



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Engineering and Information Sciences -
Papers: Part A

Faculty of Engineering and Information Sciences

2014

Large-scale immersive video conferencing by altering video quality and distribution based on the virtual context

Farzad Safaei

University of Wollongong, farzad@uow.edu.au

Pedram Pourashraf

University of Wollongong, pedram@uow.edu.au

Daniel Franklin

University of Technology, Sydney, daniel@uow.edu.au

Publication Details

F. Safaei, P. Pourashraf & D. Franklin, "Large-scale immersive video conferencing by altering video quality and distribution based on the virtual context," *IEEE Communications Magazine*, vol. 52, (8) pp. 66-72, 2014.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:
research-pubs@uow.edu.au

Large-scale immersive video conferencing by altering video quality and distribution based on the virtual context

Abstract

Current video conferencing applications do not scale to support a large number of participants. This article describes an IVC system that combines the best attributes of video conferencing and multi-user virtual environments. It is shown that each participant of IVC has a virtual context that is defined by his/her perspective and perception about the quality and relevance of video and audio of others. The virtual context determines both the visibility status and the required quality of videos of participants. This information can be used to dynamically alter the multicast trees that are formed among clients for the purpose of multimedia dissemination so that only the relevant videos are transmitted to end users. In addition, it is possible to reduce the video quality of a given user in response to the virtual context without the degradation having any perceptual impact. The combination of these factors reduces the required upload and download bandwidth of clients by more than 90 percent on average, making IVC highly scalable to support very large gatherings.

Disciplines

Engineering | Science and Technology Studies

Publication Details

F. Safaei, P. Pourashraf & D. Franklin, "Large-scale immersive video conferencing by altering video quality and distribution based on the virtual context," *IEEE Communications Magazine*, vol. 52, (8) pp. 66-72, 2014.

LARGE-SCALE IMMERSIVE VIDEO CONFERENCING BY ALTERING VIDEO QUALITY AND DISTRIBUTION BASED ON THE VIRTUAL CONTEXT

Farzad Safaei^{*}, Pedram Pourashraf^{*}, Daniel Franklin[%]

^{*}ICT Research Institute, University of Wollongong, Australia,

[%]University of Technology, Sydney, Australia

ABSTRACT

Current video conferencing applications do not scale to support a large number of participants. This paper describes an *immersive video conferencing* (IVC) system that combines the best attributes of video conferencing and multi-user virtual environments. It is shown that each participant of IVC has a *virtual context* that is defined by his/her perspective and perception about the quality and relevance of video and audio of others. The virtual context determines both the *visibility* status and the required *quality* of videos of participants. This information can be used to dynamically alter the multicast trees that are formed among the clients for the purpose of multimedia dissemination so that only the relevant videos are transmitted to end-users. In addition, it is possible to reduce the video quality of a given user in response to the virtual context without the degradation having any perceptual impact. The combination of these factors reduces the required upload and download bandwidth of clients in excess of 90% on average, making IVC highly scalable to support very large gatherings.

1. INTRODUCTION

Increasingly, users regard the Internet as a *meeting place*, where they can form communities and interact with groups of people as part of their work, play, education, or social interaction with family and friends. This phenomenon is likely to create a significant demand for multiperson-to-multiperson video communications. However, the conventional video conferencing systems cannot scale to support a large number of participants. This lack of scalability is partly *technical* and partly *cognitive*. The technical barrier stems from the fact that the number of required video streams transmitted over the network grows as the *square* of number of participants. Consequently, most current video conferencing solutions impose a rather modest upper limit on the number of end points supported. On the other hand, the application utility may not improve by increasing the number of participants, even if the bandwidth problem could be solved. The common practice of displaying videos of participants as rectangular tiles on the screen, the so-called Brady-Bunch model, is very restrictive and cannot scale. By showing the ‘relevant’ participant only, such as the one with audio activity, the scalability is improved but often resulting in strained social protocols and cognitive fatigue for the users.

The cognitive problem can be overcome by combining the concepts of video conferencing and distributed virtual environments, where the video of participants are shown on the front surface of their respective ‘avatars’. We refer to this as an Immersive Video Conferencing (IVC) system, where, in essence, the real life characteristics of a human gathering are being emulated (Figure 1). Similarly to other virtual environments, each participant is represented by an avatar and can roam

freely in a 3D space. However, unlike graphical avatars, the participants' avatars in an IVC display their real-time video and their voice travels in the virtual environment in accordance to the expected properties of the propagation of sound. The combination of rich visual and aural scenes creates a sense of immersion and provides a comfortable space for users to interact and communicate with each other. In particular, the natural human behavior of 'mingling' in a crowd becomes possible, where *multiple simultaneous conversations* can take place, participants can have peripheral awareness of surrounding conversations and dynamically move from one to another. Indeed, certain functions, such as locating other people, exchanging business cards, creating display boards on the walls and projecting content onto these boards, become easier than a physical gathering.



Figure 1 - Immersive video conferencing

Fortunately, the IVC model of video conferencing can also help with the technical aspects of scalability. This is because each participant in an IVC has a *virtual context* in relationship to others. Here, we define the 'virtual context' as those aspects of being in a multi-user virtual environment that *affect one's perspective and perception about the quality and relevance of video and audio of others*. Let us focus on the video distribution, which places more stress on the underlying network. The perspective of each user within the virtual environment determines which videos are relevant to this participant. Consequently, the irrelevant videos need not be transmitted to this user. We refer to this as *visibility culling* in response to the virtual context. Moreover, the acceptable quality level of each video stream, in terms of its resolution and frame rate, also depends on virtual context. Hence, it is possible to judiciously reduce the video quality and save bandwidth with no perceptual impact.

This is referred to as *quality culling* in this paper.

