INFOMMMI (Multimodal Interaction) 2017-2018

Exam questions, part 2 (lectures 5-7, W. Hürst)

(max. 40 points)

Comments

Some comments on possible solutions.
Notice that these are incomplete and
for some questions other answers exist
that might give full credit, too.

Question 2-1: AR/VR comparison (max. 2 points)

Assume you are an AR/VR developer who is developing VR applications for the HTC Vive (which is a head-mounted display VR system) and AR applications for the Microsoft HoloLens (which is a head-mounted see-through AR system).

(Note: In the following, a short statement can be sufficient to get full credit. No detailed explanation is needed.)

Give one issue, characteristic, or aspect that is more difficult to deal with in VR (with the HTC Vive) than in AR (with the HoloLens) and shortly state why.

This question was basically about addressing differences between VR and AR, so any of the issues listed on, for example, slide 35 from the 5th lecture would be correct (but other differences exist that are correct, too; in fact, many people wrote something about cyber sickness or navigation, which was nice, since it shows that you made a connection between Peter's and my lectures). Because the HoloLens, which has an OST display, is explicitly stated, some (but not all) of the differences between VST and OST displays that we discussed might also be correct here (because VR also uses a "full" video display), although few people did that.

Give one issue, characteristic, or aspect that is more difficult to deal with in AR (with the HoloLens) than in VR (with the HTC Vive) and shortly state why.

Same as above.

Question 2-2: Displays / depth perception cues (max. 5 points)

In the lecture, we said that depth cues for 3D perception can be categorized in four groups:
Oculomotor cues, binocular depth cues, motion related cues, and pictorial depth cues.

(Note: Below, an explanation is not needed. It is sufficient to state the display type or characteristic that the display(s) have to fulfill to be able to use these depth cues.)

For the 3 answers below, many gave informal (and correct) descriptions, which got them full credit, too.

What kind of display do you need if it is essential for your application to support oculomotor cues?

Light field displays

(Because oculomotor cues address accommodation and convergence, which are only resolved by these displays. Note: some people wrote something like "displays with eye tracking", which is actually correct, too. And "retinal display" gave full credit, too, although there might be some that don't solve this, but we didn't discuss this in the course, so it was counted as correct here.)

What kind of display do you need if it is essential for your application to support binocular depth cues?

Stereoscopic displays

(Because binocular depth cues result from binocular disparity, which is basically stereovision. Notice that you can also achieve this with regular displays if you are wearing 3D glasses or lenses, so this answer would have been correct, too)

What kind of display do you need if it is essential for your application to support pictorial depth cues?

Any

(Okay, that was a bit of a trick question (but we saw some examples in the lecture, so it wasn't that mean to ask it). Pictorial depth cues are the ones we can get with computer graphics, so any display that renders pixels, i.e., all of them, is sufficient. Many wrote something like "a standard display", which is therefore totally correct, too.)

(Note: Below, a short phrase or sentence illustrating this advantage can be sufficient to get full credit. A detailed explanation is not needed.)

Give one advantage that using an *OST (Optical-See-Through) display* might have compared to using a VST (Video-See-Through) display.

See lecture 6, slides 42 and 46.

Give one advantage that using a *VST (Video-See-Through) display* might have compared to using an OST (Optical-See-Through) display.

See lecture 6, slides 42 and 46.

Question 2-3: AR interaction (max. 3 points)

Interaction devices or concepts that we can use in AR include *special devices for 3D/6DOF interaction*, *6DOF hand tracking*, and *Tangible User Interfaces*.

(Note: Below, a short phrase or sentence illustrating this advantage can be sufficient to get full credit.
A detailed explanation is not needed.)

Give one advantage that *special devices for 3D/6DOF interaction* have compared to the other two approaches.

See lecture 7, slide 35 (but other correct answers exist and got full credit, too)

Give one advantage that *6DOF hand tracking* has compared to the other two approaches.

See lecture 7, slide 35 (but other correct answers exist and got full credit, too)

Give one advantage that *Tangible User Interfaces* have compared to the other two approaches.

This was not stated on the slides, but we shortly discussed advantages of TUIs, and if you understood the basic idea behind them, it is easy to find reasons or situations for which they are preferable to the other two options.

Question 2-4: True AR (max. 8 points)

In their paper "Breaking the Barriers to True Augmented Reality", Sandor et al. describe their vision of "True AR" as an augmented reality that is indistinguishable from reality. They define it informally as "a modification of the user's perception of their surroundings that cannot be detected by the user."

