Shared Visualization Spaces for Environment to Environment Communication

Hamed Pirsiavash, Vivek Singh, Aditi Majumder, Ramesh Jain
{hpirsiav, singhv, majumder, jain} @ics.uci.edu
Department of Computer Science, University of California Irvine, Irvine CA

ABSTRACT
In this paper, we will describe the creation of shared visualization spaces which provides a natural means of sharing, interacting and working with the collection of media artifacts, which may not exist in totality in any one physical environment. The idea is to create a virtual collaborative space which takes inputs from multiple physical spaces. Each user environment has a projector-camera system. The projector shows the Shared Visualization Space in each user's environment and each user can make changes to it using natural hand gestures which are being captured by a camera.

Keywords
Environment to environment communication, hand gesture recognition, projector-camera systems.

1. INTRODUCTION
The importance of newer, enhanced tele-communication methods cannot be overstressed. We see increasingly many applications like tele-medicine, collaborative art performances, virtual social gatherings and office collaboration which require users to interact across physical spaces. However, the current state of technology only allows users to interact either via very structured email/ document based methods, or forces them to undertake video-conferencing.

However the video conferencing systems typically just connect two devices and in the process restrict user physical movement due to line of sight, field-of-view, coverage range etc. Further, they also do not provide ready mechanisms for sharing of information in multiple modalities (images, text, notes, audio etc.) or collaborating in a natural manner across the entire user environment. Thus users typically end up sitting at specific locations and converting most of the useful data into text.

It is such dominance of the medium over message which prompted McLuhan to state 'Medium is the message' [1]. However, with the impending ubiquity of sensors, we aim to move the focus back to the users. Thus the sensors are made to fulfill user requirements rather than the other way round. Thus our goal is to intelligently organize multiple sensors around users so as to allow natural experience sharing across connected environments. Obviously the information capture and presentation need to be mediated by a sentient information system which understands the events happening in the connected environments and hence performs the best routing, organization and transfer of the relevant information. Thus the overall idea (refer Figure 1) is to allow users to interact naturally in their regular *environments*, without being constrained by any devices. Over a longer period, the vision is to create a web of such connected environments, wherein user experiences are created, captured, and shared in an event-driven framework. The first two nodes of such an Event-web have been setup across two buildings in University of California, Irvine and two more nodes are currently being setup at National University of Singapore and Eurecom, France [2].

In this paper, we discuss an important component of the E2E systems which deals with providing natural means of sharing, interacting and working with the collection of required media artifacts, which may not exist in totality in any one particular environment. This work is inspired by the observation that one of the biggest drawbacks of the current office collaboration systems is that often *the critical media components just exist in separated silos in each environment* and are rarely transmitted or shared across different user environments.

For example, in a telemedicine application, it may be important for the patient, nurse, MRI technician, physician and the specialist to share their respective accounts, notes, X-ray images, MRI data, and expert inputs onto a common collaborative space, which may not exist in reality in any one particular environment. While the individual components may have existed in parts in multiple environments across the whole globe, their real power and utility lies in the combination and the ability of all users to visualize it together in a shared virtual space. Further, for the visualization of such a common space to be useful, each user should be able to share his/her media artifacts, and be able to modify, augment or comment on other shared media in a natural and intuitive manner.

Hence, in this paper we describe a projector-camera based method for creating such joint shared visualization spaces wherein users can see the currently active visualization space using a projector and add, remove, and edit media artifacts to it by using natural hand gestures.

2. RELATED WORK
Other related work to ours includes Microsoft's surface computing system [3] which makes the interaction between users easier, but still all users have to present in one environment physically and interact through some specific equipment. Hence, it is not a suitable system for environment to environment connection applications.

Some specific equipment e.g. Wii remote, RFID trackers etc. can also be used for user interaction in different scenarios, but we wanted the user interaction to be 'natural' in the sense that they do not need to use any external devices.

Zhang at al., [4, 5, 6] have done some research in connecting two projector-camera whiteboards so that users from different environments can interact with each other through their writing on the board. Basically, they have introduced some algorithms to cancel the visual echo which occurs when they make a loop between two projector-camera systems. That system is very interesting, but the shared component is limited to what is handwritten on the boards. While we do not rule out text, we also want to include other media like printed brochures, X-ray, MRI, books, etc.

Hand gesture recognition research has made significant progress in recent years (e.g. [7, 8]). Typically an image of the hand with a particular gesture is used and the system tries to recognize it. The main application of these algorithms is in human computer interaction systems. Here also we are using it to make the collaboration easier and natural, but extending the ideas to connect across multiple environments.

