

# Creating Serendipitous Encounters in a Geographically Distributed Community

Adithya Renduchintala    Aisling Kelliher    Hari Sundaram  
Arts Media and Engineering, Arizona State University  
email :{adithya, aisling.kelliher, hari.sundaram}@asu.edu

## Abstract

This paper is focused on the development of serendipitous interfaces that promote casual and chance encounters within a geographically distributed community. The problem is particularly important for distributed workforces, where there is little opportunity for chance encounters that are crucial to the formation of a sense of community. There are three contributions of this paper. (a) development of a robust communication architecture facilitating serendipitous casual interaction using online media repositories coupled to two multimodal interfaces (b) development of multimodal interfaces that allow users to browse, leave audio comments, and asynchronously listen to other community members, and (c) a multimodal gesture driven control (vision and ultrasonic) of the audio-visual display. Our user studies reveal that the interfaces are well liked, and promote social interaction.

## Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems] *Audio input/output, Evaluation/methodology, Video*, H.5.3 [Group and Organization Interfaces] *Asynchronous interaction, Collaborative computing, Computer-supported cooperative work, Evaluation/methodology*

**General Terms:** Design, Experimentation, Human Factors

## Keywords

Serendipitous interaction, mediated communication, image repository, online media repository, remote interfaces, social computing

## 1. Introduction

In this paper we present our approach to creating serendipitous encounters in a geographically distributed community. The problem is important as in such communities, colleagues working together across distributed physical environments rely heavily on scheduled appointments and meetings to interact in an efficient and productive manner. While regular meetings are organized to optimize available time, such meetings leave little room for casual interactions. Such interactions are essential for the development of a shared sense of community, where swapping stories about last night's game or office romance rumors help foster bonds of companionship and shared understanding. We shall build upon recent research on sociable media computing [5,9] to address these concerns.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HCM'06, October 27, 2006, Santa Barbara, California, USA.  
Copyright 2006 ACM 1-59593-500-2/06/0010...\$5.00.



**Figure 1:** Interacting with the developed interface. Above- recording a comment, below- flipping through images.

The purpose of our project is not to replace current real-time communication schemes, but to design systems that serve to promote casual serendipitous encounters. Let us examine the role of traditional communication frameworks in promoting casual interaction.

Real-time, high bandwidth communication frameworks such as video phones are not particularly appropriate mechanisms to promote serendipitous interaction. In real-time communication frameworks using videophones, users are often obliged to respond actively and with full attention to sustain the interaction. This inherent constant pressure creates high cognitive overhead for the user and can deter remote interaction [8]. The video phone display metric equally contributes to a distancing effect between members known as the "window frame" effect, where in the physical separation and lack of interaction due to this separation is highlighted [10].

## 1.1 Related Work

Sociable media computing is an emerging area of research that focuses on design strategies and engineering solutions for mediated human-human interaction, typically across remote location. There have been several projects involving

communication across remote ‘mediated spaces’. Initial work to create social interaction between distributed work groups such as the ‘Media Spaces’[5] project utilized a combination of streamed video and audio data. The purpose of the Media Space project was not only to maintain social connectivity between remote participants, but to provide a structure that can support remote collaboration of work. There has been prior work on combining audio streams and video displays. The Telemurals [9] project uses audio-visual media to create instances of interaction between remote participants who would otherwise lack the motivation to do so. Interaction is promoted by progressively increasing the Visual and audio details with increased participation. Projects such as ‘Somewire’ [12] concentrate on audio streams to create opportunities for social interaction. ‘Somewire’ focused on tangible non-screen based interfaces for communication. In the “Chit Chat Club” [11] project, physical anthropomorphic sculptures equipped with video and audio channels and a controllable ‘expressive’ display are used to bring remote participants together. The chit chat club project combines advantages of online communication such as minimizing geographical constraints, with those offered by interacting in a common physical space.

Many of the current schemes developed, motivate the remote participation by focusing on the *current moment of interaction*. We adopt a complementary strategy where simultaneous presence are not a prerequisite for casual encounters.

## 1.2 Our Approach

We propose to address the problem by building two multimedia stations (with multimodal sensors and displays) that encourage serendipitous encounters across space and time [3]. Our interfaces (ref. Figure 1) foster community interaction and nurture community dialogue by allowing members to navigate images contributed to an online media repository (a Flickr pool [1]), and leave audio comments that can be heard by other members, asynchronously. We now highlight our specific contributions.

**Coupling:** We developed robust TCP/IP communication mechanisms that allow the two remote displays to communicate with each other with local and online media repositories and give rise to opportunities for casual interaction within remote participants. The local database is used to store details of activity at each station, as well as the audio comments at each station. Each multimedia station can also communicate with the community image pool at Flickr, where community members regularly upload photos.

**Multimodal display:** We developed an audio-visual display for browsing a community media repository. Our interface allows users to quickly focus on the recently active image and listen to (and respond with) audio comments left by other community members. The interface uses standard cultural icons for media navigation, and additionally reveals the system state, thus minimizing confusion [6]. The display also reveals awareness of local and remote activity.

**Gesture based control:** We developed a multimodal gesture based control of the audio-visual display. Simple vision based techniques are used to control imagery, while ultrasonic proximity sensors control the playback and recording of audio comments as well as for recognition of local and remote presence. Simplicity of control gestures and a small gesture lexicon were key requirements of the gesture based interface, to

allow unfamiliar users to learn its usage quickly. We developed a simple proximity sensor / camera based system to sense the gestures.

