AFRL-RH-TR-WP-2012-0113

FEASIBILITY OF ARTIFICIAL ATTENTION AT BEYOND-HUMAN-SCALES

Dr. David Woods
Dr. Alexander Morison

Cognitive Systems Engineering Laboratory
Ohio State University
470 Hitchcock Hall
2070 Neil Ave.
Columbus OH-43210

Final Report
April 2012

AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (https://www.dtic.mil).

AFRL-RH-TR-WP-2012-0113 has been reviewed and is approved for publication in accordance with assigned distribution statement.

//signed//	//signed///
Nicole M. McCammon, 1st Lt Work Unit Manager Human-Analyst Augmentation Branch Human Effectiveness Directorate 711th Human Performance Wing Air Force Research Laboratory	Louise Carter. PhD., DR IV Chief, Forecasting Division Human Effectiveness Directorate 711th Human Performance Wing Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-Apr-2012	Final	10/19/10 – 4/1/12

4. TITLE AND SUBTITLE

Feasibility of Artificial Attention at Beyond-Human-Scales

5a. CONTRACT NUMBER
FA8650-09-D-6939 0026

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)

Dr. David Woods
Dr. Alexander Morison

5d. PROJECT NUMBER
5328

5e. TASK NUMBER
0026

5f. WORK UNIT NUMBER
5328X04B

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cognitive Systems Engineering Laboratory
Ohio State University
470 Hitchcock Hall
2070 Neil Ave
Columbus OH 43210

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Materiel Command
Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Forecasting Division
Human-Analyst Augmentation Branch
Wright-Patterson AFB OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

AFRL-RH-WP-TR-2012-0113

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A: Approved for Public Release: Distribution unlimited; 88ABW/PA Cleared 21 Sep 12; 88ABW-2012-5091

13. SUPPLEMENTARY NOTES

Data overload is no longer the exception; it is the rule. Unfortunately, there are few instances where an overwhelming amount of data collected is managed successfully, more typically work-arounds such as filtering, ignoring, deleting, or increase manpower are used to continue to make progress toward goals. One exceptional and successful instance of managing data overload from sensing is human visual attention. Coincidentally, advances in technology over the past 30 years have led to computational models of attention that make it possible to simulate attention processes. In the present work we develop and build a computational model of attention, called Artificial Attention, that operates over a network of sensors like those referred to by the United States Air Force as layered sensing systems. The present work differs from previous computational models of attention that only operate on a single sensor with a narrow and fixed field-of-view by being scalable to operate over networks of sensors with no predefined field-of-view. The computational model of attention is used to examine what conceptual and practical advances are needed to scale computational models of attention to handle the multiple sensor feeds found in layered sensing systems. The results of preliminary testing show that several hidden assumptions behind current computational models of attention block scaling Artificial Attention to layered sensing system scales.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UNCLASSIFED					1st Lt Nicole McCammon
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR	51	19b. TELEPHONE NUMBER *(include area code)*
U	U	U			NA

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

The Cognitive Systems Engineering Laboratory at Ohio State University would like to acknowledge Air Force Research Laboratory for supporting this effort. In addition, we would also like to acknowledge SRA International Inc. for their contract support in completing the effort.

1.0 SUMMARY

Data overload is no longer the exception; it is the rule. New sensor technology, new sensor platforms, and increasing connectivity are driving this flood of data. There are few instances where the overwhelming amount of data collected is managed successfully. More typically, work-arounds such as filtering, ignoring, deleting, or increased manpower are used to continue to make progress toward goals. One successful instance of managing the potential for data overload is human visual attention. In general, the human eye is exposed to a multitude of surfaces, objects, and events in the visual field, and yet the visual system is able to attend and reorient to the critical surfaces, objects, and events at any given moment in time. The visual system does so, in part, through the operation of attention, which refers to the process of focusing cognitive resources on selected aspects of the environment while ignoring others.

Currently, studies of human attention rely upon a number of approaches, such as behavioral experiments and neurophysiological investigations. A more recent approach for studying human attention is computational modeling, which entails executable models that produce simulated behavior on real or simulated input. These computational models of attention use camera or video imagery as input and produce as output a sequence of simulated fixations.

The technology of computational models of attention make it possible to develop neurologically-inspired artificial attention systems (Woods and Sarter, 2009) that redirect their gaze as inputs change in order to handle the data overload problem (Woods et al., 2002). Unfortunately, the use of narrow field-of-view cameras and video imagery as the input has hindered the expansion of these techniques to the scale associated with layered sensing systems. At the scale of a layered-sensing system, data overload arises from feeds coming from multiple points of observation/sensors. Moreover, current models of attention have a hidden assumption - they process all of the pixels in the feed from the sensor. This assumption of complete access to the full image (i.e., all pixels) is invalid; the real need is to decide where to direct and redirect processing resources when the input image is too large to process completely, as in the case of wide-area surveillance.

The present work examines what conceptual and practical advances are needed in order to scale computational models of attention so that artificial attention systems that handle multiple sensor feeds, as found in layered sensing systems (i.e., wide area surveillance from multiple sensor-feed platforms), can be developed. To identify and demonstrate the needed advances, a test case of a computational model of attention was developed and programmed. As input, the computational model was given images from a simplified simulated environment. This allowed us to test the feasibility of scaling an artificial attention technology to a simulated layered sensing-system environment composed of multiple points-of-observation with wide field-of-view imagery.

