Agents of MASK

Mobile Analytics from Situated Knowledge

Neven ElSayed, Know Center, Graz
Eduardo Veas, Graz University of Technology
Dieter Schmalstieg, University of Stuttgart

With the emergence of digital sensor platforms and infrastructures, we are seeing a massive increase in the amount of data being gathered from objects, processes, and spaces. For example, social media interactions, data collected from IoT devices, financial transactions, transportation-related data, governmental records, and scientific datasets all make substantial contributions to the expanding pool of data sources.

To be made available for personalized, effective, timely, and responsive services, this growing volume of data needs to be fielded, sorted, and rendered accessible, anytime and anywhere. This will allow users to interact with adaptive physical objects, spaces, and their associated data [1]. Recently, *immersive* [2] and *situated analytics* [3] have been introduced to expand the visual analytical space using virtual and augmented reality technologies. These technologies build analytics on the user's capacity to move between different environments and scenarios in the physical world. However, a critical challenge arises when deploying analytics in linked query sessions that are connected to different virtual or physical worlds. In these cases, users need to switch between dynamic display spaces, various interaction modalities, and fragmented analytics sessions.

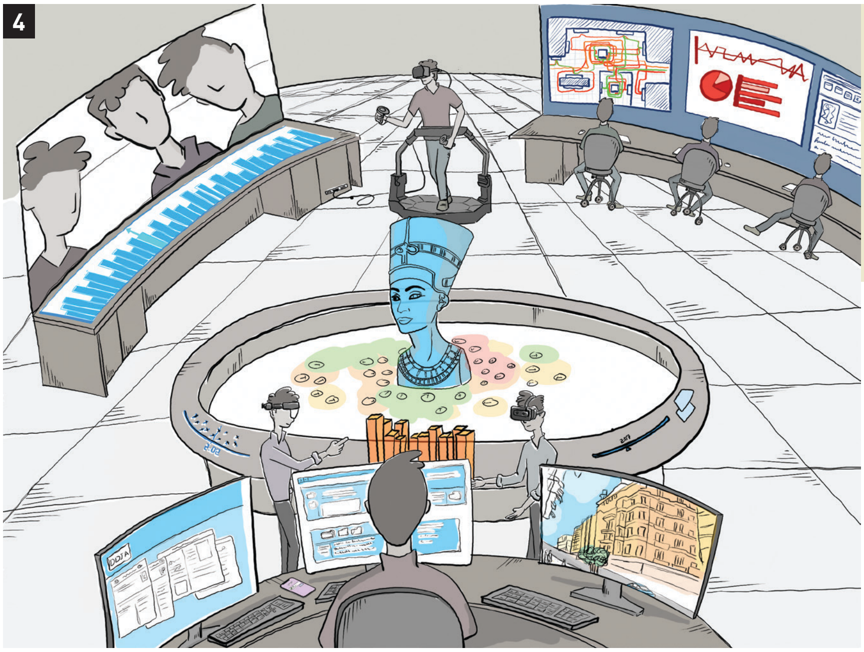
The emerging field of mobile analytics advocates the immediacy of adapted situated analytics based on knowledge of the surrounding environment. To showcase the challenges (C1–C6 below) and research directions (RD1–RD6 below), we present an illustrated use case for immersive and situated analytics, highlighting the technology limitations that prevent the deployment of analytics in different environments. Examples that we illustrate include: 1) illumination, 2) clutter, 3) static vs. dynamics of multiple objects in motion, 4) close-up views, 5) wide-open views, and 6) first-person motion. These aspects make it challenging to deploy arbitrary mixed reality techniques in real-life connected scenarios.

Our illustration, which takes the form of a speculative design scenario, presents a vision that merges *situated analytics* with *artificial intelligence* and *behavior analysis* to empower analytics *on the go*. Mobile analytics supports instant data analysis, while augmented reality seamlessly blends information into the surrounding physical environment. We illustrate the techniques and methods that, along with AI, are needed to offer a fully immersive experience, providing a cohesive, personalized presentation for *mobile analytics*.

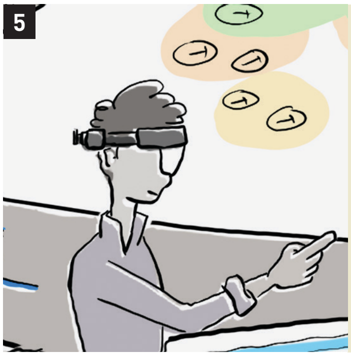
THE STORY OF A MASK

It all begins when an important piece is stolen from a museum. Detectives need to solve the case quickly and accurately, as any incorrect analysis path could lead to the piece being permanently lost. It's a typical visual analytics case, where detectives need to analyze different datasets using diverse models (AI engines), supported by visual representations [4].

The case is assigned to the Digital Detective Intelligence Agency (DDIA). The team gathers in the main control room, equipped with the most advanced *immersive analytics* tools.



The team uses the Revealing Pool, a table-like device for real-time holograms, much like a workbench or showcase. The holograms reveal the digital twin of the missing piece—an almost physical but ghostly object. Instantly, the space around the object illuminates with crowd data, illustrating patterns of motion at different intervals of time. The pool's surface appears liquid, but it is actually tangible, made of innumerable particles that rise and sink or change colors according to the data.



