# ANALYSING DRIVER (IN)ATTENTIVENESS

## / Towards a Cognitive Complexity model combining Visuospatial and Interactional parameters /

[ https://www.codesign-lab.org/cognitive-vision/ ]

VASILIKI KONDYLI - MEHUL BHATT

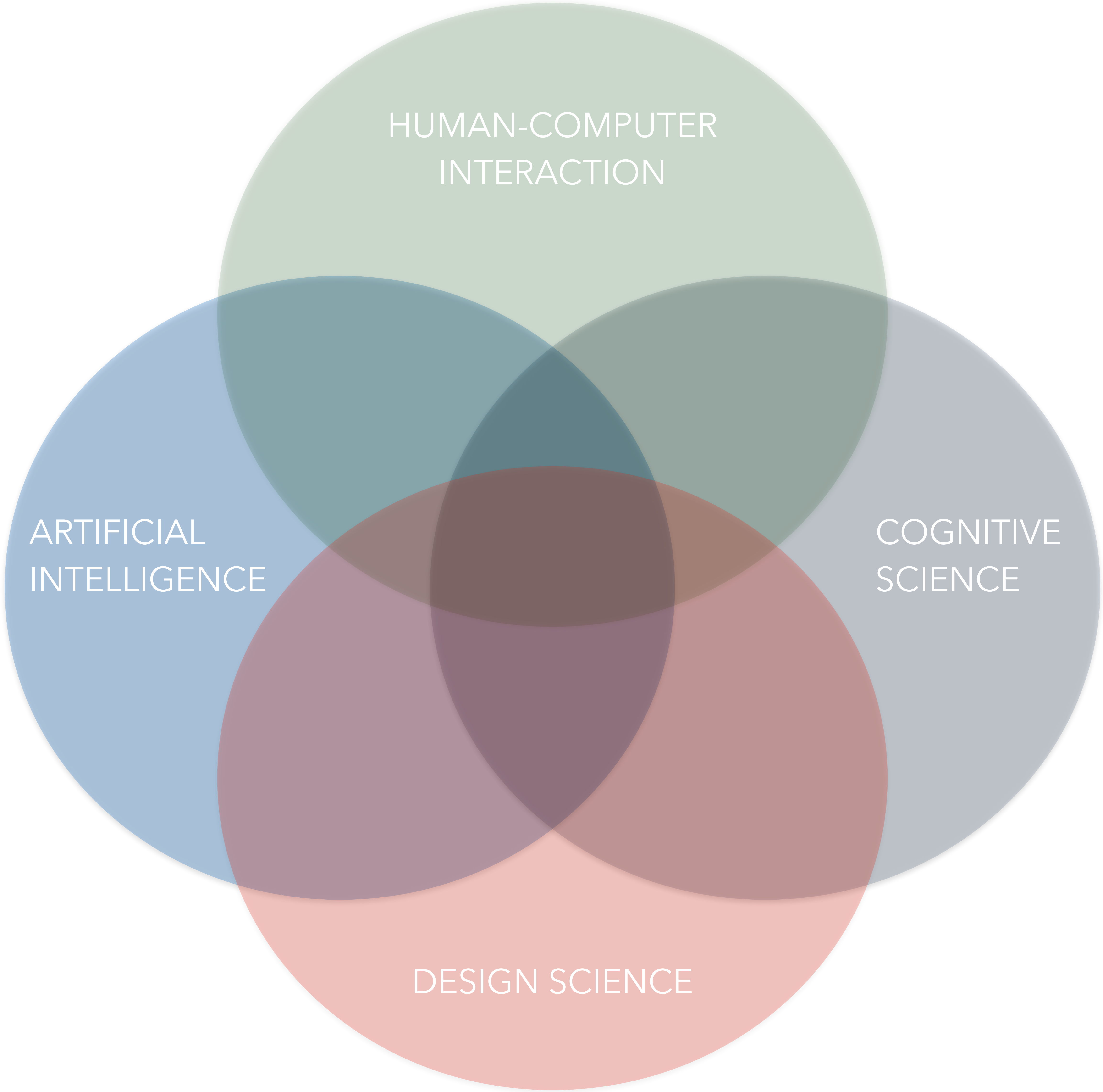
ÖREBRO UNIVERSITY

CODESIGN LAB EU / COGNITION. AI. INTERACTION. DESIGN.

DDI 2022 / OCTOBER 19 2022

DesignSpace

ÖREBRO UNIVERSITET

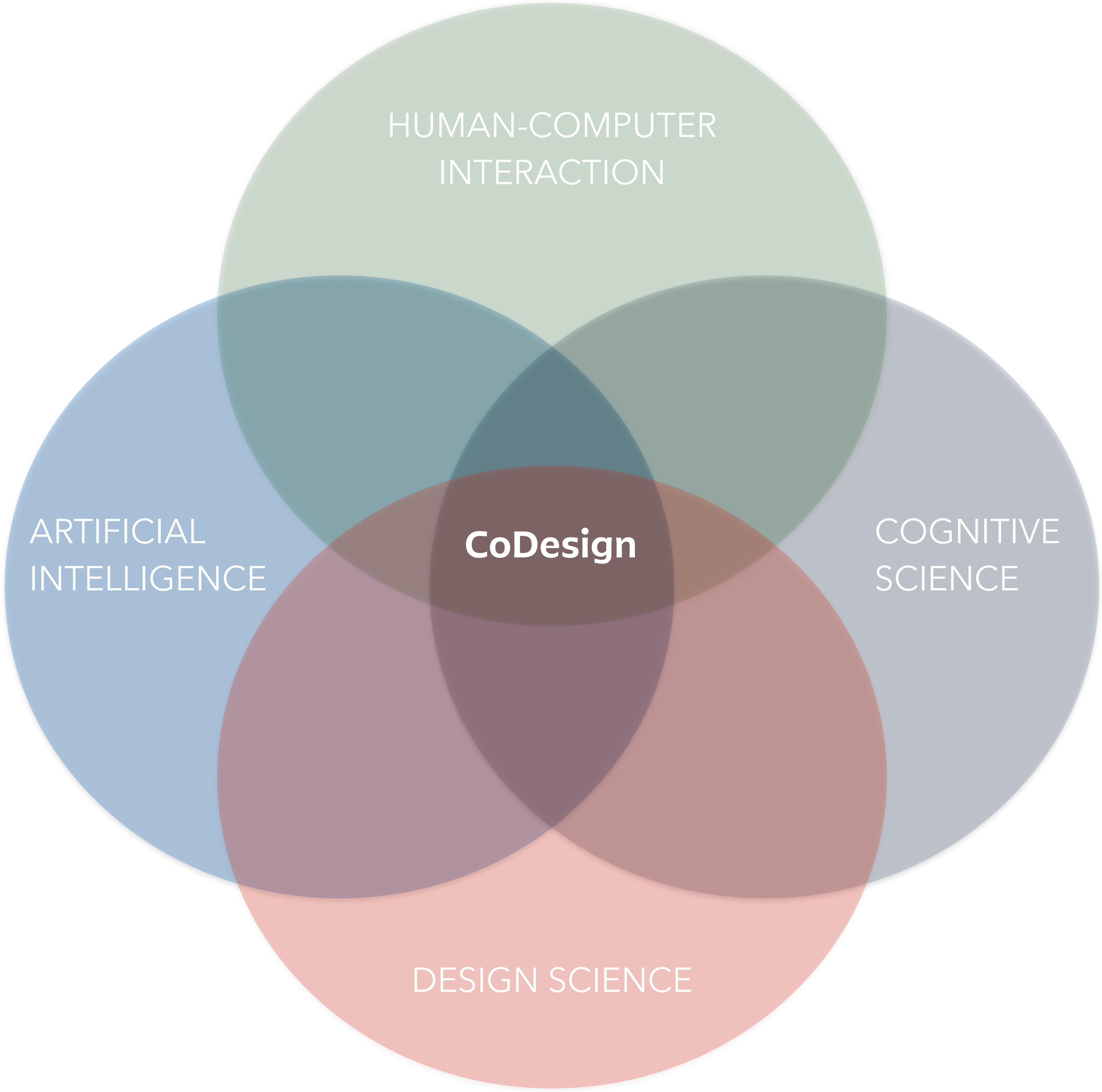

HUMAN-COMPUTER
INTERACTION
ARTIFICIAL
INTELLIGENCE
COGNITIVE
SCIENCE
DESIGN SCIENCE

# CODESIGN LAB

COGNITION. ARTIFICIAL INTELLIGENCE. INTERACTION. DESIGN.

CoDesign is CONFLUENCE of (four) research areas, driven by bottom-up interdisciplinarity, and a mixed-methods mindset for identifying and pursuing human-behavioural research.

Focus on human-centred, socio-technologically motivated opportunities where basic research in Artificial Intelligence and Spatial Cognition, and Technological Artefacts emanating therefrom, ought to demonstrate their application, impact, and societal significance.



"Develop a framework that serves as a uniform and shared reference for (replicable) experiments in visual (in)attention under ecologically valid naturalistic conditions encountered in everyday activities and related to decision-making."

INTERPERSONAL COMMUNICATION

EMBODIED MULTIMODAL INTERACTIONS

VISUOSPATIAL COMPLEXITY - CONTEXT

COMPLEXITY OF HUMAN PERFORMANCE

- How do people establish joint attention during interpersonal communication in the driving scene?
- How does the visuospatial complexity of the environment affect gaze behaviour of a driver?
- How can we model cognitive human factors so that we can systematically study the combination of aspects as found in naturalistic driving scenes?
- How a holistic model of visuospatial complexity can be useful for the development of novel autonomous driving systems?

# RELATED PUBLICATIONS / AI AND COGNITION

[ https://www.codesign-lab.org/cognitive-vision/ ]

[ https://codesign-lab.org/select-papers/ ]

# VISUO-LOCOMOTIVE PERCEPTION / KEY PUBLICATIONS

CO DESIGN LAB EU https://codesign-lab.org/

## Towards a Human-Centred Cognitive Model of Visuospatial Complexity in Everyday Driving

Vasiliki Kondyli Mehul Bhatt Jakob Suchan

Örebro University, Sweden
University of Bremen, Germany
CoDesign Lab > Cognitive Vision
www.codesign-lab.org/cognitive-vision

### Abstract

We develop a human-centred cognitive model of visuospatial complexity in everyday, naturalistic driving conditions. With a focus on visual perception, the model incorporates quantitative, structural, and dynamic attributes identifiable in the chosen context; the human-centred basis of the model lies in its behavioural evaluation with human subjects with respect to psychophysical measures pertaining to embodied visuoauditory attention. We report preliminary steps to apply the developed cognitive model of visuospatial complexity for human-factors guided dataset creation and benchmarking, and for its use as a semantic template for the (explainable) computational analysis of visuospatial complexity.

