





# ‘Diagrams’: A Hybrid Visual Information Representation and Reasoning Paradigm Towards Video Analysis

Chayanika Deka Nath<sup>1</sup>  and Shyamanta M. Hazarika<sup>2</sup> 

<sup>1</sup> Biomimetic and Cognitive Robotics Lab, Computer Science and Engineering,  
Tezpur University, Tezpur 784028, India

[cnath.nath@gmail.com](mailto:cnath.nath@gmail.com), [cnath@tezu.ernet.in](mailto:cnath@tezu.ernet.in)

<sup>2</sup> Mechanical Engineering, Indian Institute of Technology, Guwahati 781039, India  
[s.m.hazarika@iitg.ernet.in](mailto:s.m.hazarika@iitg.ernet.in)

**Abstract.** This paper presents a comprehensive representation for video analysis combining qualitative reasoning with diagrammatic reasoning. The hybrid approach is motivated by the power of diagrams that allows explicit relational representation of entities involved. Perception of qualitative information over the underlying representation, employment of inter-diagrammatic reasoning approach and their combined relevance for temporal abstractions holds key to the analysis. Activity recognition over selected videos from J-HMDB dataset are performed and encouraging results are achieved.

**Keywords:** Qualitative spatial and temporal reasoning  
Diagrammatic representation and reasoning  
Inter-diagrammatic reasoning

## 1 Introduction

Humans and animals perceive spatio-temporal information about spatial entities directly through vision or basic senses. Such information many a times is incomplete and imprecise. An intelligent brain manipulate these with cognitive faculties and experiences. Therefore, an intelligent vision system needs comprehensive representation with cognitive knowledge processing abilities. Qualitative spatial and temporal reasoning (QSTR) [7] is an established area boosting qualitative information abstractions over spatial substrate for everyday reasoning. QSTR can be further enriched with explicit representation power of diagrams or mental images to deduce conclusions within unique observed relations. Diagrammatic representation and reasoning (DR) [9] supports diagram based representation of a situation with manipulation ability to perceive new information.

This paper presents a hybrid approach- for powerful and expressive relational representation of video data through diagrams and reasoning via manipulation of qualitative spatio-temporal relations among tracked video objects. QSTR onto

DR techniques are used as cognitive elements for space-time knowledge acquisition. For proof of concept, the proposed technique is evaluated on videos of J-HMDB dataset [12] for activity recognition and inspiring results are obtained.

## 2 Related Work and Motivation

*Qualitative Spatial and Temporal Reasoning:* QSTR for spatial knowledge processing is a thriving area of research. Many formalism have been developed and established: Region Connection Calculus (RCC) by Randell, Cardinal Direction Calculus (CDC) by Frank, Rectangle Algebra (RA) by Balbiani and Allen’s interval logic (IA) by Allen [1] to deal with variety of spatio-temporal circumstances. A comprehensive representation, CORE9 have been forwarded by Cohn et al. for video analysis. Ah Lian Kor proposed an improvised hybrid cardinal direction model using RCC and CDC.

*Diagrammatic Reasoning:* Cognitive knowledge processing through analogical representation in terms of mental maps, diagrammatic representations and mental images has emerged as a promising area. REDRAW [18] a qualitative structure analyzer is based on diagram manipulation with logical reasoning. Anderson [2] defined inter-diagrammatic reasoning (IDR) for spatio-temporal abstractions over defined ‘diagram’ based representation. Narayanan, proposed abstraction of motion based relations over spatial structure through predefined knowledge.

*Motivation:* Power of heterogeneous framework, combining diagrams and formal logic has influenced multidisciplinary research findings; mathematical theorem prover [19], spatial problem solver [3]. A combined diagrammatic and sentential representation is suggested by Gottfried in [10, 11] to enrich QSTR for results within confined relational subset. Freska [8] introduced need of comparison between formal and DR processes for same underlying problems. [13] established conceptual knowledge as a common language to generalize formal and diagrammatic approaches. Motivated by these facts, the authors aim to unify conceptual and formal problem solving techniques for video data analysis. In recent work QSTR over DR techniques are employed for motion event detection and activity recognition [14, 15]. This paper focus to elaborate QSTR onto DR techniques for a comprehensive representation and reasoning for video data analysis supported by activity recognition. Interval relations [1] over DR are exploited to provide a common visual background for spatial and temporal knowledge acquisition.

