The Effects of Filtered Video on Awareness and Privacy

Michael Boyle[1], Christopher Edwards[2] and Saul Greenberg[1]
[1]Department of Computer Science and [2]Department of Psychology
University of Calgary, Calgary, Alberta, Canada T2N 1N4
+1-403-220-6087
[boylem,saul]@cpsc.ucalgary.ca; cjedward@ucalgary.ca

ABSTRACT

Video-based media spaces are designed to support casual interaction between intimate collaborators. Yet transmitting video is fraught with privacy concerns. Some researchers suggest that the video stream be 'filtered' to mask out potentially sensitive information. While a variety of filtering techniques exist, they have not been evaluated for how well they safeguard privacy.

In this paper, we analyze how a blur and a pixelize video filter might impact both awareness and privacy in a media space. Each filter is considered at nine different levels of fidelity, ranging from heavily applied filter levels that mask almost all information, to lightly applied filters that reveal almost everything. We examined how well observers of several filtered video scenes extract particular awareness cues: the number of actors; their posture (moving, standing, seated); their gender; the visible objects (basic to detailed); and how available people look (their busyness, seriousness and approachability). We also examined the privacy-preserving potential of each filter level in the context of common workplace activities. Our results suggest that the blur filter, and to a lesser extent the pixelize filter, have a level suitable for providing awareness information while safeguarding privacy.

Keywords
Media spaces, always-on video, awareness, casual interaction, distributed groupware.

INTRODUCTION
The backbone of everyday coordination and work between co-located team members is *casual interaction*, the spontaneous and one-person initiated meetings that occur over the course of the day [11]. The glue behind these interactions is *informal awareness*, where people track and maintain a general sense of who is around and what others are up to as they work and mingle in the same physical environment [11,3].

Yet casual interaction is problematic in distributed communities. It is no surprise that casual interaction drops exponentially with distance [11]; awareness of others and

consequently opportunities for interaction diminish considerably when people are out of sight. Substituting an electronic communication channel is not enough: while groupware is readily available, people still have considerable trouble establishing real-time electronic contact [3]. If casual interaction is to be supported, systems must also provide community members with a measure of awareness of who is around and how available that person is for conversation, as well as a very lightweight means to move from that awareness to an encounter to communication and work.

This explains the popularity of Instant messenger systems (e.g., ICQ™, MSN Messenger™), for they provide a very simple form of on-line presence that can be acted upon. However, these systems lack fidelity, and as a result one person may interrupt others at inappropriate times or when they are, in fact, absent.

Many researchers [2,3,6,7,8,10,12,13,14,15] have suggested that always-on video may be a means of providing distributed workgroups with high-fidelity presence and availability awareness. Such systems use open video links that, once set up, connect remote places into a single, shared video media space. In practice, these systems do promote casual interaction. The problem is that, even in these relaxed, research-oriented environments, obvious concerns over privacy and solitude exist [1,2,7,8,9,12,13,15].

To counter concerns about privacy, others have investigated techniques for protecting privacy by altering what appears on the video [4,8,10,12,13,15]. One common technique, often seen on news broadcasts, is to 'pixelize' sensitive areas of the image—for example, a person's face can be replaced by large pixels (squares). While we can tell that it is a face and how it is moving, identity is masked. Other standard image processing techniques have been applied as well, such as full scene pixelization, a scene altered to show only edges, posterizing effects, blurring, venetian blinds, and so on [4,8,12,13,15]. Advanced techniques have also been developed. For example, Hudson and Smith [10] describe a shadow-view filter that gives the effect of a ghostly shadow moving about a static scene. Crowley et. al. [4] uses eigen-filters to analyze a scene and reconstruct its images in a 'socially-correct' form.

Though filtering techniques are now being deployed in a number of prototype systems, there has been only one evaluation of how filters reveal or mask awareness information. Zhao and Stasko [15] examined four filters on brief five-second video clips shown at two sizes (80×60 and 320×240 pixels). Their study asked volunteers to identify which of five of actors were featured in a clip and which of four activities the actors were engaged in. Volunteers were primed with key information ahead of time. They were shown portraits of the actors as well as the kinds of activities to look for. The results for activity recognition show that, with the exception of the shadow-view filter, all the filters tested supported high (>90%) activity recognition. Actor identity was more difficult to recognize, with only the pixelize filter and uniform lens combination able to support moderate (>75%) actor identity recognition levels. From a separate qualitative study that evaluated the techniques in the context of a video media space application, it was found that actor identity recognition improves with familiarity.

Yet Zhao and Stasko's study is flawed in four ways. First, the authors made no specific mention of the filters' effects on privacy. Thus it is difficult to tell if they offer any advantage over an unfiltered scene. Second, the authors only investigated one level of these filters, even though most filters can be adjusted to give varying degrees of fidelity. An example of what we mean is illustrated in Figure 2, where we see two filters (a blur and a pixelize filter) applied at nine different levels. At low levels, the filter masks a great deal of information, while the high levels show the image at reasonable fidelity. We cannot really pre-judge a particular filter's ability to show awareness information while safeguarding privacy unless we considered it across a continuum of levels. Third, the scenes are not described; we have no idea of where cameras were positioned, how close actors were to it, and so in. Finally, the subjects were primed with all possible selections ahead of time: this meant they only had to discriminate between a few choices instead of interpreting a scene.

In this paper we consider the effects of a blur and a pixelize video at various levels for their impact on both awareness and privacy. We begin by introducing the study, where we describe both our methodology and the filtering effects in detail. We then present our results, including descriptive statistics to illustrate our points. We close by discussing where these filters appear effective, and by indicating some outstanding issues.

METHODOLOGY

Hypothesis
The first hypothesis considers the effect video filtering may have on a person's ability to extract awareness information from a scene, while the second hypothesis considers the effect filtering may have on a person's perception of privacy afforded by that filter.

Hypothesis 1. There is no difference in a viewer's ability to correctly identify particular awareness cues from one of five different QCIF-size video scene sequences that have been altered by either a pixelize or blur filter, where these filters were applied at ten levels ranging from heavily obscured to unfiltered effects.