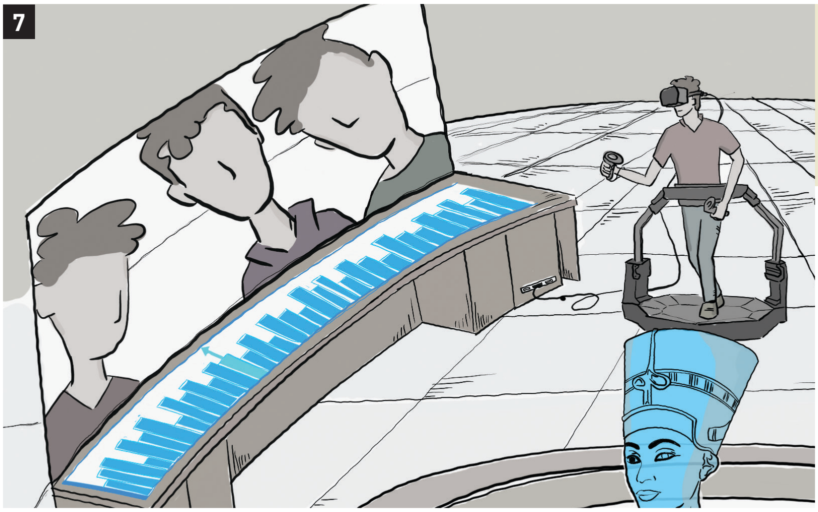
An agent touches the surface with a finger, drawing a shape that encircles several data points, then grabs and drops them in midair between him and his collaborator.

The agent analyzes the crowd's dispersal in the museum and uses a collaborative mixed reality system for fluid transitions between augmented reality (AR) and virtual reality (VR) modes. Analysts begin creating collaborative 3D and 2D visualizations in a shared workspace.

Two experts are asked to join the team—a crime investigator and an Egyptologist—to work together and solve the case. The crime investigator observes the Egyptologist's actions, focus, and gestures. As the collaboration progresses, the workspace provides cues to establish a shared understanding of the team's analysis and actions. The two experts start working in a shared view and create individual private views to evaluate some preliminary analyses.

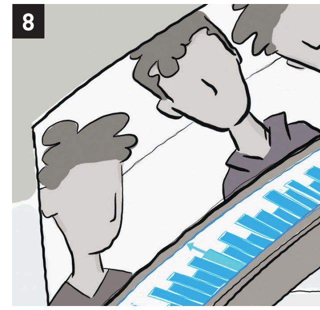
ILLUSTRATIONS BY NEVEN ELSAYED

7



The DDIA's Ghost without Shell (Gwish) is a remote agent who uses VR to navigate in "first person" mode through the museum's digital twin, generated with a dynamic neural radiance field. The Gwish observes museum visitors from the statue's perspective, analyzing their facial expressions for emotions such as nervousness or excessive seriousness.

8

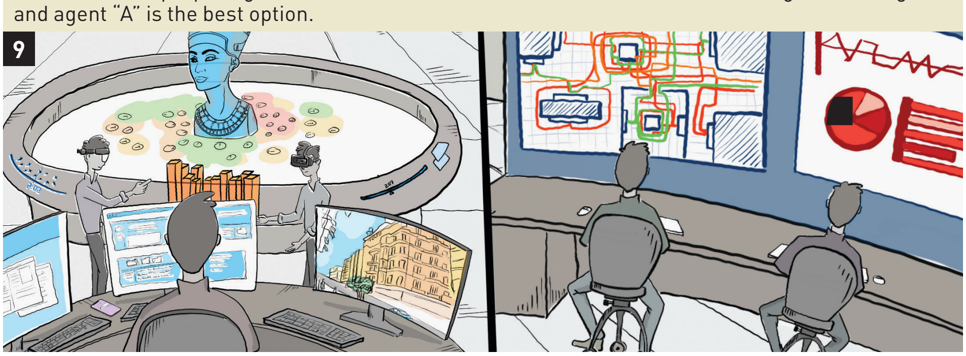


The interface allows the Gwish to navigate forward and backward through time. With the advantage of infinite locomotion, the Gwish can explore the museum like an ordinary visitor at the time of the crime.

The Gwish interleaves activities in connection with the analysts' queries. The potential suspects are highlighted on a large screen, while desktop analytics display facial-expression heatmaps in conjunction with locomotion patterns. Concurrently, trajectories and other demographic statistics of the subjects appear in an immersive multiuser tiled display. Analysts go through the spatiotemporal data combined with demographics, pulling visitors' data and routes.

The head of DDIA monitors various feeds in an old-fashioned multi-display analytics setup. With the list of suspects narrowed down, a list of locations, objects, and museum visitors emerges. Just as the media is preparing to make an announcement, the decision is made to bring in a field agent, and agent "A" is the best option.

9



10

So as not to isolate agent A from DDIA's control room, he is given the MASK (Mobile Analytics from Situated Knowledge). The MASK is built for AR mobile analytics, considering physical-world challenges such as:

