Abstract—Visualization of big data using mixed reality is challenging and promising at the same time. In this work, we present a comprehensive overview of the existing work in this area with a focus on several aspects of visual analytics. Even if the methods used in different scientific papers and projects are similar, the approaches and results differ very often, due to specific goals of the visual representations. Using our work the professionals can better understand the reality-virtuality continuum and choose an appropriate approach more easily in order to solve their specific big data analytical visualization problem.

Index Terms—Augmented Reality, Big Data, Mixed Reality, Virtual Reality, Visual Analytics, Visualization, Visual Interaction

I. INTRODUCTION

Visualization has always been an important part of data exploration. Every analytical procedure requires some illustration of concepts, techniques, results, etc. This includes primarily presentation of mostly abstract data or workflows for data processing in visual form for easy understanding. Visualization went through many phases over time and has become more and more diverse due to new types of data and manifold data processing techniques. Starting with cave paintings and maps, visualization evolved into simulations and 3D representations. Forms, types and displays changed due to a better understanding of human seeing and perception. A human eye possesses approx. 70% of all human sensory receptors, and the data is passed to the brain with the highest bandwidth. Our brain is able to process visual information in parallel using large arrays of neurons, extracting features from every part of the visual field simultaneously. Treisman described in [1] the result of brain processing as a set of feature maps. This parallel processing proceeds whether we like it or not and is mostly independent of what we choose to pay attention to (although not where we look). It is also rapid. Colin Ware indicates in his book [2] that if we want people to understand information quickly, we should present it in such a way that it can be easily detected by these large, fast computational systems in the brain.

The process of thinking or exploring that accompanies human conscious perception, cannot be done entirely inside peoples heads [3]. Mostly, an interaction with cognitive tools is necessary to support the process. These tools can be of different kinds, e.g., pencils and paper, calculators, or computer-based tools and information systems. Computer tools alone are insufficient, human interaction is highly important. Especially, visual analytics strongly relies on effective interaction of a human and a machine. As we described in [4], the Human in the Loop (HiL) concept is characteristic for the continuous support of machine processing by human feedback. In the context of Visual Analytics, HiL stands for providing continuous feedback, correcting algorithmic approaches and selecting appropriate techniques during the analytical process.

A. Visualization and Visual Analytics

Visualization and Visual analytics are both dealing with visual representation. Their scope, application and impact are, however, different. Visualization provides techniques for presentation of data or relationships for the purpose of explanation, interpretation, communication etc. Visual analytics encompasses a process of knowledge discovery by supporting the analyst to discover patterns in data, building formal models that can be processed by machines, and developing new hypotheses. This domain is often defined as an interdisciplinary approach to support exploratory knowledge discovery especially regarding large and complex data sets [5], [6], [7].

Visual Analytics focuses on the whole analytical process and is not limited to visualization and automated analysis. It also includes the entire infrastructure for creating visual analytics tools. Processing power and capacity of existing technologies allow to implement visual analytic techniques to huge amounts of heterogeneous and dynamic data (Big Data), where visualization cannot be used. Several visual analytics tools were developed in the last years, e.g., [8], [9] or [10]. The purpose and effectiveness of these tools varies depending on utilization scenarios, provided visualization techniques and user knowledge.

B. Visual Analytics with Mixed Reality

The application area of Visual Analytics can be roughly divided into two parts [11] scientific area and information representation for special purposes. The mixed reality techniques are already widely used in the scientific area [12]. However, the development of such techniques for information representation in general remains challenging. Due to the complex concepts behind mixed reality, it is more appropriate
to implement them to domains with highly dimensional and unstructured data where other techniques cannot be used.

The idea of using virtual reality for visualization of huge datasets is not new. Steve Bryson pointed out already in 1996 the challenges, possibilities and opportunities of such approaches [13]. Over time, the technical requirements and system capabilities changed enormously, and new concepts, such as Be the Data [14], emerged in this field.

Instead of two-dimensional objects, three-dimensional objects are used that are recognized by a camera and annotated with comments in augmented and virtual reality.

James A. Walsh and Bruce H. Thomas [20] developed a portable AR system for the visualization of real-time data from various sensor data. Their approach allows the visualization of large amounts of data, but it does not support multidimensional analyses.

### III. VISUAL ANALYTICS FOR BIG DATA

Humans use their cognitive perception and visual intelligence to generate meaning from data from their surroundings [21]. Various visualization methods are available for this purpose, for example in the form of diagrams. Interactive technologies not only make it possible to visualize data, but at the same time, for example, enlarge or reduce selected areas, or perform other manipulations. This means that Visual Analytics is not a specific tool, but an agile process in which the focus is not only on technology but more on human interaction [22].

