

Let's Meet at the Mobile – Learning Dialogs with a Video Conferencing Software for Mobile Devices

[doi:10.3991/ijim.v3i3.773](https://doi.org/10.3991/ijim.v3i3.773)

Hans L. Cycon¹, Thomas C. Schmidt², Gabriel Hege², Matthias Wählisch^{2,3}, and Mark Palkow⁴

¹ FHTW Berlin, Berlin, Germany, ² HAW Hamburg, Hamburg, Germany

³ link-lab, Berlin, Germany, ⁴ daViKo GmbH, Berlin, Germany

Abstract—Mobile phones and related gadgets in networks are omnipresent at our students, advertising itself as *the platform for mobile, pervasive learning*. Currently, these devices rapidly open and enhance, being soon able to serve as a major platform for rich, open multimedia applications and communication. In this report we introduce a video conferencing *software*, which seamlessly integrates mobile with stationary users into fully distributed multi-party conversations. Following the paradigm of flexible, user-initiated group communication, we present an integrated solution, which scales well for medium-size conferences and accounts for the heterogeneous nature of mobile and stationary participants. This approach allows for a spontaneous, location independent establishment of video dialogs, which is of particular importance in interactive learning scenarios. The work is based on a highly optimized realization of a H.264 codec.

Index Terms—Mobile video-based learning, H.264/MPEG-4 AVC software codec, mobile conferencing, peer-to-peer group communication, distributed SIP conference management.

I. INTRODUCTION

Remote and distance learning aims at facilitating a process of immersive qualification and life-long personal development, and does not primarily focus on a supplementary aid for memorizing facts. Neither the simple, archaic look-up model of Wikipedia, nor the schoolmasterly presentation of instructional films does advance an active, involved process of learning, which is demanding and seldom arises without incentives. Nevertheless, the motivation of students may be stimulated by problem oriented tasks and imaginative environments.

The discursive acquisition of knowledge and understanding requires dialogs, which significantly gain intensity by the synchronous video-based presence among participants. Live Q&A sessions based on mobiles enable students in an inexpensive way to talk with experts as presented in [18]. Mobile phones and their wide usage allow to get in contact with almost all professions and thereby grant a "real-life" experience to active learners.

It is important to enrich discussion processes by group orientation, but unnatural to restrict interaction to two-party dialogs. The idea of video conferencing in groups is around for some time, now, but only the flexibility of the Internet generated a noticeable deployment. As compared

to audio, video processing places significantly higher demands on end system and network transmission capabilities. The rapid evolution of networks and processors have paved the way for realistic group conferences conducted at standard personal computers, combining about a dozen visual streams of Half-QVGA (240 x 160 pixel @ 15-30 fps) resolution.

Mobile phones and networked consumer portables are now on the spot to deliver sufficient performance for rich multimedia applications and communication, as well. Videoconferencing though, which requires simultaneous decoding and encoding in real-time, poses still a grand challenge to the mobile world. Limited and expensive wireless channels on the one hand, high consumer demands on visual quality on the other, advise applications to take advantage of the latest standard for video coding H.264/AVC [1].

H.264/AVC provides gains in compression efficiency of up to 50 % over a wide range of bit rates and video resolutions compared to previous standards. While H.264/AVC decoding software has been successfully deployed on handhelds, high computational complexity still prevented pure software encoders in current mobile systems. There are, however, also fast hardware implementations available, which give rise to an increasing offer of device- and operator-bound video services.

In this work we first introduce a pure *software* solution for real-time video communication on standard smartphones in section 2. These mobile clients extend a lightweight, feature rich conferencing application developed for an infrastructure compliant use on standard PCs. This software includes ad hoc group building capabilities, as well as an application sharing facility. It is the vision that future learning environments will seamlessly integrate the conferencing software, so that students can collaborate on topics and tasks while at home or moving, or professional technicians, physicians, etc. can discuss problems 'in the field' and receive support from back-office experts at difficulties encountered. It is expected for many cases, that the ability to transmit visual streams from any location will greatly simplify distant collaboration and personal understanding. At the present of technical emergence, though, it is too early to report on practical real-life experiences, which is reserved for the future time of deployment. However, in the next section we will give some indications about future learning on mobiles.

In the second part of this paper we present the underlying peer-to-peer group communication scheme, which performs well for medium-size conferences and accounts for the heterogeneous nature of mobile and stationary participants, cf. section 2 and 3. This includes on the one hand SIP [2] standard compliant session signaling with respect to group communication, and on the other hand efficient, serverless media distribution, self-adjusting to the actual network infrastructure support. Conclusions and an outlook on future work follow in the final section.

