Evaluating Image Filtering Based Techniques in Media Space Applications

Qiang Alex Zhao, John T. Stasko
Graphics, Visualization, and Usability Center
Georgia Institute of Technology
Atlanta, GA 30332-0280 USA
{azhao, stasko}@cc.gatech.edu

ABSTRACT
Media space applications that promote informal awareness in an organization confront an inevitable paradox: the shared video connections between offices and rooms that promote informal awareness also can rob individuals of privacy. An important open problem in this area is how to foster awareness of colleagues while minimizing the accompanying loss of privacy. One proposal put forward is to filter the communicated video streams rather than broadcasting clear video. Such a scheme may facilitate awareness while helping to alleviate some aspects of the privacy loss. In this article, we describe several image filtering techniques that provide awareness in informal group communication applications while blurring the details of an individual's activities, thus potentially preserving more privacy. We describe studies to quantitatively and qualitatively assess the degrees of awareness and accuracy that these filtering techniques provide.

KEYWORDS
Informal group awareness, privacy, video, image filter, media space, real-time groupware

INTRODUCTION
Using real-time audio and video transmission together with other human-computer interface techniques, media space applications provide virtual human-to-human interaction spaces to people at physically separated locations [3]. In addition to supporting planned, formal interactions, great efforts have been made to support casual, informal group communications in media spaces [4, 5, 13]. In the synchronous communication realm, where events occur at nearly the same time, co-located informal group interactions are often implicit and serendipitous [9]. For example, an important conversation between two colleagues might be caused by merely "bumping into" each other. One simple method to support informal interactions beyond physical proximity is to continuously provide awareness information such as video and audio signals to the entire group rather than on an explicit, by-request basis. It is then up to the user to determine if a remote participant in the media space is available, and whether it is appropriate to start a conversation or not.

This continuous-access requirement poses a difficult problem in media space applications: how to balance awareness and privacy. On one hand, users must have access to awareness information about other users. Guided by social protocols, this awareness information provides the context that people utilize to start interactions. For example, a person may need someone's presence to remind him or her of the possibility and appropriateness for interaction.

On the other hand, disclosing awareness information about oneself inevitably compromises the individual's own privacy. How much privacy is compromised may vary, however. Privacy is a complex issue and has many aspects. Particular dimensions of privacy include [2]:

- Knowing where an individual is, i.e. "tracking";
- Hearing what someone is saying;
- Recording and manipulating audio, video, and other information without consent of the user;
- Identifying what someone is doing, that is, their "activity";
- Identifying who is meeting with a person;
- Seeing details of someone's actions, for example, watching as they change clothes to go play tennis;
- Seeing how someone looks, what their mood is, what they are wearing, etc.

Clearly, these are just a few of the possible privacy issues that a person may encounter. Further amplifying the complexity of the privacy issue is the fact that different people have widely varying levels of concern in this regard. Nonetheless, opening a constant video view into an individual's office as done in some media space applications probably compromises all these dimensions of privacy. So then, how can one promote awareness while minimizing people's feelings of the loss of privacy?

One proposed approach to this issue is to use abstract, iconic representations of users instead of video [1, 7, 14]. However, it is still debatable whether sacrificing the relatively richer contextual information in video images is necessary [11].

Another proposed approach is to transmit modified video data instead of raw video at low frame rates. For example, media space software may slowly transmit lower-than-ordinary resolution video images so that remote users are able to sense the presence of the owner of the video and her movement, but they are less likely to recognize the details in the video. High resolution, high frame-rate video might be reserved for users engaged in focused interactions.

Clearly, the use of modified video addresses only some of the dimensions of privacy. Individuals still may feel "tracked" and this approach does nothing with the audio channel. Modified video may help blur the details of an individual's activities, however, thus helping to protect privacy in the latter four of the privacy dimensions listed above.

For example, consider the case of a person who changes clothes in his or her office to go running, and who forgets to "turn off" the video feed of the media space office-share application. The use of a modified, altered video feed might blur the images enough such that the images could be transmitted and the person would not feel that her privacy was violated. Similarly, modified video might show that a person is meeting someone else in her office, but determining the identity of that other person may be impossible.

Image filters are natural candidates for transforming video images. After images are captured by a hardware device and before they are transmitted to other parties, image filters can change the contents of the images. Examples of image filters that hide details include a blurring filter, an edge-detection filter, and so on. Depending upon which filter is used in a video stream, users receiving the video may perceive more or less information about the person who is present in the video.

