meetings)
- Partner visits: 4 - 4/05, 6/05, 9/05, 10/05
- Conferences attended: 7 - VS-PETS, ICCV, CAMP, AAAI-05 workshop, BMVC05, HAREM, Ljubljana Cog. Systems Workshop (underline=papers)

Year 3 Costed Person-Months (Est)

Workpackage	UEDIN	INRIA	IST	TOTAL
Features: WP 1			4	4
System Arch: WP 2	17		2	19
Attention: WP 3		2	1	3
Recognition: WP 4	12	8.5	3	23.5
Model Learning: WP 5		17		17
Perform Eval: WP 6	11.5	3	2	16.5
Dissem: WP 7	5		1	6
Management: WP 8		1.5	1	2.5
TOTAL	45.5	32	14	91.5
PROPOSED	24	30	11	65

Est. Cumulative Costed Person-Months

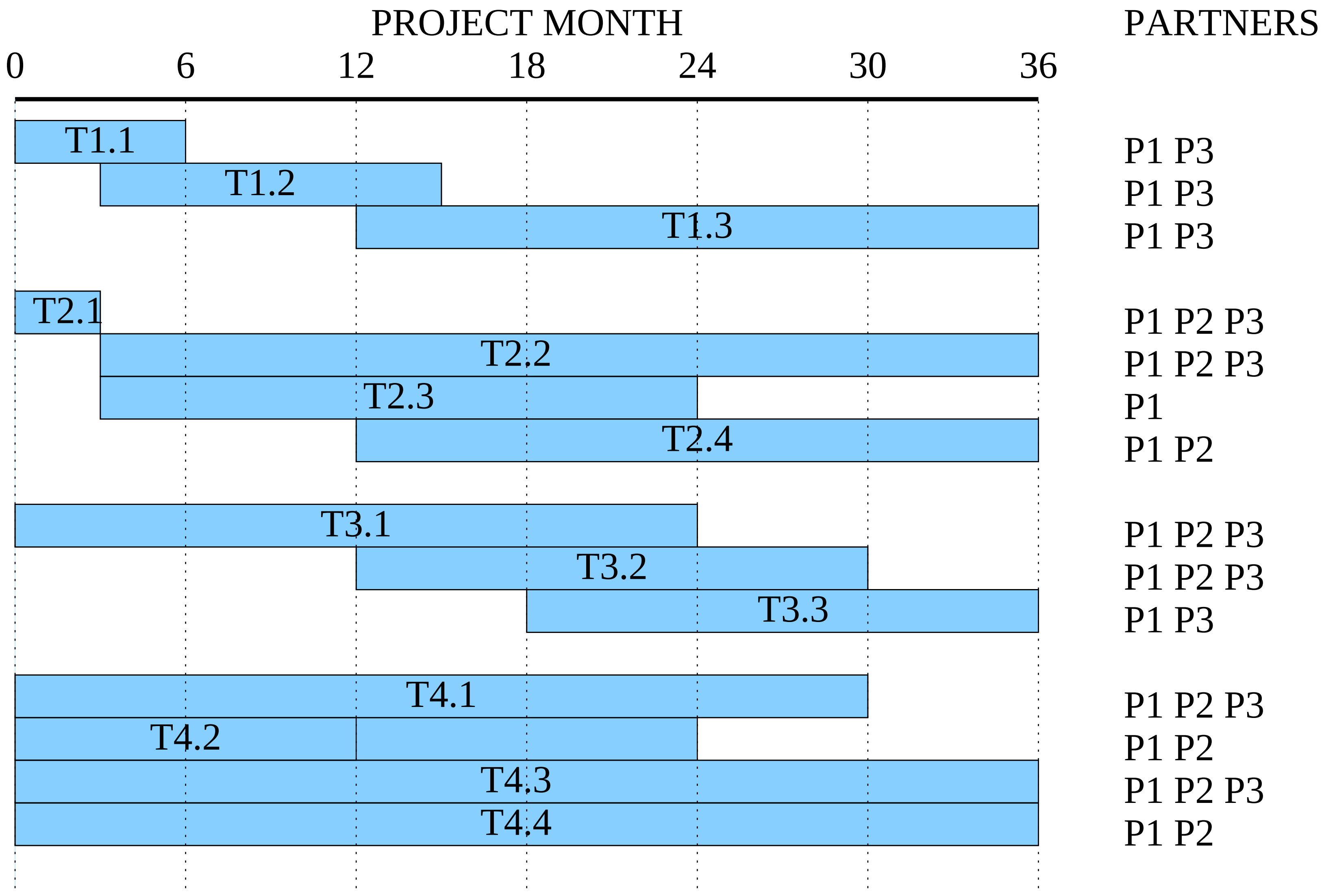
Workpackage	UEDIN	INRIA	IST	TOTAL	PROP
WP 1	4		13.75	17.75	24
WP 2	27	8.5	6.5	42	33
WP 3		6	4	10	18
WP 4	21	31.5	5	57.5	46
WP 5	4	37.75		41.75	29
WP 6	17.5	6	4.5	28	19
WP 7	6	5	3.5	14.5	17
WP 8	1	3.5	3	7.5	9
TOTAL	80.5	98.25	40.25	219	
PROPOSED	72	90	33		195