## 3 ‘Diagrams’: The Proposed Methodology

The paper forwards a hybrid representation and reasoning paradigm of video data through defined diagrams. *Diagrams* are image matrix representation of video frames in a 2-D frame of reference (x-y axes coinciding camera axes) with tracked objects, their properties and relations. QSTR and DR techniques are

implemented on diagrams through methodologies involving diagram creation, perceptual and diagram modification for spatio-temporal knowledge acquisition. Figure 1 shows a conceptual architecture of the proposed approach. The ‘diagram’ components and defined methodologies are illustrated in this section.

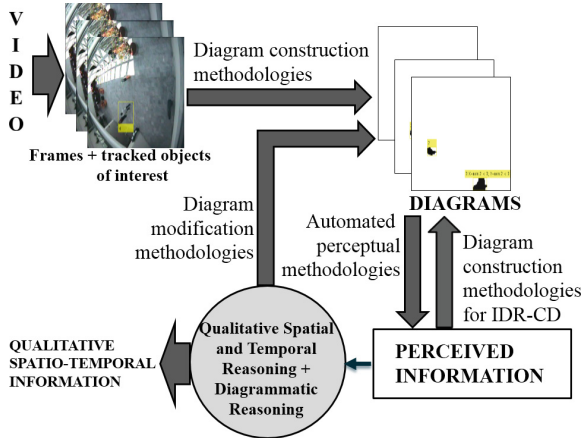
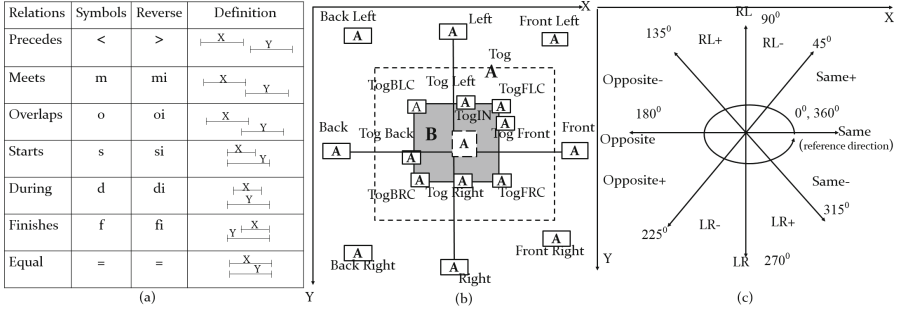


Fig. 1. Conceptual architecture of proposed methodology.

*Diagrammatic Objects:* Diagrammatic objects are considered to be of three types. Tracked video objects of interest are represented as ‘closed polygons’, distance among tracked object pairs as ‘lines’ and their direction of displacement during motion are represented as ‘rays’. Object’s properties such as, minimum and maximum extends of polygonal object along axes and length of line objects are maintained during their origination.

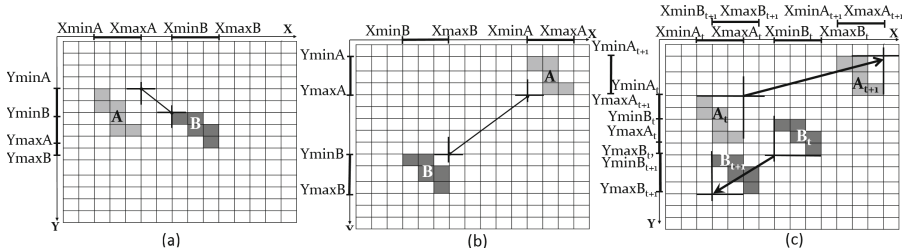
*Relations:* Along with object properties, diagrams maintain certain qualitative spatial and temporal relational information among polygon-polygon and ray-ray object pairs. These include: 13 basic Allen’s interval relations [1] along axes, 18 basic relative position relations, 3 basic relative distance relations  $\{-, +, 0\}$ , 8 basic displacement direction relations and relative displacement direction relations derived from  $QD_8$  [4] among polygon pairs. Figure 2 shows the AI relations, relative positions and  $QD_8$  relations.