Each of the image filtering techniques obscures details to a certain degree while providing some level of presence information. However, it is not obvious how these image filtering techniques compare against each other, and whether they can give the user the flexibility of controlling presence and clarity.

The focus of this paper is to evaluate the effectiveness of several filtering techniques for communicating status and supporting presence. We sought to understand how well video-filtering techniques convey or hide details of the images' contents, that is, who is present and what they are doing. First, we describe several image filters to introduce the reader to the set of techniques. Next, we describe a comparative study to evaluate how well the different filtering techniques suppress identity and activity in the video. Finally, we describe a prototype media space application that supports these different filtering modes and that has been deployed locally. We provide early user experiences with the system and describe how people have been using it.

IMAGE FILTERING BASED TECHNIQUES

Generally, the more clear that a transmitted video image is, the more an observer will be able to perceive details in the image. An ideal media space application would convey only the details necessary to promote awareness and the intended use of the tool, while suppressing other details. Such an ideal application probably never will exist, but our goal was to understand how different image filtering techniques may or may not convey status details such as identity and activity.

The NYNEX Portholes system uses blur filters to process video images before making them available on the network [10, 11]. A blur filter usually refers to a process that averages neighboring pixels in an image to produce a new, blurry image. Repeated applications of a blur filter produce incrementally blurrier images. A NYNEX Portholes user can control the cloudiness parameter of video images of her being transmitted to other users.

Other filtering techniques already exist and are worth considering. Specifically, we have been experimenting with the following image filters: a pixelization filter, an edge-detection filter, a shadow-view filter and one of its variations. To illustrate the differences among these techniques, we present images shot from the same scene using the different filters (Figure 1-a is a regular image not processed by any image filters).

Pixelization Filter

A simple pixelization filter divides an image into a grid of eight-pixel wide by eight-pixel high blocks. Then within each block, the filter calculates the average intensity and color values, and assigns them to all the pixels in that block. The effect is that the result image appears to be made of many uni-color squares, and some details in the original image are lost (Figure 1-b).

Edge-Detection Filter

An edge-detection filter produces a new image that only includes edges in the original image (Figure 1-c). An edge is a boundary of sudden intensity changes. A pixel is likely to be on an edge if at that location, the maximum rate of intensity change per unit distance in all directions is great. A simple way to compute an approximation of edges in an image is to apply the Sobel operators to the original image [6].

Shadow-View Filter

The shadow-view filter [8] assumes that the camera position and orientation do not change. At a user-specified time, usually when the field of view of the camera is empty, the filter saves the image of the scene as background. After the background image is taken, the algorithm compares live video images against the

background image. If there is something that comes into the scene and does not appear to be part of the background, the live video image should differ than the saved background image. Therefore, the shadow-view filter records average intensity changes in the grid of eight pixel by eight pixel blocks to create a new image: for each eight-by-eight pixel block in image space, if the intensity change is small, the filter copies the block of data in the background image to the corresponding position in the new image; if the intensity change is greater than a threshold, the filter copies a pixelized representation of this block in the background image to the new image. In addition, if the changed block in the background image is dark, the filter brightens its pixelization in the new image. Likewise, if the block is bright, the filter darkens it. The resulting synthesized image is then transmitted instead of live video, giving the visual effect of a ghostly shadow of the new object (Figure 1-d).

(a) unfiltered image

(b) pixelized image

(c) edge-detected image

(d) shadow image

(e) live-shadow image

Figure 1. A regular image (a) and the images resulting from the application of four different filtering techniques (b)–(e)

The shadow-view algorithm saves the intensity changes of each block with regard to the background image and reduces them periodically. It brightens or darkens each changed block according to the combined amount of new and old changes. This process effectively produces a vague motion trail of the "shadow" in the scene. The motion trail gradually fades away if the object or person in the scene remains motionless.

Global illumination changes, occurring for example when a light is turned on, can cause false motions in the images. One solution is to pre-process each image with a blur filter then followed by a histogram equalization filter [6] to lessen the effect of lighting change. Another option is to provide the user with an easy way to re-take the reference image. For example, after the user issues the reset command, the application pauses, letting the user to leave the field of view of the camera, then the program takes a snapshot of the new scene as the new reference image.

