Multi-User Systems

Distributed Systems Seminar FS2013

Alexander Grest
agrest@student.ethz.ch

1 INTRODUCTION

From school and university projects, to starting a business or running a business, all these tasks have been shown to be done more effectively by a group of people, rather than by an individual alone. Collaboration has been described as the driving factor behind human evolution [9]. It reduces the amount of time needed for task completion and improves the result by enabling team members to share ideas and learn from each others mistakes.

Humans have a long history of using tools to enable or improve collaboration. The telephone is a classical example. It enables collaboration between two individuals that are not physically present in the same room. The flip chart is another example. It is used in the typical meeting room for graphical support.

As we make progress and new technologies emerge, there is also potential for new collaboration tools. This seminar report focuses on the design, development and and evaluation of such tools that enable or support humans to collaboratively solve a task.

2 TELEPRESENCE

Telepresence is a revolutionary visual collaboration technology. It creates the illusion of physical presence of a person that is potentially located far away. Ideally, telepresence is indistinguishable from actual physical presence, thus enabling two individuals to collaborate as if there were physically present in the same room.

Face-to-face relationships (or the illusion thereof) are of crucial importance to business. Traditionally business has relied on air travel to bring together team members, sales people with clients, etc. Unfortunately, air travel is expensive (and sometimes annoying). Apart from the air fare, costs appear for the lost productivity of being inaccessible to colleagues and away from information and corporate resources. Moreover, opportunity costs of doing something else while being in an air liner or jet lagged arise.

As business continues to globalize, the ability to hold cost-efficient meetings with remote participants at a moment's notice will be crucial and thus make telepresence a key technology, with the potential of exponential growth.

2.1 Contemporary Videoconferencing

Commercial videoconferencing has been available since 1964, when AT&T installed its earliest Picturephone units (depicted in Figure 1) in booths that were set up in New York, Washington D.C. and Chicago [11]. The system was the result of decade long research at Bell labs and transmitted uncompressed video over multiple telephone lines. Poor video quality and high prices limited its success and it was closed in 1968.

Figure 1: A 1969 AT&T videophone [12].

Videoconferencing has steadily improved in capability and functionality since then, but users haven't fully embraced it. Due to issues like tiny remote participants, jerky motion and poor audio, most people still prefer face-to-face meetings. Therefore, traditional videoconferencing could not establish itself as an alternative to an in-person meeting.

Most globally operating companies nowadays have so-called ”telepresence systems” deployed in their office building. These system are often installed in a dedicated room with studio quality lightning and acoustics. Multiple large high-definition monitors are mounted on the wall and show remote participants. They generally improve over traditional videoconferencing systems by offering life-size participants, fluid motion, accurate flesh-tones, etc [6]. But they still leave something to be desired - the user must suspend disbelieve to some degree. As they only offer a 2D video feed, they posses inherently limited expressiveness.

2.2 Cave automatic virtual environment (CAVE)

The first step of creating a realistic telepresence experience is placing the user in an immersive virtual reality environment. This allows to create the illusion of being located in a virtual meeting room, seeing other users, etc. An immersive virtual reality environment minimizes the degree to which the user must suspend disbelieve. In other words, it provides a good visual simulation of the virtual reality.

The Cave automatic virtual environment belongs to the most popular. CAVE is a cube-like structure with screen faces surrounding an user. It is coupled with a head-tracking device, making sure the correct perspective appears on the screens as

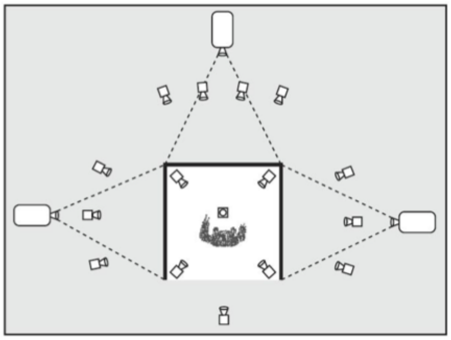

Figure 2: Illustration of the blue-c set-up. The user is located in a three-sided cube-like structure [2].

the user moves within the bounds of CAVE.

