# Seeing Through Multiple Sensors into Distant Scenes: The Essential Power of Viewpoint Control

Alexander M. Morison(✉), Taylor Murphy, and David D. Woods

The Ohio State University, Columbus, OH, USA
{morison.6,murphy.1018,woods.2}@osu.edu

**Abstract.** Sensors are being attached to almost every device and vehicle and integrated together to form sensor systems that extend human reach into distant environments. This means human stakeholders have the potential to see into previously inaccessible environments and to take new vantage points and perspectives. However, current designs of these human-sensor systems suffer from basic deficiencies such as an inability to keep pace with activities in the world, the keyhole problem, high re-orienting costs, and the multiple feeds problem. Principled approaches to the development of human-sensor systems are necessary to overcome these challenges. Principles for viewpoint control provide the key to overcome the limitations of current designs.

**Keywords:** Human-sensor systems · Viewpoint control · Perspective control · Human-robot interaction · Data overload · Keyhole effect · Wide-area surveillance · Sensor systems · Virtual environment · Multiple feeds problem

## 1 Introduction: Current Human-Sensor Systems

Sensors are sprouting up everywhere. They are being attached to almost every device and vehicle and can be found in many public and private spaces. Often, multiple sensors are integrated together to form sensor systems that extend human reach into distant environments. This means stakeholders have the potential to see into previously inaccessible environments and to take new vantage points and perspectives. All of the data feeds from sensors provide the basis for people to take action, authorize actions by other systems, or delegate authority to increasingly autonomous systems.

Many different types of sensor systems can be found in a wide range of applications like telemedicine, wide-area surveillance, emergency response, and oil exploration. In telemedicine alone several types of sensors systems can be found. Robot's move equipment and material between rooms, allow medical personnel to monitor patients remotely, and act as surgical aids in the case of robotic-assisted surgery. For emergency response, sensors provide the potential for new access to previously inaccessible areas. This inaccessibility can arise from limited physical access, in the case of a collapsed building, or because of danger in the case of a suspicious package. In each of these examples, and all of these domains, stakeholders and sensor systems form human-sensor systems that allow stakeholder's to see and act in environments in which they are not present [1].


M. Kurosu (Ed.): HCI 2016, Part II, LNCS 9732, pp. 388–399, 2016.
DOI: 10.1007/978-3-319-39516-6_37

Despite the advances in sensor technologies, actual deployments of more and more sensors have produced surprising challenges as the scale of data and reach of these systems have grown. One of the challenges is the inability to keep pace with activities in a remote environment, in particular when the sensor system is a network of sensors like in wide-area video surveillance [2]. A second challenge related to pacing is a lack of peripheral awareness, which is where a narrow "keyhole" view makes it impossible for a stakeholder to see what is just beyond their current view [3]. A third challenge is the difficulty shifting viewpoints across sensor feeds to find informative views, called the multiple feeds problem [4]. A fourth challenge for human-sensor systems arises from confounding sensor control with action potential in the design of the sensor platform. This confounding creates the possibility for sensor platforms, like robot's, to get stuck, lost, or bump into things.. At the heart of these observations is a single basic challenge for human-sensor systems using today's technologies. As sensor systems widen our ability to monitor distant environments and the amount of sensor data increases by orders of magnitude, human observers are less able to shift their viewpoint and the viewpoint of the sensor systems to keep pace with interesting activities.

This paper uses basic principles about viewpoint control derived from human perception to explain why the above observation recurs with current technologies. The paper also introduces viewpoint control technology – the basic principles for designing interactions that allow people to shift viewpoint fluently. Fluent viewpoint control is a prerequisite for people to track what is interesting at different scales as the data streams about activities in the world change.