Hypothesis 2. There is no difference in how a viewer rates these filtered video scenes in terms of how they protect privacy.

Materials: The Video Sequences
We created 95 video sequences by applying two filters at nine different filter levels to five different scenes (the 10th level is the unfiltered video). These are described below.

Scene 1 (42s): *a coffee room.* A seated male actor reads a newspaper. A 2nd male enters, gets coffee, talks to the 1st person while gesturing over the newspaper, and then leaves. A 3rd female walks by the doorway.

Scene 2 (25s): *personal office, side view.* A male actor, glancing occasionally at his computer screen is eating a snack. He drinks from a can, then reaches over to grab and open a bag of potato chips.

Scene 3 (31s): *personal office, view from doorway.* A single male actor is seen talking on the telephone while looking at his computer screen. The clip ends just after the actor hangs up and starts working on his computer.

Scene 4 (86s): *large public office plus counter.* A male actor arrives at the counter, and the female actor gets up from her desk to serve him. She retrieves something off screen, and returns. He leaves just as a 3rd male comes to the counter. A 4th male walks by in the background.

Scene 5 (27s): *over the shoulder, personal workstation.* Female actor looks at workstation while typing and moving the mouse. On-screen activities, such as window resizing, are visible.

Figure 1. The five scenes used in this study.

Scenes. We videotaped five video sequences portraying 'typical' media space settings (Figure 1). Sequences were shot using a high-quality Canon XL-1 digital video camera. While we could have used a teleconferencing PC camera, its lesser quality could have compromised our study with inferior video; we also expect camera technology to improve over time. No special lighting conducive to filming was done, as we felt this would give unrealistic footage. Thus similar to many media spaces, some scenes are poorly lit e.g., backlighting, glare, etc. We converted sequences to a frame rate of 24 frames per second at standard QCIF videoconferencing image size (176×144 pixels, 24 bits per pixel), saved as an AVI file suitable for replay on a PC.

Video filters. Using our own special video filtering software, we pre-processed each scene at nine levels of the pixelize and blur filters (Figure 2). Both filters use standard image processing algorithms, as summarized below.

The *pixelize filter* produces a "mosaic" effect across an image (lower images in each row of Figure 2). It works by re-sampling an image (a single video frame) using a coarser grid. The image is divided into rectangles of equal area, where all of the pixels in an area are reset to the mean of their original colors. The pixelize filter is computationally inexpensive and can be applied in real

time to video stream.

The *blur filter* produces a smooth blurring effect across an image (Figure 2, upper images in each row). It is a box filter, meaning that a pixel's filtered color is the mean of the neighborhood of pixels surrounding it. Unlike the discrete regions seen with the pixelize filter, the image smoothly changes from one region to the next. Our implementation of the blur filter was compute-intensive; similar but computationally inexpensive filters (e.g., Deriche IIR [5]) could be used instead for real time video manipulation.

These effects can be applied at different *filter levels*. Each level is associated with a "box size", chosen so that they nicely divide into the 176×144 pixel dimensions of the QCIF image. For example, the level 2 box size of 88x72 (see Figure 2) divides the 176x144 pixel image into four boxes. For the pixelize filter, this box size corresponds to the dimensions of each "macro-pixel" in the filtered image. For the blur filter, the box size corresponds to the neighborhood of pixels used to determine a pixel's value.

We applied the two filters at 9 different filter levels to each of the five scenes (Figure 2 illustrates one frame from one scene using both filters at all levels). Levels 1-9 range from heavily to lightly filtered. Level 10 is the unfiltered sequence. This gave 95 sequences: 5 (video scenes) x 2
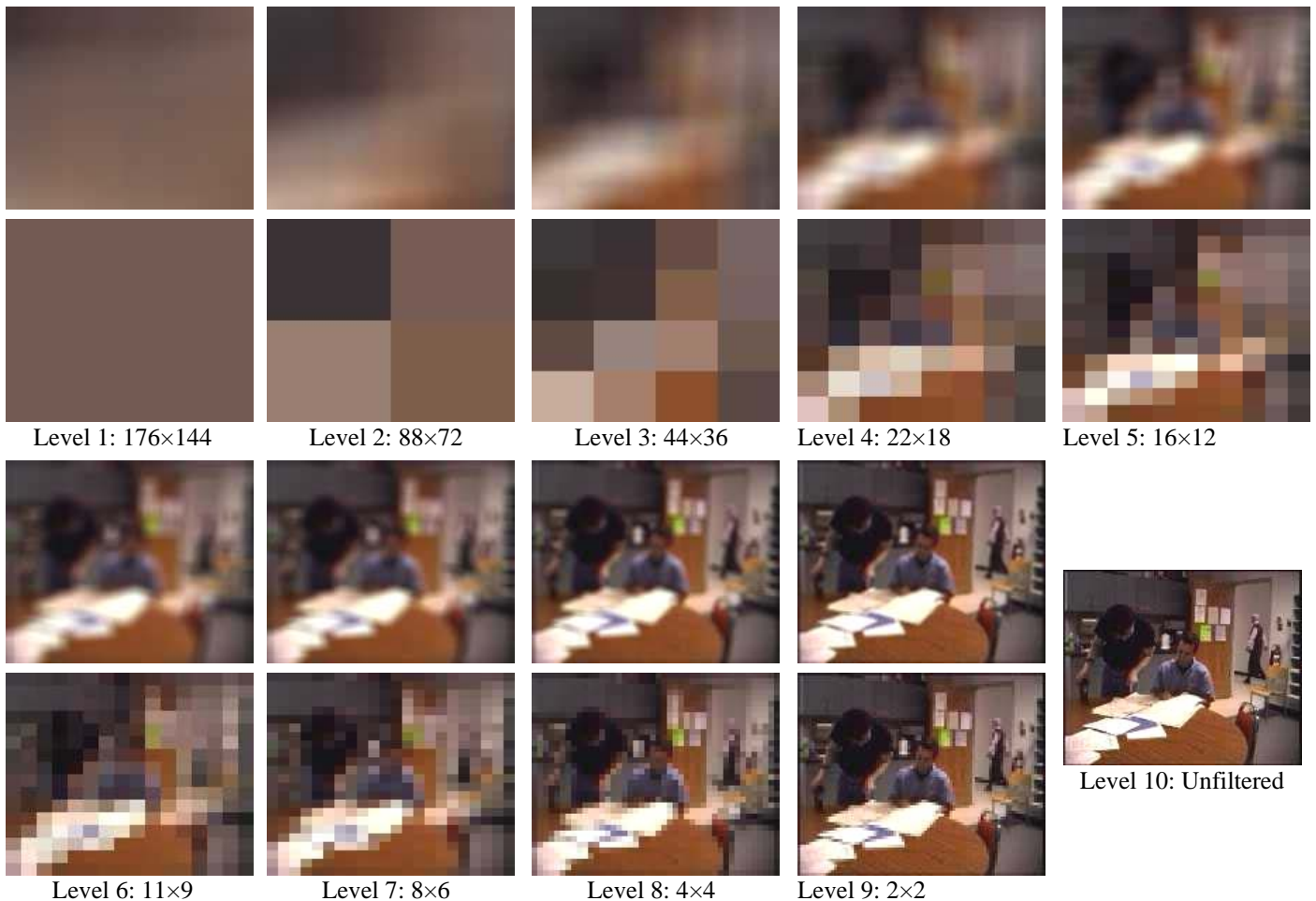