*Diagram Construction Methodology:* Methodologies are defined for automatic construction of diagrams corresponding to video frames with tracked objects of interest. The procedure assigns each cell of such a diagram with specific gray intensity which are exploited by IDR-OR operator for creating *IDR-combined diagrams* (IDR-CD) for motion related information like direction of displacement. Closed polygons representing tracked objects are assigned unique pixel



**Fig. 2.** (a) Allen’s interval relations, (b) 18 basic relative position relations, and (c) 8 basic QD<sub>8</sub> relations.

intensity and the remaining pixels contain WHITE values. A sequence of diagrams called *key diagrams* ( $\mathcal{K}$ ) are selected with difference in relative position and relative distance among interested object pairs. Each pair of  $\mathcal{K}$ s are sequentially combined using IDR-OR operator for IDR-CDs. IDR-CDs maintain objects and their relations as *union of diagrammatic objects and relations* in both participating key diagrams. Figure 3(a) and (b) shows a pair of diagrams automated from video frame at time point ‘t’ and ‘t + 1’ respectively, with IA relations and distance information as ‘line’ objects among polygons A, B (tracked objects). Figure 3(c) represents IDR-combined diagram of (a) and (b) with IA relations and ‘rays’ depicting direction of objects A and B traversing from time frame ‘t’ to ‘t + 1’.



**Fig. 3.** Example video frame diagrams (a) at time point ‘t’ (b) at time point ‘t + 1’ with objects A, B showing distance information (‘lines’), and (c) IDR-CD of (a) and (b) with displacement ‘rays’.

*Automatic Perceptual Methodology:* Visual information from ‘diagram’ are perceived through certain perceptual mechanisms. During *diagram creation* and *modification* perceptual information are manipulated for automation of object relations. Qualitative relations are perceived through analysis of quantitative object properties. IA relations among pairs of ‘polygon’ along x-y axes are automated based on their extent along the axes. These *IA relations* are the core

of visualizing *relative positions* and initiating diagram modification for *distance* and *displacement relations*. Figure 4 represents the defined look up table for automation of relative positions based on perceived IA relations. The length of ‘line’ objects are analyzed for relative distance relations. Displacement directions are perceived by analyzing angle between ‘ray’ objects and imaginary ‘rays’ originating at considered ray origin, parallel to x-axis; and the clockwise angle among ‘ray’ object pairs are exploited for relative displacement directions among associated ‘polygons’.

$\begin{matrix} \text{Iax} \rightarrow \\ \text{Iay} \downarrow \end{matrix}$	<	>	m	mi	o	oi	d	di	s	si	f	fi	=
<	Back Left	Front Left	Back Left	Front Left	Back Left	Front Left	Left	Left	Left	Left	Left	Left	Left
>	Back Right	Front Right	Back Right	Front Right	Back Right	Front Right	Right	Right	Right	Right	Right	Right	Right
m	Back Left	Front Left	Back Left	Front Left	Back Left	Front Left	Left	Left	Left	Left	Left	Left	Left
mi	Back Right	Front Right	Back Right	Front Right	Back Right	Front Right	Right	Right	Right	Right	Right	Right	Right
o	Back Left	Front Left	Back Left	Front Left	Tog BLC	Tog FLC	Tog Left	Tog Left	Tog Left	Tog Left	Tog Left	Tog Left	Tog Left
oi	Back Right	Front Right	Back Right	Front Right	Tog BRC	Tog FRC	Tog Right	Tog Right	Tog Right	Tog Right	Tog Right	Tog Right	Tog Right
d	Back	Front	Back	Front	Tog Back	Tog Front	TogIN	TogIN	TogIN	TogIN	TogIN	TogIN	TogIN
di	Back	Front	Back	Front	Tog Back	Tog Front	Tog	Tog	Tog	Tog	Tog	Tog	Tog
s	Back	Front	Back	Front	Tog Back	Tog Front	Tog	Tog	Tog	Tog	Tog	Tog	Tog
si	Back	Front	Back	Front	Tog Back	Tog Front	Tog	Tog	Tog	Tog	Tog	Tog	Tog
f	Back	Front	Back	Front	Tog Back	Tog Front	Tog	Tog	Tog	Tog	Tog	Tog	Tog
fi	Back	Front	Back	Front	Tog Back	Tog Front	Tog	Tog	Tog	Tog	Tog	Tog	Tog
=	Back	Front	Back	Front	Tog Back	Tog Front	Tog	Tog	Tog	Tog	Tog	Tog	Tog

**Fig. 4.** Relative positions of an object with respect to a reference object based on their interval relations along xy-axes.