2.3 blue-c

To create a telepresence experience, the immersive virtual reality environment must be combined with 3D scene acquisition, allowing the seamlessly and realistically integrate remotely located users into the synthesized virtual world. This poses great technically challenges. In particular, the problem of simultaneous and bidirectional image projection and acquisition is tough.

blue-c is an ETH project published in 2003 that offers a new solution to this problem [2]. Beside being rather old, it is still of interest because it contains important ideas and influenced many present systems. 20 researchers from such fields as computer graphics, vision, communication engineering, mechanical engineering, and physics build it over a time frame of three years.

One of the main ideas behind blue-c is to use time multiplexing between image acquisition and image projection. The use of an actively shuttered projection screen that can be switched from an opaque state (for projection) to a transparent state (for acquisition) allows video cameras to ”see through walls”, leaving much more freedom in optimizing camera positions with regard to 3D reconstruction of the scene.

2.3.1 Setup

The user is located in a three-sided cube-like structure as depicted in Figure 2. Each wall of the cube-like structure consists of three panels that contain a phase dispersed liquid crystal layer. This layer can be electrically switched from an opaque state to a transparent state. Each of those panel costed approximately USD 3’300, making blue-c an expensive project.

Three twin LCD projectors with additional LC shutters are utilized to generate a CAVE-like immersive display with active stereo. The projectors project images from the rear and are synchronously shuttered along with the screens. The user wears wireless shutter glasses that were modified to produce an opaque phase during image acquisition.

11 of the 16 video cameras used for image acquisition are placed outside of the cube-like structure. The five remaining cameras are attached to the four upper corners of the screen and to the ceiling.

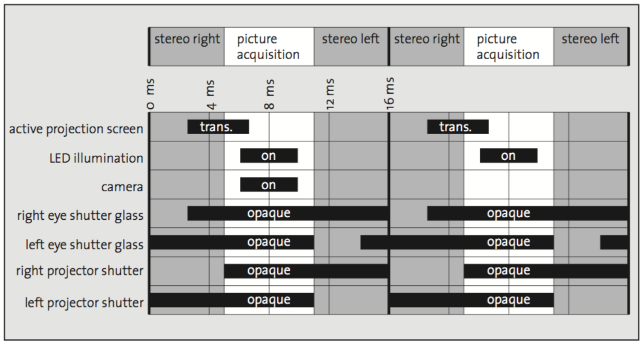

Figure 3: Timing diagram for all actively synchronized hardware components in blue-c [2].

2.3.2 Image Acquisition

The video frame is acquired in the small time slot between the projection for the left and right eye. As depicted in Figure 3, the projection screen is switched to transparent during a short time slot. During image acquisition, the shutters for both the user’s eyes and for the projectors are set to opaque.

The user is actively illuminated during image acquisition. This is necessary for proper background separation, texture acquisition and 3D reconstruction. Illumination is done using a total of 10’000 white LEDs grouped into 32 clusters. The light flashes are not noticed by the user because the shutters for both his eyes are opaque.

2.3.3 3D Video Processing

A 3D video representation of the user is computed in real time on a Linux PC cluster. In a first step, foreground pixels are extracted from the background. This tasked is simplified by placing a blue anti-reflective curtain around the cube-like structure. In a next step, a multi-scale silhouette of the user is computed. Finally, these silhouettes are used as input to extract a 3D point-based video representation.

The 3D point-based video representation is streamed to a remote location using a specially designed update scheme. For each point, a specific operator such as insert, update or delete is transmitted. At the remote site, a flat data representation is dynamically updated and displayed using point-based rendering.

2.4 DepthCube

One of the main ideas behind blue-c was to use an actively shuttered projection screen. This idea can also be used to build a volumetric display system.

The DepthCube [10] is a multi-panar volumetric display system that consists of two main components: a high-speed DLP video projector and a multiplanar optical element composed of a stack of 20 LC shutters (see Figure 4). The high-speed video projector projects a sequence of slices of the 3D image into the optical element. At any point in time, all of the LC shutters are transparent except for one that halts the currently projected slice at the proper depth. Multiplanar anti-aliasing algorithms are used to create continuous appearing 3D im-

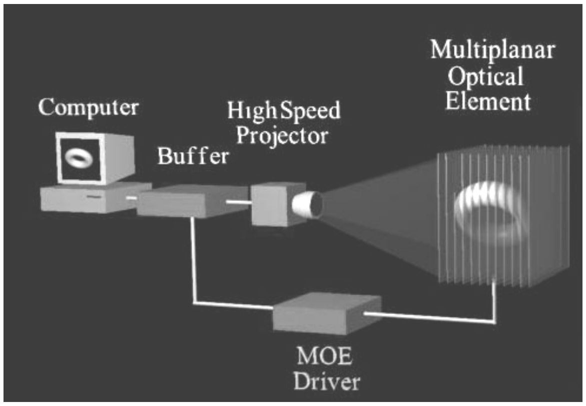

Figure 4: Schematic diagram of the DepthCube [10].

