

Evaluation of a Scalable and Distributed Mobile Device Video Recording Approach for Accessible Presentations

Raja S. Kushalnagar*
Information and Computing Studies
Rochester Institute of Technology
Rochester, NY 14623-5604
rskics@rit.edu

Jehan-François Pâris
Department of Computer Science
University of Houston
Houston, TX 77204-3010
paris@cs.uh.edu

Abstract—We evaluate the scalability of a multiple view distributed approach to recording presentations that utilizes camera equipped mobile devices. We show this approach is scalable in terms of video resolution, bandwidth and power. This scalability enables users with a broader range of mobile devices to effectively participate in presentations. We show our approach is less expensive, more scalable, flexible and easier to deploy than traditional presentation recordings. In addition, our solution supports accessible views, which enables participation by deaf and/or blind participants.

Keywords- multiple video recordings, accessible technology, simulation

I. INTRODUCTION

Consumers have embraced the freedom and flexibility of untethered wireless access, which drives the demand for smaller but more capable mobile devices. In the United States, smart phones that typically include video recording and viewing features have already captured 25% of the mobile phone market, and are on track to capture 50% by end of 2011 [1]. Current smart phones often come with dual cameras, one in the front and one in the back for real-time video conferencing and recording. These devices are being released partly in response to the deployment of 4G, the latest generation of mobile networks that has enough bandwidth to support two-way video streaming.

These mobile devices have tremendous potential to record presentations, lectures and talks flexibly, cheaply and inclusively. However, these devices, without further adaption are ill-suited to record presentations such as classroom lectures for the entire duration or with enough resolution. The reason is that smart phones generally are limited by processing power, memory, and battery energy. In addition, they have limited camera resolution, viewing screen resolution and network bandwidth. They also lack processing power to apply powerful compression algorithms that can encode video to fit their available bandwidth rates for distribution. Finally, as bandwidth and screen resolution capacities increase faster than battery capacity, the mobile device recording and viewing bottleneck shifts from bandwidth to power consumption. Therefore new

approaches that require less processing cycles and power consumption, while being equally efficient in video encoding compression and transmission are needed. By splitting the presentation view into multiple smaller views and adjusting the recording rate we can parallelize video capture cheaply and reduce power consumption by using a simpler video compression approach.

Online presentations and lectures are normally recorded by a single video camera manually aimed at the presentation. Although this model has worked well for simple talks, it does not work well with modern presentations that include multiple sources such as whiteboards, presentation slides, demonstrations, and may include teleprompts, text captions for deaf participants, or audio descriptions for blind participants. As shown in Figure 1, the participant's multiple views rarely overlap and are typically small in relation to the entire presentation view. A single view that attempts to capture all views generates a static, boring view that is hard to follow [2], and misses contextual information for comprehension [3]. The traditional solution to the above problems is to have a camera operator zoom in on the auxiliary information source as needed. However, this traditional solution requires use of a camera operator and professional camera equipment, which are time and labor expensive, and are not passive or portable solutions. In addition to the above issues, the camera zooming technique does not fully utilize the fact that these auxiliary information sources have smaller regions of interest, and that some information sources such as slides, change slowly.

We propose and evaluate a Multiple View Perspectives (MVP) [4] recording and encoding approach in terms of processor and viewer scalability. This multiple view approach also works well with P2P streaming approaches to send multiple views to heterogeneous devices in terms of bandwidth or resolution [5]. Although mobile devices have lower bandwidth and resolution as compared with a high end telepresence solution, previous work has shown that participants using personal laptops preferred multiple targeted smaller views of a presentation rather than to a large single high definition view [6]. In addition, we investigate the benefits of adapting frame rate recording to each view depending on its characteristics with respect to mobile recording devices, by measuring three metrics: video file sizes, viewer preferences and processor cycles.

* This work was done while the author was with the Computer Science Department at the University of Houston.

II. RELATED WORK

A. Presentation Recording

1) Capture Flexibility

Studies of presentation video recording and distribution tradeoffs indicate that recording, broadcasting and viewing solutions should be portable, passive, require no pre- or post-processing [7]. The solution should also be able to record synchronized auxiliary presentation sources such as slides or whiteboard that may be widely dispersed within the room [8]. Since it is difficult to keep track of the widely dispersed sources, Cavender et al. [8] investigated the possibility in consolidating multiple video sources onto a single screen, and found that to be effective. Other approaches have tried to automate presentation recordings by harnessing advances in motorized cameras in improving single video recording of meetings by automating zoom techniques [3]. Gluckman and Nayar used an array of board cameras to record a wide angle field [9]. However, none of these set ups were flexible or easy to quickly deploy. We previously tried using camera equipped web camera laptops to record multiple views, and found that it was flexible and easy to set up, although not quick to deploy [6].

