Intelligent switch:*
An algorithm to provide the best third-person perspective in augmented reality

P. Salamin, D. Thalmann, and F. Vexo
Virtual Reality Laboratory - EPFL
EPFL-IC-ISIM-VRLab Station 14
1015 Lausanne, Switzerland
email: patrick.salamin@epfl.ch
www: http://vrlab.epfl.ch

Abstract

Augmented reality (AR) environments are suffering from a limited workspace. In addition, registration issues are also increased by the use of a mobile camera on the user that provides a first-person perspective (1PP). Using several fixed cameras reduces the registration issues and, depending on their location, the workspace could also be enlarged. In this case of an extended workspace, it has been shown that third-person perspective (3PP) is sometimes preferred by the user. Based on the previous hypotheses, we developed a system working with several fixed cameras that can provide 3PP to a user wearing a video see-through HMD. Our system uses an "intelligent switch" to propose our "best view" to the user, i.e. avoiding markers occlusion and taking into account user displacements. We present in this paper, such a system, its decision algorithm, and the discussion of obtained results that seem to be very promising within the AR domain.

Keywords: Augmented reality, User context awareness, Third-Person Perspective, Best view, Video see-through HMD

Introduction

During our augmented reality experiments with a video see-through Head-Mounted Display (HMD), we always try to provide the best view to the user. As shown in [7], the best perspective depends on the performed action: first-person perspective (1PP) for manipulations and third-person perspective (3PP) for moving actions. In order to propose a 3PP to the user, we need at least a second camera that follows him/her when he/she moves in the environment. Moreover, within the framework of augmented reality, it has been proven that fixed cameras avoid lots of registration issues [2]. Finally, if there are multiple cameras, occlusion problems can also be reduced [5].

Based on the previous researches, we propose in this paper a system with several fixed cameras combined with a mobile one on the user to provide the different perspectives to the user who wears a video see-through HMD. We such a kind of system, we should have better results and provide a better comfort in almost every simulation with augmented reality. Moreover, working with several cameras allows us to enlarge the work area to a building (or at least two rooms in this paper). In order to manage this system, we implemented an "intelligent switch" that chooses the "best view" depending on the user context (location, movement, performed action, and occlusions).

*This research has been partially supported by the European Coordination Action: FOCUS K3D (http://www.focusk3d.eu).

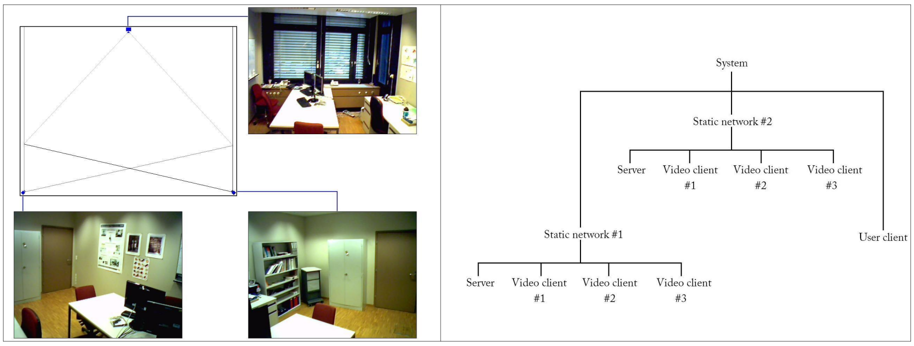

Figure 1: Left: Schema of the cameras (in blue) disposition in the room in order to cover the whole space (their angle of view is represented by the line in different colors). Each camera is coupled with a picture representing a snapshot of its video flow sent onto the network; Right: Schema of the system architecture.

Related work

A well-known way to reduce the marker occlusions consists in working with multiple cameras[3]. Indeed, even if the marker is hidden for one camera, the other cameras (due to their strategic position) should still be able to detect the marker. The second issue, the registration, is one of the main issues in AR software and may be a source of motion sickness.Notice that it has been proven in [2] that the use of a fixed camera considerably reduces this lack.