Live-Shadow Filter

While hiding some level of details, the shadow-view may not be sufficient to provide enough awareness information about the user. Under normal conditions, a viewer could have problems recognizing the moving object in shadow-filtered video. If instead of darkening or brightening the pixelizations of changed blocks in the background image, we blend them with pixelized blocks of the most recent live video image, the resulting image exposes more information about the current scene. This variation of this shadow-view algorithm is thus called the "live-shadow" technique (Figure 1-e).

One simple way to think about the shadow algorithms is that they are like pixelizers for objects that move or are alien to the usual background scene which itself is presented unmodified.

We were curious about how these different image filters might affect the utility of a media space application. Understanding and assessing the effectiveness of the image filtering techniques include evaluating how much detail one can perceive from watching a sequence of filtered video images, and comparing this with the amount of information needed to communicate status.

Use of the Portholes system [4] showed that in a media space environment, people often are interested in other people's presence, availability, and interruptability. This kind of information can sometimes be inferred from seemingly unsubstantial artifacts in the environment. For example in a private office, if the occupant is on the telephone or talking with a guest, it is usually impolite to interrupt her. Depending on her work habits, an empty office with the door open may signal that she is around and will return soon.

In our study, we sought to learn how well people could interpret a scene viewed through the different image filters. For example, could a viewer detect the presence of a person in the video? If so, could they identify the person and/or the person's activity? If the viewers can recognize the activities in the video streams, they can probably infer availability and interruptability information. This recognition, however, also potentially conveys private information. Conversely, if viewers cannot recognize and identify activity without contextual clues such as whose office it is, a bit more of the individual's privacy may have been preserved. Though in this case, one must check whether the original purpose of the viewer has been fulfilled, such as determining whether a person is available or interruptible.

QUANTITATIVE STUDY

We conducted a study to help evaluate the effectiveness of the specific image filters in delivering information about a remote space. This was done by showing filtered video segments to users, then determining whether the users could perceive awareness-related information correctly.

We recruited five students as actors in preparing the videos for user testing (Figure 2). The individuals were chosen so that all were moderately similar in appearance (males, relatively short hair, no glasses, and white shirt), thus simulating a type of worst-case scenario for identification purposes.

For each of the actors, a portrait picture was taken, then a series of video segments were shot. The office set up for filming the video included a table, a workstation with a monitor, a camera, and two chairs. The camera was placed next to the monitor, facing the primary chair where a user of the workstation would likely sit, with a glance of the doorway. The image of an actor's head occupied roughly one-ninth of the total image area in the video.

We filmed video segments of each actor performing four different but typical office activities: looking at a computer monitor at eye level, talking on the phone, meeting with a second actor while facing away from the camera (the second actor sat further away from the camera so the subjects would know which to identify), and flipping through a magazine on his lap. We also recorded a special video segment of an office without any occupants, in which the door was open and an actor passed by the doorway. Each video segment lasted 15 seconds to leave enough footage for editing. Then each segment was processed through the four filters: the pixelizer, the edge detector, the live-shadow filter, and the shadow-view filter. Finally, five seconds of the most representative portion of each processed video segment was saved to disk, along with the unfiltered version.

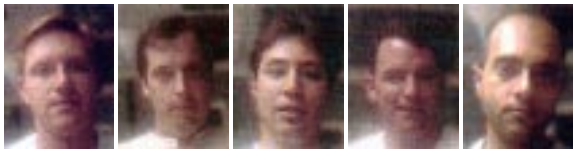

Figure 2. The five student actors

Subjects with adequate or corrected vision participated in the study. Before each set of tests, a subject was given the five portrait images of people possibly in the video streams. We showed five randomly selected warm-up video segments with brief verbal explanations prior to the formal tests to allow the subject to become familiar with the different filters, and to help reduce misinterpretations.

The subjects in the study viewed twenty-one video clips in a session (five actors doing each of the four activities plus the one empty room clip). Each subject viewed the same order of these actor-activity-pairing segments. The image filter utilized on the video clip and the image size (80 by 60 pixels or 320 by 240 pixels) of the clip were varied randomly, however. For each video segment, we asked the subject questions according to the decision tree in Figure 3.

Results

Twenty people participated in the study. All of the subjects except one were unfamiliar with the student actors in the videos. The exception only knew one of the five actors very well. Some of the subjects may have seen some of the actors before, but they did not know the actors personally.

