Integrated Design of Intelligent Surveillance Systems and their User Interface.

Conference Paper · January 2005

Source: DBLP

Alexander Toet
TNO, Soesterberg, The Netherlands

In: H.R. Arabnia & R. Joshua (Eds.), *Proceedings of the 2005 International Conference on Artificial Intelligence IC-AI'05*, Vol. I (pp. 17-23). Las Vegas, Nevada, USA: CSREA Press.

Integrated Design of Intelligent Surveillance Systems and their User Interface

Alexander Toet
TNO Human Factors
P.O. Box 23, 3769 ZG, Soesterberg, The Netherlands
Email: lex.toet@tno.nl

Abstract *– Modern complex surveillance systems consisting of multiple and heterogeneous sensors, automatic information registration and data analysis techniques, and decision support tools should provide the human operator an integrated, transparent and easily comprehensible view of the surveyed scene. Displayed visual information should be optimally tuned to the characteristics and limitations of the human visual system. Interaction with the displayed information should be intuitive and effortless so that operators can focus their attention on their task. We present an overview of (1) display techniques that can be used to optimally adjust visual information presentation to the capabilities of the human visual system and the momentary direction of viewing, and (2) a range of techniques that allow optimal interaction with the displayed information. We argue that experts from the application domain, system designers and human factor specialists should all be involved in the design process of a complex surveillance system that incorporates these techniques, to ensure that the final product fully exploits all the technological opportunities and provides a cognitive fit between the displayed information and the operator's mental model.*

1 INTRODUCTION

A human operated surveillance system should provide transparent, accurate, meaningful and timely security-related information. An intelligent surveillance system could support the operator by guiding his attention to potentially interesting objects or events, and by providing a clear explanation why it has classified these items as being interesting (i.e. it should provide easy access to the metadata information associated with registered events). The construction of such a system requires an integrated design of the user interface to ensure that the user can optimally interact and interrogate the available information.

Human operators performing surveillance tasks should have a complete and fully understood image of the situation at hand and knowledge of similar situations that occurred in the past as a basis to comprehend the current situation and all its implications. This basis is essential to develop an effective strategy to deal with the observed situation and possibly to neutralize evolving threats in an early stage. If this basis does not exist, an individual will resort to basic instinct and reaction. The literature gives many examples of incidents and accidents, which could have been avoided if operators had recognized the situation in time.

Sophisticated modern surveillance systems can provide the operator with a huge repository of data and information about the environment, and assist the operator through intelligent information presentation modes and decision support tools. However, to the user more data doesn't mean more information. In fact, in some situations, most of the data is merely distracting noise for the operator. Human cognitive capacities that are essential for performing surveillance tasks such as visual attention and information processing resources (e.g. working memory) are inherently limited. Only the information that is relevant for the situation, the tasks and the decisions that need to be made should be presented to the user in a manageable form. Since "information is largely in the eye of the beholder" [18] extracting information from data that supports the needs of the user requires knowledge of the cognitive processes of the user. Task switching increases cognitive workload. However, task switching is less demanding when the tasks are more related. Systems should prioritize information presentation, minimize task switching, guide the user's attention via different modalities, maintain attention, and optimize situational awareness (SA). By integrating cognitive support into the user interface information acquisition and recall can be enhanced [46].

In recent years, a wealth of methods for the efficient and comprehensive visualization of complex and heterogeneous information has been developed. This includes a range of visual dynamic information display techniques that enhance the user's visual capacity by optimally tuning the displayed visual information to the characteristics and limitations of the human visual system. Also, several new interaction modalities have become available. The integration of these new display and interaction techniques will enable intuitive and efficient interaction with the displayed information, and may serve to optimize the user's resources by (1) reducing the attentional workload, (2) adapting the displayed information to the task, and (3) attracting or guiding the user's attention.

In the next sections we will first present an overview of the available display and interaction modalities. Then we will discuss how these techniques can be integrated in perceptual and attentive interfaces. Finally we will argue that a successful integration these techniques depends critically on a close collaboration between experts from the application domain, system designers and human factor specialists [58].

2 DISPLAY TECHNIQUES

Effective visual information presentation strategies may help to improve observer performance in tasks that depend critically on visual information transfer. The total amount of information that can be presented simultaneously by a display medium is restricted by the area of the screen. Traditional display techniques apply windowing techniques and scroll/zoom features in an attempt to overcome this problem. A serious drawback of these methods is that the zoomed regions are displayed in separate windows. This may result in a loss of context (situational awareness), which may in turn degrade user performance in for instance navigational tasks. The amount of detail needed to perform a certain task may vary from different locations on the display. Regions of interest (focus points) should be displayed with more detail. On the other hand, the background serves to provide the global context needed to relate them to each other and can therefore be displayed with less detail. This suggests the use of variable scale displays that simultaneously present enlarged views of the regions of interest with a less detailed (but topologically correct) representation of their context. The scale (the level of detail) of the presented information may vary either gradually or discontinuously over the screen area. Because the display area is fixed variable scale displays inherently induce image distortions. In this section we will briefly review a range of different visual information presentation modes that were designed to optimize visual information transfer on displays with limited screen areas (for an extensive review see: {7007}).

2.1 NON-DISTORTION ORIENTED TECHNIQUES

Initially two main strategies were designed to cope with limited display size. The first approach is to partition or *tile* the screen into a number of non-overlapping windows. This strategy is not feasible when a proliferation of information items needs to be displayed (e.g. windows, menus, dialog boxes, or tool palettes). A non-overlapping tiling may result in items that are simply too small to resolve. The second strategy is to use overlapping windows. Only the top one is visible at any given time, and a mechanism is provided to rapidly change which window is visible (temporal sequencing). This strategy is also undesirable, since overlapping opaque objects obscure portions of information we may need to see. Frequently, a hybrid of the first two strategies is used.

