VR-ProcessMine: Immersive Process Mining Visualization and Analysis in Virtual Reality

Roy Oberhauser[0000-0002-7606-8226]
Computer Science Dept.
Aalen University
Aalen, Germany
e-mail: roy.oberhauser@hs-aalen.de

***Abstract*—Repeatable processes are fundamental for describing how enterprises and organizations operate, for production, for Industry 4.0, etc. As digitalization and automation progresses across all organizations and industries, including enterprises, business, government, manufacturing, and IT, evidence-based comprehension and analysis of the processes involved, including their variations, anomalies, and performance, becomes vital for an increasing set of stakeholders. Process Mining (PM) relies on logs or processes (as such evidence-based) to provide process-centric analysis data, yet insights are not necessarily visually accessible for a larger set of stakeholders (who may not be process or data analysts). Towards addressing certain challenges described in the Process Mining Manifesto, this paper contributes VR-ProcessMine, a solution for visualizing and interacting with PM results in Virtual Reality (VR). Our realization shows its feasibility, and a case-based evaluation provides insights into its capabilities.**

Keywords-virtual reality; process mining; process analysis; business process mining; business process management.

I. Introduction

The digital transformation sweeping through society affects businesses and organizations everywhere, resulting in an increased emphasis on business agility and automation. Business Processes (BPs) or workflows are one significant automation area, as evidenced by the Business Process Management (BPM) market, which is forecast to grow from \$8.8B in 2020 to \$14.4B by 2025 [1]. Each execution of such a process leaves a digital footprint of process-related events and the timepoint of execution, typically contained in various log files across the various IT systems (business, manufacturing, etc.) involved in an enterprise. BPs are a way for ordering the activities involved in and executed in an enterprise, be they automated, semi-automated, or human-driven, and thus BPM is where much of the value generated by an enterprise is achieved.

Process Mining (PM) is a sub-field of data science specifically focused on analyzing event data generated when (business) processes are executed [2]. Because PM relies on event logs of actual process executions, it is evidence-based (or fact-based). This analysis can provide essential insights for understanding and optimizing (business) process execution. When referring to processes we assume BPs to be a subset of the more abstract term and will use both terms interchangeably. One process variant represents a set of process instances that resulted in the same sequence of events.

The Process Mining Manifesto [3] describes eleven challenges for PM. Two of these, *C10: Improving Usability for Non-experts*, and *C11: Improving Understandability for Non-experts*, are a primary motivation for our work. A secondary effect is to address *C9: Combining Process Mining with other Types of Analysis*.

In general, visualization remains a challenge when dealing with large data sets that involve relations and different variation sets. As data and processes become more relevant to the digital enterprise and stakeholders more digitally savvy, it is all the more relevant and challenging to include non-expert enterprise stakeholders in process analysis. By leveraging Virtual Reality (VR), BP analysis can be made more accessible to a wider set of stakeholders, such that not just process modeling specialists, but also those directly involved in executing a BP or observing an automated BP can view und gain insights to various issues regarding a BP of interest, including the combination with other relevant enterprise models.

In prior work, we developed VR-BPMN [4] to visualize business processes in VR based on the BPMN notation. Our VR-EA [5] contributed a VR solution for visualizing, navigating, annotating, and interacting with ArchiMate EA models. And VR-EAT [6] presented our VR-based solution for visualizing dynamically-generated EA diagrams from EA tools. This paper contributes VR-ProcessMine, a solution for visualizing and interacting with PM results in Virtual Reality (VR). Our prototype realization shows its feasibility, and a case-based evaluation provides insights into its capabilities for addressing the aforementioned challenges.

The remainder of this paper is structured as follows: Section 2 discusses related work. In Section 3, the solution is described. Section 4 provides details about the realization. The evaluation is described in Section 5 and is followed by a conclusion.

II. Related Work

PM is supported by various tools. Open source tools includes the ProM Framework [7], Apromore [8], and PM4Py [9]; commercial options include products from over 35 vendors, including Celonis, Disco, UiPath, ARIS, and PAFnow. These tools typically provide a 2D user interface with some being Web-based interfaces (e.g., Celonis, UiPath, Apromore), whereas we provide a VR-based solution.

