Development of an OTA (Over the Air) Mobile Learning Telepresence Platform

by

Kyle Galvin
HBSc Computer Science, Lakehead University, 2013

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MSC COMPUTER SCIENCE
in the department of Computer Science

Supervisory Committee

Development of an OTA (Over the Air) Mobile Learning Telepresence Platform

by

Kyle Galvin

HBSc Computer Science, Lakehead University, 2013

Supervisory Committee

Jinan Fiaidhi and Sabah Mohammed
Supervisor
Co-Supervisor or Departmental Member
Departmental Member
Outside Member

Contents

Chapter 1

Lightweight Telepresence Technologies

Microprocessors have shaped the world over the last century. Reducing in size over time exponentially, we are now able to achieve things that would have been unimaginable in the past. We can squeeze more bits per volume, transport more information and crunch more data each second than ever before. With this explosion of portability and connectivity comes a renaissance of technological growth that is unfolding before our eyes.

Density of information and computation as well as the speed of communication are at the core of modern digital technology, yet focusing on these features displays a very hands-on 'white box' approach. There is also much to be learned with respect to the interaction between digital components and their environments, which could be considered more of an external 'black box' style description. The interface and sensors a device supplies for others to interact with is just as important as the computational and communicative abilities the device has internally to process the environment around it.

By extrapolating on current computational growth trends, we can easily imagine the capabilities we will soon wield while developing applications to improve our everyday lives. By studying these applications (both mundane and whimsical alike) we are likely to find many exciting ideas which are attainable much more immediately than they first appeared as well as many which are just around the bend.

Arthur C. Clarke once wrote "Any sufficiently advanced technology is indistinguishable from magic", and I am inclined to agree. Indeed many amazing discoveries can find roots in sci-fi and futuristic predictions which push the boundaries of our collective knowledge and explore the potential and logical conclusions of current technological progress. The most recent ideas which are moving from science fiction to science fact are telepresence and augmented reality.

[CFB$^+$ar]

To introduce these ideas, I will start by defining telepresence:

> ”Telepresence systems provide a human operator with the feeling of actual presence in a remote environment, the target environment. The feeling of presence is achieved by visual and acoustic sensory information recorded from the target environment and presented to the user on an immersive display.”
>
> [AH11]

Augmented Reality (which is a branch of virtual reality (VR) technology) is a multifaceted topic that not only lends itself extremely well to telepresence applications but extends far beyond it. Here we will focus primarily on the intersection of the two ideas despite each existing independently of the other. Applications in this domain will not only use telepresence technologies to unite two remote locations, but they will use augmented reality techniques in order to make up for the loss in fidelity that occurs when we replace the user's natural environment with a virtual representation. Three sensory obstacles causing loss of information have been identified:

1. The detection of sensory signals
2. The feedback of real-time sensory information to the operator
3. The presentation of this information in a form that can be easily detected, processed by the brain as a reflex action and responded to, since an excessive need for thought would detract from performance of the primary task.

[CWKG96]

It is important to note that this lack of fidelity is a huge obstacle to telepresence adoption. Why would I send an e-mail to somebody if I know I can call them directly? Why would I call them if they were available in person? In order to be useful, telepresence needs to solve a direct need. Typically, the following circumstances have allowed telepresence to improve our ability to interact, especially with the addition of telepresence robots:

1. The robot must operate in environments that are hazardous to human health.
2. The robot must operate at a scale that is much smaller or larger than the human size and scale.
3. The robot must operate in a location where it would be too costly for the human to be present (in terms of budget, timing requirements, and human safety).

[AWZ98]

1.1 Emerging Mobile Technologies

As microchip density increases, so does the mobility of computational and processing devices. While PDA and hand held gaming devices have been around for decades, the advance of cellular networks which allow for on-the-go personal telecommunications and widely dispersed access to internet services has really driven the shape and design of the current generation of mobile devices.

Smart phones and telecommunications aren't the only technology in this arena, but they are certainly the largest and most influential. Other devices to consider when discussing telepresence devices are lightweight microprocessors and system on a chip designs. These devices can allow industry and hobbiests alike to create a wide array of telepresence hardware that is capable of interacting with the environment around it on another's behalf. In this case, we are now less bound by strict computational limits and are now merely bound by the sensors, motors, and analog/digital conversions available to read from (and interact with) the environment around us.

When we combine our new-found freedom to invent any sort of sensory device with our fully connected and always online 'internet of things', we can begin to explore and create all sorts of ideas that were inaccessible to the real-world and thus bound to the realm of fiction, futurism, and sci-fi.

1.1.1 Microcontrollers & customized System on a Chip Components

Microcontrollers have become the de-facto platform for lightweight, mobile, and miniaturized devices. Much of the power of microcontrollers has been achieved by clever specializations and optimizations of the CPU unit. By diverging from the Desktop model where raw power takes precedence over power consumption, there are now a variety of architectures and specialized components that are well-suited to mobility. In fact there is now a wide spectrum of hardware ranging from high performance to low power consumption. With server and desktop hardware on one end of the spectrum, we have recently expanded the power efficient end of the spectrum with a range of miniaturized ARM general-purpose CPUs capable of of supporting a general-purpose operating system and related peripheral components in an extremely small enclosure. To continue down the spectrum we depart from a traditional operating system and move towards programmable integrated circuits (PIC) and pure-hardware components which perform more specialized tasks using even less space and power.

Because of this balance between portability and power, it is important to keep in mind the practical limitations on the applications a device can support. While a computationally demanding task such as image processing, we are unlikely to get satisfactory results with a PIC controller, yet the current generation of ARM controllers which are recently emerging are just beginning to practically handle these tasks, and this can be echoed by the libraries and toolkits developers have ported to the platforms. For example, there are now openCV ports on both the android and IOS platforms. This is a testament to their increasingly pervasive capabilities.