To provide a concrete example, consider the scenario shown in Figure 1. From the perspective of this user (we are looking at the scene from the perspective of the local client whose self-video is shown at the top right corner), only a subset of avatars are visible at this moment. This mimics the character of crowd interactions in the real world and is an acceptable constraint for the users. Among the visible avatars, however, the virtual distance of avatar C is larger than avatar A. Because avatar C's video is rendered on a smaller surface area, the required resolution is lower, as determined by perspective geometry. However, our user studies have confirmed that it is possible to reduce *both* the resolution and frame rate of C's video for this client *above and beyond* what is determined by perspective geometry without the effect being noticeable.

The virtual context, therefore, can reduce the network load of the IVC dramatically. However, to achieve this goal, two mechanisms must be developed. Firstly, the encoded video stream should be designed in such a way so that quality of video can be adjusted in response to virtual context of each recipient, preferably mid-stream during the multicast phase and without going through the full decoding and encoding process. Secondly, the virtual context of each user typically changes very rapidly; hence, the multicast trees that are formed among the clients for the purpose of multimedia distribution should be able to adapt to these changes efficiently.

Our research team has developed an IVC – called iSee – which is taking advantage of these methods and can support a large number of participants in a shared virtual environment [1].

2. VIRTUAL CONTEXT AND VISIBILITY CULLING

The first technique to improve the scalability of IVC is to evaluate the virtual context of each client and determine which avatars are *visible*. Earlier research on multi-user virtual environments focused on a technique, known as *Area of Interest (AOI) management*, to reduce the amount of state information sent to each client [2]. Typically, the AOI of a client is viewed to be within a certain virtual distance of the current position. In addition, the computer graphics community has developed a number of techniques to improve rendering performance by determining occlusions and visibility [3]. Inspired by both of these efforts, we have shown that there are a number of attributes associated with the virtual context that affect video visibility [4]. These include:

- *Distance-based (DB) culling*: only those video streams that are within a given virtual distance of the viewer are sent to the client.
- *DB + view frustum culling (VFC)*: Only those video streams within a given virtual distance and also within the current view frustum are sent to the client.
- *DB + VFC + back face culling (BFC)*: In addition to DB and VFC, if the avatar is facing

away from the client, the video is not transmitted to that client.

- *DB + VFC + BFC + occlusion culling*: If the avatar is occluded by other avatars (from the perspective of the client), the video will not be sent.

Figure 2 shows the reduction in network capacity usage as a function of density of participants as we progressively incorporate more refined attributes of virtual context in visibility culling. As can be seen in this Figure, the reduction in download capacity of each client is very significant, by as much as 90% on average, when employing the above visibility culling technique [4].

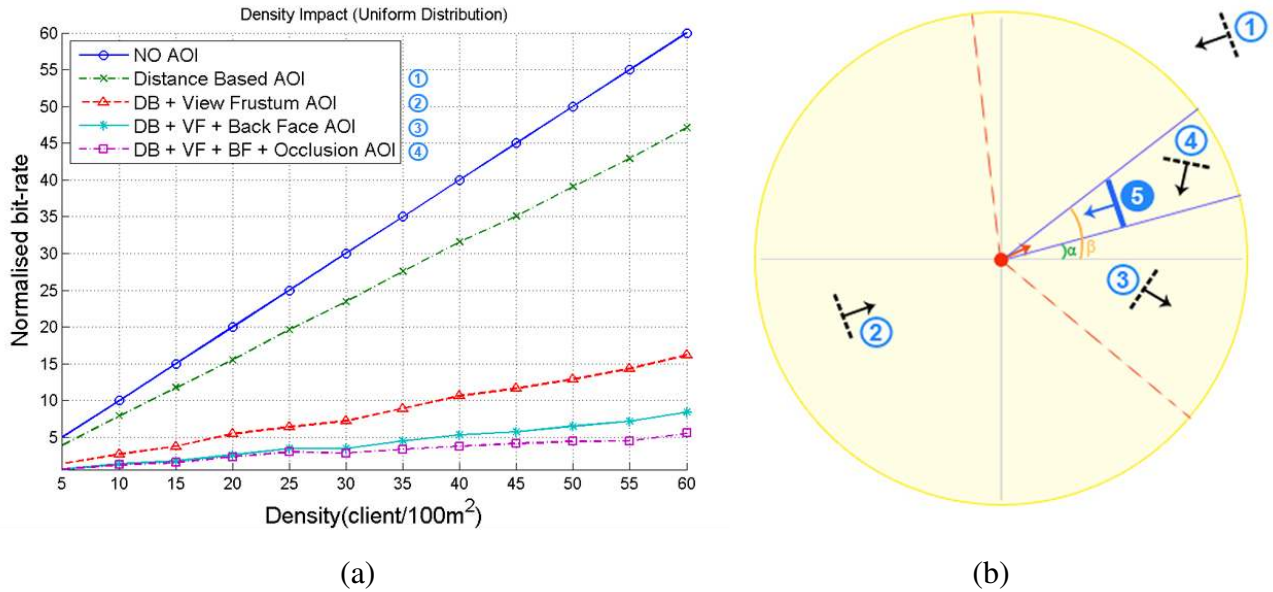


Figure 2: (a) Reduction in the mean number of downloaded video streams by each client due to visibility culling; (b) Example of visibility culling attributes: (1) outside the visible distance, (2) outside the view frustum, (3) back face culling, (4) occluded avatar. The local avatar is shown as a red dot in the centre.

While effective, the visibility culling technique introduces an undesirable side effect. The system becomes very sensitive to movements of avatars, because both translational and rotational movements of participants will change the virtual context and consequently the distribution pattern of video streams. We have introduced a prediction method, where the composition of the future view frustum is estimated based on the current motion vectors. Using a control feedback loop, the video stream distribution pattern is adjusted *ahead of time* to prepare the clients for the change of view. In a real interactive scenario, the additional bandwidth usage due to this prediction is quite modest.