2) Power Consumption

Mobile phone battery consumption is a major problem for intensive processing involved in video recording and compression. Insufficient battery life of a mobile device can destroy its usefulness if it cannot record and view for the duration of the presentation, which can go for hours. Cuervo et al. [10] found that the power consumption of a mobile device during decoding and playing a video is very computationally expensive. A simulation that played a video showed that the battery life was drained within 1 hour 20 minutes. Presentation capture involves capture, encoding, transmission, reception and decoding of video simultaneously and in real-time. Without any power optimization, battery life will be exhausted even more quickly.

Cherniavsky et al. [11] proposed a solution to conserving mobile device battery life by encoding videos at a lower frame rate, by encoding fewer frames per second. Encoding and sending fewer frames saves power and also reduces both the total bandwidth consumed by the user and total network load.

B. Presentation Viewing

1) Bandwidth and Video Resolution Scalability

Internet and wireless network bandwidth for a single device application is inherently unpredictable, because it shares its bandwidth with other applications and devices on the network. The video bandwidth synchronization is needed to match it with the varying network bandwidth. Video resolution is also unpredictable, as modern video viewing peers vary widely in download bandwidth and video resolution viewing capacity: connection speed can vary from gigabit to 4G speeds, and video resolution can

vary from high definition resolution on home theater systems to standard definition or smaller on handheld mobile peers. Supporting receivers at a single video streaming rate or resolution is not appropriate, as it can either overwhelm slower receivers, or provide insufficient quality to powerful receivers. An elegant solution to adapting to varying device and network bandwidth demands is to divide the video stream into sub-parts that can be independently viewed, and combined to reconstitute the original. One of the earliest attempts at supporting receiver heterogeneity was ESM [2], which encoded video at multiple bit-rates in parallel and broadcasted them simultaneously; it also prioritized audio over video streams, and prioritized lower quality video over higher quality video. However, this approach incurs a lot of overhead both in terms of bandwidth and stream processing. The sub-streams can be divided in terms of temporal, spatial or resolution properties. A new extension, Scalable Video Coding (SVC) [12] has been incorporated into the H.264 standard in Annex G. This extension scales bandwidth by utilizing a cumulative layered coder, which re-codes the video stream into multiple layers of a video attribute such frame rate, size, or quality. A receiver, depending on its capability, can subscribe to the base layer only with the basic playback quality, or subscribe to additional layers that progressively refine the reconstruction quality. However, this approach consumes a lot of processing power, both to encode and to decode. This approach does not easily parallelize for transmission over unreliable networks such as wireless or peer-to-peer networks, because the base layer must be received before other layers can be added on. Another new approach, called Multiple Descriptive coding (MDC) [13], is less computationally demanding and more parallelizable. This approach utilizes a multiple description coder to generate multiple streams, also referred as descriptions, from the source. Any subset of the descriptions, including each single one, can be used to reconstruct the video. MDC coding can be achieved by splitting video frames or scalar splitting. The descriptions are then distributed over multiple paths to enhance robustness and to accommodate user heterogeneity. MDC consumes less processor cycles than SVC, due to its simpler design. However, both approaches incur bandwidth penalty overhead, as SVC encoding requires extra bits for synchronizing layers, while MDC requires extra bits for each description to carry sufficient information about the original video, and these extra bits can reduce compression efficiency. Moreover, receivers need to use extra processing power to assemble and decode the video stream.

The Multiple Video Perspective approach adopts both the simplicity of MDC and the scalable resolution layered feature of SVC without using extra bits for original image reconstruction. It consumes less processing cycles and power, by reducing the amount of data processed before encoding, while retaining the temporal and resolution compression flexibility of both approaches.

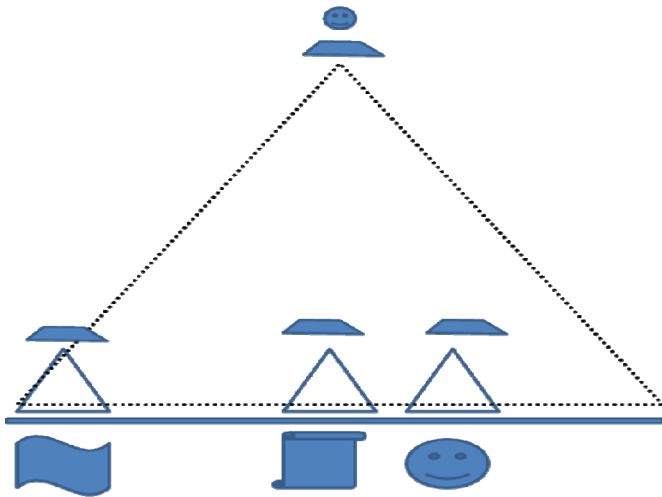


Figure 1. A presentation with presenter, slides and whiteboard information sources. The three solid triangles show camera-equipped mobile device views, while the dashed triangle shows the participant's view and single camera view.