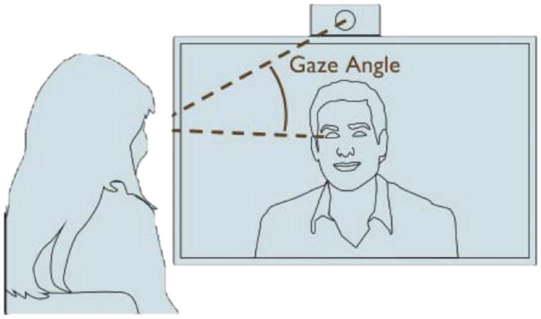

Figure 5: It is impossible to maintain eye contact with traditional videoconferencing systems [6].

ages.

2.5 Eye Contact in one-to-many Teleconferencing

Eye contact is a non-verbal ability to communicate, and it often equals to our ability to verbally express a thought [1]. Maintaining eye contact in a conversation presents an air of confidence. On the other hand, failing to maintain eye contact is often construed as a signal of untruthfulness.

As depicted in Figure 5, it is not possible to maintain eye contact with traditional videoconferencing systems. This problem becomes evident after a video conference using a tool such as Skype on a laptop. Intuitively, people focus on the eye's of the remote participant on the screen and not in the camera. But this appears as if one was looking away for the remote participant. In a video conference with multiple participants, the situation is particularly unsatisfying: when a participant looks into the camera, everyone seeing their video stream sees the participant looking toward them. When the participant looks away from the camera, no one sees the participant looking at them.

We describe next a system for one-to-many teleconferencing that accurately reproduces gaze direction and eye contact [4]. In one-to-many teleconferencing, a single remote participant attends a larger meeting with and audience of local participants.

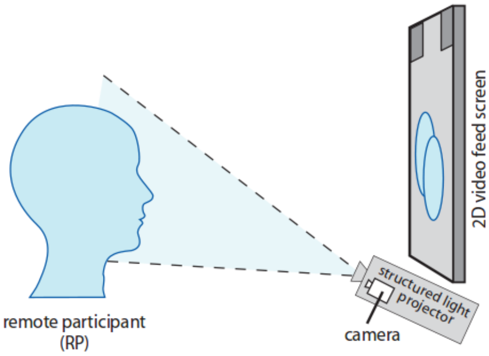

Figure 6: The real-time 3D Face Scanner [4].

In this system, the face of the remote participant is three-dimensionally scanned and then shown on a 3D display to the audience. The remote participant sees an angularly correct 2D view of the audience on a large screen. The system has two main components: A real-time 3D face scanner and an autostereoscopic 3D display.

2.5.1 *Real-time 3D Face Scanner*

The face of the remote participant is scanned at 30Hz using a structured light scanning system as depicted in Figure 6. Four repeating patterns are projected onto the face: Two 90-degree phase-shifted sinusoid patterns and two patterns that light the frame on the left and right. A monochrome camera captures frames at 120 Hz. With the help of a phase-unwrapping algorithm a depth map image for the face of the remote participant is created. This is transmitted along with facial texture images.

2.5.2 *Autostereoscopic 3D Display*

Figure 7 shows the 3D display apparatus. The main part of the display are two brushed aluminium display surfaces spinning at 900 rotations per minute. Those display surfaces are concave and reflect light in a certain direction depending on the current rotation angle. The two-sided shape provides two passes of a display surface to each viewer per full rotation. A high speed monochrome projector projects 1-bit (black or white) frames at 4,320 frames per second using a specially encoded DVI video signal. Effectively, the display projects seventy-three unique views of the scene across a 180 field of view.

