(INFOMMMI) Multimodal Interaction - 8 april 2019

Course: Multimodal interaction (INFOMMMI)

Questions 1-9 cover lectures 1-4 by Peter Werkhoven (max. 60 points).
Questions 10-13 cover lectures 5-8 by Wolfgang Hürst (max. 40 points).
Note that questions 10-13 contain multiple subquestions. Thus plan your time accordingly.

Good luck!

Number of questions: 13

You can score a total of 100 points for this exam, you need 50 points to pass the exam.

Note that this file only contains the questions covering lectures 5-8 by W. Huerst (also, some of it was written in a rush, so don't take everything for granted).

10. Comparing AR with VR

Assume you are a developer of AR and VR systems. Name one aspect that is generally more difficult to achieve with AR than with VR. Give a short explanation why. *One sentence can be sufficient to get full credits. Note that this question is about system development, not usage. Thus, make sure your answer addresses an aspect that one comes across when developing it (i.e., not a feature or usage-related aspect of the system).*

a. [max. 2 pts] Aspect that is generally considered to be **harder** to achieve **in AR** than in VR:

Different correct answers exist here; see slide from first lecture and comments made in that context.

Now name another aspect that is generally more difficult to achieve with VR than with AR. Give a short explanation why. *Again, a short answer is sufficient, and make sure to focus on a development-related aspect, not a usage-related one.*

b. [max. 2 pts] Aspect that is generally considered to be **harder** to achieve **in VR** than in AR:

Same as above

In the lecture, we compared the typical "interaction loop" of a VR system ("tracking – simulation – feedback – control") with the one for an AR system. In the following, assume we are comparing two head-mounted displays; for example, the HTC Vive for the VR system, and a Microsoft HoloLens for the AR system.

c. [max. 3 pts] For **tracking**, give one example where **there is *no* real difference**; for example, an aspect of the implementation where we can use the very same algorithms in both cases (VR and AR).

Again, different possibilities exist to answer this correctly. One example would be:

In both cases, we need to track the gaze/orientation of the user's head. No real difference exists in this context between the two scenarios, which is why the same techniques and algorithms can be used.

d. [max. 3 pts] For **simulation**, give one example where **there is a difference** between VR and AR; for example, a situation where we would need to use a different algorithm to realize it in AR, or a situation that is different in VR and AR, thus requiring us to use a totally different approach.

And again, different correct solutions exist. The most obvious aspect is probably that AR combines real and virtual, thus requiring to "match" the virtual objects to the real one. Another good answer would be stating that VR needs to render the whole world, whereas AR only needs to render parts of it.

11. Comparing different AR systems

When using **head mounted displays** for augmented reality, we can distinguish between optical see-through displays (OSTs) and video see-through displays (VSTs).

a. [max. 2 pts] Give one occlusion problem that appears with OSTs but **not** with VSTs. Shortly explain your answer (*2 short sentences could be enough to get full credits*).
When a real object (or any part of the real environment) is occluded by a virtual object, VSTs can just show the virtual one. But OSTs cannot "subtract" the light coming from the occluded part of the real environment towards the eye.

b. [max. 2 pts] Give one occlusion problem that does appear with **both** OSTs and VSTs. Shortly explain how one usually deals with it (*2 short sentences could be enough to get full credits*).
When a virtual object is occluded (from the viewer's perspective) by a real one, we should not render the occluded parts of said virtual object. This can be done if we have full 3D depth information of the real world and objects.

Instead of AR with head mounted displays, we can also create so-called **handheld AR** by using our smartphone. Here, the live video stream of the phone's camera is enriched (i.e., augmented) with graphics integrated into the live video (i.e., reality).

Assume the university asks you to develop an AR information browser that gives people on campus live information about the buildings. That is, people walking around on campus can start the app on their phone, point it at a building, and then get graphical or textual information associated to all buildings that are currently shown in the live video stream.

c. [max. 3 pts] What kind of sensors from the phone would you use to implement this? Name each sensor and specify the data that you would use from it (*short phrases for each sensor could be sufficient to get full credits; no lengthy explanation is needed*).

Note that the AR app only needs to work outside. It is not expected that it works indoors. Further note that there are obviously different ways to achieve this. Make sure that your solution is sufficiently reliable and efficient (i.e., uses the minimum amount of resources of the device).

- Accelerometer / Gryroscope to get the orientation of the device
- Compass to know where the device is heading / to get the direction of the device
- GPS to know where in the world the device is located

(Listing the camera to create the AR in the first place is correct, too, but was not necessary to get full credits.)

In the lecture, we saw an approach where your smartphone is used to create an OST head mounted display by putting it into a cheap “frame”, similarly to how you can use a Google Cardboard to create a cheap virtual reality headset (to refresh your memory: the device we looked at was called *Aryzon* and used mirrors, reflective glasses, and stereoscopic lenses in a cheap frame to achieve this).

Assume a museum wants to develop an AR app that allows people using such a device to get additional information for some pieces of their exhibition. That is, people looking at, for example, an art piece via these goggles get additional graphical or textual information associated with it.

d. [max. 3 pts] Shortly state why this situation is different from the one above (i.e., the university’s AR information browser). How would you realize this app, i.e., what information would you need to display the augmented information in the app and how would you get it? *(A short explanation is sufficient. No need for lengthy elaborations.)*

Because we are inside, we need more accurate information about the room. Because we only want to use it at dedicated places (the artworks), we can create a "prepared" environment. Both reasons suggest that a computer vision-based system should be used; could be either marker-based or natural feature tracking (*both have pros and cons, but that was not asked for here*). This is sufficient to get the information which artwork we are looking at and its 3d orientation and location in the real world (*which is needed to correctly register and render the virtual parts*).