III. ANALYSIS

A. MVP Presentation Recording

1) Viewer Benefits

The popularity of untethered, inexpensive mobile phones or netbooks that can record, download and play back video, has created a need for cheap scalable video feeds to and from such devices. Our MVP approach can fill much of this need. These devices are manually aimed at an optimal viewing angle and resolution at the areas of interest. This approach is portable and quickly deployable by users, which makes it time and labor inexpensive and enables the recording to be both passive and flexible. This enables deployment in presentations without infrastructure support.

MVP can be regarded as a content-aware video encoding and delivery technique that reduces video processing and bandwidth demand by capturing and delivering multiple non-overlapping small targeted video streams. This approach is effective because presentations normally have multiple non-overlapping regions of interest that have different sizes, resolutions and angles as seen from the viewer's perspective. It ignores less-interesting regions for further file space and video processing savings, as shown in Figure 1.

This approach has several advantages. First, each camera view extracts the presentations' region of interest to reduce the participant's field of view, which reduces information loss. Placing cameras closer to the information source improves video resolution and bandwidth [2]. This approach eliminates the need for a camera operator to zoom in and focus on the currently active region of interest. The approach shows all views simultaneously with better angles and higher resolution, as shown in Figure 1.

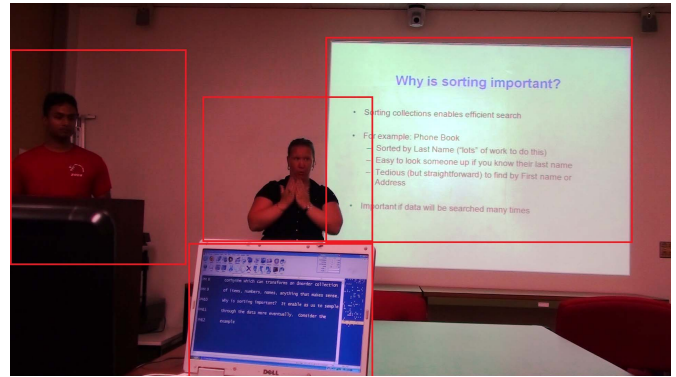


Figure 2. An accessible presentation: instructor, interpreter, captions and slides. These regions of interest are shown in red boxes that have little overlap.

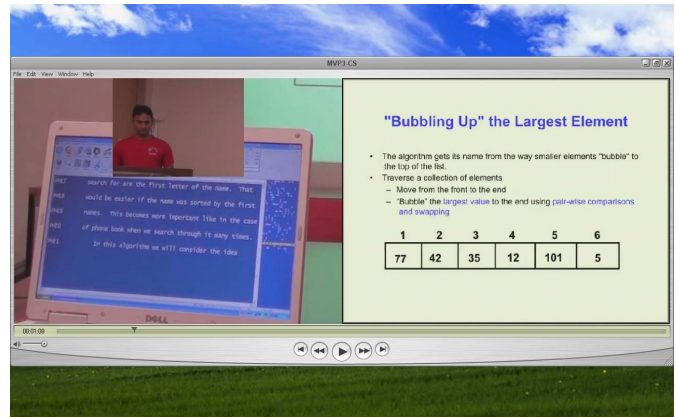


Figure 3. MVP with only the most important information sources presented visually closer together.

Second, participants take control of their learning process by prioritizing their views of the multiple video streams to accommodate their visual learning preferences, rather than depending on the camera operator's preferences. Third, grouping the regions of interest on their computer reduces their visual dispersion demands, especially for deaf participants dependent on visual translations of audio information [3]. Fourth, participants can rewind in real-time a view to review missed information while still keeping up with the presentations in other views. Fifth, view property awareness such as rate of change can significantly optimize video compression. Sixth and finally, additional accessible streams can be obtained and streamed by transforming the instructor's audio stream or overhead information to text streams. This approach enables viewers to add searching and indexing capabilities for effective review [14]. An additional audio descriptions stream of the overhead slides could enhance blind participants' accessibility.

Collectively these optimizations reduce the total amount of recorded and encoded information. These optimizations reduce the field of view and angles for the participants in Figures 2 and 3.

2) View compression and scalability

Video compression discards information that is either redundant or noticeably unimportant to human viewers. Stated another way, compression looks for similar information and collapses them together. Modern video compression approaches, such as H.264 reduce file size and bit rate by first eliminating spatial redundancy in individual picture frames, which yields a compression ratio of up to 25:1. These approaches then eliminate temporal redundancy, that is, the similarities between series of pictures in the video, which yields a compression ratio of up to 200:1.