To render vertical perspective accurately to multiple viewers, the viewers position is tracked using a face detection algorithm. Because of the concave shape of the display surfaces, each projector frame can be assumed to address just one audience member. Based on the current rotation angle, the tracked audience member who is closest to the central reflected ray of the display surface is determined. Then the face of the remote participant is rendered corresponding to the height and depth of the closest audience member.

3 MULTI-USER STEREOSCOPIC DISPLAY

Multi-user displays enable co-located collaborative work in shared virtual environments. But contemporary 3D television sets and 3D cinemas display only a single stereoscopic image stream, and thus there is only a single location from which a person observes a perspectively correct view of the

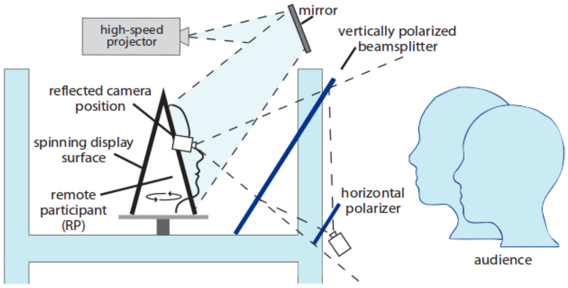

Figure 7: The 3D display apparatus showing the two-sided display surface and the high-speed video projector [4].

displayed scenes [5]. All of the other spectators perceive the 3D scene distorted. This may not matter when watching a Hollywood blockbuster, but it significantly hampers with the acceptance of 3D technology in many other application areas. To avoid distorted perspectives, each person must be provided with individual stereoscopic image pairs that are rendered for the exact position of the user in front of the display.

C1x6 is a projection-bases stereoscopic display for up to 6 users [5]. It consists of six customized DLP projectors, each of which projects six fast time-sequential images in one of the primary colors. By differently polarizing the light output of the first set of three single color projectors (red, green, blue) than those of the second set, twelve separable full-color images are projected onto a screen.

3.0.3 Projector array

The DLP projector array is capable of displaying twelve full color image streams at 360 Hz, resulting in 60 Hz per user. To achieve that, 6 DLP projectors were modified. In particular, the color wheel was removed, allowing to project three monochrome time-sequential views instead of the different primary colors of a single view. Furthermore, the projectors were modified to accept a 120 Hz input signal. This was necessary because although most DLP projectors rotate the color wheel at least twice per video frame and are thus effectively running at 120 Hz, they do not accep a 120 Hz input signal. In order to achieve 12 views, polarization is used in combination with shuttering.

3.0.4 Shutter glasses

Participants must wear shutter glasses in order for this approach to work. The shutter glasses have to work at 360 Hz and the left and right eyes need to be differently polarized.

Because of the high frame rate, regular off-the-shelve liquid crystal shutters are not suitable for this system. They have asymmetric opening and closing properties and open very slowly (longer than 2 ms). One approach to circumvent this problem would be to use ferro-electric shutters with symmetric opening and closing times of less than 0.1 ms. However, ferro-electric shutters are much more expensive than liquid crystal shutters and are very fragile.

As an alternative to ferro-electric shutters, custom shutter glasses that consist of two layers of differently configured liquid crystal shutters were build. The first layer is a regularly cross-polarized liquid shutter (NW), which is transparent if no voltage is applied and closes quickly. The second layer has equally oriented polarization filters on both sides and therefore is opaque (NB) if no voltage is applied and opens quickly. This combination allows to open and close the shutter quickly.

4 TILED DISPLAY WALLS

We certainly live in the Information Age. More and more data is collected on a daily basis and has to be interpreted. Often this is done by bringing people together into teams for collaboratively solving problems. But as our cognitive ability is finite, we need the help of tools in order to be able to deal with such an amount of data. For example, visualizations greatly facilitate data handling.

Consider the typical meeting room. In the past, it was equipped with a flip chart with large sheets of paper around. One could tear off a sheet of paper and mount it somewhere on the wall. After some time, there might be multiple sheets of paper hanging on the wall that could be readily consulted. Today's typical meeting room is equipped with a single projector, severely limiting the amount of information that can be displayed at any given point in time.

Classrooms suffer under a similar problem. Old classrooms were equipped with multiple black boards, often wrapping around the whole room. This enabled the professor to write and not erase during the entire lecture. In today's version, only a single projected image is available. Faced with this limitation, many professors developed a habit of flipping back and forth through the slide deck.