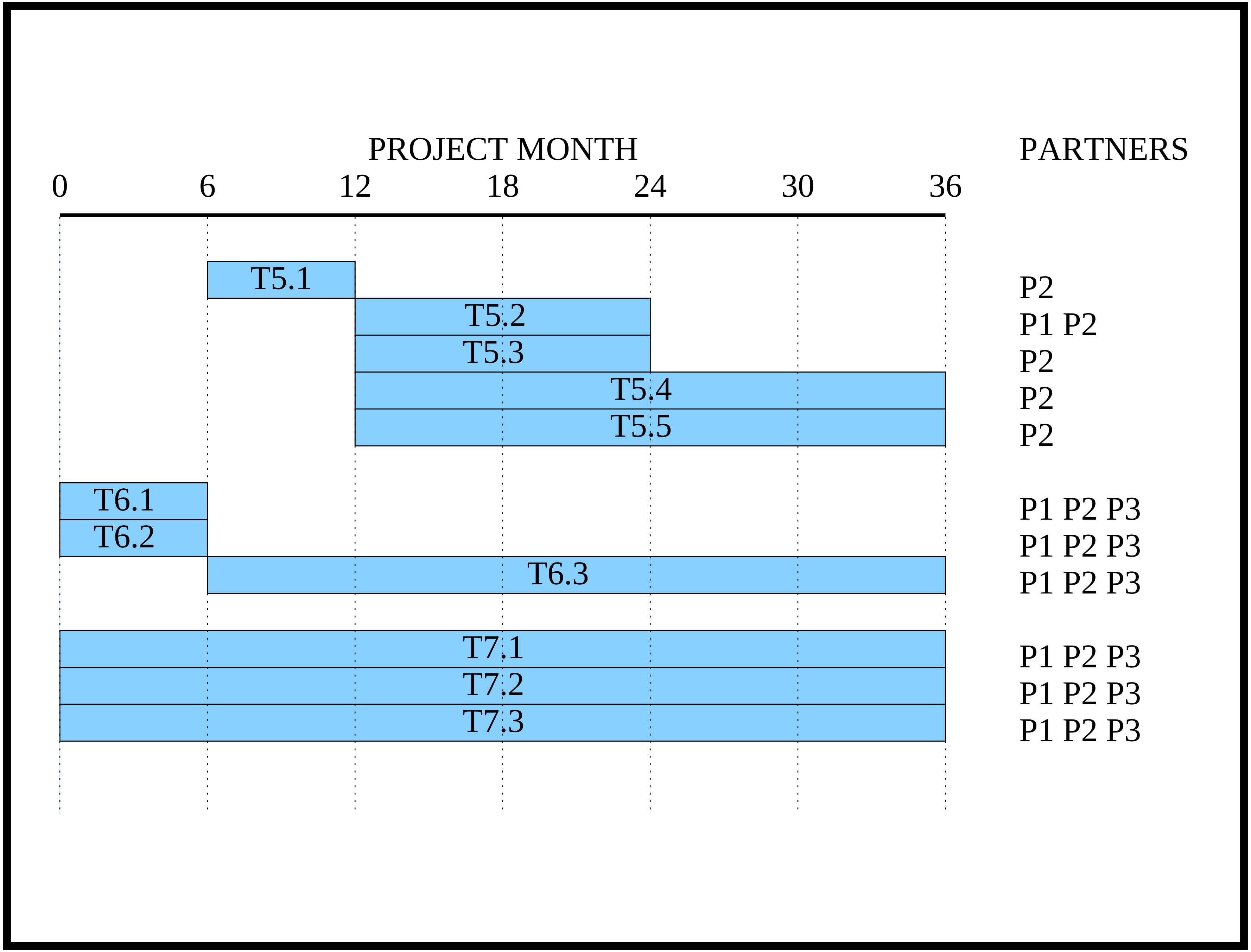
Deliverable Status

Item	Deliverable name	WP	Leader	Status
14	Feature-based salience report	3	IST	OK
16	Situation modeling tool	5	INRIA	OK
19	Feature-based salience software	3	IST	OK
20	Perceptual grouping software	1	IST	OK
21	Perceptual grouping report	1	IST	OK
22	Property observation report	4	INRIA	OK
23	Top-down primed salience report	3	UEDIN	OK
24	Property observation software	4	INRIA	OK
25	Top-down primed salience software	3	UEDIN	OK
D27	Context model learning report	5	INRIA	OK

Item	Deliverable name	WP	Leader	Status
28	Distributed control report	2	UEDIN	OK
29	Error recovery report	4	INRIA	OK
30	Federation assembly report	5	INRIA	OK
31	Object recognition report	4	INRIA	OK
32	Context model learning sw	5	INRIA	OK
33	Distributed control software	2	UEDIN	OK
34	Error recovery software	4	INRIA	OK
35	Federation assembly software	5	INRIA	OK
36	Object recognition software	4	INRIA	OK
39	Annual Project report	8	UEDIN	OK
40	Project Presentation	8	UEDIN	THIS
42	Technology Impl. Plan	8	INRIA	PARTIAL

GANTT CHART


PROJECT MONTH
PARTNERS
0
6
12
18
24
30
36
T1.1
T1.2
T1.3
T2.1
T2.2
T2.3
T2.4
T3.1
T3.2
T3.3
T4.1
T4.2
T4.3
T4.4
P1 P3
P1 P3
P1 P3
P1 P2 P3
P1 P2 P3
P1
P1 P2
P1 P2 P3
P1 P2 P3
P1 P3
P1 P2 P3
P1 P2
P1 P2 P3
P1 P2