In presentation recording, in addition to the above redundancies in video capture, we take advantage of the fact that vision is directional, as opposed to sound that is omnidirectional. In presentations, each view may have a different rate of change profile as well. Therefore, in our MVP approach we profile each view and apply the appropriate profile. For example, in a presentation with an instructor and slides, the instructor and slides views each have different characteristics and can be treated separately. This approach avoids the global constraint problem, here the minimum frame rate that depends on the rate of content change. A global view recording has to set the minimum frame rate to a global minimum that can accurately record rapidly changing regions, regardless of the rate of change in other views. On the other hand, local view recordings can set more optimal minimum frame rates for each region of interest depending on its rate of change. For example, to accurately record an instructor, a view with rapidly changing content, the minimum frame rate cannot be lowered from the full recording rate. Conversely, to accurately record slides, white boards or real time transcripts, that are all views with slowly changing content, lower minimum frame rates can be used. This directional view profile compression can yield a compression ratio of up to 30:1 or more, in addition to the usual video spatial and temporal compression.

B. MVP: Presentation Viewing

1) Scalability, flexibility and power consumption

The multiple views, by exploiting vision directionality, enable participants to choose and display only views that are interesting. These videos consume less overall bandwidth, resolution and power, especially when content specific video compression settings are optimized. This provision of multiple smaller video views enables viewers to select appropriate resolution videos to subscribe, and to reduce their aggregate bandwidth. The approach can also minimize processor load by effortlessly discarding unneeded information to reduce both computation and bandwidth without significantly impairing presentation intelligibility. The approach can also be combined with other data partitioning approaches that are orthogonal, such as multiple description codes. This approach is largely orthogonal to

```
for each stream in streams
  if ( client.bw ≥ stream.bw &&
      server.available_bw ≥ stream.bw )
    client.get(stream)
    client.bw -= stream.bw
  end if
end for each
```

Figure 4. The stream prioritization policy.

Multiple Description Coding. In MDC, a single stream is encoded into multiple unique descriptions or streams that can be independently displayed. These descriptions are usually generated by a global video division method, such as separating by time frames, i.e., interleaving, or by spatial frequency distribution. Like MVP, MDC is also tolerant and scalable, for as more descriptions are received, video quality improves. MVP ensures high fidelity, both spatial and temporal, of critical sections of a presentation's view, whereas MDC ensures minimum quality of the overall view. Therefore, combining MVP and MDC increases tolerance and scalability over networks with low bandwidth or resolution, for viewers can choose the number of descriptions and views proportional to their inbound bandwidth or resolution.

Multiple Video Perspectives has the property of being tolerant and scalable, for as more views are received, video coverage improves. This frame rate scaling enables us to provide meaningful quality of service for end-users who are viewing presentations with multiple regions of interest, even if they have limited bandwidth and screen resolution, as it is often the case for smart phones and notebooks.

When peers or receivers are heterogeneous, video bit rate and bandwidth has to adapt to both receiver device bandwidth and resolution. In terms of receiver bandwidth, one or more combinations of multiple video views, multiple description coding or scalable video coding adds degrees of scalability to the video such that each device can receive enough video streams that fit within its bandwidth. In terms of receiver resolution, one or more combinations of multiple video views or scalable video coding adds degrees of scalability to the video such that each device can receive fewer video streams that result in fewer views, and still be able to fit within its bandwidth. The following client scheduling algorithm efficiently utilizes MVP by comparing its own bandwidth against the server's bandwidth, to determine how many views it can receive. The pseudo-code is shown in Figure 4.

The client scheduling algorithm can also be used to optimize the presentation view on resolution-limited mobile devices by enabling participants to prioritize their preferred views. The participant's mobile device can then select and receive the most important views that fit within its resolution capacity, which reduces the participant's perceptual information loss.

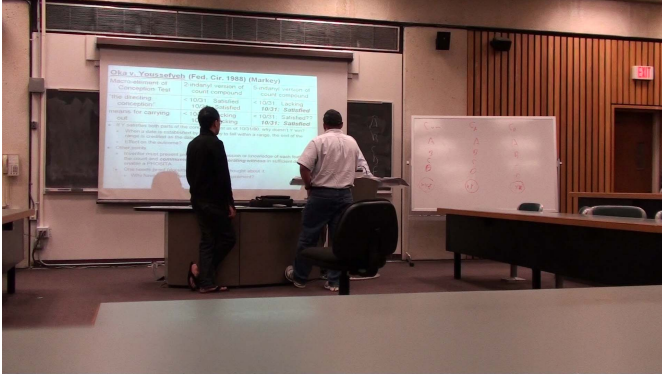


Figure 5. HD perspective of the classroom that is readable on a big screen (large monitor/display).

