Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 7 Issue 10, October - 2020

A Conceptual Framework For AR/VR Integration Into Our Every Day Lives

Ayse Kok Arslan

***Abstract*—Given the proliferation of AR (augmented reality)/VR (virtual reality) technologies in our everyday lives, it is no longer an illusion to live in a world where every part of a user's daily experience is augmented by the intelligent signals collected from the world around her. In that world, there is a crucial technology often in the form of a digital 'Assistant'- which is more useful than any assistant that has ever existed. This conceptual paper maps a primary way for how individuals can interact with particular AR/VR technologies to enhance their communication with colleagues and friends in such a way so that the world knowledge is proactively provided in a highly personalized manner. It provides the road map for the convergence of AR and VR technologies so that they become contextually intelligent. The main purpose of this study is to define the north star for this convergence: its mission, its key components, example use cases, and a roadmap.**

Introduction

Given the proliferation of AR (augmented reality)/VR (virtual reality) technologies in our every day lives, it is no longer an illusion to live in a world where every part of a user's daily experience is augmented by the intelligent signals collected from the world around her. In that world, there is a crucial technology often in the form of a digital 'Assistant'- which is more useful than any assistant that has ever existed.

Contextual intelligence — understanding of the circumstances and setting within which we experience things, along with our knowledge and memories — is what enables us to take in all of the sensory inputs around us and react to them intelligently in real time as we navigate the world.

Overview

In the quest to "Build the ultimate contextually intelligent assistant that gives people superpowers in a digitally augmented world," many tech companies try to envision an assistant-hosted experience.

Technology companies can personally reconcile the idea of a rich assistant experience with the requirement to not absorb entire apps into the assistant, by allotting themselves ownership over the perception and clarification of the request. In some cases, we might "call" an app and surface the result, or in some cases, we might launch an app with an attached payload. Yet, until technology companies reach that point, the user is in an augmented aka 'assistant' experience.

An intelligent agent that users can interact with seamlessly through voice and multimodal interactions, like they would with another human being." Although "clarifying parameters" does not sound lofty, we can view them as rich multimodal ways to understand what the user wants to do, and dispatch a request on their behalf.

Conceptual Framework

The mission of such a VR/AR technology can be stated as follows:

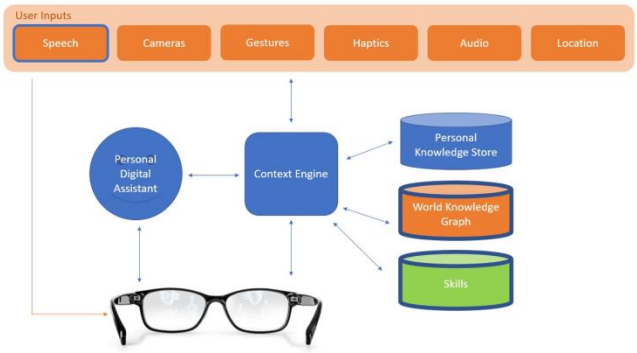
Build the ultimate contextually intelligent assistant that gives people superpowers in a digitally augmented world.

Assistant will be controlled mainly through the use of natural language, and will also work in the background to inform all parts of the AR experience, proactively mediating so the individual can focus on what matters in the moment.

Assistant will leverage understanding of the environment (audio and visual perception), and Social networking platforms' unique knowledge about the people, places, events and things that matter to an individual, and apply that understanding to make the AR/VR experience the *most* personalized, relevant and meaningful in the market. Without rich contextual intelligence and a proactive assistant to manage it, the AR/VR experience risks being noisy and often off the mark. With it, people will feel they have superpowers.

Assistant processes and stores highly personal knowledge of the person and her world, so as technology companies hold the person's privacy paramount, and understand, they must earn trust. Control of personal information and data must be clear and predictable. Integrity must be iron-clad, a user should be able to control their assistant experience and not be unpleasantly surprised with content, suggestions, or experiences.

The image below illustrates how we envision these pieces working in concert:



Assistant Platform & Runtime

The assistant technical stack can be separated to the online and offline components. The online components participate in the runtime serving path, and it is critical to get service reliability, latency, scalability, and security right. The offline components include tooling, data annotation, model training, end-to-end testing, etc. They are the enablers of the online components. Some online components run on device, while others will run in social networking platforms data centers. The offline components run primarily in social networking platforms data centers. We'll focus on the online components in this section, not because offline components are less important, but because the online components' behavior specification ultimately determines the requirements for the offline components.

Analytical Framework

Let’s look at the aspects of the AR/VR convergence on an analytical scale now:

On-device vs. cloud

For AR scenarios, battery life is the most important design consideration among all factors. Therefore, energy consumption is the most important optimization goal for on-device vs. cloud decision. Sending 1 byte and sending 10K bytes over the network consumes roughly the same amount of energy. There is a need to optimize firstly the duration to keep radio and network stack in deep sleep state, then secondly the transmission size. If some computation can be done on-device and it uses less energy compared with sending network traffic, we will do the computation on-device. Otherwise we do it in the cloud. Intuitively, if the task completion requires the cloud anyway (e.g. to call 3rd-party cloud APIs, to retrieve world knowledge, or to access user memory and social graph), then we do as much compute in the cloud as possible in the processing. Otherwise, technology companies can use on-device compute and avoid waking up the antenna and network stack completely, if the total energy consumption is less this way.

This means on-device compute will handle relatively simple tasks and the cloud service will handle more complex tasks. Given the relative difference in complexity, the primary engineering focus for the Assistant provider company should be cloud service, with on-device compute playing a very important but subservient role.

Given the energy consumption and latency requirements, the on-device compute implementation will utilize custom ML silicon as much as possible. To compute that there must exist general purpose CPUs which mostly utilize the primary programming language C. The on-device components can include ASR, NLU, dialog engine, NLG, TTS, and arbitration service that determines whether to use the cloud.