An interesting idea to consider is delegation of the heavy processing to a more capable device. However, this strategy simply moves the problem from the computational I/O boundaries of the device towards the networking I/O boundaries. To achieve practical results, we can use a hybrid approach by allowing the device to pre-process the data (assuming such a pre-processing can reduce the size of the raw data, preferably an order of magnitude or more) before sending the derived result over the network to a more capable device.

Moving from our internal capabilities towards our external interactivity, we are capable of hooking up a wide array of sensors, displays, and feedback devices. From motors to spectroscopy sensors to audio capture and processing (and many, many more) the possibilities are limited only by the fidelity/accuracy of the sensors around us and our own creativity.

It begins to become clear that tele-robotics is not an idea that would lend itself to any situation. The reduction in fidelity we introduce by using a proxy or avatar robot is severe enough that we have only found successful use-cases in well-defined environments such as laboratories and factories where there are little or no unknown variables. Situations where there are people in close proximity to the robots are often too dangerous for the technology to be applied, and furthermore the decrease in the awareness a user would have with their remote surroundings makes the potential for interaction very limited.

[CWKG96, EAM12]

The idea of presence fidelity can be considered a continuum. We can consider verbal descriptions and printed material on the low-fidelity end of the spectrum, while actual presence would be on the highest end. [AWZ98]

1.1.2 Cellular Phones & Mobile Collaboration

When discussing mobile telepresence, the recent emergence of smart phones is undeniably the most important event to occur to date. In a learning environment they are indispensable for personal organization, time management, and real-time updates. It has been stated by others that they can now allow students to "...be informed of all necessary notices, assignment deadlines and supervisor advices during their busy schedule". In fact, "Mobile education is defined ... as any service or facility that supplies a learner with general electronic information and educational content that aids in acquisition of knowledge regardless of location and time." [Hos07, Bin11]

1.2 Telepresence & Real-Time Communications

The need to provide high fidelity information in real-time is the defining challenge of telepresence systems. Current telepresence infrastructure comes with restrictive limitations due to network latency and available throughput. These challenges have been met with several techniques such as lowering the fidelity rate transmitted, compressing the data in transit, and prioritizing which data is most relevant to the situation.

1.2.1 Audio/Video Compression

Within the context of mobile devices we need to consider not only the computational complexity and bandwidth consumption of video streaming, but the power consumption of the encoder as well. [AKK09]

1.2.2 Cellular Network Bandwidth Flow & Optimization

Network traffic can be prioritized using QoS (Quality of Service) classification. By prioritizing traffic types into real-time and non-real-time categories we can decrease the average latency on timing-critical services. By extending these concepts and including cooperative game theory strategies (by exploring Nash Bargaining Systems). A game theory strategy can be broken down into three components: Players, Strategies, and Interactions (or game utilities). If we can determine a metric for success in the context of each player, than we can create strategies which each player can use to interact with the system in a way which optimizes the collective success of all the players. Applied to bandwidth optimization, each network device is a player which utilizes a strategy for sharing the limited resources of the network.

The trend among wireless networks is an increased number of cells over smaller and smaller areas serving users. By reducing this cell size, we introduce an increased number of hand offs when the user is mobile. When a user passes from the range of one operator to the next, this hand-off should not interrupt the user's communication. Because of this, our QoS strategies should include a percentage of bandwidth reserved for hand-off services.

With many different use-cases and bandwidth categories our goal is to create a strategy which optimizes the usability of all devices on the network. By ordering our bandwidth categories by highest to lowest priority we can allocate each users traffic into their respective categories while also dynamically adjusting the maximum flow of each category to reflect (as well as prioritize) the immediate demand in real-time [Kim11, KV04, LYC04]

1.2.3 Security

As telepresence technology matures, we will find ourselves relying on it in many new ways. The amount of information each device will be capable of collecting from our every day lives underscores the importance of security policies and data management. We may, for instance, be comfortable sharing a live video stream with a friend, yet that same video stream may reveal information about our location or activities which we would not want to share with others. The need for confidentiality is important to both individuals as well as businesses and institutions.

To address these circumstances, access policies must be put into place. The use-cases for these access policies can be quite diverse, and any proposed system must properly address them all in order to be widely accepted. For instance, if I have a telepresence system at home it would be reasonable to allow my own cellular device to view my house at any time. However, another user trying to view my home environment should be denied access unless I have explicitly granted it to them. Once their session has ended, their access should again be revoked.

In access control systems, the concept of groups is not new. However, in telepresence systems groups apply to more than just users, they apply to the devices as well. If I have two or more telepresence devices in my home or an organization, I should be able to facilitate another user to view from either environment (and even switch between seamlessly) for the duration of their session. In this way, the remote user is less constrained by the software and can have the flexibility of interacting wherever the hardware allows.

So far we have used the term 'session' in two different scenarios, but have left the details a little vague. A session may not be limited to a series of sequential actions with one agent as in the traditional sense of the word. A session can move beyond the hardware it was created on and travel across several telepresence devices in a group.

The access control should be aware of other variables, such as time and geo-location. This allows for much more dynamic and easily defined roles.

Now that we've provided a background on our improved access control model, we can come up with some use-cases to describe how this control system should work.

1. Trigger Duration 6:00PM to 8:00PM - Allow All in User Group: Family Access to Device Group: MyHousehold

2. Trigger Location Office And Trigger Duration 9:00AM to 5:00PM - Allow All in User Group: Peers Access to Device MyPhone

3. Trigger Duration 5:00PM to 10:00PM - Allow All in User Group: Friends Request Device Group: MyHousehold

The first item states that anyone in the 'family' group can access my household telepresence devices freely for a short duration after dinner. The second states that my work peers can instantly appear on my phone at their own discretion when I am at the office during working hours. The third states that friends can call me between 5 and 10 PM, but I must answer for them to connect.