(Note: A detailed elaboration is not needed for full credits in the following 2 questions as long as your example is convincing.)

Give an example or use case where such a True AR will likely be beneficial compared to the one we get with today's systems and shortly state why.

Many possible examples exist here (some people came up with pretty good ones actually).

Give an example or use case where such a True AR is not needed, not wanted, or could even be harmful and shortly state why.

Same as above.

The authors also discuss how an AR Turing Test could be designed (i.e., a test to verify if or to what degree an AR system could be classified as True AR). They propose three "dimensions along which the interactions in a useful test scenario can be restricted." What are these three dimensions?

(Note: It is sufficient to write down the three phrases that are used in the paper. Yet, if you don't remember the exact phrasing, you can also get full credit by giving a short informal description of the three aspects that one would need to test.)

The correct phrases, which would have given full credit are:

- Which senses can be used
- Consistency of the simulated world with actual reality
- Transition between actual and augmented reality

Hardly anyone got all three completely right, but many used phrasings for some that got them partial or sometimes even full credit; e.g., some wrote "modalities", which is similar to senses (thus, full credits), some wrote interaction, which is related to consistency (although not exactly the same; thus, partial credits, depending on how it was phrased).

The authors argue that Light Field Display technology is needed to achieve True AR for the visual channel. Give one reason why.

(Note: A short phrase can be sufficient to get full credit.)

"They resolve the accommodation-convergence conflict" *would be an answer that gives you full credit (but other correct answers are possible, too; the paper goes much more into detail here. Note: many people didn't name this conflict correctly but informally described it, which got them full credits, too (if phrased correctly))*

Question 2-5: AR systems (max. 8 points)

AR systems can be realized in various ways by using different display technologies. One example is handheld AR (when using a mobile phone), another is projected or spatial AR (when using a projector).

(Note: For the 2 questions below, you don't have to describe a full game. Listing an aspect or characteristic that is likely easier to realize or that is a clear advantage compared to the other system can be sufficient to get full credits.)

Give one characteristic or advantage why a company developing AR applications might chose to develop an AR game *for a handheld device* and not for projected AR.

Again, many possible answers exist here that could be correct. Notice that these could be based on the technology (e.g., projected AR is more sensitive to lighting conditions), characteristics and possible usages (e.g., multi-user support might be easier to achieve in projected AR (and it's more interesting for bystanders and to attract attention)), but also economic reasons (e.g., it might be cheaper to develop a downloadable app than installing the hardware for a projective system). We haven't addressed all of these issues in the course, but if they were correct, they obviously got full credits, too (e.g., some people talked more about interaction-related issues or the FOV, which are good points, too).

Give one characteristic or advantage why a company developing AR applications might chose to develop an AR game *for a projected AR system* and not for handheld AR.

See comments above.

Now we want to compare a handheld AR system that is created with a regular phone using the front facing camera, and an immersive AR system that is created with a head-mounted display such as the HoloLens using a See-Through-Display and various cameras for tracking. Assume that you want to implement an AR tower defense game. For this game, players have to place a marker on a table. The AR system shows a tower at the position of the marker. It also shows tanks approaching the tower from all sides. The goal of the game is to shoot and destroy all tanks before they reach the tower.

Note: some people misinterpreted what I meant with front-facing camera. Admittedly, this could have been better phrased in the question, sorry. Thus, they got full credits, too, if they provided a correct answer for the other scenario

Give one problem that is hard or even impossible to solve when using *a regular phone* to implement this game, but that does not appear or is easier to resolve when using a HoloLens.

The key to this and the next question is that with a regular phone's camera, you just get a single image, so it is very difficult to get accurate information about the environment. With the tracking cameras integrated into the HoloLens, you get also depth information about your environment (we talked about this when discussing the HoloLens / see-through displays, but also in the tracking lecture). Thus, a tank passing behind a real object (e.g., a coffee mug standing on the table) is hard to visualize correctly on the phone (occlusion problem), but easier on the HoloLens. Likewise, you don't know with the phone where the table ends, so a tank can drive on the same surface but "float in the air". Other correct answers to this question exist of course.

Give one problem that is hard or even impossible to solve when using *a HoloLens* to implement this game, but that does not appear or is easier to resolve when using a regular phone.

An obvious answer, also addressed in the lectures, is again the occlusion problem, but this time the other way around, i.e., when a tank passes in front of a real object. The HoloLens only allows to "add light", so with a bright mug, it will "shine through" the tank ("ghost image"), whereas the phone allows us to "remove" the mug by just not drawing the pixels in the video, but replacing them with the tank.