II. FUTURE LEARNING WITH MOBILE PHONES

Mobile phones are widely spread. They are continuously carried by children, adults as well as ancients, who all are familiar with its usage. The well-known employment of mobiles as well as their tight integration makes these devices a promising tool for upcoming learning scenarios. In the following, we leave aside technical issues as low battery power or poor network connectivity, and sketch two straightforward learning scenarios for mobile phones equipped with a video transmission component.

In contrast to desktop PCs or notebooks mobile phones possess the advantage of a very small chassis, which let them inconspicuous and thus applicable to experiments based on observations. For example, the biology subject in elementary schools can be enhanced, when pupils use their mobile phone to keep track of domestic animals. Mobiles will be distributed over the schoolyard or in specific vivariums. Based on the integrated cam, the video stream will be transmitted to a classroom without disturbing the natural behavior of the animal.

Mobile phones displace increasingly dedicated cameras in the context of tourism. Keeping in mind that mobiles are almost online, recorded data of sights can be sent immediately to online portals or web pages. Embedding such information in encyclopedia sites, may provide the learner with an up-to-date view of the location under consultation. Recalling the social change from privacy to publicity, e.g., while people upload holiday photos to portals like Flickr, it is not unlikely to assume that tourists are frank to share their current experiences. Video conferencing enforces the liveliness between the learner and the people on site.

III. THE DAVIKO VIDEOCONFERENCING SOFTWARE

In this section we give an overview of our reference implementation, a digital audio-visual conferencing system, realised as a serverless multipoint video conferencing software without MCU developed by the authors [3]. It has been designed in a peer-to-peer model as a lightweight Internet conferencing tool aimed at email-like friendliness of use.

The system is built upon a fast H.264/MPEG-4 AVC standard conformal video codec implementation [4] called DAVC. By controlling the coding parameters appropriately, the software permits scaling in bit rate from 48 to 1440 kbit/s on the fly.

Audio data is compressed using a 16 kHz speech-optimized variable bit rate codec [5] with extremely short latencies of 40 ms (plus network packet delay). All streams can be transmitted by unicast as well as by multicast protocols. Within the application, audio streams

are prioritized over video since user experience is usually more sensitive to losses in audio packets than those of video packets, which both may result from transmission errors or network congestions.

An application-sharing facility is included for collaboration and teleteaching. It enables participants to share or broadcast not only static documents, but also any selected dynamic PC actions like animations. All audio/video-streams including dynamic application sharing actions can be recorded on any site. This system is equally well suited to intranet and wireless video conferencing on a best effort basis, since the audio/video quality can be controlled to adapt the data stream to the available bandwidth.

The daViKo conferencing system is available for desktop computers running MS-Windows or Linux and on handhelds with Windows Mobile operating system.

A. The DAVC Codec

DAVC codec, the core of the videoconferencing system, is a fast, highly optimized H.264/MPEG-4 AVC standard implementation. It realizes a Baseline Profile, optimized for real-time encoding and decoding by means of a fast motion estimation strategy including integer-pel diamond search as well as a fast sub pel refinement strategy up to $\frac{1}{4}$ pel motion accuracy. Motion estimation includes the choice of several different macroblock (MB) partitions and multiple reference frames, as permitted by the H.264/MPEG-4 AVC standard. For choosing between different MB partitions for motion-compensated (i.e. temporal) prediction and MB-based intra (i.e. spatial) prediction modes, a fast rate-distortion (RD) based mode decision algorithm with early termination conditions has been employed.

In comparison to the well-known open source H.264/MPEG-4 AVC encoder implementation of x264 [6], our DAVC encoder implementation achieves up to 0.5 dB PSNR better RD performance and a considerable increase in encoding speed when using comparable encoder settings. For selected RD points we measured 284 encoded frames per second (fps) as compared to 210 fps for x264 on an Akiyo CIF test sequence. Comparing with the H.264/MPEG-4 AVC Joint Model (JM) reference software implementation (Baseline profile settings, using a high-complexity RD-based mode decision strategy neglecting any real-time constraints) our DAVC achieves only about 0.5 dB PSNR less on a speed of 284 fps. Similar results were also achieved for other test sequences.

The DAVC codec along with the H.264/AVC design also includes some suitable mechanisms to quickly recover from video packet loss.

B. Mobile Video Codec

In ongoing work, the DAVC codec has been adapted to sustain real-time performance on mobile devices. The mobile codec version operates at reduced complexity for motion compensation with a highly optimized code base for the target platform. Motion compensation has been limited to work on 16 x 16 pixel blocks, only. The code tuning includes the efficient use of the wireless MMX instruction set available at the target system. Portability is sustained by an ANSI compliant C version, to be augmented incrementally by platform specific injections.



