A Human-Centered Approach for Bootstrapping Causal Graph Creation

Minh Q. Tram
minh.tram@mavs.uta.edu
The University of Texas at Arlington
Arlington, TX, USA

Nolan B. Gutierrez
nolan.gutierrez@mavs.uta.edu
The University of Texas at Arlington
Arlington, TX, USA

William J. Beksi
william.beksi@uta.edu
The University of Texas at Arlington
Arlington, TX, USA

ABSTRACT

Causal inference, a cornerstone in disciplines such as economics, genomics, and medicine, is increasingly being recognized as fundamental to advancing the field of robotics. In particular, the ability to reason about cause and effect from observational data is crucial for robust generalization in robotic systems. However, the construction of a causal graphical model, a mechanism for representing causal relations, presents an immense challenge. Currently, a nuanced grasp of causal inference, coupled with an understanding of causal relationships, must be manually programmed into a causal graphical model. To address this difficulty, we present initial results towards a human-centered augmented reality framework for creating causal graphical models. Concretely, our system bootstraps the causal discovery process by involving humans in selecting variables, establishing relationships, performing interventions, generating counterfactual explanations, and evaluating the resulting causal graph at every step. We highlight the potential of our framework via a physical robot manipulator on a pick-and-place task.

KEYWORDS

Human-Robot Interaction, Augmented Reality, Causal Graphs

1 INTRODUCTION

Causality allows for deciphering complex relationships and drawing informed conclusions. This foundational understanding is beginning to make significant contributions in the field of robotics, reshaping the path of its advancement [10]. However, constructing accurate causal models, including causal inference and graph representations, is a daunting task. This stems from the inherent intricacy of the interactions, often compounded by a myriad of variables, which often remain elusive or hidden in the environment.

A causal graphical model, aka causal graph, serves as a visual mathematical representation of the relationships between variables
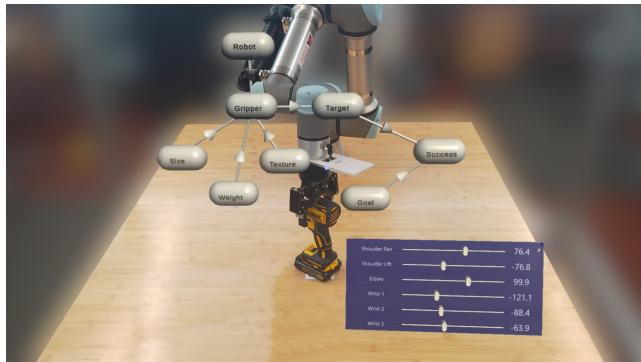
Fig. 1: The point of view of the operator from the AR overlaid physical workspace. The operator can interact with the robot and provide additional context to aid the robot's understanding via on-the-fly construction of causal graphs.

in a system. Delineating the direction and structure of these relationships allows for identifying potential causes and effects, and disentangling confounding variables [14]. By employing causal graphs, robots can predict the outcomes of their actions more accurately, make informed decisions in dynamic environments, and adapt more fluidly to new situations with a deeper comprehension of the underlying mechanisms [5, 9, 16].

Parallel to these advancements, the use of simulations and virtual interfaces offers a more economical means of gathering diverse and valuable data, which is especially conducive to improving machine learning algorithms [11]. For example, the integration of virtual, augmented, and mixed reality (VAMR) has shown promising results in fostering a more intuitive and enhanced interaction between human operators and robots (e.g., [12, 13, 15]). VAMR interfaces provide a way for operators to have a more immersive control experience, bridging the gap between the digital and physical realms.

In this work we present preliminary results on a new framework that leverages VAMR technologies to address the challenge of causal graph construction, Fig. 1. To summarize, our contributions are the following.

- We create an augmented reality (AR) interface that allows an operator to naturally construct a causal graph on the fly.
- We provide a human-centered approach to highlight critical information when establishing a causal graph of a scene.
- Our system ensures that a robot prioritizes human insights during the initial stages of scene comprehension.
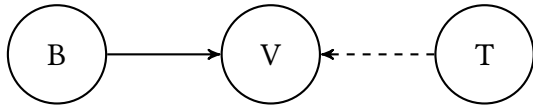
Fig. 2: A causal graph representing the relationship between a robot's battery level B, the terrain roughness T, and the robot's velocity V during navigation. The solid edge represents a well-established cause and effect, while the dashed edge represents an indirect or latent confounding variable.