3. VIRTUAL CONTEXT AND QUALITY CULLING

In addition to visibility, which is a ‘yes’ or ‘no’ decision, the virtual context also affects the required *quality* of the visible videos. Avatars within one’s view frustum are typically at different (virtual) distances and orientations with respect to the viewer (see Figure 1). As the virtual distance

of an avatar increases, it occupies a smaller area of the screen, which means the required spatial resolution of video is less than a nearby avatar. In addition, if the orientation of an avatar with respect to the viewer is not frontal (for example, Avatar B in Figure 1), the projection of video is distorted, which again presents an opportunity to reduce bit rate.

A number of studies have investigated the perceptual impact of spatial and temporal resolutions of video [5]. However, to our knowledge, none of these address the sensitivity of human perception to variations of *virtual* context. Therefore, we conducted a user study involving around 230 participants [6], which has confirmed that people are less sensitive to video quality degradation when the virtual distance increases. This perceptual tolerance is *in addition* to the perspective-induced reduction of the video size. Figure 3 demonstrates the outcome of our user study on the required resolution versus the virtual distance. This Figure confirms that it is possible to degrade the quality of video selectively in a way that is imperceptible to the participants.

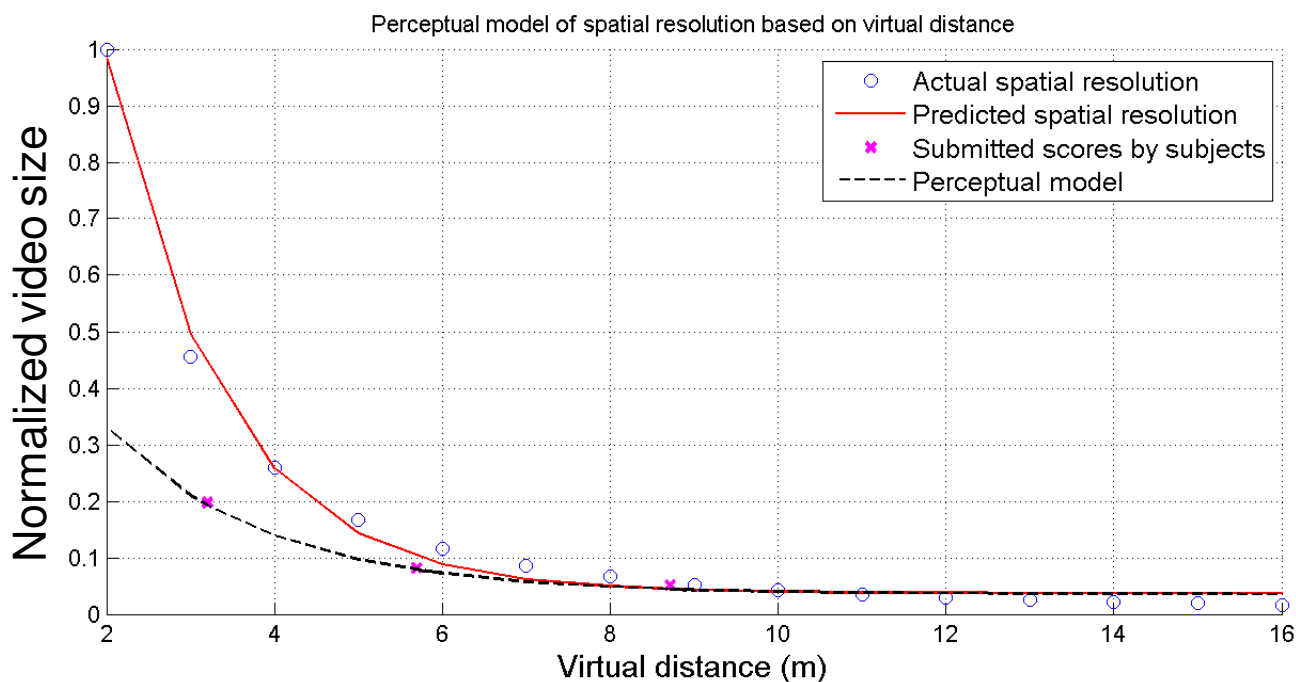


Figure 3: Reduction in the size of video as a function of virtual distance. The solid line represents perceptually ‘unnoticeable’ change and the dashed line is ‘slightly noticeable’.

In the same scenario shown in Figure 2, the additional bit rate saving due to quality culling would be between 3 – 7% [6]. However, the key advantage of this technique is for those situations where visibility culling is not effective. For example, in a virtual lecture theatre, many avatars would be visible to participants. In this scenario, employing quality culling is the main avenue of reducing bit rate, resulting in up to 70% reduction [6]. An example of adjusting quality in response to the virtual context can be seen in [1].

4. ADJUSTING VIDEO DISTRIBUTION AND QUALITY

In some sense, evaluation of the virtual context is easier and more practical than the physical context. There is no need for deployment of sensors and devices to track objects and people [7], since the information about locations, orientations, movements, occlusions and the map of the environment is already available to the client as part of the state information exchange necessary for maintaining the distributed virtual environment. The difficulty in this case is the *rapidity* by which the virtual context continually changes. The main challenge, therefore, is not so much to compute the virtual context, but to develop suitable mechanisms that can adjust video quality and distribution patterns in response to the *dynamics* of this context.

4.1. Video quality adjustment technique

For the purpose of this paper, the video of each participant is a 2D video shown on a flat surface of the avatar. However, the avatar is free to move and rotate within the 3D space; hence, the video will be rendered at different 3D orientations and distances relative to the viewer.

It is desirable to develop suitable techniques so that individual video bit streams can be *pruned* before transmission to the receiving client based on the required quality. This pruning process can take place either at the source in a unicast video distribution model or at the branching points of a multicast tree (see Figures 5 and 6). Note, however, that each participant will have a unique perspective, and a particular video stream may be required at different quality levels by a number of participants. Hence, unlike a point-to-point video telephony scenario, it is not possible for the source to simply adjust its video coding parameters based on the receiver's device.