The tests showed that several implicit assumptions behind many current computational models of attention impede our ability to scale an artificial attention system to match the scale of a layered sensing system (i.e., a system with more than one sensor). The tests also showed that the concepts developed by Woods and Sarter (2002) for modeling

attention, and the extended perception concepts developed by Morison (2010) and Morison and Woods (in press), provide one basis for developing an artificial attention system that can work at the scale of a layered sensing system.

The concepts from previous research that were tested in order to demonstrate the feasibility of an artificial attention system include: (1) multiple interdependent active sampling processes, rather than a single, active sampling process found in current models of human attention; (2) a dynamic panorama as an emergent parameter rather than the fixed-extent input images in previous models.

It also was critical to test simulations of attention under conditions with only partial observability of the environment of interest and to assess the synchronization of the attention mechanism to the pace of activities and changes in the environment - a pacing metric (Woods and Hollangel, 2006).

The tests demonstrated that

- artificial attention systems can be scaled so as to function at the scale of a layered sensing system (i.e., they can handle feeds from multiple sensors);
- artificial attention has the potential to compensate for data overload;
- pacing measures are the key metrics for evaluating and comparing the performance of different automated and human-machine systems that process multiple sensor feeds.

2.0 INTRODUCTION

Data overload is a current and growing challenge for the United State Air Force (USAF). The term 'data overload' captures the commonly held notion that the amount of data exceeds the human capacity to process it and therefore people feel overwhelmed. One result of data overload is warfighters end up processing much less of the available data as they try to complete their mission. For this reason, data overload implies an underuse of existing data; a missed opportunity to achieve or do more. Data overload is therefore a useful description of warfighters who feel that a given opportunity was missed. However, as a diagnosis, data overload falls short because it provides no treatment or guidance on how to seize the missed opportunity to achieve more. Presently, the only solution to avoid the sense of a missed opportunity and to take greater advantage of collected data is to increase manpower, which the USAF acknowledges is not a sustainable solution to data overload.

A new framing is necessary that will lead to meaningful and useful interventions. The new framing is called Extended Perception and treats ubiquitous sensing and autonomous platforms as extending human reach to new and multiple scales (Woods et al., 2004; Woods and Sarter, 2010; Morison, 2010; Morison and Woods, in press). More sensor platforms, new viewpoints, and new sensor technology are all capabilities that expand humans' sensing capabilities into new scales that were previously inaccessible (i.e., layered sensing system scales). Examples of new sensing capabilities include wide-area imagery (spatial scales), change detection (temporal scales), and hyperspectral sensors (modality scales). We call these "beyond human scale" systems.

If sensor technology expands human spatial/temporal scales to beyond normal human scales, then we must also expand human perceptual abilities to recognize patterns, explore, and re-focus to layered sensing system scales (Morison and Woods, in press). For the challenge of finding relevance in the captured data, specifically, expanding human attention to new scales is key (Woods et al., 2002). This is because human attention is the functional mechanism that allows higher-level cognitive systems to transform an overwhelming amount of data into a coherent whole that is tuned to the pace of activities in the environment and is sensitive to surprise. For these reasons, deploying scalable computational models of attention at layered sensing-system scales is a prerequisite for supporting the ability of operators and analysts to find meaning and relevance in the growing mass of data.

Developing computational models of human attention that scale to layered sensing system scales is possible, in part, because there has been recent progress in building and testing computational models of human attention for the single agent or sensor case (Itti and Koch, 2001). However, current computational models of attention are simulating human attention and have been tested only at a single, narrow point of observation. Unfortunately, all of the current computational models make several common assumptions that make them inapplicable to the wide area surveillance and multiple sensor feeds that are central to layered sensing system scales.

This work examines the feasibility of scaling computational models of human attention to layered sensing system situations. The first step is to reject inappropriate assumptions of current computational models of attention, which will allow the establishment of a new perspective for developing scalable computational models. Defining a scalable computational model of attention will rely on characteristics of layered sensing systems. The second step is to identify the additional capabilities needed to begin to address the wide area surveillance and multiple sensor feeds of concern to the Air Force. The third step is to perform some initial tests to examine the potential of Artificial Attention systems to cope with data overload.

The report begins by presenting current computational models of attention focusing on the assumptions that limit scalability and applicability to layered sensing system situations. The report then presents the tests that demonstrate how current models can be scaled and extended to work at layered sensing system scales.

3.0 RELATED WORK

Studies of human attention have a long history. Early efforts that continue today are empirical investigations of the behavior of the human attention system to uncover its underlying components. The typical approach to understanding these components is by decomposing attention into its components and then studying them individually. For instance, Posner(1971), categorized the components of attention into alertness, selectivity, and processing capacity. In the end, the components of attention have been categorized by researchers depending on their particular interests, capabilities, and ability to test certain phenomena, resulting in many different methods and tests that fall under the broad field of attention. Reductionist approaches for studying human attention is not the focus of the present work. Instead, these approaches highlight a more recent trend in studying human attention based on modeling and simulation. Modeling human attention follows from work on processing capacity approaches to human attention, such as the limited-capacity channel model of Broadbent, 1958, shown in Figure 1.