The user studies indicate that the system performs well – there was clear evidence of an emergent dialogue and the users had very positive comments on the interface design, and interaction. We intend to further test the effectiveness of the design in community building by conducting extensive user studies over longer time frames.

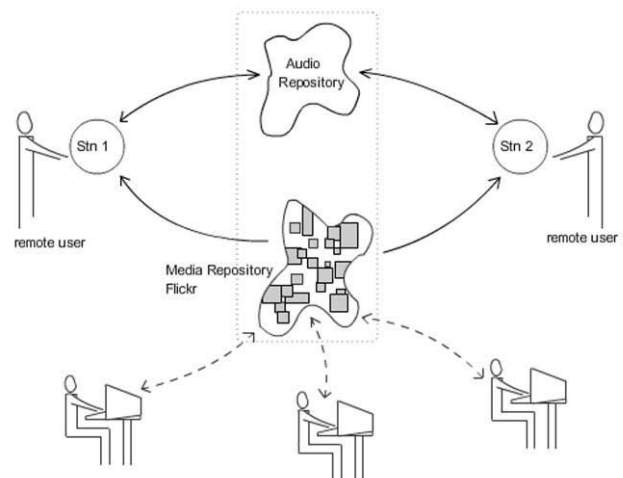
The rest of this paper is organized as follows. In the next section we present the overall system design. In section 3, we discuss serendipitous interfaces where we present our approach to visual interface design, gesture based control coupling of control to the interface. In section 4 we present our communication mechanism and in section 5, we present user studies and discussion. Finally in section 6, we present our conclusions.

## 2. The System Design

We constructed two interfaces composed of sensors and a display; we will refer to each such interface as a station. Both these stations access a pool of online images that form our media repository. Users at each station can listen to and record audio comments using a multimodal gesture based interface. The audio comments are stored online, creating an audio repository linked to the image collection. The schematic structure of the system is represented by Figure 2. The diagram shows users independently contributing to, and sharing data of an online repository. This online data set is accessed by the remote stations, through which users can record and listen to audio comments for each image forming the audio repository.

### 2.1 Online Media Repositories

Photographs have traditionally been used as an engaging and informative medium to communicate everyday experiences as well as unique events [4]. We chose Flickr as a media repository. This was because the AME Flickr group pool has emerged as a place where members of our research group share their experiences. They upload pictures and actively interact by leaving tags, notes and text comments on the Flickr site. In the current project, community members can send images from their

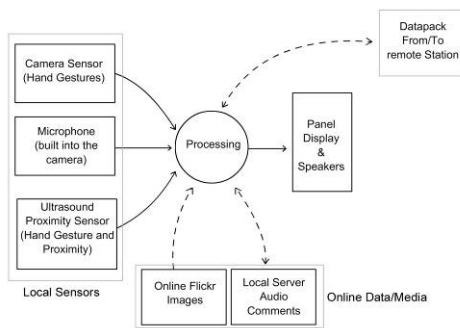


**Figure 2:** Schematic Structure of the System. Stn1 and Stn2 are the two remote stations, which access an online media repository and a audio repository. The Media repository can be added to by group users independently.

Flickr accounts to a group photo pool created for AME. This Flickr group pool formed the media repository for this project. The system pulls images off the Flickr repository once they have been uploaded by users to the AME pool. Uploading images is not a prerequisite to record and playback comments at either of the interfaces. This allows visitors to AME to interact with the interfaces and leave audio comments.

## 2.2 The Components

We now present a brief overview of the components at each station. Each station is equipped with a large LCD panel display. The display has a built-in set of speakers thus each station can provide audio-visual feedback. The station has three sensors – a camera, a microphone and an ultrasound proximity sensor, to control the interface. Users can navigate through recent images in the repository by performing simple hand gestures that the

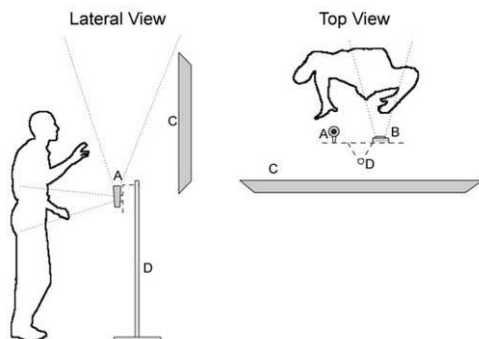


**Figure 3:** Block Diagram of Sensors-Display Coupling and External Communication. The solid lines are wired coupling, and dotted lines are TCP/IP socket connections.

station’s camera and proximity sensors detect. The sensing and display at each end are controlled by our application that runs on an Apple G5 computer. The communication between the stations is handled through TCP/IP socket connections. Figure 3 shows the sensor-display coupling as well the communication modules at each interface.