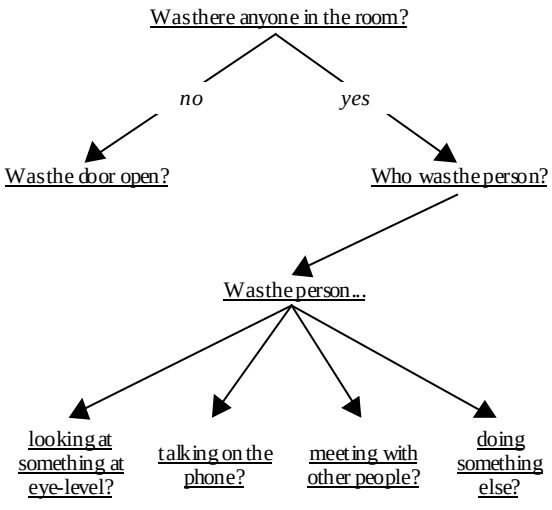

Figure 3. Decision tree for asking questions

Regardless of the image filter used, all of the subjects responded correctly to the empty room scene and noted that the door was open. In less than one percent of all trials with an actor present, a subject identified the room as being empty. This occurred because the actor was not moving much and the shadow-view filter was used, transforming the actor semi-transparent. However, in all these cases, subjects did guess correctly that the door was closed.

The chart in Figure 4 lists the correct activity recognition totals and percentages for the different image filters. Note that this is cumulative data summed over all different actors and activities. Note how all the filters supported high activity recognition levels (90% and up) except for the shadow-view filter. With it, subjects identified the correct activity about 60% of the time with both image sizes. The data is also broken out in Figure 5 according to the different activities (for simplicity, the chart for the empty room scene is not shown).

The chart in Figure 6 lists the correct actor identification totals and percentages for all the different image filters. Again, this data is summed over all actors and activities. The data is broken out by activity in Figure 7.

Note how identity was uniformly more difficult to recognize than activity (Figure 6 vs. Figure 4). As occurred for activity recognition, the shadow-view filter again exhibited the lowest correct actor identification percentages. Here, however, the other filters exhibited correct actor identifications below the 90% level found in activity recognition. The live-shadow filter showed a marked difference between the two correct recognition rates, particularly at the small image size (95% correct activity recognition vs. 53% correct actor recognition). This may be important for an application seeking to transmit activity information while suppressing individual identification, such as a view into a common area, for example a departmental copy-machine room. If an application seeks to suppress both actor and activity information while still conveying whether an individual is present, clearly the shadow-view filter would be the best.

Finally, the chart in Figure 8 lists the totals and percentages for a correct identification of *both* actor and activity in the same scene for all the different filters.

Please note that our research, and this study in particular, did not directly address the tie between the conveyance of particular types of information (identity, activity, presence, etc) and the more subjective, personal notion of the loss of privacy. This study simply assessed how effective the different image filters were at conveying or suppressing different types of information over a video stream. However, it should not be difficult to imagine how changes in the information conveyed could affect the notion of more or less privacy being surrendered.

In the next section, we describe the deployment of a simple media space application armed with these different filtering

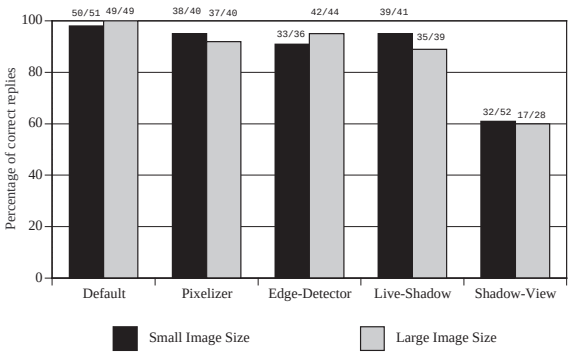

Figure 4. Activity recognition (collapsed across all different actors and scenes)