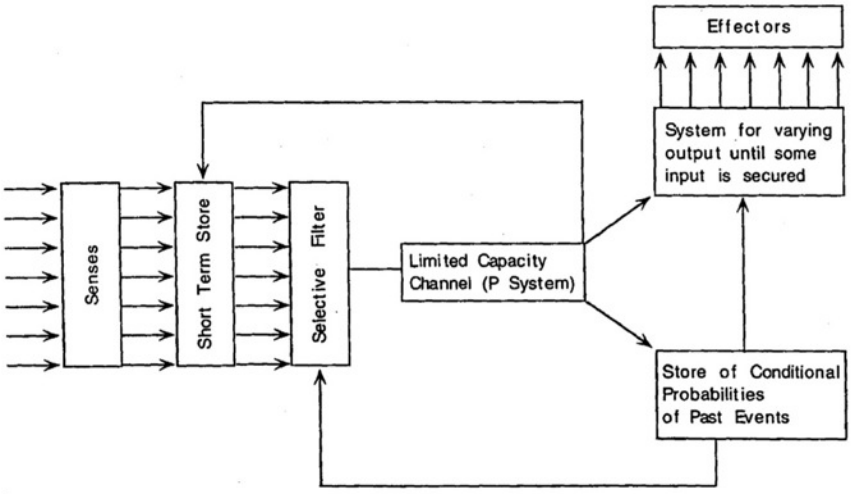

Figure 1: A Limited Capacity Channel Processing Model of Attention
(from Broadbent, 1958)

3.1 Computational Models of Attention

This trend in modeling attention has evolved into the development of a number of computational models that can be simulated, i.e., encoded as an algorithm that takes input, which simulates sensory data such as light or sound. Several researchers are developing these computational models of attention and simulating their performance on specific input and then, for example, comparing the output of the algorithm to human performance (Woods and Sarter, 2010). An example of simulated eye tracking from a computational model of attention is shown in Figure 2 from the work of Itti and Baldi(2006).

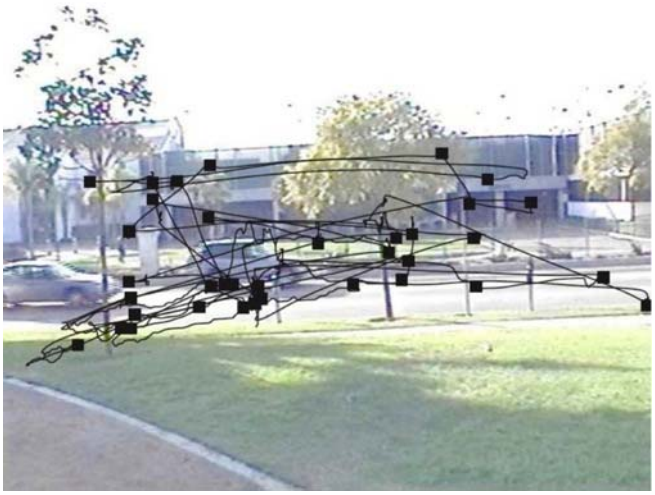

Figure 2: Simulated Eye Track from a Computational Model of Attention
(The squares represent fixation points of the algorithm and the lines indicate the scan path)

Other computational models include Koch and Ullman(1984), which is a precursor to the model developed by Itti and colleagues. Treisman and Gormican(1988) has a processing model to describe basic empirical findings of attention and perception of objects. Other computational models of attention from Tsotsos et. al.(1995), Le Meur et. al. (2006), Frintrop, et. al. (2007), and Wickens et. al. (2003) are more instances of attempts to design a model of attention that can be simulated and then compared to human performance on similar data.

In addition to measuring performance, describing or representing a computational model of attention is important. There are at least two standard approaches to representing these models of attention.

3.2 Representations of Computational Models of Attention

Each of the computational models of attention listed in the previous section use multiple representations to communicate the details of the respective model. Nearly all of the models use a block diagram representation to describe computational units, the interconnections between components, the data that is processed at each component, and the data communicated between components. An example of this type of model is shown in Figure 3 from Itti and Koch(2001). Often, in additional to the block diagram representation, a second representation is used that demonstrates how the algorithm behaves for a particular input. The representation is typically an input image/video frame of reference with an overlay; for instance, one such representation involves a sampling symbol and connecting lines referred to as a simulated eye-track; see Figure 2.