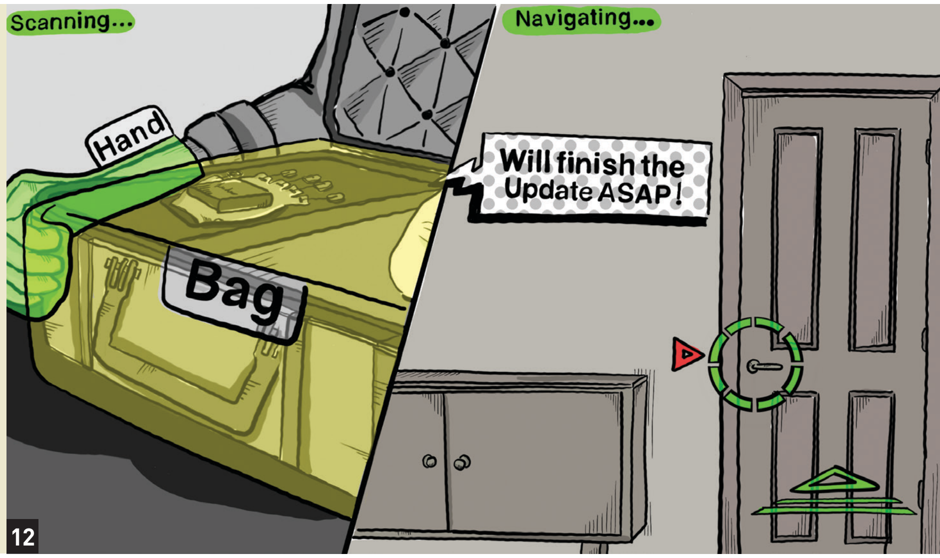
- Tracking objects of interest among multiple moving targets
- Multimodal clutter of the environment
- Depth perception of augmentations
- Visual coherence (e.g., color and illumination)
- The graphical widget relationship with physical objects
- Occlusion of the real world with augmentation
- Attention to and distraction from augmentations.

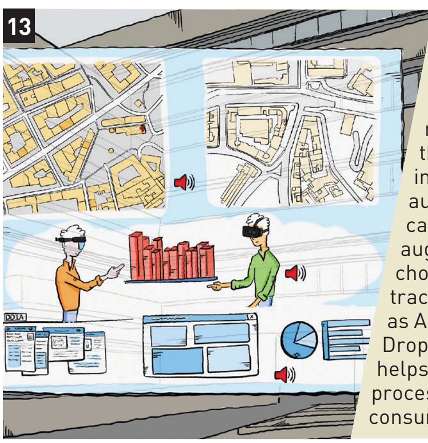
11

Agent A picks up the MASK, a video see-through display supporting AR and VR modes. It also includes multiple sensors for field data collection.

Agent A puts on and then calibrates the MASK. The MASK has eye tracking and connects to haptic gloves. The eye-tracking controls the presented content based on A's performance and is used for gaze interaction. The haptic gloves provide vibration, thermal, and kinesthetic feedback and track A's hands. As the actuators and sensors drain the gloves' power, the actuators are disabled and can be activated manually if needed.

The MASK supports data analyses in immersive environments: virtual reality (immersive analytics) [2] and augmented reality (situated analytics) [3]. With a hand gesture, A initiates the mission on the MASK. The accurate tracking supported by the gloves enables A to carry out the interactions in secret. DDIA asks A to collect data on two suspicious visitors from the field that cannot be collected digitally.