### 1 Introduction

Autonomous driving research has received enormous academic & industrial interest in recent years. As the self-driving vehicle industry develops, it will be necessary —in a manner similar to sectors such as medical computing, computer aided engineering and design— to have an articulation and community consensus on aspects such as representation, interoperability, data archival & retrieval mechanisms, and human-centred performance benchmarks. We posit that this will be necessary towards fulfilling essential legal and ethical responsibilities, such as those pertaining to representation and realisation of rules and norms, verifiable performance of normative behaviour, and technology capabilities vis-a-vis human-centred expectations.

**Human-Centred Benchmarking and Standardisation.** Within autonomous driving, the need for standardisation and ethical regulation has most recently garnered interest internationally, e.g., with the Federal Ministry of Transport and Digital Infrastructure in Germany (BMVI) taking a lead in eliciting 20 key propositions (with possible legal implications) for the fulfilment of ethical commitments for automated and connected driving systems [21]. In spite of major investments in self-driving vehicle research, issues related to human-centred design, human-machine interaction, and standardisation have been barely addressed, with the current focus in driving research primarily being on two basic considerations: *how fast to drive, and which way and how much to steer*. This is necessary, but inadequate if autonomous vehicles are to become commonplace and function with humans: not everything in autonomous vehicles is about realtime control/decision-making; several human-machine interaction requirements (e.g., for diagnostic communication, universal design) also exist. The 20 key propositions elicited by the German federal ministry BMVI highlight a range of factors pertaining to safety, utilitarian considerations, human rights, statutory liability, technological transparency, data management and privacy etc. Ethically driven standardisation and regulation will require addressing challenges in foundational human-centred AI technology,

In: STAIRS at ECAI 2020: European Conference on Artificial Intelligence 2020.

*DHM2020*
*L. Hanson et al. (Eds.)* 131

*doi:10.3233/ATDE200018*

## Multimodality on the Road: Towards Evidence-Based Cognitive Modelling of Everyday Roadside Human Interactions

Vasiliki KONDYLI, and Mehul BHATT
*Örebro University, Sweden*
*CoDesign Lab – Cognitive Vision*
www.codesign-lab.org/cognitive-vision

**Abstract** We propose an evidence based methodology for the systematic analysis and cognitive characterisation of multimodal interactions in naturalistic roadside situations such as driving, crossing a street etc. Founded on basic human modalities of embodied interaction, the proposed methodology utilises three key characteristics crucial to roadside situations, namely: explicit and implicit mode of interaction, formal and informal means of signalling, and levels of context-specific (visual) attention. Driven by the fine-grained interpretation and modelling of human behaviour in naturalistic settings, we present an application of the proposed model with examples from a work-in-progress dataset consisting of baseline multimodal interaction scenarios and variations built therefrom with a particular emphasis on joint attention and diversity of modalities employed. Our research aims to open up an interdisciplinary frontier for the human-centred design and evaluation of artificial cognitive technologies (e.g., autonomous vehicles, robotics) where embodied (multimodal) human interaction and normative compliance are of central significance.

**Keywords.** multimodal interaction, interpersonal communication, naturalistic perception, joint attention, virtual reality, autonomous driving

### 1. Introduction

Interpersonal communication and interactions are vital for safe and effective coordination of actions in everyday roadside engagements: walking around, driving, riding a bike etc. Failure in interpersonal communication leads to a lack of mutual understanding of a situation and it is responsible for a great number of roadside accidents [1, 2]. With further strides in the autonomous vehicles industry and the present impetus on high-level visual intelligence technologies [3, 4], it will therefore be necessary to account for the role of interpersonal communication on the street and articulate human-centred performance benchmarks, e.g., from the viewpoint of training, testing and validation as part of statutory compliance measures.



## Visuo-Locomotive Complexity as a Component of Parametric Systems for Architecture Design

Vasiliki Kondyli, Mehul Bhatt (Örebro University, Sweden)
Evgenia Spyridonos (University of Stuttgart, Germany)

**Abstract** A people-centred approach for designing large-scale built-up spaces necessitates systematic anticipation of user's embodied visuo-locomotive experience from the viewpoint of human-environment interaction factors pertaining to aspects such as navigation, wayfinding, usability. In this context, we develop a behaviour-based visuo-locomotive complexity model that functions as a key correlate of cognitive performance vis-a-vis internal navigation in built-up spaces. We also demonstrate the model's implementation and application as a parametric tool for the identification and manipulation of the architectural morphology along a navigation path as per the parameters of the proposed visuospatial complexity model. We present examples based on an empirical study in two healthcare buildings, and showcase the manner in which a dynamic and interactive parametric (complexity) model can promote behaviour-based decision-making throughout the design process to maintain desired levels of visuospatial complexity as part of a navigation or wayfinding experience.

**Keywords:** Visual Perception – Environmental Psychology – Architecture Design – Parametric Design – Cognitive Computational Modelling – Spatial Cognition – AI and Design

### Publication Note

This is a preprint of the contribution published as part of the proceedings of ICoRD 2021: 8th International Conference on Research into Design, IDC School of Design (IIT Mumbai, India).
ICoRD 2021, www.idc.iitb.ac.in/icord2021/.

The overall scientific agenda driving this research may be consulted here:
The DesignSpace Group / www.design-space.org

Select publications related to this research are available at:
www.codesign-lab.org/select-papers (under DesignSpace)

# Grounding Embodied Multimodal Interaction

Towards Behaviourally Established Semantic Foundations for Human-Centred AI

Vasiliki Kondyli[1], Jakob Suchan[2] and Mehul Bhatt[1]

[1] *Örebro University, Sweden*
[2] *German Aerospace Center (DLR), Germany*

*CoDesign Lab » Cognition. AI. Interaction. Design.*
*info@codesign-lab.org / https://codesign-lab.org*

**Abstract**

We position recent and emerging research in *cognitive vision and perception* addressing three key questions: **(1)** What kind of relational abstraction mechanisms are needed to perform (explainable) grounded inference –e.g., question-answering, qualitative generalisation, hypothetical reasoning– relevant to embodied multimodal interaction? **(2)** How can such abstraction mechanisms be founded on behaviourally established cognitive human-factors emanating from naturalistic empirical observation? and **(3)** How to articulate behaviourally established abstraction mechanisms as formal declarative models suited for grounded knowledge representation and reasoning (KR) as part of large-scale hybrid AI and computational cognitive systems.

We contextualise (1–3) in the backdrop of recent results at the interface of AI/KR, and Spatial Cognition and Computation. Our main purpose is to emphasise the importance of behavioural research based foundations for next-generation, human-centred AI, e.g., as relevant to applications in Autonomous Vehicles, Social and Industrial Robots, and Visuo-Auditory Media.

**Keywords**

Multimodal Interaction, Commonsense Reasoning, Declarative Spatial Reasoning, Declarative AI, Explainable AI, Cognitive Human-Factors, Cognitive Systems

## 1. Motivation

Multimodality in interaction is an inherent aspect of human activity, be it in social, professional, or everyday mundane contexts. Next-generation AI technologies, aiming for compliance with human-centred ethical and legal requirements, performance benchmarks, and inclusive usability expectations will require an inherent foundational capacity to analyse –e.g., understand, explain, anticipate– everyday interactional multimodality in naturalistic settings involving technology mediated collaborative assistance of humans. Amongst other things, this necessitates that the foundational building blocks of such next-generation systems be semantically aligned with the descriptive complexity of human task conceptualisation and performance expectations.

**Declaratively Mediated Multimodality.** The significance of "*grounding*" in semiotic construction, e.g., enabling high-level meaning-making, has been long-established in Artificial

*KR4HI 2022: Knowledge Representation for Hybrid Intelligence, June 14, 2022, Amsterdamm, The Netherlands*
vasiliki.kondyli@oru.se (V. Kondyli); jakob.suchan@dlr.de (J. Suchan); mehul.bhatt@oru.se (M. Bhatt)

CEUR Workshop Proceedings (CEUR-WS.org)

---

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

ELSEVIER

# Commonsense visual sensemaking for autonomous driving – On generalised neurosymbolic online abduction integrating vision and semantics ☆

Jakob Suchan[a,c], Mehul Bhatt[b,c], Srikrishna Varadarajan[c]

[a] *University of Bremen, Germany*
[b] *Örebro University, Sweden*
[c] *CoDesign Lab (EU), Cognitive Vision, Sweden*

A R T I C L E I N F O

*Article history:*
Received 19 July 2020
Received in revised form 29 April 2021
Accepted 30 April 2021
Available online 7 May 2021

*Keywords:*
Cognitive vision
Deep semantics
Declarative spatial reasoning
Knowledge representation and reasoning
Commonsense reasoning
Visual abduction
Answer set programming
Autonomous driving
Human-centred computing and design
Standardisation in driving technology
Spatial cognition and AI

A B S T R A C T

We demonstrate the need and potential of systematically integrated *vision* and *semantics* solutions for visual sensemaking in the backdrop of autonomous driving. A general neurosymbolic method for *online* visual sensemaking using answer set programming (ASP) is systematically formalised and fully implemented. The method integrates state of the art in visual computing, and is developed as a modular framework that is generally usable within hybrid architectures for realtime perception and control. We evaluate and demonstrate with community established benchmarks KITTIMOD, MOT-2017, and MOT-2020. As use-case, we focus on the significance of human-centred visual sensemaking —e.g., involving semantic representation and explainability, question-answering, commonsense interpolation— in safety-critical autonomous driving situations. The developed neurosymbolic framework is domain-independent, with the case of autonomous driving designed to serve as an exemplar for online visual sensemaking in diverse cognitive interaction settings in the backdrop of select human-centred AI technology design considerations.