| Level 1: 176×144 | Level 2: 88×72 | Level 3: 44×36 | Level 4: 22×18 | Level 5: 16×12 |

| Level 6: 11×9 | Level 7: 8×6 | Level 8: 4×4 | Level 9: 2×2 | Level 10: Unfiltered |

Figure 2. Examples of the two filters and the nine filter levels tested applied to a frame from scene 1. Level 10 is the original scene.

(filter types) x 9 (filter levels) + 5 (unfiltered video scenes).

Materials: Questionnaires

We designed three questionnaires to gather experimental data.

Pre-test questionnaires captured standard data about each volunteer, including age, gender, and computer and groupware experience. The questionnaire also queried the subjects about:

- their willingness to give out personal information over the Internet (e.g., date of birth, credit card number);
- their perception on whether their privacy is protected when using a computer;
- their willingness to let personal acquaintances view their video image on another computer;
- their willingness to let strangers view their video image on another computer, given that reciprocity would be enforced.

During-test questionnaires captured how a person perceived a single video sequence using a particular filter at various filter levels. We developed a special web-based system that automatically displayed the appropriate video clip, where subjects could fill in the questionnaire by selecting a combination of radio buttons, sliders, and by typing into text fields[1]. Figure 3a illustrates a screen snapshot. The questionnaire asked people to identify (when feasible) scene features including:

- the number of actors present;
- whether each actor is seated or standing;
- whether each actor is moving or still;
- the gender of each actor;
- the objects in the scene and their location within it;
- the general activity of each actor;
- how busy or idle each actor appears;
- the 'tone' of the activity i.e., from casual to serious;
- how approachable or withdrawn each actor appears.

For each of these items, the questionnaire also asks people to rate their confidence of their assessment.

The software also included a special 'final' question that asked them to rate the privacy-preserving potential of a still snapshot taken from the middle of a video sequence and filtered at different levels i.e., from unprotected to protected (Figure 3b). Subjects could replay the filtered sequence by clicking a given snapshot.

Post-test questionnaires captured the volunteer's perceptions of the filter they tested.

- likes and dislikes;
- the situations where they would enable the distortion and what level(s) they would set it to; and

---

[1]The video unfiltered video scenes and this part of the experiment can be viewed at http://sern.ucalgary.ca/grouplab/filteredvideo/. When prompted, enter TEST-XXX, and select a scene number from 1-3, 5 or 6. For filter, chose P (pixel) or B (blur). We discard data beginning with this TEST prefix.

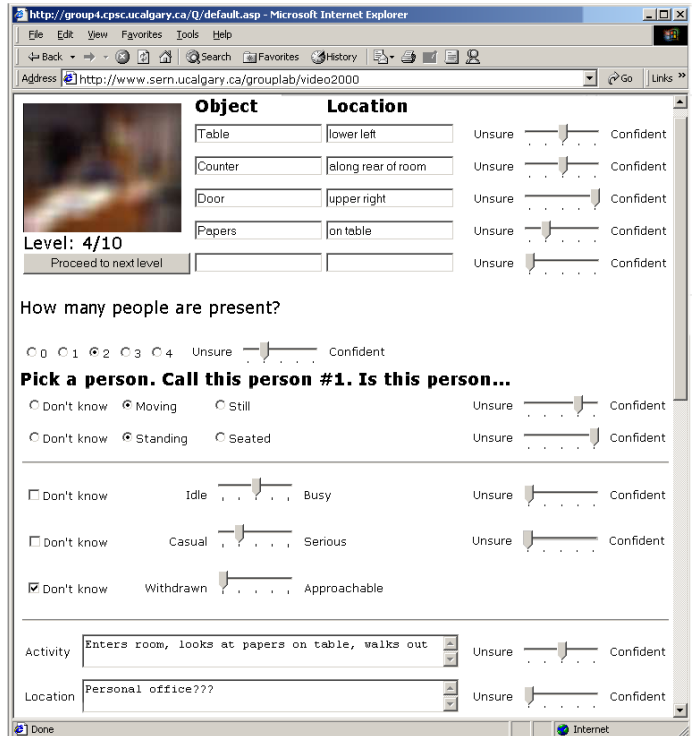- if they would use an open video link if it was filtered.