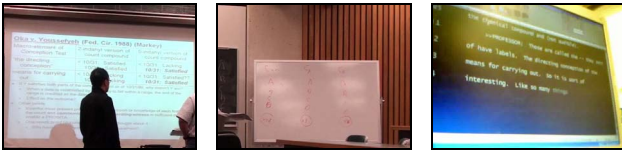


Figure 6. The following regions of interest of the overhead, whiteboard, and real-time transcript respectively.



Figure 7. Personalized MVP Solution Example.

2) Accessibility

As participants cannot always rely on the presenter to provide optimal video streams, our approach enables participants to supplement instructor provided video streams with their own personal video streams. These personal video streams may offer better viewing angles. Disabled participants can also capture video streams of accessible accommodations such as captions that can be shared with other participants. This personalized approach empowers participants in many subtle ways. For example, optimal seating choices are typically limited which can lead to information loss for sensory impaired participants like hard of hearing students who need to be close to the presenter. The MVP approach enables participants to obtain and view simultaneous multiple dispersed perspectives without losing essential information, by placing them side by side on a single screen. This approach increases presentation

universal access in that every participant can benefit from the video streams, including the presenter, other participants and accessibility providers. Figure 5 and Figure 6 illustrate the need for adjusting size, resolution and angles in recording presentation regions of interest.

Another advantage of personal MVP set ups is that they are very flexible and can be easily adapted to a wide range of presentation settings, whether in viewing arrangements or lighting. For example, the personal set up could include a light or infrared camera during low light presentations when lights are dimmed for slide shows [15]. Figure 7 shows a student using a personal MVP system during a presentation. The solution shown in the picture utilizes three web cameras and a camera equipped mobile phone.

For single high definition views, most participants' laptop screens neither have the size nor resolution of a high definition television set. Therefore they cannot view everything with enough resolution. By utilizing the fact that they are not interested in the areas that fall outside the regions of interest, they need not record these areas. The collection of views has equal resolution and quality as the original high definition view, but consumes less screen space, video file size and draws the views closer.

IV. EXPERIMENT DESIGN

A. Presentation Recording

We recorded a lecture on Bubble Sort which is a simple computer science concept. We chose the Bubble Sort concept because it is a non-trivial problem that is not very technical and highly visual. We recorded both a single HD and MVP versions of the lecture. We divided the 9-minute long lecture into four equal segments so that we could show each view in a balanced, repeated measures design without modifying the actual order of the lecture. Participants were also shown a short 30 second introductory video to familiarize themselves with the view. The lectures were made accessible for deaf participants by a professional sign language interpreter and captioner. We recorded the instructor, classroom power-point, captions and sign language interpreter using multiple camera equipped mobile phones (HTC EVO) that recorded these regions of interest to MP4 video at 640x480 pixels at 30 frames per second.

We first generated several versions of recordings of each view that differed only in the recording rate, i.e., frames per second and bit rate for use in the evaluation of context aware compression. The slides perspective video was recompressed with frames per second set to 1, 10 and 15 fps and the bit rate reduced proportionally. This resulted in considerable file savings. For example, re-encoding at 1 fps yielded 95% compression for the slides podcast file size, to 3 MB. We recompressed the real-time captions, instructor and interpreter perspectives with the frames per second set to 1, 10 and 15 fps, which yielded proportional file savings and then combined each of the perspective video files into a complete MVP video file.

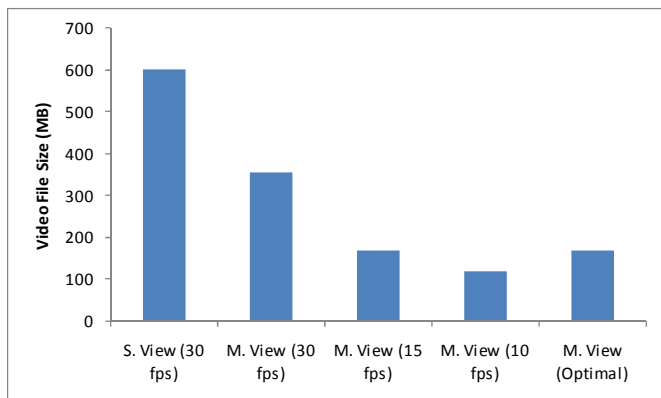


Figure 8. File size (MB) for each presentation at specific video recording rates (fps).

For comparison purposes in terms of file size savings, we recorded the entire lecture simultaneously using a HD camera at a resolution of 1920x1080 at 30 frames per second and the resulting file size was 602 MB. Each recording of the four views was 90 MB, one-sixth that of the recorded single HD view file. We then combined these views into a single MVP video file, which yielded a file size of 360 MB, which is about 60 percent of the single HD view file size. Figure 8 shows how recoding at lower fps reduces file size. However, this results in unsatisfactory perceptual satisfaction at slower rates. Based on participant responses discussed in the next section, we recoded at more perceptually optimal rates of 15 fps each for the instructor and interpreter, 10 fps for captions, and 1 fps for the slides. This recoding resulted in a file size of about 168 MB, which is 75 percent of the single HD view as shown in Figure 8.