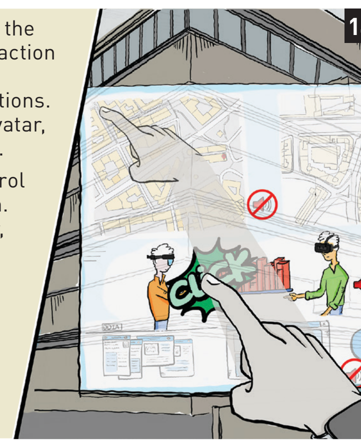


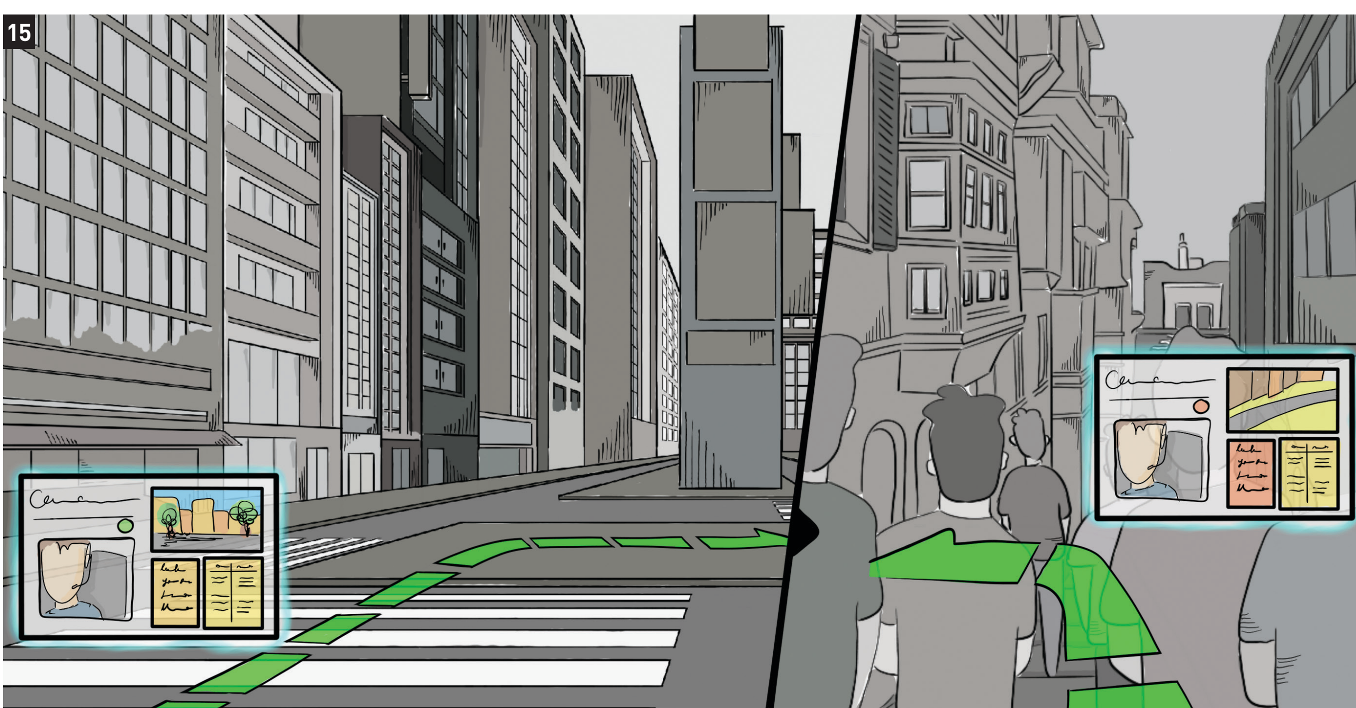


Once A leaves the research lab, the MASK invokes the outdoor tracking mode, presenting the task's information on an augmented 2D canvas. The augmented window is chosen to reduce the tracking challenges, as A will be walking. Drop tracking further helps reduce the processing consumption.

Agent A uses midair interaction with the presented content. To keep the interaction inconspicuous, A can reach far-off sections with gaze and glove interactions. The MASK supports A with a hand avatar, enhancing the near-field interaction.

Audio streaming from the DDIA control room uses spatial sound localization. DDIA headquarters is busy, however, which produces a cluttered audio environment and disrupts the automatic sound localization. A opens and controls the audio manually, switching between audio sources to communicate with different teams in the control room.



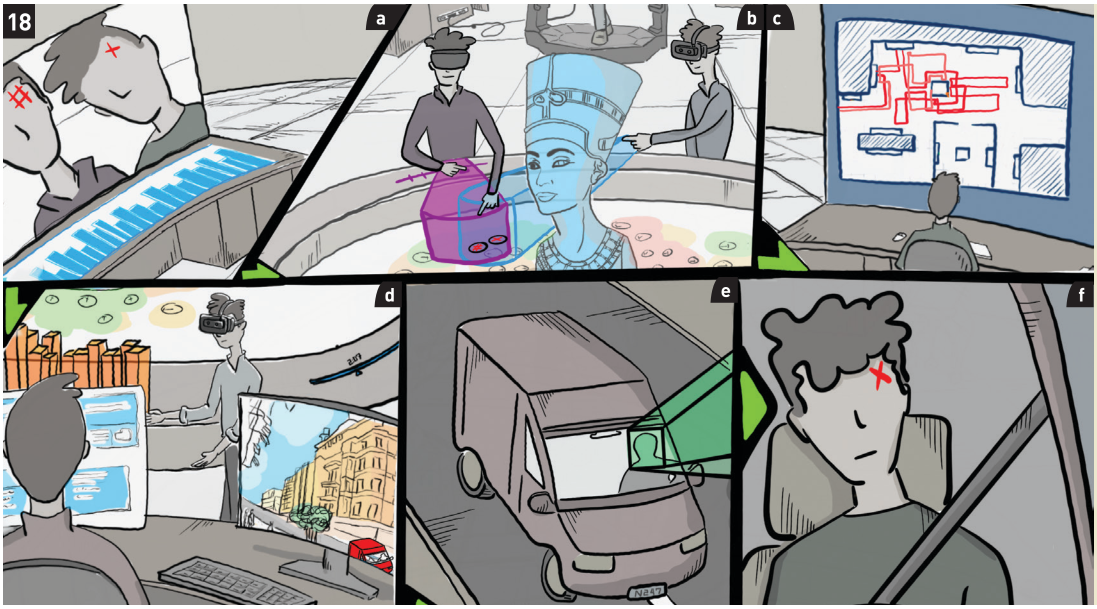


Agent A is guided by the MASK through the city with virtual cues while collaborating with DDIA HQ. When A arrives at a busy street, however, the navigation arrows start to become cluttered with the surroundings, occluding the physical scene. Moreover, multiple objects moving in the dynamic environment as well as A's own movements start to confuse the alignment and contrast of the situated visualization. The MASK uses a context-aware technique to update the appearance and location of the annotations; however, as A and the scene are rapidly changing, the adaptation frequency is very high.

Agent A is tasked with checking for traces of phosphorescent paint on two suspicious individuals (S1 and S2), who have been selected based on the analysis session in the control room.

Agent A finds S1 while the suspect is having coffee. A scans for a trace of phosphorescent paint on S1 using the MASK's built-in ultraviolet (UV) camera, but no trace is found.

When A locates S2, however, he finds a trace of the paint on the suspect's hand and clothing. The computer vision algorithm in the MASK recognizes S2 and enables the analysis mode. In this mode, A can filter the data in the DDIA HQ remotely. By selecting S2 in the physical world, he is identified as a major suspect. This situated analytics brushing action updates the data in the DDIA HQ. A massive cascade of smart filters takes place at all stations.