1.2.4 Privacy

Encrypting live video streams is computationally expensive. Because of this, there is a trade off between bit rate and security that must be addressed. On one hand, we want to be reasonably confident that our communications are not being intercepted. On the other hand, we want the highest quality video and the longest use of our mobile device batteries. There is much to be gained by exploring this continuum, as well as the possibility of selective encryption to hide only the most sensitive data. [Feh13]

With the advent of Location Based Services (LBS) it is becoming increasingly difficult to control the extent in which a users locational information is used. Mix networks use short-term psudo-anonymous names to mask the identities of participants. With this mechanism, it becomes much more difficult for an adversary to correlate which actions in the system were performed by which users of the system. Since mix networks leverage multiple proxy servers to achieve anonymity, this is going to cause a tremendous spike in latency and bandwidth utilization. [LL12, FRF$^+$07, PHE02]

The ubiquity of cameras and CCTV devices along with the proliferation of digital signal processors and facial tracking techniques are making the collection and centralization of user activities increasingly simple. Proposed countermeasures include using DSP techniques to selectively scramble identifying information such as license plates and faces to preserve user privacy, however these techniques rely on the cooperation of the surveillance administrators as well as the addition of costly components to the surveillance system. Currently the only countermeasure users have to prevent privacy intrusion is to opt-out of an otherwise useful service. In many cases such as public CCTV opting out is difficult if not entirely impossible. [Cav07, HR13]

Efforts to create a system which allows for users and even objects (in the case of license plates and sensitive documents) to register a preference for privacy and be 'scrubbed' from any published video have been explored. This does, however, require the cooperation of video producers and publishers by running their video through a central 'privacy scrubbing' service in a process that is comparable to a telemarketing 'do-not-call' list. While there is no technical reason preventing people from ignoring the video scrubbing process, publishers could be subjected to social and legal pressure to respect the privacy preference of others.

Ironically, the need for users to identify their privacy preference as well as their timestamped location into the central database dictates that they must give up their locational privacy in order to preserve their video privacy. [Bra05]

I propose that with additional countermeasures, the location of a user wishing to remain anonymous can be effectively verified while their identity remains a secret, provided several non-anonymous users whom the central privacy authority trusts each independently verifies the locational claim of the anonymous user via radio signal triangulation. In this way, the anonymous user can verify their location at a particular time without revealing any digital fingerprint or certificate to other nodes in the network including the central database.

1.2.5 Cloud Based Assisted Technologies

With the advance of internet connectivity, it is rare that modern mobile cellular devices are off line. If we are constantly in communication by means of a global (universal) IP address, It follows that we can achieve two things which we previously could not.

First, we can synchronize our local data with the data of others in real-time. This lends itself to instant news aggregation, social media, e-mail, instant messaging, and even VoIP technologies. This is not in itself extremely surprising as these features have existed in the scope of the desktop application since the dawn of always on high-speed broadband connections, but the ability to bring this to a widely distributed array of mobile devices brings the connectivity of our society (and the speed of information travel as a result) to an all new level.

Second, we can now outsource services which are not desired or capable of running on the mobile device to another computer. While this is typically (as of yet) a cloud service provider's machine, it is reasonable to consider that over time software will develop which allows users to host their own content from a simple always-on home computer which serves as a personal hub for content including but not limited to public social media, geo-secure proxy access, private home surveillance, and data storage. By using strict private/public tags, and 'group' authentication on a server's data as well as RSS-style content aggregators, it should be possible to design a decentralized 'home cloud' service which can serve many useful purposes to a mobile user in the field.

1.3 Digital Identification and Modeling

As our current trend of miniaturized mobile networked devices continues, the ability to stream high definition real-time media from many simple cellular devices is beginning to unfold. With higher quality cameras emerging in consumer devices as well as faster network connectivity emerging in the forms of 3G and 4G telecommunication services it is completely feasible (even today) to use two phones to achieve a long distance video conversation.

It is interesting to consider that if we can stream enough information between two points to effectively allow a user to 'see' and 'hear' what is in another location, then theoretically speaking we must have processed and moved (via the VoIP phone system) enough information about the two locations to effectively understand (to the same degree the two communicating people were capable) what is happening at each of the locations. We as humans don't think about the image processing we do on a daily basis. We take for granted the fact that, somewhere between the rods and cones of our eyes and the high-level understanding we have of our surroundings, a lot of information was processed, stored, and acted upon.

While the topic of Artificial Intelligence and Computer Vision is far from bridging this gap, impressive results have been found by re-thinking the ways we can facilitate enhanced modelling and identification techniques.

1.3.1 Bar Codes & QR Codes

QR codes (or two-dimensional bar codes) can reference nearly anything. Ranging in size from 25x25 to 177x177, they are most often used to redirect a user to a URL containing anything from videos to product info to social media content [ALYY11] Basic compression is done using run-length coding, where sequences of identical values are replaced with a single instance of that value followed by the repetition count. [ALYY11]

1.3.2 RFID; NFC

[PJ09]

1.3.3 Real-Time Digital Modeling

Triangulation & Accelerometer based Orientated Positioning

If many radio-enabled devices (whether via wifi, bluetooth, gps, or other means) are within communication range, it is possible to use time-synchronized signals to triangulate the positions of the devices with respect to each other. If any of these devices were equipped with a GPS, it could communicate this information and allow neighbors to take their local position data (respective to each other) and place them globally.

If any of these devices were equipped with an accelerometer and a camera, it is theoretically possible to calculate a position and orientation vector for the camera, effectively letting the device give extremely precise descriptions of what region of space is being recorded. This information (spatial meta-data, as well as the raw audio/video data) can be combined with other similar information from the region in order to provide high-fidelity reconstructions of recorded events. This information can be further augmented by other sensory input such as spectroscopes, depth-sensors, and environment monitors.