Figure 3a. The during test software. This screen snapshot of the questionnaire (modified for clarity as a figure), shows Scene 4, level four using the blur filter. The subject has indicated what objects they could see, attributes of the person and their activities, and their confidence in their statements.
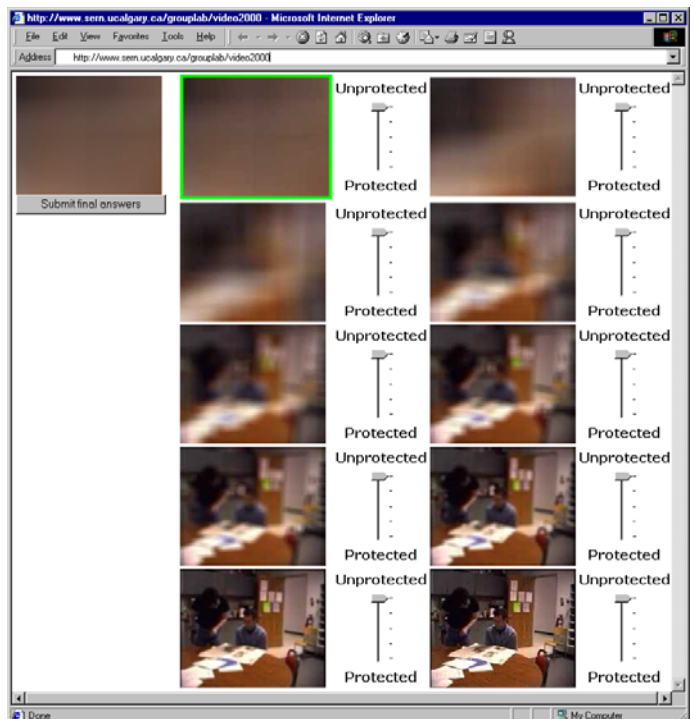

Figure 3b. The during test software. This last question shows a single frame from the middle of the scene, at different filter levels (in this case Scene 1 using the blur filter). Clicking a frame replays the filtered sequence. The subject can adjust the slider to rate the level of privacy offered at a particular level.

Experimental Design

Variables. The independent variables are scene type (5) x filter type (2) x filter levels (10). The dependent variables, captured via the questionnaires, were people's ability to identify correctly particular awareness cues, their confidence in their identification, and their perceptions of how privacy is maintained. Particulars were listed in the questionnaires section.

Subjects were recruited from junior-level university courses, mostly from computer science with some from other disciplines. They comprised 10 male and 10 female volunteers between the ages of 17 to 32 years. All subjects had good or corrected eyesight.

Subject assignment. We used a mixed design. Scenes and filter type were between subject: each volunteer was randomly assigned to one of the five scenes, and to one of the pixelize or blur filter. Filter levels were within subject: each volunteer saw that scene ten different ways: the 9 filtered levels and the 10th unfiltered view. Filter order was fixed, always beginning with a heavily filtered scene and ending with an unfiltered scene.

Procedure

Testing was largely self-directed and self-paced. Each subject was first assigned to a particular scene and filter. After completing a consent form and pre-test questionnaire, the subject started our software on a 19" 1280x1024 display. The software loaded the video scene filtered at level 1 (heavily filtered) and the 'during-test' questionnaire form for that scene (similar to Figure 3a). The subject then viewed this scene as many times as desired: the video clip operated on a loop, and the scene faded to back between loops. The subject then tried to identify features in the scene and their confidence in their reply by filling in the during-test questionnaire. When satisfied with their answers, the subject proceeded to the next filter level and questionnaire form. (Answers provided at the current level were stored for data analysis). Because previous responses were automatically copied to the new instance of the form, subjects would just modify their answers as they identified further features at each new level.

Once the subject completed all ten levels, the software asked the subject to rate the privacy-protecting potential of each filter level. Similar to what is seen Figure 3b, this form displayed a single frame extracted from the middle of each scene, but shown at the ten filter levels. Subjects were asked to imagine themselves as the main actor in the scene, and then to rate how they felt their privacy would have been protected had someone else been observing them. While we deliberately left the identity of the observer somewhat vague, only a few subjects requested clarification: these were told that the observer was a peer e.g., a co-worker.

RESULTS

Pre-test questionnaire.

In the pretest questionnaire, 12 of the 20 subjects answered 'yes' when asked if their privacy was protected when using a computer, and if they would be willing to provide personal information over the Internet. When asked if they would be comfortable if someone they knew could view them through an always-on computer-based video connection, 13 out of 20 responded 'yes'. Similarly, when subjects were asked if they would be comfortable if a stranger could view them over a reciprocal video channel—where they could see the stranger at the same time—12 responded 'yes' to the question. The subjects who responded negatively to these questions raised several concerns. Some said their answer would depend on the situational context i.e., whether they were in a business, academic or home setting. For video links, they were concerned about how well they knew the individual on the other side. Some displayed a general unwillingness to let others know what they were doing. Others stated a sense of discomfort with the idea of always-on video. What was somewhat surprising is that these answers were spread across subjects i.e., while it seems that about 12 subjects always answered positively, it was not the same 12 across the questions.

Identifying the Number of Actors in a Scene

We wanted to know how well subjects could correctly identify the number of actors in a particular video scene under different filters and levels. Since the number of actors featured varied with the scenes, we transformed our result into a normalized metric that would facilitate comparison across scenes. The normalized number of actors r was computed for each response i with respect to the correct number of actors n for that scene:

$$r = 1 - ((i\text{-}n) / n)$$

Using this metric, values equal to 1 mean that the subject correctly identified the number of actors present. Values less than 1 meant that they identified fewer actors than actually present, while values greater than 1 mean they overestimated the actors.

Figure 4 plots the mean normalized number of actors observed at each level for the two filters tested across all scenes; the error bars show the standard deviation. Although there was much variation, we see that subjects using the blur filter were more willing to take a guess at the number of actors early on. Subjects could also assess the number of actors in the scene with accuracy at 0.8 or above as early as level 3 with the blur filter, but not until level 5 with the pixelize filter. The amount of variation drops considerably with the blur filter by level 5.

