Background

The principal strength of robots is that robots can be deployed where humans cannot or should not be deployed. Correspondingly, humans’ principal strengths are that we perform well in complex and unstructured environments where robotic technologies are limited.

Through co-robotic systems, human beings can operate in the vacuum of space\(^1\); can be scaled to penetrate much smaller spaces than humanly possible\(^2\); and to operate at much larger than human scales\(^3\).

Approach

Our primary goal is to remove humans from hazardous environments by providing a means of performing the required tasks remotely. This will allow them to function more effectively, working from an environment of personal safety. To achieve this goal we are developing integrated co-robotic systems within a Co-Robotics Telesupervision Architecture that supports augmenting human capabilities. Our Architecture supports:

- incorporating autonomous agents to monitor for safety, and to assist the worker;
- situation awareness through immersive multi-sensory high-fidelity telepresence;
- capability to scale the effective size of the human worker to better match each task;
- capability to augment the spectra of vision, audition, spatial orientation, and proprioception;
- precise physical interaction through haptic teleoperation.

Complex tasks in unstructured environments are more difficult to characterize than repetitive tasks in well-structured settings. A co-robotic system through which humans operate remotely is ideal for examining the multi-modal sensory stimuli and sensorimotor command data as they are relayed between the robotic agents and the human. Therefore, our Architecture for deploying human expertise through co-robotic systems will facilitate monitoring these data, and will support progressive sensory and cognitive augmentation.

By inserting monitoring agents in the remote sensory channels, concurrent analysis of the environment can be automatically conducted. Possible examples include: building a 3D map of the environment that is viewable by the operator in an accessory window; identifying anomalous features and overlaying these on the remote three-dimensional view for the operator to consider more closely. In a similar way, inserting monitoring agents into the movement (or motor) control channels allows monitoring of the human's intended remote movement and modifying it. Possible examples include scaling movement, or limiting accelerations, or applying soft limits for physical 'stay-out areas' to prevent collisions. This augmentation applies both to manipulators and vehicles.

We will prove the system through conducting formal experiments that extend human senses: modifying sensory spectra and altering geometric scale; and that extend human cognition with intelligent agents that autonomously monitor operations, augmenting human reasoning to reduce cognitive load.

Co-Robotics Telesupervision Architecture

Of the many forms that humans working with robots may take, we focus on a systemic approach to augment human capabilities. This includes extending human senses and reach physically; modifying human senses in scale both geometrically and spectrally; and expanding human cognition.

Recognizing the limits of autonomy that preclude direct leaps from the majority of human-accomplished tasks to fully-automated tasks, we propose a tractable approach to selectively integrate augmentation of human sensory and cognitive capabilities. Within our open architecture of human/autonomous cooperation, we support a

---

1 NASA’s upgraded Robonaut 2, <robonaut.jsc.nasa.gov>  
3 T-52 Enryu Rescue Exoskeleton, <www.enryu.jp/t52>
progression from direct human teleoperation; to augmented human operations; to high-level human supervision of autonomous actions. Our architecture provides a framework within which co-robotic assembly and inspection systems can be continuously improved as intelligent autonomous agents are developed, proven robust, and integrated.

Figure 1. Co-Robotic Telesupervision Architecture with autonomous agents for human augmentation.

The primary elements of our co-robotic telesupervision architecture (Figure 1) as embodied by a system for remote operation are: the distal robotic sensory and manipulation tools; and the proximal immersive telepresence and manipulation controls of the telesupervisor's workstation.

Distal Robotic Systems

The co-robotic systems include multi-modal teleperception sensors for binocular stereoscopic vision, binaural stereophonic audition, force-reflecting haptic manipulation, and proprioception for vehicle attitude and accelerations. These physical robotic elements are deployed remotely into the field using robotic vehicles adapted to the space requirements. For example, relatively flat terrain may use a wheeled robot vehicle, while access to a cliff face may require a specialized climbing vehicle.

Intelligent Assisting Agents

The high-fidelity sensory data for immersive telepresence and teleoperation is transmitted between the remote robotic assets and the telesupervision workstation that is situated in a human-safe environment. These data are available to the Intelligent Assisting Agents that can autonomously monitor, interpret, indicate, automate, and limit. The architecture supports development of autonomous agents as each co-robotic system task domain is analyzed and defined allowing their modular development, testing, and incorporation.

High-level autonomous agent planning and monitoring is supported by our Telesupervision Architecture. Some existing autonomous agents are relatively mature, such as some robot navigation techniques. Others are less robust or are developed per application such as task-specific planning and monitoring. Our co-robotic telesupervision architecture will support incorporation of additional and improved autonomous agents as they are developed, tested, and proven.