However, projectors have become quite affordable in recent years. The motivation behind tiled display walls is to combine multiple projectors to form a single large, high-resolution, wall-sized display surface, thus significantly increasing the area available for collaboration [7].

We present a rear-projected tiled display that can scale with multiple projections, users and applications [8]. The display is built by a distributed network of plug-and-play projectors (PPPs). A plug-and-play projector consists of a projector, a camera and a computation unit.

The display is created by aligning multiple plug-and-play projectors in a rectangular array as depicted in Figure 8. The PPPs are aligned casually and overlap with their neighbours. Using visual communication via the cameras, a plug-and-play projector is able to detect its neighbours whenever its camera perceives another plug-and-play projector in the overlapping area. Moreover, the plug-and-play projectors are also in a IP multicast group to communicate.

4.0.5 Neighbour detection

The PPPs are initially unaware of the configuration of the array that they are arranged in. Using a distributed registration technique each PPP can discover its neighbour, the total number of projectors in the display and its own coordinates in the array of PPPs.

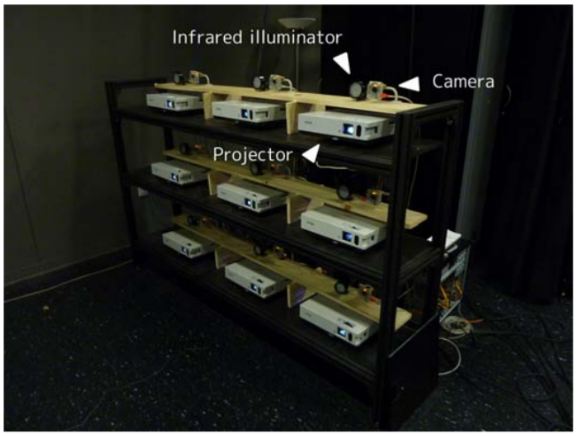

Figure 8: The display is created by aligning multiple plug-and-play projectors in a rectangular array [8]

.

When a plug-and-play projector is turned on, it projects four QR codes around the center. The QR codes contain the IP address and port of the PPP, as well as location of the QR Code (2D coordinates of its top left corner in the projector coordinate system). Since the camera of each PPP sees more than their own display, they see the neighbors QR code along with their own.

Each PPP detects the QR codes from its neighbors to find out which of the left, right, bottom and top neighbors exist and creates the local connectivity graph of itself with its neighbors. Next, they decode the QR code to find out the exact IP-address of each of their neighbors. Finally they broadcast the location of each of their neighbors along with the associated IP-address to all the PPPs. When each PPP receives this information, it augments its local connectivity graph using this information. Therefore each PPP knows the total number of projectors in the display and their configuration in the end.

4.0.6 Geometric registration

A key challenge when building a tiled display is to combine multiple projectors is such a way that the resulting image is correct and seamless. When multiple projectors are set-up casually, they may be misaligned, leading to visible breaks in the image content. One way to solve this problem is to warp the projected image based on feedback from a camera. The mapping between the planar display coordinates and the projector's coordinates is estimated as accurately as possible using feedback from the camera. This mapping can be expressed by a 3 x 3 matrix H, often called a planar homography (see Figure 9). The inverse mapping is then applied to the image in software. Therefore, the end result is a correct and seamless image.

For that purpose, all QR codes are augmented with blobs embedded in the quite zone. In a first step, each PPP detects the blobs in the quite zone of the QR codes it projects itself. This allows it to determine the mapping between the coordinates of its camera and its projector, also called the self-homography. In a next step, the PPP detects the homographies with its adjacent projectors, using blobs detected in their QR codes. Finally, it concatenates its self-homography with the homography with the adjacent projector to create the local homography $H_{i \to j}$.

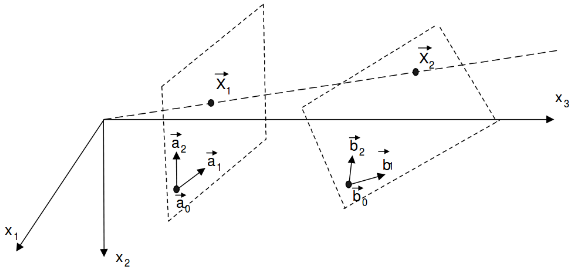

Figure 9: The mapping between the planar display coordinates and the projector's coordinates can be expressed by a 3x3 matrix called planar homography [3].

