Photo-Conferencing: a novel approach to interactive photo sharing across 3G mobile networks

Kharsim Yousef
Department of Computer Science
University of Bath,
Bath, BA2 7AY.
kam26@bath.ac.uk

Eamonn O'Neill
Department of Computer Science
University of Bath,
Bath, BA2 7AY.
eamonn@cs.bath.ac.uk

ABSTRACT
This paper describes a novel interaction technique that allows mobile phone users to exchange photos interactively during an active phone conversation. Unlike existing lab based solutions, the one presented in this paper works in the field across existing mobile 3G networks. We believe this technique has the potential to revolutionise media exchange practices and the future of in call mobile interactions.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *User-centered design.*

General Terms
Documentation, Design, Reliability, Experimentation, Theory,

Keywords
Mobile conferencing, 3G, photo sharing, camera phone.

1. INTRODUCTION
New and improved third generation mobile networks are constantly on the horizon, from GPRS, UMTS to HSDPA and upcoming fourth generation WiMax offerings. Mobile operators and device makers are constantly striving to overcome bandwidth constraints and restrictive data connections. In this paper we present a prototype that demonstrates the use of existing 3G networks to overcome the constraints of an existing mobile service (MMS) and to demonstrate the potential available in today's 3G networks to create compelling mobile to mobile interactive services.

The business cases for Multimedia Messaging Service (MMS) and 3G services were predicted by analysts in 2002 to at least partially justify the heavy investments made by operators in Europe for 3G licences and services, which totalled over $100 billion across Western Europe. Recent evidence suggests that despite the heavy investments into 3G networks, new services such as MMS have failed to take off and picture messaging has been described as "a flop" by The Economist [9], with SMS still remaining the dominant data application globally for 2006 [11].

Recent research into the mobile capture-culture has identified numerous advantages, from providing opportunistic, spur-of-the-moment capture [18] to enabling the creation of "life documents" Plummer [17]. However, despite the 3G capabilities provided by devices for potentially sharing captured images, cameraphone use largely remains restricted to 'capturing' rather than sharing photos by means such as MMS. Kindberg et al [13] note that the majority of photos are not sent immediately via MMS, but rather shared directly on the mobile device's screen when the participants are physically co-present. Reporting camera phone adoption in Japan, Okabe [16] presented a similar finding that users tend not to email images to one another, but preferred to share directly using the handset screen itself and that 'People are in fact frustrated when trying to share images remotely and interactively' [16].

The need for interactivity can also be traced back to earlier ethnographic studies of domestic photography by Chalfen, in which he argued that "[domestic photographs] are meant to be shared, and they are meant to prompt interaction" [5]. Existing research into mobile media sharing has identified limitations to MMS services [12, 14, 16] and presented alternative solutions [15, 18]. However these solutions don't differ significantly from MMS and have not sought to provide remote interactivity. In this paper we seek to present one such solution to achieving remote interactional photo sharing over existing 3G networks.

2. PROTOTYPE SYSTEM
To investigate our goals, we developed a complete mobile photo-conferencing system (see Figure 2), comprising remote mobile to mobile session initiation protocols, client based software and interactive screen adaptation. The client software was developed to run on most Windows Mobile devices (Smartphones and Pocket PCs), supporting the .net compact framework 2.0.

2.1 Mobile P2P Networking
The emerging body of research on cameraphone use suggests that people want to share images, however mobile users are frustrated when trying to share images remotely and interactively [1]. This section covers the key networking technologies used in our prototype to achieve remote interactive image sharing. The initial prototype underwent several major networking overhauls over the course of the development, as selecting the correct networking approach was vital to creating a working prototype.

Simtech'07, November 26-27, 2007, Melbourne, Australia.

Initiating interactive user sessions requires a means to connect multiple participants. SIP (Session Initiation Protocol) is one such signalling mechanism used in the establishment, modification and termination of networked sessions between one or more participants. Early iterations of our client used standard TCP connections over WiFi to enable SIP functionality. This approach is in many ways similar to approaches used on desktop machines. The inbuilt WiFi chips in the devices we used provided the IP addressability needed to allow SIP's session management protocols to establish the required connections and UDP/TCP to route the data back and forth between multiple mobile devices.

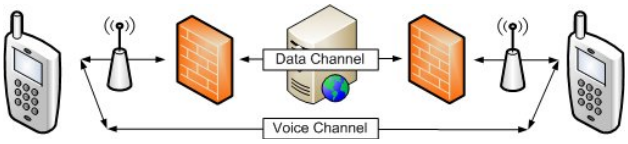

Figure 1. Mobile to mobile connections illustrating voice and data channels, the latter restricted by firewalls.