Time-synchronization techniques are limited by the sample speed of the measuring apparatus' clock. Determining distance by means of the received signal strength do not require a clock, however the distance cannot be deduced from the signal strength alone without a calibration phase which takes into account the variance of each device's radio signal.

Stitching Multiple Images Together

In cases where images are not aligned, unwanted artifacts can be produced. Color and lighting inconsistencies can also be introduced which would create an unbalanced effect. A lack of references and identifiable control points can also make it difficult to correctly position image fragments. [JT08, XP09]

Depth Maps for 3D Imaging

Depth imaging techniques have only previously existed in costly special-purpose applications. With the spread of large-scale production on Time-Of-Flight sensors (specifically, the Microsoft Kinect) the accessibility and spread of these devices has grown considerably. Time-Of-Flight technology involves measuring the delay between sending and receiving an infrared signal. With this information, triangulation techniques can be used to measure the depth between the device and the target.

These devices have been designed with object recognition in mind and are not particularly suited for 3D scanning applications. Low resolution and a large amount of noise are certainly factors when re-purposing these technologies for scanning [CST+13]

1.3.4 Image Recognition and Classification

Image recognition is a complex problem often approached with a neural network model or similar fuzzy categorical organizer. If RGB information is augmented with depth information we can achieve much better results than if we were to rely on RGB information alone.

Vector Quantization

Uncertainty / Fuzzy Logic

Improved Accuracy Through Domain-Specific Environments

telepresence surgery model - live data can be collected, then used to simulate the procedure virtually. This effectively allows us to generate training programs which are extremely accurate within the domain of the live collected data.

Imagine a doctor in front of device operates instrumentation which performs surgery on a remote patient. instrumentation includes a wide variety of I/O (controls and sensory output via microphone, video, and even tactic feedback) Assume access to highly detailed descriptions of our I/O over the duration of many operations (live experience captures) The challenge is to make a virtual model of the operating procedure in which the doctor can interact with a virtual patient in a way which is synonymous with the standard interactions they would encounter with a live patient.

The challenge is considered 'solved' when the doctor cannot differentiate between a live patient telepresence experience and a simulated patient telepresence experience I call this challenge the "Telepresence Turing Test", and it can be applied to any activity or domain in which telepresence can augment.

This challenge has a few interesting unknowns.

How 'synonymous' with live data can we realistically make the experience?

What are the most important factors we need to capture in our training data? What instrumentation can best capture those factors?

From the training data (live experience captures), how can we best create and improve upon a simulation model? [GHJS95]

1.4 Identifying the Major Elements for a Lightweight Mobile Telepresence System

Lightweight mobile telepresence systems is a rapidly-evolving concept. Traditionally we have been bound by heavy and cumbersome desktop hardware, high-latency, and low network throughput. As these barriers have been reduced and removed, we have begun to redefine what it means to be connected.

Always-Available network communication

Real-Time video streaming

Geolocational services

Screen sharing (also useful for remote presentation)

File sharing (p2p for reducing infrastructure and bottle necks)

User/Group management and authentication

1.5 Summary

In conclusion, we are capable of much more than what is currently offered in terms of increasing the fidelity of our telepresence systems. On top of the increase in raw information storage capabilities, improvements in sensors as well as interactive peripherals have reshaped the way we use technology. If current trends continue we will soon find ourselves in a high density and highly distributed network of miniature devices, both as stand alone technologies (such as currently emerging smart phones) as well as embedded into every day consumer objects (as is the case with RFID tagging, QR coded items, and micro-controller enabled electronics). With this emerging paradigm it becomes much easier for computers to identify and process the objects around them, leading the way to many new modelling and digitizing techniques.

Chapter 2

Designing OTA Mobile Telepresence Services

2.1 Authentication Server

The authentication server is in charge of identifying and connecting users. When a user connects to the website, a websocket connection is made between the authentication server and the client. Along with this connection a session ID is provided which we can use to identify the user over multiple transactions. When a user logs in, this session ID is coupled to their user name and marked as an active session. This allows other users to identify the user as online, as the session ID can be decoupled from the user name the moment the socket connection is closed.

2.2 Server Side Command Dispatcher

The authentication server receives data from the user in a 3-tuple form: ¡sender, command, args¿. The sender includes detailed information about the request origin. Not only are the user credentials included, but also which of the client application widgets made the request. This is to allow us to create a return route for the request. The command field is the name of the function we wish to invoke, and the args field is an array of arguments to pass to the function. The commands for authentication have been kept intentionally simple, consisting of operations to sign up, log in, and request contacts.

2.3 Server Side RESTful Model

The authentication server works as a simple RESTful model. RESTful operations are primarily invoked from the server side command dispatcher which houses the business logic of the application. Our RESTful API only consists of four commands: 'create', 'read', 'update', and 'delete'. The user name of the client invoking the request is also included, allowing us to create permissive rules to limit who can access or modify groups of data.

2.4 Client Side Command Dispatcher

The client side command dispatcher is broken down into multiple parts. Commands may be launched from external sources such as the authentication server (eg. Describing bad log in credentials, or supplying a list of online users) or a P2P connection (such as a remote user initiating a call). Commands may also be launched from the clients web GUI.

Each client widget has it's own command API, and the application has naming and routing mechanisms to relay commands from these various sources to the appropriate widget. The widget then handles the command however it pleases.

2.5 Client Side Widget Factory

As the list of client side widgets grow, the widget factory serves as a way to create and register new widgets. The widget factory is in charge of creating unique names for each widget for message routing.