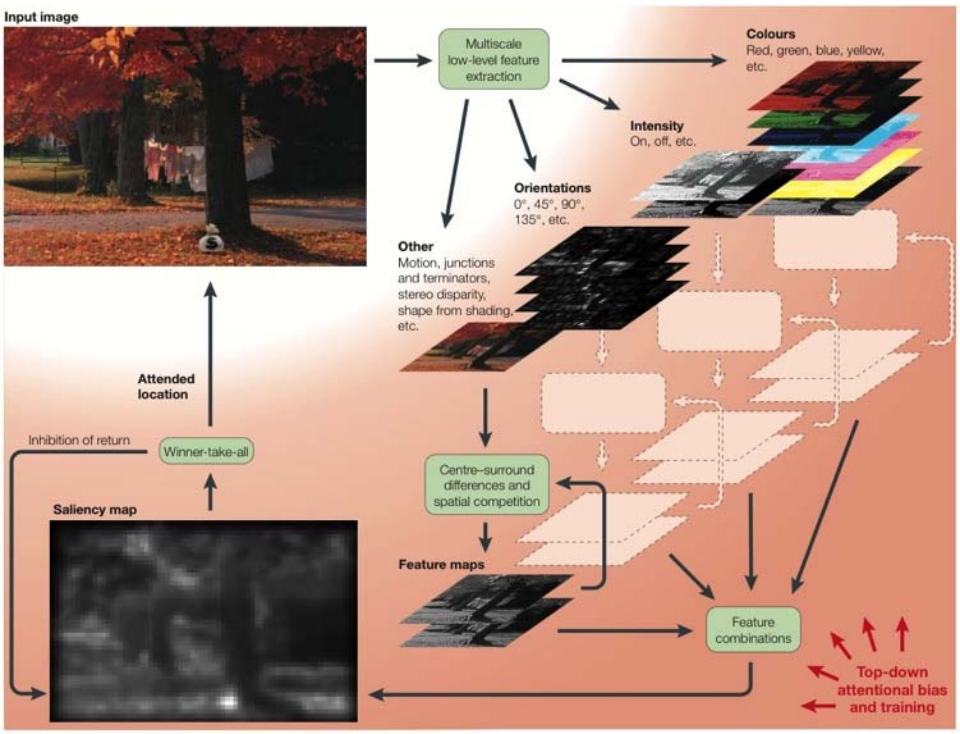

Figure 3: Block Diagram Representation of the Itti and Koch, 2001 Computational Model of Attention with Input Image, Multi-Scale Feature Extraction, Local Center Surround Differencing, and Feature Combination

(The resulting saliency is used to determine the next fixation location using a winner-take-all selection mechanism)

Other more dynamic representations have also been used to demonstrate algorithm performance. For instance, using an image frame-of-reference, a video from iLab translates the video to move the image fixation position to the center of the video window. This dynamic representation is important for several reasons. First, the representation demonstrates the several frames-of-reference involved in computational models of attention; in this instance, the representation entails a display window, an input image of the same size, and the fixation position. Second, the representation selects a pseudo-eye frame-of-reference, where the representation functions as if the eye is fixed and observing a moving world. In other words, the movement of the eye is recoded as movement of the world. Relative to other representations, the eye frame-of-reference is unusual and potentially confusing, however it does capture the dynamics and fluidity of an attention process.

Although the specific details of each model vary, in general, all of these computational models follow a general form, shown in Figure 4. The algorithm steps are:

1. Begin with a fixed-extent input image, e.g., a 640 x 480 image
2. Apply multiple machine vision algorithms to extract features
3. Normalize and combine all features to produce a probability map
4. Probabilistically select a fixation location, and repeat from (1).

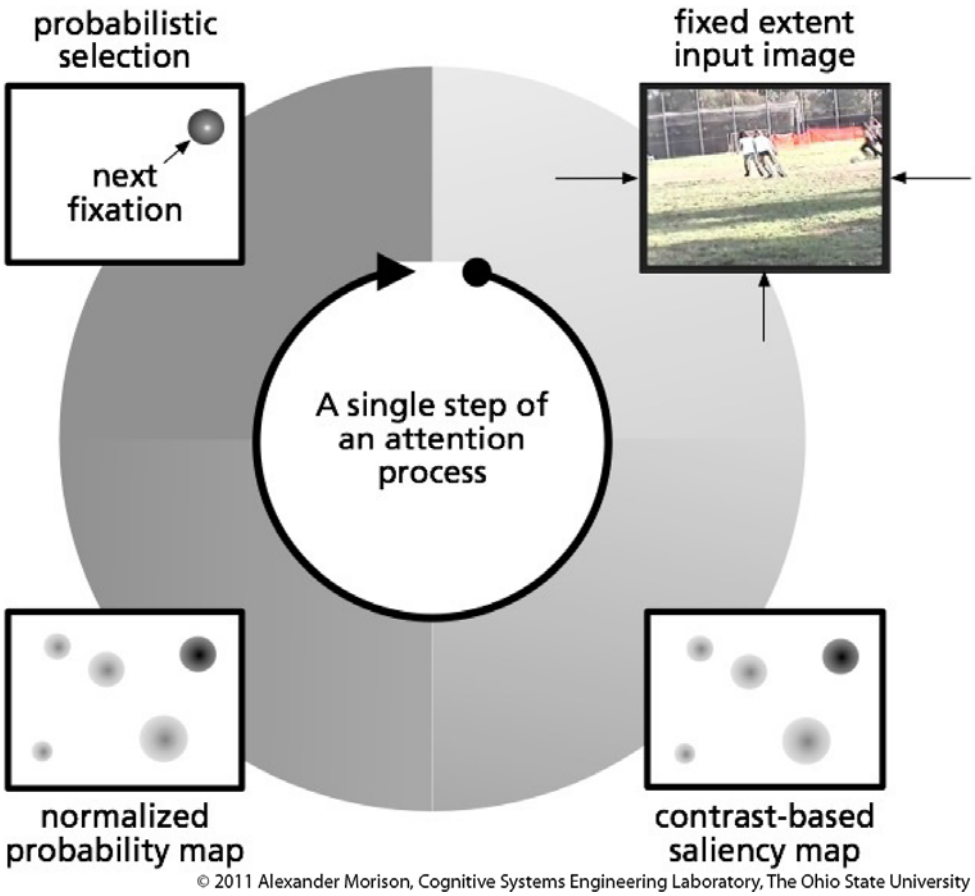

Figure 4: An Abstracted Representation of Existing Computational Models of Human Attention