Figure 1. The Mobile Video Application

The application was tested on a 520 MHz Xscale processor built in an Asus P735 system. Thereon it can reliably encode and decode a QCIF video stream in parallel at 5/10 fps, without CPU exhaustion or frame dropping. Real-time encoding rate increases up to 10 fps for moderate video complexity. QCIF @15 fps is the maximal image feed that can be obtained from the front camera in our test equipment. Performance values for the mobile encoder are moderately lower as compared to the results for the full DAVC.

Reduced coding complexity results in an enhanced data rate sent by the mobile, but the gross total rate for a bidirectional video exchange at 10 fps complies to 3GPP/UMTS bandwidths constraints. A sample picture of the mobile conferencing application is shown in figure 1. For further information we refer to [19].

IV. DISTRIBUTED POINT-TO-MULTIPOINT CONFERENCING

Our application aims for simple, flexible, and cost-efficient ad-hoc conferencing functions, which scale appropriately well, but avoid any infrastructure assistance. Such a solution requires group session management and media distribution at peers, which for the sake of standard compliance we realize with group conferencing functions in SIP, cf. [7-9]. Implemented as pure software on standard personal devices, user agent peers are exposed to severe restrictions in real-world deployments: Often they are located behind NATs and firewalls with network capacities confined to asymmetric DSL or wireless links. Capacity constraints and resilience to node failures require peer-managed ad-hoc conferences to organize in a distributed multi-party model. As a key component, the heterogeneity of clients must be accounted for, whereas the range of scalability is limited to about a dozen parties in videoconferences.

A. P2P Adaptive Architecture

A peer-to-peer conferencing system faces the grand challenge to be robust with respect to the infrastructure. The role a user agent is able to attain in a distributed scenario needs to be adaptively determined according to constraints of its device and current network attachment. In a simplified scenario, clients may be divided into two groups, distinguished by their ability to act as a SIP conference focus or not. A focus must be globally addressable and have access to necessary processing and network resources.

This elementary adaptation scheme can be based on individual decisions of user agents and gives rise to a hybrid architecture of super peers, chosen from potential focus nodes, and remaining leaf nodes. To decide on its potential role of building a focus, a client at first needs to determine NATs and firewalls. Aside from address evaluation, this is done by a simple probe packet exchange. As the implementation is CPU-type aware, processing restrictions are easily evaluated, as well. However, an *á priori* judgement on available network bandwidth is not easily obtained. An evaluation of the local link capacity is frequently misleading, as wireless devices may be located behind wired transmitters of lower, asymmetric capacity such as in ADSL. Current experiments to quickly retrieve reasonable estimates of up- and downstream capacity are ongoing on the basis of variable packet size, nonintrusive estimators, cf. [10]. Note that network capacity detection is of vital use for temporal adaptation of the video codecs, as well.

Leaf nodes attach to super peers in subordinate position, whereas potential focus nodes may be assigned to be super peers or leaves. Super peers provide global connectivity among each other and NAT traversal assistance to leaves (Fig. 3), while leaf nodes experience super peers in different roles: A leaf nodes sees its next hop super peer as

the conference focus, while the remote super peers act as proxies on the path to the leaves behind.¹ This set-up corresponds to the well known architecture of Gnutella 0.6 and successive hybrid unstructured peer-to-peer systems, cf. [11]. Despite its architectural analogy, our routing layer for real-time group applications follows a different, next-hop design.

The daViKo video conferencing implementation is designed along the above hybrid network architecture with some additional “low number of peers features”. It works as a pure peer-to-peer network as long as there are less than 4 peers in a conference. There is a super peer which works in this case only as address reference in a LDAP server scenario [15, 17]. If there are 4 or more peers the super peer will channel the audio/video streams in a multicast manner (see Fig. 2). If all peers however are located within an intranet and connected with sufficient bandwidth, the super peer will, connect all participants in a peer-to-peer topology. A SIP stack implemented in the super peer provides additional daViKo connectivity to customary SIP based video conferencing systems.

V. CONCLUSIONS & OUTLOOK ON FUTURE WORK

We have presented a hybrid peer-to-peer software for high-quality videoconferencing on mobiles, admitting utmost flexibility with respect to end systems, operators and network provisioning. To the best of our knowledge, this was the first software implementation of an H.264 video encoder that operates in real-time on mobile phones at the time it was presented on CeBIT (March 2008). An adaptive, fully distributed conference management scheme with SIP has been developed as part of the multi-party scenario. This hybrid peer-to-peer model accounts for client capabilities as well as network attachment, and does scale well beyond standard use.

