[20230322] INFOMMMI - Multimodal interaction - 3 - USP

Course: BETA-INFOMMMI Multimodal interaction (INFOMMMI)

Note: This file only contains the questions covering lectures 5-7 (W. Hürst)

Duration: 2 hours

Number of questions: 12

[20230322] INFOMMMI - Multimodal interaction - 3 - USP

Course: Multimodal interaction (INFOMMMI)

EXAM CONTENT AND DURATION

Questions 1-6 cover lectures 1-4 by Peter Werkhoven (max. 50 points).
Questions 7-12 cover lectures 5-7 by Wolfgang Hürst (max. 50 points).

The maximum time to answer all questions is two hours. It is up to you how much time you spend on which questions.

Good luck & success!

Number of questions: 12

7 Defining AR & its characteristics

R. Azuma introduced three characteristics to define Augmented Reality (AR). (*Note: Read all sub-questions before answering the first one. This is basically just one question split into three, so it is easier to grade and easier to give you partial credits if some parts are not fully correct.*)

1 pt. **a.** [max. 1 point] Explain one of them in your own words. (*"In your own words" means that an informal description is sufficient.*)

2 pt. **b.** [max. 2 points] Give an example of a useful system that is commonly referred to as an AR system but does not fulfil this criterion or does so in a very basic, limited way.

2 pt. **c.** [max. 2 points] Give an example of a useful system that is commonly referred to as an AR system and fulfils this criterion to a better degree than the one you described in the previous sub-question.

Comments on potential solutions

For a.

1. Combines real and virtual.
2. Is interactive in real time.
3. Is registered in three dimensions.

(Note: The question asked for an informal explanation. It is not given here, since there is an infinite number of ways to informally describe these characteristics, but if it as missing in the exam, 0.5 points were be deducted.)

Good examples for 1. (interaction) include things that just show virtual objects (e.g., information browsers, such as the Layar browser, or "the Beatles crossing Abby Road" example from the lecture) for b. and ones that let you actively interact with them (e.g., the UFO defender or tower defense game examples from the lecture) for c.

Good examples for 2. (registration) in 3D include things that just float in the air and are superimposed on the real world (e.g., the UFO defender game) for b. and ones that show them at a fixed location (e.g., the tower defense game) for c.

Good examples for 1. (combination of real/virtual) are a bit trickier to find, since basically anything that is not fully VR or fully real is a combination of both. Some tried it though and got a reasonable answer, which, even if not fully correct, still got them full credits if it reflected a good understanding of the matter.

(Note that there is an infinite number of examples that could be used here, and depending on which one is picked, it could very well be that something listed under b. or c. above is a correct answer for c. or b., respectively, in another context.)

8 Tracking

When building AR systems, ideally, we want to be able to track all six degrees of freedom (6 DOF).

2 pt. **a.** [max. 2 points] What are these 6 DOFs?

5 pt. **b.** [max. 5 points] Not all AR systems need to track all 6 DOFs. Give an example for a useful AR system that only needs to track 3 DOFs. Mention also what DOFs these are and what kind of sensor you would use to track them. Justify why you need these sensors and why it is a good choice to use them (and not others).

2 pt. **c.** [max. 2 points] Give an example for an AR system that would require 6 DOF. (*No explanation about sensors is needed here. Just explaining the system is sufficient.*)

a. Location (x/y/z position in space) and orientation (rotation around each of the three axis). (*The technical terms for orientation are pitch, yaw, and roll. Describing them informally without stating their name, as done above, would have given full credits, too.*)

For b. and c., various examples exist. The easiest for b are probably the ones where you only have virtual objects that do not have an orientation (e.g., a 3D mouse pointer, a ball (unless you consider the ball's texture)) and thus only need spatial tracking. Other examples are systems where the virtual objects are placed in a 2D plane or fixed distance away from the user and thus do not require tracking in all directions. Good examples for c are basically everything where both the location and orientation of a user's motions or virtual objects in the 3D space are needed.

A few people forgot to describe the used sensors in b and thus only got half of the credits. (Hint for future exams: Make sure to read the exam question carefully and don't waist valuable credits by overlooking parts of what was asked for.)

9 Display characteristics: FOV and FOR

AR displays have a field of view (FOV) and a field of regard (FOR).