Integration with app model

Besides the runtime integration work to enable apps to use the Assistant service, there is also a need to release the Assistant Studio as a development platform, for developers to define intents that are specific to their app (e.g. in-game actions that only makes sense in the context of the game's world), as well as to define the related dialog policies and task completion behaviors, and to annotate data for model training.

Integration with the shell

The Assistant can be an integral part of the “Shell” (or “Launcher”). The user should be able to use natural language as well as other means to launch apps, invoke system functionality, and control device behavior. The Proactive component in Assistant should be the cloud counterpart of the on-device notification management subsystem, to provide content (e.g. calendar events, reminder, etc.), and interfacing with the notification queuing and transport system to provide situational intelligence (e.g. whether the user and the device are in interruptible state, whether it's appropriate to show certain notifications, and the best output modality for the notifications).

Security and privacy

Since the AR device will be present in the most intimate and private parts of a user's life, it should be self-evident that we need high-level security and privacy guarantees around Assistant in AR. Here we list the principles around privacy. There can be a secure operating system which provides a strong foundation for us to implement these principles.

1. Without explicit triggers, the system must not persist captured content, except into a private vault that only the user has access to (specifically, social networking platforms cannot decrypt content in users' private vaults for any purpose).

2. With an explicit trigger, the system is allowed to persist captured content, accessible by Social networking platform. The trigger can be in different forms; wake word and gestures are commonly understood ones. Other triggers can be used as well, as long as users have a clear understanding of the implications of the triggers.

3. Aggregated statistics at a coarse level are allowed to be captured and accessible by Social networking platforms.

The “private vault” can take the form of on-device storage, or more likely, an encrypted form in a social networking platforms data center which requires a per-device key to decrypt (and which social networking platforms does not have access to at rest). Platforms can stream contextual data to the Assistant and store and learn user memory from the stream, as long as they are stored securely such that only the user can decrypt the content. It does have implications on how platforms implement user memory.

Multi-layer inference and fusion, and Assistant's responsibilities

Developing a contextual ‘Assistant surely requires ML inferences and a fusion of multiple signals together to produce higher-level semantic understanding.

Yet, it should be clear that the AR system has ML inference in multiple components at multiple layers outside the Assistant, and signal fusion can happen at multiple layers outside of Assistant as well. For example, the Assistant will take the result of computer vision as input. Yet, the Assistant is not where computer vision happens, nor the place where location information is fused into computer vision to aid object recognition.

Assistant will need to implement *part* of a multi-sensor fusion architecture, which will have both device- and server-resident components. To enable smooth integration of gesture, gaze, location, and computer-vision inputs, these components will need to be integrated at a fairly low level in the device operating system. This suggests that the “intent detection” component of the Assistant can become a feature of AR development platforms, and further suggests that high-level intents will be a significant source of control events for user-facing applications. This has implications for the AR development platforms, Application Model, the user interface toolkit, the service discovery framework, and the software distribution model of the platform; for all of these the Assistant team must help to craft a multi-modal and contextual experience.

World knowledge

The Assistant provides the bridge between the AR device and the wealth of world knowledge, to give the user super-human knowledge. The knowledge graph subsystem in the Assistant has offline components that gather and understand facts, and online components that retrieves knowledge.

User memory

The Assistant provides the bridge between the AR device and the Social networking platforms social graph, and it remembers relevant personal experiences and events for the user, to give the user super-human memory. The user memory subsystem determines the relevance of signals, filters the signals based on relevancy, and stores the signals with useful form and granularity. It also provides search and retrieval of the memory, as well as proactive functionality to show notifications, recommendations and reminders at appropriate time in appropriate context.

Physical and virtual world task completion

Examples of physical world task completion include “calling Uber” and “ordering pizza.” Whereas virtual world task completion are app specific, such as opening a blast door in a game (related to the app model integration described above). The Assistant has specific agents to handle different categories of physical and virtual world task completion. The Assistant will solve integration problems, and implement agents, to interoperate with home and automotive IoT platforms as well.

Evaluation

Once the ideas in this framework could be realized, AR/VR convergence would happen to a great extent. Yet, what does success look like?

At a basic functional level, without an amazing voice-powered contextual Assistant, there would be no reasonable way to interact with AR glasses for several hours every day without interfering with other aspects of life. In this sense, success means being able to leave the stage in your bag and not having to stop to invoke virtual keyboards and other input mechanisms when on the go throughout the day (aside from certain productivity / work / gaming scenarios of course).

Ultimately, success means that the assistant present in AR glasses feels like a trusted companion and guide, that anticipates each person's needs and makes their experience in the world feel richer, smarter and more rewarding. It is indispensable in day-to-day life and achieving desired outcomes and goals within each person's ecosystem of friends, family and co-workers.

Conclusion

Realizing the roadmap for AR/VR convergence requires technology companies to develop the following items:

1. **A way to operate and get the most out of the device** - the assistant will be a primary interface to control the functions and settings of the device.

2. **Memory superpowers -** long- and short-term memory and recall powers that are essentially unlimited, including past experiences, preferences, calendar, important places and people.

3. **World knowledge superpowers -** information should be immediately accessible through voice and gaze. Assistant should inform and delight

with information, entertainment and more, anchored in the context of the world around a person at that moment in time.

4. **A personalized experience of the world -** that integrates real-time inputs, personal context and knowledge of the world to enable AR experiences that are highly relevant and help people achieve their goals, especially as they relate to family, friends and colleagues (delivered through the “context engine” in the picture below).

5. **A personal digital assistant with a voice and personality** - an embodiment of a trusted guide that the person can interact with through a combination of modalities.