B. MVP Viewing Scalability

We evaluated the compression of the file size of video podcast files that were encoded at successive levels of compression using multiple video perspectives (MVP) and multiple description codes (MDC).

MVP recordings were obtained by using four mobile device cameras to record the 9 minute long Bubble Sort lecture. Each recorded view of the presentation was stored as a separate video podcast file.

MDC files were generated by splitting each video podcast file into four equal resolution and bandwidth description perspective podcasts by extracting every 4th frame and re-encoding at a fourth of the original bit rate. Four descriptions were used to encode each view. The reason for picking four descriptions was that four descriptions offered better scalability than encoding with two or three descriptions. Five or more descriptions resulted in diminishing scalability returns. A server can stream simultaneously all video files of a description of a view. A viewing client can fetch and display the video as long as it receives at least one description of one view.

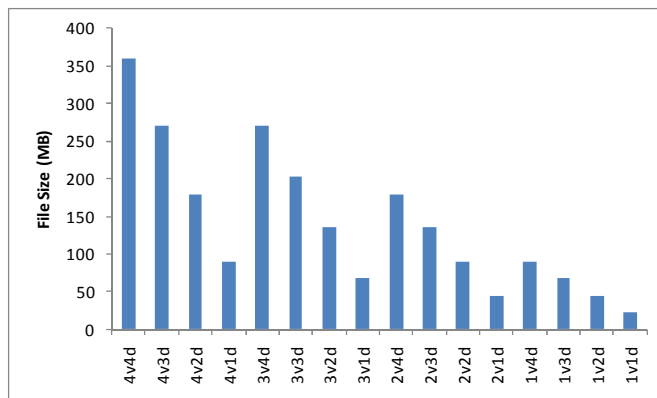


Figure 9. File size (MB) of the multiple views as views and descriptions are gradually reduced. X-axis labels indicate the number of views and the number of descriptions per view.

Four views and four descriptions offered four levels of scalability each and 16 levels of scalability combined. Figure 9 charts the reduction in file size of a multiple view podcast that totaled 360 MB. It shows the reduction in file size as the number of descriptions for each view is reduced first and then the number of views is reduced next. Although this nested loop does not capture all possibilities, it does show how file size can be reduced to as low as 22.5 MB for a single description of a slide view. For the slides view, participants were able to follow the view according to the study evaluation shown below.

C. MVP Viewing Evaluation

To evaluate the efficacy of context-based view compression, we recorded four views of a presentation on Bubble Sort, focused on the instructor, slides, captions and sign language interpreter. The captions and interpreter views are typical in an accessible presentation where a captioner transcribes the audio for deaf participants and indexing and sign language interpreters provide access for deaf participants. We then created four compression profiles for each view by re-encoding the original 30 frames per second recording at 15, 10 and 1 frame per second respectively. We compared participant responses to three questions after they watched each of the compression profile views.

We recruited 18 participants ages 20-45 (11 female) for the study. Each participant downloaded and watched a video clip consisting of four balanced views of the presenter, power point, captions and sign language interpreter views. Each view was randomly set to 30, 15, 10 or 1 frame per second. After viewing each video clip view, each of the participants was asked to respond to three preference questions using a Likert scale from 1 to 5. Each participant was asked to rate the views they watched on the basis of the following questions: 1) “Is the video easy to follow?”, 2) “Is the video quality sharp?”, and 3) “Is the video motion smooth?”.

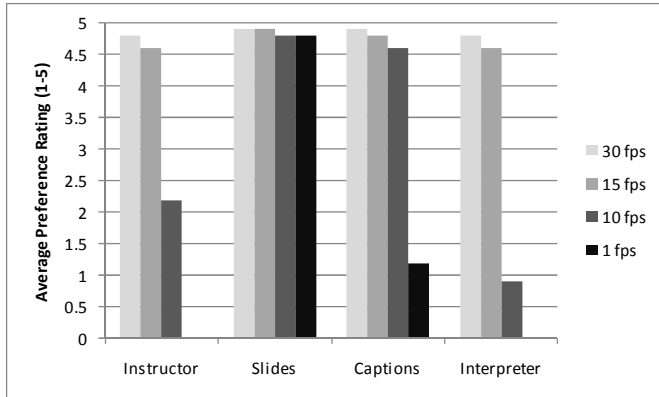


Figure 10. Average preference responses (on a scale of 1-5) for the question “Easy to follow”.

Respondents gave similar ratings for all three questions. Figure 10 charts the participant responses for the question “Is the clip easy to follow?” for each view at a specific fps. For the slides perspective, there was no significant preference difference between viewing at 1 fps or higher. For the captions perspective, there was no significant preference difference between viewing at 10 fps or above. For both the interpreter and instructor perspectives, there was no significant preference difference between viewing at 15 fps or above.