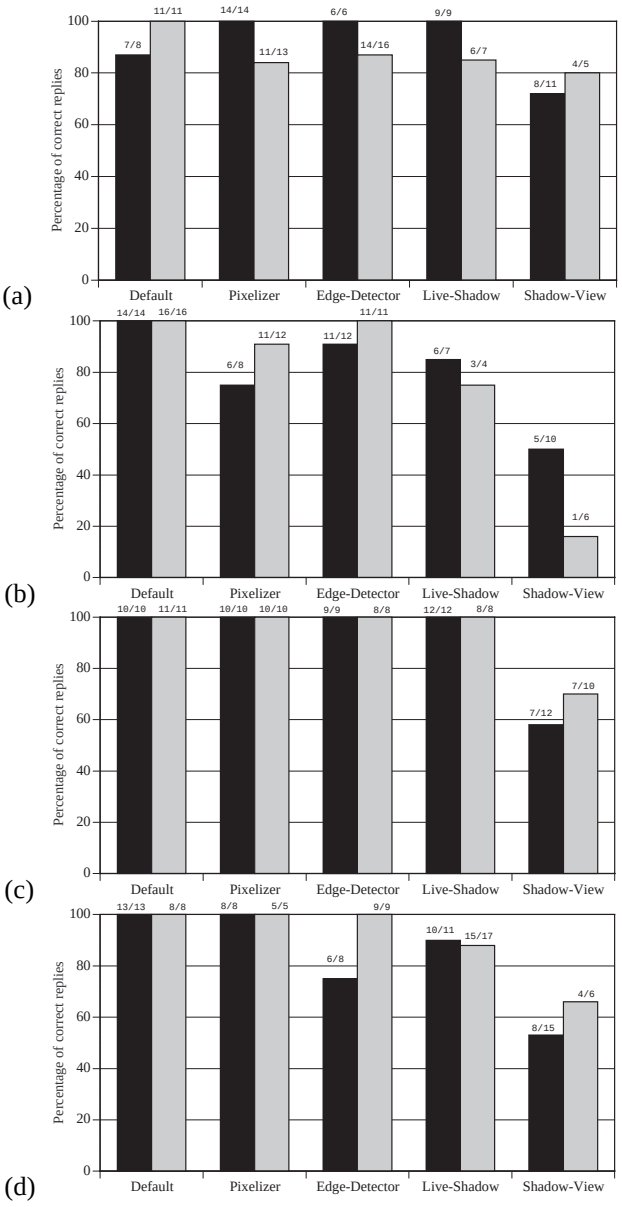

Figure 5. Activity recognition: charts (a)-(d) represent the computer scene, the phone scene, the meeting scene, and the magazine scene, respectively

tools, as a way of gauging individuals' views on the effectiveness of the image filtering techniques in video space applications.

QUALITATIVE STUDY

In order to acquire more subjective data about the use of different video techniques in a casual group awareness system, we prototyped a video space application based on a modified version of "vic" [12], a popular Internet video conferencing tool.

On start up, the program displays a collection of thumbnail images (80 by 60 pixels) of available video sources in a video space (Figure 9). By clicking on the "Capture" button, the user can start or stop capturing and transmitting video from the camera connected to his or her computer. The "Options" pull-down menu allows the user to switch among the different image filtering modes plus the no filtering mode, and built-in static image-notes such as "at a meeting" or "do not disturb". It was important to us to allow a user to select the video filter being used to broadcast her signal to all other users. This menu also has an option to pop up the control panel, which allows the user to fine tune parameters of the transmission, such as the video capture hardware device to use, the frame- and bit-rate bounds, etc. The "Members" button pops up a participant list that shows everyone known to the current video session, including those not sending a video stream. This participant list also allows the local user to choose which video streams to receive and which video streams to ignore.

If the user clicks on a thumbnail image, a larger window of the same video source pops up. The size of this window varies depending on the type of hardware its source uses, but is usually close to 320 by 240 pixels. This larger view also includes other information about the video sources, such as a text note posted by its owner, frame rate and other statistics (Figure 10).

Adding the image filters accounted for a majority part of the modifications done to the original vic. Besides the previously described four images filters, we added an "Activity Only" mode as an extreme in providing the least amount of information about a user. In this mode, a bar chart of overall image intensity differences was transmitted instead of video (similar to that in [11]).

Other modifications to the original vic are mostly user interface related. For example, the original vic displays transmission statistics of each video source along with the thumbnail video images in the main application window. When this tool is used in a media space setup, the statistical information is less frequently needed in the main display area. To keep the top-level application window small and well utilized, we moved statistical information displays to the detail windows associated with individual video sources.