## 2.3 Station Location and Physical Setup

The location of each station within the physical space is critical to aiding interaction. We experimented with different physical locations. We looked for spaces with apt special affordances and spaces frequented by people. While the location of the station



**Figure 4** Diagram of physical setup of a Station. A- Camera Sensor, B- Ultrasound Proximity Sensor, C- Display Panel and D- Sensor Stand.

should be one that is easily accessed, we noted that audio output from the system can be at times disturbing and intrusive. Secondly when we noted that when one of the interfaces was placed in a corner, but still visible to people walking by, the users were less inclined to go forward and play with the interface. Hence we decided to place both interfaces on a wall that faced a corridor, or a highly frequented path. Our interface is designed to react with a visual change, when it detects people walking past it. We placed one of the stations in the office space for students and researchers. The other station was positioned in the lobby area of another research building.

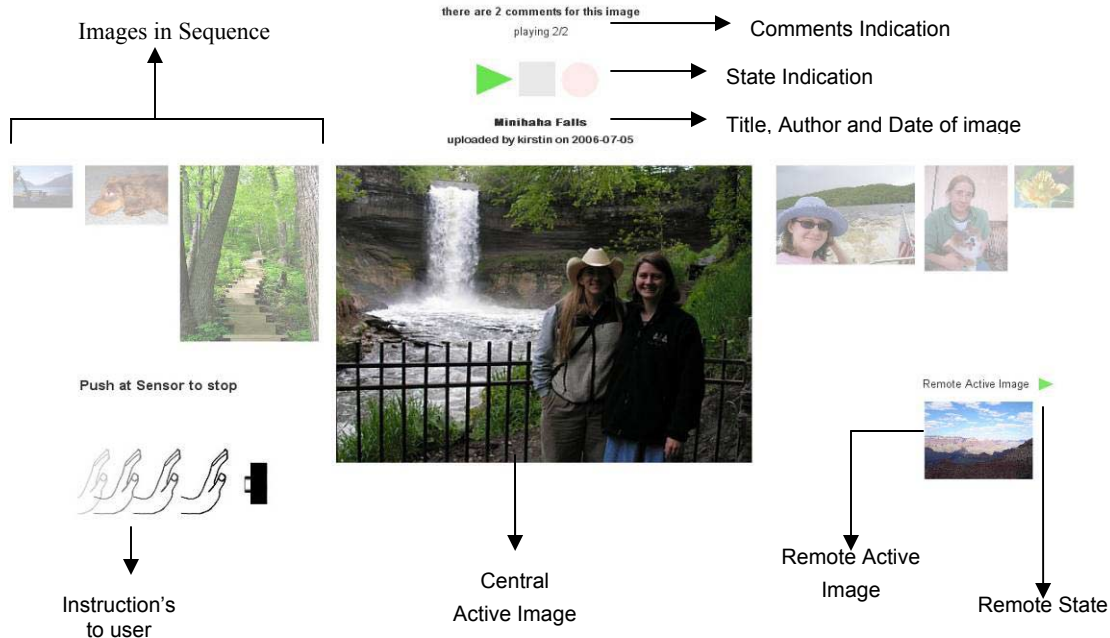
A simple gesture driven interface was designed to explore the image repository and help build the audio repository. The display device is a wall mounted flat panel LCD screen. The panel is fixed at a convenient height; the center of the screen is five and a half feet from the floor (ref. Figure 4). The camera sensor and the proximity sensor are fixed on a custom made stand positioned in front of the center of the panel. In the current set-up the sensors are positioned four feet from the floor, and one and a half feet from the panel. The stand was used as it is easily adjusted and sensors can be relocated without much difficulty. As the diagram (ref. Figure 4) shows the camera is oriented upwards and the proximity sensor detects presence perpendicularly away from the LCD panel.

## 3. Serendipitous Interfaces

In this section we present the visual design, control and navigation of the stations. By the term ‘serendipitous interface’ we envision systems that are non-cumbersome to interact with, for short and repeated periods. Such an interface should allow users to access required data quickly and easily, without requiring them to use a traditional keyboard or a mouse. Thus we focused on ways to use the “body as an input device” [7] worked towards creating a gesture driven interface. In the next three sections we discuss in detail, visual interface design, the control of the interface through gestures, and navigation and interaction paradigms.

### 3.1 The Visual Interface Design

The visual interface was designed to be simple and uncluttered. To inform users that the images are structured in a sequence we chose to display a set of images. There is an active image, and there are three images to the right and left of the active image, thus creating a sense of sequence. This would make searching for a particular image easier as users can have a sense of where an image is located with relation to others. The seven images are arranged spatially to resemble the view through a convex magnifying lens. We intuitively understand that the central portion of a magnified image is of most importance. There is a decrease in visual clarity towards the edge of the magnified image. We intended to incorporate this idea in our visual display. The images towards the sides are faded to a greater extent than the images towards the center. The central image is the largest and clearest.