(From Woods and Morison, 2011)

(The representation highlights the limiting assumptions including, the fixed map extent, the use of all data at each time-step, the single process, and the lack of an active sampling)

3.3 Bottom-Up and Top-Down

Research on computational models of attention should include a section on top-down attention, although this is not the focus of the present work. An important aspect of human attention is the simultaneous use of top-down gist with bottom-up processing (Woods and Sarter, 2010). Yarbus(1967) beautifully demonstrated the influence of top-down gist in human attention, where a subject was primed with a sequence of different questions about the content of the painting and the resulting eye tracking data demonstrated different eye tracks (see Figure 5). That is, context influences eye movement (cf. also, Itti and Baldi, 2006). And since eye movement is approximately coincident with attention, context influences attention (Woods et al., 2002). Even though top-down is a fundamental component of attention, our understanding of what constitutes the top-down component, and our understanding of how top-down processes influence the active sampling processes, is extremely limited.

The computational models of attention noted thus far almost exclusively rely on bottom-up data processing. The top-down gist of a scene, and the corresponding influence of knowledge and expectations on sampling of an environment, are almost entirely absent from such models (e.g., Christoffersen et al., 2007). In the few instances where a top-down component to the attention process was cited, the implementation is a modulation of parameters within the model, e.g., thresholds on features, feature weighting, or altering probabilities. The deeper question of how and why those modulations occur is almost always underspecified.

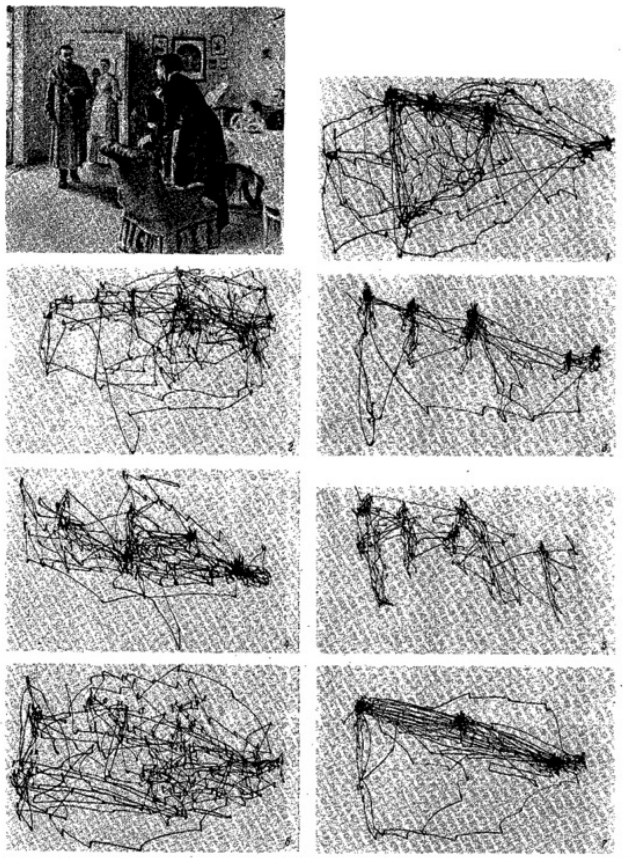

Fig. 109. Seven records of eye movements by the same subject. Each record lasted 3 minutes. The subject examined the reproduction with both eyes. 1) Free examination of the picture. Before the subsequent recording sessions, the subject was asked to: 2) estimate the material circumstances of the family in the picture; 3) give the ages of the people; 4) surmise what the family had been doing before the arrival of the "unexpected visitor"; 5) remember the clothes worn by the people; 6) remember the position of the people and objects in the room; 7) estimate how long the "unexpected visitor" had been away from the family.

Figure 5: The Work of Yarbus (1967) Demonstrating the Immediate Top-Down Influence on Eye Scan Path

(This work demonstrates that the top-down properties of attention are not slow sequential build-up but are immediate and drive eye saccades)

The next section describes the methods, assumptions, and procedures used to construct and simulate a scalable computational model of attention for layered sensing system situations. The approach eliminates the general assumptions that constrain the relevance

of current computational models of attention covered above. The simulation method is also described in the next section, which is fundamental to extending computational models of attention and current technology for 3-dimensional virtual environments. Finally, the general procedure used to produce the results and demonstrate feasibility is described.

4.0 METHODS, ASSUMPTIONS, AND PROCEDURES

The methods, assumptions, and procedures used to demonstrate the feasibility of scaling Artificial Attention to scales attendant to a layered sensing system all follow from the opportunities and challenges created by ubiquitous sensing in combination with extensive connectivity and platform autonomy. As the name suggests, layered sensing systems indicates an environment or space which is not necessarily experienced directly by an individual physically present in that environment. Advances in technology are creating access to layered sensing system environments in at least two critical ways 1) through new sensor technology providing access to previously inaccessible physical environments or viewpoints; and 2) by creating new environments or spaces, such as in the case of cybersecurity. In order to understand the applicability and usefulness of scaling computational models of attention to layered sensing system scales, a relevant environment is necessary, that is, an environment that illustrates one of the above classes involving beyond human scales.