*Diagram Modification Methodology:* During diagram creation certain information is not visually available. Diagram modification techniques were introduced with abilities to endow new information through automatic insertion of new diagrammatic objects like ‘lines’ for distance information and ‘rays’ for displacement information based on IA relations among associated polygon objects. Diagram modification techniques for ‘line’ object endpoints determination are based on IA relations among ‘polygon’ pairs along the axes. In the same way, ‘ray’ objects end points relay on IA relations along x-y axes, among considered polygons at two different time frames combined together in IDR-CDs. Figures 5 and 6 represent tables based on IA relations among polygons along axes for diagram modification to endow distance ‘lines’ and displacement direction ‘rays’.

$I_{Ax} \rightarrow$ $I_{Ay} \downarrow$	<	>	m	mi	o	oi	d	di	s	si	f	fi	=
<													
>													
m													
mi													
o					0	0	0	0	0	0	0	0	0
oi					0	0	0	0	0	0	0	0	0
d					0	0	0	0	0	0	0	0	0
di					0	0	0	0	0	0	0	0	0
s					0	0	0	0	0	0	0	0	0
si					0	0	0	0	0	0	0	0	0
f					0	0	0	0	0	0	0	0	0
fi					0	0	0	0	0	0	0	0	0
=					0	0	0	0	0	0	0	0	0

Fig. 5. Line object's end points determination during diagram modification for distance information between two polygons based on their interval relations along xy-axes.

$I_{Ax} \rightarrow$ $I_{Ay} \downarrow$	<	>	m	mi	o	oi	d	di	s	si	f	fi	=
<													
>													
m													
mi													
o													
oi													
d													
di													
s													
si													
f													
fi													
=													

Fig. 6. Ray object's end points determination during combined diagram modification for displacement information of polygons from one time frame to the next based on their corresponding interval relations along xy-axes in time 't' and 't + 1'.

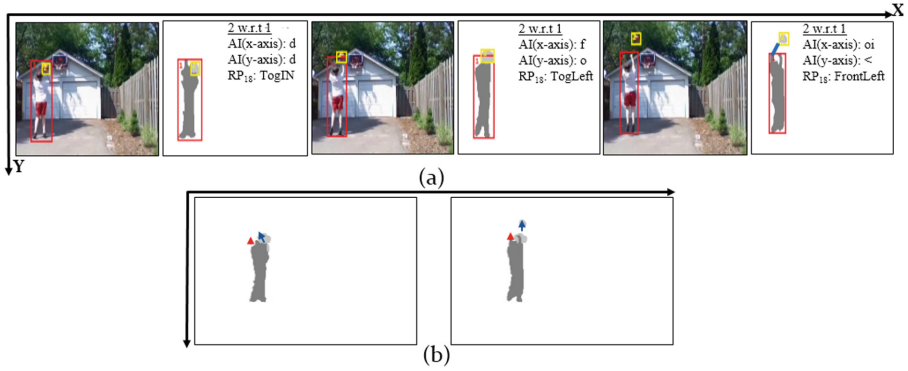
## 4 Application: An Example with Evaluation Details

Video analysis has tremendous applications over automatic video surveillance system, starting from event detection, extending to human activity detection, animal behavioral and movement ecology, and acquiring disaster management scenarios. This paper advocates the proposed video representation methodology through its application over video analysis for activity recognition.