PROJECT MONTH
PARTNERS
0
6
12
18
24
30
36
T5.1
P2
T5.2
P1 P2
T5.3
P2
T5.4
P2
T5.5
P2
T6.1
P1 P2 P3
T6.2
P1 P2 P3
T6.3
P1 P2 P3
T7.1
P1 P2 P3
T7.2
P1 P2 P3
T7.3
P1 P2 P3

Year 2 Review Action Items

1. **How to convince about the system architectural investigations**: 1) 2 papers on plug-and-play + 1 on controller architecture, 2) Few hours integration and 3) Planned dynamic system control experiments.

2. **Promote the CVML standard further**: GT datasets and CVML interface to CVonline; several publication mentions; public API on SourceForge; extended CVML for heads; direct emails. CVML API: about 50-100 accesses/month since Nov 2004.

3. **Joint publications**: 1) A paper comparing 6 target detections algorithms involving all 3 groups. 2) A second paper (INRIA, UEDIN) on intelligent system control in progress.

CAVIAR linked PhD theses

H. Tran - +3 months

N. Gourier - +6 months

P. Moreno - +6 months

T. List - +18 months

P. Ribeiro - +18 months

Life After CAVIAR

1. **Commercial Exploitation**: Blue Eye Video (increasing market for activity recognition); Observit links
2. **Scientific Exploitation**: INRIA (new research on autonomous architectures); INRIA (Brdiczka, Zaidenberg to CHIL); UEDIN (BEHAVE); IST - internal
3. **Further Research**: FP7 Surveillance?