We then analyzed the subjects' confidence of their guesses at each level of a filter. Not surprisingly, a person's confidence increases along with the filter level i.e., as the clarity of the image improves. There was little difference
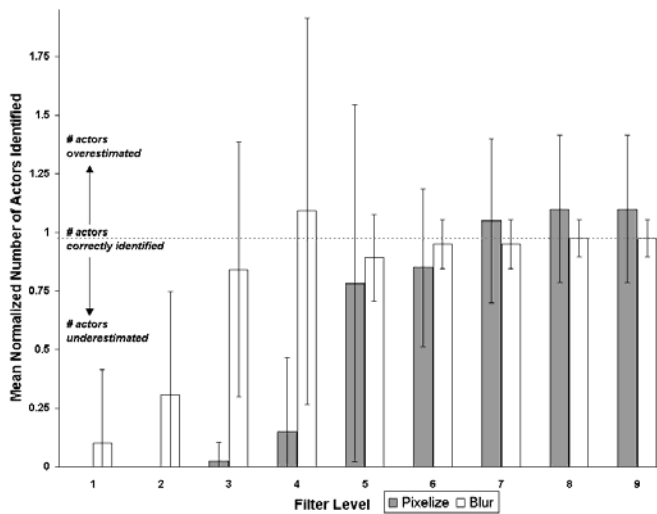
Figure 4. Normalized number of actors identified vs. filter level.

between the reported confidence and filter type; people seemed moderately confident in their answers by level 5.

We then looked at these results on a per-scene basis. In particular, scenes 1 and 4 contained background actors that were in view for only a brief period of time, and we wanted to know if subjects' guesses differed in these scenes. These scenes proved to be the ones that contributed most to under-estimation errors. However, most subjects correctly spotted this person by level 6 with both filters.

Identifying Posture

We wanted to know if subjects could identify the posture of the principle actor in a scene i.e., whether they were seated or standing, and whether they were moving around or stayed fairly still. We analyzed the data and determined, for each subject, the level (threshold) in a given scene where they correctly identified these posture attributes. Results are plotted in Figure 5 for each scene.

As seen in this figure, subjects correctly determine posture fairly early on with the blur filter: levels 2-3 for movement, and levels 3-4 for their seated / standing position. With the pixelize filter, subjects do not do this until later on: about level 5 for both movement and seated / standing position. In both cases, most everyone correctly states the posture of the main actor featured by level 6.

With both filters, people seemed to be able to identify posture at the same time they correctly identified the number of people in a scene. There is a strong correlation (0.97) between determining someone's seated or standing position with their ability to identify the number of actors in the scene.

Identifying Gender

We wanted to know how well subjects could assess the gender of the principle actor in each scene. This proved perhaps the most problematic category, in part because this is difficult to determine even at full fidelity because of the small size of actors portrayed and the poor lighting in some
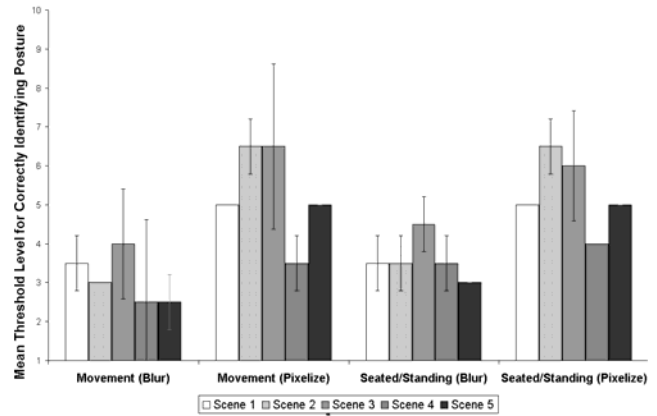

Figure 5. The mean threshold level where subjects were able to identify correctly an actor's posture, as measured by whether they could tell if the actor was moving, and if they were seated or standing.

scenes. While we do not show our analysis graphs here, we found that subjects are unable (or unwilling) to assess gender for the pixelize filter until levels 6 and 7, and even then only 30 – 40% did so correctly. By level 8, however, all subjects are able to correctly assess gender. By contrast, subjects appear more willing to assess gender earlier when viewing the scene under the blur filter, although they are often wrong. At levels 6-7, 60-65% of the subjects correctly assess gender, and most everyone gets it by level 8.

Identifying Objects in a Scene

As subjects viewed each scene across the different filter levels, the images would progressively reveal more information. We asked subjects to try and identify any objects visible within the scene as soon as they could. We then analyzed the raw data, where we looked at each subject's answer and decided at what level (i.e., the threshold) they had roughly and correctly identified some of the basic objects in the scene, and at what level they correctly identified some of the more detailed objects.

The results are plotted per scene in Figure 6. While there are differences between scenes, most subjects could
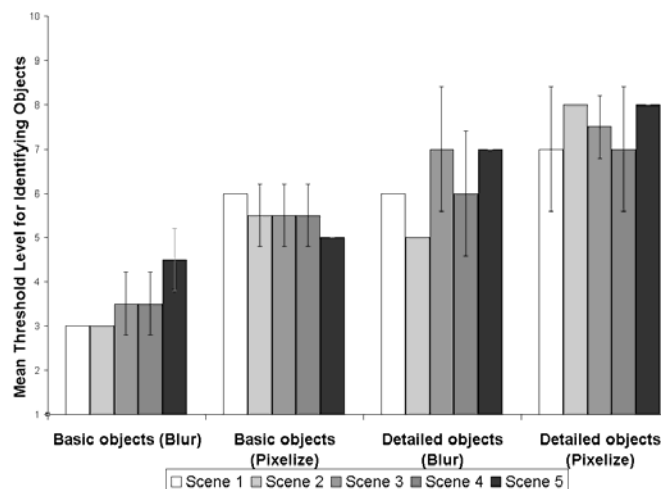

Figure 6. The mean threshold level where subjects were able to identify basic and detailed activities of actors in a given scene.
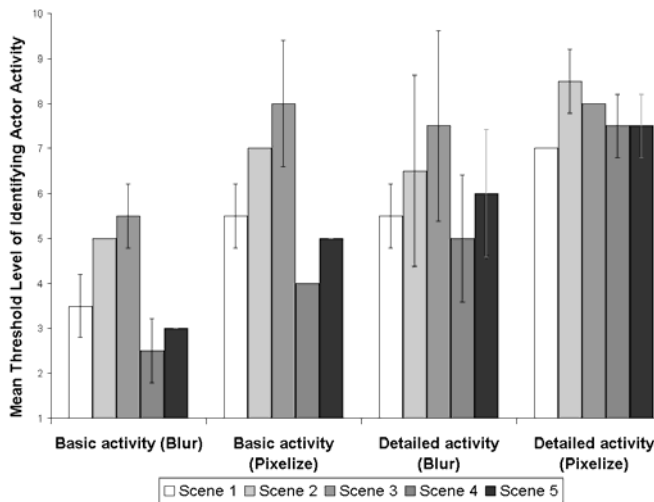
Figure 7. The mean threshold level where subjects were able to identify basic and detailed activities of actors in a given scene.