Another approach to optimise visual information presentation on limited display areas is the use of semi-transparency or multi-layer displays [3, 26, 27]. Semi-transparency can create the impression of multiple interface layers on a conventional display. People are able to shift their focus rapidly between the two views, to the point of initiating an action in the foreground layer and continuing it in the background layer. Semi-transparency was readily and early adopted in computer games like Diablo II [10] and Everquest [3]. In this type of games the degree of transparency of a component depends on the amount of attention paid to it by the user.

Semi-transparency has been used to create a multi-layer zoomable focus+context display [56]. In this approach the user can selectively enlarge interesting parts of the display, while the initial top-level overview remains visible as a transparent overlay. Thus, the user remains aware of the context while zooming and scrolling through the display.

Semi-transparency also effectively encodes the spatial relations (depth cues) between objects in 3D space. As a result, human performance in interactive 3D computer graphics environments improves when using semitransparent tools [79].

Special (auto-) stereoscopic displays can be deployed to present semi-transparent interface layers on different perceived depth planes. In this case, the user looks through the interface presented in the foreground to see the interface presented in the background. Ideally, more information will be simultaneously visible using this approach. However, overlapping windows may also interfere with each other.

2.2 DISTORTION-ORIENTED TECHNIQUES

Focus+context displays combine a detailed (full size or enlarged) representation of the regions of interest with a less detailed (compressed) representation of the remaining regions (the background). The low-resolution representation of the background provides the context for the focus regions. Several different types of focus+context displays have been developed (for an overview and taxonomy of distortion oriented display techniques see: [6, 37]).

When focus+context displays are used in an interactive mode users can navigate through the information space by selecting and moving (dragging) the regions of interest, for instance by using a pointing device [20] or by finger pointing or rubbing on a touchscreen [49].

2.3 HYBRID TECHNIQUES: MAGIC LENS FILTERS

Magic lenses are small arbitrarily shaped windows which the user can move around (e.g. by mouse, keyboard- or gaze- control) over the display area in order to inspect certain areas of the display in more detail or to locally reveal additional information [66]. The operators associated with these windows or magic lenses can range from simple image processing operations like local image enlargement (in order to see more detail), contrast enhancement, noise or clutter reduction, to quite general operations, for instance to view different types of information related to the area that is being inspected. In video surveillance applications, additional information may be shown through a magic lens positioned on a tracked subject, whereas the rest of the scene remains unchanged. This serves to prevent severe clutter of the display that would occur if the additional information was shown for all targets in the scene. Also, it eliminates the need for the operator to divert his gaze to a separate screen with additional information. Multiple operators can be applied in a single magic lens. In the previous example, the magic lens positioned on a tracked target may not only represent additional (alpha numeric) intelligence information (e.g. identity, age, nationality), but may also serve to enhance the image (contrast, magnification), or represent the information from different sensor modalities (e.g. millimetre wave imagery showing hidden metal objects).

Magic lenses have a number of potential advantages over traditional methods of generating alternate views and filtering information. Binding the filter to a spatially bounded, movable region creates an easily understood user model based on experience with physical lenses (using these filters is just like moving a looking glass over a newspaper). Limiting the view to a local region preserves context and can reduce clutter. Lenses can be parameterized, and can have arbitrary shape. The user can apply different operators simultaneously over different parts of the displayed information to get multiple, simultaneous views. Lenses that overlap combine their effects (e.g. by multiplying their magnification), making it easy to create visual macros [7]. These macros can be temporary, or can be "welded together" to create a compound lens that encapsulates a set of operators and parameters. Finally, the magic lens metaphor can be used uniformly across applications.

2.5 Gaze Contingent Multiresolution Displays

Gaze-contingent multiresolutional displays center high-resolution information on the user's gaze position, matching the user's area of interest ([53, 57]; for recent overviews see: [15, 53, 57]). Image resolution and details outside the area of interest are reduced, lowering the requirements for processing resources and transmission bandwidth in demanding display and imaging applications. Gaze-contingent displays integrate a system for tracking viewer gaze position (by combined eye and head tracking) with a display that can be modified in real time to center the area of interest at the point of gaze. Online analysis of the observer's patterns of eye movements may allow more sophisticated interactions than simple foveation, like zooming-in and other computer interface controls [24].

2.6 Gaze Contingent Stereoscopic Displays

A novel gaze contingent display technique uses real-time stereoscopic gaze-tracking to measure the observer's 3-D fixation point and enhance the appearance of the volumetric image representation of the region around this point [34].

2.7 Gaze Contingent Multimodal Displays

Displays with integrated eye trackers are highly suitable for gaze contingent display of multimodal images. An observer can for instance use such a system to inspect a night vision surveillance scene. His fixation location on the screen is registered by the system and can for instance drive a magic lens, that can locally display each of the available individual image modalities or fused combinations thereof. In combination with another input interface modality (e.g. speech or gesture recognition, a pointing device or just keyboard buttons) the observer can indicate what information about the fixated region should be displayed.

2.8 Gaze Contingent Multilayer Displays

Another option is to equip a two-layer display with an integrated eye tracker. With such a configuration the observer can easily indicate the region of interest merely by looking at it and tell the system (e.g. through speech or by pressing some buttons) to place this region in a different depth layer to facilitate later retrieval.