**Figure 5:** In this figure we examine the details of the visual interface design

Important details about an image such as the image title and its author are displayed above the active image. All the textual information on the screen is for the central ‘active image’. Above the title is a space for icons to indicate the states of the station (i.e. wait, play, stop or record). As the state names suggest, wait implies the user must wait at the image for three seconds before the station switches to the “play” state. Here the audio comments that the image has received will be played reverse-chronologically. The “stop” state is the system stopping

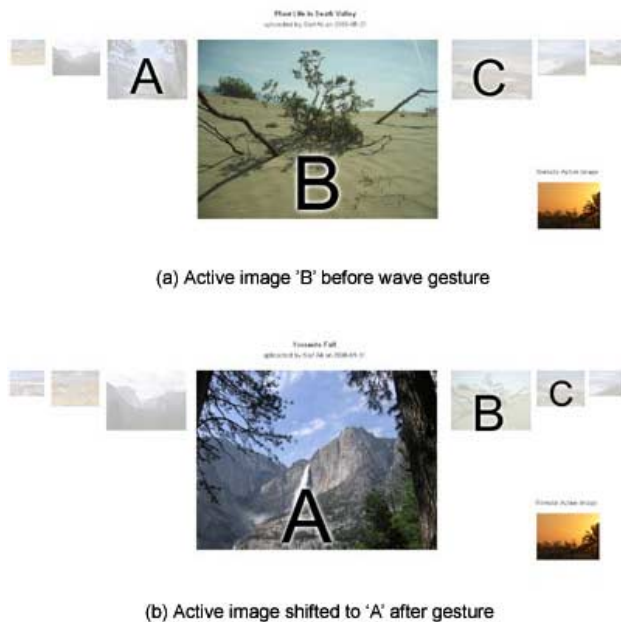
the audio, and waiting for the gesture to record. In the record state the microphone is initiated and begins recording audio data. The play, stop and record states are depicted by traditional media control icons: a green triangle, a black square and a red circle respectively. The number of comments for the active image is displayed above the state indicators. The display contains an instruction message box that prompts users are encountering the media station for the first time. For example if the station is in the ‘play’ state, the message in the instruction message box is ‘push toward the sensor to stop’. When playback of comments has been stopped the instruction message reads ‘Push and Hold to record a new comment’. The remote active image and the state of the remote station are displayed in the bottom right corner of the screen (ref. Figure 5).

### 3.2 Multimodal Gesture Control

We designed an interface that permits a user to access the media repository quickly, and through using simple hand / body gestures. This was important to maintaining a casual, low-cognitive load interaction. We deliberately decided not to use traditional input devices (e.g. keyboard / mouse) or buttons in the interface. Note that we decided to have a simple gesture lexicon. The goal was to enable unfamiliar users to quickly learn to interact with the interface. Our design for gesture analysis was based on this lexica. We acknowledge that a more complex gesture lexicon will require sophisticated sensing; however such lexica are unlikely to be of use our system. In this section we shall present our approach to visual interface control through vision and proximity sensor based sensing.

#### 3.2.1 Camera based Gesture Detection

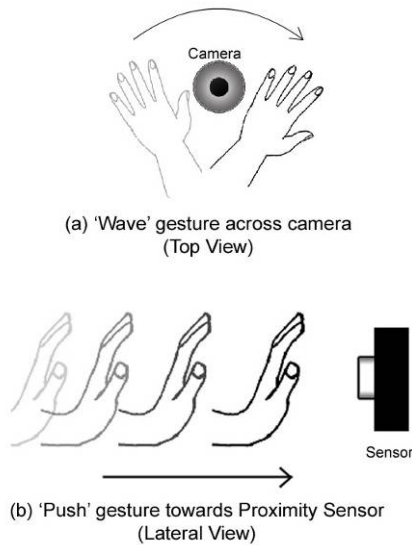
In this section we discuss a simple camera based control of the visual interface. The camera oriented upward is programmed to identify pixel changes from the left to right and vice-versa (ref. Figure 8(a)). A control signal is sent to the display application when such a pixel change is recognized by the camera across its



**Figure 6:** Above 7(a): Initial Screen ‘B’ is the active image, Bottom 7(b): after receiving control signal ‘A’ is the active image.

entire field of view. The camera is oriented in the upward direction, therefore people passing by will not trigger false positives. The camera casing also encloses a microphone that records the audio comments.

The Camera and the proximity sensor are controlled by Max/MSP runtime program [2]. Max/MSP is a graphical programming paradigm that allows real-time control of various sensor inputs. The camera is a commercially available iSight camera, which has a built in microphone. The Max/MSP program down samples the camera input to 25 x 6 resolution



**Figure 8:** (a) The top view schematic of a wave gesture from left to right, over the camera’s filed of view. (b) Lateral view diagram of a ‘push’ towards the sensor.

from 640 x 480 VGA. The 24bit color output from the camera is also converted to a 256 level grey scale image. Three clusters of pixels centered at (1, 3), (12, 3) and (24, 3) are scanned for changes over a threshold value. If there is a change in all three pixels in order within a time threshold, the application detects a gesture. As each frame is relatively low resolution, we chose to refresh the frames at 30 Hz the maximum possible by the camera (ref. Figure 8(a)).

### 3.2.2 Proximity sensor Based Gesture Detection

In this section we describe how we use an ultrasonic proximity sensor for controlling the audio comments (ref. Figure 8(b)). The ultrasound proximity sensor detects any object located along a perpendicular angle away from the plane of the wall. This sensor is used to control the ‘recording’ and ‘playback’ of comments and detects any physical presence within three to four feet. The gestures and usage of these sensors is explained in the following section.

The ultrasound proximity sensor is controlled through a microcontroller and a Max/MSP application. The microcontroller triggers the sensor to sense presence at 33.33 Hz. Once the sensor receives the trigger pulse, an output pulse is generated at its output pin, and the sensor transmits an ultrasound wave and waits to detect the reflected wave. The width of this pulse is proportional to the distance of the object. The microcontroller measures the pulse width and generates a

numeric result in inches. The microcontroller sends this information to another Max/MSP program via a midi Cable. The rate of change of distance is computed by the Max/MSP program. When this change is above a threshold, a control signal is sent to the display application. The communication between Max/MSP and the display applications is through TCP/IP socket connections (ref. Figure 8(b)).

