Cognitive Interpretation of Everyday Activities — Toward Perceptual Narrative Based Visuo-Spatial Scene Interpretation

Mehul Bhatt[1,2], Jakob Suchan[1,2], and Carl Schultz[1,2]

1 Cognitive Systems
University of Bremen
Bremen, Germany
{bhatt,jsuchan,cschultz}@informatik.uni-bremen.de

2 Sonderforschungsbereich Transregional Collaborative Research Center 8:
Spatial Cognition
University of Bremen
Bremen, Germany

Abstract

We position a narrative-centred computational model for high-level knowledge representation and reasoning in the context of a range of assistive technologies concerned with *visuo-spatial perception and cognition* tasks. Our proposed narrative model encompasses aspects such as *space, events, actions, change, and interaction* from the viewpoint of commonsense reasoning and learning in large-scale cognitive systems. The broad focus of this paper is on the domain of *human-activity interpretation* in smart environments, ambient intelligence etc. In the backdrop of a *smart meeting cinematography* domain, we position the proposed narrative model, preliminary work on perceptual narrativisation, and the immediate outlook on constructing general-purpose open-source tools for perceptual narrativisation.

1998 ACM Subject Classification I.2.0 Artificial Intelligence: General – Cognitive Simulation, I.2.4 Knowledge Representation Formalisms and Methods, I.2.10 Vision and Scene Understanding: Architecture and control structures, Motion, Perceptual reasoning, Shape, Video analysis

Keywords and phrases cognitive systems; human-computer interaction; spatial cognition and computation; commonsense reasoning; spatial and temporal reasoning; assistive technologies

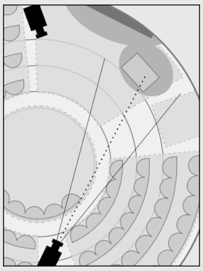
Digital Object Identifier 10.4230/OASIcs.CMN.2013.24

1 Introduction: Cognitive Interpretation by Narrativisation

Narratives have been a focus on study from several perspectives, most prominently from the viewpoint of language, literature, and computational linguistics; see for instance, discourse analysis and computational narratology [1, 13, 14, 12]. From the viewpoint of commonsense reasoning, and closely related to the computational models of narrative perspective, is the position of researchers in logics of *action and change*; here, narratives are interpreted as "*a sequence of events about which we may have incomplete, conflicting or incorrect information*" [16, 18]. As per McCarthy [15], "*a narrative tells what happened, but any narrative can only tell a certain amount. A narrative will usually give facts about the future of a situation that are not just consequences of projection from an initial situation*". The interpretation of narrative knowledge in this paper is based on these characterisations, especially in regard to the commonsense representation and reasoning tasks that accrue whilst modelling and reasoning about the perceptually grounded, narrativised epistemic state of an autonomous

Workshop on Computational Models of Narrative 2013.
Editors: Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister; pp. 24–29
OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Listing 1. Smart Meeting Cinematography



The smart meeting cinematography domain focusses on professional situations such as meetings and seminars. A basic task is to automatically produce dynamic recordings of interactive discussions, debates, presentations involving interacting people who use more than one communication modality such as hand-gestures (e.g., raising one's hand for a question, applause), voice and interruption, electronic apparatus (e.g., pressing of a button), movement (e.g., standing-up) and so forth. The scenario consists of people-tracking, gesture identification closed under a context-specific taxonomy, and also involves real-time dynamic collaborative co-ordination and self-control of pan-tilt-zoom (PTZ) cameras in a *sensing-planning-acting* loop. The long-term vision is to benchmark with respect to the capabilities of human-cinematographers, real-time video editors, surveillance personnel to record and semantically annotate individual and group activity (e.g., for summarisation, story-book format digital media and promo generation). [5]

agent pertaining to *space, actions, events, and change* [2]. In particular, this encompasses a range of inference patterns such as: (a) spatio-temporal abduction for scenario and narrative completion [3]; (b) integrated inductive-abductive reasoning with narrative knowledge [7]; (c) narrative-based postdiction for abnormality detection and planning [8].

Perceptual Narratives. These are declarative models of visual, auditory, haptic and other observations in the real world that are obtained via artificial sensors and / or human input. Declarative models of perceptual narratives can be used for interpretation and control tasks in the course of assistive technologies in everyday life and work scenarios, e.g., behaviour interpretation, robotic plan generation, semantic model generation from video, ambient intelligence and smart environments (e.g., see narrative based models in [10, 11, 17, 3, 7, 8]).