## 2 Basic Deficiencies in the Design of Current Human-Sensor Systems

There are several important deficiencies with the existing designs of human-sensor systems. The first of these is the inability of these systems to keep pace with activities in the distant environment. The second deficiency is that keyhole effects are everywhere and in many cases sensor systems experience multiple keyholes either from multiple sensors or the poor design of navigation mechanisms – like in the case of large field-of-view sensors described in this section. A related deficiency to the keyhole effect is the limited navigation support that exists for looking across multiple sensor feeds.. And lastly, sensor system designers have confounded viewpoint control and action capability to the point that even simple tasks become complex.

### 2.1 Inability to Keep Pace with Activities

The pace or tempo of activities and events is a fundamental concept for any human-technology system including human-sensor systems. As a temporal and relational property it is usually absent from the design and analysis of interactive systems. Pace is a difficult concept to include in studies and designs because it captures a dynamic relationship between an environment and an observer(s). Pace depends on both the activities and events in the world, but also the knowledge, experience, and other tasks of the observers. The relationship between these temporal processes is also dynamic, which of these processes leads a change in pacing and which follows depends on context. In the end, an observer's ability to track events must match the tempo of activities and events in the world of interest.

The limited ability of human-sensor systems to keep pace with activities can be illustrated at the scale of a network of multiple sensors. There are several sources of complexity that make tracking activities across multiple sensors difficult within a wide area video surveillance network (Fig. 1A). A few of these sources include different perspectives, overlapping and non-overlapping sensor coverage, and sensor data displays that remove spatial relationships between sensors. The net result is that wide area surveillance systems are most valuable in forensic analysis after events have already occurred. In forensic analysis, observers slow down the flow of sensor data as a work-around to escape from the inherent pacing limitations of the human-sensor system. Current designs impose high cognitive load as the observer must keep a model in their head of the relationships across cameras, the directions the cameras point, the orientation of the activity to be monitored, and the gaps in camera coverage.

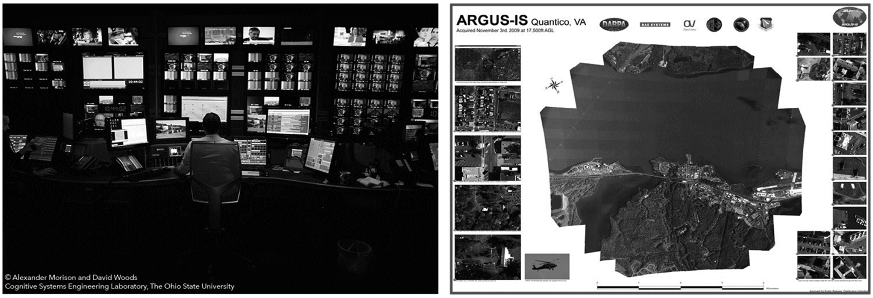


**Fig. 1.** A. Wide area video surveillance network. B. Image from a large, high resolution field-of-view sensor.

Another trend in the development of human-sensor systems is to create extremely large field-of-view sensors, which is a response to the challenges associated with narrow field-of-view sensors (Fig. 1B). These sensors eliminate the need to control multiple sensors at the same time. However, these sensor feeds are so large that people can only take in a portion of the total field-of-view at any one moment. As a result, people shift their focus sequentially over the available field at a scale that is less than the total available field-of-view. This typically means a person must scale and translate – zoom and pan – a virtual view or open up a series of windows each at an appropriate resolution for the task. Neither of these trends in human-sensor system design help users know where to look or focus next as events occur in the scene of interest.

These examples illustrate some basic findings about viewpoint control in human perception and in human-sensor systems [4, 5]. The process of perceiving is an active sampling process. There is no perfect single viewpoint; rather, the view from any single point of observation simultaneously reveals and obscures properties of the scene of interest so that comprehension requires shifts of viewpoint [4]. The key is supporting the control of this sampling process to track what is “interesting” over time.