The geometric registration algorithm starts from a reference projector which is considered as the only registered PPP initially. In each subsequent step S, PPPs with Manhattan distance S from the reference join the set of registered PPPs by aligning themselves with the PPPs who joined the registered display in step S-1. The process stops when all the projectors belong to the set of registered projectors.

4.0.7 Gesture Management

The tiled display wall is not only able to display an image, but it can also react to gestures. A gesture is defined as a sequence of actions that are detected by the system. Gesture management happens completely distributed, as there is no centralized server. Each PPP is responsible for managing the actions that occur within its domain. When an action occurs in the overlapping area of multiple PPPs, only the PPP with the smallest ID handles the action.

If an action is close temporally and spatially to another action, it is assumed to belong to the same gesture. If an action is temporally or spatially far away, it is considered the commencement of a new gesture. This can happen when two users are interacting simultaneously with the display. The end of a gesture is detected by a timeout.

When a PPP is tracking the gesture and finds it is to move into the neighborhood of an adjacent PPP, it sends an anticipatory message to notify the neighboring PPP about a gesture coming its way. This message contains all the necessary data to handle the continuation of a gesture. Later, when the adjacent PPP detects an action in the neighbourhood of the location predicted by an anticipatory message, it identifies the action as part of a continuing gesture.

4.0.8 Reaction Management

Reaction management involves processing actions and reacting with an particular event. The reaction to a particular gesture is application specific. However, the set of PPPs that have to react to an action may be bigger than the set of PPPs that registered the action. For example, in a map application, one can move the map by a sweeping gesture that spans just a few PPPs, but all PPPs must react and change which part of the map they display. To solve this problem, messages are broadcasted to all PPPs.

5 REFERENCES

1. Jan Castagnaro. Communication skills: The importance of eye contact. http://voices.yahoo.com/communication-skills-importance-eye-contact-631932.html, November 2007.

2. Markus Gross, Stephan Würmlin, Martin Naef, Edouard Lamboray, Christian Spagno, Andreas Kunz, Esther Koller-Meier, Tomas Svoboda, Luc Van Gool, Silke Lang, et al. blue-c: a spatially immersive display and 3d video portal for telepresence. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 819–827. ACM, 2003.

3. Allan Jepson. Planar homographies. http://www.cs.toronto.edu/~jepson/csc2503/tutorials/homography.pdf, 2013.

4. Andrew Jones, Magnus Lang, Graham Fyffe, Xueming Yu, Jay Busch, Ian McDowall, Mark Bolas, and Paul Debevec. Achieving eye contact in a one-to-many 3d video teleconferencing system. *ACM Transactions on Graphics (TOG)*, 28(3):64, 2009.

5. Alexander Kulik, André Kunert, Stephan Beck, Roman Reichel, Roland Blach, Armin Zink, and Bernd Froehlich. C1x6: a stereoscopic six-user display for co-located collaboration in shared virtual environments. In *ACM Transactions on Graphics (TOG)*, volume 30, page 188. ACM, 2011.

6. Howard S Lichtman. Telepresence, effective visual collaboration and the future of global business at the speed of light. *HPL, Human Productivity Lab Magazine*, 2006.

7. Aditi Majumder and Michael S Brown. *Practical multi-projector display design*. AK Peters, Ltd., 2007.

8. Pablo Roman, Maxim Lazarov, and Aditi Majumder. A scalable distributed paradigm for multi-user interaction with tiled rear projection display walls. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1623–1632, 2010.

9. Keith Sawyer. Collaboration drives human evolution. http://keithsawyer.wordpress.com/2011/04/05/collaboration-drives-human-evolution, April 2011.

10. Alan Sullivan. 58.3: A solid-state multi-planar volumetric display. In *SID Symposium Digest of Technical Papers*, volume 34, pages 1531–1533. Wiley Online Library, 2003.

11. Wikipedia. Videoconferencing — Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Videoconferencing, 2013.

12. Wikipedia. Videophone — Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Videophone, 2013.