Published in the CHI'99 adjunct proceedings, p. 13-14, Copyright © 1999 by the ACM

CoMedi: Using Computer Vision to Support Awareness and Privacy in Mediaspaces

Joëlle Coutaz, François Bérard, Eric Carraux, William Astier
CLIPS-IMAG, BP 53
38041 Grenoble Cedex 9, France
+33 (0)4 76 51 48 54
{Joelle.Coutaz, Francois.Berard}@imag.fr

James, L. Crowley
INRIA Rhône-Alpes
655 avenue de l'Europe
38330 Montbonnot St-Martin, France
+33 (0)4 76 61 53 96
jlc@imag.fr

ABSTRACT
CoMedi is a mediaspace prototype that uses computer vision to provide new solutions to the problems of visual discontinuity and privacy. CoMedi includes a robust face tracker based on cooperation of multiple vision techniques, a tele-exploration tool based on a multi-resolution fovea, and new privacy filter using eigen-space coding.

Keywords
Computer mediated communication, mediaspace, awareness, privacy, computer vision, fovea, eigen-space.

INTRODUCTION
Mediaspaces are supposed to facilitate informal communication and group awareness while assuring privacy protection. However, low bandwidth communication is a source of discontinuity resulting in a loss of peripheral awareness. Privacy is often implemented as an accessibility matrix coupled to an all-or-nothing exposure of personal state. CoMedi is a mediaspace prototype that addresses these problems using a combination of Computer Vision and Speech. We first present the overall design of CoMedi, then discuss the original contributions in more details.

COMEDI: OVERALL DESIGN RATIONALE
The design of CoMedi is based on technical, functional and interactional requirements. In order to reach a critical mass of users for realistic social studies, a technical requirement for CoMedi is to accommodate heterogeneous platforms (MacOs, Irix, Windows NT). To this end, CoMedi has been built using Java despite differences in performance and appearance on different platforms. The functional and interactional requirements have led us to organize the user interface of CoMedi around the porthole metaphor enriched with an optional fisheye presentation. A slot of the porthole displays the personal information that the corresponding remote user has accepted to make observable through an accessibility matrix coupled with publication filters. This information can include name, video scene, level of availability, and a "message of the day". When the fisheye feature is enabled, selecting a slot using either the mouse or a spoken command, provokes an animated distortion of the porthole that progressively enlarges the selected slot.

The motivation for a fisheye porthole is three-fold: It supports lightweight glancing. No explicit action is required from the user except looking at the porthole. The fisheye technique provides detailed rendition in context: the selected slot is enlarged in a way that details about the remote activity are revealed without loosing awareness of the activities in the peripheral slots. The porthole promotes scalability: up to sixty slots can be accommodated appropriately on a 17" high resolution screen. For a larger number of users or a smaller screen, the porthole can be amplified with holophrastic rendering techniques (in-place expansion and contraction of information). Further-more, users can build their own porthole, keeping their "most important" people in the mosaic. In addition to social customization, reducing the number of slots minimizes the need for local computer memory and network bandwidth.

VISUAL DISCONTINUITY
Although perceptual bandwidth may be unimportant for loosely coupled activities, it becomes vital for conjoint actions such as V-phone connections and tele-explorations. Multiple views on remote sites improve the information bandwidth of a single static channel, but require the user to link different views [3, 7]. Another approach is the concept of "Virtual Window' in which a remote camera is slaved to a person's head movements [4]. However, the lack of robustness of the vision technique developed in [4] (image differencing, assumption of upright head angle) imposes restrictions on the user. In CoMedi, we have designed Fovea to resolve the problem of discontinuity apparent in Extra-Eyes [7]. In addition, we have developed a face tracker that integrates multiple computer vision techniques to provide the robustness and efficiency imposed by human requirements.

Fovea
An image from Fovea is shown in figure 1. As in Extra-Eyes, Fovea is inspired by the architecture of the human retina. A foveal image, provided by a high-resolution steerable camera, is blended with a peripheral image given by a low resolution fixed camera, using alpha channel coding. The user may choose a circular or rectangular blending function for the foveal image. The foveal image may be displaced and zoomed using the mouse.

In preliminary user studies, six users were asked to find random targets. Users unanimously stated a preference for

Published in the CHI'99 adjunct proceedings, p. 13-14, Copyright © 1999 by the ACM

the circular foveal image. However, performance was most efficient with a rectangular fovea, less efficient with Extra-Eyes and least efficient with a circular foveal image.