2 BACKGROUND AND MOTIVATION

2.1 Causal Graphs and Causal Inference

To establish a foundation for causal inference, Pearl introduced a number of mathematical formulations, notably structural causal models (SCMs), intervention and do-calculus, and counterfactuals, to rigorously define the relationships between variables. For example, consider a mobile robot navigation scenario. The causal graph that represents the relationship between the robot's velocity, the battery level, and the roughness or difficulty of the terrain can be represented as a directed acyclic graph (DAG), Fig. 2. In this scenario, we can represent the relationships between the variables using an SCM f as

$$V = f(B, T, U_V), \tag{1}$$

where V is the velocity, B is the battery level, T is the terrain roughness, and U_V is an internal variable that is unobserved, but has an effect on the velocity of the robot. An intervention in this situation would be to charge the robot's battery regardless of any other factor hence ensuring it is always full. This can be represented using do-calculus, i.e.,

$$P(V \mid do(B = \text{full})) = \sum_T P(V \mid B = \text{full}, T = t)P(T = t). \tag{2}$$

In (2), estimates of the speed of the robot are given through the intervention by averaging over all possible values of T. Assuming that there are no confounding relationships between the variables, we have effectively broken the natural causal relationship between B and V, while leaving the causal relationship between T and V intact. This is where counterfactuals can help infer the causal relationship between B and V during an intervention, which can aid in reinforcing or disproving causal relationships between variables. This is important for both pre-planning, on-the-fly decision making, and insights toward post-incident analysis. By using this mathematical framework to understand the causal relationships between variables, we can make informed predictions about a system. For instance, in the robot navigation scenario we can use these relationships to pinpoint the root cause of performance issues, pre-charge the robot's battery if it is expected to go through rough terrain that will drain the battery, or adjust the robot's behavior to achieve a desired outcome.

2.2 Motivation

The use of causal inference is not new in robotics [4, 6–8]. *Nonetheless, automating the creation of a viable causal graph remains an underexplored area of research.* Constructing a causal graph is hard due to the following reasons.
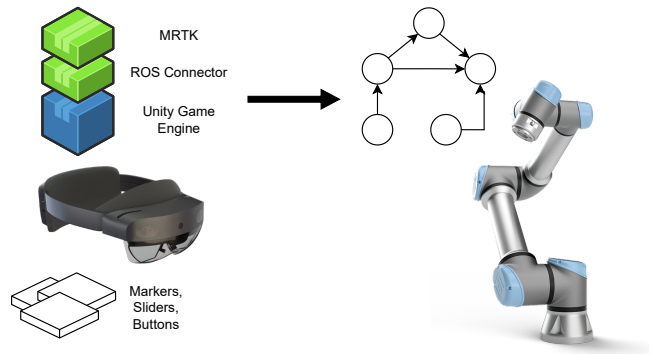

Fig. 3: The visualization and interaction pipeline of our proposed framework. The operator can interact with the robot using an overlay interface via various modes and provide context hinting directly onto the workspace to aid the robot in its understanding of the scene.

- **Complexity of the environment:** Variables and their causal relationships must be identified in dynamic environments while accounting for confounding variables.
- **Causation vs. correlation:** The distinction between causation and correlation from observational data is not always clear, and definitive proof of causation often requires more data.
- **High dimensionality:** The number of variables in an environment can be large, which can lead to complex and/or multiple causal graphs.

3 HUMAN-CENTERED CAUSAL GRAPHICAL MODELS

In this section, we delineate the specific methodologies and procedures employed to actualize human-centered causal graphs. First, we present a brief overview of our framework's components. Then, we provide a detailed description of each component and the communication pipeline between them.

3.1 Component Overview and Consideration

Our framework to facilitate robotic understanding through human-centered causal graph construction is composed of three primary components: (i) an AR-capable headset, (ii) a software interface to visualize the robot's perception and allow the operator to interact with the robot, and (iii) a control interface to realize the operator's commands and interventions. Each of these components plays a pivotal role in the seamless execution of the system, providing both unique and complementary functionalities. Fig. 3 details the overall system architecture and the transmission pipeline between the components.