3 Interaction Techniques

Visual information display techniques may be more effective when the user can dynamically adjust the display mode to his momentary information requirements during each stage of a task. The interactive adjustment can for instance be done through keyboard commands, mouse or hand pointing, voice commands, or gaze directing. The use of multimodal interfaces, that combine two or more input modalities in a coordinated way, may help to reduce the ambiguity of the individual input devices and to enhance the interaction speed [50-52]. Different input modalities that have recently been applied to interact with dynamic visual information displays are: gaze control, gesture, speech and facial expression recognition.

Gaze control is significantly faster than by mouse control for object selection [48], especially with large display screens and virtual environments and for relatively large targets [61, 62]. However, functions like dragging and double clicking are better performed by manual pointing [33]. It is far more natural to manipulate an object with the hand than with gaze. This suggests that gaze and mouse control should be used in a complementary fashion [59]. In this way, mouse pointing may serve to disambiguate user input, e.g. when selecting small targets [77]. In this case eye gaze controls the "ballistic" movements of the cursor from one screen location to a region of tolerance in which the final target point is located, whereas manual control serves to compensate for over- and undershoots and lock the cursor onto the target [78, 80].

Eyesight and speech are two channels that humans naturally use to communicate, especially when their hands are occupied. Recognition ambiguities of speech and gaze inputs are inevitable. Combining the two overcomes imperfections of recognition techniques, compensates for drawbacks of single mode, supports mutual correction of individual input signals, reduces the error rate and improves interaction efficiency [81]. For instance, eye gaze provides additional information that can be used to disambiguate speech, allowing an operator's verbal commands to be directed to the appropriate receiver [23, 43]. Another benefit of a combined gaze and speech multimodal input device is that it allows simplified descriptions, which contributes to both error avoidance and user's acceptance. Voice commands in combination with speech recognition can for also be used to zoom and scroll images, or direct the focus of interest in focus+context displays.

Gaze direction and facial expression that is subserved by the underlying facial muscle activity are used frequently and fully automatically in human interaction. Some recent studies investigate the feasibility of combining voluntarily eye movements and voluntarily produced changes in the level of electrical activity of facial muscles as a new human–computer interaction technique [54, 67]. The voluntary use of facial muscle *corrugator supercilii* works well for clicking as a counterpart for the mouse button press. With the new technique the user's hands are left free for other purposes.

Pointing in 3D to interact with objects or to change the viewpoint of cameras can be done with 3-D tracked mouse devices, with data gloves, or through hand tracking using 3-D computer vision techniques. In virtual environments hand-based pointing is faster than gaze-based pointing for distant objects [9].

4 INTEGRATED INTERFACES

The main console of a complex surveillance system should provide an intuitive, meaningful overview of the situation that can easily be absorbed and assimilated by the human operator in time critical situations. It should also provide an integrated access to the heterogeneous sources of detailed information, together with intuitive and efficient tools to arbitrarily navigate and make semantic connections between these different information sources. To reduce the operator' workload, systems should prioritize information presentation, minimize task switching, and guide the user's attention via different modalities. Finally, it should include efficient techniques to present the information to the observer. To some extent these goals can be achieved by integrating the presentation and interaction techniques presented earlier in this study. In the rest of this section we present three different and complementary integration approaches: a simple combination of different interaction and presentation techniques, perceptual interfaces that observe the operator and respond to his actions, and attentive interfaces that also model the user to anticipate his needs.

4.1 MULTIPLE INTERFACES

Complex surveillance systems generally consist of a main console that provides a global representation of the situation, in combination with multiple peripheral displays that provide detailed information on different aspects of the observed scene. The integration of a magic lens filter or focus+context function with such a system allows the operator to display and inspect on a peripheral screen details that have been observed and indicated on the main console. For instance, a suspect individual that is indicated by the operator on an overview of a scene may be displayed in detail and tracked on a separate screen. Or, when using an intelligent system with information support, the operator may indicate multiple individuals and the system can visualize (e.g. alpha numerical or graphical data) the interrelations of these persons on a separate screen.

In practice surveillance operators frequently use the logical relationships between different sensors [39]. A surveillance system should therefore allow sets of (disperse and heterogeneous) sensors to be grouped [72]. A system with multiple screens allows the simultaneous representation of signals of multiple logically related sensors and their geographical or logical relation. For instance, the location of all sensors and their coverage (overlap) may be indicated on a schematic map of the surveillance zone, and the operator can select a subset of these sensors (cameras) e.g. by clicking on their icons on the schematic map to view their output (images) on additional screens [72].

Multiple coordinated visualizations enable users to rapidly explore complex information. An system that enables operators to rapidly and dynamically mix and match visualizations and co-ordinations to construct custom exploration interfaces is therefore a powerful surveillance tool [47].

A system using multiple displays may deploy tools to attract and guide the attention of the operator. For instance, when the system detects a suspicious event it may blink an icon on the main console to attract the attention of the operator, and it may guide her attention to the peripheral screen that represents the event in more detail (e.g. though intensity modulations of that screen, or through 3D audio cues).

4.3 PERCEPTUAL INTERFACES

A perceptual interface allows the user to interact with a system through other means than the normal keyboard and mouse. These interfaces are realised by giving the computer the capability of to observe and interpret the user's (eye-)movements or voice commands.

One of the most promising interaction techniques is based on the registration of the operator's gaze behavior [68]. The human gaze reveals information about the user's intention and attention. It is a potential porthole into his current cognitive processes. The human fixation behavior over time reveals information on the cognitive state of the user such as confusion or fatigue, or on his degree of expertise. If a computer knows where the user fixates it can react by taking appropriate actions like presenting the information in an adapted form, presenting additional information or activating fixated items. Gaze directed displays are therefore an important enabling technology for attention aware systems .