### 3.2.3 Detection of local Presence

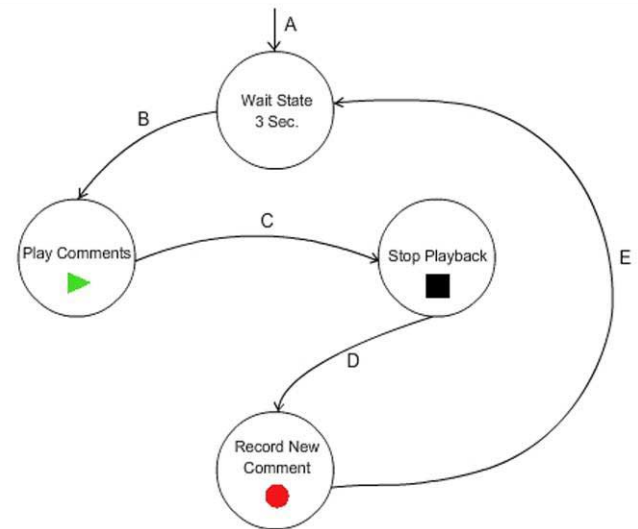
We wanted the system to be able to attract attention of those who are near the LCD panel. As mentioned previously the proximity sensor sensing the distance of any object from its location. This made it relatively easy to detect presence within the range of the proximity sensor. If the sensor reports a presence between three to four feet, the system recognizes it as a local presence.

## 3.3 Exploring the Repository

In this section we show how the multimodal sensing affects the visual interface. The display panel shows a sub-section of the twenty-five most recent images that have been uploaded to the Flickr pool. Seven images are displayed at a time, and the active image is displayed in the center, while the other images are arranged on either side. The active image can be explored to a greater level; previous comments for that image can be played. The browsing user can also choose to record a new comment. We now describe the coupling between sensing and feedback in the next three subsections.

### 3.3.1 Visual Media Navigation

We shall now discuss how the local visual display is controlled through hand gestures. Control signals from the camera sensor switch the active image. When a ‘left to right’ movement is detected i.e. when the user waves his/her hand from the left to right across the camera, the image changes as explained in Figure 6. Initially image ‘B’ is the active image. Now, the user can switch the active image ‘A’ by simply waving his/her hand from left to right across the camera.



**Figure 7:** The control loop of operations. The State transitions are, A- Selected an Active Image, B- No activity for three seconds, C- ‘Push’ Gesture detected by proximity sensor, D- Another ‘push’ gesture and hand held near sensor, and E- Hand released.

### 3.3.2 Playback / recording of audio comments

Once an active image is selected, the system enters a wait state. If no activity is detected for a period of three seconds the comments that have been posted for the active image begin to play. The user can continue to listen to all the comments or can intervene and stop playback. To stop the audio, a user needs to move his/her hand towards the sensor. This movement resembles a ‘push’ as explained in section 3.2.2. When this gesture is performed with a speed above a threshold, the system identifies it as a valid ‘push’. A control message is then sent to the display application, which stops playback of all the comments associated with the active image. Another ‘push’ enables the microphone to record, and the user can now leave a new comment. While recording, the user needs to place his/her hand in close proximity to the sensor because the recording continues as long as the sensor detects presence within a few inches. Once a comment is recorded it is updated to the server and the system loops back to the wait state. Figure 7 depicts this operation of this control loop.

### 3.3.3 Displaying local presence

When the proximity sensor detects any presence within three to four feet, the images on the screen (of the local station) begin to flip automatically. The quick automated flipping of images continues until the image that has been commented on most recently occupies the central position. Effectively the system informs the person in proximity to the panel that this image has recently been commented on. The system chooses this image as the active image.

### 3.3.4 Displaying Remote Activity

When there is any form of interaction with one of the stations, actions are communicated to the other station. These control messages trigger display indicators that reflect remote actions. This feature is critical in letting users get a sense of remote activity [6]. The background color of the screen also indicates the state of activity at the other remote station. When control messages are being sent to and from the sensors and display of the remote unit, the background color flicker to grey and back to white. The background color of the display screen flickers from white to grey and to white again when anyone is using a remote station. The background color flickers when a remote user is flipping through images, listening to comments or recording a new comment. This is a non intrusive way to attract attention of people who are near the station, but not close enough to notice the remote active image.

The active image of a remote station is displayed in the bottom right corner of the screen (ref. Figure 5). This display of the remote active image is updated with every gesture given as input at the remote station. Further more the state of the remote station is also indicated near the remote active image. For example, if a remote user is recording a comment, a small red circle appears by the remote image. By these display indicators the users can build a clear picture of the activity occurring at the remote station. Thus local proximity to any of the stations triggers display responses in both stations. This feature coupled with constant mutual awareness indicators can initiate remote interaction.

In this section we have presented our approach to the design of visual interface of the station, sensing mechanisms for gesture /

body movement based interface control and the coupling of sensing to the media feedback.