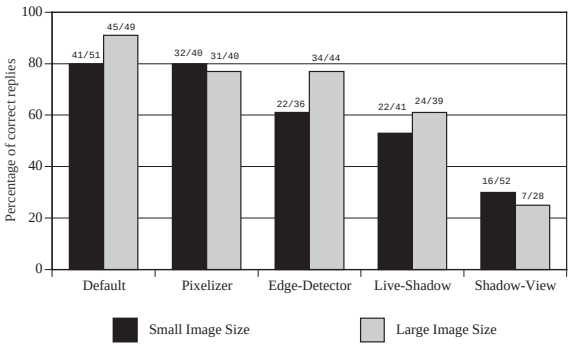

Figure 6. Actor identification (collapsed across all different actors and scenes)

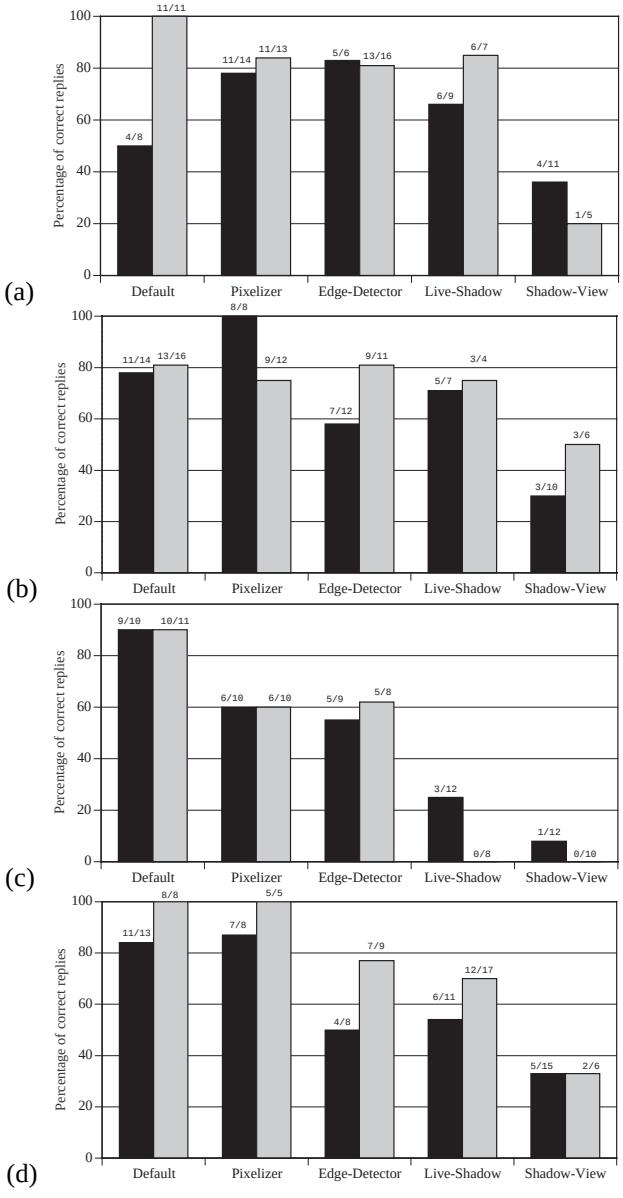

Figure 7. Actor identification: charts (a)-(d) represent the computer scene, the phone scene, the meeting scene, and the magazine scene, respectively

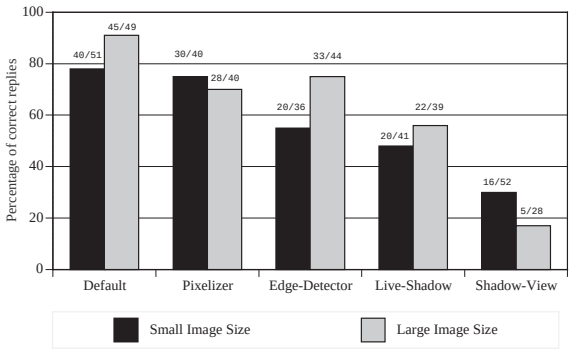

Figure 8. Accuracy levels in correctly identifying both activity and actor

Usage Feedback

We set up a special media space called the "Electric Lounge". A number of volunteers within the local community participated in this study and connected to the Electric Lounge using our enhanced system. However, due to video capturing equipment shortage and user on-line time variations, the number of simultaneous users in the Electric Lounge ranged from four to about ten.