Experimenting with our WiFi based solution however uncovered many limitations to this approach. Requiring WiFi coverage was an obvious drawback to this system given the currently patchy nature of WiFi coverage. But as we discovered, stringent networking restrictions (see Figure 1) over academic and corporate networks, some of which operate port-restricted address translators and traffic-restricted firewalls, prevented devices running on separate WiFi networks from connecting to one another.

Initially we examined several methods using WiFi to overcome these limitations, such as disguising our traffic in the form of Skype calls [2] that could bypass some firewalls. However we also wanted to overcome WiFi coverage restrictions, which would prevent our solution being used by field engineers whom we consider to be among our target users. We therefore re-examined the mobile data technologies that are currently available, such as 2G (e.g. GPRS) and 3G networks (e.g. HSDPA, UMTS, EDGE).

On first examination connection speeds appear to be the main differentiator between the two technologies, with 3G networks boasting faster connection speeds than 2G networks. But 2G and 3G networks differ in one very important aspect: unlike their predecessors, 3G devices are able to concurrently establish data and voice connections, enabling the exchange of data during active phone conversations. This ability is central to achieving mobile conferencing functionality.

Firewalls and restrictions imposed on WiFi networks are still a concern in 3G networks where mobile operators have installed firewall systems and ingress filtering to prevent inbound data connections to mobile devices. Emerging technologies as IP Multimedia Subsystem (IMS) [4], could overcome these limitations by delivering internet protocols to mobile users and enabling point to point connections. However IP based platforms such as IMS are not available to the majority of mobile phone users, limiting its current applicability to lab based scenarios.

The challenge then is how to overcome network restrictions (see Figure 1) and enable SIP functionality over IP-less connections. We experimented with a number of different approaches. The most successful were based on existing web principles such as AJAX [19] and the use of 'XMLHttpRequest' to synchronously exchange data between clients.

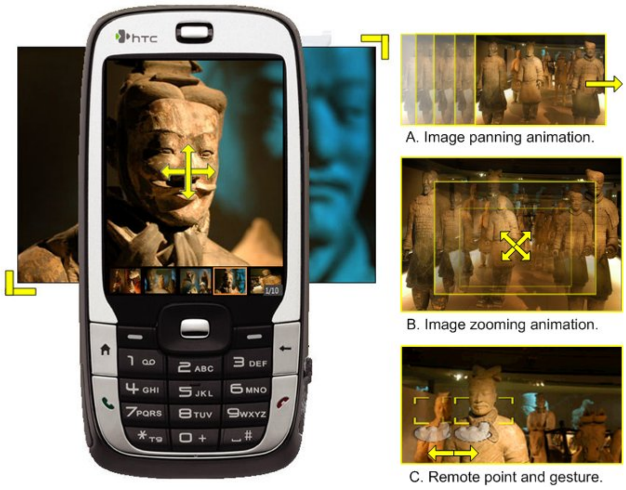

Figure 2. Photo-Conferencing user interface, supporting remote Pan, Zoom and Gestural interactions.

The majority of firewalls used by mobile operators are typically set up to block specific data packets such as UDP. These firewalls don't however block the HTTP traffic typically associated with web browsing. By using HTTP 1.1 GET requests and tunnelling techniques it's possible to traverse firewalls. In this approach data is packaged into HTTP streams which are sent to a relay, that extracts and delivers the packets to the desired endpoint.

This approach offers some advantages: it can operate over existing 3G and WiFi networks, with the relays providing additional buffering to cope with connectivity loss that typically occur in mobile environments. However the disadvantage of this approach is that HTTP requests, in contrast to TCP, are not primarily designed for real time traffic. Therefore additional application specific optimisation techniques are required to achieve optimal results.

2.2 Shared Mobile Interactions

The conferencing solution consists of a rich user interface (see Figure 2) that can be initiated at any point during an active voice conversation to enable instant media exchange, and when idle to view prior sessions. The user interface has been designed to support conferencing 'What You See Is What I See' (WYSIWIS) functionality, in which media content and gesture interactions are replicated across all connected devices.

The interface currently supports a number of remote media gesturing techniques: Panning, Zooming, Pointing and Motion. These provide the mechanisms through which users can indicate focus during a conferencing session and construct what Crabtree et al. [7] describe as "a host of fine grained grammatical distinctions".

The WYSIWIS synchronisation of on screen elements is achieved through a process of 'state' propagation. Due to the roaming capabilities of mobile users, wireless 3G/WiFi connectivity is far more susceptible to signal drops which are a common cause of data loss.

In a typical networking scenario it would suffice to propagate each event across the network e.g. pan left, right, zoom in, etc. However, a single data loss could render several remote mobile clients out of sync. To overcome this problem, we adapted our protocols to exchange 'state' information rather than 'event' data. State information consists of significant attributes pertaining to active components, e.g. the displayed image's zoom length and x,y co-ordinates. This allows the system to be far more resilient to packet loss and out-of-order events. It also enables throttling of incoming packets that can reduce CPU utilisation during animation. One drawback to this approach is that in comparison to single event transmission, it incurs additional data overhead that may require compression.