 ISBN: 978-1-61208-986-7

Work involving PM with VR include Vogel & Thomas [10] show groundwork and an architecture concept, yet no prototype is described nor are VR screenshots provided. Other work combining PM with VR is often specialized to processes in certain sectors, such as training for factory or manufacturing, logistics, safety, or education and learning, or the health sector. For instance, Roldán et al. [11] describe a complex assembly training system for Industry 4.0 operators.

III. Solution

VR is defined as a "real or simulated environment in which the perceiver experiences telepresence" (Steuer 1992), a mediated visual environment, which is created and then experienced. VR provides an unlimited space for visualizing a growing and complex set of enterprise models and processes and their interrelationships simultaneously in a spatial structure. As enterprise models grow in complexity and reflect the deeper integration of both the business and IT reality, an immersive digital enterprise environment provides an additional visualization capability to comprehend the "big picture" for structurally and hierarchically complex and interconnected diagrams and digital elements, while providing an immersive experience for digital process model visualization and analysis in a 3D space viewable from different perspectives.

As to benefits of an immersive VR experience vs. 2D for an analysis task, [12] investigated a software analysis task that used a Famix metamodel of Apache Tomcat source code dependencies in a force-directed graph. They found that VR does not significantly decrease comprehension and analysis time nor significantly improve correctness (although fewer errors were made). While interaction time was less efficient, VR improved the UX (user experience), being more motivating, less demanding, more inventive/innovative, and more clearly structured.

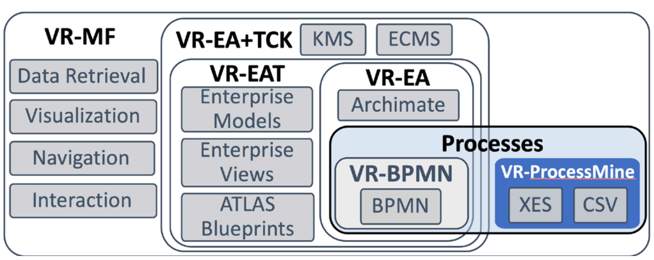

Figure 1. The VR-ProcessMine solution concept (blue) in relation to our prior VR solution concepts.

Our generalized solution concept for VR-ProcessMine is shown in Figure 1. VR-ProcessMine utilizes our generalized VR Modeling Framework (VR-MF) [5], which provides a VR-based domain-independent hypermodeling framework, which addresses four primary aspects that require special attention when modeling in VR: visualization, navigation, interaction, and data retrieval. VR-EA [5] provides specialized direct support and mapping for EA models in VR, including both ArchiMate as well as BPMN via VR-BPMN [4]. VR-EAT [6] extends this further with integration of EA tools for accessing dynamically generated diagrams and models from an EA tool in VR. VR-EA+TCK extends these capabilities by integrating further enterprise knowledge, information, and content repositories such as a Knowledge Managemement System (KMS) and/or an Enterprise Content Management System (ECMS).

In order to support PM and visual analysis, the VR-ProcessMine solution should exhibit the following capabilities:

- *Log file import*: event logs in different event log formats can be imported and processed;
- *Multiple analyses*: multiple event logs can be loaded in order to compare them directly;
- *3D visualization*: elements should be depicted in 3D to support an immersive observation experience;
- *Free element placement*: an individual analysis should be movable in VR space so that they can be compared in locality with another analysis of interest;
- *Hide/show analyses*: to minimize visual clutter, analyses can be hidden and then seen again;
- *Trace detection*: relations between events should be clearly visible; and
- *Colored hot spots*: events are colored to indicate their relative frequency.

A. Visualization in VR

In order to differentiate process variants (depending on the analysis being done), our visualization concept for VR depicts each of these on separate vertical plates standing on a common hyperplane representing a single process. This permits any plate to be selected, moved, and compared with others of interest. Furthermore, since the number of process variants can be very large, it leverages the unlimited space in VR, allowing the hyperplane to depict many process variants at once. All process variants are initially equally spaced on the hyperplane and can be compared with each other.

