DOVE: Drawing over Video Environment

Jiazhi Ou, Xilin Chen, Susan R. Fussell, Jie Yang
School of Computer Science
Carnegie Mellon University
Pitsburgh, PA 15213 USA
{jzou, xlchen, sfussell, yang+}@cs.cmu.edu

ABSTRACT
We demonstrate a multimedia system that integrates pen-based gesture and live video to support collaboration on physical tasks. The system combines network IP cameras, desktop PCs, and tablet PCs (or PDAs) to allow a remote helper to draw on a video feed of a workspace as he/she provides task instructions. A gesture recognition component enables the system both to normalize freehand drawings to facilitate communication with remote partners and to use pen-based input as a camera control device. The system also embeds some tools, such as controlled video delay, gesture delay, and remote camera pan-tilt-zoom control. The system provides a software environment for studying multimodal/multimedia communication for remote collaborative physical tasks

Categories and Subject Descriptors
H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces - *Collaborative computing, Computer-supported cooperative work.*

General Terms
Design, Human Factors.

Keywords
Gestural communication, gesture recognition, video stream, video conferencing, video mediated communication, computer-supported cooperative work, multimodal interaction

1. INTRODUCTION
DOVE (Drawing Over Video Environment) is a system to support multimodal/multimedia communication during collaborative physical tasks—tasks in which two or more people interact with real objects in the 3D world. Collaborative physical tasks play an important role in many domains, including education, industry, and medicine. For example, a remote expert might guide a worker's project, or a surgical expert might assist in a medical procedure at another location. Because the expertise required to perform collaborative physical tasks is becoming increasingly distributed, there is a critical need for technologies to support their remote accomplishment. Despite this need, however, the majority of previous systems for remote collaboration have been designed to support activities that can be performed without reference to the external spatial environment. Consequently, these systems have limited application in contexts in which physical objects play a key role. Our goal is to allow remote collaborators to communicate about their physical world through speech and gesture with the same ease as they can do so when co-located. And it can be easily incorporated into existing video conferencing systems.

The system we will demonstrate supports remote interaction using gestural communication over video streams using video cameras, tablet PCs, PDAs, and desktop PCs. The system allows collaborators to share the workspace through video connections. It also provides remote support for pointing and representational gesture by overlaying pen-based gestures on video streams.

DOVE further provides support for gesture recognition, both to enhance interpersonal communication and as a camera control device. Unlike existing gesture recognition systems used for human computer interaction, which support recognition only of predefined gestures, our system supports recognition of predefined gestures, freehand drawing, and a combination of the two.

2. SYSTEM DESCRIPTION
DOVE facilitates gesturing over video within the context of collaborative physical tasks where two or more people interact with real objects in the 3D world. In this type of task, some participants are co-located with task objects in the workspace ("workers"), whereas others participate from a distance ("helpers"). In our initial experimental setup, pairs are collaborating to build a large toy robot; however, the technology can be generalized to any type of collaborative task, such as telemedicine or distance education, in which a remote party needs to refer to physical objects in a workspace.

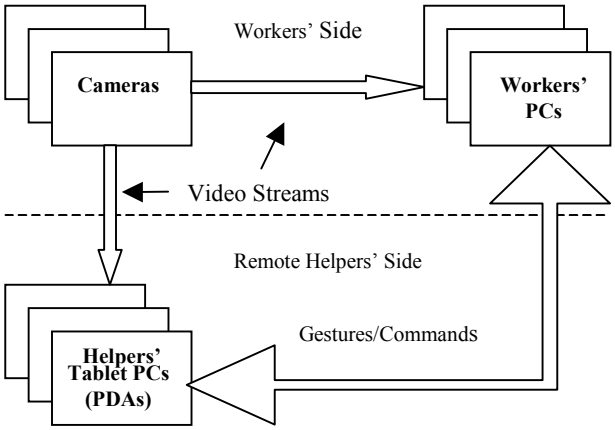

Figure 1. Overview of system architecture.

The DOVE architecture is shown in Figure 1. The workspace is visually shared through video cameras and equipped with tablet PCs, PDAs, desktop PCs or other handheld devices. Real-time video streams from the cameras are sent to collaborators'

MM'03, November 2-8, 2003, Berkeley, California, USA.
ACM 1-58113-722-2/03/0011.

computing devices. Remote participants can make freehand drawings and pen-based gestures on the touch sensitive screen of a computing device, overlaid on the video stream, just like using a real pen on a piece of paper in a face-to-face setting. The results are observable by all collaborators.