Operators in complex event-driven domains face considerable and often competing attentional demands. The capacity to attend to several objects simultaneously decreases as a function of visual complexity [1]. Adding more detail to a display (thereby increasing its complexity or the amount of clutter) places a greater demand on the user to actively ignore task-irrelevant features, which in turn can lead to a decreased awareness of all unattended features [44, 63]. In situations where multiple events occur simultaneously, operators can fail to detect important changes even when they are not fatigued, stressed or multitasking (change blindness, e.g. [16]). A gaze-based perceptual interface may diminish these effects by temporarily eliminating less relevant peripheral details (reducing the amount of clutter) during changes, thereby increasing the chance that they will be noticed by the user. In general, dynamical filtering of information on the basis of user interest allows cognitive load associated with complex displays to be managed more effectively.

A semi-transparent or multi-layer display (e.g. [11]) can be turned into a perceptual interface by integration with a gaze tracking device. In case of semi-transparent displays, the transparency level of regions on which an observer dwells may automatically change (turning it more opaque) in response to the user's viewing behaviour. In case of multiple depth layer displays, the depth layer on which the observer actually fixates may turn opaque, whereas the layers in front may become fully transparent. Also, fixated objects may be transferred to a different (more prominent) depth plane. This may be a useful feature for surveillance systems and complex control displays. For instance, when monitoring a large crowd, previously inspected suspect individuals may be transferred to a front plane, making them easier to track. Another example is a complex plant control display, where it may be useful make control items that require frequent inspection more prominent by transferring them to separate depth planes.

By integrating a gaze tracking device with a surveillance system that deploys multiple displays, the fixation behavior of the user on the central display can be used to select the additional task supporting information that is provided on the peripheral displays [71]. Crucial information on the central display remains visible at all times, whereas the supporting information can be updated and altered either automatically (using gaze registration) or interactively (e.g. though a combination of button pressing or voice commands and gaze control).

4.4 Attentive Interfaces

An attentive (or reactive) interface dynamically prioritizes the information it presents to its users, such that information processing resources of both user and system are optimally distributed across a set of tasks [73, 74]. More precisely, attentive user interfaces

- monitor user behavior,
- model user goals and interests,
- anticipate user needs,
- provide users with information, and
- interact with users.

Attentive user interfaces are related to perceptual user interfaces, which incorporate multimodal input, multimedia output, and human-like perceptual capabilities to create systems with natural human-computer interactions [52, 70]. Whereas the emphasis of perceptual user interfaces is on coordinating perception in human and machine, the emphasis of attentive user interfaces is on directing attention in human and machine. For a system to attend to a user, it must not only perceive the user but it must also anticipate the user. The key lies not in how it picks up information from the user or how it displays information to the user; rather, the key lies in how the user is modelled and what inferences are made about the user.

Attentive displays are particularly useful in domains with tasks that require visual, spatial and causal reasoning. These domains share five characteristics [45]:

1. objects of the domain are spatially distributed;
2. the domain is dynamic, i.e. objects and their properties change over time;
3. objects causally interact with each other;
4. such interactions can be traced along chains of cause-effect relationships that branch and merge in spatial and temporal dimensions; and
5. predicting the future evolution of a system in the domain requires reasoning from a given set of initial conditions and inferring these causal chains of events.

The domain of visual surveillance satisfies all these criteria. An example of task in video surveillance is tracking a dispersed group of suspect individuals, where a range of different intelligence information from a variety of different sources (cameras, flight plans, biometric sensors) may be used to construct the overall recognised picture. In these types of tasks, an attentive interface must leverage knowledge about the task that the user is engaged in and the trajectory of the user's attention shifts in order to provide the right information in the right place and at the right time.

User behaviour may be monitored, for example, by video cameras to watch for certain sorts of user actions such as eye movements [33, 80] or hand gestures [4], by microphones to listen for speech or other sounds [52], by monitoring heart rate variability or motor activity using electroencephalogram analysis [8], or by a computer's operating system to track keystrokes, mouse input, and application use [29, 38, 41, 42]. User goals and interests may be modelled using Bayesian networks [29], predefined knowledge structures [60], or heuristics [41, 42]. User needs may be anticipated by modelling task demands [60]. Information may be delivered to users by speech or by text [41, 52], and users may interact directly through eye gaze, gestures or speech [4, 33, 65, 80]. By statistically modelling the interactive user behaviour, attentive displays may establish the urgency and relevance of the displayed information in the context of current activity. They may use this information to adjust their renderings to provide peripheral context in support of focused activity. The optimal allocation of attentional resources requires careful interruption management [8]. Poorly designed attentive interfaces can be counterproductive if they distract the user or make false inferences about the user's needs and goals.

Gaze contingent displays can be deployed to actively guide the attention of observers to certain parts of a scene. In surveillance applications a gaze contingent display may (1) draw the operator's attention to unusual activities or suspect individuals, (2) alert the observer in case his attention fades (which the system can assess by monitoring his fixation behavior over time), and (3) provide additional information related to the inspected location.