2 pt. **a.** [max. 2 points] Explain what FOV means in this context. *(A short, informal description is sufficent.)*

2 pt. **b.** [max. 2 points] Explain what FOR means in this context. *(A short, informal description is sufficent.)*

2 pt. **c.** [max. 2 points] Give an example of an AR display where the FOV and the FOR are the same.

2 pt. **d.** [max. 2 points] Give an example of an AR display where the FOV and the FOR are *not* the same.

2 pt. **e.** [max. 2 points] Give an example of a use case or situation where the issue above (FOV is not the same as FOR) could cause a problem or result in a negative user experience.

See lecture notes for a. and b.

Probably the most obvious example for c. is spatial AR (because the whole projection can be seen at any time, and this is also the only area where virtual elements can appear). *(Explanation was not required.)* Other examples include the magic mirror we saw in the lectures or any other kind of static display.

Examples for d. are head-mounted displays (both optical see-through and video see-through), since they have a rather small FOV but can display virtual elements at random positions around you. The same goes for phones or tablets used for handheld AR. *(Explanation was not required.)*

Good examples for e. are situations where relevant information is not noticed because it is outside the FOV (or when elements are cut off because of the limited FOV – remember the example discussed in the lecture of a virtual avatar standing close to you but only parts of the avatar's body can be seen due to the limited FOV).

10 Display characteristics: Optical vs. video see-through displays

One potential issue with optical see-through glasses for AR are so-called “swimming artifacts.”

2 pt. **a.** [max. 2 points] Explain what is meant by that.

Text from the slides, under “OST disadvantages”: Lag & jitter between real & virtual images (”*swimming artifacts*”)

1 pt. **b.** [max. 1 point] Does this effect also happen for video see-through displays? Shortly explain your answer.

Text from the slides, under “VST advantages”: Matchable time delays (compensation for lag between real and virtual images, no “swimming artifacts”)

Another unwanted effect often seen with optical see-through glasses for AR are so-called “ghost images.”

2 pt. **c.** [max. 2 points] Explain what is meant by that.

Text from the slides, under “OST disadvantages”: HMD only adds light to physical image (“*ghosts*”, problem with darker colors)

1 pt. **d.** [max. 1 point] Does this effect also happen for video see-through displays? Shortly explain your answer.

Text from the slides, under “VST advantages”: True occlusion (light can be added and subtracted, no “ghosts”)

Any sentence(s) explaining these phrases (or expressing them in a reasonable way) would have given full credits. Note that important aspects needed to be provided (e.g., the difference between real & virtual for swimming artefacts). I graded generously though (e.g., “lag” and “jitter” are two different things, since the first one happens when people are moving but the second can happen also when they are not – people still got full credits though if only one of them was mentioned).

2 **Display characteristics: Depth perception & vergence-accommodation conflict**

Humans perceive depth in various ways. One of them are pictorial depth cues.

2 pt. **a.** [max. 2 points] Name two pictorial depth cues. (*Just stating the name or phrase is sufficient. No explanation needed.*)

- Occlusion
- Linear perspective
- Size gradient
- Relative height
- Texture gradient
- Relative brightness
- Aerial perspective
- Depth-of-focus
- Shadows
- Shading

In the paper "The Effect of the Vergence-Accommodation Conflict on Virtual Hand Pointing in Immersive Displays" from 2022, Batmaz et al. studied the effect of vergence and accommodation, which are non-pictorial depth cues, on virtual hand interaction.

1 pt. **b.** [max. 1 point] Name one other non-pictorial depth cue, i.e., other than vergence and accommodation. (*Just stating the name or phrase is sufficient. No explanation needed.*)

- Binocular disparity (stereopsis)
- Motion parallax
- Kinetic depth

2 pt. **c.** [max. 2 points] Shortly explain the vergence-accommodation conflict.

Since only an informal description was needed, there is an infinite number of ways to answer this correctly. For example: Virtual objects are all rendered at the same distance, which creates a mismatch of accommodation and vergence. While accommodation (that is, informally, how your lenses adjust when focusing) adjusts to this fixed rendering distance, vergence (that is, informally, how your eyes rotate when focusing) still adjusts to the "virtual" position the object would have in the real world.

