Software Systems for Vision-Based Spatial Interaction

Jason Corso, Guangqi Ye, Darius Burschka and Gregory D. Hager,
Department of Computer Science, Johns Hopkins University

The VICs project is exploring the development of modular, adaptable software systems for vision-based, human-computer interaction. The key idea of this approach is to use local visual interaction cues (VICs) on a video stream shared between the user and the machine. A VIC consists of a graphical representation (e.g. an icon) superimposed on the video stream (thus visible to the user), associated image processing algorithms for activating the cue, and other application-specific code. The video stream could be monocular or stereo, enabling 2-D and 3-D interaction and may be combined with speech or haptics to provide enhanced interaction capabilities.

VICs are intended to be used in situations where large-scale spatial motion, particularly hand-eye coordinated motion, is essential. For example, manipulating large volumes (visualized graphically) can be done by annotating the model with icons that can be grasped and released. Interaction with real physical systems is also possible. For example, a surgeon viewing a retinal image through a stereo microscope may lay out icons on the surface of the retina to mark areas of damage. These icons may be moved and manipulated using the gestural cues (now performed with surgical tools), and may ultimately be used to target therapeutic drugs or other interventions.

VICs were designed around local (in the visual field) visual cues to avoid the problem of general tracking of human motion. This approach strongly limits image processing, allows that image processing to be dynamically driven by the interaction, and it provides an easily parameterized, modular basis for software system development. Another important notion in VICS is that the user is intentionally attempting to interact with the system. To this end, training is performed to recognize intentional gestures by the user (e.g. pressing a button) vs. unintentional motions (e.g. moving a hand over a button). Finally, the idea of statically typed, time-invariant behavior specification is being explored as a software basis for VICs and their compositions. Canonical libraries of vision algorithms for different classes of VICs are being developed.

To date, we have reported on two specific aspects of VICS. For systems that paint an interface on a surface, a recent report [1] details methods for fast, direct stereo for surface registration and tracking. For the specific case of a planar surface, we have shown that we can register the incoming streams through slew rates exceeding 1350 deg/sec and translations of up to 3 m/s using less that 20% of a standard PC. Thus, we can maintain a consistent view during normal user head motion. Once two incoming data streams are registered, a foreground/background segmentation is used to drive training and subsequent application of a hidden Markov model for recognizing gestures. We have shown [2] that it is possible to train HMMs to detect button press gestures with more than 98% accuracy and with extremely low (< 0.5%) false positive rates.

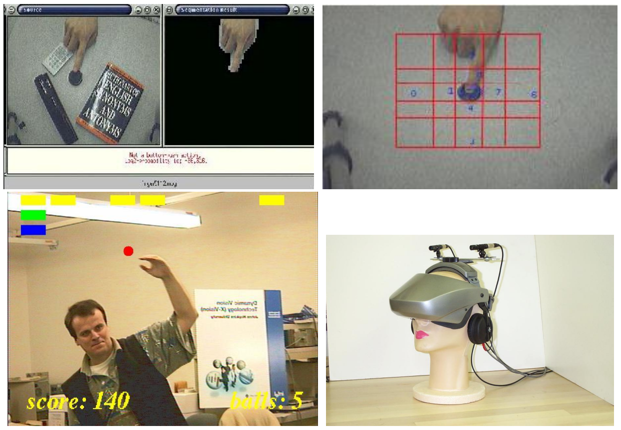

Figure 1: The upper row illustrates how images are segmented and HMM training and recognition is performed on an abstracted location grid. The lower row shows a dynamic 2D VICs system and our prototype 3D acquisition and display system.

We are developing two classes of demonstration systems. The first uses the concept of a video mirror to implement a two-dimension version of VICS. The second class of systems use an HMD combined with a head-mounted stereo camera to create a three-dimensional interface. Video of preliminary VIC systems can be found at our WEB site (`http://www.cs.jhu.edu/CIRL/projects/`).

References

[1] Jason Corso and Gregory D. Hager. Planar surface tracking using direct stereo. Technical report, The Johns Hopkins University, 2002. CIRL Lab Technical Report.

[2] Guangqi Ye and Gregory D. Hager. Appearance-based visual interaction. Technical report, The Johns Hopkins University, 2002. CIRL Lab Technical Report.

This research is supported by NSF ITR grant 0112882. Any opinions, finding, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.