Intelligent user interfaces are human-machine interfaces that aim to improve the efficiency, effectiveness, and naturalness of human-machine interaction by representing, reasoning, and acting on models of the user, domain, task, discourse, and media (e.g., graphics, natural language, gesture). Intelligent user interfaces are multifaceted, in purpose and nature, and include capabilities for multimedia input analysis, multimedia presentation generation, model-based interfaces, agent-based interfaces, and the use of user, discourse and task models to personalize and enhance interaction. Such an intelligent interface might anticipate the information required for a particular task that the user is performing and provide the user with only the necessary information in the most appropriate manner, perhaps modified by known user preferences for receiving data of that sort. The goal would be to ensure that the user retains an overall situational awareness of the incident progression, while at the same time having access to the detailed information required to solve a particular problem or perform a specific task in the most efficient and effective manner. Intelligent displays based on the information available at any particular time thus would filter, order or highlight that information to allow it to be used most effectively by the particular responder. The underlying idea is that the user and the agents collaborate effectively and in complete harmony to achieve the user's goals.

For instance, an intelligent attentive interface may dynamically mix and match visualizations and coordinations to construct custom exploration interfaces like multiple coordinated visualizations that enables the operator to rapidly explore complex information [47]. Such a system may notice a potential threat and proactively present relevant related information in optimal display modes.

Perceptual intelligent interfaces are attentive multimodal interfaces that learn. They adapt their behaviour to suit the user, rather than the other way around, and they do so by paying attention to the user and his surroundings in the same way another person would do [55]. They enable rich, natural, and efficient interaction with computers, by leveraging sensing (input) and rendering (output) technologies [69]. These interfaces learn by tracking the observer and his interactions with the environment over time. As a result they can anticipate the user's actions and his information requirements.

5 Concluding Remarks

We presented an overview of (1) display techniques that can be used to optimally adjust visual information presentation to the capabilities of the human visual system and the momentary direction of viewing, and (2) a range of techniques that allow optimal interaction with the displayed information. We discusssed some approaches to integrate these techniques in the design of complex surveillance systems.

Interface design is the process of shaping displays and controls so that they provide information or interaction capabilities for a user [76]. How the interface tools are built can have serious consequences for data manipulation. Controls need to be matched to the task. The ability to locate information depends on the ease with which large data-sets can be manipulated. In order for people to use computer databases for retrieving relevant information in a timely fashion, the information structures must fit with user's expectations about the space. It is critical to make relevant information salient without producing information overload, and that requires substantial attention to the structure of information displays or other forms of communication. The way in which displays and functions interact has been shown to have life and death consequences in such diverse areas as aviation and medicine. To know what information and interactions are needed or helpful, the designer must know what data are pertinent to the observer's needs, intentions, expectations, and interests, in interaction with some system and in what order and relations [75]. Current trends in interface design have progressively recognized the importance of cognitive factors. People should be able to operate systems in a natural (intuitive) way so that they can focus their attention on their task instead of struggling with the technology itself.

The main challenge for designers of complex surveillance systems is to determine the appropriate information users need to achieve SA, and to deliver the right information and support in the right form at the right time. Therefore the system should be designed such that it integrates and represents the available data in meaningful way so that it can be easily absorbed and assimilated by the human operator in time critical situations. A structured approach is required to incorporate SA considerations into the design process, including a determination of SA requirements (cognitive task analysis), designing for SA enhancement [17], and measurement of SA in design evaluation (preferably during the design process) [18]. A primary benefit of examining system design from the perspective of operator situation awareness is that the impact of design decisions on situation awareness can be objectively assessed as a measure of quality of the integrated system design when used within the actual challenges of the operational environment. This may provide useful diagnostic information that can be used to determine what kind of support operators need. Examples of successful application of cognitive engineering are for instance the design of a naval combat management workstation [71] and a new air control system [19]. In both cases, novel information methods were recommended and designed to enhance the operator's situational awareness. Situation awareness is the *perception* of the elements in the environment within a volume of time and space, the *comprehension* of their meaning and interrelations, and the *projection* of their status in the near future [17]. In other words, SA simply means that one knows exactly what is going on around oneself. In operational terms, this means that one has received and comprehended all the information that is relevant for a given task. The basis of SA is knowledge of the objects in a given region. Although knowledge of the objects is essential, it does not by itself constitute SA. It is also necessary to know all the relations between the objects that are relevant to the current operation. This requires situational analysis: a full analysis of all perceived objects and their interrelations [58]. The number of relations increases exponentially with the number of objects. Although modern surveillance systems include a range of decision support tools, the full recognized picture (full SA) is still only achieved in the mind of the human operator. In other words, SA is a product of mental data fusion. The system interfaces should therefore be designed such that the human operator can achieve SA with minimal workload in a timely fashion. This implies that the data should be presented in a transparent, intuitive and easily accessible way. The presentation of new information should minimally interrupt task performance. The system should provide tools to interact with the data in an ergonomic and intuitive way. The interaction modes may include such modalities as: gesture/speech recognition, eye gaze, lip reading and even biofeedback mechanisms attached to the operators themselves. The integration of multimodal inputs for human machine interaction can be approached from the viewpoint of multiple information source fusion. Different information sources can be related to different interface modalities. Selection of the appropriate modalities to integrate and at what level to integrate them should be part of the process of developing an interoperable multi-disciplinary situational awareness fusion system [14]. Currently, the component technologies of intelligent surveillance systems are developed in isolation, and the user interface is usually not considered in the design process. Today's user interfaces are often the bottleneck in effectively and efficiently utilizing the available information flow for decision making. As a result, operators are forced to function as human data integrators rather than as decision makers. New capabilities must be developed to visualize and interact with information in order to provide an integrated situation awareness picture to the operator. To ensure that the design of a complex surveillance system both exploits all the technological opportunities and to ensure the cognitive match between the system and its operators the design team should include experts from the application domain, system designers and human factor specialists [58].