Currently, there are two key mechanisms for video quality differentiation: Hierarchical Video Coding or Multiple Description Coding. But unfortunately, these techniques focus on the *overall* video quality, equally in every part of the video. In IVC, the quality differentiation may not be uniformly distributed over the spatial extent of a video. For example, the video may be partially occluded or the avatar may be rotated and some parts of the video requiring lower quality due to perspective variations.

Our proposed 'perceptual pruning' process, in contrast, enables controlled degradation of video quality at the scale of a *video coding block*. Modern video codecs, like H.264/AVC, partition the video into a number of macro- or micro-blocks, such as 16x16, 16x8, 8x8, 4x4 pixel blocks. In IVC, as a result of variations in the virtual context, the video surface may be distorted, where some blocks in the frame are squeezed (texture minification) while others are stretched (texture magnification).

In perceptual pruning, the spatial resolution of each block (or a group of blocks with the same distortion pattern) is adjusted according to the projected size of the block on the screen by setting a

suitable number of high frequency coefficients in the Discrete Cosine Transform (DCT) domain to zero. This is referred to as DCT down-sampling. In other words, the projected size of each block is calculated according to the virtual context of the receiver and the vertical and horizontal frequencies outside of the projected boundaries are zeroed out. For example, if an 8x8 block is projected onto a 6x2 block, the DCT coefficients in the remaining region could be replaced by an ‘all-zero’ symbol. This scheme allows a fine degree of control over the quality. In practice, however, a few ‘levels’ of degradation is sufficient, as there will be diminishing returns in using finer granularity. In particular, with respect to virtual distance, four levels are identified: *Near*, *Medium*, *Far*, and *Occluded* (the latter means that the whole block is occluded by another object and can be set to all-zero). In addition, three levels are identified with respect to angular orientation: *Frontal*, *Medium-Rotation*, and *Large-Rotation*. The source encoder will arrange the DCT coefficients and identify the appropriate boundaries for these levels, which can then be replaced by an ‘all-zero’ symbol as the video stream travels down the multicast tree.

Figure 4 shows an experimental verification of DCT down-sampling where the degraded images are compared to the references. Despite the fact that DCT down-sampling results in significant distortion *if the video was shown normally*, due to 3D rotation, the reference and degraded videos are perceptually similar as verified by both the structural similarity index (SSIM around 0.99) and Peak to Signal Noise Ratio (PSNR around 46 – 49 dB).



Figure 4: Reference (Ref) and Degraded images (Deg) after Perceptual Pruning for various virtual orientations

The perceptual pruning can take place in the source (in case of unicast), a server, or an intermediate node during multicast and should be done for each recipient independently. However, if two or more recipients have similar perspectives of a given video, then the same stream may be used for those. The predicted frames of a video (P frames¹) must have the same pruning process as their corresponding reference frames (e.g. the I frame). So any changes in perspective that requires modification of the pruning mask will be applied at the next I frame.

4.2. Video dissemination multicast trees

The combination of visibility and quality culling techniques provide a powerful basis to make IVC scalable. However, these techniques require the formation of a number of *point-to-multipoint* video distribution trees, which are highly dynamic and are undergoing continuous change both in terms of their composition (identity of participating nodes) and the required video rate of each leaf. To our knowledge, the degree of agility demanded of the underlying network infrastructure in responding to changes in the virtual context far exceeds the capabilities of existing mechanisms that adapt to the physical context.

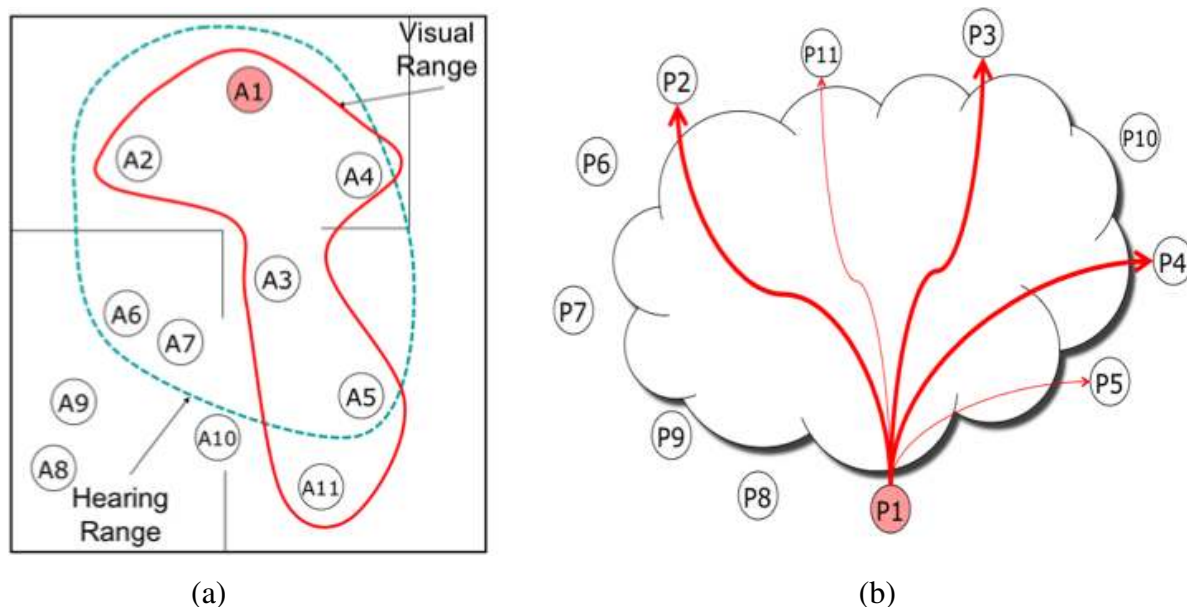


Figure 5: (a) The location of A1, the virtual representation of P1, with respect to the virtual crowd. (b) The unicast dissemination of video originated at peer P1 is shown. A thinner arrow signifies the delivery of content at lower resolution or frame rate.