Figure 8. The mean threshold level where subjects indicated how busy, serious and approachable people appeared in a given scene.

roughly identify several basic scene objects by level 3 using the blur filter, but not until about level 6 with the pixelize filter. Similarly, subjects began identifying detailed objects by level 6 with the blur filter, but not until scene 7-8 with the pixelize filter. From our own personal observations and further analysis of subjects using the pixelize filter, we noticed there was a very small gap between when people first identified the basic objects in a scene, to where they were able to identify the objects in detail; that is, there was only about a one level difference.

Identifying Actor Activity in a Scene.

Similar to how we analyzed object identification, we analyzed activity identification by deciding at what level had subjects correctly identified the basic and then the detailed activities of actors. Results are plotted in Figure 7. Again we see that people identify basic activities using the blur filter earlier (around levels 3-4) than when they use the pixelize filter (levels 5-7). Similarly, they see detail by around levels 5-6 with the blur and levels 7-8 with the pixel.

Identifying Busyness, Seriousness and Approachability in a Scene

While there are many ways for people to determine availability, we feel that a person's estimates of an actor's busyness, seriousness and approachability are reasonable indicators. Yet a person's determination of these availability metrics is highly subjective: even given perfect video fidelity, people may make quite different judgments. Consequently, we analyzed *when* subjects were willing to make a judgment of availability without considering whether this judgment was correct.

Figure 8 plots the thresholds that subjects appeared willing to commit themselves to a decision of busyness, seriousness, and approachability. Results are highly variable both between subjects and across scenes. Still, people seem to declare busyness at around level 5 (blur
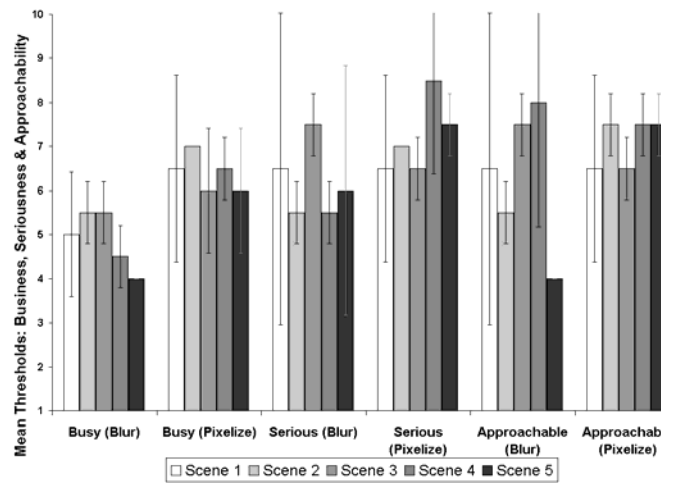
filter), and level 6 (pixelize filter). Seriousness and approachability seems to demand more fidelity—perhaps to make out actors faces and expressions—with people making judgments by about levels 5-6 (blur filter) and levels 6-7 (pixelize filter).

These judgments are fairly stable: once made, subjects rarely changed them even as fidelity increased. As would be expected, confidence in judgments increased with fidelity. For example, while people were only moderately confident of their first attempt to assess busyness, this quickly increased. In particular, more than half the observers were very confident about the accuracy of their assessments by level 5 with the blur filter, but only somewhat confident by level 7 with the pixelize filter.

Rating Privacy

We had asked people to imagine themselves to be the main actor in the scene, and then to rate how well a filter at each filter level would protect their privacy. Ratings went from 1 (unprotected) to 5 (protected). In Figure 9, we plot people's
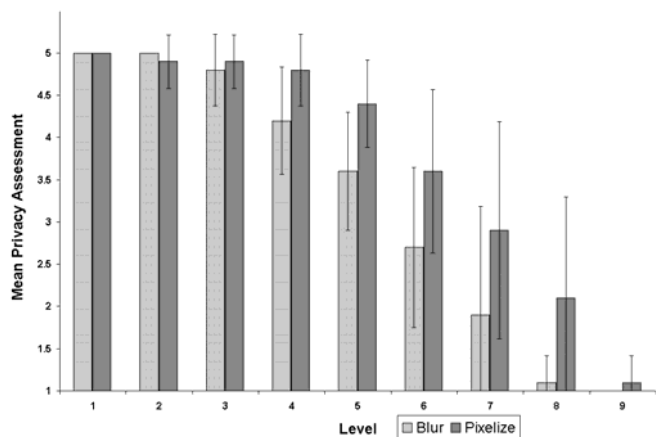

Figure 9. The mean privacy assessment level where subjects felt that privacy would be protected. 1 is unprotected and 5 is highly protected.

privacy assessment by levels, averaged across all scene types. We see that both filtration techniques do conserve the privacy factor at more opaque filter levels. Including the standard deviation, subjects give ratings of 3 or higher (which means privacy is moderately to fully protected) between levels 1-5 with the blur filter, and between levels 1-6 with the pixelize filter.

We also noticed that the scene type somewhat affected subjects' privacy assessments. In general, people were more relaxed about privacy in Scene 1 (the public coffee room) but more concerned about Scene 2 (someone eating a snack) and scene 4 (viewing a public serving area from afar, which usually implies that some people will not be aware of the camera).

Post-Test Questionnaire

Post–test questionnaire items asked subjects various questions regarding the filtering technique they were presented with.