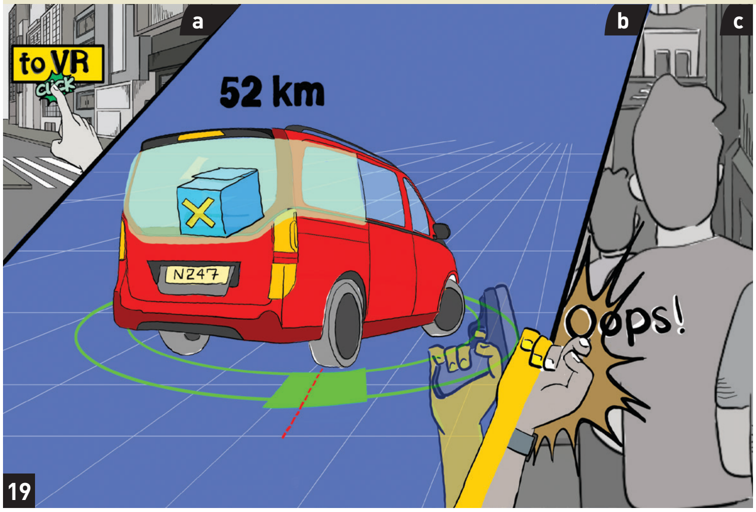
The museum video footage is then trimmed to include only the shots with S2. The DDIA agent at the Revealing Pool marks a third suspect, S3, who is in close proximity to S2 most of the time. Selecting S3 filters the spatiotemporal data in front of the analysts as well as the trajectory data on the wall display.

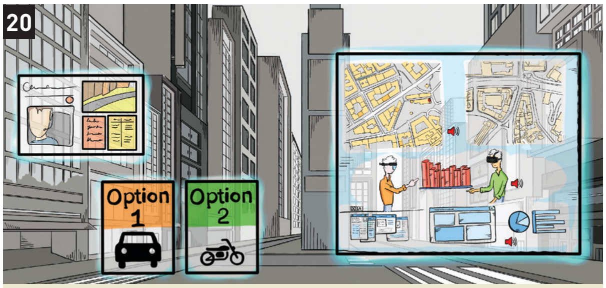
Real-time videos of the traffic-light CCTV cameras are processed, and S3 is found stopped at a light. The team processes the predicted trajectories between S3 and S2, estimating the path of S3's car, which probably contains the stolen piece.

The vehicle type is recognized in CCTV frames. The team feeds the vehicle's parameters and physics to a 3D simulator to estimate the vehicle's load and position of the stolen piece in the vehicle. Simultaneously, AI algorithms are running analyses using mesh deformation models to figure out how to stop the car with the least impact on it and the stolen piece inside.

The predictive simulation calculates the optimal hitting point, angle, and car speed. The car has to be shot in the rear right tire, at a 37-degree angle, with the car going 52 km/h. The mission of pulling the car over is assigned to A, who is the one nearest to S2's location. However, agent A needs to be trained on the correct shooting angle, as any mistake could cost DDIA the stolen object. A receives a call to join a VR training session through the MASK with a DDIA expert advising how to stop the car safely. A accepts the training and starts the VR session. The MASK enables A to visualize and interact with the car's digital twin.

The VR training uses hand gestures and haptics to increase engagement. While aiming and shooting the car tire in the virtual training, isolated from the real scene, A accidentally bumps into a pedestrian walking in the street, which momentarily disconnects him from the VR training. A finishes the training, and the mission changes from exploratory analysis to operational analysis.





Agent A is given two options to chase the suspicious car, by car or by motorbike. The MASK presents the impact of each option on time; however, the massive amount of presented information clutters the physical world. Such clutter representation usually uses interactive visualization, clustering the data into an overview and detailed on-demand view. That is challenging, however, when A is walking, especially in a dynamic and cluttered environment.

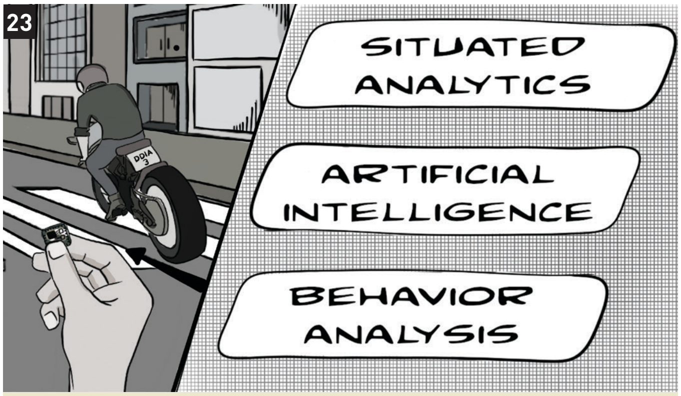
21 So far, Agent A has had to put up with auditory clutter, visual clutter, and self-management of interaction with multiple remote parties. Both the real world and the virtual world are competing for A's attention.

With the existing features in the MASK, A is provided with interactive visualization and connected to the DDIA room. However, there is a list of challenges, including:

