A MOBILE AR KIT AS A HUMAN COMPUTER INTERFACE FOR COGNITIVE VISION

Hannes Siegl and Axel Pinz

Institute of Electrical Measurement and Measurement Signal Processing,
Graz University of Technology, Austria,
email: siegl, pinz@emt.tugraz.at

ABSTRACT

Existing Augmented Reality (AR) applications suffer from restricted mobility and insufficient tracking (head-pose calculation) capabilities to be used in fully mobile, potentially outdoor applications. We present a new AR-kit, which has been designed for modular and flexible use in mobile, stationary, in- and outdoor situations. The system is wearable and consists of two independent subsystems, one for video augmentation and 3D visualization, the other one for real-time tracking fusing vision-based and inertial tracking components. Several AR-kits can be operated simultaneously, communicating via wireless LAN, thus enabling in- and outdoor applications of mobile multiuser AR scenarios. In the European cognitive vision project VAMPIRE (IST-2001-34401), our AR-kits are used for interactive teaching of visual active memory. This is achieved via a new kind of 3D augmented pointing, which combines inside-out tracking and 3D stereo HCI, and delivers approximate scene coordinates and extent of real objects in the scene.

keywords: 3D interaction device, real-time pose computation, augmented reality, mobile system, mobile AR

1. AUGMENTED REALITY

In contrast to virtual reality which is widely used for simulation, augmented reality (AR) enriches the perceived reality by additional information with a representation ranging from video annotation or highlighting to projections of complex 3D objects. This technique is used as visual aid for medical and military purposes, for entertainment, for assembly processes or for engineering design [1, 6] or for interactive teaching of visual active memory described in this paper.

For this purpose, AR applications must be able to recover the position and orientation of the user's head. This localization task requires high accuracy and low jitter tracking methods performing at real-time.

Existing AR applications suffer from restricted mobility and insufficient tracking (head-pose calculation) capabilities to be used in fully mobile, potentially outdoor applications.

The mobile augmented reality system (MARS) by Höllerer et al. [4] uses an inertial/magnetometer orientation tracker(Intersense) and a centimeter-level / real-time kinematic GPS position tracker which is dependent on a base station providing correction signals. Another wearable AR-kit based on GPS for position tracking and on a magnetometer for orientation tracking is Tinmith by Piekarski and Thomas [7]. Behringer et al. presented a similar approach in [2].

We present a new AR-kit, which has been designed for modular and flexible use in mobile, stationary, in- and outdoor situations. The system is wearable and consists of two independent subsystems, one for video augmentation and 3D visualization, the other one for real-time tracking fusing vision-based and inertial tracking components. Several AR-kits can be operated simultaneously, communicating via wireless LAN, thus enabling in- and outdoor applications of mobile multiuser AR scenarios.

2. AR COMPONENTS

A typical AR-kit consists of components for (self-) localization and pose recovery such as vision-based tracking or inertial tracking devices, a head mounted display (HMD) for visualization of information and most of the times a human computer interface for communication with the system providing information. In figure 1 the sketch of a system designed for the EU Cognitive Vision project Vampire is shown.

A laptop is used for rendering information and serving the HMD with the video stream captured from a stereo pair consisting of two fire-wire cameras. A custom CMOS camera and an inertial tracker are used for hybrid tracking.

Laptop and single board computer (SBC) are mounted on a backpack (see fig. 2.b) and are connected via LAN (direct link). HMD and tracking sensors are mounted on a helmet (see fig. 2.a and 2.b).

3. VISUALIZATION SUBSYSTEM

A laptop with OpenGL graphic chip (nVidia Quadro) is employed for rendering of the graphics for the head mounted displays (HMD). For prior projects we used optical see-through HMDs such as Sony Glasstron LDI-D100BE or

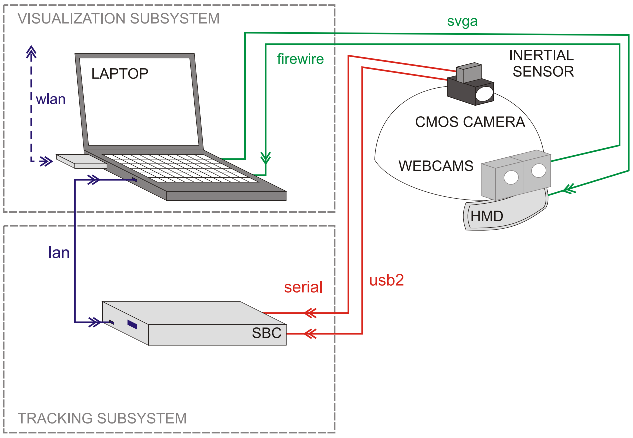

Fig. 1. AR-kit components: For visualization a high end laptop and a custom stereo video see-through HMD are employed. An inertial sensor and a custom high speed CMOS camera are used for tracking.