REFERENCES

[1] Alvarez, G.A. & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15(2),* 106-111.

[2] Baudisch, P., Good, N. & Stewart, P. (2001). Focus plus context screens: combining display technology with visualization techniques.*Proceedings of the 14th annual ACM symposium on user interface software and technology* (pp. 31-40). New York, USA: ACM Press.

[3] Beverly, L., Harrison, B.L. & Vicente, K.J. (1996). An experimental evaluation of transparent menu usage.*Proceedings of the SIGCHI conference on Human factors in computing systems: common ground* (pp. 391-398). New York, USA: ACM Press.

[4] Bolt, R.A. (1980). "Put-that-there": Voice and gesture at the graphics interface.*Proceedings of the 7th annual conference on Computer graphics and interactive techniques* (pp. 262-270). New York, USA: ACM Press.

[5] Carpendale, M.S.T., Cowperthwaite, D.J. & Fracchia, F.D. (1997). Extending distortion viewing from 2D to 3D. *IEEE Computer Graphics and Applications, 17(4),* 42-51.

[6] Carpendale, M.S.T. & Montagnese, C. (2001). A framework for unifying presentation space.*Proceedings of ACM Conference on User-Interface Software Technology* (pp. 61-70). New York, USA: ACM Press.

[7] Carpendale, S., Ligh, J. & Pattison, E. (2004). Achieving higher magnification in context.*Proceedings of the 17th annual ACM symposium on User interface software and technology* (pp. 71-80). New York, USA: ACM Press.

[8] Chen, D. & Vertegaal, R. (2004). Using mental load for managing interruptions in physiologically attentive user interfaces.*Extended Abstracts of ACM CHI 2004 Conference on Human Factors in Computing Systems* (pp. 1513-1516). New York, USA: ACM Press.

[9] Cournia, N., Smith, J.D. & Duchowski, A.T. (2003). Gaze- vs. hand-based pointing in virtual environments.*Proceedings of the Conference on Human Factors in Computing Systems* (pp. 772-773). New York, USA: ACM Press.

[10] Cox, D.A., Chugh, J.S., Gutwin, C. & Greenberg, S. (1998). The usability of transparent overview layers.*CHI 98 conference summary on Human factors in computing systems* (pp. 301-302). New York, USA: ACM Press.

[11] Deep Video Imaging Ltd. (2004). *Interactive dual plane imagery*. Website: http://www.Deepvideo.com.

[12] Dorr, M. (2004). Effects of gaze-contingent stimuli on eye movements. Lübeck, Germany: Institut für Neuro- und Bioinformatik, Universität zu Lübeck.

[13] Dorr, M., Martinetz, T., Gegenfurtner, K.R. & Barth, E. (2004). Guidance of eye movements on a gaze-contingent display.*Proceedings of the 5th Workshop on Dynamic Perception* (pp. In press- Tübingen, Germany.

[14] Drozd, A.L., Carroll, C.E., Kasperovich, Jr., Kasperovich, I., Varshney, P.K., Weiner, D.D. & Alford, M.G. (2001). Towards the development of multi-sensor integrated display systems.*Proceedings of the 4th International Conference on Information Fusion* (pp. ThC24-1-ThC24-8). Montreal, Canada: International Society of Information Fusion.

[15] Duchowski, A.T., Cournia, N. & Murphy, H. (2004). Gaze-contingent displays: a review. *CyberPsychology & Behavior, 7(6),* 621-634.

[16] Durlach, P.J. (2004). Change blindness and its implications for complex monitoring and control systems design and operator training. *Human-Computer Interaction, 19(4),* 423-451.

[17] Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37(1),* 32-64.

[18] Endsley, M.R. (2001). Designing for situation awareness in complex system.*Proceedings of the second international workshop on symbiosis of humans, artifacts and environment* (pp. 1-14). Kyoto, Japan: Japan Society for the Promotion of Science (JSPS).

[19] Endsley, M.R., Mogford, R.H. & Stein, E.S. (1997). Controller situation awareness in free flight.*Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 4-8). Santa Monica, CA: Human Factors and Ergonomics Society.

[20] Flider, M.J. & Bailey, B.P. (2004). An evaluation of techniques for controlling focus+context screens.*Proceedings of the 2004 conference on Graphics Interface* (pp. 135-144). New York, USA: ACM Press.

[21] Furnas, G.W. (1982). *The FISHEYE view: A new look at structured files* (Report Technical Memorandum 82-11221-22). Bell Laboratories.

[22] Furnas, G.W. (1986). Generalized fisheye views.*Human Factors in Computing Systems CHI '86* (pp. 16-23). Boston, USA.

[23] Glenn, F.A., Iavecchia, H.P., Ross, L.V., Stokes, J.M., Weiss, D. & Zakland, A.L. (1986). Eye-voice controlled interface. *Proceedings of the Human Factors Society,* 322-326.

[24] Goldberg, J.H. & Schryver, J.C. (1995). Eye-gaze-contingent control of the computer interface: Methodology and example for zoom detection. *Behavior Research Methods, Instruments and Computers, 27(3),* 338-350.

[25] Gutwin, C. (2002). Improving focus targeting in interactive fisheye views.*Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'02)* (pp. 267-274). New York, USA: ACM Press.

[26] Harrison, B.L., Ishii, H., Vicente, K.J. & Buxton, W.A.S. (1995). Transparent layered user interfaces: an evaluation of a display design to enhance focused and divided attention.*Human Factors in Computing Systems: Proceedings of CHI'95.* (pp. 317-324). New York, USA: ACM Press.