The proposed technique is implemented over videos of J-HMDB dataset<sup>1</sup> involving selected verbs {catch, throw, shoot ball, push and pull up}. Accurate tracking is essential for a reliable video analysis. Due to various factors like, occlusion, presence of noise, various lighting conditions 100% accuracy in tracking results is itself a challenging task. For evaluation of the proposed QSTR and DR mechanism, tracking is achieved through manual labelling of objects of interest; focus is only to validate the proposed methodology, attempts at improvement of tracking being outside the scope of this work. Labelers guiding or refining labels for accuracy might end up with inaccurate tracking information which conflicts the fact about inter-dependability among accuracy in tracking and reliable video analysis. A ‘diagram’ sequence is automated based on tracked objects information in extracted video frames. QSTR techniques over DR are employed for automatic abstraction of spatio-temporal relations among object pairs in forward moving time in terms of displacement direction (Di) of individual objects, their relative positions (RP) at two consecutive time frames, relative displacement directions (RD<sub>i</sub>) and their relative distances (RD<sub>t</sub>). Based on these relations, sequence of certain short term activities (STA) are formulated among objects while traversing among consecutive key diagrams. This STA sequence is considered as a feature vector and a standard supervised machine (SVM) is used for associated activity classification. For example, Fig. 7 shows (a) video frames and associated key diagrams with IA relations and distance information (‘lines’ in blue) and (b) IDR-combined diagrams with direction of displacement information (red ‘rays’ depict displacement of object 1 and blue ‘rays’ depict displacement of object 2) in a *shoot ball* video from J-HMDB dataset. In the example, the abstracted sequence of relations among object 1 and 2 from the IDR-CDs are as shown in Table 1. Based on these relations a sequence of STAs obtained is: {AB:togIN-Tog, ABapart:TogLeft} which constituting the minimal sequence for *shoot ball* activity among object 1 and 2. ‘AB:togIN-Tog’ infers that objects A, B both are in motion and move from completely inside (TogIN) position to partial overlap (TogLeft) position; ‘ABapart:TogLeft’ infers that objects A, B both are in motion and move from completely partial overlap position (TogLeft) to disjoint (Left) position.

The performance of the proposed video analysis methodology for activity recognition is shown in Table 2 in terms of per class accuracy, precision, recall and F1-score. A rough comparison of recognition accuracy of the five considered activities with that of state-of-the art performances is presented. Since, performance accuracy of the activities in published state-of-the-art techniques are

<sup>1</sup> <http://jhmdb.is.tue.mpg.de>.



**Fig. 7.** Sequence of video frames corresponding to selected key diagram from a shoot ball video in JHMDB dataset, with (a) key diagrams with distance information and (b) IDR-CDs with displacement information. (Color figure online)

**Table 1.** Perceived qualitative relations of object A w.r.t. object B in IDR-CDs of example in Fig. 7(b).

	Relations as a set of {Di of A, Di of B, RP of B w.r.t A in time frame ‘t’, RP of B w.r.t A in time frame ‘t + 1’, RDi of B w.r.t A, RDt of B w.r.t A}
<b>IDR-CD<sub>1</sub></b>	{RL, RL+, TogIN, TogLeft, Same+, 0}
<b>IDR-CD<sub>2</sub></b>	{RL, RL, TogLeft, Left, Same, +ve}

**Table 2.** Activity recognition performance on videos of J-HMDB dataset reported in terms of accuracy, precision, recall and F-score. A comparison with state-of-the-art performances is presented as per class accuracy.

Activities	STA sequence + SVM				Stacked feature vectors [16]	Pose-based CNN [6]	Dynamic programmed SVM [17]	GRP [5]	GRP w/o constraints [5]
	Accuracy	Precision	Recall	F-score					
Catch	87.2	70.0	70.0	70.0	38.0	51.0	22.0	50.0	56.0
Throw	95.3	100	80.0	88.9	35.0	38.0	18.0	72.0	72.0
Shoot ball	91.1	87.5	70.0	77.8	77.0	80.0	35.0	45.0	18.0
Push	97.6	90.9	100	95.2	81.0	95.0	60.0	75.0	75.0
Pull up	89.1	69.2	90.0	78.2	100	95.0	86.0	100	100

computed over all the 21 verbs of J-HMDB dataset a coarse comparison of the recognition performance is difficult. However, performance seems to be inspiring for the underlying video analysis methodologies. Better performance may be achieved through consideration of more relational information among STAs during recognition or via some other recognition techniques.