2.6 Client Side Widget Structure

Each widget should be contain a function called 'widget' which will be invoked when the client creates a new widget instance. As input, we pass in any external dependencies the widget requires. This function is in charge of initializing the widget, building the widgets DOM and layout, as well as defining the external API the widget provides.

2.7 P2P Connection

A key component of this application will be the P2P widget. The P2P widget will be in charge of facilitating P2P connections with other users. Through this connection, users will be able to share VoIP communications, screencasts, and share files. The P2P connection will be in charge of connectivity, P2P routing, as well as stopping/starting telepresence calls.

2.8 Goals

In order to improve on current telepresence technologies, we must revisit the main sensory obstacles in telepresence interaction and modify our implementations in a way that reduces such obstacles. To reiterate, these obstacles are:

1. The detection of sensory signals
2. The feedback of real-time sensory information to the operator
3. The presentation of this information in a form that can be easily detected, processed by the brain as a reflex action and responded to, since an excessive need for thought would detract from performance of the primary task.

[CWKG96]

The goal of this project is to create a system which automates and facilitates the above in a helpful and intuitive way via data augmentation of the live stream. These three points will each be addressed individually by it's own software module placed between the telepresence device and the user's device.

A good example would be to recognize and highlight a QR code should one come into the visual frame. The users options could be as simple as 'open URL in browser' or 'ignore'.

A second example would be to alert the user when a presenter has entered the frame after a lengthy absence.

Finally, if a group of two or more people collaborate to make a 'digital audience', tags, notes, and comments can be passed among audience members in real time. For example, if the professor leaves a particularly difficult equation on the board, audience members could be enabled to directly click on the location of the equation within the video in question and communicate their interest in a real time collaborative way. This information can be saved along with the video for the professors review at a later time.

2.9 Assumptions

There are a wide variety of telepresence devices available, each with it's own features and limitations. This project will focus on the most common telepresence software available, which is the modern smart phone. We can assume that the telepresence device will come equipped with a camera and a microphone, as well as a wireless network connection capable of simultaneously transmitting and receiving audio/video data with latency low enough to qualify for 'real-time communication'. If required, benchmarks and data may later be supplied to demonstrate that this is indeed a common feature of modern hardware.

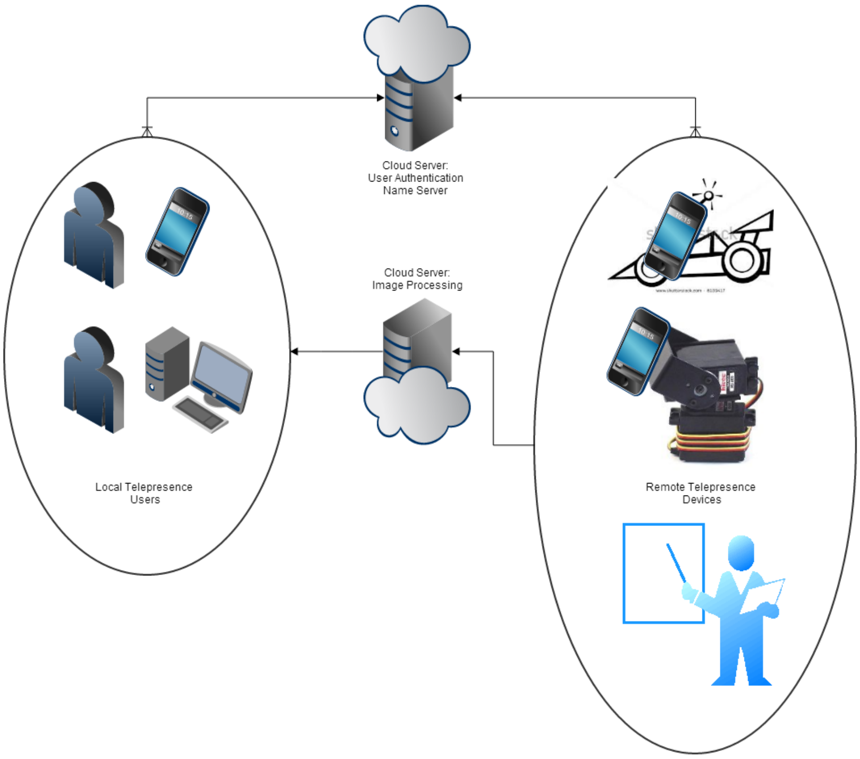
2.10 Design

In order to achieve our goals, the approach I intend to take involves scanning the raw telepresence data between clients in order to parse and extract some of the high level information out of it. If we can automate the detection of useful signals, gestures, queues, and events in the raw audio/video data, we can highlight and emphasize this information to the user. Furthermore, if the signals we parse can be responded with by a minimal subset of actions we are entirely capable of supplying these options in real-time and allowing the user to quickly decide how they choose to respond.

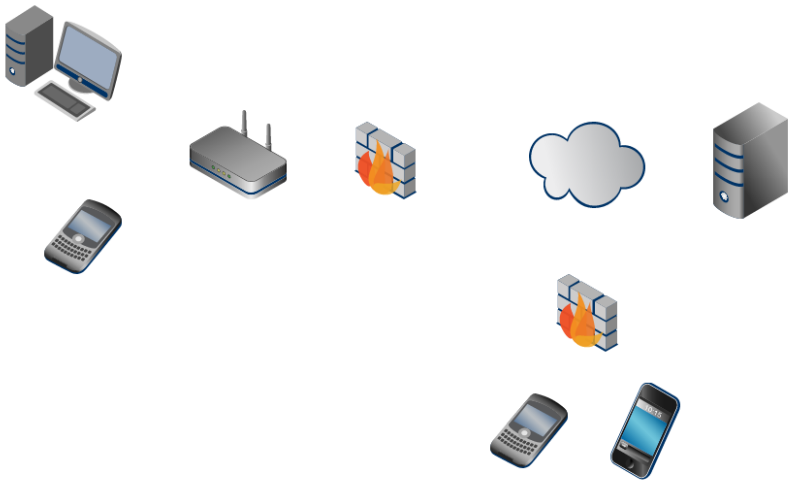
If the mobile telepresence devices which collect and transmit the raw audio/video data are most likely to be modern smart phones, than we would be wise to treat these devices as thin clients in the sense that their purpose is limited to data collection and user interaction. The heavy processing involved with computer vision over large streams of image matrices would quickly overwhelm such devices.