For the AR headset, we opted for the Microsoft HoloLens 2 due to its unparalleled capabilities. Its commercial availability ensures ease of access and its extensive API support, via Microsoft's Mixed Reality Toolkit (MRTK) [1], offers a versatile platform for development and integration. The HoloLens 2 is a self-contained head-mounted display (HMD) that is capable of spatial mapping and tracking, while also providing a projected holographic display. It permits the operator to operate in a hands-free, untethered manner. This

feature is crucial for our work for two primary reasons. First, being able to superimpose additional information directly onto the workspace enhances the operator's understanding of the robot's perception of the environment. This is vital when visualizing the implications of the causal graph on the robot's physical actions in real-time. Second, the ability to operate in a hands-free manner through different modalities (i.e., direct-hand input, eye-tracking, and voice commands) allows the operator to swiftly and naturally interact with both the robot's actions and the underlying causal structure. This ensures that the operator can manipulate the causal graph, adjust parameters, and interact with robot perception and environment understanding without the constraints of traditional input methods.

3.2 Interaction and Visualization Modalities

Relying on the HoloLens 2's hand-tracking capabilities, our system allows the operator to interact with the robot and environment through a number of mechanisms. This includes direct manipulation of the robot's joint positions via virtual sliders and command buttons, providing a planning context such as desired end-effector positions and approach vectors, and the ability to query forward and inverse kinematics solutions. Furthermore, compared to traditional VR devices where operator interactions are bound to handheld controllers and the operator's vision is obstructed by a display, the HoloLens 2 offers greater interaction and visualization capabilities.

3.3 Causal Graph Construction and Intervention

In our human-centered framework, the focus is primarily on the dynamic construction and intervention of causal graphs. To this end, we leveraged Unity [3] to render causal structures as 3D objects in the AR space allowing the operator to interact with the causal graph directly. These graphical constructs, composed of nodes and edges, offer a visual guide that aids the operator in deciphering the complex interplay between robot actions and environmental cues.

Nodes, visualized as tangible entities within the AR landscape, symbolize specific events or actions that the robot can detect or perform. Their design incorporates distinct visual features, such as shades, opacities, and pulsations, communicating their present state or relevance. On the other hand, edges provide visual insights into the direction, strength, or likelihood of causal linkages. These are often characterized by variations in their thickness, texture, or color gradient, allowing users to quickly comprehend causal dependencies.

Within the immersive world facilitated by the HMD, operators can actively modify and intervene in the causal relationships by adding, removing, or modifying the nodes and edges of the current causal structure. The device's advanced hand-tracking feature supports intuitive gestures, such as pinching, swiping, or rotating, enabling users to adjust causality strength, node prominence, or even the direction of a causal link. Furthermore, the mobility of the HMD lets an operator freely move around the environment and provide context hints directly onto the robot workspace, accommodating seamless adjustments and experiments with the causal graph.

3.4 Physical Robot Actuation and Simulation

The physical robot actuation and simulation components of our system, driven by ROS 2 Humble [2], facilitate communication between the robot and Unity software layers. This allows the operator to interact with the robot and environment through the AR interface. Concretely, it enables the translation of intricate directives from a causal graph into precise physical movements of the robot, ensuring adaptability and dexterity in diverse real-world scenarios.

4 EVALUATION

4.1 Causal Graph Creation

Using a UR5e robot arm provisioned with a Robotiq 2F-85 gripper and tasked with picking up a finite set of known objects and placing them at a desired location, we construct a causal graph to represent the relationship between the robot's actions and the environment. To achieve this, the potential variables related to our task are represented as follows.

(1) *Robot*: The robot state.
(2) *Gripper*: The gripper state.
(3) *Target*: The robot's motion planning to achieve the goal position.
(4) *Goal*: The desired object placement location.
(5) *Type*: The object's type.
(6) *Weight*: The object's weight (light - heavy).
(7) *Size*: The object's size (small - large).
(8) *Texture*: The object's texture (smooth - rough).
(9) *Rigidity*: The object's rigidness (soft - hard).
(10) *Success*: The success rate of a pick-and-place sequence.

For the purpose of demonstration, we construct this causal graph under the assumption that external factors (e.g., robot power loss, environmental disruptions, etc.) do not occur and are insulated from the causal representation. As shown in Fig. 4, the relationships between the variables can be visually represented as a graph for an easier understanding of potential causality and the interplay between nodes. Within the graph, each edge represents a direct causal relationship between the robot and the operating environment. For example, if the weight of the object changes then it could influence the behavior of the gripper (e.g., the gripper force or grasp vector).

Our framework can directly create the causal graph in the AR space, Fig. 5. The operator interacts via the AR interface to label the nodes and edges, set the content of each node, and adjust the strength of each edge. The operator can also communicate with the robot and the environment through the AR interface to obtain additional information and context to aid in the construction of the causal graph. The causal graph can then be parsed and interpreted by the motion planning backend during the planning step to ensure that the robot's actions are consistent with the state of the graph. Algorithm 1 demonstrates how the robot consumes a causal graph generated by the operator during each motion planning step.