## 4. Communication between Remote Stations

We now describe system level communication framework between the two remote stations. When there is any form of interaction with one of the stations, actions are communicated to the other station. These control messages trigger display indicators that reflect remote actions.

The communication framework is based on TCP/IP socket connections. Both the stations are connected through a TCP/IP server that is running in ‘relay’ mode. This server is also running on a Max/MSP program. Both the display applications (one at each station) connect to this relay server. When one sends a data packet to the server, the server then automatically relays it to all other connected applications. This gives flexibility to expand to a greater number of remote stations in the future.

There are three fields in every data packet sent across remote stations. Table 1 shows the three fields and the possible data strings for each field. The first field specifies the location from which the message is being transmitted. ‘‘Mat’’ denotes the Matthews Center Building and ‘‘Byc’’ denote the Brickyard Building. The second field contains information on the action that has been performed at the interface. When a user flips to a new image, this field holds the string ‘‘image’’. When a user stops the audio comments, the field holds ‘‘stopped’’ string. Similarly ‘‘record’’ is transmitted when a user begins recording a new comment, and ‘‘stopped’’ is also transmitted when the user completes recording a comment. The third field is a parameter field that is ‘‘null’’ in all cases except when the action is ‘‘image’’, in which case the parameter field holds the Flickr URL of the image being viewed. This URL is used to display the Remote Active Image on the LCD Panel.

**Table 1:** Data Fields composing a Data a Packet and possible data string for each field.

Location	Action/Activity	Parameter
‘‘Mat’’	‘‘image’’	Image URL
‘‘Byc’’	‘‘recording’’ ‘‘stopped’’ ‘‘play’’	Or ‘‘NULL’’

Each message packet has the three following fields separated by a ‘‘\*’’ separator, thus a possible data packet could be as follows.

- *mat\*\*recording\*\** :This would communicate that the message is from ‘mat’ abbreviation for Matthews Center Building and a remote participant is currently recording a comment for an image.
- *byc\*\*image\*\*http://static.flickr.com/70/159063225\_7248b5dd5c.jpg*: This packet’s data fields indicate that a remote participant at The Brickyard Building (byc) has presently navigated to an image with flickr URL is <http://static.flickr.com/70/159063225728b-dd5c.jpg>

The display interface was made with Adobe Flash software. This application also initiates recording of sound from the microphone and the playback of comments. Audio comments are directly recorded into a server as ‘.flv’ files using Flash Media Server. The display application calls PHP scripts that update the database holding information such as number of

comments per image, name and location of each comment on the server. A java script is run to search for new uploads to the flickr image pool, and information on the new images such as image title, image author, and image URL are added to another database.


## 5. Usage Scenarios and Experiments

We will now describe the typical usage scenarios for our system and the experiments that were conducted to test the effectiveness of the two remote stations. We will then discuss the feedback we received from the participants in the study.


### 5.1 Typical Scenarios

In this section we will present two typical usage scenarios using the remote panels.


- Case 1: Each remote station can act as a standalone interface, with a large number of images and comments that a single user can browse through. The user can respond to these comments, knowing that they can be retrieved by other users at a later time. In this way a station can bring people who work at different times into ‘contact’.
- Case 2: Another typical mode of communication is when two users are simultaneously using both remote stations. As described through the paper, the display allows the users to be mutually aware of each others activity. Icons present near the remote active image indicated whether the remote user is either listening to comments or recording a new comment. When one user realizes that the remote user is recording a new comment, he/she can flip to that comment and post a reply. Authors of an image can leave initial comments explaining the image or leave short anecdotes related to the image.



*“This sixteen year old Pomeranian lost all her teeth, so her tongue won’t stay in her mouth!”*



*“aa... Are they ALL holding hands!?”*



*“This picture makes me want to take a week of from work and go exploring the canyon...”*

**Figure 10:** Examples of initial audio comments. The comment for the first image was left by the author, the others however are comments that are either inquisitive in nature or are a response to an others image

## 5.2 Experiments and Results

We conducted the user study in accordance with the two scenarios described in the previous section. Occupants who work at the Brickyard building and Matthews Center were asked to interact with the interface at their respective workplaces. Six participants were involved in the user study. Two were graduate students at Matthews Center, and two members from the administration who work at the Brickyard Building, and two professor’s one from each location. We spread the user study over two days. All the users were requested to upload new images to the Flickr pool. We observed that, the images uploaded were mainly of trips the participants made during the summer.



*Stjepan (M at 1:10 pm on 6<sup>th</sup> July): What is that thing for anyway?*

*Kirstin (B at 2:45 pm on 6<sup>th</sup> July): Kikona had surgery so she had to wear this cone, so she wouldn’t chew on her stitches...*

*Siew (M at 6:17 pm on 6<sup>th</sup> July): It looks so sad!*

*Stjepan (M at 1:42 pm on 7<sup>th</sup> July): Is that the reason why the other dog got its teeth knocked out!?*

