TUTORIAL ON:

COGNITIVE VISION

ON DEEP SEMANTICS FOR EXPLAINABLE VISUOSPATIAL COMPUTING

TUTORIAL PRESENTERS

Mehul Bhatt, Jakob Suchan
Örebro University (Sweden) – University of Bremen (Germany)

TUTORIAL SUMMARY

The tutorial on cognitive vision addresses computational vision and perception at the interface of language, logic, cognition, and artificial intelligence. The tutorial focusses on application areas where the processing and explainable semantic interpretation of (potentially large volumes of) dynamic visuospatial imagery is central, e.g., for commonsense scene understanding; visual cognition for cognitive robotics / HRI, autonomous driving; narrative interpretation from the viewpoints of visuoauditory perception & digital media design, semantic interpretation of multimodal human-behavioural data.

The tutorial highlights Deep (Visuospatial) Semantics, denoting the existence of systematically formalised declarative AI methods –e.g., pertaining to reasoning about space and motion– supporting semantic (visual) question-answering, relational learning, non-monotonic (visuospatial) abduction, and simulation of embodied interaction. The tutorial demonstrates the integration of methods from knowledge representation and computer vision with a focus on (combining) reasoning & learning about space, action, motion, and (multimodal) interaction. This is presented in the backdrop of areas as diverse as autonomous driving, cognitive robotics, eye-tracking driven visual perception research (e.g., for visual art, architecture design, cognitive media studies), and psychology & behavioural research domains where data-centred analytical methods are gaining momentum. The tutorial covers both applications and basic methods concerned with topics such as: explainable visual perception, semantic video understanding, language generation from video, declarative spatial reasoning, and computational models of narrative. The tutorial will position an emerging line of research that brings together a novel & unique combination of research methodologies, academics, and communities encompassing AI, ML, Vision, Cognitive Linguistics, Psychology, Visual Perception, and Spatial Cognition and Computation.

ABOUT THE TUTORIAL

Cognitive vision as an area of research has gained prominence, with several recent initiatives addressing the topic from the perspectives of language, logic, and artificial intelligence. There has also been an increased interest from the computer vision community to synergise with cognitively motivated methods for perceptual grounding and inference with visual imagery. Most recently, neurosymbolic integration of vision and semantics has gained momentum to cater to explainability requirements.

This tutorial focusses on application areas where the processing and semantic interpretation of (potentially large volumes of) highly dynamic visuo-spatial imagery is central: embodied cognitive vision for robotics; dynamic imagery & narrativity from the viewpoint of research visual perception and embodiment multimodal interaction; commonsense scene understanding etc. In the backdrop of applied areas as diverse as autonomous driving, (evidence-based) architecture design, cognitive media studies, cognitive robotics, this tutorial will pursue a twofold objective encompassing applications and basic methods concerned with:

- semantic video understanding
- embodied grounding and language generation from video
- declarative spatial reasoning
- computational models of narrative
- moving image (analysis)
- semantics with large-scale dynamic imagery
- visual perception (eye-tracking with a focus on media, and wayfinding research for architecture design)
- evidence based studies in visual perception

DEEP VISUOSPATIAL SEMANTICS. The high-level semantic interpretation and qualitative analysis of dynamic visuo-spatial imagery requires the representational and inferential mediation of commonsense abstractions of *space, time, motion, and interaction*. In this backdrop, deep (visuo-spatial) semantics denotes the existence of systematic formalisation and declaratively grounded models —e.g., pertaining to space and motion— that can be used to perform reasoning and query answering, relational learning, embodied grounding and simulation etc.

The tutorial will emphasise the integration of knowledge representation and reasoning methods with computer vision with a focus on the development of hybrid AI and cognitive interaction technologies where reasonign about space, action, and change is central. The tutorial will cover both theory and applications:

1. Theory, methods, and tools. demonstrating KR-based general methods and tools for commonsense reasoning about space and motion in the context of visual data (e.g., image, video, point-clouds); the focus will be on semantic Q/A with video
2. Applications. highlighting a general class of scenarios where high-level cognitive vision or human-like perceptual sense-making of visual input is crucial (e.g., for analysis, interpretation, relational learning)

1. THEORETICAL FOUNDATIONS

Commonsense spatial, temporal, and spatio-temporal relations and patterns (e.g.,"left", "overlap", "during", "between", "separation", "collision") serve as powerful abstractions for the spatio-linguistic grounding of visual perception and embodied action & interaction. Such spatio-linguistic primitives constitute the basic ontological building blocks of visuo-spatial computing in diverse areas. With regard to foundational theory and methods / tools, the tutorial will present computational *visuo-spatial representation and reasoning* from the viewpoint of the research areas of *artificial intelligence*, *commonsense reasoning*, and *spatial cognition and computation*. The focus will be on mixed qualitative-quantitative reasoning about space, space-time, and motion.

2. APPLICATIONS

From an application viewpoint, the tutorial will present an overview of emerging research and application outcomes concerning reasoning about moving objects in autonomous driving, moving image analysis & cognitive vision, visual semantics & embodied grounding, computational models of narrative for the moving image, and visuospatial perception & cognition in the context of a range of areas concerned with evidence-based studies in visual perception. Select case studies developed as a part of an ongoing large-scale experiments in HMI and cognitive media studies will be presented.

PRESENTER BIOGRAPHIES

Prof. Dr. Mehul Bhatt
AASS – School of Science and Technology
Örebro University, SWEDEN

CoDesign Lab EU. www.codesign-lab.org
mehul.bhatt@oru.se | www.mehulbhatt.org

Mehul Bhatt is Professor within the School of Science and Technology at Orebro University (Sweden). His basic research focusses on formal, cognitive, and computational foundations for AI technologies with a principal emphasis on knowledge representation, semantics, integration of commonsense reasoning & learning, explainability, and spatial representation and reasoning. Mehul Bhatt steers CoDesign Lab (www.codesign-lab.org), an initiative aimed at addressing the confluence of Cognition, Artificial Intelligence, Interaction, and Design Science for the development of human-centred cognitive assistive technologies and interaction systems. Since 2014, he directs the research and consulting group DesignSpace (www.design-space.org) and pursues ongoing research in Cognitive Vision (www.codesign-lab.org/cognitive-vision) and Spatial Reasoning (www.spatial-reasoning.com).

Jakob Suchan
Human-Centred Cognitive Assistance Lab. (AG HCC)
Department of Mathematics and Informatics
Bremen, GERMANY
Tel: +49 (0)421-218-64197 | jsuchan@uni-bremen.de
http://hcc.uni-bremen.de/ | www.codesign-lab.org/cognitive-vision

Jakob Suchan is researcher within the Human-Centred Cognitive Assistance Lab at the Faculty of Mathematics and Informatics, University of Bremen, Germany. His research is in the area of cognitive vision (www.cognitive-vision.org), particularly focussing on the integration of vision and AI (specifically, KR) from the viewpoint of computational cognitive systems where integrated (embodied) perception and interaction are involved. Jakob is also a member of the DesignSpace Group (www.design-space.org).