A second limiting constraint that follows from scaling attention to layered sensing systems is that attention affects what is sensed or captured from sensors. This constraint puts significant demands on collection mechanisms because what data are collected is dependent on the output from the attention algorithm. That is, where a sensor is oriented now and what is observed from that point of view in the scene of interest will, in part, specify a new view direction - where to look and where to focus next (Woods et al., 2002; Woods and Sarter, 2010). In principle, the feasibility testing could use wide-area video surveillance imagery; however, no such data were readily available and the complexity imposed by the form of such data was deemed unnecessary for demonstrating feasibility.

In combination, these constraints lead to a simulation-based approach to demonstrating feasibility. A simulation-based approach is similar to the approaches presented previously for computational attention systems that work at a human scale. Following a description of the simulation method, the assumptions underlying the scaling of an Artificial Attention model, tailored for layered sensing system situations, are described. These assumptions differ from the assumptions of current computational models of attention. Finally, some comments are made about the procedure used to demonstrate feasibility and about the results from the feasibility demonstration.

4.1 Simulation Method

A simulation-based approach is used to demonstrate the feasibility of applying and scaling a computational model of attention to the scale of a layered sensing system. Before feasibility can be demonstrated at a layered sensing system scale, however, a computational model of attention must be demonstrated at a human scale. In the present context, human scale means a single point-of-observation in a physical environment. Even at a human scale, there exist many limitations inherent to using a non-simulation-based approach. For instance, there are physical sensing constraints, sensor limitations, conceptual constraints on attention processes, practical challenges of collecting data, and design challenges in representing and evaluating algorithm output.

An example of a physical constraint is the velocity limitation of standard video sensors. A key aspect of attention processes is that they are tuned to the pace of activities in an environment, a point that will be elaborated in greater detail later. This means that, for a video sensor monitoring an area with human activity, the sensor must sample, i.e., orient and grab an image, at a rate that is commensurate with standing, walking, running, bicycling, or driving. Unfortunately, standard video sensors are not presently designed as sampling devices. Instead, video sensors are devices designed for staring and function best with relatively slow movements on the order of 10º/sec. Typically, speeds faster than 10º/sec induce sensor artifacts like motion blur.

One potential solution not explored in the present work is to slow down the pace of activities in the physical environment to match the maximum sampling rate of a video sensor. Slowing down an environment in this manner would be composed of a set of contrived activities, and therefore this option would be no better than a simulated world. Moreover, using a physical environment would increase the challenge of conducting a controlled set of trials. Undoubtedly, variations in pacing across trials would cause significant difficulties in comparing sampled behaviors from trial to trial.

Another potential solution to the physical limits of using real sensors and a physical world, which is the solution implemented in the present work, is a simulation-based approach to the environment and sensors. At human and layered sensing system scales, a simulation-based approach presents several advantages, including experimental control, alleviation of physical sensor constraints, repeatability, and a flexible approach to varying the richness of the events and activities being simulated.

The simulation-based approach also introduces some challenging questions with respect to fidelity and representativeness of the simulation. For instance:

- What simulated space or environment is suitable for testing an attention process at a human scale and at a layered sensing system scale, for example, a 2-dimensional image space or a 3-dimensional virtual environment?
- What is the set of prototypical objects and behaviors relevant to testing an attention process at a human scale and at a layered sensing system scale, for example, what objects, surfaces, activities, or events would be relevant?
- How do you represent an environment or an attention process at a layered sensing system scale?

The next section explores some possible answers. With the goal of demonstrating the feasibility of scaling computational models of attention to a layered sensing system, finding and illustrating one or more possible solutions to these questions is sufficient for success at this stage of the development of technology for Artificial Attention. The simulation-based approach provides the power necessary to explore a set of possible solutions. The exploration with both successes and failures reveals the possibilities and constraints of using a simulation-based approach as well as the opportunities for scaling computational models of attention to a layered sensing system.

4.2 Assumptions of Attention Models

Previous sections provided an overview of current computational models of attention, and introduced a generalized version of these models in Figure 4. This generalization is useful because it captures in a concise manner some of the assumptions of current computational models of attention. Surprisingly, some of these assumptions are relatively poor for modeling human attention at a human scale. Moreover, many of the assumptions of current models make the model itself inapplicable to the scale of a layered sensing system, where practical issues like data overload are inescapable. Current models of attention assume:

- access to the entire environment at all times, which is inconsistent with data overload.
- a single attention process, which is inconsistent with human physiology and functional mechanisms of focusing and reorienting.
- a single, linear, temporal scale, which is inconsistent with both human attention and with layered sensing system environments.
- an apriori well-defined attention boundary with fixed extent, i.e., a camera image, which is inconsistent with human attention.
- a single point-of-observation that is potentially reorientable or movable, which is inconsistent with layered sensing system environments.

In order to operate at the scale of a layered sensing system, these assumptions must be relaxed. The constraints of creating an attention process that operates at multiple points-of-observation (i.e., a layered sensing system) are the following (cf., Woods and Sarter, 2010, who provide the basis in terms of neurobiology and situation awareness):

- The environment or space over which the attention process operates is only partial observable, i.e., not all data is accessible all the time. Partial observability is part of the definition of data overload.
- A new sample location changes the portion of the environment observable.
- An attention process is composed of two separate but interdependent processes: a center and a surround. The center is a focusing process and the surround is a reorienting process.
- Attention operates over at least two (potentially more than two) time scales, e.g., for human attention saccades and perceived fixations.
- The result of an attention process is a dynamic emergent panorama, which has neither an apriori well-defined boundary, nor a fixed extent.