[27] Harrison, B.L., Kurtenbach, G. & Vicente, K.J. (1995). An experimental evaluation of transparent user interface tools and information content.*Proceedings of the 8th annual ACM symposium on User interface and software technology* (pp. 81-90). New York, USA: ACM Press.

[28] Horvitz, E. (1999). Principles of mixed-initiative user interfaces.*Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit* (pp. 159-166). New York, USA: ACM Press.

[29] Horvitz, E., Breese, J., Heckerman, D., Hovel, D. & Rommelse, K. (1998). The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users.*Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* (pp. 256-265). San Francisco, CA: Morgan Kaufmann.

[30] Hyrskykari, A. (1997). Gaze control as an input device. In K.-J. Räihä (Ed.), *Proceedings of ACHCI'97: Advanced Course on Human-Computer Interaction* .

[31] Jacob, R.J.K. (1990). What you look at is what you get: eye movement-based interaction techniques.*Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people* (pp. 11-18). New York, USA: ACM Press.

[32] Jacob, R.J.K. (1991). The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems, 9(2),* 152-169.

[33] Jacob, R.J.K. (1993). Eye movement-based human-computer interaction techniques: toward non-command interfaces. In H.R. Hartson & D. Hix (Ed.), *Advances in Human-Computer Interaction* (pp. 151-190). Norwood, NJ, USA: Ablex Publishing Co.

[34] Jones, M.G. & Nikolov, S.G. (2000). Volume visualisation via region-enhancement around an observer's fixation point.*Proceedings of the International Conference on Advances in Medical Signal and Information Processing (MEDSIP 2000)* (pp. 305-312). Herts, UK: The Institution of Electrical Engineers IEE.

[35] Kadmon, N. & Shlomi, E. (1978). A polyfocal projection for statistical surfaces. *The Cartographic Journal, 15(1),* 36-41.

[36] Leung, Y.K. (1989). Human-computer interface techniques for map based diagrams. In G. Salvendy & M. Smith (Ed.), *Proceedings of the Third International Conference on Human-Computer Interaction* (pp. 361-368). Amsterdam, The Netherlands: Elsevier.

[37] Leung, Y.K. & Apperley, M.D. (1994). A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction, 1(2),* 126-160.

[38] Linton, F., Joy, D. & Schaefer, H. (1999). Building user and expert models by long-term observation of application usage.*Proceedings of the Seventh International Conference on User Modeling* (pp. 129-138).http://www.cs.usask.ca/UM99/papers.shtml.

[39] Luff, P., Heath, C.C. & Jirotka, M. (2000). Surveying the scene: Technologies for everyday awareness and monitoring in control rooms. *Interacting with Computers, 13(2),* 193-228.

[40] Mackinlay, J.D., Roberston, G.G. & Card, S.K. (1991). The Perspective Wall: detail and context smoothly integrated.*Proceedings of the Conference on Human Factors in Computing Systems `91 (CHI '91)* (pp. 173-179). New York, USA: ACM Press.

[41] Maglio, P.P., Barrett, R., Campbell, C.S. & Selker, T. (2000). SUITOR: an attentive information system.*Proceedings of the 5th international conference on Intelligent user interfaces* (pp. 169-176). New York, USA: ACM Press.

[42] Maglio, P.P., Campbell, C.S., Barrett, R. & Selker, T. (2001). An architecture for developing attentive

information systems. *Knowledge-Based Systems, 14(1-2),* 103-110.

[43] Maglio, P.P., Matlock, T., Campbell, C.S., Zhai, S. & Smith, B.A. (2000). Gaze and speech in attentive user interfaces.*Proceedings of the International Conference on Multimodal Interfaces* Springer-Verlag.

[44] Most, S.B., Simons, D.J., Scholl, B.J., Jimenez, R., Clifford, E. & Chabris, C.F. (2001). How not to be seen: the contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychological Science, 12(1),* 9-17.

[45] Narayanan, N.H. & Yoon, D. (2003). Reactive information displays.*Proceedings of INTERACT 2003: Ninth IFIP TC 13 International Conference on Human-Computer Interaction* (pp. 244-251). Amsterdam, The Netherlands: IOS Press.

[46] Neerincx, M.A. (2003). Cognitive task load analysis: allocating tasks and designing support. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 283-305). Mahwah, NJ: Lawrence Erlbaum Associates.

[47] North, C. & Schneiderman, B. (2000). Snap-together visualization: a user interface for coordinating visualizations via relational schemata.*Proceedings of the working conference on Advanced visual interfaces* (pp. 128-135). New York, USA: ACM Press.

[48] Ohno, T. (1998). Features of eye gaze interface for selection tasks.*Proceedings of The Third Asia Paicific Computer Human Interaction APCHI'98* (pp. 1-6). Washington, USA: IEEE Computer Society.

[49] Olwal, A. & Feiner, S. (2003). Rubbing the fisheye: precise touch-screen interaction with gestures and fisheye views.*Conference Supplement of UIST '03 (ACM Symposium on User Interface Software and Technology)* (pp. 83-84). New York, USA: ACM Press.

[50] Oviatt, S. (1996). Multimodal interfaces for dynamic interactive maps.*Proceedings of the SIGCHI conference on Human factors in computing systems: common ground* (pp. 95-102). New York, USA: ACM Pres.

[51] Oviatt,S. (2003). Multimodal Interfaces. In J.Jacko & A.Sears (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications.* (pp. 286-304). Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

[52] Oviatt, S. & Cohen, P. (2000). Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM, 43(3),* 45-53.