When asked what they liked and disliked about the filtration method they used, subjects generally gave more 'likes' comments to the blur filter as compared to the pixelize filter. They liked the way the blur filter concealed identity, they liked how they could determine movement while still masking details, and they felt it was visually 'smooth'. They also felt the blur filter had potential to regulate privacy. While there were similar positive comments about the pixelize filter, they disliked that it was hard to see who was there and that it was often difficult to tell what kind of scene/environment was captured by the camera.

Another question asked whether people would leave a video link on themselves if it were filtered. Responses differed depending upon which filter the person had used. Six out of 10 replied that they thought a pixelize filter would be sufficient, while nine out of 10 thought that a blur filter would be effective enough to leave the video always on.

DISCUSSION

Given these results, we can reject the two null hypotheses. Of course, this comes as no surprise. We expect that people can clearly identify more awareness cues as fidelity increases across filter levels, just as we would expect protection of privacy to decrease. A more interesting question is to consider if there is any filter type and level that provides some basic level of awareness while still providing a reasonable safeguard to privacy.

We answer this by considering all our results together, as summarized in Table 1.

Privacy is a key issue when judging the filtering techniques used in the present study. While privacy is best protected when filtration effects are strong, this comes at the cost of a person's ability to identify information that could be crucial for accurately determining availability. With both filters, however, there appears to be a filtration level that provides

		Blur	*Pixel*
Number of people		3	5
Posture	movement	2-3	4-5
	seated/standing	3-4	5
Gender		6-7	~7
Objects	basic	3	6
	detailed	6	7-8
Actor activity	basic	3-4	5-6
	detailed	5-7	7-8
Availability	busyness	5	6
	seriousness	5-6	6-7
	approachability	5-6	6-7
Privacy protected to		5	6

Table 1. Thresholds for identifying awareness cues and for determining privacy protection.

safeguards to privacy while still providing basic awareness information.

As indicated in Table 1, this balance of awareness and privacy occurs at level 5 with the blur filter. Subjects chose this level as the highest level that still provides some safeguard to privacy. Yet we saw that subjects could assess basic awareness information quite early on (number of people, posture, basic scene objects, basic actor activity) and finer attributes at the threshold (the basic availability parameters). We also noticed that subjects were generally willing to speculate on key awareness cues as early as level 3, even when this came at the expense of accuracy. At level 5, subjects had more difficulty assessing detailed information that could contribute to privacy violations i.e., they could not describe scene objects in detail, nor could they assess gender, and they were just on the edge of identifying actor activity details.

The pixelize filter provides a more precarious balance between awareness and privacy, this time at level 6. What differs is that the level at which privacy safeguards are acceptable is very close to the level that people are just identifying some basic awareness cues. In particular, the threshold for identifying most awareness information is at levels 5-6, and subjects appeared more reluctant to hazard guesses earlier on. In practice, this probably means that users of the pixelize filter will sometimes find it difficult to identify accurately certain cues i.e., when events in the scene happen further away from the camera; when lighting is poorer, and so on. That is, there is more danger that the pixelize filter at level 6 will operate poorly in some settings because awareness cues are difficult to extract from the scene. On the flip side of the coin, another danger is that it is very close to level 7, which subjects said (in the post-test questionnaire) pretty well displayed almost all there was to reveal in the scene. That is, when it is possible to tell anything from a pixelized scene, it soon becomes possible to tell most everything. We checked this by examining our data closely: indeed, the pixelize filter tended to be

characterized by large jumps across these threshold levels. When taken together, all these points suggest that, although the pixelize filter is widely used in existing prototypical privacy-preserving video media space applications, it may in fact be a poor choice of filter.

The better ratings of the blur filter over the pixelize filter are echoed in the post-study questionnaire. Perhaps the most telling response was that almost all of the blur filter subjects felt that they would use an always-on video link if the blur filter were part of it. While pixelize filter users were also positive, their response was not as overwhelming.

All these results come with large caveats. First, there is high variability in subjects' responses across scenes, across filters, and particularly across the levels tested. We believe that no single level of filtration can guarantee privacy safeguards in all cases for all people. Second, the size of objects within a scene can affect how people view it. Our scenes were mostly medium-range to far shots. If people position their cameras so that certain objects appear large, they will likely be more identifiable. Third, people's willingness to expose themselves to others will depend greatly on where their camera is situated. For example, people will be more sensitive to privacy concerns if they are being viewed in their casual home office (where other members of the family may enter in various states of undress) than in a public work office. Fourth, people's perception of where projecting awareness begins to violate privacy will depend on their relationship with the person on the other side of the video link. While we believe that many people will be receptive to video links with intimate collaborators—close colleagues, good friends, etc.—only a minority of people would be willing to expose themselves to more distant people (e.g., supervisors) and to the world at large (e.g., where the video can be viewed by anyone on the world wide web).

All this implies that, while a balance between privacy and awareness may be possible, it will be a precarious balance at best. Filtering alone may be adequate for some situations, but certainly not all. Our subjects were well aware of this fact, and many underscored the need for 'blocked' modes, where a person could completely cut off the video.

APPLICATION

We are now incorporating and fine-tuning a privacy-preserving filter into several tele-presence support tools. One of the issues we are interested in is how to create lightweight techniques so that people can easily control and mutually negotiate their levels of privacy and awareness.

Nanana (Figure 10) is our experimental two-party video media space application. We plan to embody it into a small device that can be situated anywhere in a person's environment [8,12]. In its current form, Nanana adjusts awareness and privacy needs on the fly by dynamically altering both the level of a pixelize filter (we plan to

replace this with a blur filter) and the frame rate. Nanana is reciprocal: the filtration level and frame rate viewed by one party is always the same as that viewed by the other.
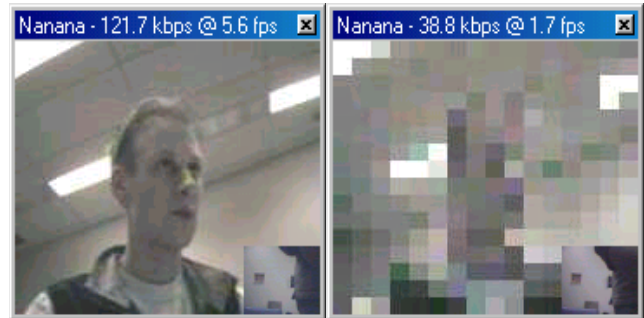
a) Nanana's video is shown unfiltered at a high frame rate (left) and filtered at a low frame rate (right).

b) Blocked video as seen by the blocking (left) and by the blocked (right) parties. Note video in corners can still pass through.