### 2.2 The Keyhole Problem

An old challenge that re-emerges for human-sensor systems is the keyhole effect [6], which has been referred to as the “soda straw effect” for sensors on robotic platforms [7]. For example, navigating a robot through broken terrain is difficult when the robot handler is stuck looking through the ‘soda straw’ of a single camera. The human handling the robot can easily miss important features of the terrain that are difficult to discern, like landmarks along a path [8]. This is one example of how the keyhole makes robot handlers vulnerable to missing important properties of an environment that are just beyond the boundary of the camera’s current field-of-view. Expanding the awareness of the robot handler requires moving the camera’s view, but in what way and when? This challenge can be illustrated through real-world examples.

**Example: Robotic Response at Fukushima.** The robot operators at the 2011 Fukushima Daiichi nuclear power plant disaster experienced a perceptual shortfall when using just the sensors located on the robot. The limits imposed by the keyhole effect and lack of peripheral awareness left them unable to supervise and tele-operate the robots in the damaged environment successfully. But as responsible problem holders, the roboticists and engineers adapted their method of deploying robots [9]. Instead of sending a single robot into the nuclear power plant, robots were deployed in pairs (see Fig. 2). The first robot performed the planned tasks, such as opening doors and turning valves. The second robot of the pair was used to provide a view of the first robot relative to the local environment. This second view provided the first robot handler the necessary context around the first robot to successfully navigate doorways and passageways [10]. The adaptations during the response to the Fukushima nuclear power plant disaster provide a real example where resources were reallocated to provide a necessary viewpoint to overcome perceptual limits in order to complete critical tasks and achieve pressing goals.

**Fig. 2.** At the Fukushima Daiichi nuclear power plant disaster two robots were used to overcome the keyhole effect with a single robot.

### 2.3 Multiple Feeds Problem

There are several natural responses to the limited ability to perceive through the sensors on robots. These include adding cameras in the scene of interest, adding cameras to the platform, or to add more platforms in the scene of interest. These workarounds quickly result in a new challenge: how to integrate and navigate these diverse sensor feeds now available to remote observers. Shifting viewpoints to find informative views or view transitions for a particular context can be a challenging activity. Viewpoint control is a skill that people exhibit naturally as they move through and explore scenes (e.g., [11]). Re-establishing a basis for fluent skilled control of viewpoint for human-sensor systems extended over these new scales is an extremely difficult design challenge.

### 2.4 Confounding Viewpoint Control and Action Capability

The challenge of designing viewpoint control is particularly clear when robotic platforms confound action capability and viewpoint control. This confound occurs for any sensor platform where viewpoint control is dependent on moving the platform through the world. When desired actions of the platform produce unintended viewpoint shifts (or the reverse), operators are forced to balance an extremely difficult trade-off. There are many circumstances where operators may wish to move sensors independent from the action potential. Nearly all existing sensor platform system designs fail to design for independent viewpoint control or coordinated interplay between viewpoint shifts and action capability.

An example of how this confound undermines the fluency and capability of the human-robot system is shown in Fig. 3. In this case, a robot operator is attempting to place a cylinder inside a circular aperture as part of a manipulation task. This robot operator quickly realized that there was a large amount of uncertainty about how to grip the cylinder and when to release the cylinder in the aperture accurately. When gripping the cylinder, it was difficult for the operator to determine whether enough of the cylinder had been grasped so that it would not slip through the gripper and fall to the ground. Then, when attempting to release the cylinder in the aperture, it was difficult to determine whether the cylinder would remain in the aperture, or not, after release. All operators struggled with these two decisions and the relationship between them, meaning the best solution for gripping the cylinder to avoid dropping was the worst solution for gripping the cylinder to make sure it remained in the aperture, and vice versa. The irony of this example is that a child easily completes this exact task by the age of 2. This confound occurs commonly in all types of action capable sensor systems such as ground robots, manipulators, and submersibles.

