Engineering Platform and Workflow for Cognitive Vision

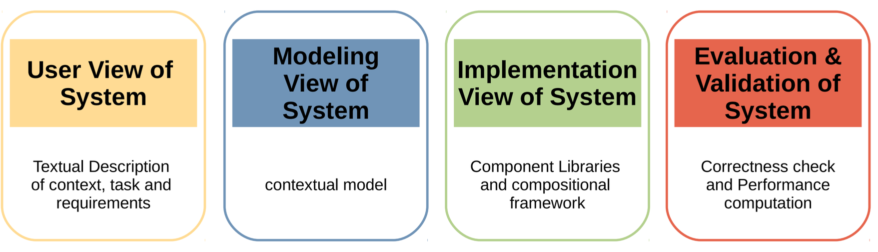
Tobias Weis[2], Martin Mundt[1], Patrick Harding[2], Visvanathan Ramesh[1,2]

[1] *Frankfurt Institute for Advanced Studies, Frankfurt am Main, Germany*
[2] *Johann Wolfgang Goethe Universität, Frankfurt am Main, Germany*

Overview

As part of the Bernstein Focus in NeuroTechnology initiative in Frankfurt, a trans-disciplinary project involving neuroscience, psychology, computer science and engineering, we have developed engineering platforms, simulation tools, and application case studies in cognitive vision following systems engineering principles.

System Engineering Methodology



Our engineering platform effort lies on the modeling, implementation and validation views of the system. Mainly on core data structure support for algorithms, distributed run-time execution of modules, enhanced debugging of system pipeline execution states, contextual modeling and computer graphics simulation of data, machine learning, optimization, and performance characterization of systems [1, 2, 3]. Given the rapidly emerging open-source efforts in Machine Learning and AI, the platform effort is planned to be shifted to a startup in engineering platforms.

Inspirations from Cognitive Architecture

- Our effort is to realize the proposed cognitive architecture [1], with its major components: hypothesis generators, deliberation and knowledge update.
- Visual intelligence is seen as context and task sensitive indexing followed by detailed estimation or deliberation (Indexing involves decomposition of input data into quasi invariant features as suggested in [4]). The indexing sub-modalities (e.g. color, texture, motion, illumination, etc.) are complementary in nature.
- Detailed state estimation can be implemented by a variety of schemes - distributed fusion, belief propagation, markov-chain monte carlo methods, or deliberation and reasoning [1, 5, 6, 7].

Model based system design thus involves translation of appropriate scene priors and task requirements to perform quick hypothesis generation and fusion.

Goal

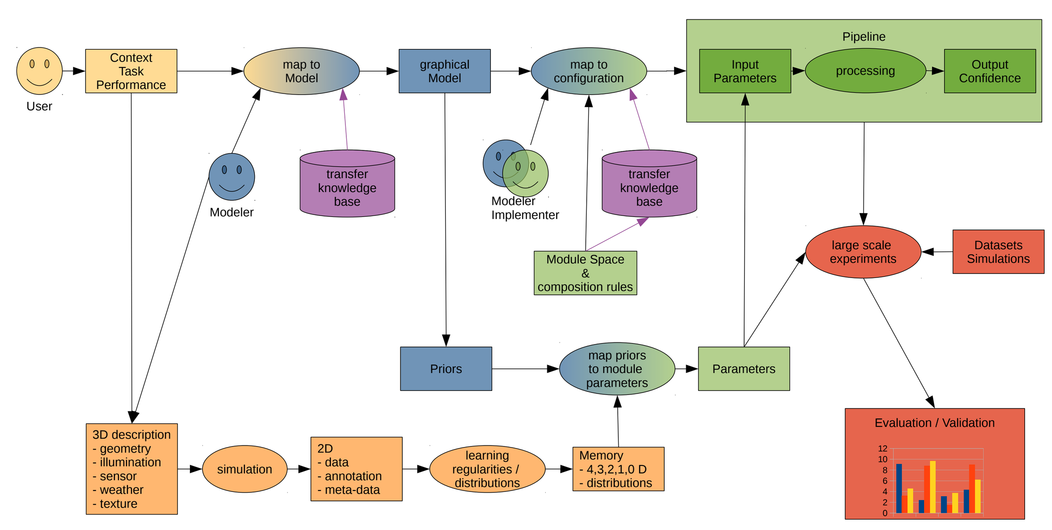
The startup venture will focus on next generation engineering tools that will **simplify design, operation, and validation** of large scale vision and cognition applications. The emphasis is on aggregating open-source platforms and on differentiating technologies in systems science and engineering outlined above.