4.2 Causal Graph Intervention

For a pick-and-place task, we use our constructed causal graph to evaluate whether the relationships between the defined variables are valid, which in turn can be inferred by the robot's performance. For instance, if the robot consistently fails to pick up a cordless drill
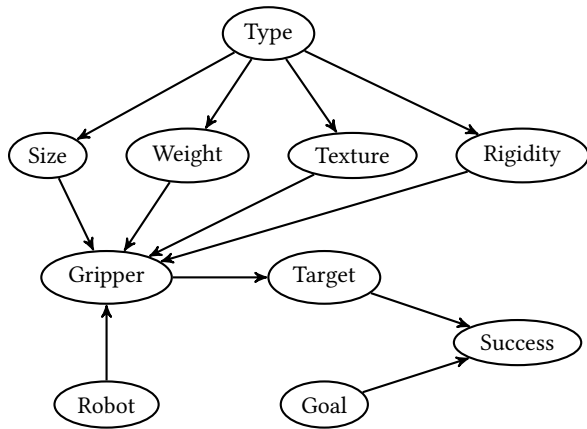
Fig. 4: An example of the relationships between the operator-identified variables for a pick-and-place task. Each edge corresponds to a direct causal relationship between the connected variables. Changes in the source of an edge have a direct consequence on the target of an edge.

Algorithm 1 Human-in-the-loop Causal Graph Iteration

Require: Human has initialized the causal graph
Ensure: Causal graph is DAG, all nodes have properties
1: **while** true **do**
2:     **if** operator is modifying causal graph **then**
3:         **Wait**
4:     **else**
5:         **Execute** with guidance from causal graph
6:         **Check** execution status
7:         **if** status is SUCCESS **then**
8:             **Continue**
9:         **else**
10:            **Notify** Operator
11:            **Ask** Continue?
12:         **end if**
13:     **end if**
14: **end while**

---

but it has no issues with other objects, then we can be relatively certain that several known properties or relationships (e.g., weight, texture, etc.) may be detrimentally affecting the robot. With this human-in-the-loop approach, we can quickly devise an intervention procedure to decide if certain variables are causing degraded performance. For example, we can intervene by setting the object's weight to a specific value and then benchmark the robot's performance under the assumption that the robot did not account for the object's weight properly.

If the robot's performance is influenced by the weight of the object, then we can make adjustments to the robot's behavior to ensure optimal performance (e.g., adjusting the pick points, etc.). Alternatively, should the intervention be inconclusive, then we can intervene by fixing other object properties to a specific value and then benchmark the robot's performance. However, further investigation may be required to understand the exact reason. For example, the gripper might not be suited for the cordless drill's
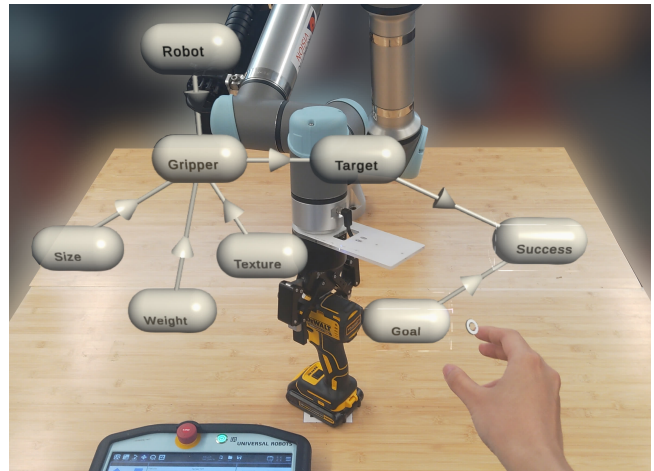

Fig. 5: The operator interacting with the robot and environment through the AR interface to construct a causal graph. Nodes and edges are visualized and can be manipulated directly in the AR space.

shape, or the perception system may not be identifying the drill's correct pick points. Since our framework has the capability to adjust causal graph structure on the fly, we can quickly update the graph to reflect the new structure thus allowing the robot to make more informed decisions.

4.3 Discussion

Our framework provides advantages over traditional methods for constructing causal graphs. The AR interface allows an operator to interact with the robot and the environment directly and more intuitively, making it easier for humans to understand the system's dynamics. It also presents a visual understanding of causal relationships, which allows operators to make informed decisions on how to optimize the system. For example, the operator can visualize in real time where interventions might be needed, or where potential failures could occur.