22

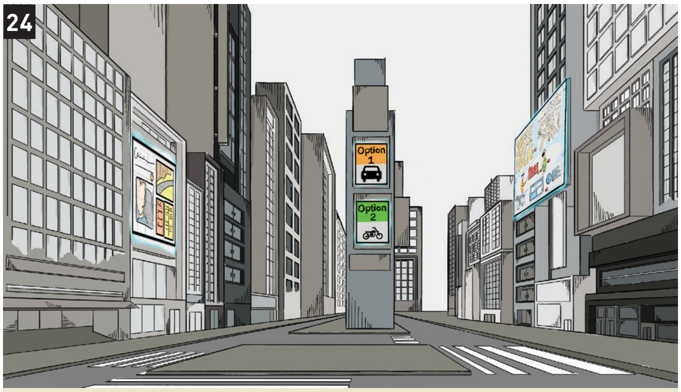
SITUATION	DESCRIPTION	CHALLENGES
A opens and controls the audio manually, keeping the manual switch between audio sources to communicate with different teams in the control room.	**C1: Manual adaptation:** User required to adapt interactive dashboards' settings in limited physical interaction space.	Manual controls, limited dexterity, and out-of-reach controllers
The MASK uses a context-aware technique to update the appearance and location of the annotations. However, as A and the scene are rapidly changing, the adaptation frequency is very high.	**C2: Occlusion:** The presented information was occluding physical-world objects, which were moving faster than the system could adapt to.	Information overload, adaptation speed, latency, and dynamic canvases
A is tasked with checking for traces of phosphorescent paint on two suspicious individuals (S1 and S2), who have been selected based on the analysis session in the control room.	**C3: Disconnected sessions:** While the user scanned the physical world, the streamed visualization was disabled, helping them focus on the surroundings. However, this action disconnected the user temporarily from the remote collaboration.	Limited view space and asynchronous remote collaboration
The team processes the predicted trajectories between S3 and S2, estimating the path of S3's car, which probably contains the stolen piece.	**C4: View representation:** All analysts, including those with different backgrounds such as Egyptologists and detectives, were exposed to the sam e visual representation, which influenced their behavior during the collaborative analytics session.	Adaptive visual representation and behavior awareness
A accidentally bumps into a pedestrian walking in the street, which momentarily disconnects him from the VR training.	**C5: Isolated view:** The user was isolated in VR mode with no awareness of the surrounding dynamic physical environment.	Peripheral awareness and view's dynamic transition
The massive amount of presented information clutters the physical world.	**C6: Clutter view:** The flood of information added to the physical surroundings created a cluttered view that distracted the user.	Information overload, density of presentation, and augmentation layout

23



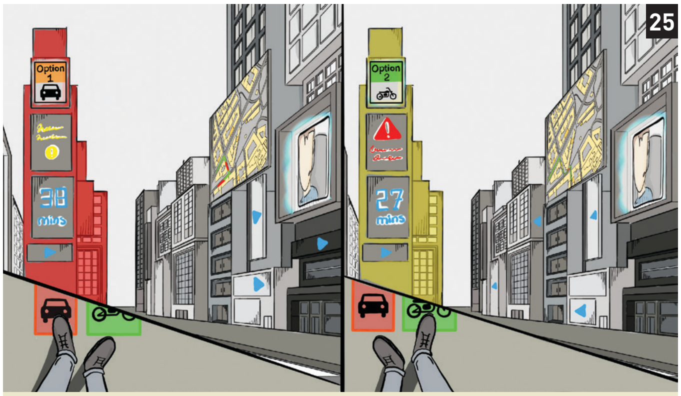
Agent A's trip is easier using the MASK; however, some of the techniques are not effective *in the wild*. So, A goes to a rendezvous point and receives an update for the MASK's software. This update enhances the AI's situated mobile analytics with a dynamic analytical interface based on user behavior, the situation, and surrounding knowledge, including the physical world's features and display mode.

24

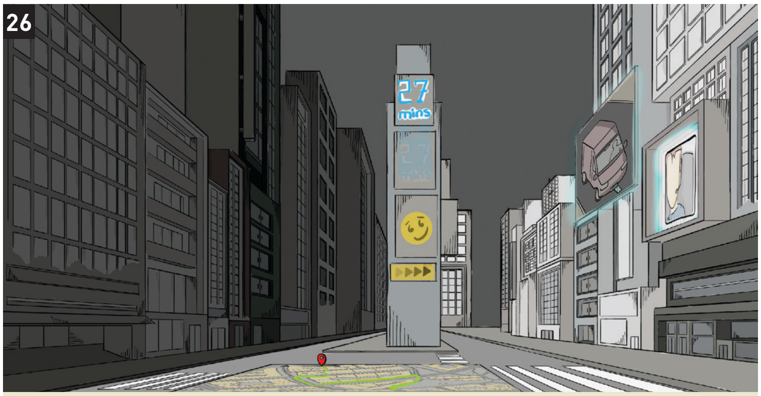


Using the update, the visualization blends into the physical world, reducing the scene clutter. The visualization aligns and scales based on the task and is allocated to the scene based on its expected purpose. For instance, important data presents as fixed-location abstract data, while complex graphs present as situated visualizations, which can be scaled to enhance visual perception.

25

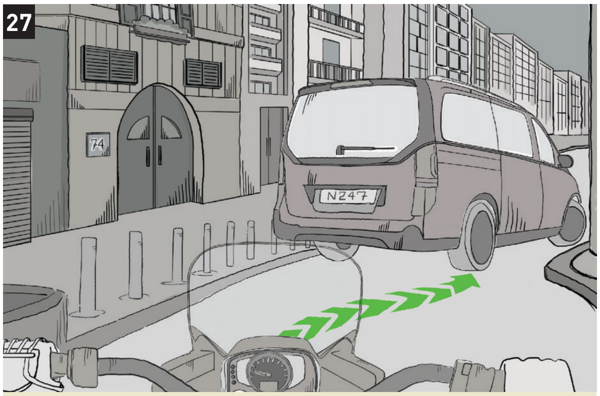


When Agent A arrives at the location of S3's car, the MASK switches to its secret interaction mode, changing from hand gestures to embodied interaction, keeping A inconspicuous. A analyzes two choices that were calculated remotely in the control room, selecting, filtering, and analyzing data nodes or selecting and filtering in real time using foot interaction. A chooses option 1, viewing the data's details on demand. The detailed view appears as canvases blended into the surroundings; the overview is presented as blended color codes. A considers the car option, which is faster and more secure for chasing another car. However, based on A's field observation, presented traffic map, and VR training, it would be difficult to obtain the required shooting angle using the car. Therefore, he chooses the motorbike.