## 1. Motivation

Autonomous driving research has received enormous academic & industrial interest in recent years (Sec 5). This surge has coincided with (and been driven by) advances in *deep learning* based computer vision research. Although end-to-end deep learning based vision and control has (arguably) been successful for self-driving vehicles, we posit that there is a clear need and tremendous potential for hybrid visual sensemaking solutions that integrate *vision and semantics* towards fulfilling essential legal and ethical responsibilities involving explainability, human-centred AI (Artificial Intelligence), and industrial standardisation (e.g, pertaining to representation, realisation of rules and norms, fulfilling statutory obligations).

☆ This paper is an invited contribution based on a distinguised paper nomination at the 2019 International Joint Conference on Artificial Intelligence (IJCAI-19).

*E-mail address:* info@codesign-lab.org (M. Bhatt).
*URL:* http://www.codesign-lab.org (M. Bhatt).

https://doi.org/10.1016/j.artint.2021.103522

# A COGNITIVE MODEL OF VISUOSPATIAL COMPLEXITY / INTERACTIONS

**A. Mode**

Explicit Interaction
Implicit Interaction

**B. Method**

Formal Device-based
Informal - Device-based
Formal Body-based
Informal Body-based

**C. Level of Social Attention**

Individual
Monitoring
Common
Mutual
Joint (Shared)

**Modalities**

Speech
Head Movements
Facial Expressions
Gestures
Body Postures
Gaze
Auditory Cues

Kondyli, V. & Bhatt, M. (2020). Multimodality on the Road: Towards Evidence-Based Cognitive Modelling of Human Interactions in Everyday Roadside Situations. Advances in Transdisciplinary Engineering, II, Proceedings of DHM 2020: 6th International Digital Human Modelling Symposium 2020 (DHM 2020), Skövde, Sweden (pp. 131 - 142), IOS Press.

**A. Mode**

Explicit Interaction
Implicit Interaction

**B. Method**

Formal Device-based
Informal - Device-based
Formal Body-based
Informal Body-based

**C. Level of Social Attention**

Individual
Monitoring
Common
Mutual
Joint (Shared)

**Modalities**

Speech
Head Movements
Facial Expressions
Gestures
Body Postures
Gaze
Auditory Cues

**JOINT ATTENTION**

**INFORMAL - BODY-BASED** **EXPLICIT**

Kondyli, V. & Bhatt, M. (2020). Multimodality on the Road: Towards Evidence-Based Cognitive Modelling of Human Interactions in Everyday Roadside Situations. Advances in Transdisciplinary Engineering, II, Proceedings of DHM 2020: 6th International Digital Human Modelling Symposium 2020 (DHM 2020), Skövde, Sweden (pp. 131 - 142), IOS Press.

**A. Mode**

- Explicit Interaction
- Implicit Interaction

**B. Method**

- Formal Device-based
- Informal - Device-based
- Formal Body-based
- Informal Body-based

**C. Level of Social Attention**

- Individual
- Monitoring
- Common
- Mutual
- Joint (Shared)

**Modalities**

- Speech
- Head Movements
- Facial Expressions
- Gestures
- Body Postures
- Gaze
- Auditory Cues


Kondyli, V. & Bhatt, M. (2020). Multimodality on the Road: Towards Evidence-Based Cognitive Modelling of Human Interactions in Everyday Roadside Situations. Advances in Transdisciplinary Engineering, 11, Proceedings of DHM 2020: 6th International Digital Human Modelling Symposium 2020 (DHM 2020), Skövde, Sweden (pp. 131 - 142), IOS Press.

**A. Mode**

Explicit Interaction
Implicit Interaction

**B. Method**

Formal Device-based
Informal - Device-based
Formal Body-based
Informal Body-based

**C. Level of Social Attention**

Individual
Monitoring
Common
Mutual
Joint (Shared)

**Modalities**

Speech
Head Movements
Facial Expressions
Gestures
Body Postures
Gaze
Auditory Cues

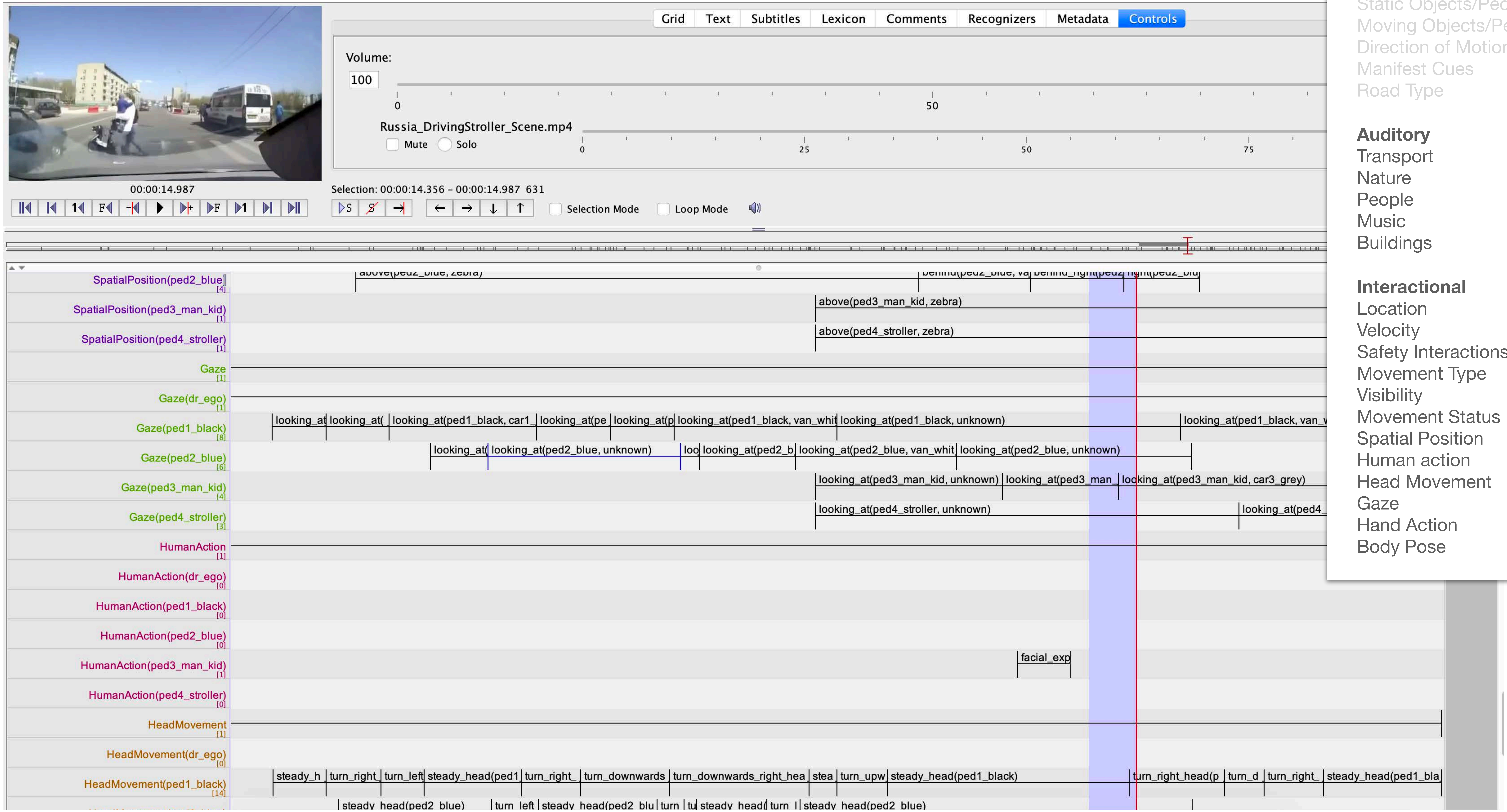
## MUTUAL ATTENTION

## INFORMAL - BODY-BASED

## IMPLICIT


Kondyli, V. & Bhatt, M. (2020). Multimodality on the Road: Towards Evidence-Based Cognitive Modelling of Human Interactions in Everyday Roadside Situations. Advances in Transdisciplinary Engineering, 11, Proceedings of DHM 2020: 6th International Digital Human Modelling Symposium 2020 (DHM 2020), Skövde, Sweden (pp. 131 - 142), IOS Press.

# INTERACTIONAL ATTRIBUTES / SEMANTIC ANNOTATIONS




**Visuospatial**
Static Objects/People
Moving Objects/People
Direction of Motion
Manifest Cues
Road Type

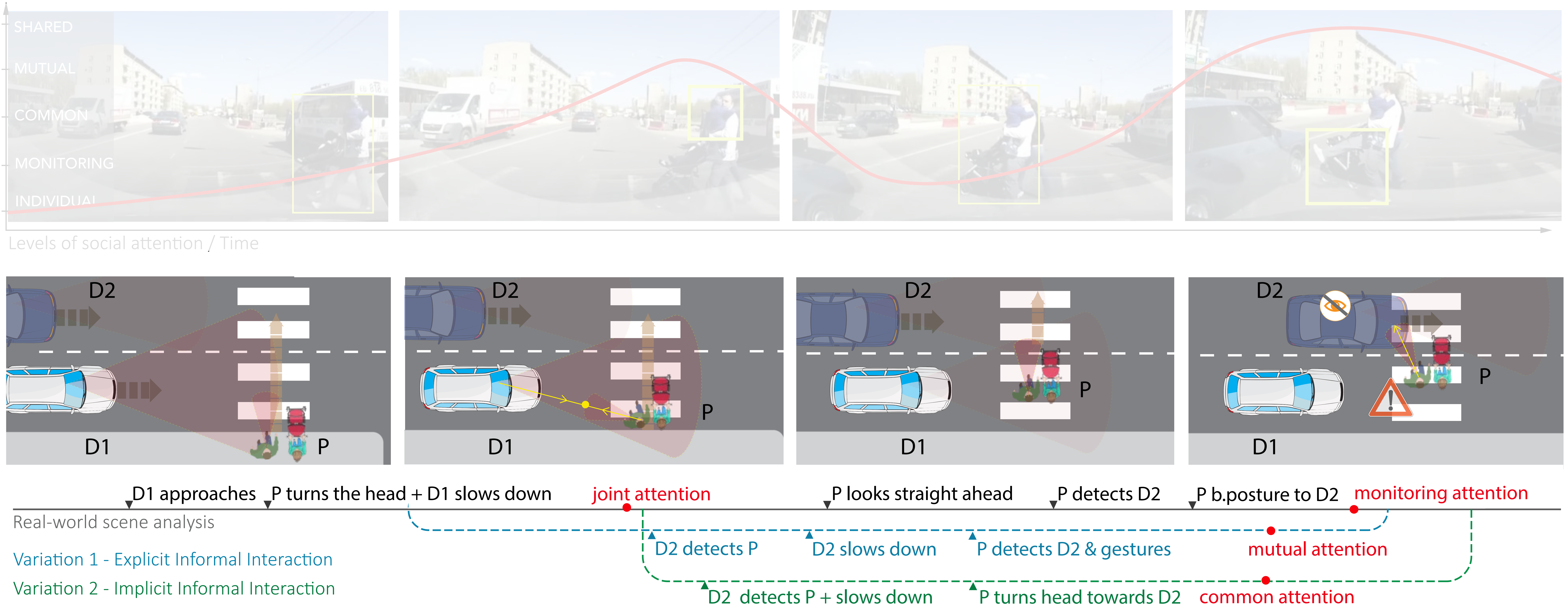
**Auditory**
Transport
Nature
People
Music
Buildings