Because of this limitation I plan on deploying a cloud server which intercepts the data streams between clients in order to process the large amounts of raw data. This cloud will also store the streams and act as a database of past transmissions along with all the key meta-data we accumulate through our processing techniques.

2.11 Networking



2.11.1 NAT traversal



(Graph to be completed soon)

2.11.2 I/O Limitations

The hardware needed to deploy such a system requires high bandwidth and low latency. The proxy characteristics of the proposed computer vision cloud server will negatively effect latency, but should be appropriate for live communications in most cases.

While I suspect treating the mobile end-point devices as a thin client (by offloading the live signal processing to a cloud server) is an obvious gain, the increased latency of adding a proxy server could theoretically be measured against the reduced computational power of the smart phone devices with respect to computer vision processing. For bandwidth reduction, a hybrid approach can be imagined in which the visual stream is pre-processed on site and only the derivative data is sent to the server directly, however this leaves the server without raw video data which would force the users to result to more contrived and elaborate means of seeing the presenters environment.

In the situation where large groups of users wish to view a single location at the same time, it is increasingly likely that the host cannot serve the entire crowd at once. There are several ways to deal with this, the most obvious being that the host serves to one or more mirror sites which divide the user base amongst themselves. A more collaborative method would be to use the users as peer to peer mirrors, but the bandwidth available to the average peer could limit their outbound streaming ability past the point of usability.

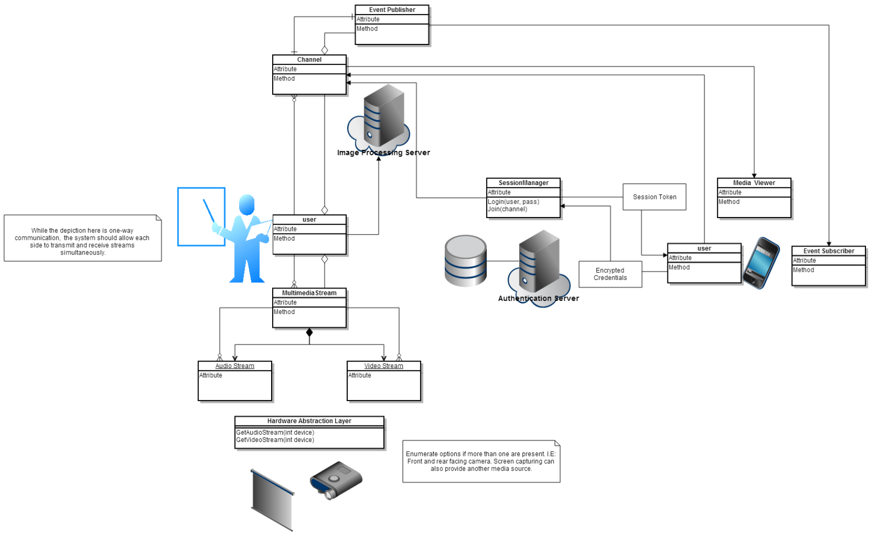
Optimizing Performance

Using the above architecture, the cloud server receives live audio and video data from the presentation room and transmits it to each of the connected clients simultaneously. This data is further expanded with the augmented meta-data and communication between peers. A worthwhile study would include measuring the bandwidth this meta-data occupies, and deciding if there would be significant performance gains by offloading the task of relaying this information between peers to another device.

Another consideration is which codec to use when compressing the audio and video data. Industry standards are heavily pushing the mpeg4 h264 codec, and after a brief overview I am inclined to agree with their methodology.

Finally, the digital signal processing of the audio/video data on the server is certainly the most complex component of this project. The best technique to extract and process the incoming data is a vast and complex topic; One that is extremely circumstantial and unlikely to have a single clear-cut solution. Furthermore, the ability to accelerate real-time video data crunching via gpgpu and parallel techniques is extremely promising in an application such as this.

2.11.3 Message Passing



While the depiction here is one-way communication, the system should allow each side to transmit and receive streams simultaneously.

Enumerate options if more than one are present. I.E: Front and rear facing camera. Screen capturing can also provide another media source.

Plenty of data needs to be transmitted through this system. Not only do we have the raw audio and video data, but we have to arrange a communication protocol which does at least two things. First, it needs to create and manage user sessions in a way that allows them to reliably connect to each other. Second, it needs to communicate the information we derive from each streaming session back to the user associated with the session.

Because the two message passing use-cases are so mutually exclusive, it is entirely possible to divide the workload across two different servers; the first of which would handle authentication and user sessions and the second of which would be tasked with using visualization techniques on the streams and returning any results back to the users.