B. Navigation in VR

The immersion afforded by VR requires addressing how to navigate the space while reducing the likelihood of potential VR sickness symptoms. Thus, two navigation modes are included in the solution: the default uses gliding controls, enabling users to fly through the VR space and view objects from any angle they wish. Alternatively, teleporting permits a user to select a destination and be instantly placed there (i.e., by instantly moving the camera to that position); while this can be disconcerting, it may reduce the susceptibility to VR sickness for those prone to it that can occur when moving through a virtual space.

C. Interaction in VR

Since interaction with VR elements has not yet become standardized or intuitive, in our VR concept, user-element interaction is handled primarily via the VR controllers and a virtual tablet. Our VR-Tablet provides detailed context-specific element information, and can provide a virtual keyboard for text entry fields (via laser pointer key selection) when needed.

IV. Realization

Our solution prototype is partitioned into the Data Hub, a backend for data processing and PM, and the front end responsible for VR visualization (see Figure 2).

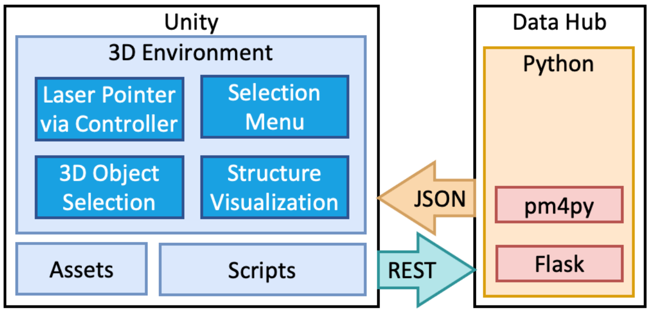

Figure 2. VR-ProcessMine logical architecture.

A. Process Mining

The Data Hub, based on Python 3.9, prepares datasets for visualization. Flask was used to provide REST APIs with JSON for frontend integration. The python library pm4py (Process Mining for Python) [9] is used to convert imported log files into data objects and data frames. The result depends on the type of graph desired. A Directly-Follows Graph (DFG) algorithm provides a summary of all process event transitions and variants and how often each process variant was executed. Listing 1 shows the result of the parsed dataset providing the number of transitions occurring between two events.

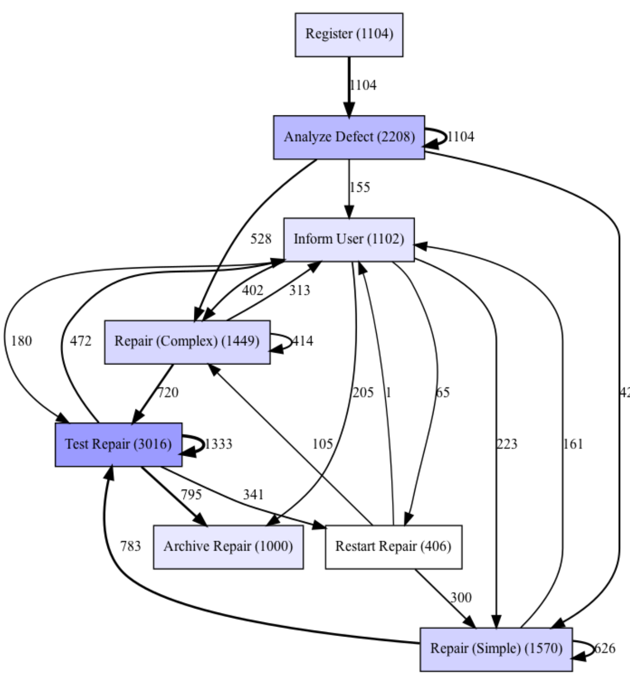

Figure 3. Example DFG-based process map result from pm4py.

Figure 3 shows a DFG-based process map visualization result using pm4py. A node represents an event. A graph consists of a set of transitions between a set of nodes.