Figure 1. On the left, the zoomed-in circular fovea; on the right, the rectangular fovea with a low zoom factor.

The Face Tracker

The face tracker developed for CoMedi uses a pan-tilt-zoom color camera. Tracking employs an architecture in which a supervisor activates and coordinates three complementary visual processes: eye blink detection, skin color detection, and cross-correlation [2].

Eye blink detection provides an estimated position of the face in the image which can be used to initialize procedures for skin detection and cross correlation. The skin detector uses a 2-D histogram of luminance normalized color to generate an image in which pixels represent the probability of skin. Skin regions are detected by thresholding and connectivity analysis. The probability that a region is a face is estimated using the size, position and aspect ratio of the resulting regions. Color skin detection is computationally cheap but sensitive to camera noise resulting in jitter. A more precise, but fragile detection is provided by cross correlation.

Cross-correlation compares a reference template (e.g., piece of an eyebrow) to an image neighborhood by computing an inner product at each position within a search region. The maximum of the inner product corresponds to the detection of the reference template. While cross-correlation within a search region is fast and accurate, it will fail when the head turns or moves too fast. In such cases the system recovers the face using color skin detection and then reinitializes tracking using blink detection.

The robustness and speed of the face tracker supports “natural” displacements. The user can move away from the terminal to show, for example, information from the local environment to distant observers. Because the tracker supports mobility, the mouse and the keyboard may not be reachable. Speech recognition can then be used as a functionally equivalent modality to control the mediaspace.

PRIVACY

In current mediaspaces, privacy has been addressed using either an accessibility matrix or video filters. CoMedi exploits both techniques in conjunction. Available video filters include low resolution images [6], shadowing [5], temporal difference images, and the Eigen Space filter.

Eigen Space filter

In [1], we present an image coding technique for images using principal component analysis (PCA): an orthogonal set of basis images are determined by PCA of a large set of "socially correct" images. Live images are coded by computing the inner product with the basis images. An interesting property of Eigen Space coding is that only information within the original image set will be captured by the coding and reconstructed in the resulting images. For example, in Figure 2, the source image (left most) is showing François with his finger in his nose. This socially incorrect gesture does not belong to the basis space and is not displayed in the reconstructed image (middle left). Similarly, persons appearing in the background are not published to distant observers if not present in the basis.

Figure 2. Eigen space filtering for private video space.

Although the Eigen space filter provides a very efficient means for hiding unexpected and private events, it relies on the proper and ethical use of the basis image set. For example, in Figure 2, Francois (middle right) is using Fred’s basis set and is seen at the other end as Fred (right most). The blurry aspect of the reconstructed image can help the remote observer to detect the deception. However, the blurring may also be attributed to the low bandwidth of the network and the substitution would not be discovered.

CONCLUSION

CoMedi has been in use for six months between the ten members of the development team located at two sites. We are now in a position to deploy CoMedi and to conduct realistic social and usability studies.

ACKNOWLEDGMENTS

This work has been supported by France Télécom-CNET.

REFERENCES

1. Coutaz, J., Crowley, J. and Bérard F. Eigen space Coding as a Means to Support Privacy in CMC, *INTERACT’97* (1997), Chapman & Hall Publ.
2. Crowley, J. L. and Bérard F. MultiModal Tracking of Faces for Video Communications, in *Computer Vision and Pattern Recognition CVPR '97*, IEEE Publ.
3. Gaver, W., Sellen, A., Heath, C. and Luff P. One is not Enough: Multiple Views on a MediaSpace, in *Proc. INTERCHI'93* (1993), ACM Press., 335-341.
4. Gaver, W., Smets, G. and Overbeeke K. A Virtual Window on a Media Space, in *CHI'95*, 257-264.
5. Hudson, S., and Smith, I. Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems, *CSCW’96*, 248-257.
6. Lee, A., Girgensohn, A. and Schlueter K. NYNEX Portholes: Initial User Reactions and Redesign Implications, in *Proc. GROUP’97*, ACM Press.
7. Yamaashi, K., Cooperstock, J., Narine, T. and Buxton, W. Beating the limitations of Camera-Monitor Mediated Telepresence with Extra Eyes, *CHI’96 (1996)*, ACM Press, 50-57.