In addition to this new baseline set of properties for an Artificial Attention model, some additional properties include the different temporal sampling rates of the center and the surround processes, which are discussed in the results section. Also, the spatial relationship between these processes is another assumption in the current model. Similar to human physiology, the center and surround process are coupled; however, functional descriptions suggest that center and surround attention processes are not absolutely

coupled to eye movements (i.e., covert shifts of attention while maintaining fixation on a target). This is not reflected in the current Artificial Attention implementations.

4.3 Procedures

The simulation-based procedure is similar to those of other computational models of attention. The Artificial Attention algorithm will be executed over the same input for multiple iterations. Over several iterations a few key parameters will be varied to gain insight into the sensitivity of the attention process to different configurations. Given the probabilistic nature of the attention process, the actual performance of the algorithm is non-deterministic and therefore many iterations of the algorithm are necessary to categorize performance.

In the current work, demonstrating feasibility is illustrated using a limited number of runs. The limited number of runs is mostly constrained by challenges in implementing the algorithm using several different interacting components and the complications that rise from such endeavors. In particular, one complication is the memory leaks that exist in the current implementation of JavaCV, which does not correctly free memory in the underlying OpenCV implementation. This memory issue is the major bottleneck to collecting additional data. Fortunately, sufficient iterations have been collected to demonstrate the basic capability of the Artificial Attention approach and illustrate the feasibility of scaling Artificial Attention technology to a layered sensing system.

5.0 RESULTS

The goal of the present work is to demonstrate the feasibility of scaling a computational model of attention to a layered sensing system. Achieving this goal required implementing a simulated 3-dimensional test world, a computational model of attention, and a simulated layered sensing system sensor. The results of the current work demonstrate that scaling a computational model of attention to a layered sensing system is possible. In addition, the results show that the key concepts for Artificial Attention technology promise major gains in dealing with data gluts. These key concepts include:

- Attention arises from an actively sampling process—deciding where to sample next.
- Attention requires a minimum of two separate but interdependent sampling processes (i.e., a center and a surround).
- A sampling process must be tuned to the pace of activities in the environment.
- Samples build up a dynamic panorama that has neither a fixed extent nor a static shape (including a decay function, as samples fade out of the panorama).

These assumptions and the results from the simulation of Artificial Attention at the scale of a layered sensing system are described in detail. Prior to the layered sensing scale results however, a detailed description of the Artificial Attention computational model, and the results from applying the model to a human scale (single sensor), are provided. The computational model and human scale output provide a necessary foundation for understanding the output of the model at the scale of a layered sensing system.

5.1 Artificial Attention Model

There are many steps and details in how a computational model of attention transforms a dynamic environment into a sequence of sample view directions from a single point of observation, that is, into a set of orientations akin to an eye track.

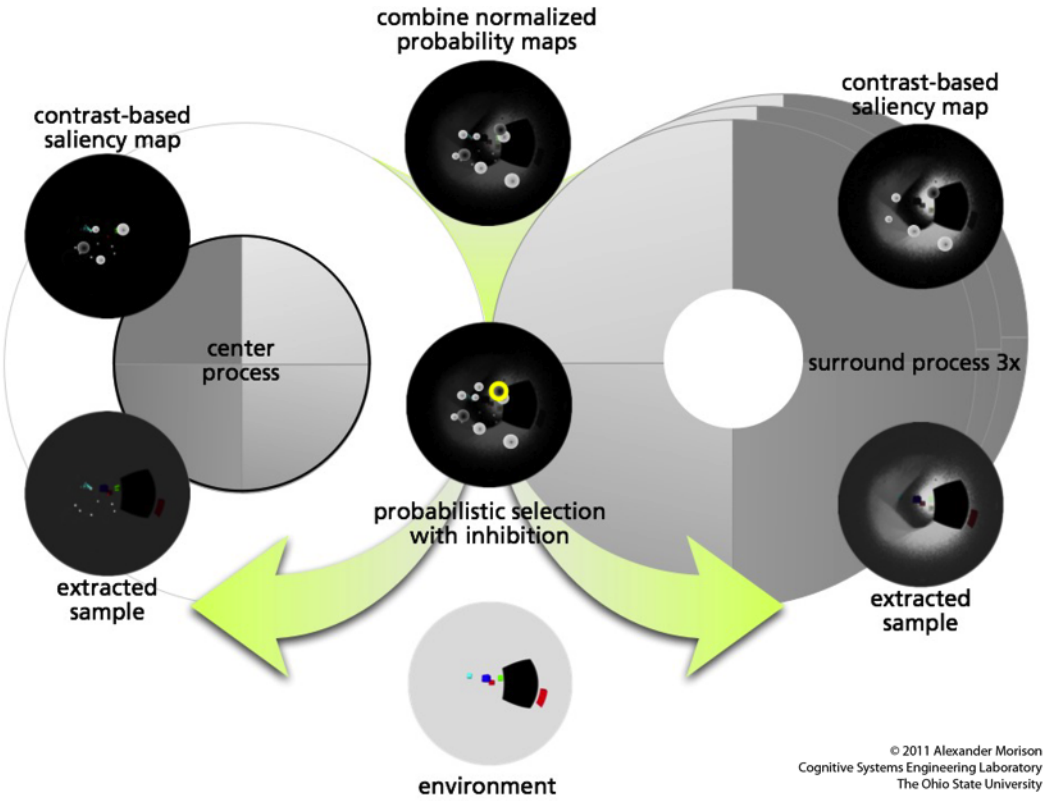

Figure 6: Diagram Representation of the Artificial Attention Model

(From Woods and Morison, 2011)