The confound of viewpoint control with platform action assumes that a change in sensor positions will not negatively influence the positioning of action capability, and vice versa. This assumption is violated when viewpoint and action demands occur simultaneously, as they did in the example above. This design makes the assumption that all viewpoint movements are equally informative; an assumption that is not true usually. In the manipulator test, operators had to make complex manipulator arm movements to try to find valuable viewpoints. This is far from how people re-orient as they


**Fig. 3.** A human-ground robot attempting to insert a cylinder through an aperture. After releasing the cylinder it tilts back into the gripper, away from the aperture (lower right).

move through a scene. These examples illustrate how current designs of human-sensor systems leave observers to deliberate about properties of the scene via slow, high workload, error prone processes. These weaknesses led designers to attempt to develop an additional layer of external cognitive aids to repair the weaknesses in these deliberative processes. Ironically, the development of human-sensor systems appears to reprise ancient debates about indirect versus direct approaches to explain how human perception works. There is a common theme running through these deficiencies – a lack of viewpoint control that would assist people make comprehensible perspective shifts to track what is informative as events occur and situations change.

## 3 Perspective Taking Concepts

Human-sensor systems are not designed to help stakeholders shift perspectives, transition across viewpoints, or see what other viewpoints could be taken next. These systems have undermined the skilled perspective shifts that people exhibit naturally as they move through and explore scenes (e.g., [11]). Indeed, a closer examination of sensors systems reveals the ad hoc nature of how these systems are designed. At best, sensors are positioned based on a specific need supporting a single function. At worst, sensors are positioned based on irrelevant factors like mounting space and cabling requirements.

Overcoming the challenges described above requires a principled approach to the design of viewpoint control for human-sensor systems. The works of [1, 12] developed the principles of viewpoint control using findings from visual perception, ecological perception, and the neurobiology of attention as these have been applied to human-technology systems [2, 5, 6].

The first basic principle is that there is no single best view: the view from any point of observation simultaneously reveals and obscures aspects of the scene of interest [13, 14]. As a sensor moves relative to the scene, some properties of that scene, and the objects, people, and activities present, are directly and accurately perceived, while the same perspective shift makes other properties more difficult to perceive accurately [13, 15]. Fundamentally, it is through shifting perspective that observers come to comprehend activities and status of the scene of interest.

The second principle is that getting lost and keyhole effects are diminished or disappear when you can provide two kinds of views effectively in parallel. One view is what can be seen from the current sensor position (its point of observation and its view direction). The second provides the view of a sensor's position and orientation (view direction) relative to the scene. This principle was established by applying results from perception to information system design [6]. The usefulness of the second view increases as the number of sensor viewpoints over which an observer can see and explore increases (see illustrations in [4]). The Fukushima case described earlier is a locally innovative but crude example of providing two views in parallel.

The third principle is that viewpoint control requires both egocentric and exocentric frames-of-reference and a mechanism to transition between these frames-of-reference. An additional constraint on these frames-of-reference is that they are spherical coordinate systems. These coordinate systems provide the foundation for creating comprehensible shifts in perspective – changes in point of observation – and constrain or eliminate poor changes in point of observation.

Fourth, viewpoint control utilizes the principle of center/surround from human perception to break down the keyhole. Center/surround requires a high-resolution central view with a larger (even if lower resolution) surround area. The center specifies the current viewpoint of interest and the surround provides re-orienting cues about where to focus next in the search for interesting structures or activities across the scene of

**Fig. 4.** A panoramic representation created from a narrow field-of-view PTZ camera (left) and two views of a viewpoint control input device or 3-dimensional joystick (right).

interest. This principle has been found to be critical for exploration in virtual environments and underlies the natural organization of the human visual system.

These principles are applied to the design of a wide-area video surveillance system. This system combines center-surround representations of video surveillance feeds, a viewpoint control device that embodies the egocentric and exocentric frames-of-reference, and a virtual three-dimensional environment that acts as the medium where a virtual viewpoint can be navigated over the wide-area sensor network.