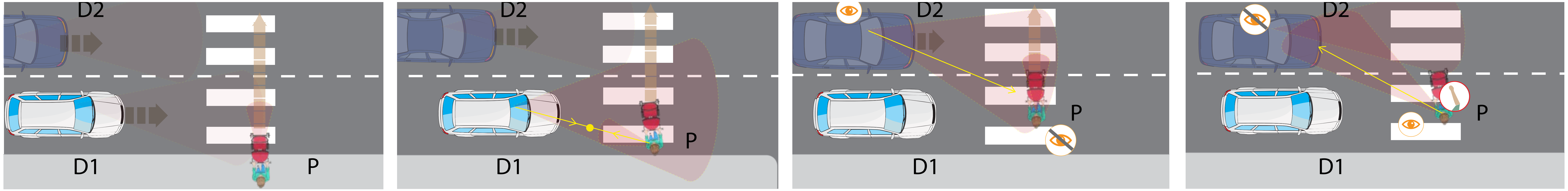
**Interactional**
Location
Velocity
Safety Interactions
Movement Type
Visibility
Movement Status
Spatial Position
Human action
Head Movement
Gaze
Hand Action
Body Pose

**SP** (Speech) — **HM** (Head Movement) — **FE** (Facial Expressions) — **GE** (Gestures) — **BP** (Body Postures) — **GZ** (Gaze) — **AU** (Auditory Cues)

| | | MODALITIES |
|---|---|---|
| **A** | P with a kid's stroller is crossing a two-lanes road while D1-D2 are approaching. | HM, BP, GZ, FE |
| **B** | C changes lane / turns in front of a car. | HM, GZ, AU |
| C. | Inattentive group of P crossing the street, D approaches seeking attention and signalling | GE, GZ, BP |
| D. | P looks at the traffic light that turns red, signalling and taking to other P on the other side of the street and crossing inattentively, D approaches the crossing | SP, FE, GE, BP, GA |
| E. | P emerges between parked cars and enters a parked car. D approaches | HM, GZ, AU |
| F. | P on wheelchair approaches a zebra crossing, D and C approach from different sides and give priority to P | HM, FE, GE, GA |
| G. | D turns to the street and P who are walking on the street move to the side | AU, HM, BP |
| H. | P (or group of pedestrians) cross half way a double-way street, they do not check the second lane, D approaches | AU, BP |
| I. | Low traffic road, P on the side of the street negotiate crossing with D, while M and C are passing between stopped cars | GZ, AU, HM |
| J. | P exits a shop/parking slot and walks on the street, D approaches | HM, BP, GA |
| K. | P is close to a zebra crossing, talking on the phone or texting with no clear intention to cross, D approaches | BP, FE, SP |
| L. | C standing close to a bike, and get on the bike, with no clear intention to start driving | BP, HM |
| M. | P steps on the road because of an obstacle on the pavement, C avoids pedestrian and changes lane, while D approaches | HM, GE, BP, GA |
| N. | M overtakes a car, looking for occluded pedestrians, and gives priority to P who is crossing | GZ, HM |
| O. | Policemen regulates traffic, instruct D for the direction too follow | BP, AU, GE |

INTERACTIONAL ATTRIBUTES / FROM REAL-WORLD TO VR






Variation 1 -- Top view

_D1 approaches _P turns the head + D1 slows down joint attention _P looks straight ahead _P detects D2 _P b.posture to D2 monitoring attention

Real-world scene analysis

Variation 1 - Explicit Informal Interaction

D2 detects P D2 slows down P detects D2 & gestures mutual attention

Variation 2 - Implicit Informal Interaction

D2 detects P + slows down P turns head towards D2 common attention

# A COGNITIVE MODEL OF VISUOSPATIAL COMPLEXITY / ENVIRONMENTAL ATTRIBUTES

**Quantitative**

SIZE

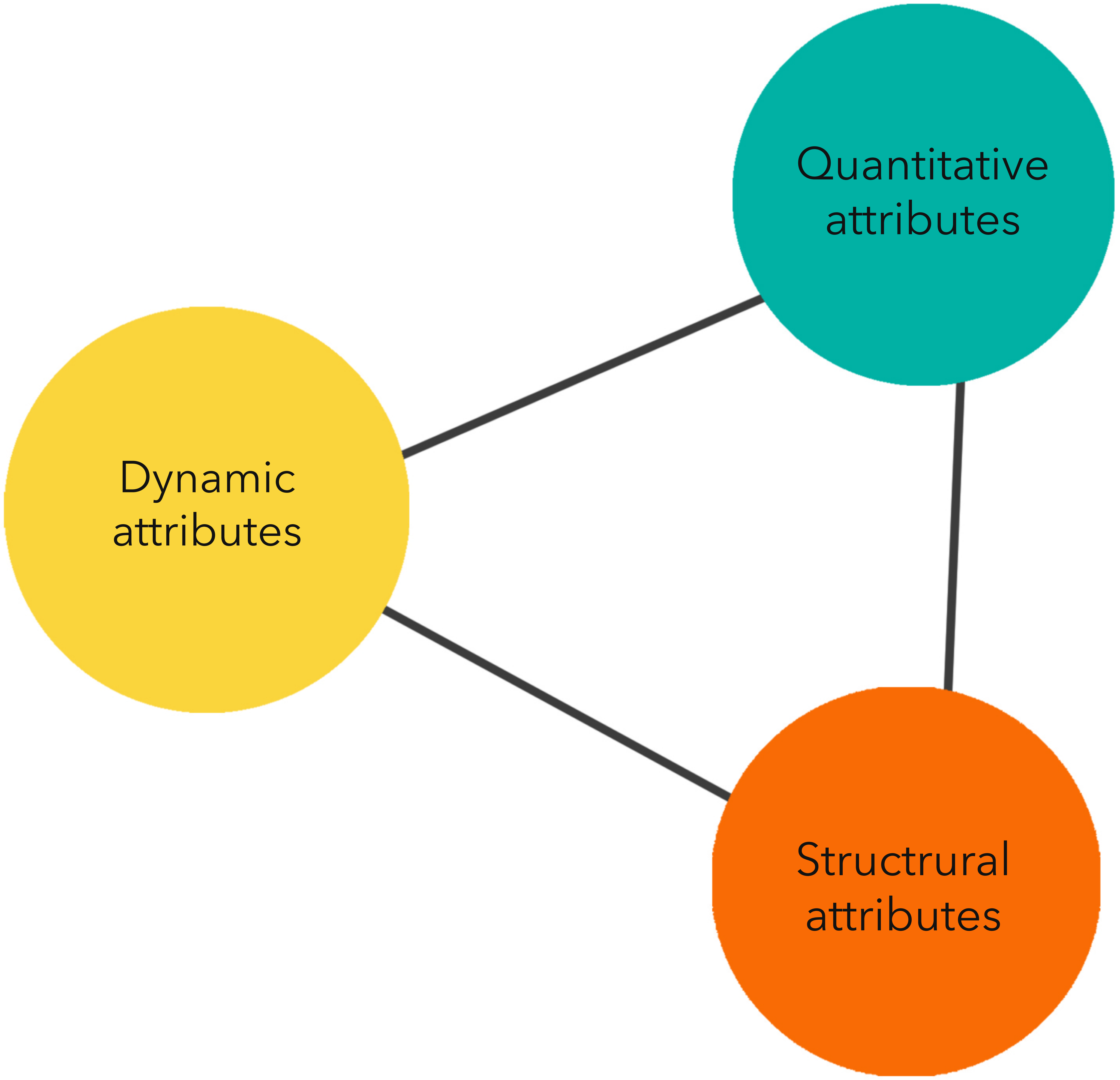
CLUTTER
- Quantity
- Variety of Colors
- Variety of Shapes/Objects
- Objects Density
- Edges Density
- Luminance
- Saliency
- Target-background similarity