Figure 10. The Nanana media space.

Changes in video clarity are mutually decided between parties by the way both people position their mouse pointer inside or outside the video window. While one party may raise or lower what they can see by one level without another person's permission (so they can get a better sense of awareness or so they can restrict what others see), both parties must cooperate to take it to either extreme. Table 2 describes this behavior.

Pointer positions	Frame rate	Filter level
Both inside	High	None
Inside and outside	High	Moderate
Both outside	Low	High

Table 2. Mapping of pointer positions (relative to each party's video window) to quality of service factors.

When we move Nanana off the desktop and onto the dedicated small device, we will replace this explicit and somewhat unnatural pointer-based interaction with proximity sensors that detect how close a person is to the device. We believe this will provide a socially natural and implicit way to balance what parties can see. In essence, the closer both parties are to the Nanana device (and thus to each others' images) the more Nanana will let them see

into each other's space by continuously but reciprocally adjusting the filtration levels [8,12].

Nanana also allows a person to easily "block" the video. A person just covers the camera with their hand (we use a simple algorithm to detect when the image goes black for a few frames), which toggles the blocking mode. When one person blocks, she sees the image of the back of a hand superimposed over the remote party's video image; the remote party instead sees the image of the palm of a hand (Figure 10b). Both parties may block at any time—indeed even at the same time—but each is responsible for removing the block. We have chosen this approach to closely resemble one's tendency to cover the camera in "dire" circumstances.

CONCLUSIONS

In this study we examined the blur and pixelize filters at various levels for their impact on awareness and privacy using video scenes typical to media space settings. In particular, we looked at: how the accuracy and confidence with which an observer of filtered video can extract awareness cues on the number of actors in a scene, their posture (moving, standing, seated), their gender, the objects visible (basic to detailed), and how available people looked (busyness, seriousness and approachability). We also looked at the privacy-preserving potential of each level in the context of common workplace activities.

Our results show that the filter and the level it is operated at do have an impact on privacy and awareness. It is possible to filter a scene so that some aspects are discernable, but others are not. This is precisely what a privacy-preserving distortion filter would need to accomplish for it to be successful.

In particular, it appears that a balance may be possible using the blur filter at around level 5, and to a lesser extent (because of increased risk) with a pixelize filter around level 6.

While this study provides a good starting point, it is limited. We need to consider filter behavior in extremely privacy-sensitive situations. We also need to understand how people normalize their use of these systems, i.e., where, when and with whom they connect, and the implications of the social factors mentioned in the caveats discussed in the previous section. Of course, we also need to design, build, and deploy a filtered video media spaces into real field settings to see how these filters are used in practice.

We expect that many people will be strongly opinionated on this type of technology [9]; some lay people have described our work as 'dangerous'. Experiences show that fears of transforming this kind of technology into surveillance devices are well founded. However, we feel that we cannot put our heads in the sand. There is little question that computer-based video communication will soon be available to most consumers: video technology is now cheaply available on the desktop, and high-speed internet connections are rapidly becoming affordable. We believe that it is part of our social responsibility to develop technology that builds in safeguards that protect our users.

REFERENCES

1. Bellotti, V. (1996) What you don't know can hurt you: Privacy in collaborative computing. *Proc HCI'96 Conference on People and Computers XI*, 241-261, Springer.

2. Buxton, W. (1997) Living in augmented reality: Ubiquitous media and reactive environments. In / K. Finn, A. Sellen and S. Wilbur (eds) *Video Mediated Communication*, Hillsdale, NJ.

3. Cockburn, A. and Greenberg, S. (1993) Making contact: Getting the group communicating with groupware. In *Proc ACM COOCS'93*, 31-41.

4. Crowley, J. Coutaz, J., and Berard, F. (2000) Things that see. *Commun. ACM*, 43(3), 54-64.

5. Deriche, R. (1990) Fast algorithms for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1), 78-87, January.

6. Dourish, P. and Bly, S. (1992) Portholes: Supporting awareness in a distributed work group. *Proc ACM CHI'92,* 541–547.

7. Gaver, W., Moran, T., MacLean, A., Lövstrand, L., Dourish, P., Carter, K and Buxton, W. (1992). Realizing a video environment: EuroPARC's RAVE system. *Proc ACM CHI'92,* 27-34.

8. Greenberg, S. and Kuzuoka, H. (2000). Using digital but physical surrogates to mediate awareness, communication and privacy in media spaces. *Personal Technologies*, 4(1), January, Elsevier.

9. Harper, R. (1995) Why people do and don't wear active badges. *CSCW Journal*, 4(4), p297-318

10. Hudson, S. and Smith, I. (1996) Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. *Proc ACM CSCW'96,* 248-257.

11. Kraut, R., Egido, C. and Galegher, J. (1988) Patterns of contact and communication in scientific collaboration. *Proc ACM CSCW'88*, 1-12.

12. Kuzuoka, H. and Greenberg, S. (1999) Mediating awareness and communication through digital but physical surrogates. *ACM CHI'99 Video Proceedings and Conference Extended Abstracts*.

13. Lee, A., Girgensohn, A. and Schlueter, K. (1997) NYNEX Portholes: Initial user reactions and redesign implications issues in technology supporting learning. *Proc ACM GROUP'97*, 385-394

14. Tang, J., Isaacs, E. and Rua, M. (1994) Supporting distributed groups with a Montage of lightweight interactions. *Proc ACM CSCW'94,* 23–34.

15. Zhao, Q. and Stasko, J. (1998) Evaluating image filtering based techniques in media space applications. *Proc ACM CSCW'98,* 11–18.