Data is converted to a dictionary, whereby all duplicates are removed so that each node exists only once in the graph. The recursive list of fan-in relations (nodes reaching this node) together with their occurrence frequency provides the basis for a weighted directed graph as shown in Listing 2. Aggregating the total occurrences across all incoming transitions of a node (event) provides a total frequency of that event across all process instances.

```
Counter({
    ('Test Repair', 'Test Repair'): 1333,
    ('Register', 'Analyze Defect'): 1104,
    ('Analyze Defect', 'Analyze Defect'): 1104,
    ('Test Repair', 'Archive Repair'): 795,
    ('Repair (Simple)', 'Test Repair'): 783,
    ('Repair (Complex)', 'Test Repair'): 720,
    ('Analyze Defect', 'Repair (Complex)'): 528,
    ...
}
```

Listing 1. Snippet of a parsed DFG dataset.

```
{
  "Register": {
    "start":
  },

  "Analyze Defect": {
    "Register": 1104,
    "Analyze Defect": 1104
  },

  "Repair (Complex)": {
    "Analyze Defect": 528,
    "Repair (Complex)": 414,
    "Inform User": 402,
    "Restart Repair": 105
  },
    ...
}
```

Listing 2. Recursive list of fan-in relations.

B. Virtual Reality

VR realization is handled by the frontend, based on Unity 2021 using Steam VR and implemented in C#. VR hardware consisted of an HTC Vive. Via our VR-Tablet concept, a user can specify a log file to be processed. Currently CSV and XES formats are supported. Once loaded, the first plate shows a DFG with entire set of nodes and transition frequency (see Figure 4), with the plates behind it showing the different process variants (if any).

To support interaction, an affordance in the form of an anchor (sphere) is provided on a corner of each vertical plate or hyperplane, which if selected can be used to reduce visual clutter by collapsing (hide) or expanding (show) that object, or the anchor can be used to place the object elsewhere.

The total number of input transitions to a node represent the total number of times that event occurred. Thus, the higher this number, the higher the frequency. To represent this visually, a ten-step color scale was used to map the frequency between low activity (blue) and high activity (red), analogous to mapping temperature (see Figure 5). This can be used to quickly identify frequently occurring events in a process and help focus analysis and potential optimizations.

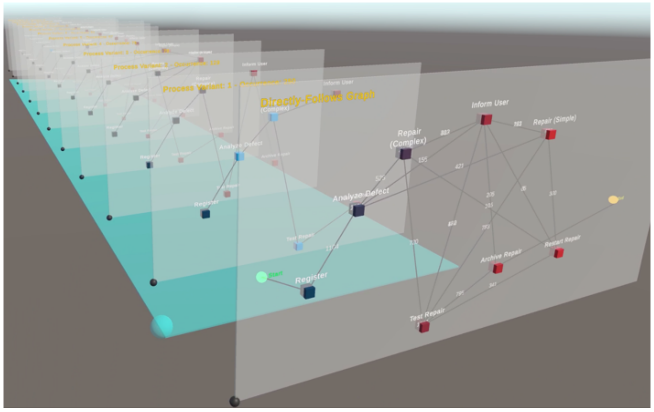

Figure 4. Initial plate shows DFG with entire set of nodes and transition frequency. Each plate thereafter represents one process variant.

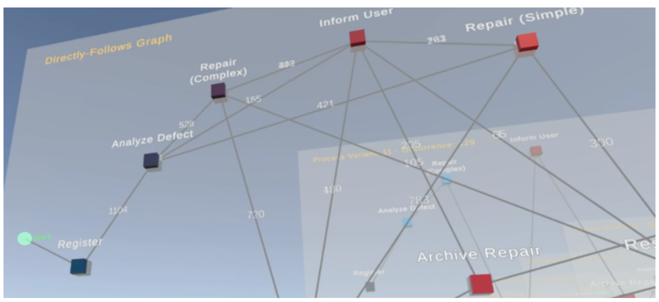

Figure 5. Node color and edges in a DFG.

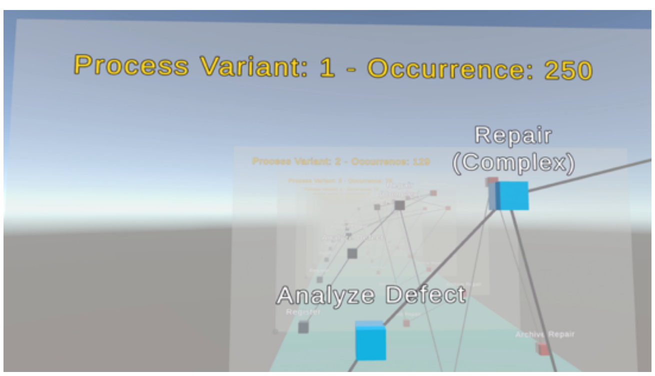

Figure 6. Partial process variant showing occurence frequency.