Constructing causal graphs this way introduces a level of modularity and scalability. For instance, causal elements can be easily adjusted to fit the needs of the system or the entire graph can be replaced by constructing a replacement from scratch. Yet, this methodology is not without its limitations. A criticism of causal graphs is that they can be an oversimplification of the environment and interplay between variables. Real-world robotic operations may involve numerous subtle interactions, not all of which can be accurately captured in a graph.

5 CONCLUSION

This paper provided preliminary results on a human-centered approach for automating the construction of causal graphical models. Our framework combines the strengths of VAMR and simulation technologies to address the dilemma of bootstrapping the creation of a causal graph. Furthermore, we highlighted the ability to visualize, intervene, and update causal graphs on a physical robot for a real-world pick-and-place task. In the future we will work on addressing the challenges of oversimplification and iteratively updating causal graphs.

REFERENCES

[1] 2024. *Microsoft Mixed Reality Toolkit.* https://github.com/microsoft/MixedRealityToolkit-Unity

[2] 2024. *Robot Operating System.* https://www.ros.org

[3] 2024. *Unity 3D Game Engine.* https://unity.com

[4] Jake Brawer, Meiying Qin, and Brian Scassellati. 2020. A Causal Approach to Tool Affordance Learning. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems.* 8394–8399.

[5] Yu Cao, Boyang Li, Qian Li, Adam Stokes, David Ingram, and Aristides Kiprakis. 2021. Reasoning Operational Decisions for Robots via Time Series Causal Inference. In *Proceedings of the IEEE International Conference on Robotics and Automation.* 6124–6131.

[6] Maximilian Diehl and Karinne Ramirez-Amaro. 2023. A Causal-Based Approach to Explain, Predict and Prevent Failures in Robotic Tasks. *Robotics and Autonomous Systems* 162 (2023), 104376.

[7] Wenhao Ding, Haohong Lin, Bo Li, and Ding Zhao. 2023. CausalAF: Causal Autoregressive Flow for Safety-Critical Driving Scenario Generation. In *Proceedings of the Conference on Robot Learning.* PMLR, 812–823.

[8] Seng-Beng Ho, Mark Edmonds, and Song-Chun Zhu. 2020. Actional-Perceptual Causality: Concepts and Inductive Learning for AI and Robotics. In *Proceedings of the IEEE Symposium Series on Computational Intelligence.* 442–448.

[9] Tabitha E. Lee, Jialiang Alan Zhao, Amrita S. Sawhney, Siddharth Girdhar, and Oliver Kroemer. 2021. Causal Reasoning in Simulation for Structure and Transfer Learning of Robot Manipulation Policies. In *Proceedings of the IEEE International Conference on Robotics and Automation.* 4776–4782.

[10] Yangming Li. 2023. Deep Causal Learning for Robotic Intelligence. *Frontiers in Neurorobotics* 17 (2023).

[11] Mayank Mittal, Calvin Yu, Qinxi Yu, Jingzhou Liu, Nikita Rudin, David Hoeller, Jia Lin Yuan, Ritvik Singh, Yunrong Guo, Hammad Mazhar, Ajay Mandlekar, Buck Babich, Gavriel State, Marco Hutter, and Animesh Garg. 2023. Orbit: A Unified Simulation Framework for Interactive Robot Learning Environments. *IEEE Robotics and Automation Letters* (2023).

[12] David Puljiz, Franziska Krebs, Fabian Bosing, and Bjorn Hein. 2020. What the HoloLens Maps Is Your Workspace: Fast Mapping and Set-up of Robot Cells via Head Mounted Displays and Augmented Reality. In *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems.* 11445–11451.

[13] David Puljiz, Erik Stöhr, Katharina S. Riesterer, Björn Hein, and Torsten Kröger. 2019. General Hand Guidance Framework Using Microsoft HoloLens. In *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems.* 5185–5190.

[14] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. 2021. Toward Causal Representation Learning. *Proc. IEEE* 109, 5 (2021), 612–634.

[15] Minh Q Tram, Joseph M Cloud, and William J Beksi. 2023. Intuitive Robot Integration via Virtual Reality Workspaces. In *Proceedings of the IEEE International Conference on Robotics and Automation.* 11654–11660.

[16] Caiming Xiong, Nishant Shukla, Wenlong Xiong, and Song-Chun Zhu. 2016. Robot Learning with a Spatial, Temporal, and Causal and-or Graph. In *Proceedings of the IEEE International Conference on Robotics and Automation.*