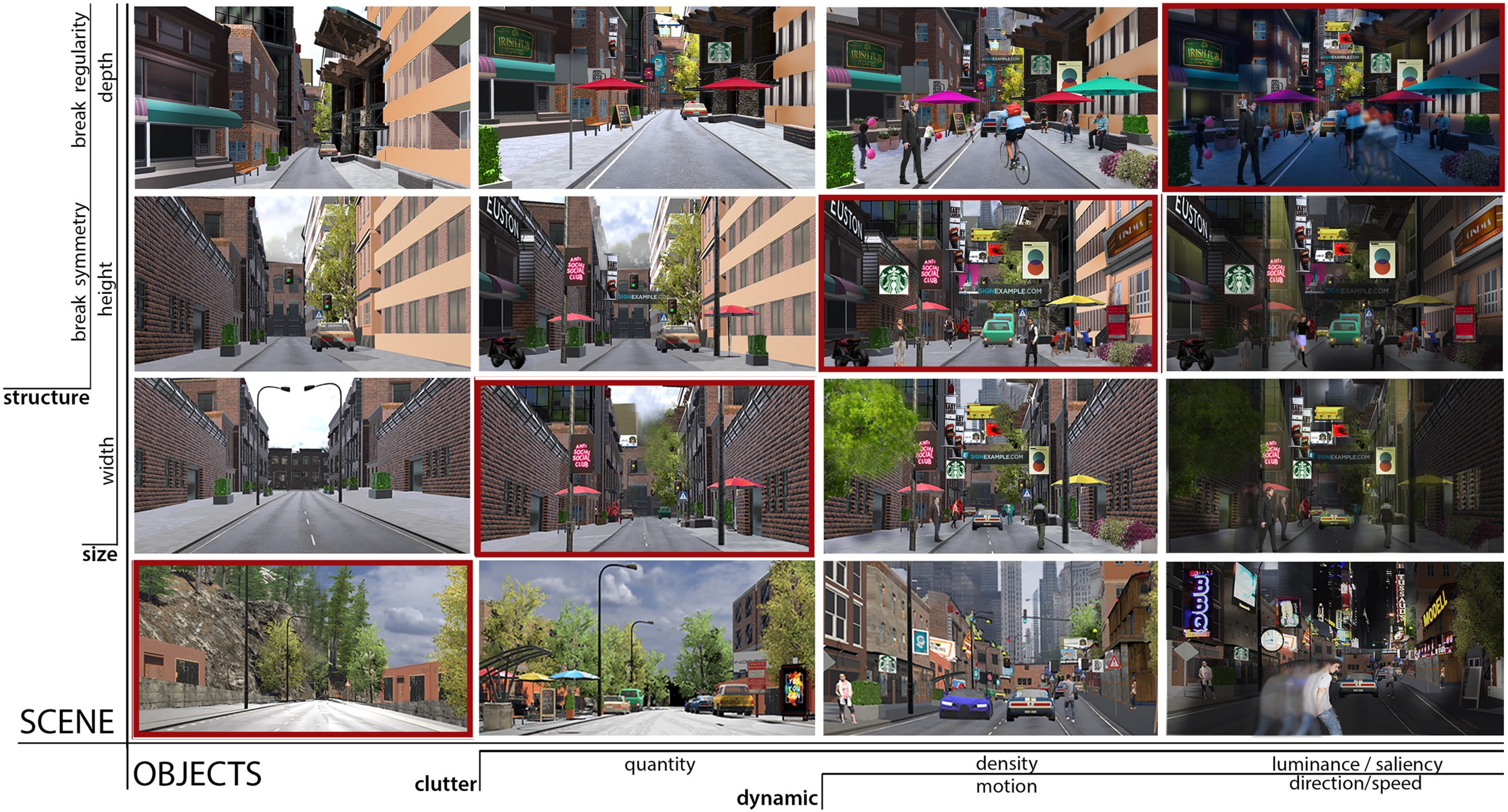
**Structural**

- Repetition
- Symmetry
- Order
- Homogeneity/Heterogeneity
- Regularity
- Openness
- Grouping

**Dynamic**

- Motion
- Flicker
- Speed / Direction

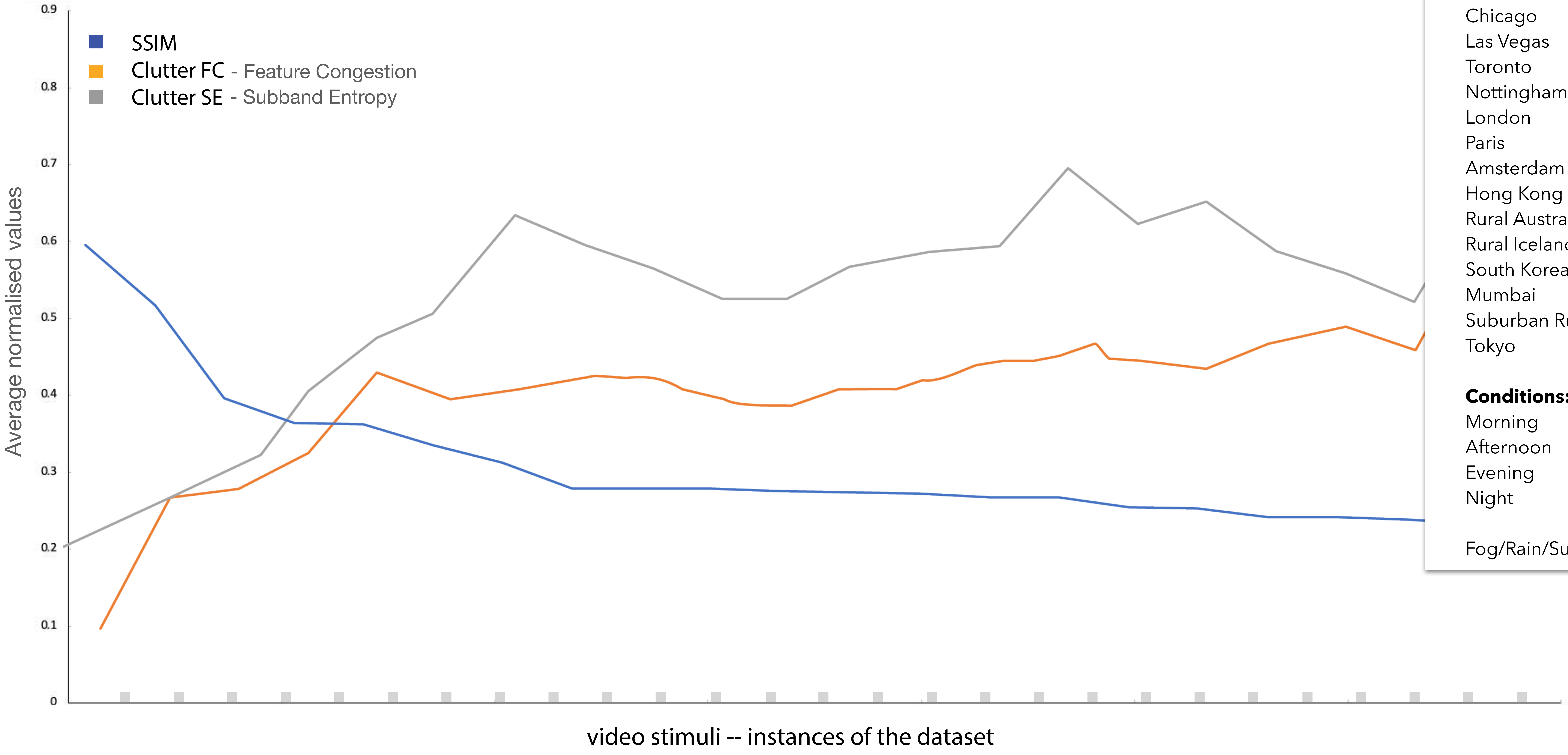



break regularity
depth
break symmetry
height
structure
width
size
SCENE
OBJECTS
clutter
quantity
density
luminance / saliency
dynamic
motion
direction/speed

# A COGNITIVE MODEL OF VISUOSPATIAL COMPLEXITY / EVALUATION


AXIS
BBQ
TUSSAUDS
MODELL
SPORTING GOODS
TRAITEUR GREC
AU BOUT DU CHAMP

# VISUOSPATIAL COMPLEXITY / VISUAL CLUTTER & STRUSTURAL SIMILARITY



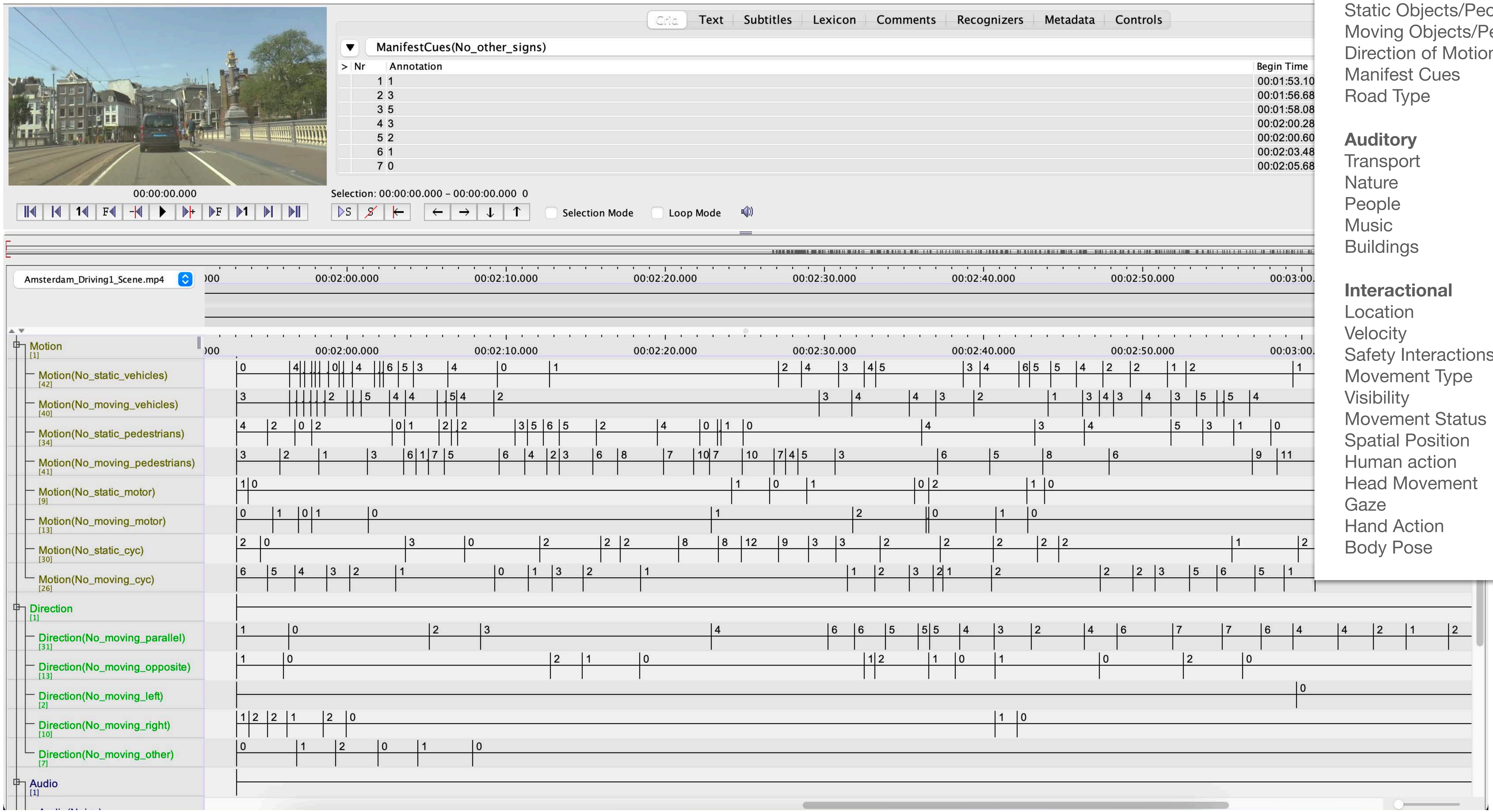

**27 VIDEOS**

**Locations:**
New York
Chicago
Las Vegas
Toronto
Nottingham
London
Paris
Amsterdam
Hong Kong
Rural Australia
Rural Iceland
South Korea
Mumbai
Suburban Russia
Tokyo