We are currently experimenting with an early deployment in teaching and learning scenarios. Thereby we target on augmenting the Hypermedia Learning Object System (hylOs) [12] by mobile and group communication tools. It is our vision that eLearning Objects accessed from this eLearning Content Management can be shared and discussed in video conferencing, while user presence can be equally supported on desktops and handhelds [13].

We are also about to extend our wired scenarios on distributed asynchronous and synchronous learning system on mobiles. An asynchronous mobile learning system has been deployed already using our so called Nibbler server [16] (see Fig. 4). This is a rapid deployment system for mobile learning content. Technically it is a multiformat transcoding system which semi automatically transcodes video and audio content into a large variety of different formats for web cast and streaming or download and display on almost any mobile device. Thus video coded learning content can be provided for pull on demand on mobiles. (a feature we may call lectures2go).

¹ This architecture relies on the presence of at least one globally addressable, sufficiently powerful peer. As there are many scenarios, where this is likely to fail, we advise for and offer a permanently deployed ‘silent’ relay-peer at some unrestricted place.

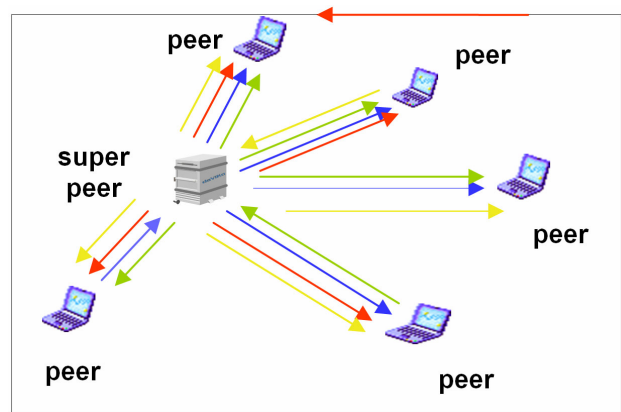


Figure 2. Super Peer-centric Multipoint Videoconferencing

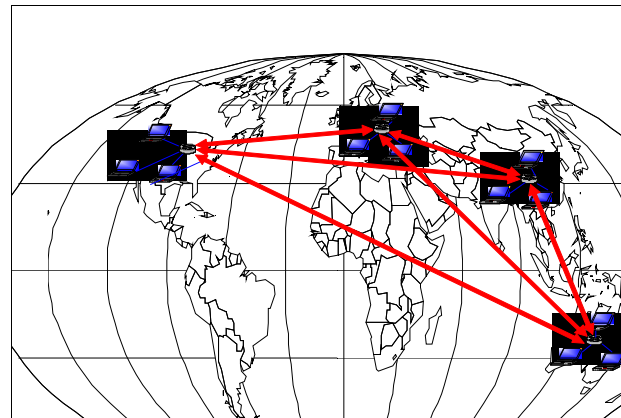


Figure 3. Global Hybrid P2P Network

REFERENCES

- [1] ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, Advanced Video Coding for Generic Audiovisual Services, ITU, Tech. Rep., 2005, draft Version 3.
- [2] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, SIP: Session Initiation Protocol, IETF, RFC 3261, June 2002.
- [3] M. Palkow, The daViKo homepage, 2008, <http://www.daviko.com>.
- [4] J. Ostermann, J. Bormans, P. List, D. Marpe, N. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, Video Coding with H.264/AVC: Tools, Performance and Complexity, IEEE Circuits and Systems Magazine, vol. 4, no. 1, pp. 7-28, April 2004. (doi:10.1109/MCAS.2004.1286980)
- [5] The Speex projectpage, <http://www.speex.org>, 2008.
- [6] VideoLan: x264 - a free h264/avc encoder, <http://www.videolan.org/developers/x264.html>, 2007.
- [7] A. Johnston and O. Levin, Session Initiation Protocol (SIP) Call Control – Conferencing for User Agents, IETF, RFC 4579, August 2006.
- [8] R. Mahy, R. Sparks, J. Rosenberg, D. Petrie, and A. Johnston, A Call Control and Multi-party usage framework for the Session Initiation Protocol (SIP), IETF, Internet Draft - work in progress 9, November 2007.
- [9] T. C. Schmidt and M. Wählisch, Group Conference Management with SIP, in SIP Handbook: Services, Technologies, and Security, S. Ahson and M. Ilyas, Eds. Boca Raton, FL, USA: CRC Press, pp. 123-158, November 2008, on invitation.
- [10] R. Prasad, C. Dovrolis, M. Murray, and kc claffy, Bandwidth Estimation: Metrics, Measurement Techniques, and Tools, IEEE Network, vol. 17, no. 6, pp. 27-35, November-December 2003. (doi:10.1109/MNET.2003.1248658)