2.3 Mobile Content Adaptation

Enabling mobile to mobile connections, creating shared interaction spaces and carefully optimizing the client software has allowed us to extend our photo-conferencing capabilities across a large number of mobile devices currently available on the market, from low-end Smartphones to more powerful Pocket PC devices. In today's mobile market consumers are however presented with a greater choice of devices, form factors and screen resolutions to meet their individual needs (see Figure 3). These variations present new challenges in screen adaptation that mobile services need to overcome in order to succeed.

Figure 3. Illustration of some of the different mobile devices and resolutions available.

The first wave of MMS suffered from interoperability issues, in which messages created by some MMS phones were not entirely compatible with the capabilities of recipient MMS phones [3, 6]. Although MMS interoperability issues still exist today, mobile operators were quick to learn from their mistakes and introduced dynamic content adaptation techniques such as MMSC [8] to rectify initial user experiences and encourage the adoption of MMS services.

Key to the photo-conferencing solution presented in this paper is the maintenance of a shared visual space and deictic referencing, in which the mechanics of collaboration [10] can be applied. For such a solution to succeed, it needs to overcome such interoperability issues. In the following section we present two preliminary techniques, 'content framing' and 'content transformation', that enable cross-device content adaptation during photo-conferencing sessions.

2.3.1 Content Transformation

Content transformation is a technique in which the source (original) image is modified to accompany variations in the target device's screen orientation and resolution (see Figure 4) whilst maintaining deictic referencing. The transformation consists of varying the image's dimensions and aspect ratio in order to apply stretching across the available display space on each device.

The top half of Figure 2.2 illustrates the image content as it would appear on the screen of a 240x320 (Portrait QVGA) display, with the bottom half illustrating how it would appear on a 320x240 (Landscape QVGA) display. These are two common screen resolutions, found on many of the latest mobile devices such as the HTC S710 and the Motorola Q9 (see Figure 3) respectively.

The advantage of this approach is that it utilizes all of the mobile device's screen real-estate, whilst maintaining an acceptable level of support for deictic referencing, in which a question such as "What colour is the flag in the bottom right?" would return the same answer with both display resolutions (see Figure 4). Additionally, when performing transformations to displays which are variant multiplications of the source display, for example displaying the content from a 240x320 [QVGA] device to a 480x640 [VGA] display found on many Pocket PCs such as the iPAQ 200 (see Figure 3), there is no image skewing during transformation since both screens have identical aspect ratios.

However, not all content is suitable to undergo transformations in which skewing occurs. Schematic and textual content may become much harder to read, which can affect usability.

2.3.2 Content Framing

This technique provides an alternative to content transformations and is more suitable for sharing textual and schematic contents across mobile devices as no transformation or skewing is applied to the original image and aspect ratios are maintained. Content framing uses subtraction method A ∩ B (see Figure 4) in which both screens permit shared content to be viewed, shading out areas not viewable on both device screens. This allows both participants to interact around shared visual content, without incurring any distortions.

In comparison to content transformation, content framing doesn't make the most of the entire pixel repertoire provided by the mobile device, which can be a restricting factor given the already limited pixel range available on most mobile devices.

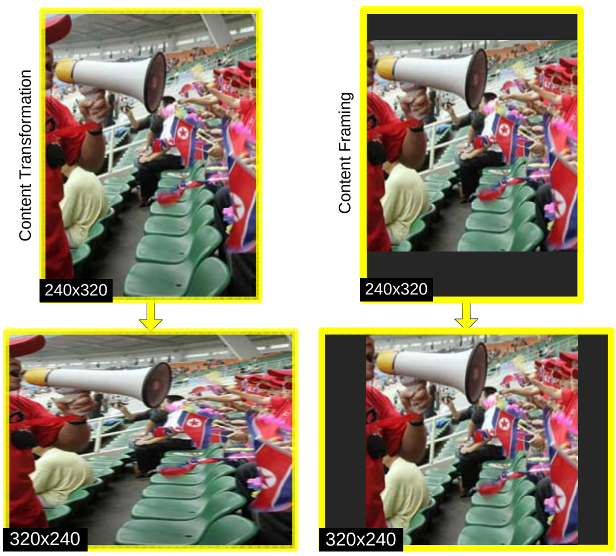

Figure 4. An example of content transformation (left) in comparison to content framing (right).

3. EVALUATIONS

The solution presented in this paper has undergone an extended development cycle that was essential to the development of the custom networking protocols, mobile client and strict optimizations required to enable concurrent voice, data exchange and animations to operate smoothly on Smartphone devices.