Case studies

In this presentation, anomaly detection from video for an automotive application is used as a case study to demonstrate our engineering workflow and platform capabilities [8]. We highlight how to translate user requirements to graphical models for contexts that are in turn translated to approximate inference engines. We also illustrate how our platform is used for evaluating the performance of the inference engine on real as well as synthetic data.

Workflows

Systems Engineering Workflow illustrates the link between User, Modeler, Implementer and Validation viewpoints.



Case study: brake light detection in automotive video sequences

User view

- Task: brake light transition detection system on monocular video sequences. Detect brake light state changes including confidence level of detections.
- Context: Traffic settings: 'urban areas', 'country roads' and 'highways' with multiple lanes, traffic densities and junction settings. Illumination & Weather Conditions: "day", "night", "dawn", "dusk", "overcast", "spray", "fog", "raining", "frost", "snow", etc.
- Performance: Near-human level performance with self-diagnostics. Real-time implementation on embedded platforms. Academic prototype's focus is on demonstration on methodology.

Modeler view

The modeler postulates the low entropy factors in the OOBN, what factors require invariance (i.e. what sub-modalities are relevant for task) and chooses invariant modules.

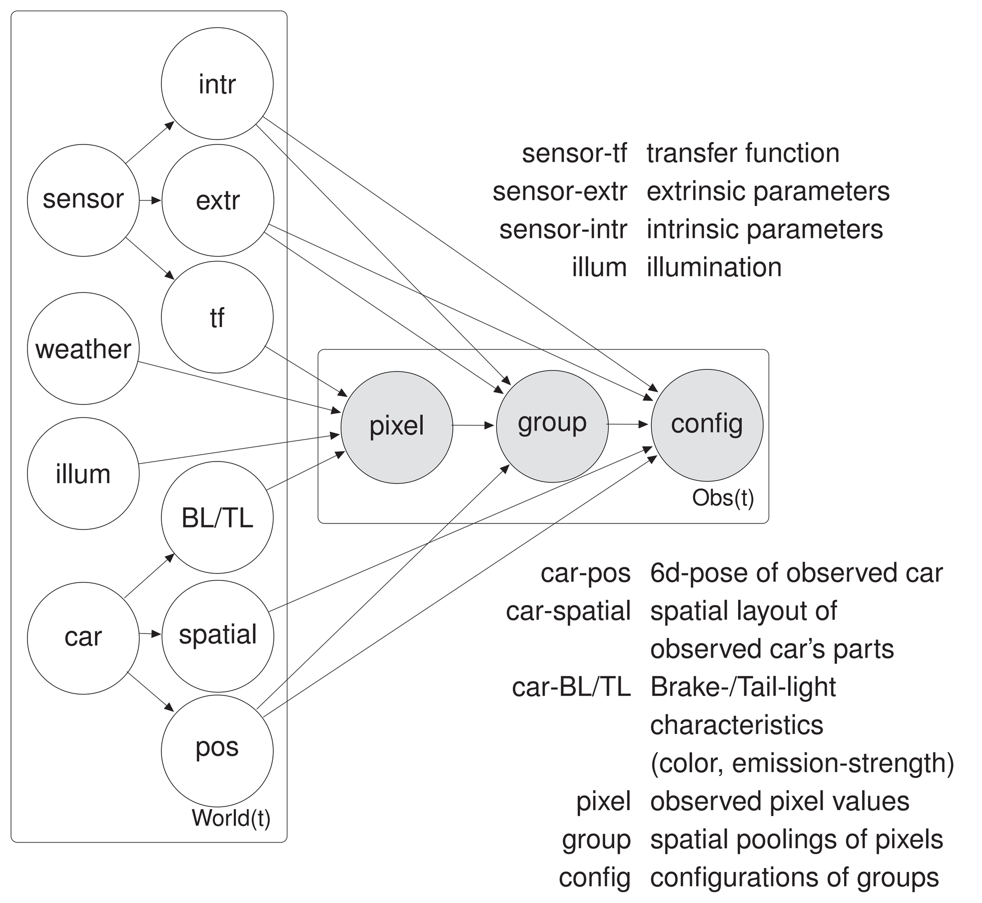
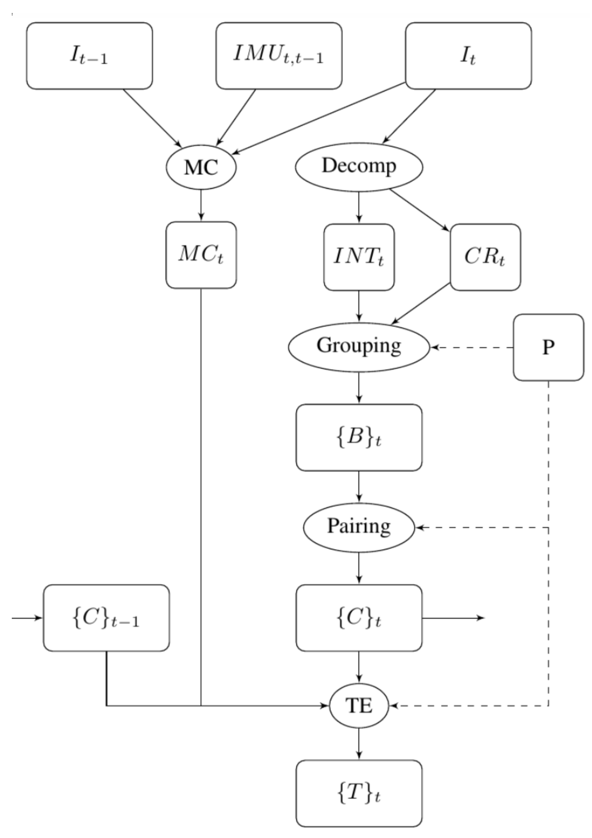