The researches cited above propose a system providing for augmented reality experiments into a static user working on a very small area like a desktop in indoor, or for a user carrying a camera outdoor[1]. Unfortunately, in this last case, users are carrying a camera, which is then not at a fixed location. And as shown above, this induces registration issues.

We will then use several fixed cameras to provide 3PP when the user is moving, and 1PP for the fine manipulation with a camera coupled on the users' HMD.

Description of the system

The goal of our system is to provide the "best view" to a user who can move in several rooms and manipulate objects in augmented reality. Based on previous studies of Salamin et al. [7][6], we know that these two actions require different perspective: third-person perspective (3PP) for navigation tasks, respectively first-person perspective (1PP) while manipulating an object with the hands. In our case, instead of having a camera that follows the user for the 3PP, we decided to have multiple fixed cameras. Consequently, the user will not need to matter about collisions of a cumbersome backpack with wall, ceiling, doors, etc.

As there are multiple cameras, we need a system that will automatically detect which camera needs to be activated for the user best view. For this simulation, we will work in an area of two adjacent rooms in which we already put three cameras at strategic positions (see left pic of Figure 1). Our system (right pic of Figure 1) considers that there are two networks of three cameras (one network per room). Our system then first have to localize the user, i.e. in which room he/she is currently. Once done, depending on the visibility of the markers and on the displacement of the user, our system will choose which video stream to provide to the user.

Video clients are linked to a TRUST Webcam (30fps at 1280x1024). They have three tasks to perform in parallel: acquire webcam video flow with the help of the Digital Signal Video Library (DSVL) and process it with ARToolKit to detect if there are visible markers; stream via RTP this JMF-processed video flow onto their network in a continuous way; and connect to the server of their network to transmit the markers visibility status (ARToolKit).

The servers also have three main tasks to perform: detect and accept all the three video clients (plus the user) of its network to receive their information about their markers visibility status;

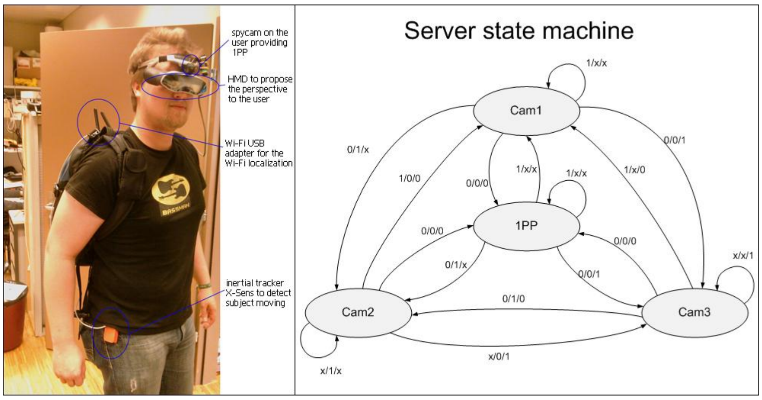

Figure 2: Left: User fully equipped; Right: Schema defining the switch of perspective in a room. The states correspond to the active camera (providing the video flow to the user) and the values on the link between them represent if they "can see" the ARToolKit marker ('1') or not ('0'). (Notice that the value 'X' means that any value can fit)

choose the best video stream to indicate to the user client; and inform user client the video stream to connect to.

The user client is composed of a notebook (Windows Vista for network tricks with a Wi-Fi antenna to connect to the network server) linked to a Trust Webcam, an XSens inertial tracker (user movements), a Wi-Fi USB adapter (user location), and a SONY Glasstron video see-through HMD (user video feedback).