## 4 A Viewpoint Control Demonstration Using a Wide-Area Video Surveillance System

This wide-area video surveillance network system shows how the viewpoint control principles apply to the design of a human-sensor system. This demonstration illustrates how egocentric and exocentric frames-of-reference can be used to control the orientation of a sensor. This example also presents one method for making a smooth transition from an egocentric to an exocentric frame-of-reference. The demonstration also provides an example of how selecting and viewing different sensor feeds can be designed as virtual navigation across sensor views. Lastly, this demonstration establishes how properties of a virtual viewpoint, like virtual distance, can be leveraged to encode sensor network properties like sensor control.

The demonstration is built with three main components, a new viewpoint control device (a 3-dimensional joystick), a panoramic representation built for a pan-tilt-zoom (PTZ) capable video camera, and a 3-dimensional virtual environment. The viewpoint control device and the panoramic representation are shown in Fig. 4. The 3-dimensional virtual environment is shown in Fig. 5. This virtual environment is populated with four video cameras each depicted by a panoramic representation. Since the video camera can only see a narrow field-of-view, the panorama is updated by moving the camera around the scene. The design of the current system balances the update of the panorama with a decay that removes stale data. This update can be seen in the panoramic representation in the lower left area of Fig. 5, where the top half of the panorama is updated, but the bottom half appears decayed. The panoramic representation in the 3-dimensional virtual

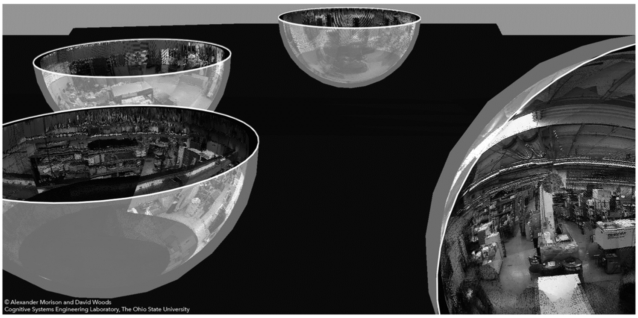


**Fig. 5.** A 3-dimensional virtual environment populated with four panoramic representations created by PTZ cameras.

environment allows an operator to look at current time, but also the recent past. This illustrates a new way of addressing the issue of pacing previously described.

The updating of each panoramic representation is an automatic behavior of the PTZ camera. When an operator needs to track an activity or event of interest, looking at a physical location in the distant environment involves aligning the 3-dimensional joystick, the 3-dimensional virtual viewpoint, and the orientation of the PTZ camera. This alignment is shown in Fig. 6 with the controller in an egocentric configuration. The virtual viewpoint is shown on the left at the center of the panoramic representation, as if at the position of the PTZ camera. The configuration of the joystick and the orientation of the joystick viewpoint are shown on the right of the figure. The relationship between the center of the controller (the white dot annotated as the point-of-observation) and the view direction block to the lower right (the white dot below the point-of-observation) define an egocentric frame-of-reference. The movement of the controller to the left or right creates a similar change in the orientation of the virtual environment viewpoint and the orientation of the PTZ camera.

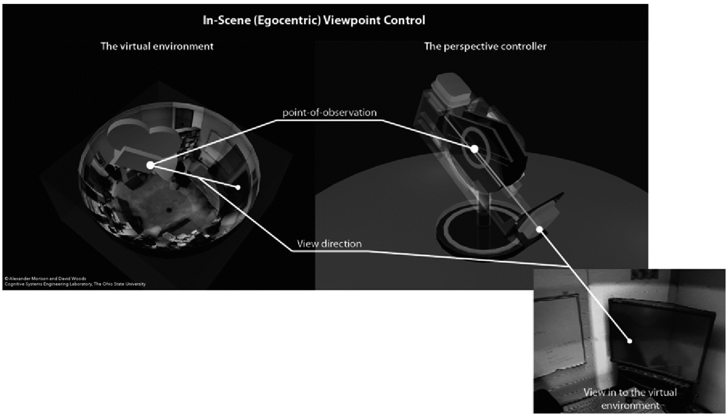


**Fig. 6.** The egocentric relationship between a virtual camera in a virtual environment with a panoramic sensor representation (left) and the viewpoint input control device (right).