[53] Parkhurst, D.J. & Niebur, E. (2002). Variable-resolution displays: a theoretical, practical and behavioral evaluation. *Human Factors, 44(4),* 611-629.

[54] Partala, T., Aula, A. & Surakka, V. (2001). Combined voluntary gaze direction and facial muscle activity as a new pointing technique. In M. Hirose (Ed.), *Proceedings of INTERACT 2001* (pp. 100-107). Amsterdam, The Netherlands: IOS Press.

[55] Pentland, A. (2004). Perceptual intelligence. *Communications of the ACM, 43(3),* 35-44.

[56] Pook, S., Lecolinet, E., Vaysseix, G. & Barillot, E. (2000). Context and interaction in zoomable user interfaces.*Proceedings of the working conference on Advanced visual interfaces* (pp. 227-231). New York, USA: ACM Press.

[57] Reingold, E.M., Loschky, L.C., McConkie, G.W. & Stampe, D.M. (2003). Gaze-contingent multiresolutional displays: an integrative review. *Human Factors, 45(2),* 307-328.

[58] Roy, J. (2001). From data fusion to situation analysis.*Proceedings of the 4th International Conference on Information Fusion* (pp. ThC21-1-ThC21-8). Montreal, Canada: International Society of Information Fusion.

[59] Salvucci, D.D. & Anderson, J.R. (2000). Intelligent gaze-added interfaces.*Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 273-280). New York, USA: ACM Press.

[60] Selker, T. (1994). COACH: A teaching agent that learns. *Communications of the ACM, 37(1),* 92-99.

[61] Sibert, L.E. & Jacob, R.J.K. (2000). Evaluation of eye gaze interaction.*Proceedings of the ACM CHI 2000 Human Factors in Computing Systems Conference* (pp. 281-288). New York, USA: Addison-Wesley/ACM Press.

[62] Sibert, L.E., Jacob, R.J.K. & Templeman, J.N. (2001). *Evaluation and analysis of eye gaze interaction* (Report NRL Report NRL/FR/5513--01-9990). Washington, D.C.: Naval Research Laboratory.

[63] Simons, D.J. & Chabris, C.F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception, 28(9),* 1059-1074.

[64] Spence, R. & Apperley, M. (1982). Database navigation: an office environment for the professional. *Behaviour and Information Technology, 1,* 43-54.

[65] Starker, I. & Bolt, R.A. (1990). A gaze-responsive self-disclosing display.*Proceedings of the SIGCHI conference on human factors in computing systems: empowering people* (pp. 3-10). New York, USA: ACM Press.

[66] Stone, M.C., Fishkin, K. & Bier, E.A. (1994). The movable filter as a user interface tool.*Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence* (pp. 306-312). New York, USA: ACM Press.

[67] Surakka, V., Illi, M. & Isokoski, P. (2004). Gazing and frowning as a new human-computer interaction technique. *ACM Transactions on Applied Perception, 1(1),* 40-56.

[68] Toet, A. (2005). Gaze directed displays as an enabling technology for attention aware systems. *Computers in Human Behavior, In press.*

[69] Turk,M., & Kölsch,M. (2004). Perceptual interfaces. In G.Medioni & S.B.Kang (Eds.), *Emerging Topics*

in Computer Vision. (pp. 455-519). Prentice Hall PTR.

[70] Turk, M. & Robertson, G. (2000). Perceptual user interfaces. *Communications of the ACM, 43(3),* 32-34.

[71] van Delft, J.H. & Schraagen, J.M. (2004). Decision support interface.*Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics* (pp. 827-832). Washington, USA: IEEE Press.

[72] Velastin, S.A., Boghossian, B., Lo, B. & Vicencio-Silva, M.A. (2005). PRISMATICA: Toward ambient intelligence in public transport environments. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans, 35(1),* 164-182.

[73] Vertegaal, R. (2002). Designing attentive interfaces.*Proceedings of the symposium on eye tracking research & applications* (pp. 23-30). New York, USA: ACM Press.

[74] Vertegaal, R. (2003). Attentive user interfaces. *Communications of the ACM, 46(3),* 30-33.

[75] Woods, D. (1986). Paradigms for intelligent decision support. In E. Hollnagel, G. Mancini & D. Woods (Ed.), *Intelligent decision support in process environments* (pp. 153-173). New York, USA: Springer-Verlag.

[76] Woods, D. (2005). The cognitive engineering of problem representations. In G. Weir & J. Alty (Ed.), *Human-computer interaction and complex systems* (pp. 169-188). London, UK: Academic Press.

[77] Yamato, M., Monden, A., Matsumoto, K., Inoue, K. & Torii, K. (2000). Button selection for general GUIs using eye and hand together.*Proceedings of the working conference on Advanced visual interfaces* (pp. 270-273). New York, USA: ACM Press.

[78] Zhai, S. (2003). What's in the eyes for attentive input. *Communications of the ACM, 46(3),* 34-39.

[79] Zhai, S., Buxton, W. & Milgram, P. (1996). The partial-occlusion effect: utilizing semitransparency in 3D human-computer interaction. *ACM Transactions on Computer-Human Interaction, 3(3),* 254-284.

[80] Zhai, S., Morimoto, C. & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing.*Proceedings of the CHI'99: ACM Conference on Human Factors in Computing Systems* (pp. 246-253). New York, USA: ACM Press.

[81] Zhang, Q., Imamiya, A., Go, K. & Mao, X. (2004). A gaze and speech multimodal interface.*Proceedings of the 24th International Conference on Distributed Computing Systems Workshops (ICDCSW'04)* (pp. 208-214).IEEE Computer Society.