**Conditions:**
Morning
Afternoon
Evening
Night

Fog/Rain/Sun

# VISUOSPATIAL COMPLEXITY / SEMANTIC ANNOTATIONS

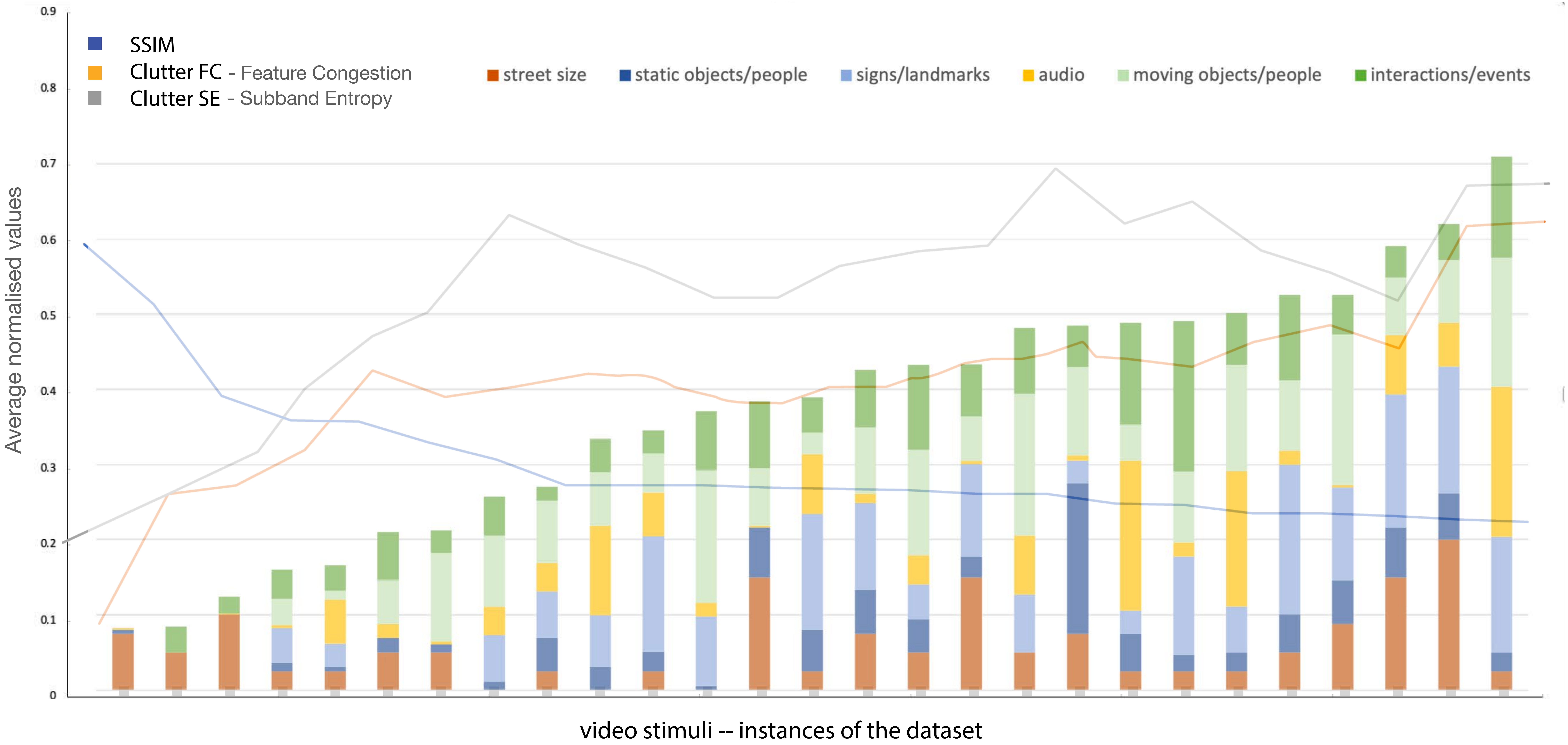



**Visuospatial**
Static Objects/People
Moving Objects/People
Direction of Motion
Manifest Cues
Road Type

**Auditory**
Transport
Nature
People
Music
Buildings

**Interactional**
Location
Velocity
Safety Interactions
Movement Type
Visibility
Movement Status
Spatial Position
Human action
Head Movement
Gaze
Hand Action
Body Pose

# VISUOSPATIAL COMPLEXITY / LOW & HIGH LEVEL ATTRIBUTES



# APPLICATIONS

## LARGE-SCALE VISUAL PERCEPTION STUDIES IN DRIVER (IN)ATTENTIVENESS

## Inattention Blindness

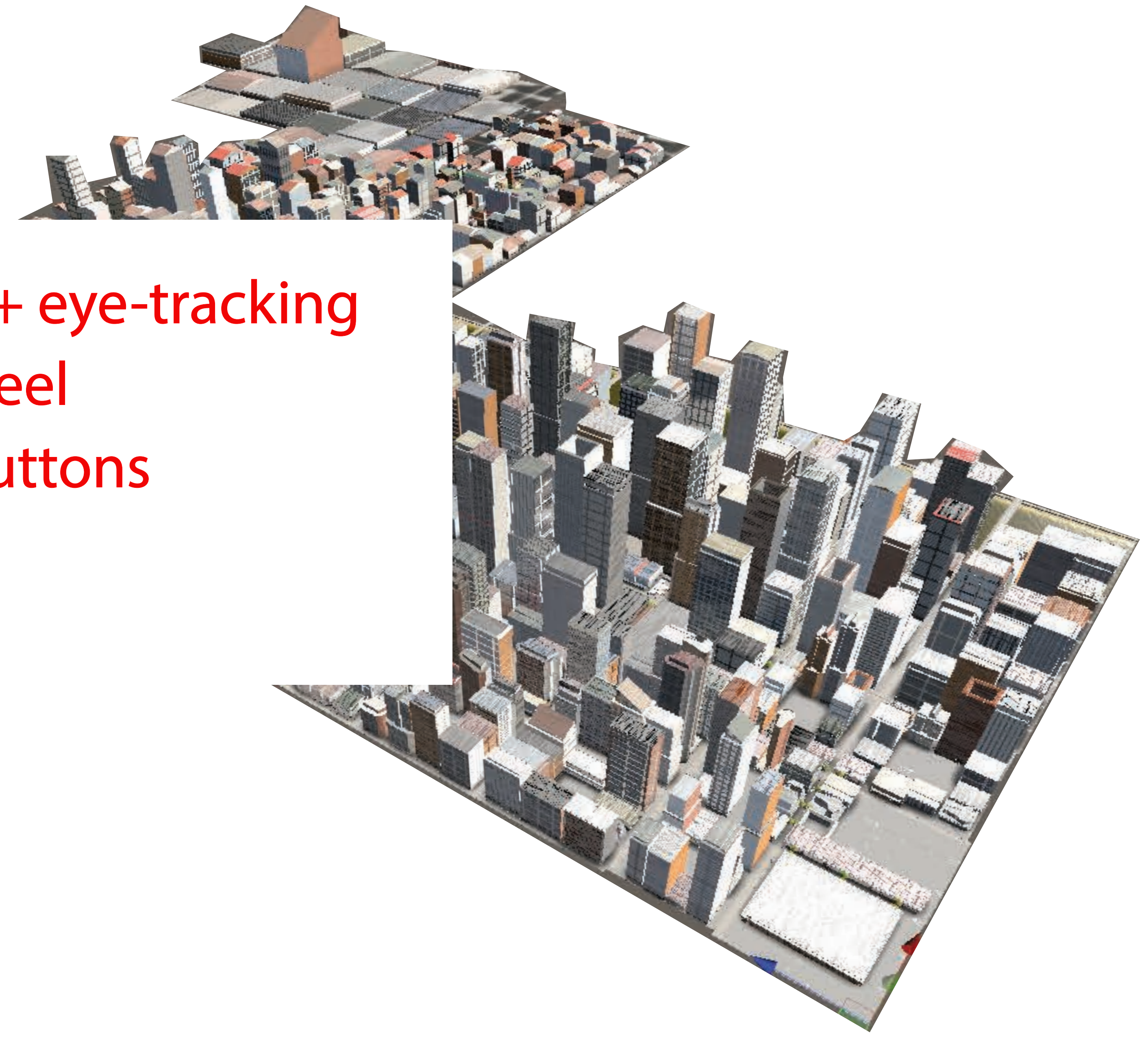
80 participants

1 hour task

36 interactions

12 levels of visuospatial complexity

52 building blocks

CO DESIGN LAB EU https://codesign-lab.org/

## Towards a Human-Centred Cognitive Model of Visuospatial Complexity in Everyday Driving

Vasiliki Kondyli Mehul Bhatt Jakob Suchan

Örebro University, Sweden
University of Bremen, Germany
CoDesign Lab > Cognitive Vision
www.codesign-lab.org/cognitive-vision

### Abstract

We develop a human-centred cognitive model of visuospatial complexity in everyday, naturalistic driving conditions. With a focus on visual perception, the model incorporates quantitative, structural, and dynamic attributes identifiable in the chosen context; the human-centred basis of the model lies in its behavioural evaluation with human subjects with respect to psychophysical measures pertaining to embodied visuoauditory attention. We report preliminary steps to apply the developed cognitive model of visuospatial complexity for human-factors guided dataset creation and benchmarking, and for its use as a semantic template for the (explainable) computational analysis of visuospatial complexity.