To illustrate this point, Figure 5-(a) shows a small section of a virtual crowd. Consider avatar A1 corresponding to peer P1. The visual and hearing ranges of A1 are shown in this Figure as areas enclosed by solid and dashed curves respectively. Consequently, the voice and video of P1 must

¹ For real-time video, the B frames are typically not used due to latency constraints.

somehow reach all the peers residing in these ranges for inclusion in their aural and visual scenes. This corresponds to P2–P7 for voice and P1 to P5 + P11 for video. In addition, A3, A5 and A11 could tolerate a lower resolution and frame rate for A1’s video because of the larger virtual separation. These quality variations, however, depend on the current arrangements in the virtual crowd and could change with the passage of time.

In the absence of native multicast support at the IP layer, a unicast model for dissemination of content can be considered as shown in Figure 5-(b). To avoid cluttering this Figure, we have only shown the video distribution of P1. Each peer, however, has to unicast both voice and video to the subset of peers in its hearing or visual ranges. For the whole virtual crowd, therefore, we would require mDN unicast flows on average, where m is the number of media types (e.g. voice and video), D represents the average number of participants within the communication range of each peer, and N is the total number of participants.

Focusing on P1 and assuming that r is the bit rate of video, the *upload* capacity required to distribute this peer’s video will be rD on average. As discussed before, it is possible to reduce this by the quality culling process, where the unicast flow to peers who are further away in the virtual world would contain smaller-sized video streams. Assume for clarity that the bit rate of video can be decreased by a factor α for each level of virtual distance increase (from *Near* to *Medium* or from *Medium* to *Far*). Then video streams to P5 and P11 require a lower bit rate (either αr or $\alpha^2 r$ depending on their virtual distance to P1), shown as thinner arrows in the Figure. On the receiving side, each peer receives a number of unicast flows from everyone in its hearing and visual ranges. On average, this *download* capacity will be the same as the upload capacity in the case of unicast distribution.

If native multicast support is available at the IP layer, then P1 only needs to send one video flow at the maximum quality as a multicast flow, which would reduce the upload capacity requirements significantly. The download capacity, however, would remain the same, and indeed if the multicast system did not provide quality culling, it could worsen.

Native inter-domain multicast may not be available. In this case, a peer-to-peer *overlay* multicast (where the multicast trees are created between the peers or end points) could be used to reduce the upload capacity requirements of a client. Figure 6-(a) shows the case of creating the shortest path multicast tree using the *physical separation between the peers as the cost metric*. Most existing peer-to-peer algorithms for multicast adopt this approach [8]. The peers may use a number of algorithms to approximate the physical proximity, such as direct probing of the communication delay or reference measurements with respect to well-known landmarks. For best results, such information should be made available to every peer and kept consistent [9]. This conventional method to create

multicast trees, however, is not suitable for quality culling. For example, the virtual proximity zones of *Near*, *Medium* and *Far* are shown in this Figure. Although P5 is in the ‘*Far*’ zone with respect to P1, it must download at the full rate because it is on the path to other peers that require higher quality content.

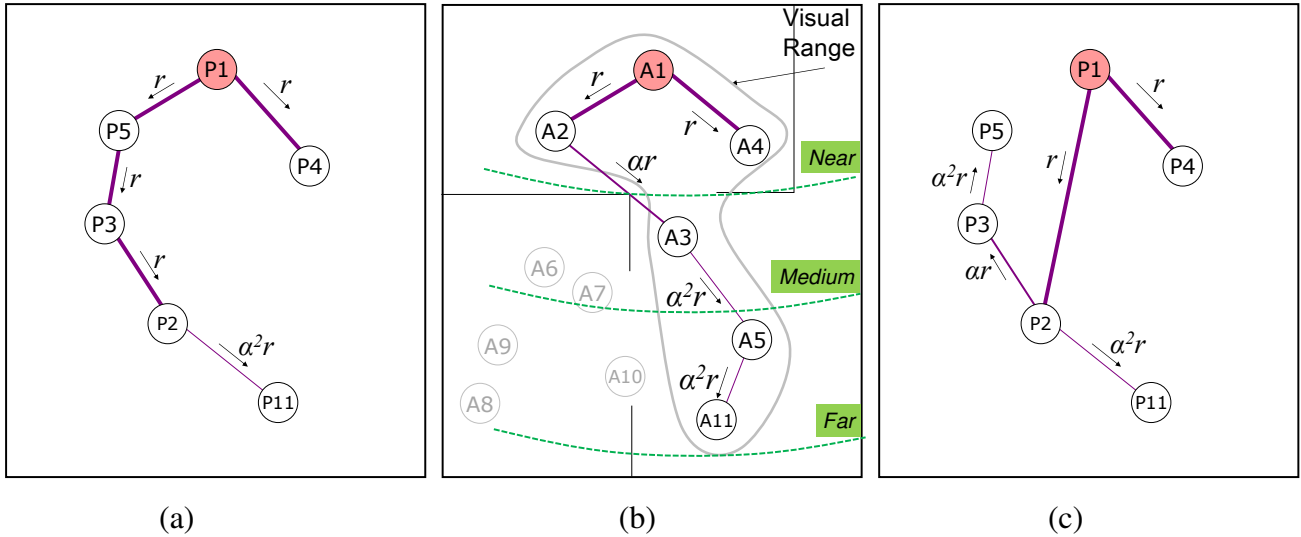


Figure 6: The multicast tree with quality culling for dissemination of peer P1 (A1) video. The shortest path tree is created using (a) the physical distance, (b) the virtual distance, and (c) a hybrid approach as the cost metric. Three zones within the visual range are shown and α represents the bit rate reduction factor associated with video quality reduction from one zone to another.