3. IMPLEMENTATION

The overall block diagram of the system is illustrated in Figure 2, and a sample image of each class of hand gesture is shown in Figure 3. A brief description for each block is as follows:

3.1 Hand Silhouette Detection

A Shared Visualization Space Model (SVSM) maintains the record of currently shared media artifacts by each user as well as their access and ownership rights. The projector in each environment takes the input from the shared visualization model and shows it. The camera present in the environment captures an image of the scene. From the SVSM, the system knows what should be shown on the projector and be present in the scene if no user interaction is taking place. Also, we know the photometric and geometric calibration parameters of the projector and camera; hence, we can synthesize to create the image that we expect the camera to capture. Since the user may show a hand gesture on the board, the synthesized image is not exactly similar to the captured image. Therefore, by subtracting these two images, we can get the silhouette of hand.

3.2 Alignment

We find the bounding box of hand silhouette image and crop it. Then we map silhouette to an ellipsoid using PCA to find its orientation and rotate it to make it up-right. This allows our system to be rotation invariant.

3.3 Feature Extraction

We extract three features from the hand silhouette:

1. (Number of pixels on hand image) / (Number of pixels on the contour of hand image)

2. (Number of pixels on the right half of bounding box) / (Number of pixels on the left half of bounding box)

3. (Number of hand pixels on the upper half of bounding box) / (Number of hand pixels on the lower half of bounding box)

3.4 Gesture Recognition

In the current implementation, we have restricted ourselves to four different classes of gestures (refer Figure 3), and have implemented an algorithm based on a *neural networks* running on the features mentioned above to detect the different gestures. The trained neural network gave accuracy of more than 95% for the considered gesture classes.

3.5 Interacting with the System

Once the user gestures are understood by the system, they are used to undertake procedures like capturing a new media artifact or choosing a media artifact to move or remove. So these actions allow natural user interaction in the shared visualization space.

3.6 Updating the SVSM

Finally, the SVSM is updated based on new information coming from the current environment and all users across different environments can see it promptly projected in their environment. Hence, all projectors are showing the same virtual space and each camera in each environment is capturing that user's gesture and processes it and updates the SVSM accordingly.

In the current implementation, the geometric calibration to find the projector area is done using a calibration pattern shown once. The photometric calibration step is circumvented by capturing a frame without hand gestures and using it as the synthetic image shown in the block diagram. We are also currently running the system on just one client but it is easily extendable to a number of clients working on the same model.

4.CONCLUSION

We have described a projector camera system which lets users connect from different environments to share their media artifacts in a shared virtual space and be able to change it interactively using their hand gesture and experience a better and natural collaboration.

5. REFERENCES

[1] Marshall McLuhan and Q. Fiore, "The Medium is the Massage: An Inventory of Effects" pub. Penguin Books, 1967.

[2] Vivek Singh, Hamed Pirsiavash, Ish Rishabh, Ramesh Jain, "Towards Environment-to-Environment (E2E) Multimedia Communication Systems," ACM International Workshop on Semantic Ambient Media Experience, Vancouver, 2008.

[3] Microsoft Surface computing, http://www.microsoft.com/SURFACE/index.html.

[4] Manning Zhou, Zhengyou Zhang, Huang, T., "Visual echo cancellation in a projector-camera-whiteboard system," International Conference on image processing (ICIP'04), Urbana, IL, USA, 2004.

[5] Miao Liao, Mingxuan Sun, Ruigang Yang, Zhengyou Zhang, "Robust and Accurate Visual Echo Cancelation in a Full-Duplex Projector-Camera System," Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), 2006.

[6] Zhengyou Zhang , "Computer Vision Technologies for Remote Collaboration Using Physical Whiteboards, Projectors and Cameras," Proceedings of the Computer Vision for Interactive and Intelligent Environment (CVIIE'05), 2005.

[7] Moaath Al-Rajab, David Hogg, and Kia Ng, "A Comparative Study on Using Zernike Velocity Moments and Hidden Markov Models for Hand Gesture Recognition," Lecture Notes in Computer Science 5098, pp. 319–327, 2008.

[8] Vassilis Athitsos and Stan Sclaroff, "An Appearance-Based Framework for 3D Hand Shape Classification and Camera Viewpoint Estimation," Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002.