(a) (b) (c)

Fig. 2. AR-kit: Our custom stereo video see-through set consisting of Fire-i firewire webcams and an I-visor 4400VPD HMD(a) the backpack with laptop for visualization (b) hybrid tracking unit consisting of custom CMOS camera and an XSens MT9 (c)

LDI-D50B. Optical see-through HMDs provide the possibility to augment the visually perceived reality of the user by a video stream while he is still able to see through the HMD goggles. Unfortunately Sony has discontinued their series of optical see through HMDs, and therefore low cost off-the-shelf stereo optical see-through HMDs with comparable video quality and field of view are not available any longer.

As a consequence we developed a custom stereo video see-through set consisting of Fire-i firewire webcams and an I-visor 4400VPD HMD (see fig. 2.a).

3.1. Laptop

We use a Dell Precision M50(Pentium 4, 1.8 Mhz, nVidia Quadro4 500 GoGL) for information rendering. It communicates via LAN with the tracking subsystem and via WLAN (Cisco Aironet 350) with collaborative AR systems or other components of the VAMPIRE project related to object recognition, learning and scene interpretation via contextual information (see sect. 5).

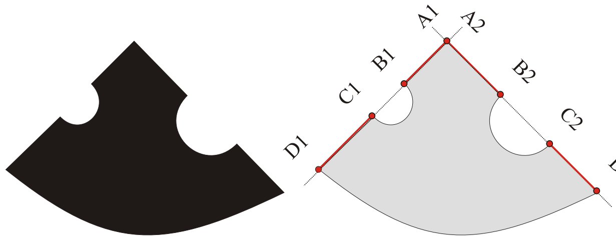

Fig. 3. This figure shows an artificial corner target [3] which is used as an intermediate step on the way to natural features. The target is identified by the perspectively invariant cross ratio of the segments on the two intersecting lines. Then the pose can be calculated by the corner positions.

3.2. HMD

The I-visor 4400VPD is able to display SVGA(stereo) in 60, 70 and 75 Hz VESA standard mode and provides 31 degree diagonal field of view per eye.

3.3. Webcams

Our stereo camera pair consists of two Fire-i CCD webcams. It is possible to grab via IEEE1394 at 30 fps (640x480, YUV 4-1-1 12 bit or gray level) with these low cost cameras. The original lens holder can be easily replaced by a standard 12x0.5 miniature lens holder (e.g. SUNEX CMT103 or CMT821) which permits experiments with different lenses. In addition, we removed one IEEE1394 interface socket on the board and connected the cameras directly.

4. TRACKING SUBSYSTEM

The tracking hardware of our AR-kits is mainly used for self-localization or inside-out tracking.

4.1. Tracking Devices

For vision-based tracking, a custom Fuga 1000 (1024×1024 pixels, 10 bit) based CMOS camera [5] with USB2 interface was implemented to gain extremely fast access of small, arbitrarily positioned image portions (see tab. 1) typically used for tracking of e.g. corners (see fig. 3).

In order to deal with fast movements of the head, vision-based tracking is fused with a commercially available inertial sensor (MT9 from XSens) by Kalman filtering.

4.2. SBC

The hybrid tracking data is processed on an Intel socket370 PIII on an Advantech PCI-9577 which seems to be a good

window side length	number of windows	requests/second
8	5	2600
8	15	2000
8	25	1300
16	5	2000
16	15	1000
16	25	660

Table 1. Request rates vs. window sizes and numbers: Request denotes a cycle consisting of window positioning and read-out.

trade off with respect to speed, power consumption and compatibility with various PCI extension boards (CAN bus, frame grabbers) we use for different experiments. The SBC is integrated in a custom aluminum case. Its power supply unit serves not only the SBC but also HMD, inertial sensors and cameras.

4.3. Power Supply

It is possible to operate our AR gear either with an AC power cord or with a battery pack consisting of Saft VH F Ni-MH cells (1.2 V nominal voltage, 13500 mAh, $\approx$ 1 hour system uptime).