Initially, users had a little trepidation about participating in the Electric Lounge. Fairly quickly, however, users became accustomed to having the video space application running.

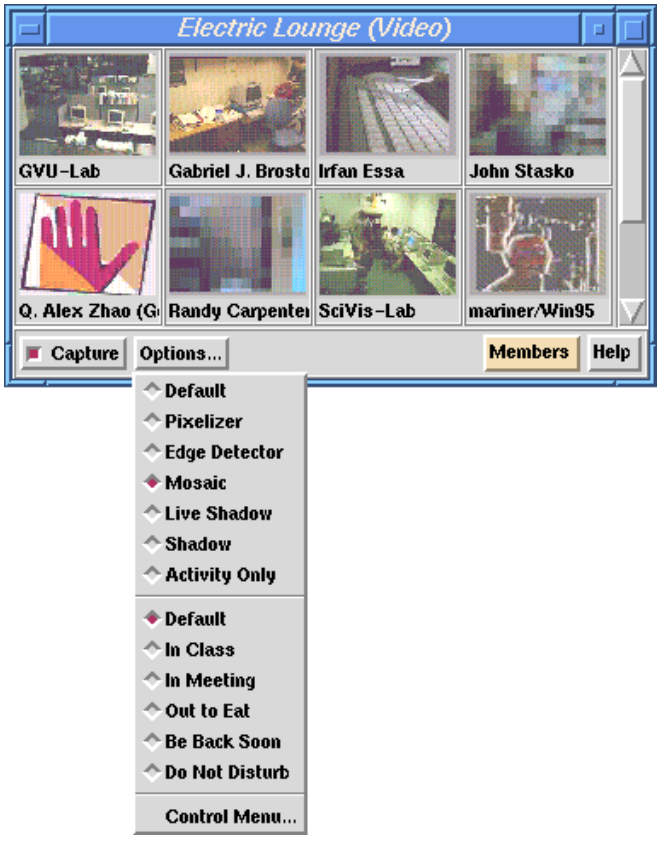

Figure 9. The main display area of the prototype video space application

As curiosity and novelty wore off, people tended to forget about the video space when working on other tasks, even if the videos were visible on the desktop. They used the video space to check the status of other people with whom they interacted. Sometimes the users would glance the gallery of video images to get a vague sense of what was happening in the virtual proximity. They also preferred to use the Electric Lounge to express some information about themselves, such as "not in office right now" or "do not disturb", rather than leaving the video space by quitting the application.

From informal interviews, we found that the possibility of identifying someone in a filtered video changed with familiarity level. For example, someone familiar with the user might be able to guess if she was in a pixelated video based on shirt color, or hairstyle, or the geometry of her face. Especially in a tightly related group of people with frequent collaborations, identifying the person in the video was not difficult because people were extremely familiar with each other, and the video thumbnails were labeled. In this case, people could reliably judge the availability and interruptability of other users by watching filtered videos, even if the details of the activity in a video were not available. Without being able to recognize most of the gestures, people used the Electric Lounge many times to watch a remote office and wait for the guest in the room to leave before making a visit or placing a call.

We also observed that users did not often change among the different filters broadcasting their signal. People tended to choose one filter and stayed with it. The shadow and live-shadow views were seldom used, usually because they involved a background setup process. The pixelization filter was often chosen – it seemed to convey a reasonable level of information while also blurring fine-grain details. The "Activity Only" bar chart filter was virtually never used.

Several users suggested that they might not be interested in what the background looked like, and it could be made even blurrier than the foreground. To test out this idea, we added a fifth image filter, the "mosaic" filter (Figure 10). Instead of using a background, the mosaic filter kept a two dimensional array of intensity change values calculated from consecutive video images. It periodically lowered the recorded intensity change levels to gradually lessen the effects of old intensity changes. Before each image redraw, the mosaic filter painted blocks that had motion in higher resolution and static blocks in lower resolution. Effectively, the mosaic filter serves as a form of pixelizer with fine-grain pixelization in areas of motion and coarse-grain pixelization in static areas. Since the mosaic filter was implemented after we started the evaluation studies, we do not have any quantitative data or user comments about this filter yet.