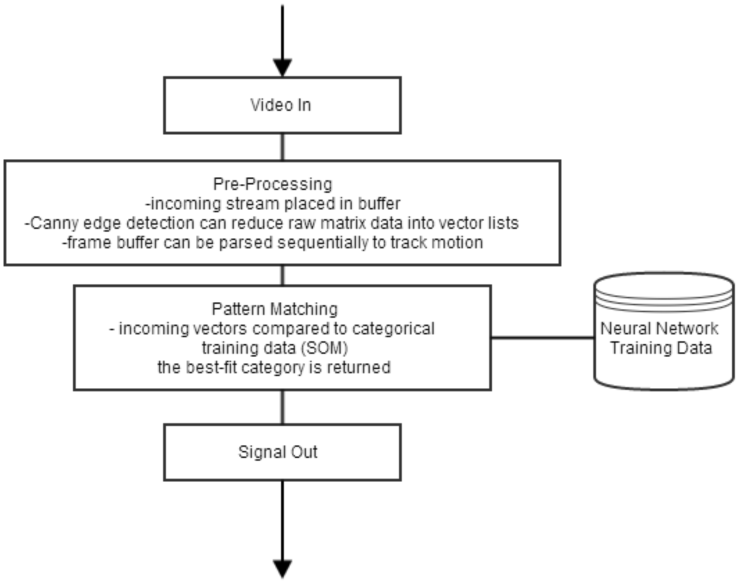
Authentication

left,right,forward,backward toggle mute connect

Visualization

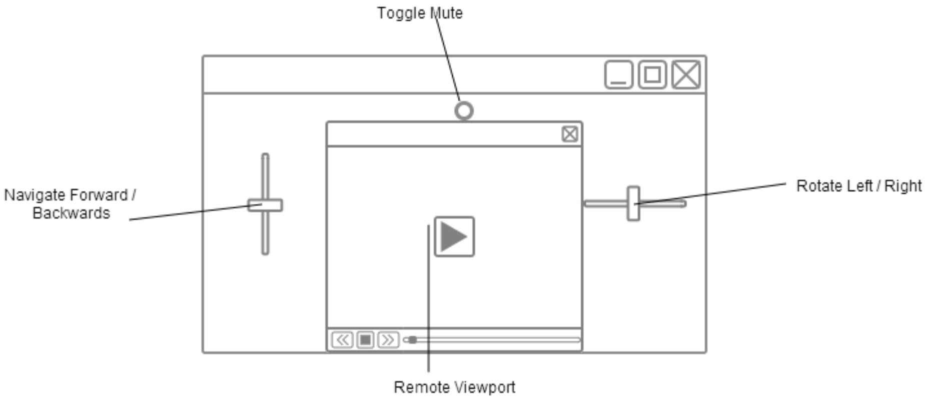
audio video augmented overlay (another video stream)

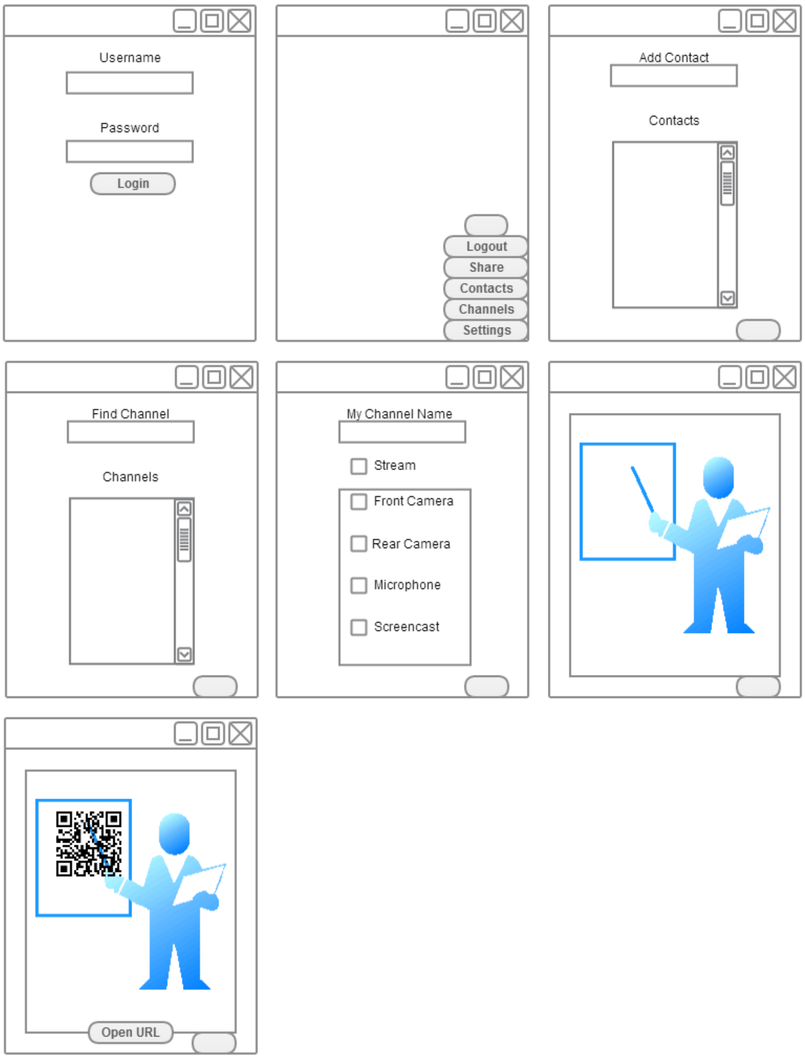
2.12 Digital Signal Processing


Video In
Pre-Processing
-incoming stream placed in buffer
-Canny edge detection can reduce raw matrix data into vector lists
-frame buffer can be parsed sequentially to track motion
Pattern Matching
- incoming vectors compared to categorical
training data (SOM)
the best-fit category is returned
Neural Network
Training Data
Signal Out

- computer vision module
- raw video stream gathered from telepresence hardware
- video analyzed using computer vision and pattern recognition (opencv)
- client connected to this module will receive events via the audio/video analyzer
- scripting language? HighGUI enough?
- client can also potentially receive augmented audio/video data with the DSP module overlay

2.13 UI


Toggle Mute
Navigate Forward / Backwards
Rotate Left / Right
Remote Viewport


Username
Password
Login
Logout
Share
Contacts
Channels
Settings
Add Contact
Contacts
Find Channel
Channels
My Channel Name
□ Stream
□ Front Camera
□ Rear Camera
□ Microphone
□ Screencast
Open URL

Chapter 3

Implementation Details

3.1 Generic Video Capture

3.2 Authentication

3.3 Video Streaming

3.4 QR Code reading

3.5 Event Handlers

Bibliography

[AH11] A.P. Arias and U.D. Hanebeck. Motion control of a semi-mobile haptic interface for extended range telepresence. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 3053–3059, 2011.