In contrast, Figure 6-(b) depicts the case where the multicast tree is constructed based on the arrangement of peers in the virtual scene, using the *virtual distance* between the avatars as the primary ‘cost’ metric. The advantage of this scheme is that as the multimedia content travels down the tree, the virtual distance is also increasing, which means that the peer can tolerate lower quality. For example, A2 and A4, who are the closest avatars to A1, receive the video with minimal delay and at the highest available quality, while the path to A5 has three overlay hops (path P1–P2–P3–P5). As video packets travel down the tree, the perceptual pruning can be applied before relaying the video to the rest of the tree. Quality culling, therefore, is naturally supported by this approach with little or no overhead. The other advantage of this approach is that there is no need to exchange topology information for construction and alteration of trees. This is because the details of the virtual topology are already available to each peer as part of their virtual context. In particular, each peer can respond to changes in multicast trees independently and consistently.

The key drawback of using the virtual distance as the basis of route optimization is that avatars that are close in the virtual environment may be physically far from each other leading to inefficient utilisation of network transmission capacity. Consequently, the third approach shown in Figure 6-(c) is a hybrid scheme to tackle the shortcomings of the previous two. In this case, we use the physical

distance to create the shortest path tree, but employ the algorithm in three successive stages according to the virtual zones.

The multicast trees in an IVC, therefore, are strongly influenced by the virtual context. First, the formation of each tree and the identity of nodes in the tree are determined by the visibility culling process. Second, the virtual distance is used as the primary metric to organize the tree, as opposed to the physical distance.

5. THE NEXT CHALLENGE – MANAGING CONGESTION

The visibility and quality culling techniques described in this paper aimed at minimizing the required bit rate, but what happens if the available network resources cannot support this minimum? The conventional TCP-style method of increasing latency to deal with congestion is not suitable for a real-time interactive service. Also, leaving it to network routers to randomly and indiscriminately drop packets from the multitude of video and audio streams could result in a poor user experience. What is needed is *judicious discarding of perceptually less salient* content, where saliency is determined by the current virtual context of each user *independently*.

There are Quality of Service (QoS) differentiation mechanisms for preferential treatment of flows within a network. The current methods typically assume a fixed relative ranking among the various streams. For example, it is possible to tag a certain video stream to have higher priority over another. However, virtual context is user-specific and also changes rapidly. In other words, the same video stream will have to receive different loss treatments depending on the destination, and this arrangement is susceptible to alteration on short time scales.

We propose that managing congestion for IVC requires further research, especially with respect to the following three key challenges:

- 1- The ability to rank multimedia flows, in real-time, based on their importance/saliency so as to minimize the perceptual impact of loss. This is a non-trivial procedure, taking into account the impact of virtual context on human perception, including the role of the audio in drawing attention to the video content, and also the interplay between video codec parameters and the impact of loss.
- 2- Develop flexible mechanisms within the network to provide QoS differentiation of flows per stream *and* per user, based on a rapidly changing saliency ranking of flows. One promising approach would be to consider Software Defined Networking for this purpose.
- 3- Develop mechanisms to mitigate the ‘control loop’ latency, that is, the discrepancy between the slower processes of changing QoS parameters within the network versus rapid fluctuations of virtual contexts.

6. CONCLUSIONS

Understanding the physical context of a user is highly beneficial for the optimization of service delivery and multimedia communications. This paper demonstrates that for a certain class of applications where real-time multimedia is critical, the virtual context plays an even greater role. The virtual context has the advantage of being easier to compute. The key challenge in this case is the rapidity by which the relevant attributes of the virtual context change. As such, fast and efficient mechanisms must be provided by the network and multimedia streaming and encoding sub-systems to adapt to these changes dynamically and efficiently.

6. REFERENCES

- [1] http://youtu.be/hdl0IO_8UGw
- [2] Boulanger, J., et al. "Comparing interest management algorithms for massively multiplayer games." 5th ACM SIGCOMM workshop on Network and system support for games, 2006.
- [3] Cohen-Or, D., et al., "A survey of visibility for walkthrough applications" IEEE Transactions on Visualization and Computer Graphics, vol. 9 , issue3, pp.412,431, 2003
- [4] Pourashraf, P., Safaei, F., and Franklin, D., "Distributed Area of Interest Management for Large-Scale Immersive Video Conferencing", ICMEW Workshop of IEEE International Conference on Multimedia and Expo ICME2012, pp. 139-144.
- [5] Lin, Weisi, and C-C. Jay Kuo. "Perceptual visual quality metrics: A survey", Journal of Visual Communication and Image Representation 22.4 (2011): 297-312.
- [6] Pourashraf, P., Safaei, F., and Franklin, D., "Minimization of video downstream bit rate for large scale immersive video conferencing by utilizing the perceptual variations of Quality", IEEE International Conference on Multimedia and Expo ICME2014.
- [7] Perera, C. et al., "Context Aware Computing for The Internet of Things: A Survey," IEEE Communications Surveys & Tutorials, vol.16, issue 1, pp 414-454, 2013.
- [8] M. Castro, et al., "Scribe: A Large-Scale and Decentralized Application-Level Multicast Infrastructure" IEEE Journal on Selected Areas in Communications, vol. 20, issue 9, pp 1489-1499, 2002.
- [9] Dowlathshahi, M., Safaei, F., "Overlay Multicasting of Real-Time Streams in Virtual Environments", Proceedings of IEEE Globecom conference, USA, 2006.

BIOGRAPHIES



Farzad Safaei graduated from the University of Western Australia with the degree of Bachelor of Engineering (Electronics) and obtained his PhD in Telecommunications Engineering from Monash University, Australia. Currently, he is the Professor of Telecommunications Engineering at the University of Wollongong. Before joining the University of Wollongong, he was the Manager of Internetworking Architecture and Services Section in Telstra Research Laboratories. His research interests include multimedia communications and immersive multimedia.