Therefore the optimal multiple view combines the four compressed podcasts at the best compression level where there was no drop in preference. The preferences indicated that the perspectives to use were the podcast slides at 1 fps, captions at 10 fps, instructor and interpreter views at 15 fps), into a single file for viewing as a multiple video perspective file. The optimal view file size of all four recorded multiple view perspectives was 168 MB, which is about 75 percent less than the recorded single view perspective file size.

D. Processing power and Power Scalability

We investigated the effect of reducing frame rate on processing cycles and indirectly on power consumption, for encoding and decoding simultaneously each video view.

We ran two simulation instances of a single view presentation recording and viewing on an official Android 2.2 emulator on a Windows 7 system. The first instance encoded an uncompressed AVI file at a bit rate of 100 Mbps for each presentation view to a MP4 file at 1.5 Mbps, using H.264/AVC CBP. This simulates the capture and conversion of video by a mobile device into video in compressed format suitable for streaming. The second instance decoded this MP4 file, which simulates the receipt and decoding of the compressed video streaming for end user viewing. Figure 11 shows that decreasing the frame rate of the video file leads to a reduction in the average number of processor cycles needed for encoding and decoding.

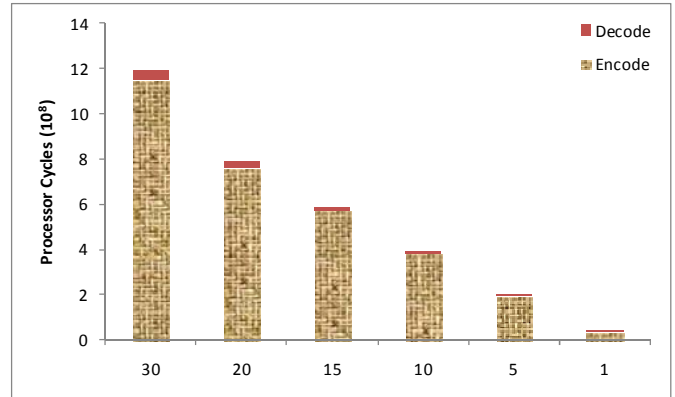


Figure 11. Average processor cycles per second (10^8) for simultaneous encoding/decoding of video of the captions at 30, 20, 15, 10, 5 and 1 frames per second.

Power consumption is reduced as well, when processor cycles are reduced. We repeated the simulation for presentation recording and viewing using multiple views on the same system. Figure 12 shows a similar and even more dramatic reduction in the average number of processor cycles and power consumption needed for encoding and decoding.

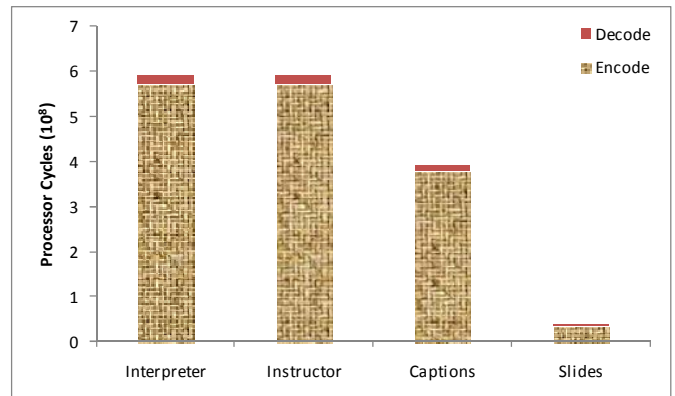


Figure 12. Average processor cycles per second (10^8) for simultaneous encoding/decoding for a customized recording of each view (interpreter at 15 frames per second, instructor at 15 frames per second, captions at 10 frames per second, and slides at 1 frame per second).

E. Transmission Scalability

Bandwidth savings are first obtained by substituting multiple standard definition perspective streams in place the single high definition single stream. In our simulation, the single H.264 HD stream bit rate was 9 Mbps, while all four H.264 stream views bit rate totaled 5.2 Mbps, a ratio slightly less than 2:1. This approach is effective because modern presentations have many non-interesting regions between regions of interest, each of which may have different sizes, resolutions and angles. When contextual

compression is applied, the total bit rate improves to about 4:1. We can apply MDC to further reduce the client minimum bit rate requirement by 4:1, although this can significantly impact perceptual quality for some views.

Our approach enables a presentation to be streamed effectively to a variety of end-consumer peers with varying bandwidth and resolutions – home theater systems with wired LAN connections, laptops with wireless LAN connections, and smart phones with mobile data connections. This approach also reduces total information recorded and transmitted as compared with traditional presentation multimedia recording approaches. It enables use of modern smartphones to lower the cost of capturing, distributing and viewing presentations. It also enables participants to reduce visual dispersion and prioritize views and sizes of perspectives on low resolution devices.