Process variants with only the relevant nodes (events) that occurred are shown uniquely in separate planes. The variants are assigned a default ID with their occurrence frequency indicated on top (see Figure 6). To support analysis and differentiation, placement of all nodes follows the initial placement in the DFG so that variants can be placed to overlap across the z-axis with the nodes in the same position, or when placed side-by-side the equivalent node locations are placed in the same relative position on each plate.

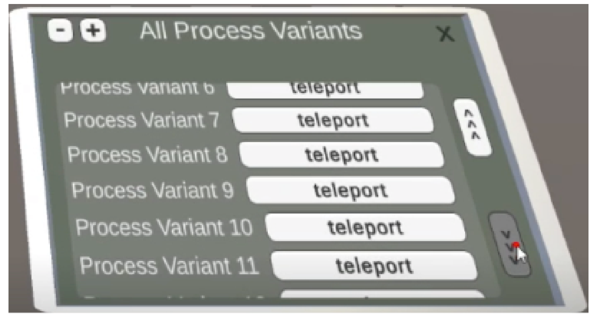

Figure 7. VR-Tablet showing scrollable variant list with a teleport option.

Besides the fly-through ability in VR using the controllers, navigation support includes the ability to quickly teleport to a specific variant by using the VR-Tablet and choosing a variant of interest as shown in Figure 7.

V. EVALUATION

To evaluate the solution, we utilize a case study focusing on three challenges identified in the Process Mining Manifesto [3]. These are:

C10: Improving Usability for Non-experts and

C11: Improving Understandability for Non-experts.

C9: Combining Process Mining with other Types of Analysis.

Our dataset consisted of randomly generated process variants based on a software defect repair process (a snippet is shown in Listing 3). A process variant represents multiple process instances that exhibited the same sequence (or node transition) order.


```
{
"variant": "Register,Analyze Defect,Analyze Defect,Inform User,Repair (Complex),Repair
(Complex),Test Repair,Test Repair,Archive Repair",
"case:concept:name": 78
},
{
"variant": "Register,Analyze Defect,Analyze Defect,Repair (Simple),Inform User,Repair (Simple),
Test Repair,Test Repair,Archive Repair",
"case:concept:name": 75
},
{
"variant": "Register,Analyze Defect,Analyze Defect,Repair (Simple),Repair (Simple),Test Repair,
Test Repair,Inform User,Archive Repair",
"case:concept:name": 67
},
{
"variant": "Register,Analyze Defect,Analyze Defect,Repair (Complex),Repair (Complex),Test
Repair,Inform User,Test Repair,Archive Repair",
"case:concept:name": 64
},
{
"variant": "Register,Analyze Defect,Analyze Defect,Repair (Simple),Repair (Simple),Test Repair,
Inform User,Test Repair,Archive Repair",
"case:concept:name": 41
},
{
"variant": "Register,Analyze Defect,Analyze Defect,Repair (Simple),Repair (Simple),Test Repair,
Test Repair,Restart Repair,Repair (Simple),Inform User,Repair (Simple),Test Repair,Test Repair,
Archive Repair",
"case:concept:name": 29
},
{
"variant": "Register,Analyze Defect,Analyze Defect,Inform User,Repair (Simple),Repair (Simple),
Test Repair,Test Repair,Archive Repair",
"case:concept:name": 28
},
{
"variant": "Register,Analyze Defect,Analyze Defect,Repair (Simple),Repair (Simple),Test Repair,
Test Repair,Inform User,Restart Repair,Repair (Simple),Repair (Simple),Test Repair,Test Repair,
Archive Repair",
"case:concept:name": 21
},
```


Listing 3. Dataset snippet of randomized process variants (in JSON).