High-Level Cognitive Interpretation and Control. Our research is especially concerned with large-scale cognitive interaction systems where high-level perceptual sense-making, planning, and control constitutes one of many AI sub-components guiding other low-level control and attention tasks. As an example, consider the *smart meeting cinematography* domain in Listing 1. In this domain, *perceptual narratives* as in figure 1 are generated based on perceived spatial change interpreted as interactions of humans in the environment. Such narratives explaining the ongoing activities are needed to anticipate changes in the environment, as well as to appropriately influence the real-time control of the camera system. To convey the meaning of the presentation and the speakers interactions with a projection, the camera has to capture the scene including the speakers gestures, slides, and the audience. E.g., in figure 1, when the speaker explains the slides, the camera has to capture the speaker and the corresponding information on the slides. To this end, the camera records an overview shot capturing the speaker and the projection, and zooms on the particular element when the speaker explains it in detail, to allow the viewer to follow the presentation and to get the necessary information. When the speaker continues the talk, the camera focuses on the speaker to omit unnecessary and distracting information. To capture reactions of the audience, e.g., comments, questions or applause, the camera records an overview of the attending people or close-up shots of the commenting or asking person.

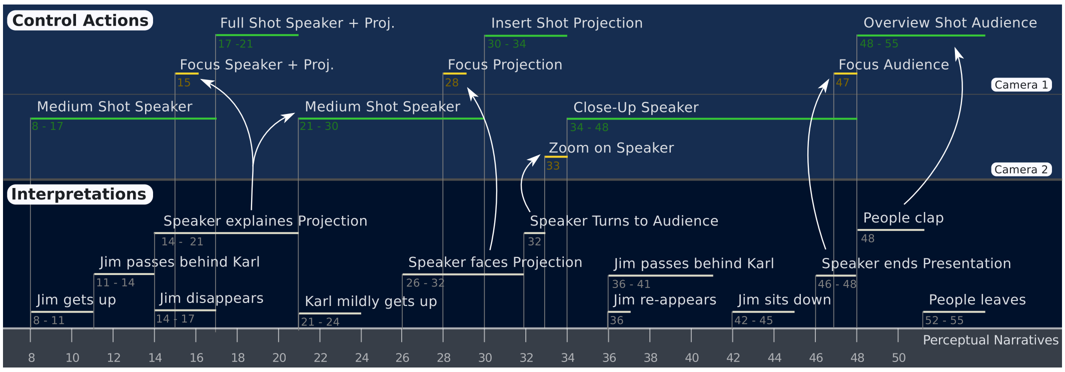


Figure 1 Cognitive Interpretation and Control by Perceptual Narrativisation.

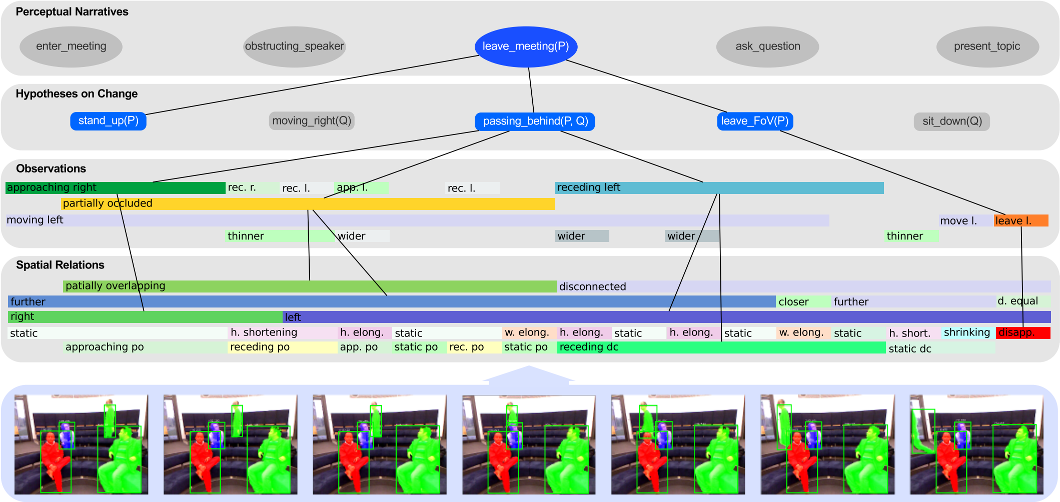


Figure 2 Perceptual Narratives of Depth, Space, and Motion.