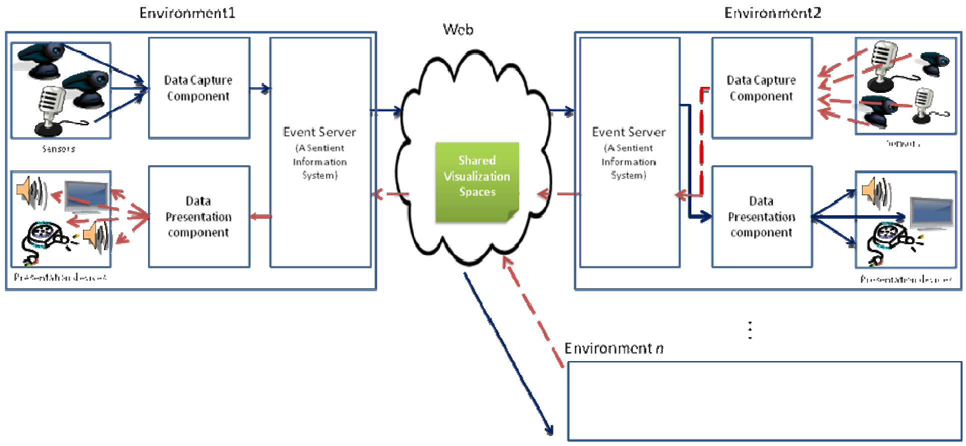

Figure 1: Overview of E2E communication systems

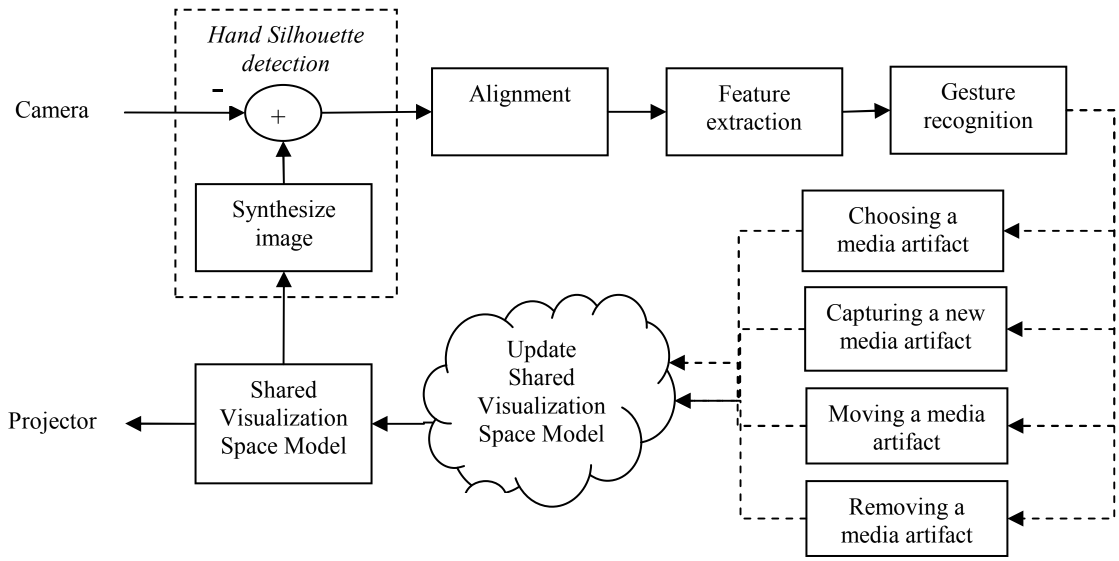

Figure 2. The overall block diagram of the system

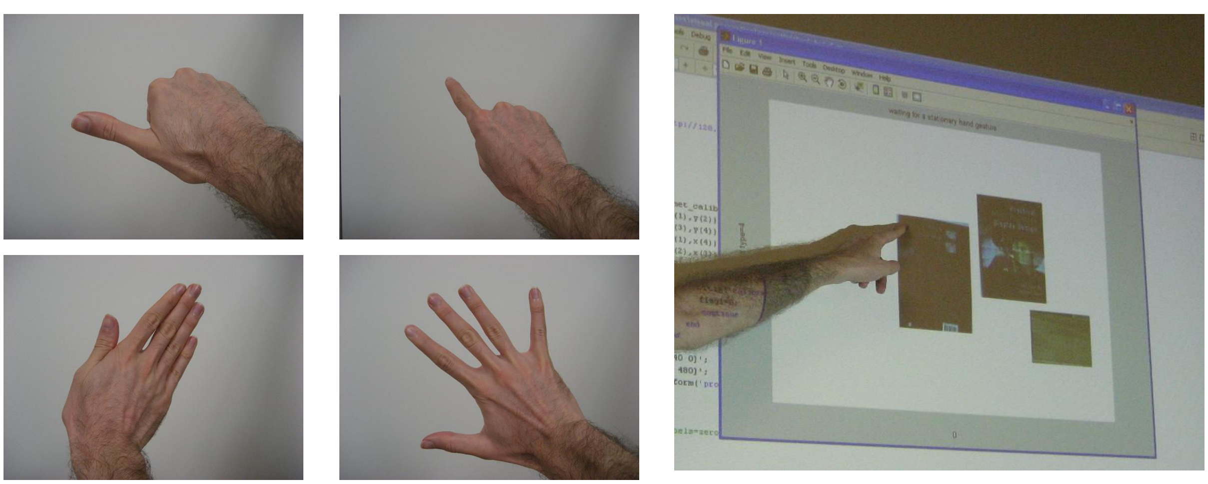

Figure 3. Some sample hand gestures and a snapshot of the working system