Question 2-6: Non-visual AR / Azuma's AR definition / sensors (max. 14 points)

In 1997, R. Azuma introduced a formal definition of AR by specifying three characteristics that each AR system should fulfill. Although his article only elaborates on visual displays, the definition was intended to cover other modalities as well. Recently, the speaker and headphone manufacturer Bose introduced AR glasses that do neither contain visual displays nor cameras, but are purely focused on audio. We shortly discussed it in the lecture. Here an excerpt from a related news article:

"Unlike other augmented reality products and platforms, Bose AR doesn't change what you see, but knows what you're looking at – without an integrated lens or phone camera," Bose said. "And rather than superimposing visual objects on the real world, Bose AR adds an audible layer of information and experiences, making every day better, easier, more meaningful, and more productive."

Discuss if Bose's AR system fulfills the characteristics of Azuma's definition by completing the sentences below (and crossing out the parts that are not correct).

(Note: The option "does only partly fulfill" has been added below because there might be cases that are debatable or when the above description might not provide all information to give a clear yes or no answer. It is more important that your answer reflects that you really understood the essence of this characteristic than just guessing the right answer.)

A first characteristic of Azuma's definition is:

Combines real and virtual

The Bose AR system [**does fulfill** | ~~**does only partly fulfill** | **does not fulfill**~~] this characteristic because:

This seems clearly fulfilled, since it combines real and virtual sounds (there is nothing in the description that suggests that real sound is blocked out). Note: many stated that it combines virtual sounds with real objects, which is correct, too (some people might argue that only combination within one modality qualifies as AR, but this is just an opinion and since Azuma's definition does not directly reflect on it, I consider both opinions to be correct).

A second characteristic of Azuma's definition is:

Is interactive in real time

The Bose AR system [does fulfill | **does only partly fulfill** | ~~**does not fulfill**~~] this characteristic because:

Given current sensor technology, it seems fair to assume that we get the information "what you are looking at" in almost real time and assuming playback of audio can also be done immediately, you should be able to implement audio playback that reacts in real time to your interactions (= head motions). If this is considered "truly interactive" is debatable, e.g., since you cannot influence the output if you stay at a fixed location (the same goes for many visual systems btw.). Therefore, depending on how you justify it, the other two options could also count as a correct answer (it really depends on how you interpret "interactivity"; if your description matched your interpretation and the latter made sense, you got full credits).

A third characteristic of Azuma's definition is:

Registered in 3D

The Bose AR system [does fulfill | **does only partly fulfill** | does not fulfill] this characteristic because:

Since this is a sole audio device, it seems likely that it supports 3D surround sound / binaural sound, in which case you can register sounds correctly in 3D (more or less). Yet, this is not clear from the description, which is why I marked the "partly fulfill" characteristic, but depending on how you argue, both other options could be equally valid. It's important though that your answer reflected the registration of the augmented parts (i.e., the sounds), not just the user.

In the news article cited above, they also list several potential use cases for such a system. For the two examples quoted below, list what kind of sensors have to be integrated into the device to realize it and shortly state why (i.e., what kind of data you get from the sensor(s) that is needed for this use case).

(Note: Only list sensors that are absolutely needed and don't just write down any sensor that comes to your mind. Adding ones that are obviously not needed will result in zero credit.)

Use case example 1: *"For travel, the Bose AR could simulate historic events at landmarks as you view them … You could hear a statue make a famous speech when you approach it."*

Sensor(s) needed and data they deliver:

GPS should be sufficient to implement such a scenario (since it is not important if you look at the statue directly; that would only be the case if it has a conversation with you).

Use case example 2: *"Bose AR would add useful information based on where you look. Like the forecast when you look up or information about restaurants on the street you look down."*

Sensor(s) needed and data they deliver:

For that we need location information (GPS) and looking direction, which we could get via accelerometer and/or gyroscope (mentioning either of these or both plus GPS gave full credits).

Bose explicitly states that their device does not include any cameras. Give one use case example for audio AR that cannot be implemented for the Bose AR glasses because a camera would be needed as sensor.

(Note: You can either shortly describe an example case like above or just list a characteristic that is impossible to realize without a camera.)

Example / characteristic:

The most obvious example that came to my mind is anything with a changing environment (e.g., a car parked in front of something), since that needs to be recognized by visuals. Other examples exist.