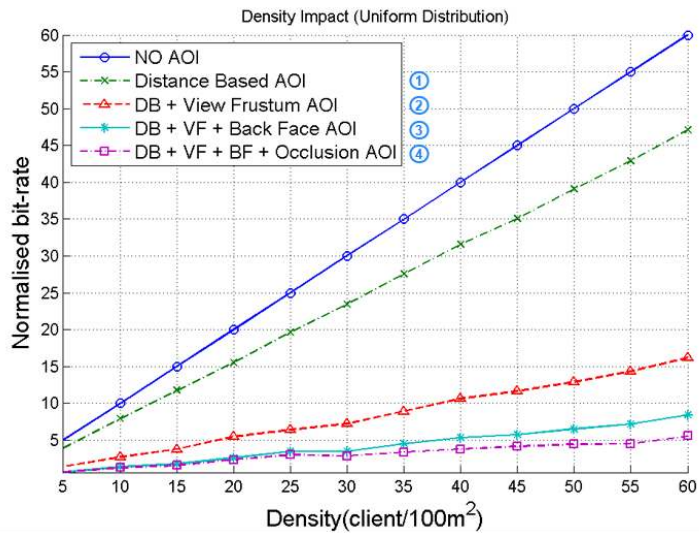
Pedram Pourashraf received his PhD degree in Telecommunications Engineering – “Immersive Multimedia Systems” from the University of Wollongong (UOW), Australia in 2014. He is currently a Research Fellow at the ICT research centre at UOW, where he is working on a large-scale immersive video conferencing technology. His research interests include video coding, perceptual video processing, 3D immersive environments and large-scale telecommunication networks.



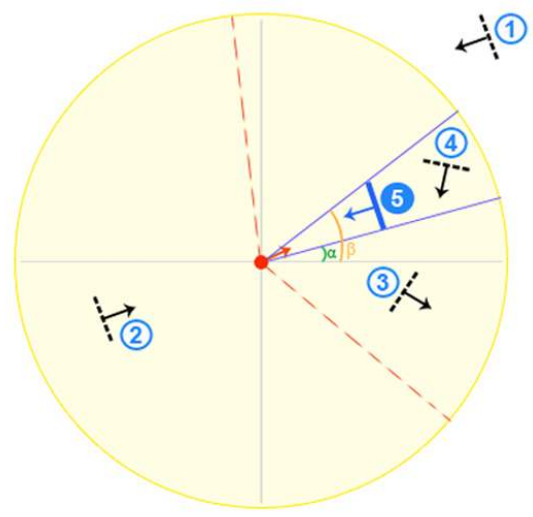
Dr Franklin completed his PhD in Telecommunications Engineering at the University of Wollongong (2007) and also holds a Bachelor of Engineering (electrical) - Honours I, University of Wollongong (1999). He is currently a Senior Lecturer in the Faculty of Engineering and Information Technology at the University of Technology, Sydney. His research and commercial interests include cooperative communications, mesh networks, software radio, embedded and real-time systems, and analog and digital electronics.



Figure 2 - Immersive video conferencing



(b)



(b)

Figure 2: (a) Reduction in the mean number of downloaded video streams by each client due to visibility culling; (b) Example of visibility culling attributes: (1) outside the visible distance, (2) outside the view frustum, (3) back face culling, (4) occluded avatar. The local avatar is shown as a red dot in the centre.

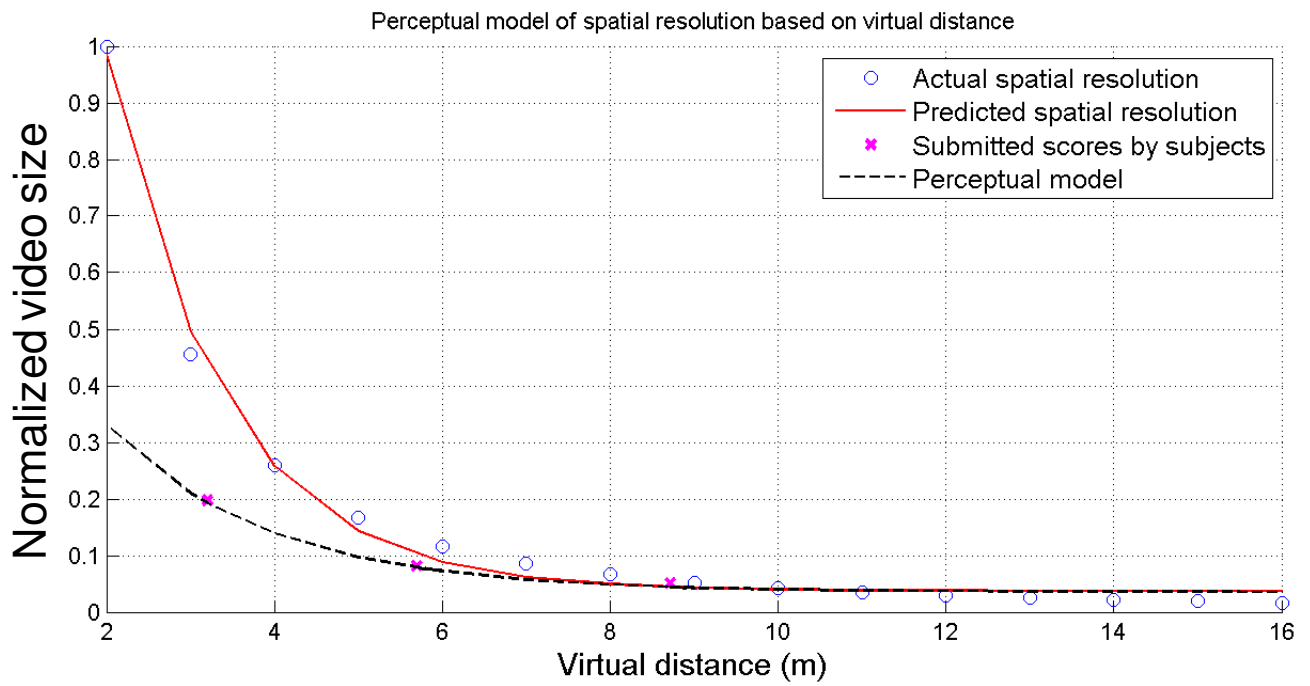


Figure 3: Reduction in the size of video as a function of virtual distance. The solid line represents perceptually ‘unnoticeable’ change and the dashed line is ‘slightly noticeable’.