### 1 Introduction

Autonomous driving research has received enormous academic & industrial interest in recent years. As the self-driving vehicle industry develops, it will be necessary —in a manner similar to sectors such as medical computing, computer aided engineering and design— to have an articulation and community consensus on aspects such as representation, interoperability, data archival & retrieval mechanisms, and human-centred performance benchmarks. We posit that this will be necessary towards fulfilling essential legal and ethical responsibilities, such as those pertaining to representation and realisation of rules and norms, verifiable performance of normative behaviour, and technology capabilities vis-a-vis human-centred expectations.

**Human-Centred Benchmarking and Standardisation.** Within autonomous driving, the need for standardisation and ethical regulation has most recently garnered interest internationally, e.g., with the Federal Ministry of Transport and Digital Infrastructure in Germany (BMVI) taking a lead in eliciting 20 key propositions (with possible legal implications) for the fulfilment of ethical commitments for automated and connected driving systems [21]. In spite of major investments in self-driving vehicle research, issues related to human-centred design, human-machine interaction, and standardisation have been barely addressed, with the current focus in driving research primarily being on two basic considerations: *how fast to drive, and which way and how much to steer.* This is necessary, but inadequate if autonomous vehicles are to become commonplace and function with humans: not everything in autonomous vehicles is about realtime control/decision-making; several human-machine interaction requirements (e.g., for diagnostic communication, universal design) also exist. The 20 key propositions elicited by the German federal ministry BMVI highlight a range of factors pertaining to safety, utilitarian considerations, human rights, statutory liability, technological transparency, data management and privacy etc. Ethically driven standardisation and regulation will require addressing challenges in foundational human-centred AI technology,

In: STAIRS at ECAI 2020: European Conference on Artificial Intelligence 2020.

*DHM2020* 131
*L. Hanson et al. (Eds.)*

*doi:10.3233/ATDE200018*

## Multimodality on the Road: Towards Evidence-Based Cognitive Modelling of Everyday Roadside Human Interactions

Vasiliki KONDYLI, and Mehul BHATT
*Örebro University, Sweden*
*CoDesign Lab – Cognitive Vision*
www.codesign-lab.org/cognitive-vision

**Abstract** We propose an evidence based methodology for the systematic analysis and cognitive characterisation of multimodal interactions in naturalistic roadside situations such as driving, crossing a street etc. Founded on basic human modalities of embodied interaction, the proposed methodology utilises three key characteristics crucial to roadside situations, namely: explicit and implicit mode of interaction, formal and informal means of signalling, and levels of context-specific (visual) attention. Driven by the fine-grained interpretation and modelling of human behaviour in naturalistic settings, we present an application of the proposed model with examples from a work-in-progress dataset consisting of baseline multimodal interaction scenarios and variations built therefrom with a particular emphasis on joint attention and diversity of modalities employed. Our research aims to open up an interdisciplinary frontier for the human-centred design and evaluation of artificial cognitive technologies (e.g., autonomous vehicles, robotics) where embodied (multimodal) human interaction and normative compliance are of central significance.

**Keywords.** multimodal interaction, interpersonal communication, naturalistic perception, joint attention, virtual reality, autonomous driving

### 1. Introduction

Interpersonal communication and interactions are vital for safe and effective coordination of actions in everyday roadside engagements: walking around, driving, riding a bike etc. Failure in interpersonal communication leads to a lack of mutual understanding of a situation and it is responsible for a great number of roadside accidents [1, 2]. With further strides in the autonomous vehicles industry and the present impetus on high-level visual intelligence technologies [3, 4], it will therefore be necessary to account for the role of interpersonal communication on the street and articulate human-centred performance benchmarks, e.g., from the viewpoint of training, testing and validation as part of statutory compliance measures.



## Visuo-Locomotive Complexity as a Component of Parametric Systems for Architecture Design

Vasiliki Kondyli, Mehul Bhatt (Örebro University, Sweden)
Evgenia Spyridonos (University of Stuttgart, Germany)

**Abstract** A people-centred approach for designing large-scale built-up spaces necessitates systematic anticipation of user's embodied visuo-locomotive experience from the viewpoint of human-environment interaction factors pertaining to aspects such as navigation, wayfinding, usability. In this context, we develop a behaviour-based visuo-locomotive complexity model that functions as a key correlate of cognitive performance vis-a-vis internal navigation in built-up spaces. We also demonstrate the model's implementation and application as a parametric tool for the identification and manipulation of the architectural morphology along a navigation path as per the parameters of the proposed visuospatial complexity model. We present examples based on an empirical study in two healthcare buildings, and showcase the manner in which a dynamic and interactive parametric (complexity) model can promote behaviour-based decision-making throughout the design process to maintain desired levels of visuospatial complexity as part of a navigation or wayfinding experience.

**Keywords:** Visual Perception – Environmental Psychology – Architecture Design – Parametric Design – Cognitive Computational Modelling – Spatial Cognition – AI and Design

### Publication Note

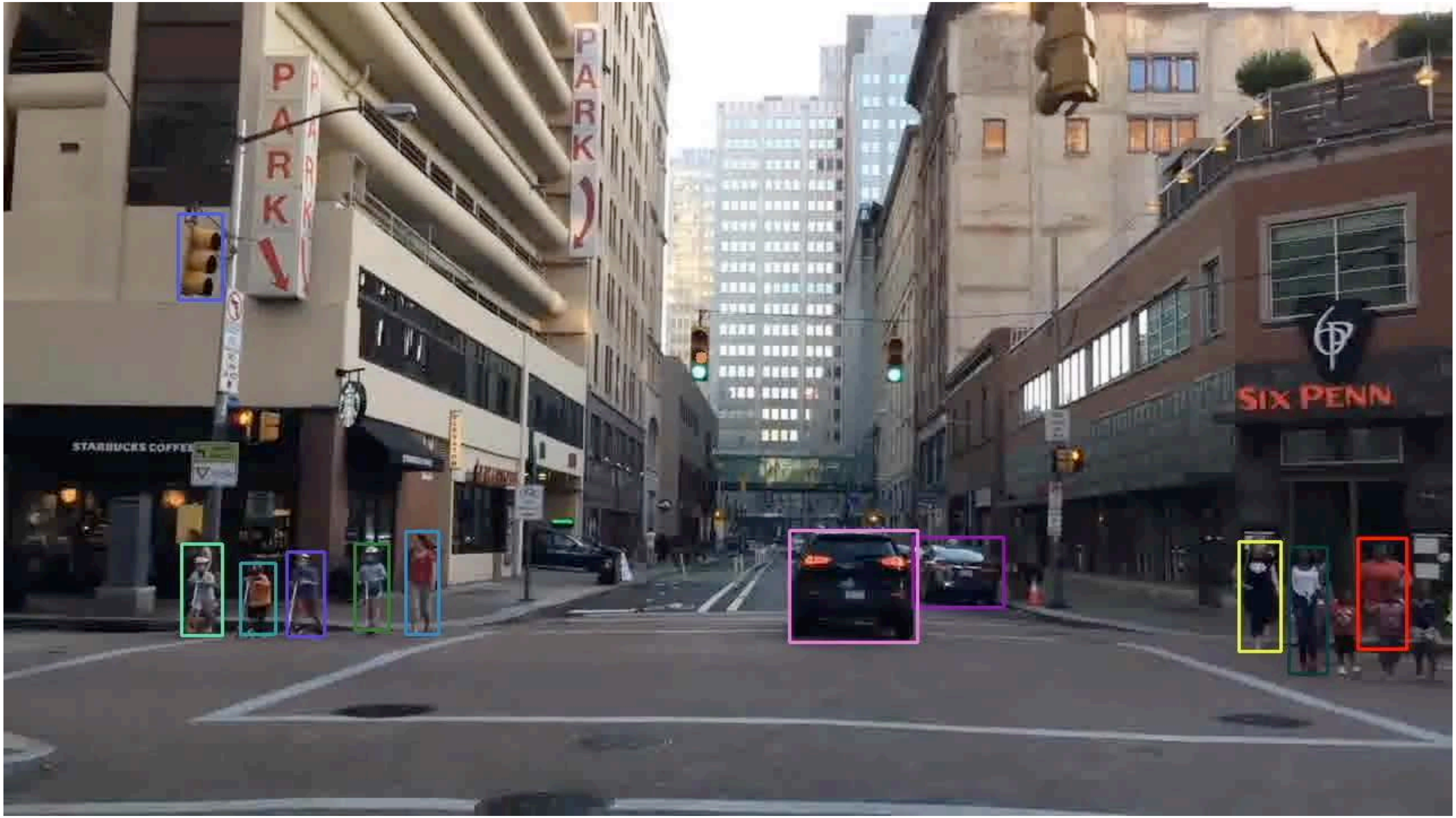
This is a preprint of the contribution published as part of the proceedings of ICoRD 2021: 8th International Conference on Research into Design, IDC School of Design (IIT Mumbai, India).
ICoRD 2021, www.idc.iitb.ac.in/icord2021/.