A. Improving Usability and Understandability for Non-experts

One aspect that these challenges intend to address is that PM results are made accessible to end-users, and that they can intuitively interact with these, and understand these in their daily work routines. As such, they require intuitive user interfaces (UIs) to support usability as detailed in C10.

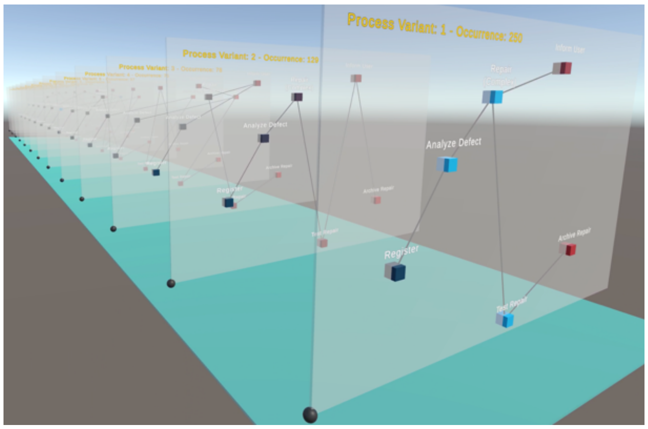

Figure 8. All process variants displayed on a hyperplane.

To support user-friendly UIs, our solution provides an immersive VR experience addressing visualization, navigation, interaction, and data retrieval. With regard to visualization, each process variant is visualized on a 2D plate (Figure 8), leveraging the third dimension for scaling to display all process variants, providing an overview of how many variants exist. Via fly-through navigation, differences can be observed. Furthermore, via our VR-Tablet concept (Figure 7), the user can instantly teleport to a specific one. For interaction, the VR-Tablet can provide details of an event or object.

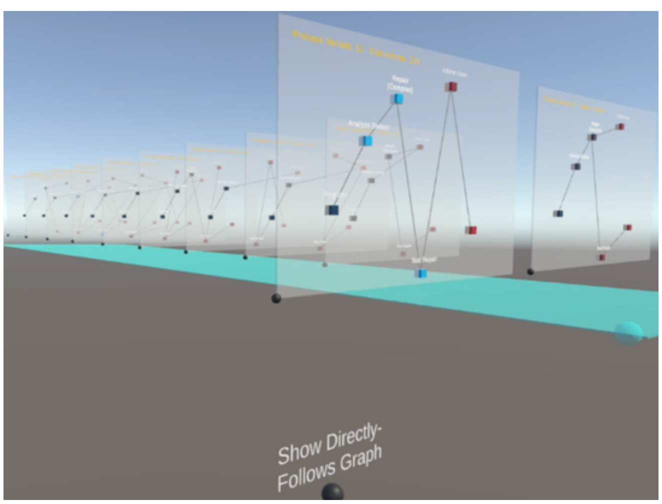

Figure 9. Anchor control for variant comparison or collapsing/expanding.

Plate anchors provide an affordance for flexibly collapsing or moving plates (Figure 9). Via data retrieval from our data hub concept, the VR-Tablet hides the sophisticated PM algorithms within a PM service, making suitable types of PM analysis accessible via the integration of the pm4py library as a service.

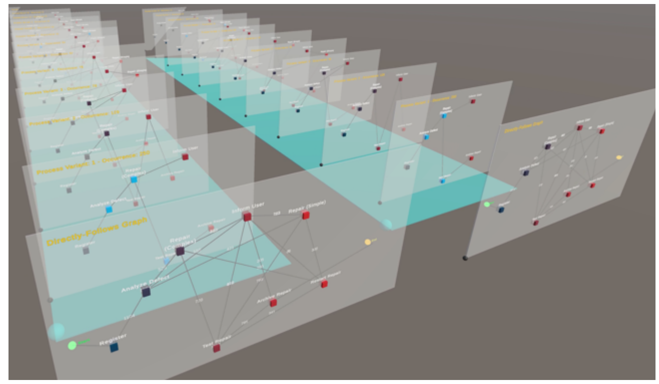

Figure 10. Different processes can be displayed via different hyperplanes.