The Institute for Visual Analytics in Vancouver [51] explains that the process starts with humans, who first have to learn to understand the context and the data. Only after the data has been cleaned and pre-processed, the visualized presentation of the data can be used to derive information and get insights. At the same time, the understanding of the context increases, which means that Visual Analytics cannot be defined as a sequential, but as an iterative process.

#### A. Challenges

Big data is often described by the three original Vs: *volume*, *velocity* and *variety*. In the meantime, new V-terms have been added, so that there exist now up to 10 Vs, see [23] and [24]. Often, however, it is only *veracity* and *value* that are added, depending on the purpose and the nature of the data. With the better ability to handle large amounts of data, on the technical and human side, big data is getting constantly bigger and the requirements are more and more increasing [25].

The challenges for the tools are constantly renewed by the constant digitization of the world, as new types of data and larger amounts are added. It is already possible to prepare data to a certain level (semi)-automatically. Many of these tools are already adapted for big data to ensure better analysis support [26].

Even if these tools can now cope with high-dimensional data sets, if the data also contains spatial and temporal dimensions, the preparation becomes much more complicated again. These dimensions have the property that in many cases they duplicate the data records for every point in time and space. The number of records increases by several dimensions. The complexity of the data grows accordingly and increases the effort to find suitable visualizations and to display all relevant dimensions. This often lead to "clutter" and to visualizations with too much information at the same time, which leads to a higher cognitive load for the user and thus to poorer results.

Comparing and recognizing trends and patterns in big data becomes then a problem [27]. This makes interactions, e.g.
selection and zoom, a necessity to be able to view the different facets of big data and to be able to choose the visualized dimensions dynamically [23].

B. Opportunities

Visualizations that show as much information as possible at once mostly have a problem with data that is both spatial and temporal, because these dimensions are difficult to dispense with. This limits the flexibility to select parameters of the visualization. Mixed reality is a possibility to address these issues. It is particularly helpful when displaying spatial data. At MIT e.g., tweets were displayed on the campus in a scanned campus environment [28].

The ability to realize more complex animations is also a plus. A perfect example of this is the Reddit Place [29]. A total of 16.5 million pixels were edited on a white 1000 x 1000 pixel canvas in 72 hours (Fig. 2), whereby every Reddit user could make a change every 5 minutes [30]. It would be difficult to place all possible information from the experiment in static images. Elements would also be lost on the desktops in 2D or 3D, since interactions are more complicated and 3D elements cannot be perceived in details. The individual information can only be presented and understood through meaningful interaction.

Greg Bahm created a visualization for VR [31] that shows some of the potential very well. It contains a mixture of animation, spatial representation, diverse interaction and suitable visualization of useful parameters. Class relationships can also be recognized well in VR models [32], since another dimension is available for distribution and 3D interactions are possible. The visualization can thus be viewed from all sides. Thus, VR brings new possibilities, but also challenges, whereby the visualizations have to be reconsidered for the proper presentations [33].

![Fig. 2. Reddit Place visualization [31].](image)

C. Spatial Presentation

Stereoscopy as a part of virtual and mixed reality allows to visualize three-dimensional data without dimension reduction. 3D displays usually use the binocular parallax cue (disparity) to create an impression of depth. The representation of depth is in this case a simulation of the depth perception process through our visual system.

There are various display technologies available on the market or developed in different scientific projects. A good overview is given in [36] and [37]. Available display technologies can be divided in stereoscopic and auto-stereoscopic displays. Stereoscopic displays require special glasses to separate the views for the left and right eye. The auto-stereoscopic displays perform the view separation on their own and can be used without glasses. Most often, the separation is done by a special lens system (Alioscopy, Dimenco) or through parallax barrier (Nintendo 3DS). The multiview auto-stereoscopic displays are able to present simultaneously multiple views, thus increasing the freedom of movement and creating more voluminous effect that greatly improves the 3D experience. E.g., headtracked auto- stereoscopic displays are multiview displays combined with a head tracking technology. Based on head position, the observer is presented with a different point of view allowing great freedom of movement, e.g., Fraunhofer HHI’s Free2C display technology [38].

The spatial cognition works better this way and spatial relationships are easier to recognize [39]. Richardson et al [40] showed in 1999 an example using maps and virtual environment (VE). Relationships, such as distances, can be better estimated in VE. However, maps also have advantages, e.g., with more precise estimations over several floors.

