INFOMMMI (Multimodal Interaction) 2016-2017

Exam questions for <u>part 2</u>

(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: Tracking (max. 4 points)

In the lecture, we compared the common "interaction loop" for VR implementations with a similar one for AR implementations. What actions in the AR loop correspond to the step "tracking" in the VR loop?

(Note: Three words should be enough to answer this question correctly.)

Calibration, registration, tracking

Give one example or reason, why tracking in AR is commonly considered more difficult than in VR.

(Note: A short answer should be sufficient to get full credit. No long explanation needed, if the difference between AR and VR is clear. You will NOT *get credits though for mentioning a general tracking issue that appears in AR and in VR.)*

Various ways exist to answer this correctly; most examples go back to the fact that it is harder because we need to combine the real world (or, e.g., a live video feed of it) with virtual elements (making, e.g., lags more noticeable and less tolerable, lead to inaccuracies, etc.)

Question 2-2: Display technologies (max. 6 points)

Complete the following sentences to create a correct statement:

(Note: Give a short explanation, and ***do not forget to cross out the part that is wrong*** *in the first part of the sentence. Just complete the sentence. There is no need to explain the terms accommodation and vergence.)*

a) Accommodation is generally **[~~a problem |~~ no problem]** with video see-through displays, because …

… everything you see is at the same focal distance.

b) Vergence is generally **[a problem ~~| no problem~~]** with video see-through displays, because …

… of possible vergence-accommodation conflicts (= mismatch of the position / viewing direction of your eyes and the focus of the eyes' lenses)

Question 2-3: Perception & UI design (max. 3 points)

AR implementations on handheld devices (e.g., smartphones) often do not register virtual objects in 3D, but just superimpose them onto the live video stream. To cope with this, pictoral depth cues are often used in interface design for such applications to create more "3D-looking" experiences. Give three examples of such pictoral depth cues.

(Note: It is sufficient to just list the names of three such pictoral depth cues to get full credit. An explanation or concrete description of how they are used in an AR interface is NOT *necessary. Only list* THREE *cues. If you write down more than three, you will get* NO *credit for this question.)*

1. Occlusion
2. Linear perspective
3. Size gradient
4. Relative height
5. Texture gradient
6. Relative brightness
7. Areal perspective
8. Depth-of-focus
9. Shadows
10. Shading

(note that the rather uncommon "no credits for more than 3 examples" was put here to avoid that people just write down random stuff in the hopes of getting at least a few right)

Question 2-4: AR interaction (max. 4 points)

In the lecture, we discussed how tangible user interfaces (TUIs) can be used for interaction in AR. Other options for AR interaction include using a dedicated (and trackable) device, such as a magic wand, or tracking a user's hand(s).

(Note: A short sentence can be sufficient to get full credit. No long explanation needed. Make sure though that your example clearly states the advantage of TUIs compared to the other method. That is, do not just list any advantage of TUIs, but one that is obviously a disadvantage of the other method.)

Name one advantage TUIs can have over a magic wand for AR interaction:

Various ways exist to answer this correctly; for example, things like TUIs = real objects integrated into real environment, direct versus indirect manipulation of virtual objects, shape (cup, paddle, …) can better match the task, etc.

Name one advantage TUIs can have over hand-tracking for AR interaction:

Various ways exist to answer this correctly; for example, tactile perception (you can "feel" the TUI), better feedback that might be easier to control (e.g., rotation angles might be easier to adjust when you have a real object in your hand than just making hand gestures mid-air), could be easier to track accurately, etc.

Question 2-5: Applying Azuma's AR definition to concrete systems (max. 11 points)

In his paper "A survey of augmented reality" (1997), R. Azuma introduced a definition of augmented reality by providing three criteria that an AR system should fulfill. Give one example of a system that one would intuitively consider to be an AR system that does NOT fulfill Azuma's definition (i.e., that according to the definition is technically NOT an AR system, although we normally still call it an AR system, AR program, or AR app. Shortly state which characteristic is violated and why.

(Note: A very short explanation should be sufficient to get full credit. No need for detailed explanation.)

Various ways exist to answer this correctly; e.g., the ones we discussed in the lecture in relation to handheld AR (smartphones, where virtual objects are only superimposed onto the live video stream , but not registered in 3D). Several people actually referred (correctly) to Pokemon Go ☺

The *Meta cookie* system by Narumi et al. (2010) describes a setup where real cookies with a trackable QR code on top are augmented by smells and related visuals displayed at the location of the barcode. These augmentations are perceived via a head-mounted device containing tubes for blowing odors into your nose and see-through glasses to show the visuals (see image).