The controller can also be configured to represent an exocentric frame-of-reference as shown in Fig. 7. The exocentric frame-of-reference is created by reversing the orientation of the view direction block (upper left of the controller) relative to the center point of the controller. In the virtual environment, the left panel of Fig. 7, the virtual viewpoint is outside of the panoramic representation. At the same time, the virtual view and camera view align to show a similar PTZ camera view as in the egocentric configuration. This exocentric viewpoint provides a wider virtual viewable field than the PTZ camera, which is an example of the fourth principle. In the current demonstration, the PTZ camera is directly controllable in the exocentric configuration, but control depends on the distance between the virtual viewpoint and the panoramic representation.

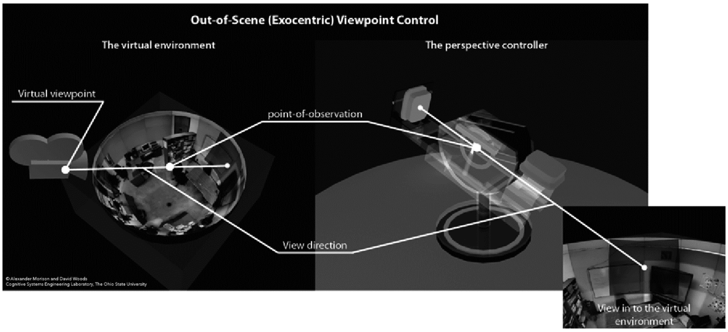


**Fig. 7.** The exocentric relationship between a virtual camera in a virtual environment with a panoramic sensor representation (left) and the viewpoint input device control device (right).

In both exocentric and egocentric configurations of the controller the distance between the view direction block and the center of the controller can be increased or decreased to create movement of the virtual viewpoint towards or away from the virtual point-of-observation. This ability to move the virtual viewpoint through the 3-dimensional virtual environment provides a unique opportunity for sensor selection.

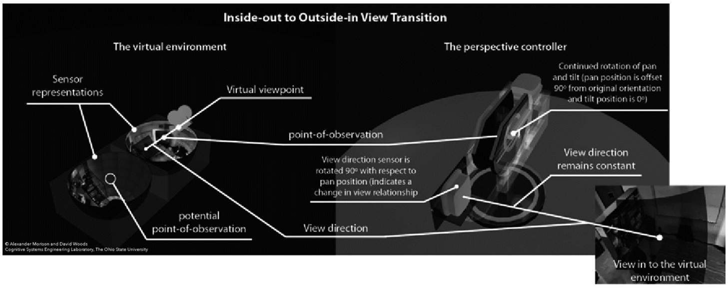


**Fig. 8.** The transition from an egocentric to an exocentric relationship between a virtual camera in a virtual environment with a panoramic sensor representation (left) and the viewpoint input control device (right).

In the exocentric frame-of-reference the distance of the virtual viewpoint from the sensor panoramic representation can be used to define sensor operator control. A sensor operator must be able to select or de-select a camera for control. Often selection of a camera implies de-selection of the currently controlled camera. In this demonstration, the virtual viewpoint distance to a sensor panoramic representation defines PTZ camera control. One option is to encode the virtual distance as a discrete, binary on-off threshold. The camera is selected if the distance between the virtual viewpoint and the panoramic representation is less than the binary threshold. Otherwise, the camera is not selected. An alternative encoding is to use distance as a continuous dimension. The closer the virtual viewpoint is to the panoramic representation the more frequently the PTZ camera samples from the desired view direction.

The remaining component of the two frames-of-reference is the mechanism for transitioning between referent frames. A snapshot of the 3-dimensional joystick's smooth transition between referent frames is shown in Fig. 8. In this snapshot the transition between the view direction block position and orientation are rotated half way between the positions in the egocentric and exocentric configurations. This transition maintains the virtual view direction while changing the configuration of the 3-dimensional joystick.