Here are the main rules of our algorithm(presented in left pic of Figure 2): no movement detected by the inertial tracker on the user directly leads to choose 1PP anddetected movement leads to propose 3PP.Notice that there is a preference for a camera located behind the user in order to avoid "mirror effect" that may introduce biases like the right-left inversions.

Obviously, our system must avoid changing too many times the chosen perspective. Indeed we hypothesize that each camera change would perturb the user.

Experimentation

We tested our system with 12 naive and voluntary participants (10 males and 2 females). They were all between 20 and 35 years old. The equipment previously described is shown in right pic of Figure 2.

Protocol

There are four main steps in the experimentation described in this paper. As written above, our working space is extended to two rooms. This means that one step will be to move from a room into the other one.

Another step of this experiment is to stay in a room with no displacement and to turn on oneself (use of 1PP).

A third step consists in walking in the room (use of a chosen 3PP).

The last step is the manipulation of an augmented object. This action can be performed while walking or staying, which means a change of perspective (3PP, respectively 1PP).

All the four steps cited above are performed several times in different orders during the experimentation. The experimentation usually lasts around twenty minutes.

Questionnaire

Once the experiment performed, we proposed a SUMI-like (Software Usability Measurement Inventory) questionnaire [4] to the users. This questionnaire is composed of two parts: user profile and software evaluation (50 statements).

Results

Globally, most of users enjoyed the system. Every step was performed by every user, even if some of them needed more time to adapt to the system. They walk a lot in the rooms looking for augmented objects and trying the perspectives. We will now first present the users' answers tothe questionnaire proposed to them after the experiment.

Our adapted SUMI questionnaire was filled by every participant. Its first part, concerning the users' profile, reveal us that twenty minutes for training was widely enough all the participants but one.

The questions of the second part reveal us that our software is very accurate and fast to leave the perspective current when the augmented object disappears. But it also informs us that the reconnection to another video flow (couple of seconds) can be very perturbing at the beginning.

Our system was then considered as very attractive and intuitive enough, even if improvements can be done for a future version.

Conclusion and further works

The obtained results confirm our hypotheses. User comfort does not suffer from the changes of perspective; some of the users even play at forcing the perspective change during the experiment. Working with an augmented environment larger than a desktop seems to be very promising for future researches in this domain.

Participants, who already took part to previous 3PP experiments with a camera coupled to a backpack on their body, especially appreciate the change of perspective that avoids occlusions.

An improvement would be to reduce the time needed for the perspectives switch and improve the image quality. A solution to this problem would be to send only the video flow chosen by the system to the access point. This would allow us to send video flows with a higher resolution onto the network.

References

[1] Cheok, A. D., Goh, K. H., Liu, W., Farbiz, F., Fong, S. W., Teo, S. L., Li, Y., and Yang, X. (2004). Human pacman: a mobile, wide-area entertainment system based on physical, social, and ubiquitous computing. *Personal Ubiquitous Comput.*, 8(2):71–81.

[2] Collins, R., Lipton, A., Fujiyoshi, H., and Kanade, T. (2001). Algorithms for cooperative multisensor surveillance. In *Proceedings of the IEEE*, volume 89, pages 1456–1477.

[3] J., K., I., C., and G., M. (2003). Continuous tracking within and across camera streams. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 267–272.

[4] McSweeney, R. (1992). Sumi – a psychometric approach to software evaluation. Unpublished MA (Qual) thesis in Applied Psychology, University College Cork, Ireland.

[5] Mittal, A., Larry, and Davis, S. (2003). M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo. In *International Journal of Computer Vision*, pages 189–203.

[6] Salamin, P., Thalmann, D., and Vexo, F. (2008). Improved third-person perspective: a solution reducing occlusion of the 3pp? In *VRCAI 2008, the 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry.*

[7] Salamin, P., Vexo, F., and Thalmann, D. (2006). The benefits of third-person perspective in virtual and augmented reality? In *ACM Symposium on Virtual Reality Software and Technology (VRST '06)*, pages 27–30.