## 5 Conclusion

This paper focus on a comprehensive representation of video data through utilization of strength of ‘diagrams’ in human problem visualization and solution strategies. Human basically convey effective solutions based on mental maps or spatial organization of problems. The proposed methodology is a step towards formalizing the use of diagrams in cognitive vision for representation and reasoning purpose. The authors strongly advocate general concept and perception about a spatio-temporal structure to be computationally effective over formal computation with detailed and complex organizational information. A novel approach of integrating DR and QSTR techniques is being presented for video data representation in a cognitive vision system. This hybrid strategy narrows the option of ambiguity in relational composition. The work presented shows how diagrams and ‘commonsense knowledge’ can be put together for a human like problem definition through procedures like: information perception, endowing new information through diagram modification and inter diagrammatic reasoning. An application over the spatio-temporal abstractions through the proposed methodology for activity recognition is being presented. Few videos from J-HMDB dataset are being evaluated. STAs are defined over the abstracted spatio-temporal information, which serve as feature vector for a supervised machine for activity recognition. Encouraging recognition results are obtained. Further improvement could be achieved by strengthening the feature vector. As an alternative, defining a formal language automata that preserves temporal co-relations among STAs together with sequence information can be considered to uplift the framework towards precise activity recognition. This is a part of ongoing research.

## References

1. Allen, J.F.: Maintaining knowledge about temporal intervals. *Commun. ACM* **26**(11), 832–843 (1983)
2. Anderson, M., McCartney, R.: Diagram processing: computing with diagrams. *Artif. Intell.* **145**(1), 181–226 (2003)
3. Banerjee, B.: Spatial problem solving for diagrammatic reasoning. Ph.D. thesis, The Ohio State University (2007)
4. Baruah, R., Hazarika, S.: Qualitative directions in egocentric and allocentric spatial reference frames. *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.* **6**, 344–354 (2014)
5. Cherian, A., Fernando, B., Harandi, M., Gould, S.: Generalized rank pooling for activity recognition. arXiv preprint [arXiv:1704.02112](https://arxiv.org/abs/1704.02112) (2017)
6. Chéron, G., Laptev, I., Schmid, C.: P-CNN: pose-based CNN features for action recognition. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3218–3226 (2015)
7. Cohn, A.G., Hazarika, S.M.: Qualitative spatial representation and reasoning: an overview. *Fundam. Informaticae* **46**(1), 1–29 (2001)
8. Freksa, C.: Computational problem solving in spatial substrates. *Int. J. Softw. Inform.* **9**(2), 279–288 (2015)

9. Glasgow, J., Narayanan, N.H., Chandrasekaran, B.: *Diagrammatic Reasoning: Cognitive and Computational Perspectives*. MIT Press, Cambridge (1995)
10. Gottfried, B.: Representing short-term observations of moving objects by a simple visual language. *J. Vis. Lang. Comput.* **19**(3), 321–342 (2008)
11. Gottfried, B.: The systematic design of visual languages applied to logical reasoning. *J. Vis. Lang. Comput.* **28**, 212–225 (2015)
12. Jhuang, H., Gall, J., Zuffi, S., Schmid, C., Black, M.J.: Towards understanding action recognition. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3192–3199 (2013)
13. Lieto, A., Chella, A., Frixione, M.: Conceptual spaces for cognitive architectures: a lingua franca for different levels of representation. *Biol. Inspired Cogn. Archit.* **19**, 1–9 (2017)
14. Nath, C.D., Hazarika, S.M.: Combining diagrammatic reasoning with qualitative spatial and temporal reasoning for motion event detection. In: *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pp. 1–4. IEEE (2015)
15. Nath, C.D., Hazarika, S.M.: Qualitative spatial and temporal reasoning over diagrams for activity recognition. In: *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, p. 72. ACM (2016)
16. Peng, X., Zou, C., Qiao, Y., Peng, Q.: Action recognition with stacked fisher vectors. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8693, pp. 581–595. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-10602-1\\_38](https://doi.org/10.1007/978-3-319-10602-1_38)
17. Soomro, K., Idrees, H., Shah, M.: Predicting the where and what of actors and actions through online action localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2648–2657 (2016)
18. Tessler, S., Iwasaki, Y., Law, K.: Qualitative structural analysis using diagrammatic reasoning. In: *IJCAI* (1), pp. 885–893 (1995)
19. Urbas, M., Jamnik, M.: A framework for heterogeneous reasoning in formal and informal domains. In: Dwyer, T., Purchase, H., Delaney, A. (eds.) *Diagrams 2014*. LNCS (LNAI), vol. 8578, pp. 277–292. Springer, Heidelberg (2014). [https://doi.org/10.1007/978-3-662-44043-8\\_28](https://doi.org/10.1007/978-3-662-44043-8_28)