2 Perceptual Narrative Generation for Activity Interpretation

Systems that monitor and interact with an environment populated by humans and other artefacts require a formal means for representing and reasoning about spatio-temporal, event and action based phenomena that are grounded to real public and private scenarios (e.g., logistical processes, activities of everyday living) of the environment being modelled. A fundamental requirement within such application domains is the representation of *dynamic* knowledge pertaining to the spatial aspects of the environment within which an agent, system, or robot is functional. This translates to the need to explicitly represent and reason about dynamic spatial configurations or scenes and, for real world problems, integrated reasoning about perceptual narratives of *space, actions, and change* [2]. With these modelling primitives, the ability to perform *predictive* and *explanatory* analyses on the basis of sensory data is crucial for creating a useful intelligent function within such environments [7].

Perceptual Narratives of Space and Motion

To understand the nature of perceptual narratives (of space, and motion), consider the aforediscussed work-in-progress domain of *smart meeting cinematography* (Listing 1). The particular infrastructural setup for the example presented herein consists of Pan-Tilt-Zoom (PTZ) capable cameras, depth sensors (Kinect), and a low-level vision module for people tracking (whole body, hand gesture, movement) customised on the basis of open-source algorithms and software. With respect to this setup, declaratively grounded perceptual narratives capturing the information in figure 1 is developed on the basis of a commonsense theory of *qualitative space* (Listing 2), and interpretation of motion as qualitative spatial change [9]. In particular, the overall model as depicted in figure 2 consists of:

Space and Motion: A theory to declaratively reason about qualitative spatial relations (e.g., topology, orientation), and qualitative motion perceived in the environment and interpret changes as domain dependent observations in the context of everyday activities involving humans and artefacts.

Explanation of (Spatial) Change: Hypothesising real-world (inter)actions of individuals explaining the observations by integrating the qualitative theory with a learning method (e.g., Bayesian and Markov based (logic) learning) to incorporate uncertainty in the interpretation of observation sequences.

Semantic characterisation: as a result of the aforementioned, real-time generation of declarative narratives of perceptual data (e.g., RGB-D) obtained directly from people/object tracking algorithms.

Hypothesised object relations can be seen as building blocks to form complex interactions that are semantically interpreted as activities in the context of the domain. As an example consider the sequence of observations in the meeting environment depicted in figure 2.

> Region P elongates vertically, region P approaches region Q from the right, region P partially overlaps with region Q while P being further away from the observer than Q, region P moves left, region P recedes from region Q at the left, region P gets disconnected from region Q, region P disappears at the left border of the field of view

To explain these observations in the ‘context’ of the meeting situation we make hypothesis about possible interactions in the real world.

> Person P stands up, passes behind person Q while moving towards the exit and leaves the room.

Listing 2. Qualitative Abstractions of Space and Motion

To represent space and spatial change we consider spatio-temporal relations [6] holding between individuals in the environment, i.e., *topology, orientation, size, movement.* Combinations of spatial and temporal relations serve as observations describing perceived phenomena in the real world. The theory is implemented using **CLP(QS)** [4], which is a *declarative spatial reasoning framework* that can be used for representing and reasoning about high-level, qualitative spatial knowledge about the world. CLP(QS) implements the semantics of qualitative spatial calculi within a constraint logic programming framework (amongst other things, this makes it possible to use spatial entities and relations between them as native entities). Furthermore it provides a declarative interface to qualitative and geometric spatial representation and reasoning capabilities such that these may be integrated with general knowledge representation and reasoning (KR) frameworks in artificial intelligence.

The **semantic interpretation of activities** from video, depth (e.g., time-of-flight devices such as Kinect), and other forms of sensory input requires the representational and inferential mediation of qualitative abstractions of space, action, and change. Such relational abstractions serve as a bridge between high-level domain-specific conceptual or activity theoretic knowledge, and low-level statistically acquired features and sensory grammar learning techniques. Generation of perceptual narratives, and their access via the declarative interface of logic programming facilitates the integration of the overall framework in bigger projects concerned with cognitive vision, robotics, hybrid-intelligent systems etc. In the smart meeting cinematography domain the generated narratives are used to explain and understand the observations in the environment and anticipate interactions in it to allow for intelligent coordination and control of the involved PTZ-cameras.

3 Immediate Outlook

The smart meeting cinematography scenario presented in this paper serves as a challenging benchmark to investigate narrative based high-level cognitive interpretation of everyday interactions. Work is in progress to release certain aspects (pertaining to space, motion, real-time high-level control) emanating from the narrative model via the interface of constraint logic programming (e.g., as a Prolog based library of depth–space–motion). We also plan to release general tools to perform management and visualisation of activity interpretation data.

References

1 Roland Barthes and Lionel Duisit. An Introduction to the Structural Analysis of Narrative. *New Literary History*, 6(2):237–272, 1975.