The overall scientific agenda driving this research may be consulted here:
The DesignSpace Group / www.design-space.org

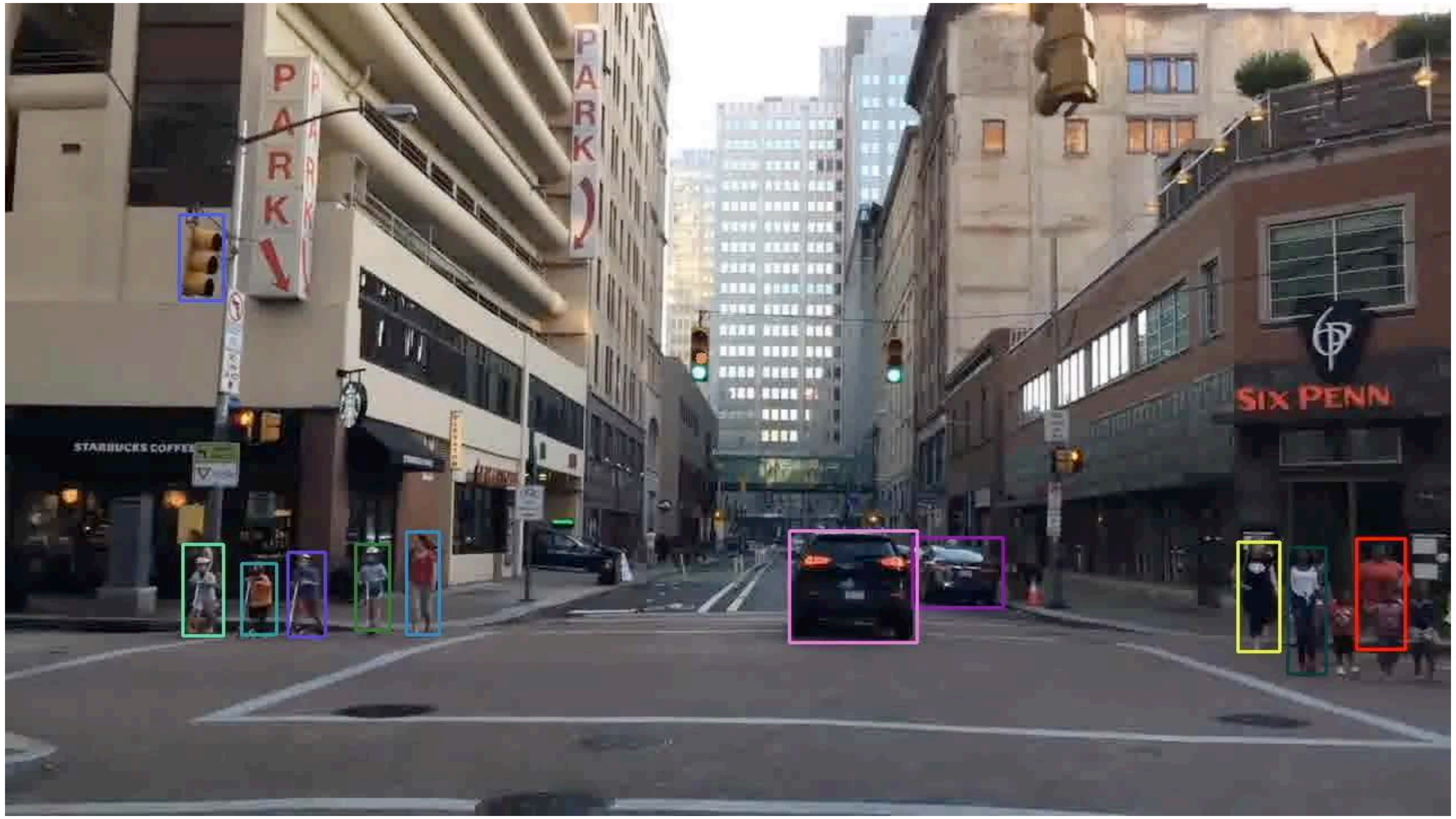
Select publications related to this research are available at:
www.codesign-lab.org/select-papers (under DesignSpace)

# APPLICATIONS

## HUMAN-CENTRED BENCHMARKING FOR AUTONOMOUS DRIVING SYSTEMS




J. Suchan & Bhatt, M. S. Varaadarajan (2021). MCommosense visual sense making for autonomous driving. On generated neurosymbilic online abduction integrating vision and semantics. Artificial Intelligence 299.




SYSTEMATIC ANALYSIS

COGNITIVE CHARACTERISATION & MODELLING

TRAINING & TESTING IN BENCHMARK DATASETS

EVALUATION SCHEMA

J. Suchan & Bhatt, M. S. Varaadarajan (2021). MCommosense visual sense making for autonomous driving. On generated neurosymbilic online abduction integrating vision and semantics. Artificial Intelligence 299.

Instance 1
Instance 10
Instance 25
Wheelchair
inattentive to
driver
pedestrian
gesture
to pedestrian
Quantitative
Structural
Dynamic
Eye-tracking/Head rotaion/Reaction time
Reation Time
Detection Rate
Fixation
Duration of Fixations
Instance 10
Average Velocity
Steering angle
Mode
Method
Social Attention
Modalities
explicit
implicit
formal
informal
Joint
Common
Monitoring
gaze
gestures
body/face expression
speech auditory
VISUOSPATIAL COMPLEXITY
HUMAN PERFORMANCE
EMBODIED MULTIMODAL INTERACTIONS

# Grounding Embodied Multimodal Interaction

Towards Behaviourally Established Semantic Foundations for Human-Centred AI

Vasiliki Kondyli[1], Jakob Suchan[2] and Mehul Bhatt[1]

[1]Örebro University, Sweden
[2]German Aerospace Center (DLR), Germany

*CoDesign Lab » Cognition. AI. Interaction. Design.*
*info@codesign-lab.org / https://codesign-lab.org*

### Abstract

We position recent and emerging research in *cognitive vision and perception* addressing three key questions: **(1)** What kind of relational abstraction mechanisms are needed to perform (explainable) grounded inference –e.g., question-answering, qualitative generalisation, hypothetical reasoning– relevant to embodied multimodal interaction? **(2)** How can such abstraction mechanisms be founded on behaviourally established cognitive human-factors emanating from naturalistic empirical observation? and **(3)** How to articulate behaviourally established abstraction mechanisms as formal declarative models suited for grounded knowledge representation and reasoning (KR) as part of large-scale hybrid AI and computational cognitive systems.

We contextualise (1–3) in the backdrop of recent results at the interface of AI/KR, and Spatial Cognition and Computation. Our main purpose is to emphasise the importance of behavioural research based foundations for next-generation, human-centred AI, e.g., as relevant to applications in Autonomous Vehicles, Social and Industrial Robots, and Visuo-Auditory Media.

### Keywords

Multimodal Interaction, Commonsense Reasoning, Declarative Spatial Reasoning, Declarative AI, Explainable AI, Cognitive Human-Factors, Cognitive Systems

## 1. Motivation

Multimodality in interaction is an inherent aspect of human activity, be it in social, professional, or everyday mundane contexts. Next-generation AI technologies, aiming for compliance with human-centred ethical and legal requirements, performance benchmarks, and inclusive usability expectations will require an inherent foundational capacity to analyse –e.g., understand, explain, anticipate– everyday interactional multimodality in naturalistic settings involving technology mediated collaborative assistance of humans. Amongst other things, this necessitates that the foundational building blocks of such next-generation systems be semantically aligned with the descriptive complexity of human task conceptualisation and performance expectations.

**Declaratively Mediated Multimodality.** The significance of "*grounding*" in semiotic construction, e.g., enabling high-level meaning-making, has been long-established in Artificial

*KR4HI 2022: Knowledge Representation for Hybrid Intelligence, June 14, 2022, Amsterdamm, The Netherlands*
vasiliki.kondyli@oru.se (V. Kondyli); jakob.suchan@dlr.de (J. Suchan); mehul.bhatt@oru.se (M. Bhatt)

CEUR Workshop Proceedings (CEUR-WS.org)

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

ELSEVIER

# Commonsense visual sensemaking for autonomous driving – On generalised neurosymbolic online abduction integrating vision and semantics ☆

Jakob Suchan [a,c], Mehul Bhatt [b,c], Srikrishna Varadarajan [c]

[a] University of Bremen, Germany
[b] Örebro University, Sweden
[c] CoDesign Lab (EU), Cognitive Vision, Sweden

A R T I C L E  I N F O

*Article history:*
Received 19 July 2020
Received in revised form 29 April 2021
Accepted 30 April 2021
Available online 7 May 2021

*Keywords:*
Cognitive vision
Deep semantics
Declarative spatial reasoning
Knowledge representation and reasoning
Commonsense reasoning
Visual abduction
Answer set programming
Autonomous driving
Human-centred computing and design
Standardisation in driving technology
Spatial cognition and AI

A B S T R A C T

We demonstrate the need and potential of systematically integrated *vision* and *semantics* solutions for visual sensemaking in the backdrop of autonomous driving. A general neurosymbolic method for *online* visual sensemaking using answer set programming (ASP) is systematically formalised and fully implemented. The method integrates state of the art in visual computing, and is developed as a modular framework that is generally usable within hybrid architectures for realtime perception and control. We evaluate and demonstrate with community established benchmarks KITTIMOD, MOT-2017, and MOT-2020. As use-case, we focus on the significance of human-centred visual sensemaking –e.g., involving semantic representation and explainability, question-answering, commonsense interpolation– in safety-critical autonomous driving situations. The developed neurosymbolic framework is domain-independent, with the case of autonomous driving designed to serve as an exemplar for online visual sensemaking in diverse cognitive interaction settings in the backdrop of select human-centred AI technology design considerations.


## 1. Motivation

Autonomous driving research has received enormous academic & industrial interest in recent years (Sec 5). This surge has coincided with (and been driven by) advances in *deep learning* based computer vision research. Although end-to-end deep learning based vision and control has (arguably) been successful for self-driving vehicles, we posit that there is a clear need and tremendous potential for hybrid visual sensemaking solutions that integrate *vision and semantics* towards fulfilling essential legal and ethical responsibilities involving explainability, human-centred AI (Artificial Intelligence), and industrial standardisation (e.g, pertaining to representation, realisation of rules and norms, fulfilling statutory obligations).

☆ This paper is an invited contribution based on a distinguished paper nomination at the 2019 International Joint Conference on Artificial Intelligence (IJCAI-19).

*E-mail address:* info@codesign-lab.org (M. Bhatt).
*URL:* http://www.codesign-lab.org (M. Bhatt).

https://doi.org/10.1016/j.artint.2021.103522
0004-3702/

# CONCLUSION | COGNITIVE-BASED VISUOSPATIAL COMPLEXITY MODEL

- Common framework for embodied naturalistic behavioural studies on driver (in)attention
- Analysis and interpretation of human interactions /visuospatial complexity
  - human-centred computational models
- Human factors not sufficiently explored in autonomous driving research
- Provide guideline to create human-centred benchmarking
- Evaluate datasets based on human-centred factors

# THANK YOU

[ https://www.codesign-lab.org/cognitive-vision/ ]

[ https://codesign-lab.org/select-papers/ ]