Gracious fall-forward/fall-back between autonomous agents and direct teleoperation is one of the key strengths of our modular augmented telesupervision.

Distant Human Expert Telecollaboration
A further expansion of the concept of the "intelligent assisting agent" that is supported by our architecture is the facility to provide a subset of the telepresence data to a distant human expert who has more specific domain knowledge than the telesupervisor operating through the co-robotic system. This is especially useful when an unforeseen condition is detected for which additional expertise is desired. By supporting this telecollaboration access to a wide variety of distant domain experts, unexpected situations can be addressed rapidly, without the time and cost to co-locate the experts for consultation.

High-Fidelity Immersive Telepresence

Situation awareness and the sense of presence requires high-fidelity capture and reproduction of sensory and sensorimotor data. Telepresence presentation to the telesupervisor includes geometrically-correct binocular stereoscopic viewing systems, and high-fidelity stereophonic audio reproduction. Force-reflecting manipulation control reflects to finger, hand, and arm exoskeletons, allowing the teleoperator to feel directly into the environment through the deployed co-robotic system. The attitude (orientation) and vibration of the co-robotic vehicle or end-effector will be relayed to the Telesupervision Workstation and reproduced by adjusting the attitude of the platform or operator chair for Vestibular Spatial Orientation and to provide 'seat-of-the-pants' proprioception.

Co-Robotic Telesupervision Workstation

By integrating operator interface components for mobility, manipulation, telesensing, autonomous agent tasking; and by providing a portal to facilitate remote experts’ telecollaboration, the Co-Robotic Telesupervision Workstation becomes the hub of planning, control, and collaboration.

• Direct human teleoperation is supported by situation awareness through immersive multi-sensory high-fidelity telepresence; and precise physical interaction through haptic teleoperation.
• Augmented human operation is supported by autonomous agents to monitor for safety, and to assist the worker; scaling the effective size of the human worker to better match each space and task; and augmenting the spectra of vision, audition, spatial orientation, and proprioception.
• High-level human supervision of autonomous actions is supported by intelligent assisting agents that incorporate greater autonomy such as: safe path planning and navigation, automatic task-specific operations, or system 'health' monitoring.

Mitigating Risks by Design

When robotic assets are deployed into a high-risk area, one into which humans should not or cannot be sent (e.g., extreme terrain, high radiation flux), then the greatest risks are to the deployed robotic assets and the environment/workspace into which they are deployed. Risks include unplanned or unintended actions: falls; collisions; becoming entangled or wedged. These are often the result of lack of perceptual awareness of the environment (by the human or intelligent assisting agents).

By designing a task- and environment-appropriate immersive, multi-sensory perceptual system, combined with force reflecting extremities, the teleoperator working through the co-robotic system will have a more faithful sense of being in the environment, and can employ the intuition and careful practices that an in-situ worker would use for ensuring safety.

As more sophisticated robot health and safety algorithms are developed, and higher-level planning agents optimize the sequence of actions, the risks of unplanned actions will be further reduced.

Co-Robotic Systems for Planetary Exploration Tasks

The following sections detail the sub-systems we can develop and integrate within the framework of our co-
robotic telesupervision architecture, and to prove our space co-robotic systems through conducting experiments with representative tasks. The sections are structured by task and related technology areas as:

- Immersive multi-sensory high-fidelity telepresence;
- Force-reflecting haptic teleoperation;
- Augmented human cognition; and
- References.

This development will provide a co-robotics telesupervision architecture that has been proven by being applied to analogous real-world tasks and formal experimentation and analysis.

**Co-Robotic System for Remote Space Operations**

Our design principles and methodology are predicated on certain fundamental requirements:

First, the physical systems of the deployed robotic equipment must be physically capable of accomplishing the domain-specific tasks.

Second, the deployed sensing capability must support perception of the surroundings and conditions with sufficient fidelity to remotely accomplish the goal tasks.

Third, the Co-Robotic Telesupervision Workstation must employ the most natural interfaces practical for sensing and control designed in accordance with Human Factors and Ergonomics design principles.

The adequacy of these system components is critical — they must be proven adequate initially with a human in the loop. If these systems are inadequate to the tasks under direct human teleoperation, it is unproven that any amount of autonomy can make them so.

**Immersive Multi-Sensory High-Fidelity Telepresence**

For effective situation awareness, high-fidelity sensory cues are required, including geometrically-correct wide-angle binocular stereoscopic vision, binaural stereophonic audition, haptic proprioception, and vestibular spatial orientation. It is critical to provide as natural as practical a rich sensory experience to allow the telesupervisor an immersive high-fidelity experience of being present remotely. Recognizing that the force-reflecting extremities, and the vestibular spatial orientation detailed in the two sections following this one provide significant physical cues to the deployed environment, immersive visual and aural sensory cues are fundamental to situation awareness. These aspects of telepresence can be addressed within a telepresence sensory "head" incorporating a geometrically-correct binocular stereoscopic camera, and an acoustically-correct binaural stereophonic microphones positioned with respect to the anthropomorphic relationships of the human head.