Furthermore, via the hyperplane concept, our solution can scale to depict a large set of different processes (and their variants) simultaneously, supporting larger cross-process analysis scenarios as seen in Figure 10.

B. Combining PM with other Types of Analysis

Although PM and analysis does provide operational insights, any resulting outcomes typically hinge on some comparison with the original process model (or schema) and potentially other enterprise-relevant knowledge or data. Furthermore, a result of a comparison to a process model may require adjustments to the process model to remove errors or for effectiveness improvements or efficiency optimizations. One advantage of VR's unlimited space is the ability to represent multiple heterogenous models simultaneously and for non-experts to immersively discover and navigate these models.

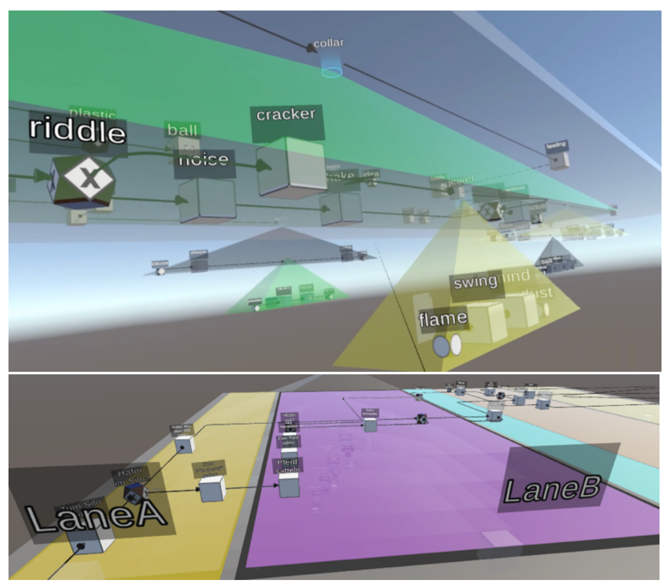

Figure 11. VR-BPMN [4]: subprocesses (top) and swimlanes (bottom) in VR.

Towards combining PM with other analysis types, where BPMN models are available, VR-ProcessMine results can be displayed side-by-side with a VR-BPMN model (Figure 11), allowing non-experts to immersively comprehend various process aspects not necessarily evident via PM, such as subprocesses and swimlanes.

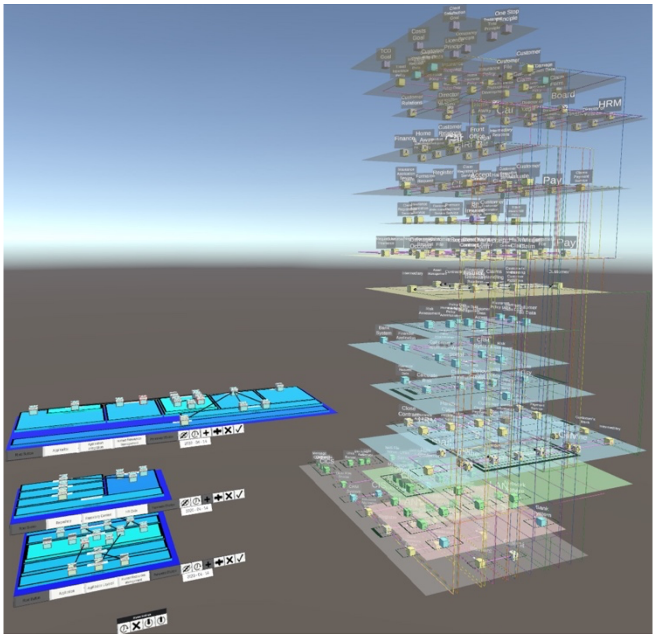

Figure 12. VR-EAT: Enterprise hypermodeling and analysis in VR.

Furthermore, as exemplified with VR-EAT [6], our VR-based heterogeneous hypermodeling capability shown in Figure 12, PM results could be visualized next to other enterprise models, making deeper cross-model and cross-domain analysis readily accessible and allowing operational insights to guide such an analysis.