V. CONCLUSIONS

This paper proposes a Multiple Video Perspective streaming approach for capturing and streaming online presentations that is less expensive and is more inclusive, scalable, flexible and easier to deploy than traditional video streaming, especially for mobile peers. We analyzed issues in extending presentation video capture and streaming to a range of consumer peers and proposed a solution of utilizing multiple view perspectives that use context aware compression.

The MVP approach yields bandwidth and video resolution scalability of up to 2:1. Incorporation of context sensitive frame rate recording yields a further bandwidth savings of up to 4:1 on average. Our user study results showed that we could obtain savings in power and bits by lowering the frame rate of certain views, such as captions or slides without significantly affecting intelligibility. Reducing the recording rate saves bandwidth and power in power limited mobile device recording and transmission. This scaling for video bit rate and resolution enables meaningful quality of service for viewing presentations with multiple regions of interest to clients with limited screen resolution and bandwidth, especially for smart phones and notebook peers. The results show that combining MVP and MDC along with a client stream prioritization policy at the video recording and encoding stages enable support of a wider range of heterogeneous bandwidth peers. This stream prioritization policy can either utilize default preferences obtained from user studies or user specified preferences in selecting the multiple sub-streams generated by the MDC or MVP for user viewing. This approach allows individual viewers to fit the streams received within their available bandwidth, thus maintaining a reasonable quality of service and an effective learning experience.

VI. REFERENCES

- [1] NielsenWire, "Android Soars, but iPhone Still Most Desired as Smartphones Grab 25% of U.S. Mobile Market," *NielsenWire*, 2010.
- [2] Q. Liu, Y. Rui, A. Gupta, and J.J. Cadiz, "Automating camera management for lecture room environments," *Proceedings*

of the SIGCHI conference on Human factors in computing systems - CHI '01, 2001, pp. 442-449.

- [3] A. Ranjan, J. Birnholtz, and R. Balakrishnan, "Improving meeting capture by applying television production principles with audio and motion detection," *Proceedings of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, 2008, pp. 227-236.
- [4] R.S. Kushalnagar and J. Pâris, "Multiple View Scalability of Presentations Distributed to Heterogeneous Devices," *Proceedings of the 2010 International Symposium on Performance Evaluation of Computer and Telecommunication Systems*, Ottawa, Canada: 2010, pp. 7-13.
- [5] R.S. Kushalnagar, "A Comparison of P2P Approaches in Handling Multiple Video Perspective Streams by Mobile Devices to Support a More Inclusive Learning Environment," *8th Annual IEEE International Conference on Pervasive Computing and Communications (PERCOM 2010)*, Mannheim, Germany: IEEE Computer Society, 2010, pp. 855-856.
- [6] R.S. Kushalnagar, A.C. Cavender, and J. Pâris, "Multiple View Perspectives : Improving Inclusiveness and Video Compression in Mainstream Classroom Recordings," *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, Orlando, Florida: 2010.
- [7] C. Zhang, J. Crawford, Y. Rui, and L. He, "An automated end-to-end lecture capturing and broadcasting system," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 4, 2008, p. 6.
- [8] A. Cavender, J. Bigham, and R. Ladner, "ClassInFocus: enabling improved visual attention strategies for deaf and hard of hearing students," *Proceedings of the eleventh international ACM SIGACCESS conference on Computers and accessibility*, ACM, 2009, p. 67-74.
- [9] J. Gluckman and S.K. Nayar, "Catadioptric Stereo Using Planar Mirrors," *International Journal of Computer Vision*, vol. 44, 2001, pp. 65-79.
- [10] E. Cuervo, A. Balasubramanian, D. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl, "Maui: Making smartphones last longer with code offload," *In Proceedings of the 2010 ACM MobiSys*, 2010.
- [11] N. Cherniavsky, A. Cavender, R. Ladner, and E. Riskin, "Variable frame rate for low power mobile sign language communication," *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, ACM, 2007, p. 170-176.
- [12] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, 2007, pp. 1103-1120.
- [13] V. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, 2001, pp. 74-93.
- [14] J. Subhlok, O. Johnson, V. Subramaniam, R. Vilalta, and C. Yun, "Tablet PC video based hybrid coursework in computer science: report from a pilot project," *Proceedings of the 38th SIGCSE technical symposium on Computer science education*, ACM, 2007, pp. 74-78.
- [15] R.S. Kushalnagar and J. Pâris, "Empowering Deaf Students with Multiple Views," *Proceedings of the 2010 Technology and Deaf Education: Exploring Instructional and Access Technologies*, Rochester, NY: 2010.