CONCLUSION

This article describes studies to evaluate the effectiveness and utility of different image filtering techniques used in

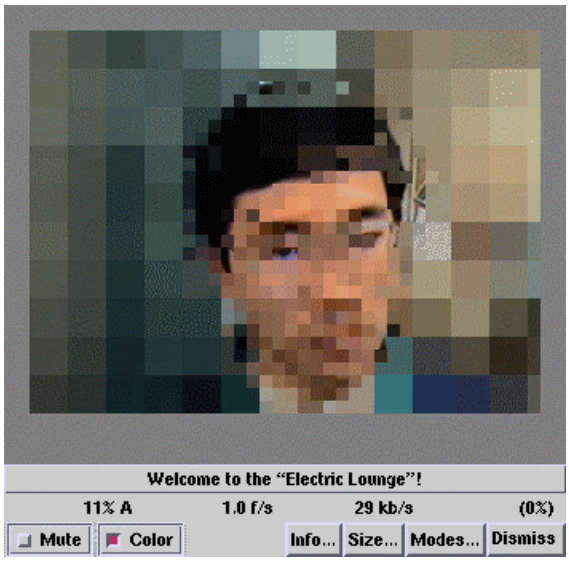

Figure 10. The close-up view of one video source (video image shown as processed by the new "mosaic" filter)

media space applications. We described a number of existing filtering techniques, and introduced two new methods. We believe that these techniques can be useful tools for software designers to help control interpersonal access in media space applications. In addition, we presented a study to quantitatively assess the ability of these different filters to convey or suppress activity, identification, and presence information over a video stream. Further, we presented a video space application that utilizes these filtering techniques, and we discussed early user feedback on the program.

ACKNOWLEDGEMENTS

We wish to thank Richard Catrambone for helpful discussions on the analysis of the data. We also thank everybody who participated in the studies, and the vic team for making the source code publicly available.

REFERENCES

1. Ackerman, M., and Starr, B. *Social Activity Indicators: Interface Components for CSCW Systems*. In UIST'95 Conference Proceeding, 159-168. ACM, 1995.
2. Bellotti, V. *What You Don't Know Can Hurt You: Privacy in Collaborative Computing.* In HCI'96 Conference Proceeding, British Computer Society. Springer-Verlag, 1996.
3. Bly, S. A., Harrison, S. R., and Irwin, S. *Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment*. Communications of the ACM, Vol. 36, No. 1, 28-47. ACM, January 1993.
4. Dourish, P., and Bly, S. *Portholes: Supporting Awareness in a Distributed Work Group*. In CHI'92 Conference Proceeding, 541-547. ACM, 1992.
5. Fish, R. S., Kraut, R. E., Root, R. W., and Rice, R. E. *Video as a Technology for Informal Communication.* Communications of the ACM, Vol. 36, No. 1, 48-61. ACM, January 1993.
6. Gonzalez, R. C., and Woods, R. E. *Digital Image Processing*. Addison-Wesley, 1992.
7. Greenberg, S. *Peepholes: Low Cost Awareness of One's Community*. In CHI'96 Conference Companion, 206-207. ACM, 1996.
8. Hudson, S. E., and Smith, I. *Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems.* In CSCW'96 Conference Proceeding, 248-257. ACM, 1996.
9. Kraut, R. E., Fish, R. S., Root, R. W., and Chalfonte, B. L. *Informal Communication in Organizations: Form, Function, and Technology*. In Oskamp, S., and Spacapan, S. (Eds.) People's Reactions to Technology in Factories, Offices, and Aerospace, The Claremont Symposium on Applied Social Psychology, 145-199. Sage Publications, 1990.
10. Lee, A., Girgensohn, A., and Schlueter, K. *NYNEX Portholes: Initial User Reactions and Redesign Implications*. In GROUP'97 Conference Proceeding, 385-394. ACM, 1997.
11. Lee, A., Schlueter, K., and Girgensohn, A. Sensing Activity in Video Images. In CHI'97 Extended Abstracts, 319-320. ACM, 1997.
12. McCanne, S., and Jacobson, V. *vic: A Flexible Framework for Packet Video*. In Multimedia'95 Conference Proceeding, 511-522. ACM, 1995.
13. Tang, J. C., Isaacs, E., and Rua, M. *Supporting Distributed Groups with a Montage of Lightweight Interactions*. In CSCW'94 Conference Proceeding, 23-34. ACM, 1994.
14. Wax, T. *Red Light, Green Light.* In CSCW'96 Conference Companion, 1-2. ACM, 1996.