We have since showcased a working demo at the "Wireless 2.0" conference, which received great interest and feedback from mobile industry participants. Additionally we are in the process of recruiting participants for lab and field based ethnographic studies to further examine interaction aspects of our mobile 3G photo-conferencing solution.

4. CONCLUSION

In this paper we presented a new mobile-to-mobile interaction technique that we believe overcomes the problems inherent in existing media sharing services. Our system demonstrates rich interactional P2P capabilities that can operate throughout existing 3G mobile networks and addresses the important issue of mobile content adaptation. Content transformation and content framing techniques are demonstrated to enable rich media sharing across mobile devices adapting to variations in screen resolution.

Some of the possible usage scenarios of this approach extend beyond that of social media sharing to DIY assistance, e.g. "which button should I press – look, they all seem to be red" and to professional field engineers, e.g. "just sent you the latest schematics, let me talk you through the new alterations before you start repairs".

This research constitutes an initial prototype of a new form of mobile-to-mobile media sharing service that is spontaneous, dynamic and can occur during an active phone conversation. Although the first wave of 3G media sharing solutions such as MMS created sometimes fractured interactions, we believe that the next wave will offer richer, more coherent user experiences.

5. ACKNOWLEDGMENTS

Thanks to Vodafone Group R&D for their support for this research.

6. REFERENCES

[1] Aoki, P. M. 2005. Turning from Image Sharing to Experience Sharing 2005), First Workshop on "Pervasive Image Capture and Sharing: New Social Practices and Implications for Technology"(PICS), Ubicomp 2005.

[2] Baset, S. A. and Schulzrinne, H. 2004. An Analysis of the Skype Peer-to-Peer Internel Telephony Protocol 2004), Arxiv preprint cs.NI/0412017.

[3] Bodic, G. L. 2003. Multimedia Messaging Service: An Engineering Approach to MMS. John Wiley and Sons.

[4] Camarillo, G. and García-Martín, M. A. 2004. The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds. John Wiley and Sons.

[5] Chalfen, R. 1998. Family photograph appreciation: Dynamics of medium, interpretation and memory 2-3 1998), Communication & cognition. Monographies, 31, 161-178.

[6] Coulombe, S. and Grassel, G. 2004. Multimedia adaptation for the multimedia messaging service 7 (Jul 2004), Ieee Communications Magazine, 42, 120-126.

[7] Crabtree, A., Rodden, T. and Mariani, J. 2004. Collaborating around collections: informing the continued development of photoware 2004), Proceedings of the 2004 ACM conference on Computer supported cooperative work, 396-405.

[8] Daniel Ralph, P. G. 2003. MMS: Technologies, Usage and Business Models. John Wiley and Sons.

[9] Economist, T. 2006 Picture messaging: Lack of textual appeal. The Economist, 380, 8489 (Aug. 2006) 56.

[10] Gutwin, C. and Greenberg, S. 2000. The Mechanics of Collaboration: Developing Low Cost Usability Evaluation Methods for Shared Workspaces 2000), Proceedings of the 9th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 98-103.

[11] IDC 2006 Survey indicates that less than 10% of users are utilizing services other than SMS. Press Release. http://www.idc.com/getdoc.jsp?containerId=pr2006_03_03_130022.

[12] Ito, M. 2005. Intimate Visual Co-Presence 2005), Pervasive Image Capture and Sharing Workshop, Ubicomp 2005.

[13] Kindberg, T., Spasojevic, M., Fleck, R. and Sellen, A. 2004. How and Why People Use Camera Phones 2004), Consumer Applications and Systems Laboratory. H&P Laboratories Bristol, HPL-2004-216, November, 26, 2004.

[14] Kindberg, T., Spasojevic, M., Fleck, R. and Sellen, A. 2005. The ubiquitous camera: an in-depth study of camera phone use 2 2005), Pervasive Computing, IEEE, 4, 42-50.

[15] Maia Garau, J. P., Scott Lederer, Chris Beckmann 2006. Speaking in Pictures: Visual Conversation Using Radar 2006), Second Workshop on "Pervasive Image Capture and Sharing: New Social Practices and Implications for Technology"(PICS), Ubicomp 2006.

[16] Okabe, D. 2005. Social practice of Camera Phone in Japan 2005), First Workshop on "Pervasive Image Capture and Sharing: New Social Practices and Implications for Technology"(PICS), Ubicomp 2005.

[17] Plummer, K. 2001. Documents of Life 2: An Invitation to a Critical Humanism. Sage Publications.

[18] Van House, N. A. and Davis, M. 2005. The Social Life of Cameraphone Images 2005), Proc. Pervasive Image Capture and Sharing Workshop at the Seventh Int. Conf. on Ubiquitous Computing.

[19] Zakas, N. C., McPeak, J. and Fawcett, J. 2006. Professional Ajax. Wiley.