**Geometrically-Correct Vision**

To provide the teleoperator the most natural, and least fatiguing telepresent experience, our high-fidelity approach to those aspects of perception that give the sense of truly “being there” are an integrated system of geometrically-correct binocular stereoscopic cameras and viewing systems. We address binocular vision with well-developed, well-reported, and well-demonstrated concepts for geometrically-correct remote visual sensing [Podnar et al, 2006, Grinberg et al, 1994].

To quickly gain accurate situational awareness, a telesupervisor's remote vision system must faithfully reproduce a view analogous to that gazed upon by the uninstrumented eyes. Any introduced distortion impairs the operator's ability to work precisely, and can cause substantial fatigue with prolonged use. Oversimplified binocular systems that merely converge the optical axes, or "toe-in" two cameras, result in horizontal and vertical misalignment distortions that increase as the gaze moves away from the center of the scene. Vertical errors are fatiguing and can cause headache, nausea, and can leave the person with temporary residual vertical phoria (eye misalignment).
The comparison between the oversimplified approach, and the geometrically-correct approach is illustrated in Figure 2. We note that the geometrically-correct camera sensors are co-planar, and that the cameras are modified to shift the center of each sensor off of the lens optical axis to shift the fields of view, allowing a visual area of the fields of view to be coincident. This modification of the cameras, and the precision with which it must be made is significant (Avoiding this modification effort may account for the popularity of the illustrated oversimplified approach.).

It is insufficient to consider only the camera in a geometrically-correct telepresence viewing system. To reproduce reality as if the viewer were gazing on the scene with uninstrumented eyes, the display system must also adhere to equivalent geometries. When camera imagery is displayed on a video monitor, it is natural to consider this monitor as a window "through" which the viewer gazes. By strictly adhering to equivalent geometries of a direct view with human eyes through a window for the binocular stereoscopic camera, and the view of a virtual image "through" the screen of a stereoscopic display system we can accurately reproduce the object scene (Figure 3).

In the diagram of Figure 3a, the interpupillary distance between the viewer’s eyes is a fixed measurement. The width of the window constrains the angle of view for each eye and defines the area of coincidence when we position the eyes such that a line drawn through the two pupils is parallel with the window, and position the cyclopean point (the point between the two pupils) normal to the plane of the window and centered on the aperture of the window.

Selection of the effective window aperture is limited by the physical width of the display screen. Incorporating the distance of the viewer’s eyes from the display screen completes the system’s geometric constraints. In Figure 3b,
the spacing of the cameras is set at the average adult human interpupillary distance of 63 mm to provide the same view as directly gazing on the scene. The area of coincidence is set at the distance of the viewer’s eyes from the display screen as shown in Figure 3c. The camera lenses must then be chosen to equal the calculated angle of view.

Geometrically-Correct Scaled Vision

We can deploy three-dimensional telesensory vision systems that are geometrically analogous to the perception of uninstrumented eyes. However, without violating these constraints, we have the ability to modify the effective scale of the human operator working through the co-robotic system. This ability to scale is a powerful augmentation of the perceptual capabilities of the human telesupervisor.

Modifying the effective scale of the remote co-robotic vision system does not involve changing the magnification of the cameras (zooming) as this introduces depth distortions along the optical axis. It is purely by changing the interpupillary distance of the camera lenses that a scaled viewing geometry is achieved. In addition to building a human scale remote camera system, we can develop a scaled camera system to effectively reduce the size of the telesupervisor when working through the co-robotic system.

Anthropomorphic Audition

In environments where sound propagates naturally through a gaseous medium, audition can be employed as an effective immersive modality. When natural audition is not possible (i.e., in vacuum) stereophonic audition can still provide very useful cues for both mechanical system operation and physical interactions with objects and the ground by employing contact microphones on the structure. We detail the former in the following paragraphs.

Many subtle depth and manipulation cues are processed subconsciously through stereophonic hearing. The cost of adding this sensory modality is relatively very small, yet the situational awareness it provides of the telepresently perceived environment is enormous. We address binaural hearing (audition) at normal human scale by integrating a commercial anthropomorphically-correct stereophonic microphone system (e.g., Neumann KU-100) at the robot end and existing commercial audio reproducers at the workstation end. All the required technology and components are commercially available, but careful system design and integration, considering the natural geometry of the eyes and ears of a human's head, shoulders, and torso, are required to maximize sensory utility and minimize operator fatigue and the possibility of operator distraction.