5. VAMPIRE SYSTEM DESIGN

The project "Visual Active Memory Processes and Interactive REtrieval" (VAMPIRE) aims at the development of an active memory and retrieval system in the context of an Augmented Reality scenario. The AR gear provides the image data perceived from the user (stereo camera), the user's head pose (inside-out tracker) and basically the human computer interface (HCI) defining actions (e.g. query, learning, naming of objects). The VAM hierarchically maintains the acquired data (image data, head pose, object locations, gesture interpretation) from all the connected modules, tries to build contextual relations (cup – coffee – sugar) and thus provides data for connected modules itself (see fig. 4).

6. VAMPIRE APPLICATION SCENARIO

Within the VAMPIRE project, we aim at mobile indoor applications in unprepared rooms. During an offline initialization phase, the scene is analyzed for promising visual landmarks, followed by a sparse reconstruction of the scene in terms of these landmarks and their scene coordinates. Afterwards, the landmarks are used for online real-time tracking of camera / head pose. The user receives visual feedback using the stereo head-mounted-display (HMD), so that the real scene can be augmented by virtual content which is perceived in 3D, in perfect registration with the real scene. Teaching of the VAM as well as interpretation of recognition results requires several new modalities of user-system interaction, some of them are outlined below:

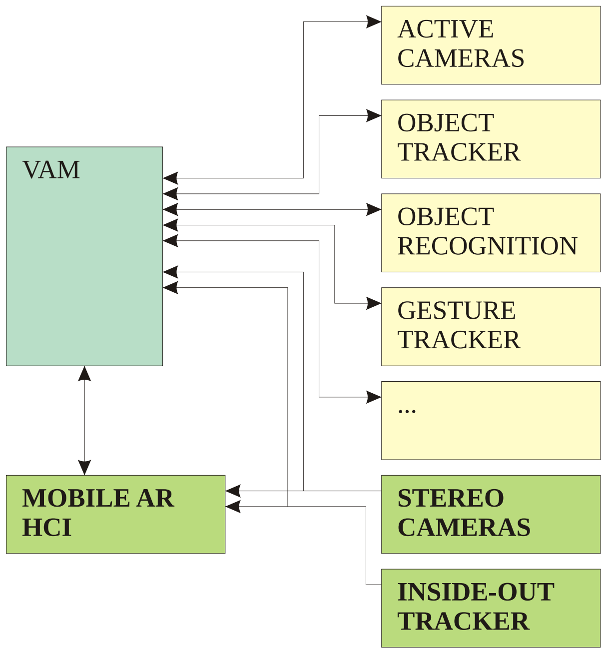

Fig. 4. Vampire System (functional sketch): The visual active memory (VAM) manages and interprets data collected from various modules. The modules AR HCI, stereo camera pair and inside-out tracker for self-localization are for technical reasons physically and functionally closer connected.

- **Pointing at objects in 3D**
  A 3D cursor was implemented which is basically operated by mouse wheel and buttons. It exploits disparity and object size to generate the perception of distance which allows – together with the head pose obtained from the tracking subsystem – to compute an estimate of an objects size and its position in the room (see fig. 5). This information can be used to support segmentation of training objects as well as initiate query processes e.g. with stationary active cameras, especially in cluttered scenes. The integration of pointing gestures will yield a more natural feel for simple scenes in the future. In figure 6 an experimental verification of this concept is outlined. 3 users placed the 3D cursor next to 3D points in the office (distance=$0.5, 1, .., 3$ m) two times. The mean disparity (see fig. 7) obtained with this subjective localization experiment is not linear as it theoretically should be. Especially for the shortest distance (0.5 m) a high (relative) standard deviation ($\sigma/\mu = 8\%$) could be observed due to differences in anatomy (for all the other inspected distances: $\sigma/\mu < 1\%$).

- **Localization of contextually related objects**
  A pointing gesture or the selection of an object with the 3D cursor can trigger a query for other objects having a contextual relation indentified by the VAM,

e.g. coffee machine – coffee –sugar – cup. Then the system helps to find these associated objects.

- **Highlighting of recognized objects**
  The head-mounted camera of the mobile AR-kit transmits a video stream which will be analyzed by object recognition processes. Once an object has been recognized, this can be reported to the mobile user on demand, and the object can be highlighted.