## 5 Conclusion

Human-sensor systems are growing in number and kind. But, despite the benefits of these systems, there are several important deficiencies with existing designs. These deficiencies include the inability to keep pace, the keyhole effect, the multiple feeds problem, and the confounding of viewpoint control and action potential. Four principles for successful viewpoint control were introduced. These principles include the need for a moving point-of-observation, to provide two views in parallel, the use of egocentric and exocentric frames-of-reference, and the design of center-surround relationships. These were demonstrated using a wide-area video surveillance network prototype system. The key components of the demonstration were described including egocentric and exocentric viewpoint control, navigation, and sensor control selection. In short, this paper argues that viewpoint control is present in all existing human-sensor systems, but it is either ad-hoc or happenstance. In order to make coherent progress on the described deficiencies of these systems, fundamental advances in human-sensor system design are necessary. The viewpoint control model and demonstration are the beginning of a principled approach to viewpoint control for human-sensor system design.

## References


1. Morison, A.M.: Perspective control: technology to solve the multiple feeds problem in sensor systems. The Ohio State University (2010)
2. Woods, D.D., Sarter, N.B.: Capturing the dynamics of attention control from individual to distributed systems. Theor. Issues Ergon. **11**, 7–28 (2010)
3. Voshell, M., Woods, D.: Breaking the keyhole in human–robot coordination: method and evaluation. In: Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (2005)
4. Morison, A., Woods, D., Murphy, T.: Human-robot interaction as extending human perception to new scales. In: Hoffman, R., Hancock, P., Parasuraman, R., Szalma, J., Scerbo, M. (eds.) The Cambridge Handbook of Applied Perception Research, 2nd edn, pp. 848–868. Cambridge University Press, Cambridge (2015)
5. Woods, D.D., Patterson, E.S., Roth, E.M.: Can we ever escape from data overload? A cognitive systems diagnosis. Cogn. Technol. Work **4**, 22–36 (2002). doi:10.1007/s101110200002
6. Woods, D.D.: Visual momentum: a concept to improve the cognitive coupling of person and computer. Int. J. Man Mach. Stud. **21**, 229–244 (1984)

7. Casper, J., Murphy, R.R.: Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. IEEE Trans. Syst. Man Cybern. Part B **33**, 367–385 (2003)
8. Hughes, S., Lewis, M.: Robotic camera control for remote exploration. Comput. Hum. Interact. **6**, 511–517 (2004)
9. Woods, D.D., Hollnagel, E.: Joint Cognitive Systems: Patterns in Cognitive Systems Engineering. CRC Press, Boca Raton (2006)
10. Guizzo, E.: Fukushima robot operator writes tell-all blog. In: IEEE Spectrum (2011). http://spectrum.ieee.org/automaton/robotics/industrial-robots/fukushima-robot-operator-diaries
11. Simons, D.J., Wang, R.F.: Perceiving real-world viewpoint changes. Psychol. Sci. **9**, 315–320 (1998)
12. Roesler, A.: A new model for perspective: the role of point of observation in virtual and remote perspective-taking. The Ohio State University (2005)
13. Tittle, J.S., Roesler, A., Woods, D.D.: The remote perception problem. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 260–264 (2002)
14. Morison, A.M., Voshell, M., Roesler, A., Feil, M., Tittle, J., Tinapple, D., Woods, D.D.: Integrating diverse feeds to extend human perception into distant scenes. In: McDermott, E.B.Y.P., Allender, L. (eds.) Advanced Decision Architectures for the Warfighter: Foundation and Technology, pp. 177–200. Alion Science and Technology, Boulder (2009)
15. Breazeal, C., Berlin, M., Brooks, A., Gray, J., Thomaz, A.: Using perspective taking to learn from ambiguous demonstrations. Rob. Auton. Syst. **54**, 385–393 (2006)