Ref-00



Ref-30



Ref-60



Ref-80



Deg-00



Deg-30

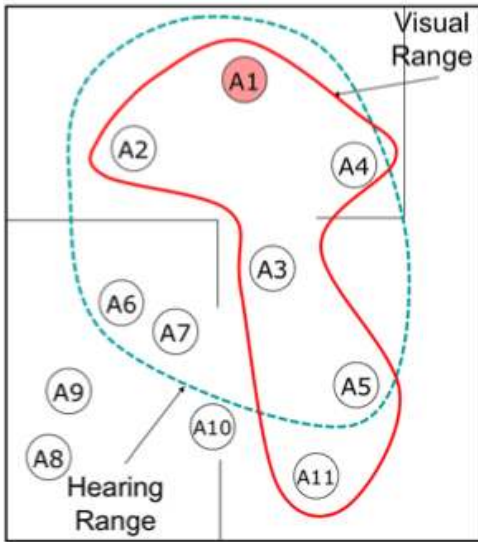


Deg-60

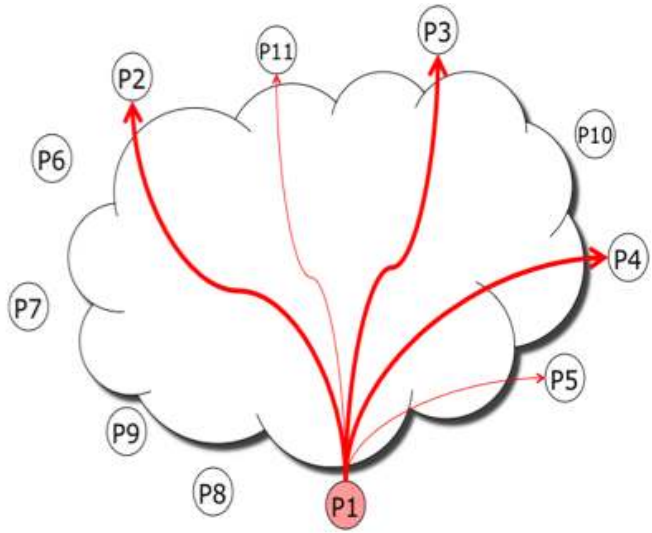


Deg-80

Figure 4: Reference (Ref) and Degraded images (Deg) after Perceptual Pruning for various virtual orientations



(a)



(b)

Figure 5: (a) The location of A1, the virtual representation of P1, with respect to the virtual crowd. (b) The unicast dissemination of video originated at peer P1 is shown. A thinner arrow signifies the delivery of content at lower resolution or frame rate.

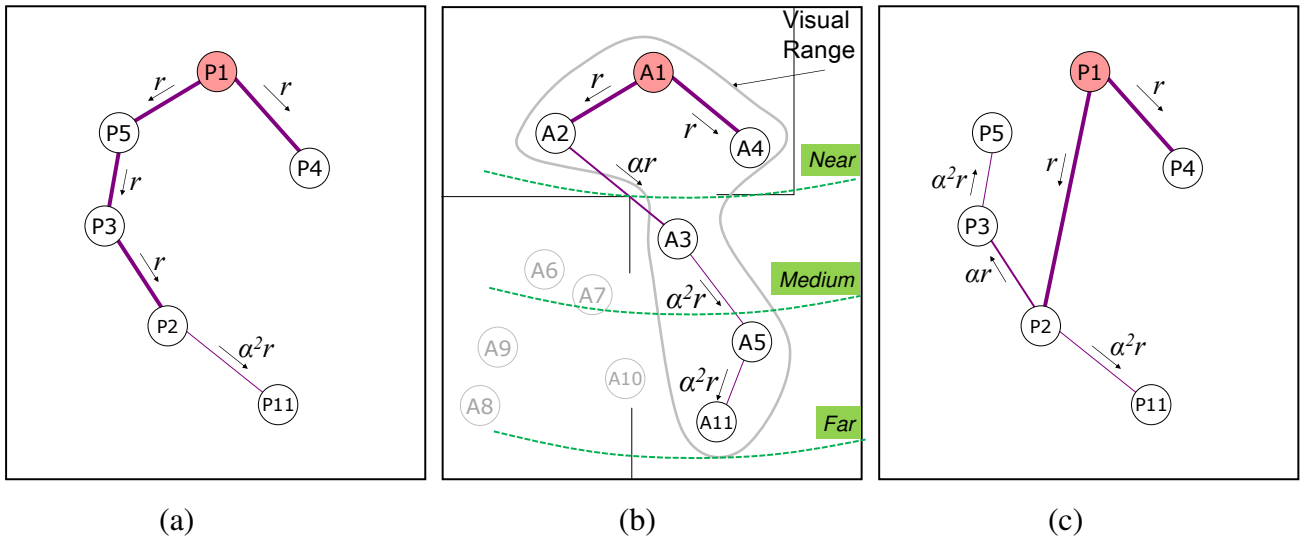


Figure 6: The multicast tree with quality culling for dissemination of peer P1 (A1) video. The shortest path tree is created using (a) the physical distance, (b) the virtual distance, and (c) a hybrid approach as the cost metric. Three zones within the visual range are shown and α represents the bit rate reduction factor associated with video quality reduction from one zone to another.