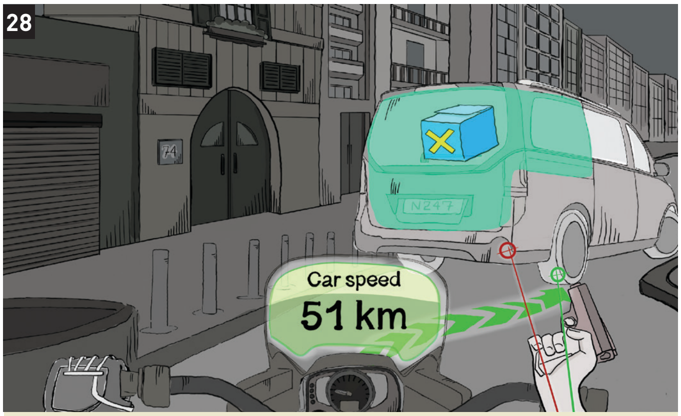
A selects the transportation option using embodied interaction by stepping on a "bike option" virtual button augmented on the ground beneath him. The MASK reverts from analysis mode to execution mode. In execution mode, the MASK allocates the canvases on the sides of buildings, reducing clutter.

The abstract data is presented on the street using a mediated reality technique, dimming the least important parts of the physical world. A follows the augmented cues, while the MASK keeps measuring A's biosignals. Sensor data is used to adapt the task sequences. The built-in eye tracking used to adapt the visualization is based on A's gaze movement. To measure his performance accurately, A's signal measurement and gaze interaction are sent to the DDIA control room, which assesses him throughout the task.

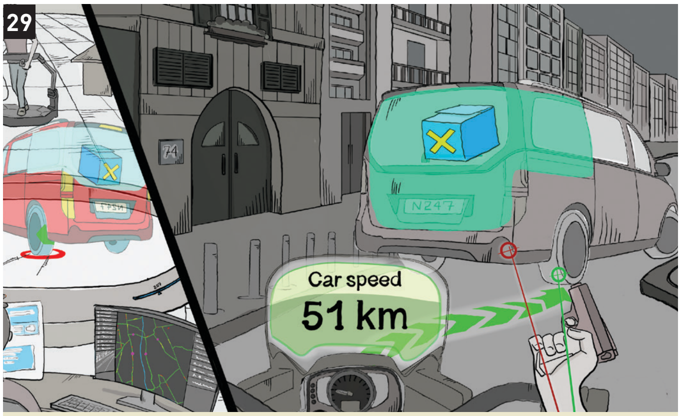


Once Agent A reaches the motorbike, the MASK switches to the rapid interactive visualization mode. The MASK enables only one selection option, which can be controlled by eye gaze and glove clicks, with blended information representation. A can select the object in front of him to explore the attached information.

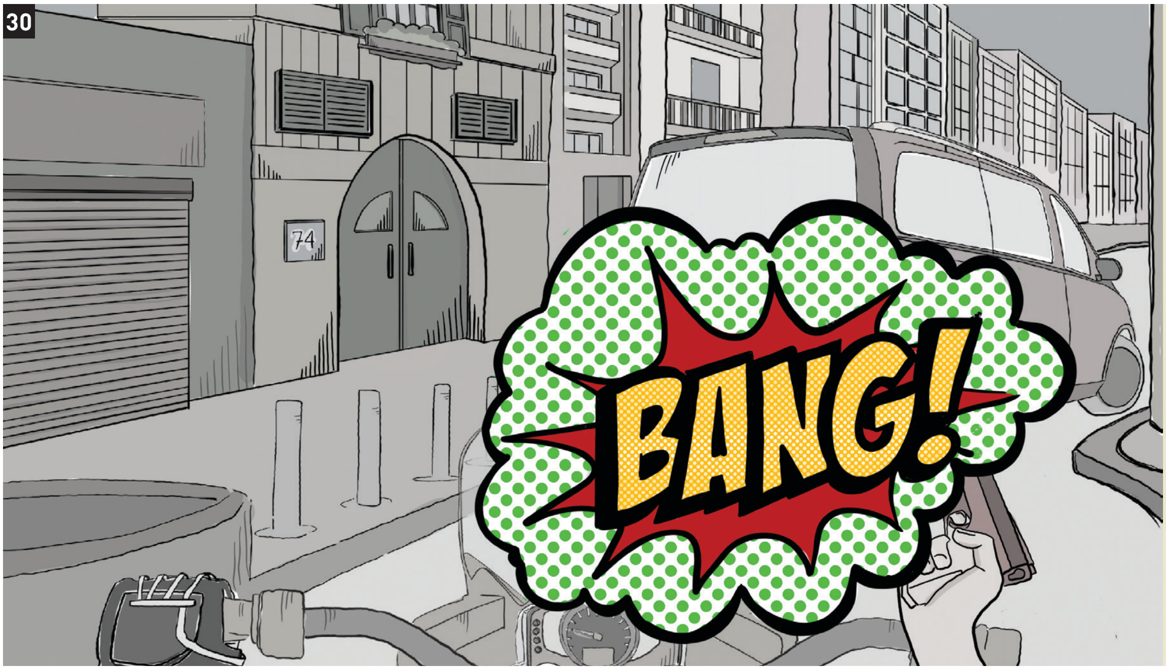
A's bike captures the car's speed using built-in sensors and the car's orientation using the MASK's camera. The captured data is then streamed to DDIA to be processed and to calculate the vehicle load's status using a physics engine, fast GPUs, and distributed rendering.