To avoid potential delay caused by a centralized video server, we use network IP cameras, each of which is a server and connected to the network independently; other computers on the network can be its clients. Once started, a network IP camera opens a TCP/IP port and waits for its clients. When a connection is established, the server's status message and the client's authentication messages will be exchanged. If the client is authenticated, video data will be sent in JPEG format upon a client's image request message. By using this technique, the video flow and the process overhead is shared by all network IP cameras.

After connecting to network IP cameras, the communication among collaborators' computing devices is also in client-server mode. For example, the worker's computer can be a server and the helper's computer can be clients. A socket is created on the worker's computer. It waits and accepts client sockets from the helper's computer. After the establishment of a connection, a helper can send remote gestures and commands through socket communication, or vice versa. The trajectories of freehand drawing and gesture recognition results are observable on all collaborators' monitors. An example of DOVE in use in a robot construction task is shown in Figure 2.

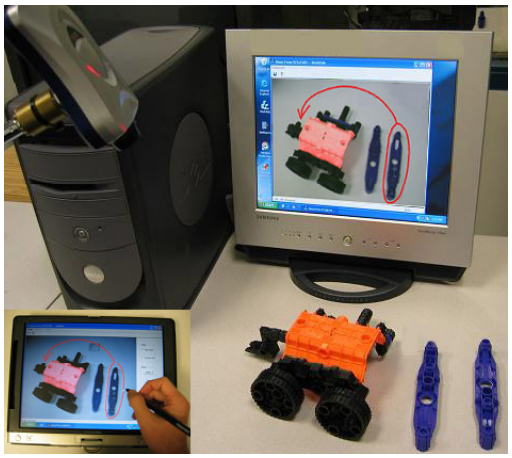

Figure 2. An example of DOVE. Worker's side is equipped with a camera and a PC. The remote helper's tablet PC, on which gestures are drawn, is shown in the lower left insert.

2.1 DOVE Drawing Modes

Pen-based gestures consist of sequences of points, starting when the pen touches the screen and ending when the pen is lifted. When the helper is drawing, the sequence of points will be added to a link list of the current gesture and sent to the worker's computer simultaneously. DOVE users can choose among three drawing modes: freehand drawing, gesture recognition, or drawing normalization. In *freehand drawing mode*, what the helper draws will appear on the worker's monitor exactly as drawn. Examples of freehand drawings from a preliminary user study are shown in Figure 3.

In *gesture recognition mode*, a predefined gesture will be recognized and a specified command will be executed. For example, a straight arrow could send a command to move the camera, indicating the direction and length of movement, whereas curved arrows might send requests to zoom the camera in and out. DOVE's gesture recognition system can support both of these functions.

Arrows Circles Sketches

Figure 3. Sample of gestures created by participants during a collaborative robot assembly task.

In *gesture normalization mode*, the current sequence of points will be sent to a gesture recognition module immediately after the user lifts the pen from the screen. By recognizing and normalizing shapes such as lines, arrows and circles and presenting the normalized images on helper and worker's computers, gesture recognition is expected to facilitate communication between remote partners (Figure 4).

Figure 4. Remote gesture recognition in a robot building task. A freehand oval and arrow (left, center) have been regularized by the recognition component (right)

2.2 Pseudo Video and Pseudo Gesture Delay

Because jitter is more likely to happen in an Internet environment, we established a wireless local area network (LAN) for our preliminary tests of the system. However, in real world environments network jitter [1] may be significant. To address this problem, we implemented pseudo video delay on helper's side and pseudo gesture delay on worker's side. In pseudo video/gesture delay, each frame/gesture is stored in a buffer and doesn't show up until a few seconds after its arrival. Experimenters can specify how long each delay is and investigate the impact of jitter through user studies.

2.3 Other Features

The current DOVE prototype provides users with five additional capabilities. First, users can set parameters for their sketches, including pen width and drawing color. Second, a set of buttons allows users to erase all gestures, their first gesture, or their latest gesture. Third, the user can specify an "automatic erase" mode, in which gestures disappear after a predefined time. Fourth, we implemented an "undo/redo" function so that a user can always undo the last action (i.e., drawing or erasure) or redo what is undone. Finally, we provided a 'snapshot' function that allows the user to keep an image at any time as a JPEG file on the local disk.

3. REFERENCES

[1] Gutwin, C., & Penner, R. (2002). Improving interpretation of remote gestures with telepointer traces. *Proceedings of CSCW 2002*. (pp.49-57). NY: ACM Press.

[2] Sezgin, M., Stahovich, T., & Davis, R. (2001). Sketch based interfaces: Early processing for sketch understanding. *Proceedings of PUI-2001*. NY: ACM Press.