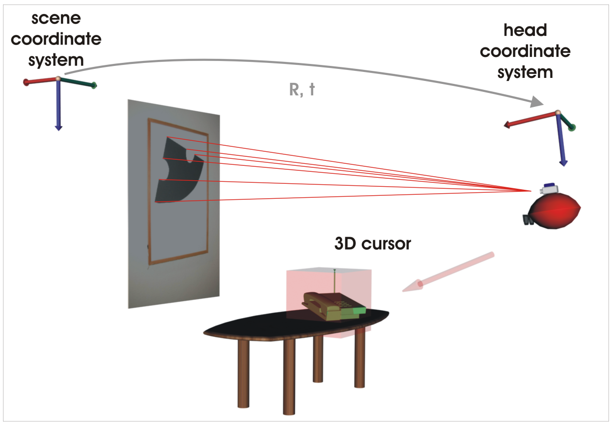

Fig. 5. Pointing with 3D cursor (sketch): The tracking system processes the corner features of the framed image for self-localization. The selection of the phone with a 3D cursor allows to estimate the position of the object.

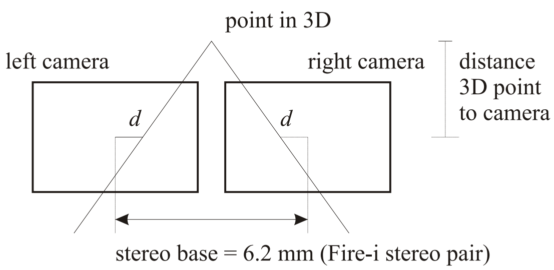

Fig. 6. Sketch of the experimental setup to determine the relation between 3D point distance and disparity

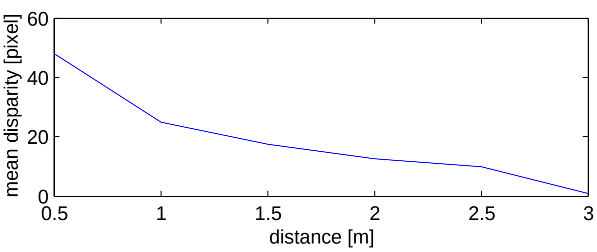

Fig. 7. Mean disparity d vs. distance experimentally obtained at the department (3 users, 2 tries each)

7. DISCUSSION AND OUTLOOK

We presented the mobile AR gear which is employed as human computer interface for the cognitive vision project VAMPIRE which tries to model human memory processes in an office environment and a new AR 3D cursor for pointing. At the moment our system is designed for a modular and flexible use in experimental environments and situations. Especially as far as hybrid tracking performed on the SBC is concerned, our architecture allows an easy integration of various components, as it provides a number of interfaces (3 serial ports, 1 parallel port, PCI interface, 2 USB ports, 5 connectors 5/12 V DC for external hardware, LAN, audio in). This flexibility has a strong impact on size and weight of the AR gear, although the technological advance in mobile computing would already allow a much more compact design.

Acknowledgement

This research was funded by *VAMPIRE* Visual Active Memory Processes and Interactive REtrieval (EU-IST Programme IST-2001-34401).

8. REFERENCES

[1] W. Bardfield and T. Caudell (eds.), *Fundamentals of wearable computers and augmented reality*, Lawrence Erlbaum, 2001.

[2] R. Behringer, C. Tam, J. McGee, S. Sundareswaran, and M. Vassiliou, *A wearable augmented reality testbed for navigation and control, built solely with commercial-off-the-shelf (cots) hardware*, Proc. of ISAR 2000 (Munich, Germany), October 2000, pp. 12–19.

[3] M. K. Chandraker, C. Stock, and A. Pinz, *Real-time camera pose in a room*, 3rd Intern. Conference on Computer Vision Systems, ICVS 2003 (Graz, Austria), April 2003, pp. 98–110.

[4] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallawa, *Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system*, Computers and Graphics **23** (1999), no. 6, 779–785.

[5] U. Mühlmann, M. Ribo, P. Lang, and A. Pinz, *A new high speed CMOS camera for real-time tracking applications*, submitted to: ICRA 2004, New Orleans.

[6] Y. Ohta and H. Tamura (eds.), *Mixed reality: Merging real and virtual worlds*, Springer-Verlag, 1999.

[7] W. Piekarski and B. Thomas, *Augmented reality with wearable computers running linux*, 2^{nd} Australian Linux Conference (Sydney), January 2001, pp. 1–14.