Based on the physics-engine calculation, the MASK augments the optimal shooting trajectory as lines situated to A's hand, to reduce the rendering cost.



In the meantime, one of the DDIA team members is manipulating traffic to slow down the suspicious car. A is streaming the live scene using the MASK's camera, combined with eye tracking.



This augmented visualization is updated based on the data collected from the physical world and simulation engines. A aims at the car's right rear tire, using real-time mobile analytics blended with visualization, while the DDIA experts provide audio guidance. Once A is confident, he aims and shoots the tire—*Bang!*—at the correct time and position. He successfully stops the car and accomplishes the task accurately using *mobile situated analytics*.

The presented storyline provides a comprehensive look at the combination of mixed reality and visual analytics for mobile analytics, emphasizing the immediacy requirements for information on the go. We advocate the need for a fully immersive experience that considers both the real-world situation and the abstract information, with artificial intelligence generating a cohesive personalized presentation of real and virtual worlds. Various paradigms for the access and analysis of digital information have been put forth: visualization [5], visual analytics [4], augmented reality [1], immersive analytics [2], situated analytics [3], and embedded data representations [6]. Mobile analytics implies an on-the-go analysis, while augmented reality seamlessly blends information into the surrounding physical environment. We draw from the body of literature the techniques and methods needed to offer a fully immersive experience, with AI generating a cohesive, personalized presentation of real and virtual. The storyline highlights the existing visualization, interaction, and analysis for mobile analytics, and its potential research directions (RD), which can be summarized as follows:

RD1. Adaptive UI blended interface based on user behavior: by blending the augmented visualization into the physical environment, enabling the system to manipulate the view based on situated knowledge (e.g., object tracking, depth, and clutter factor).

RD2. Adaptive embodied interaction: by dynamically changing its interaction mode based on the user's behavior and surrounding knowledge.

RD3. Blended visualization for perception enhancement: by employing scene-manipulation techniques to control users' attention. These models can be automatically adapted based on situated knowledge, including factors like clutter percentage, depth, and illumination.

RD4. Visual cues for dynamic situation augmentation: by utilizing predefined static models to adapt the situated visualization for fast animation, leveraging behavioral knowledge of the users' status.

RD5. Task-driven analytic interface: by adapting the situated visualization and adjusting the interaction tool based on predefined task goals.

RD6. Collaborative mobile analytics: by dynamically adapting the visual analytics session for each user independently, based on users' behavioral knowledge.

In this article we introduced, reviewed, and illustrated various parameters affecting mobile analytics *in the wild*. Through our illustrative scenario, we presented the challenges and requirements for deploying mobile analytics in real-life scenarios. We envision the impact of adapting visual (situated and immersive) analytics when considering *situated knowledge* and an *AI user interface*. Situated knowledge includes the physical environment's parameters, the analytics' tasks, and the user's behavioral data. AI engines use the collected parameters for generative models, calculating the augmentation settings and control models.

Endnotes

1. Kruijff, E., Swan, J.E., and Feiner, S. Perceptual issues in augmented reality revisited. *IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 2010, 3–12; https://doi.org/10.1109/ISMAR.2010.5643530
2. Chandler, T. et al. Immersive analytics. *Big Data Visual Analytics (BDVA)*. IEEE, 2015, 1–8; https://doi.org/10.1109/BDVA.2015.7314296
3. ElSayed, N.A.M.,Thomas, B.H., Marriott, K., Piantadosi, J., and Smith, R.T. Situated analytics: Demonstrating immersive analytical tools with augmented reality. *Journal of Visual Languages and Computing 36* (Oct. 2016), 13–23; https://doi.org/10.1016/j.jvlc.2016.07.006
4. Keim, D. Andrienko, G., Fekete, J-D., Gorg, C., Kohlhammer, J., and Melançon, G. Visual analytics: Definition, process, and challenges. *Lecture Notes in Computer Science 4950* (2008), 154–176.
5. Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. *Proc. of 1996 IEEE Symposium on Visual Languages*. IEEE, 1996, 336–343.
6. ElSayed, N.A.M., Smith, R.T., Marriott, K., and Thomas, B.H. Blended UI controls for situated analytics, *2016 Big Data Visual Analytics (BDVA)*. IEEE, 2016, 1–8; https://doi.org/10.1109/BDVA.2016.7787043

Neven ElSayed, a senior researcher at the Know Center in Graz, has specialized in augmented reality (AR) and visual analytics (VA) since 2010. ElSayed introduced "situated analytics" as a novel merging between AR and VA, earning a Ph.D. in computer science from the University of South Australia in 2017 and receiving the Michael Miller Medal for an outstanding thesis in the same year.
→ nelsayed@know-center.at

Eduardo Veas is a professor of intelligent and adaptive user interfaces at the Institute of Interactive Systems and Data Science at Graz University of Technology. He is also area manager of the Human-AI Interaction group at Know Center GmbH. He has a Ph.D. in computer science from Graz University of Technology and a master's in information science and technology from Osaka University in Japan.
→ eveas@tugraz.at

Dieter Schmalstieg is Alexander von Humboldt Professor of Visual Computing at the University of Stuttgart. His research interests are augmented reality, virtual reality, computer graphics, visualization, and human-computer interaction. He is a fellow of the IEEE and a member of the IEEE VGTC Virtual Reality Academy.
→ schmalstieg@visus.uni-stuttgart.de

DOI: 10.1145/3633521