**Figure 9** Informing personal experiences to colleagues. (M) – Matthews Center, (B) - Brickyard Building

### 5.2.1 Single User Experiments

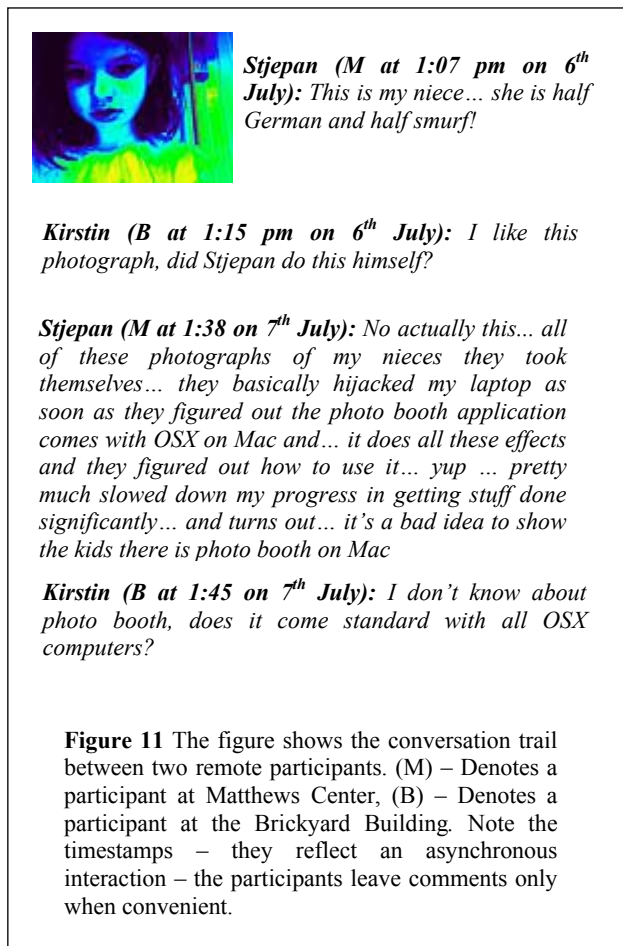
On the first day, the users were simply asked to interact with the interface, and familiarize with the sensing modalities. They were encouraged to record initial comments (note that all “comments” are audio comments) on the images that they uploaded. They were free to leave comments on other’s images as well. During these tests only one user was allowed to use interface at either location. No fixed time period for interaction was enforced, users interacted with the system for as long as half an hour, and sometime left more that five to six comments. At the end of their interaction, they were asked to fill up a questionnaire. The questions were classified in three parts, part one dealt with the physical interface, the second portion requested feedback on the visual layout and the clarity of the information disseminated on the LCD Panel. The third portion asked questions on probable effects, commenting of an image would have in building a sense of community.

We found that initial comments generally were explanatory in nature. The comments provided additional context to understand the image. Others were inquisitive about images and wanted to know where a particular image was taken, or who was in the picture. Users enjoyed leaving humorous comments on their own images, or those left by others. Figure 10 shows some of

the initial comments that some users left. At the end of the single user experiments we observed that certain users tend to engage with the system longer than others. This was expected as the potential of the system will be apparent only when users begin receiving comments from colleagues at the remote workplaces.

### 5.2.2 Interacting With Remote Users

To explore the second possible scenario, the next day we gave users a specific time to interact with the system. We set up



**Figure 11** The figure shows the conversation trail between two remote participants. (M) – Denotes a participant at Matthews Center, (B) – Denotes a participant at the Brickyard Building. Note the timestamps – they reflect an asynchronous interaction – the participants leave comments only when convenient.

times such that one user interacts with the station at the Brickyard building, and another interacts with the station at Matthews center. Again we did not impose any restrictions on the time of interaction. We found that both users were actively engaging, exchanging comments. These comments followed a narrative trail. The comments usually started with a questions pertaining to the picture and then took paths that resembled normal everyday conversations between physically co-located participants.

Figure 11 shows the conversation between two remote participants – Stjepan and Kirstin. The image is uploaded by Stjepan who works at the Matthews Center. It can be seen that Kirstin who works at the brickyard replies taking the focus of the exchange away from the photograph.

Apart from creating a base for comments, and conversation via the interface, colleagues get an insight into the author’s personal lives. The interface provided an opportunity to users to disseminate certain facts about themselves and those around them. One of the users put up an image of their pet cat, which underwent surgery. It’s unlikely that such incidents would be broached in meetings, and this interface turned to be apt for publicizing them. We found that users were far more comfortable with the interface on the second day, and were able to concentrate more on engaging with the online media and remote users.

### 5.2.3 Evaluation and Feedback

The feedback that was requested from users fell into three main categories, (a) The Ease of Using the Physical Interface (b) The Visualization Scheme and (c) Feedback on Interacting with Remote participants. All questions had seven answer choices ranging in gradation from “Very Poor” to “Very Good” from which user could pick one. Answer choices to questions regarding how often users would use the system were ranging from “Very rarely, once a week or less” to “Very often, more than a few times a day”. The results of the Interaction Feedback are tabulated in Table 2, Table 3 and Table 4.