1 pt. **d.** [max. 1 point] Different technologies exist that deal with this conflict. Name one of them. (*Hint: In the lecture, we mentioned two. The paper lists additional ones. Just stating the correct term or phrase is sufficient. No explanation needed.*)

- Eye tracking (and adjusted rendering)
- Light field displays
- Varifocal displays
- Holographic displays

(First two are from the lecture, last three are mentioned in the paper.)

2 pt. **e.** [max. 2 points] One major problem with the vergence-accommodation conflict is that it negatively affects depth perception. Mention two other negative effects of this conflict.

From the paper:

- Fatigue
- Increased eye fatigue
- Reduced performance
- Reductions in (viewing) task performance
- High cognitive load
- Reduced comfort

(These are all quotes from the paper. Other phrasings are possible.)

2 pt. **f.** [max. 2 points] Imagine an AR display and a VR display which both have resolved the vergence-accommodation conflict. Give one reason why the AR display may demonstrate a better interaction performance than the VR display, even if neither of them suffers from the vergence-accommodation conflict.

This is related to hypothesis H3 in the paper. Quotes from the paper:

"While VR headsets are designed to show virtual content, see-through AR HMDs enable users to perceive real-life environmental cues, including lighting and texture. For instance. … Thus, we hypothesize that user performance using AR multifocal HMDs is better than using VR multifocal HMDs."

"… the additional cues afforded by the AR condition together with the ability of displaying virtual elements at the same depth as physical ones allow users to better identify the target depth."

Informally, this basically says that the real-world environment can provide more and better depth cues than a fully rendered VR environment, which could potentially lead to a better performance. Any comment like this gave full credits.

3 **Interaction**

Controllers have become the de-facto standard used by most Virtual Reality systems these days. Yet, not for Augmented Reality.

1 pt. **a.** [max. 1 point] Give one convincing reason why.

Various correct answers exist. The two most used ones were variations of the following:

- Controllers are visible (in contrast to VR where you can replace them with a virtual representation that can be anything).
- Makes it difficult to interact with real and virtual objects at the same time.

2 pt. **b.** [max. 2 points] Although controllers work very well in many VR tasks, they also have limitations. Name two of such limitations. (*One short sentence for each should be sufficient to answer this question. No detailed explanation required. Hint: Remember the innovative controller solutions that we saw from Microsoft Research*.)

Various correct answers exist. Some people also came up with nice ones not mentioned in the lecture. Examples for correct answers include but are not limited to:

- One-size-fits-all solution (that does not always fit to all problems).
- No force feedback (compare to first example from MS Research).
- Only limited haptic feedback (vibration, but no texture, see third example from MS Research).
- Haptic feedback limited to certain positions or whole controller.
- No experience of shape (see last example from MS Research)
- No correlation between two controllers (see first example from MS Research)
- Special hardware needed (see slides – comparison with hand tracking)
- Unnatural (see slides -—comparison with hand tracking)
- Difference between interacting with real and virtual objects (see slides – comparison with hand tracking (applies only for VR, but gave full credits, too))
- Ergonomics, Gorilla arm

Tangible user interfaces (TUIs) are another way to interact in AR.

2 pt. **c.** [max. 2 points] Give one possible advantage that a TUI-based interaction has compared to using a controller when used in AR.

2 pt. **d.** [max. 2 points] Give one possible disadvantage or limitation that a TUI-based interaction has compared to a controller when used in AR.

One interaction technique used in VR that could also be applied to at least some AR scenarios is cone casting (aka flashlight), which is a modification of ray casting.

1 pt. **e.** [max. 1 point] Shortly explain how it works.

See lecture.

1 pt. **f.** [max. 1 point] What problem does it address that we often see with ray casting?

Objects at a distance are hard to target due to a lever effect.

1 pt. **g.** [max. 1 point] Give one common problem with cone casting that remains or is newly introduced.

If objects in the distance are very close to each other, it becomes even harder to select the right one.

There are more correct answers here because the question also asked for problems that remain. So, for example, selecting occluded objects is still difficult.

Thank you for participating in the course. We hope you enjoyed it. Please fill out the Caracal evaluation to tell us what went well and what didn't, so we can further improve it. Thanks!