2 Mehul Bhatt. Reasoning about space, actions and change: A paradigm for applications of spatial reasoning. In S. Hazarika, editor, *Qualitative Spatial Representation and Reasoning: Trends and Future Directions*, pages 284–320. IGI Global, USA, 2012.

3 Mehul Bhatt and Gregory Flanagan. Spatio-Temporal Abduction for Scenario and Narrative Completion: a preliminary statement). In Mehul Bhatt, Hans Guesgen, and Shyamanta Hazarika, editors, *ECAI 2010, Workshop Proceedings, Spatio-Temporal Dynamics*, pages 31–36, 2010.

4 Mehul Bhatt, Jae Hee Lee, and Carl Schultz. CLP(QS): A Declarative Spatial Reasoning Framework. In Max J. Egenhofer, Nicholas A. Giudice, Reinhard Moratz, and Michael F. Worboys, editors, *Spatial Information Theory, 10th International Conference, COSIT 2011,*

Belfast, ME, USA, September 12-16, 2011. Proceedings, number 6899 in Lecture Notes in Computer Science, pages 210–230. Springer, 2011.

5 Mehul Bhatt, Jakob Suchan, and Christian Freksa. ROTUNDE – A Smart Meeting Cinematography Initiative. In M. Bhatt, H. Guesgen, and D. Cook, editors, *Proceedings of the AAAI-2013 Workshop on Space, Time, and Ambient Intelligence (STAMI).*, Washington, US, 2013. AAAI Press. (to appear).

6 Anthony G. Cohn and Jochen Renz. Qualitative spatial representation and reasoning. In Frank van Harmelen, Vladimir Lifschitz, and Bruce Porter, editors, *Handbook of Knowledge Representation*, pages 551–596. Elsevier, 2007.

7 Krishna Dubba, Mehul Bhatt, Frank Dylla, Anthony Cohn, and David Hogg. Interleaved Inductive-Abductive Reasoning for Learning Event-Based Activity Models. In Stephen Muggleton, Alireza Tamaddoni-Nezhad, and Francesca A. Lisi, editors, *Inductive Logic Programming, 21st International Conference, ILP 2011, Windsor Great Park, UK, July 31 – August 3, 2011, Revised Selected Papers*, number 7207 in Lecture Notes in Computer Science, pages 113–129, 2011.

8 Manfred Eppe and Mehul Bhatt. Narrative based postdictive reasoning for cognitive robotics. In *Proceedings of the 11th International Symposium on Logical Formalizations of Commonsense Reasoning*, 2013.

9 Antony Galton. *Qualitative Spatial Change*. Oxford University Press, Oxford, 2000.

10 Hannaneh Hajishirzi, Julia Hockenmaier, Erik T. Mueller, and Eyal Amir. Reasoning about robocup soccer narratives. *CoRR*, abs/1202.3728, 2012.

11 Hannaneh Hajishirzi and Erik T. Mueller. Symbolic probabilistic reasoning for narratives. In Ernest Davis, Patrick Doherty, and Esra Erdem, editors, *Logical Formalizations of Commonsense Reasoning, Papers from the 2011 AAAI Spring Symposium, Stanford, California, USA, March 21-23, 2011*, number SS-11-06 in AAAI Technical Reports, pages 123–126, 2011.

12 George Lakoff and Srini Narayanan. Toward a computational model of narrative. In Mark Alan Finlayson, editor, *Computational Models of Narrative: Papers from the 2010 AAAI Fall Symposium*, number FS-10-04 in AAAI Technical Reports, pages 21–28. AAAI Press, 2010.

13 Inderjeet Mani. *Computational Modeling of Narrative*. Number 5 in Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.

14 Inderjeet Mani. Computational Narratology. In Peter Hühn, Jan Christoph Meister, John Pier, and Wolf Schmid, editors, *The Living Handbook of Narratology*. Hamburg University Press, 2013.

15 John McCarthy. Concept of logical AI. In Jack Minker, editor, *Logic-based artificial intelligence*, pages 37–56, Norwell, MA, USA, 2000. Kluwer Academic Publishers.

16 Rob Miller and Murray Shanahan. Narratives in the Situation Calculus. *Journal of Logic and Computation*, 4(5):513–530, 1994.

17 Erik T. Mueller. Modelling space and time in narratives about restaurants. *Literary and Linguistic Computing*, 22(1):67–84, 2007.

18 Javier Pinto. Occurrences and narratives as constraints in the branching structure of the Situation Calculus. *Journal of Logic and Computation*, 8(6):777–808, 1998.