For scaled stereophonic audition (to complement the scaled stereoscopic vision identified above) commercially-available high-fidelity miniature microphones can be incorporated into the scaled co-robotic telesensory system. While many of the same binaural localization cues (e.g., intensity, timbre, spectral qualities, reflections in the confined space) may be maintained, the timing cues and phase cues in certain frequency bands will be reduced or altered if the interaural distance is altered. (An in-depth treatment of these auditory scaling issues is a valid topic for further research.)

Force-Reflecting Haptic Teleoperation

We note that humans are very capable of navigating through dark rooms without vision using touches and gently bumping into objects. Navigating by touch is very low bandwidth. For interacting with the remote environment, force-feedback actuators and posture proprioception add to faithfully reproducing sensorimotor control. It is within this rich sensory environment that humans can effectively work remotely in hazardous environments; and the basis upon which tools for progressively augmenting and automating human tasks can be effectively developed.

A deployed robotic portion of the system “above the waist” integrates robotic hands and arms with force-reflecting exo-skeleton controls to allow the teleoperator to perform a wide variety of manipulation tasks naturally. A minimum manipulation subsystem consists of: a 4-axis force-feedback arm; two-fingered force-
feedback hand; and a force-reflecting exo-skeleton for fingers and arms at the Co-Robotic Telesupervision Workstation.
The reproduction of gross forces at the hand allows proprioception, or kinesthesia, which is the self-sense of the position of limbs and other parts of the body. This provides a significant additional cue to the immersive visual teleperception.

To address fingertip tactition, a skin-sensor and skin-display tactile system must faithfully relay the touch cues that human skin senses with at least four distinct types of tactile cells located at different skin depths and endowed with different sensor modalities – pressure, temperature, and vibration each responding to different frequency ranges. A separate document is available that goes into greater detail.

Vestibular Spatial Orientation

The attitude (orientation) of the co-robotic vehicle or end-effector can be relayed to the Telesupervision Workstation and reproduced by adjusting the attitude of the platform or operator chair for Vestibular Spatial Orientation at relatively low frequencies. These can have scaled adjustments and be limited for safety to prevent overtipping the telesupervisor while providing this cue. An inertial measurement unit incorporated into the distal robotic systems, relays acceleration and orientation data to three actuators that drive the telesupervisor's support platform.

In addition, we can relay relatively higher frequencies (i.e., collision, vibration) to provide richer 'seat-of-the-pants' proprioception. By employing low-frequency audio drivers vibrations can be reproduced. Again, these can be scalable to relay the cue without significantly impacting the telesupervisor.

Augmenting Human Cognition

Our co-robotic telesupervision architecture supports incorporation of autonomous agents applicable to the co-robotic system tasks of space operations. The support for modular insertion of these agents allows development and testing each agent without significant modification of the co-robotic system. The high-fidelity sensory data for immersive teleperception and teleoperation is transmitted between the remote robotic assets and the co-robotic telesupervision workstation. These data are available to the Intelligent Assisting Agents that can autonomously monitor, interpret, indicate, automate, and limit.

One example of a visual augmentation agent may address autonomous detection of visual 'features-of-interest', and that identifies these features to the human telesupervisor via a 3D visual overlay that aligns with the viewed environment. The telesupervisor is provided with controls to modify modes of the display (including turning it off to reduce distraction). The feature set can be selected from task-specific needs and may include automatic detection of unusual geologic formations, rocks with unexpected characteristics, and other automated detection agents as they are developed and proven.

Another example visual augmentation agent can address localization within the environment using sensors to build a three-dimensional map of the environment. This can be generated from a priori data (such as orbital imaging) and refined in greater detail from visual sensors being monitored locally. A separate virtual display can be presented to the telesupervisor that shows the mapped space and the deployed co-robotic system within the space to provide a non-immersive higher-level situation awareness (common to many video games). This can also include controls such as point-of-view adjustment.

A physical augmentation agent may be informed by the localization data above. Based on the mapped workspace and the position and posture of the deployed co-robotic system, the intelligent assisting agent can employ algorithms to determine safe workzones and 'soft' stay-out volumes to prevent unwanted collisions or damage. These volumes will be updated in real time as operations are carried out.

Telecollaboration
A telecollaboration sub-system can be managed at the Co-Robotic Telesupervision Workstation to enable a range of secondary remote experts to observe (e.g., NASA JPL “backroom”), and tertiary experts to review using conventional telecollaboration tools (e.g., via world-wide web pages), to provide domain expertise to improve decision-making associated with the situation.

References