(Note: compass is no longer needed, neither is the accelerometer, since we can get the orientation of the phone with respect to the environment from the marker or natural features, and GPS is neither needed nor will it work inside).

Fun fact: A student who previously followed this course actually did this as his thesis project. He compared usability of handheld AR versus the Aryzon approach for a museum tour in the Rietveld Schroder House in Utrecht.

Now assume we want to use a smartphone with a Google Cardboard-like frame to create a VST head mounted display.

e. [max. 2 pts] Shortly explain why it might be a good idea to use an additional camera mounted on the cardboard frame for this instead of using the one integrated in the smartphone. *(1-2 sentences could be sufficient to get full credits.)*

1. Possible distortion
   (the field of view of consumer phones' cameras might not be best for such a task)
2. Eye displacement
   (the location of the camera doesn't match to the location of your eyes)

12. AR interaction

Interaction in AR (and VR) is usually done in 3D. (*Note: also read the second sub-part of the question before answering the first one*.)

a. [max. 1 pt] Give one reason why this is generally more difficult than standard human-computer interaction (e.g., with keyboard and mouse).

Various correct answers exist here (see, for example, the list of "What makes 3D interaction difficult" on a slide from the interaction lecture.

Two techniques to interact with objects that are out of reach in VR and AR are *ray casting* and *cone casting (*aka *flashlight)*.

- [max. 2 pts] What problem does cone casting solve that commonly exists with ray casting?Shortly explain (*2 sentences should be enough to get full credits*).

Small motions of the ray can result in large motions at the distance (lever arm effect), making it hard to target objects that are far away. Cones cover a larger area the further away they are from the user, thus dealing with this issue.

- [max. 2 pts] What is a potential problem or disadvantage of this approach?

Because they select all objects within the cone, it is still hard to select the right object if they are very close to each other.

- [max. 4 pts] In their paper “Augmented Reality vs Virtual Reality for 3D Object Manipulation” from 2015, Krichenbauer et al. compared task completion time for selection and transformation tasks in VR and AR. Their results show that when using a 3D input device, completion time in AR was generally lower than in VR. Shortly state what might have been the reason for this.
(*Note: you do not have to write down the details of the results here. General statements, similarly to the one used by the authors in their hypothesis, can be sufficient to get full credits.*)

See paper. Possible reasons they suggest (although not all of them are proven) are:

- *Seeing the real world in AR might give you a more direct understanding of spatial relations*
- *Hand-eye coordination might be improved due to visual feedback*
- *Participants might have preferred the different background*
- *Participants were more effective because they were more engaged*

Note: it was not required to use exactly the same phrasing as in the paper to get full credits, as long as your description matched the described aspect well enough. Also, few people came up with other reasons that seemed plausible as well and thus also gave credits.

13. Multimodal AR (smell/olfactory AR)

In the last lecture, we talked about a device produced by the company VASQO that attaches below a VR headset and produces the illusion of smell by spraying small amounts of fragrance. Now assume the company would also make such a device for an AR headset, such as MicroSoft's HoloLens. For the olfactory part of such a multimodal AR system, list each of the three criteria from Azuma's AR definition and shortly state if they would be fulfilled or not.

(*No lengthy explanation is needed, as long as it becomes clear that you understood the definition and how to apply it to this concrete context. Note that without having tested the device yourself, you might not be able to judge certain things just based on the description from the lecture and the video that we saw. If that is the case, it is perfectly okay to start your answer with something like: "Assuming this is implemented in a way that ... "*.)

Note that different answers can be correct here, depending on a) how you interpret real/virtual, interaction, etc. (we discussed in the lecture that there are different ways to interpret the general statements of Azuma) and b) how the system actually works. Every answer that was correct based on your view (and description) of the situation here gave full credits.

a. [max. 1 pt] First characteristic according to Azuma:
Combines real and virtual

b. [max. 2 pts] Is this fulfilled by the olfactory part of the system? Shortly explain your answer.
If you consider the fragrance as a "virtual" addition to the real environment, than this criteria is clearly fulfilled (esp. since the system does not filter out real world smells). Some people argued that the augmented part must be "digital" or "computer generated". This is not the point of view we took in the lectures (remember the "augmented perception" quote), but a valid definition of AR used by others, so if you explained it correctly, you still got full credit.

c. [max. 1 pt] Second characteristic according to Azuma:
Is interactive in real time

d. [max. 2 pts] Is this fulfilled by the olfactory part of the system? Shortly explain your answer.
If the system reacts (in real time) to changes in the environment or the user's location (e.g., intensified smell if you get closer to the source of the smell, decreased if you turn your head away), this criteria is fulfilled.

e. [max. 1 pt] Third characteristic according to Azuma:
Is registered in three dimensions

f. [max. 2 pts] Is this fulfilled by the olfactory part of the system? Shortly explain your answer.
Here it depends a bit on how you interpret the "location" of smell in 3d. Smell is not perceived in 3 dimensions like, for example, audio (we can hear from which direction a sound is coming in 3d, but we cannot really judge where the source of smell is unless we see it). Some people argued that way, and if done correctly, this would be a correct answer. You could however also argue that by moving your head and body, smell does change, so the perception of smell in 3d does change actually. If the system supports this (e.g., changes the intensity of smell if you turn your head in the other direction; see comments on interactivity above), the answer is yes. Some people argued this way, and again, if phrased correctly, this is a valid answer that gave full credits.