Shortly discuss this setup with respect to the three characteristics introduced in the AR definition by Azuma. That is, list each of the three characteristics and then address if and how it is fulfilled for each of the two modalities smell and visuals.

The characteristics are (in random order):

- Combines real and virtual
- Interactive in real-time
- Registered in 3D

Notice that this question does not have clear yes/no answers but was intended to check if you really understood the content or just tried to memorize it. Thus, you generally got full credit if it was clear that you "knew what you are talking about" and discussed the most essential issues, even if some things were missing.

The reason why there is not a clear yes/no answer to some of the issues is that

a) the system description is not 100% complete (e.g., it's not clear if people can only smell the artificial odor or the real world as well; thus, some argued it combines real and virtual smell, others said it doesn't; both are correct, depending on how you interpret the system and thus got full credits)

b) Azuma's characteristics leave room for different interpretations (e.g., in a multimodal AR setting like here, does a purely virtual/artificial smell combined with a real visual fulfill the "combines real and virtual" criteria, or is that only fulfilled when real and virtual are combined within one modality? Again, both interpretations are possible and gave you full credits, as long as your description illustrated what you meant and that you understood the underlying basics and principles).

Question 2-6: Comparison of different AR systems / display technology (max. 12 points)

Assume you are wearing a head mounted video see-through AR display and you are looking at a table. On the table, there is a virtual coffee mug created by the AR system and a real coffee mug.

(Note that some of the following questions can be answered very shortly. The space between the text does not necessarily represent the size of the expected answer. But if we would make it shorter for some questions, we might give away too much of the answer already.)

Similarly to the last question, this was more about checking if you really understood the content, but not about getting a 100% correct or complete answer (e.g., many people did not list all sensors or necessary techniques, but didn't get a reduction for it). The major issue here was the occlusion problem. Some people discussed other issues, too (e.g., lower FOV for optical see-throughs), and got some (small) credit for it, too, if it made sense (i.e., it was not just a general statement with no direct relation to the described situation).

Notice that the same comment applies for the notes provided below: they are NOT complete but just illustrate the major issues; if you provided something similar, you likely got full credits, but if you phrased it differently or stated other things that are correct, too, you likely got credit for it as well.

a) Situation 1: From your perspective, the virtual coffee mug is partly behind the real coffee mug. Shortly describe what the AR system must do to make the scene look realistic. When doing this, state what information is needed for this, how to get it (e.g., with what kind of sensor), and what approach / technique / algorithm / etc. one might use for this. If the problem is not solvable with today's technology, shortly describe why.

   Parts of the virtual cup are occluded by the real one.

   Thus, we need to identify this overlap (needs, e.g., depth information from real world, e.g., via depth cameras), and only render the pixels of the virtual cup that are not occluded

   Note: many people didn't mention depth information, but still got full or almost full credit, if the rest of their description was correct. Other (correct) statements include head-orientation, etc. Some people also made comments about accommodation, which (if correctly phrased) got them some credits, too.

b) Situation 2: Assume the same scenario as in situation 1, but now you are wearing an optical see-through display. Would the situation change and how (or why not)?

   This leads to the same occlusion problem as in situation 1 => same answer

   Note: one could argue though that the same problem as in situation 4 applies here, too, with respect to the virtual cup and its background. Many people did and got full credit for it (but you didn't get any deduction if you didn't mention it and your answer was otherwise correct).

c) Situation 3: Assume the same scenario as in situation 1 (i.e., you are wearing the video see-through display again), but now the virtual coffee mug is partly in front of the real coffee mug. Shortly describe what the AR system must do to make the scene look realistic. When doing this, state what information is needed for this, how to get it (e.g., with what kind of sensor), and what approach / technique / algorithm / etc. one might use for this. If the problem is not solvable with today's technology, shortly describe why.

   Same as in situation 1, but now we have to identify the occluded pixels from the real cup and don't render those.

d) Situation 4: Assume the same scenario as in situation 3, but now you are wearing an optical see-through display. Would the situation change and how (or why not)?

   We can only add but not subtract light

   Thus, overlaying the visuals will create "ghost-like" virtual objects

   Note: some mentioned limited FOV, too, which is a good point and got credits, too (if and only if it was appropriately stated with respect to the situation, and not a general statement like OSTs have a more limited FOV than VSTs).

e) How would the above issue(s) change if you were using a Retinal display that projects directly onto your retina?

   They solve the problem in (d), because the direct projection onto the eye is able to "block out" the real light source

   Notice that this is a very informal description, which was totally okay in the exam, since we also didn't cover this in detail in the lecture, but only discussed it informally.