Figure: Graphical model for the task of brake-light detection. The **World** plate contains latent variables on which the observed variables in **Obs** are conditioned at time t.

Systems analysis is used to characterize invariant module's performance in context (either on real or simulated data) and derives an extended OOBN that includes the quasi-invariants. Deliberative reasoning may also be incorporated at this stage.

Implementer view

As described in [1], we implemented a cognitive system that consists of multimodal quasi-invariant filters that map to the variables identified in the modeling phase. A memory storage contains prior hypotheses on the world and is indexed by context and current state estimates, selecting and tuning parameters of the estimators.



Evaluation view

Based on the implementation, large-scale experiments have been conducted and analyzed.

Work	# imgs	D/N	Recog	FN	FP
[Liu2015]	1983	D	85.6%	12.52%	16.65%
DSEQ	439	D	97.3 %	2.7%	0.3%
DFP	40	D	92.5 %	7.5%	17.5%
[Chen2012]	26	N	84.6%	15.4%	N/A
[Tham2009]	45	N	86.67%	13.3%	6.665%
DSEQ	549	N	88.2%	3.3%	8.6%
DFP	16	N	93.75%	6.25%	6.25%

Table: Evaluation of state-estimation without temporal information. D/N denotes day or night conditions, Recog the detection rate, DSEQ and DFP denote our system evaluated on two datasets.

Simulation

Simulation is utilized during the modeling stage to populate prior memory representations.

C,**T**,**P** → 3D description → Simulation → Memory

These memory representations are used by the inference engines of the realtime system.