**Table 2:** User Response to the Physical Interface

Feedback Elements	Average Score
Speed of Interface	5.5 / 7
Ease of Use	5.6 / 7
Consistance and Accuracy	5.0 / 7
Gesture Selection	5.8 / 7
Control over Interface	5.1 / 7

**Table 3:** User Response to Visualization Scheme

Feedback Elements	Average Score
Layout	5.5 / 7
Information displayed for an image	5.5 / 7
Clarity of State of the system	6.1 / 7
Helpfulness of short Instuctions on screen	6.3 / 7
Clarity in Display of Remote Activity	6.3 / 7

**Table 4:** User response to Effectiveness in aiding remote Interactions.

Feedback Elements	Average Score
Would comments make you photograph appealing	5.2 / 7
Would you check if new images/Comments are present	5.3 / 7
Likelihood of discussing these and comments later, in person	5.6 / 7
Was remote activity clear during simultaneous interaction	5.7 / 7
If you noticed a remote recoding taking place would you stop to listen	6.5 / 7
Likelihood to reply	6.0 / 7



We found that the feedback on the physical interaction very positive. The response indicates that the physical interface design and visualization was well liked and promoted remote engagement.

## 6. Discussion

We now discuss improvements that we aim to implement based on the feedback from our experiments.

One prominent suggestion that we received was to provide more control over the playback of the comments. To address this we are working towards providing the users with the ability to fast forward, rewind, and delete comments. Although the system is meant for a closed group of people, users requested a display to identify who left a particular comment. We also feel that we can replace the camera sensor, with additional proximity sensor as the camera was sensitive to the ambient lighting conditions, which need not remain constant within either workplace.

The use of multiple installations has been attempted in several methods, however utilizing a common media repositories created by users of the spaces to motivate remote interaction is new. The concept of serendipitous interfaces was also well received by participants of the user study. The initial results and feedback were very promising. We found users leaving comments specifically for other colleagues, and most users felt that they were quite likely to leave personalized comments for colleagues whom they may not get to meet regularly. All users felt that the interface was fun. With the interface easy to access and the Flickr image repository increasing, we have created opportunities for the workgroup to interact with each other in a casual, fun and serendipitous manner.

## 7. Conclusions

In this paper, we presented our approach to creating serendipitous encounters in a geographically distributed community. The problem is important as distributed workforces seldom promote casual and chance encounters that are often keys to developing a strong sense of community. There are three contributions to this paper: (a) a robust communication mechanism for casual interaction, that allows two multimodal interfaces to communicate with each other, utilizing local and online media repositories. (b) developing a visual interface that allows users to quickly browse recently commented media. The interface also revealed local and remote awareness of presence, as well as the state of the system. (c) a multimodal (vision and ultrasonic sensor) based gesture recognition to control the media on display. The ultrasonic sensor is also used to control audio comments (recoding and playback) and can also detect local and remote activity. We built two interfaces and located them in two remote buildings containing members from the arts media and engineering program. The user study results are very positive, indicating that the interfaces were easy to use, and promoted interaction. We also began to see the emergence of dialogue amongst the remote community members. We plan on building upon the interface in several ways – (a) we plan to develop mechanisms to identify users unobtrusively, using RFID tags (b) adapt the visualization to analysis of user activity at the two stations.

## References

- [1] *Flickr* <http://www.flickr.com>.
- [2] *Max/MSP (media tool)* <http://www.cycling74.com>.
- [3] S. AGAMANOLIS (2003). *Designing Displays for Human Connectedness. Public and Situated Displays: Social and interactional aspects of shared display technologies*. K. O'HARA, M. PERRY, E. CHURCHILL and D. RUSSELL (eds), Kluwer.
- [4] P. APPAN, H. SUNDARAM and D. BIRCHFIELD (2004). *Communicating everyday experiences*, ACM Multimedia Workshop on Story Representation, Mechanism and Context, also AME-TR-2004-07, pp. 17-24, Oct. 2004, New York, New York.
- [5] S. A. BLY, S. R. HARRISON and S. IRWIN (1993). *Media spaces: bringing people together in a video, audio, and computing environment*. *Commun. ACM* **36**(1): 28-46.
- [6] P. DOURISH (2001). *Where the action is : the foundations of embodied interaction*. MIT Press 0262041960 (alk. paper) Cambridge, Mass. ; London.
- [7] E. HORNECKER and J. BUUR (2006). *Getting a grip on tangible interaction: a framework on physical space and social interaction*, Proceedings of the SIGCHI conference on Human Factors in computing systems, 1-59593-372-7, ACM Press, 437-446, Montreal, Canada.
- [8] E. A. ISAACS and J. C. TANG (1993). *What video can and can't do for collaboration: a case study*, Proceedings of the first ACM international conference on Multimedia, 0-89791-596-8, ACM Press, 199-206, Anaheim, California, United States.
- [9] K. KARAHALIOS and J. DONATH (2004). *Telemurals: linking remote spaces with social catalysts*, Proceedings of the SIGCHI conference on Human factors in computing systems, 1-58113-702-8, ACM Press, 615-622, Vienna, Austria.
- [10] K. G. KARAHALIOS (2004). *Social catalysts : embracing communication in mediated spaces*: 108 p., Thesis Ph. D. - Massachusetts Institute of Technology School of Architecture and Planning Program in Media Arts and Sciences 2004.
- [11] K. G. KARAHALIOS and K. DOBSON (2005). *Chit chat club: bridging virtual and physical space for social interaction*, CHI '05 extended abstracts on Human factors in computing systems, 1-59593-002-7, ACM Press, 1957-1960, Portland, OR, USA.
- [12] A. SINGER, D. HINDUS, L. STIFELMAN and S. WHITE (1999). *Tangible progress: less is more in Somewire audio spaces*, Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, 0-201-48559-1, ACM Press, 104-111, Pittsburgh, Pennsylvania, United States.