Virtual reality enhances visual representations by adding an extra dimension. Nevertheless, the third dimension is not a universal remedy, since it only adds another feature to the representation [41].

However, the advantage is visible in the Reddit Place visualization (Fig. 2). Using the third dimension for selectable parameters, e.g., a heatmap about the number of changes, lifetime, or color changes over time, attention is drawn to the points of interest.

The stereoscopic representation contributes also to a significantly better perception. Another example is visualization of insect trajectories, see Figure 5 [35]. The time axis for movement paths of insects serves here as an additional dimension. This allows an easy visual perception of pauses and speeds. One of the early approaches [34] shows the usage of a three-dimensional Treemap for visualization of the Unix file system (Fig. 3). Due to transparency of the individual cubes and their hierarchy, it is possible to get a simple overview of the system structure.

![Fig. 3. Visualization of the Unix File System [34].](image)
An important aspect here is whether the visualization is helpful and simplifies the work, at the same time meeting requirements varying from case to case. In [28] the developed visualization was helpful in order to be able to get an overview of the geographic location and averaging tweets on the MIT campus (Fig. 4).

The authors of [42] compare two- and three-dimensional visualizations. They support the statement that the increased load generated with mixed reality is justifiable due to perception simplicity and comprehensibility of data representation.

![Fig. 4. Visualization of tweets on the MIT campus [28].](image)

**D. Interaction in Higher Dimensions**

Interactions within visualizations of big data are essentially important for understanding of highly dimensional and dynamic data. Mixed reality technologies support a variety of interactions with the represented data. Simulations or animations can be seen in the same context [43]. Scaling and zooming are some of the simple examples of interactions with data. Mixed reality allows to perform also natural interaction with the visualization using gestures and making it possible to walk around the data [44]. The authors of [45] describe 5 patterns for possible interaction with mixed reality visualizations:

- Selection
- Manipulation
- Viewpoint control
- Indirect control
- Compound.

All of the options described are useful when considering data. Many interactions are, however, computationally expensive and can affect the effectiveness of the visualizations. Users should consider very carefully, which interactions can be most useful in a specific use case.

The visualization of insect trajectories (Fig. 5) supports the possibility of visual queries among other interactions with the visualization.

Interactions and manipulation of data in the analytical process is characteristic for the Human in the Loop approach. The main concept is the continuous support of machine processing by human feedback. In the context of visual analytics, this concept occurs in terms of providing continuous feedback and correcting algorithmic approaches within the analysis.

![Fig. 5. Visualization of insect trajectories with visual queries [35].](image)

**IV. DISCUSSION AND CONCLUSION**

In this paper we present an overview of mixed reality approaches and examples and explain how they can be used for further analytical processing.

The provided overview and the presented examples show that mixed reality offers additional value for data visualizations. We found out that the main advantage is better perception of data, due to stereoscopic representation. Spatial relationships can also be better understood in such visualizations.

After mastering the challenges, there are several options for using mixed reality (as well augmented and virtual). The form of visualization of big data can be optimized in such a way that multidimensional data can be presented in a more understandable way [46].

One of the examples is the Google Earth VR application, which enables a user to travel in virtual space to any place on Earth. In addition to the locations, a user recognizes distances and height differences and experiences them in a more memorable way than with a two-dimensional display [47].

Furthermore, the combination of big data with augmented reality in the form of assistants for navigation can simplify our everyday life in the future. For example, current data can be extracted from official sources as well as from social networks in order to derive load factors for roads or public transport or to make predictions in order to show the fastest route via an AR / VR interface [48].

Another important example of visualizing big data with mixed reality is to display data for applications in medicine. For example, the virtualized structures could be displayed in real time during an operation thus supporting the surgeons [49]. The combination of the mentioned technologies results also in new learning methods, which lead to a higher mental performance [50], as well as a more efficient adaptation of the cognitive load.

Our work can serve to better understand the reality-virtuality continuum and helps professionals to choose an appropriate approach to solve their specific big data analytical visualization challenge.

Although, we did our research based on a big amount of...
scientific and practical approaches, the speed with which the technologies in this field are developing is enormous. To describe all possible relevant approaches in detail, this publication would not be sufficient by far. In our future work we intend to compare some of the newer approaches more detailed and present our results to the scientific community.

REFERENCES

[41] C. Chen and T. Hsu. Game development data analysis visualized with virtual reality. In 2018 IEEE International Conference on Applied System Invention (ICASI), pages 682685, April 2018