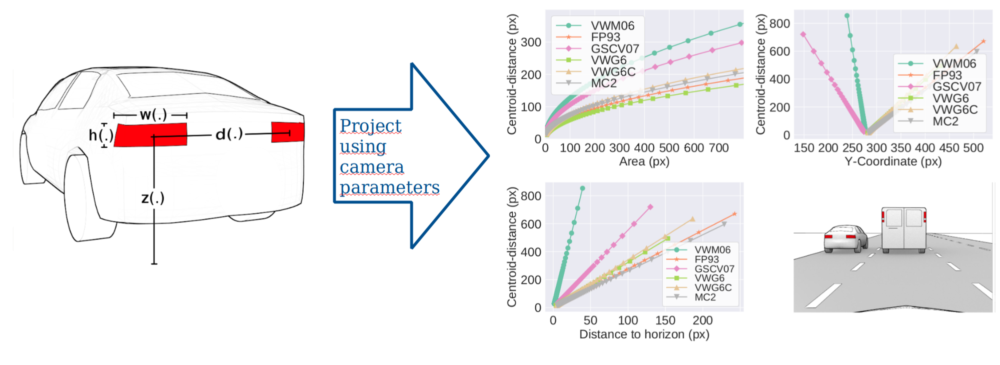


Figure: Priors on geometrical brake/tail-light statistics

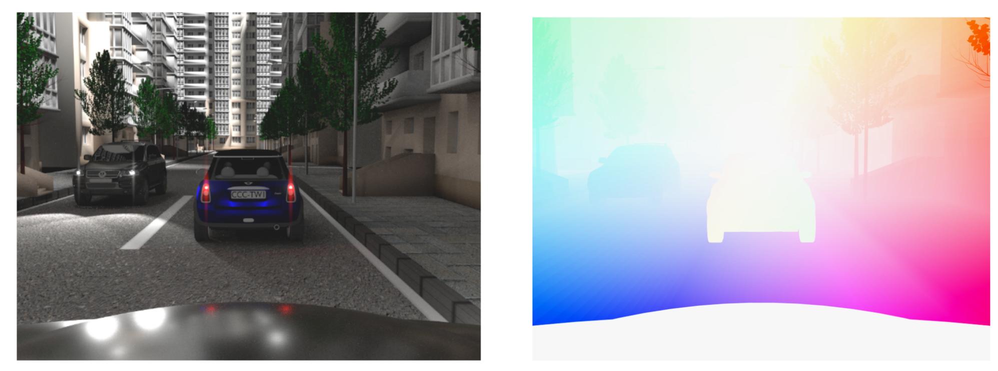
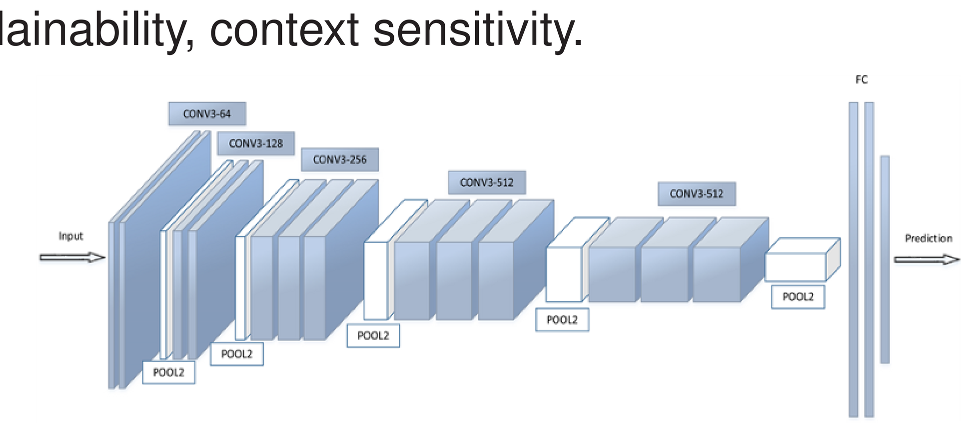


Figure: Simulation for expected flow given scene geometry and velocity

Integration and Fusion of perspectives

The systems engineering process that links explicitly context, task, performance specs to designs allows for explainability, context sensitivity.



Moreover, it allows ease of integration of Model-based perspectives with modern machine learning (deep-learning) perspectives to provide various degrees of transparency. Ongoing research in the context of industry and European Union's horizon 2020 projects address these aspects.

Results

Principled analysis and modeling of involved variables allows for the selection of suited estimators for a given task in context. Priming of estimators from given memory representations allows to set parameters that adapt to a given context.

Figure: Two frames from a braking event (top: time t, bottom: t+1). First column: input images, second column: map of c_R, third column: magnified transition estimations (T:Off-transition from on to off, T:N-no transition detected.)

We have shown the design and implementation of workflows following the explained methodologies and presented a case study.

Acknowledgments

This work was funded by the German Federal Ministry of Education and Research (BMBF), project 01GQ0840 and 01GQ0841 (Bernstein Focus: Neurotechnology Frankfurt).

Bibliography

[1] V. Ramesh and C. von der Malsburg. Systems engineering for visual cognition. bernstein conference, 2013.
[2] V. Ramesh. Performance characterization of image understanding algorithms, 1995.
[3] et al. Greiffenhagen, Mi. Design, analysis, and engineering of video monitoring systems: an approach and a case study.
[4] C. von der Malsburg. A vision architecture based on fiber bundles. bernstein conference, 2012.
[5] W. Mann. 3d object interpretation from monocular images, 1996.
[6] Zhuowen Tu, Xiangrong Chen, Alan L Yuille, and Song-Chun Zhu. Image parsing: Unifying segmentation, detection, and recognition. *International Journal of computer vision*, 63(2):113–140, 2005.
[7] Amy Perfors, Joshua B Tenenbaum, Thomas L Griffiths, and Fei Xu. A tutorial introduction to bayesian models of cognitive development. *Cognition*, 120(3):302–321, 2011.
[8] RN Hota, P. Harding, VS Veeravasarapu, and V. Ramesh. Top-down contextual modeling for vision, 2014.
[9] T.Weis, M.Mundt, P.Harding, V.Ramesh. Anomaly detection for automotive visual signal transition estimation, 2017.