[AKK09] J.J. Ahmad, H.A. Khan, and S.A. Khayam. Energy efficient video compression for wireless sensor networks. In *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, pages 629–634, 2009.

[ALYY11] Hou A-Lin, Feng Yuan, and Geng Ying. QR code image detection using run-length coding. In *Computer Science and Network Technology (ICCSNT), 2011 International Conference on*, volume 4, pages 2130–2134, 2011.

[AWZ98] A. Agah, R. Walker, and R. Ziemer. A mobile camera robotic system controlled via a head mounted display for telepresence. In *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, volume 4, pages 3526–3531 vol.4, 1998.

[Bin11] Huang Bin. The study of Mobile Education development based on 3G technique and Cloud Computing. In *Uncertainty Reasoning and Knowledge Engineering (URKE), 2011 International Conference on*, volume 1, pages 86–89, 2011.

[Bra05] J. Brassil. Using mobile communications to assert privacy from video surveillance. In *Parallel and Distributed Processing Symposium, 2005. Proceedings. 19th IEEE International*, pages 8 pp.–, 2005.

[Cav07] A. Cavallaro. Privacy in video surveillance [in the spotlight]. *Signal Processing Magazine, IEEE*, 24(2):168–166, 2007.

[CFB+ar] Samuel L. Clemens, William C. Faulkner, Elizabeth B. Browning, Judith S. Murray, Louisa M. Alcott, Harriet B. Stowe, and Carl A. Sandburg. Primarytitle. In Ralph W. Emerson, William B. Yeats, and Robert L. Frost, editors, *SecondaryTitle*, volume Volume of *ThirdTitle*, pages StartPg–OtherPg, PlaceofPub, PubDateFreeForm PubYear. AuthorAddress, Publisher. Notes.

[CST+13] Yan Cui, S. Schuon, S. Thrun, D. Stricker, and C. Theobalt. Algorithms for 3d shape scanning with a depth camera. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(5):1039–1050, 2013.

[CWKG96] D. Caldwell, A. Wardle, O. Kocak, and M. Goodwin. Telepresence feedback and input systems for a twin armed mobile robot. *Robotics Automation Magazine, IEEE*, 3(3):29–38, 1996.

[EAM12] C. Escolano, J.M. Antelis, and J. Minguez. A Telepresence Mobile Robot Controlled With a Noninvasive Brain #x2013;Computer Interface. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 42(3):793–804, 2012.

[Feh13] G. Feher. The price of secure mobile video streaming. In *Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on*, pages 126–131, 2013.

[FRF+07] Julien Freudiger, Maxim Raya, Márk Félegyházi, Panos Papadimitratos, et al. Mix-zones for location privacy in vehicular networks. In *Proceedings of the first international workshop on wireless networking for intelligent transportation systems (Win-ITS)*, 2007.

[GHJS95] P.S. Green, J.W. Hill, J.F. Jensen, and A. Shah. Telepresence surgery. *Engineering in Medicine and Biology Magazine, IEEE*, 14(3):324–329, 1995.

[Hos07] W. Hosny. Power engineering mobile education technology. In *Universities Power Engineering Conference, 2007. UPEC 2007. 42nd International*, pages 971–974, 2007.

[HR13] M.A. Hossain and S.M.M. Rahman. Towards privacy preserving multimedia surveillance system: A secure privacy vault design.

In *Biometrics and Security Technologies (ISBAST), 2013 International Symposium on*, pages 280–285, 2013.

[JT08] Jiaya Jia and Chi-Keung Tang. Image Stitching Using Structure Deformation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(4):617–631, 2008.

[Kim11] S. Kim. Cellular network bandwidth management scheme by using nash bargaining solution. *Communications, IET*, 5(3):371–380, 2011.

[KV04] Sungwook Kim and P.K. Varshney. An integrated adaptive bandwidth-management framework for QoS-sensitive multimedia cellular networks. *Vehicular Technology, IEEE Transactions on*, 53(3):835–846, 2004.

[LL12] Xinxin Liu and Xiaolin Li. Privacy Preserving Techniques for Location Based Services in Mobile Networks. In *Parallel and Distributed Processing Symposium Workshops PhD Forum (IPDPSW), 2012 IEEE 26th International*, pages 2474–2477, 2012.

[LYC04] Kam-Yiu Lam, Joe Yuen, and E. Chan. On using buffered bandwidth to support real-time mobile video playback in cellular networks. In *Multimedia Software Engineering, 2004. Proceedings. IEEE Sixth International Symposium on*, pages 466–473, 2004.

[PHE02] Sang Yun Park, Moon Seog Han, and Young Ik Eom. An efficient authentication protocol supporting privacy in mobile computing environments. In *High Speed Networks and Multimedia Communications 5th IEEE International Conference on*, pages 332–334, 2002.

[PJ09] R. Pathak and S. Joshi. Recent trends in RFID and a java based software framework for its integration in mobile phones. In *Internet, 2009. AH-ICI 2009. First Asian Himalayas International Conference on*, pages 1–5, 2009.

[XP09] Yingen Xiong and K. Pulli. Sequential image stitching for mobile panoramas. In *Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference on*, pages 1–5, 2009.