VI. Conclusion

Increasing digitalization in enterprises and organizations implies that the business and operational processes executed will increasingly also become digitally accessible, offering a significant opportunity. While current PM tools and techniques can provide valuable insights for optimizing (business) processes, these benefits can be hindered when possible insights are not readily accessible to a larger (non-expert) stakeholder set, including those directly involved in performing these processes. VR-ProcessMine contributes an immersive solution concept for visualizing and interacting with PM results in VR. Our realization shows its feasibility, and the case-based evaluation provides insights into its capabilities towards addressing certain challenges described in the Process Mining Manifesto, in particular improving usability, understandability, and the potential to combine PM with other types of analysis.

Future work includes more comprehensive PM analyses, deeper integration with our enterprise hypermodeling VR-EA-TCK and VR-BPMN solution concept, automatic filtering of process variants by a node or transition of interest, collaboration support, and a comprehensive empirical study.

Acknowledgment

The authors would like to thank Laura Berroth for her assistance with the design, implementation, evaluation, and diagrams.

References

[1] Business Process Management Market by Component, Deployment Type, Organization Size, Business Function, Industry, and Region - Global Forecast to 2025. [Online]. Available from: https://www.marketsandmarkets.com/Market-Reports/business-process-management-market-157890056.html 2022.05.10

[2] W. van der Aalst, Process Mining - Data Science in Action, Second Edition. Springer, 2016.

[3] W. V. D. Aalst, et al., "Process Mining Manifesto," In: International Conference on Business Process Management. Springer, Berlin, Heidelberg, 2011, pp. 169-194.

[4] R. Oberhauser, C. Pogolski, and A. Matic, "VR-BPMN: Visualizing BPMN models in Virtual Reality," In: Shishkov, B. (ed.) BMSD 2018. LNBIP, vol. 319, Springer, Cham, 2018, pp. 83–97. https://doi.org/10.1007/978-3-319-94214-8_6

[5] R. Oberhauser and C. Pogolski, "VR-EA: Virtual Reality Visualization of Enterprise Architecture Models with ArchiMate and BPMN," In: Shishkov, B. (ed.) BMSD 2019. LNBIP, vol. 356, Springer, Cham, 2019, pp. 170–187. https://doi.org/10.1007/978-3-030-24854-3_11

[6] R. Oberhauser, P. Sousa, and F. Michel, "VR-EAT: Visualization of Enterprise Architecture Tool Diagrams in Virtual Reality," In: Shishkov B. (eds) Business Modeling and Software Design. BMSD 2020. LNBIP, vol 391, Springer, Cham, 2020, pp. 221-239. https://doi.org/10.1007/978-3-030-52306-0_14

[7] B. F. Van Dongen, A. K. A. de Medeiros, H. Verbeek, A. Weijters, and W. van der Aalst, "The ProM Framework: A new era in Process Mining Tool Support," In: International Conference on Application and Theory of Petri Nets, Springer, 2005, pp. 444–454.

[8] M. La Rosa et al., "APROMORE: An advanced process model repository," Expert Systems with Applications, 38(6), pp.7029-7040, 2011.

[9] A. Berti, S. J. van Zelst, and W. M. P. van der Aalst, "Process Mining for Python (PM4Py): Bridging the Gap Between Process-and Data Science," In: Proceedings of the ICPM Demo Track 2019, co-located with 1st International Conference on Process Mining (ICPM 2019), 2019, pp. 13-16. http://ceur-ws.org/Vol-2374/

[10] J. Vogel and O. Thomas, "Towards a virtual reality-based process elicitation system," 40 Years EMISA 2019. LNI P-304, Gesellschaft für Informatik e.V., 2020, pp. 91-104, ISBN 978-3-88579-698-5

[11] J. J. Roldán, E. Crespo, A. Martín-Barrio, E. Peña-Tapia, and A. Barrientos, "A training system for Industry 4.0 operators in complex assemblies based on virtual reality and process mining," Robotics and computer-integrated manufacturing, 59, 2019, pp.305-316.

[12] R. Müller, P. Kovacs, J. Schilbach, and D. Zeckzer, "How to master challenges in experimental evaluation of 2D versus 3D software visualizations," In: 2014 IEEE VIS International Workshop on 3Dvis (3Dvis), IEEE, 2014, pp. 33-36.