(The key aspects of the model are two processes (center and surround), different temporal rates for the two processes, a sequence of steps for feature extraction, an environment which to sample, and view direction selection)

The computational steps that compose a single attention agent are shown in Figure 6. This representation is similar to the block diagrams in Figure 4 that generalizes across existing computational models of attention. The description of the tested Artificial Attention algorithm is organized based on this representation, which serves as an anchor for explaining the algorithm. The description begins with the starting point for the model, the input environment (the bottom middle image of Figure 6). Both the 3-dimensional simulated environment and the image representation are described below. Following this is a description of the output representation. Describing the output representation is important because the description of the algorithm will partially rely on sample output images taken from test data. Then, an overview of an Artificial Attention Agent is given and where the two fundamental attention processes are introduced. These two attention processes are shown in the left and right halves of Figure 6. Although these two attention processes are different, they share a similar structure and, therefore, in terms of implementation, both attention processes are instantiated from a single attention mechanism. A description of the Artificial Attention model applied at the scale of a

layered sensing system is then provided. Finally, the overall feasibility of Artificial Attention at the scale of a layered sensing system is discussed.

5.2 Simulated Physical Environment

The obvious starting point for a computational model of attention is the world. In this context, the world is the image captured from a 3-dimensional virtual environment. We describe the content of the simulated environment, the point of observation of the Artificial Attention agent, and the approach for capturing a sphere of view directions at the agent's current point of observation.

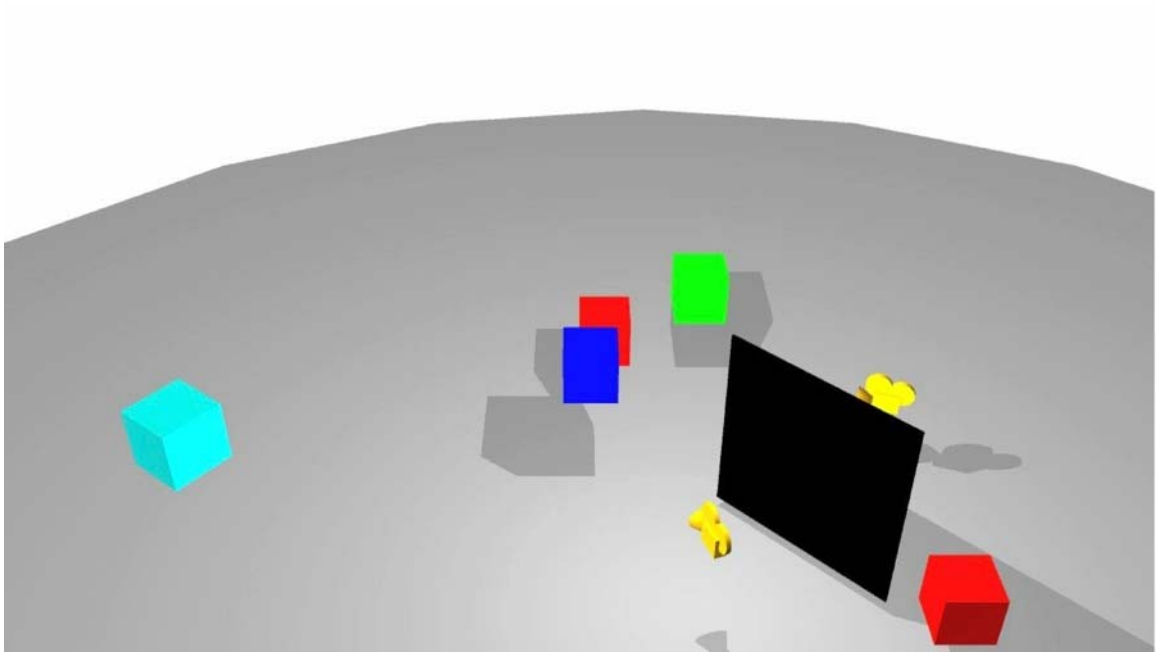

Figure 7: View of the Test Environment with Objects of Various Colors and Two Camera Positions Denoted by the Orange Camera Objects

The 3D virtual environment is a substitute for a physical environment and is populated with surfaces, actions, and events. The particular surfaces, actions, and events instantiated for demonstrating feasibility are simplified versions of real-world phenomena. For example, a substitute for people in the virtual environment are simple shapes, such as cubes and planes. Certainly people are vastly more complex than simple shapes, however, both are examples of closed surfaces that define an object. Similarly, simplified actions and event are created using these simple objects. Examples of simple actions include translating and rotating objects and examples of simple events include objects appearing or disappearing behind occluding edges or other objects (i.e., dynamic occlusions). These simple actions and events are analogies to more complex actions, such as walking or bicycling, and events, such as people moving past an occluding boundary of a building or appearing from behind an occluding boundary of a moving car. For the purpose of demonstrating feasibility of Artificial Attention, we believe that these simple actions and events are representative of the more complex actions and events.

With the environment defined, only two additional pieces of the simulation remain before describing the Artificial Attention algorithm, a point of observation in the environment and a method for collecting all potential view directions at that location. In the 3D virtual environment, these two pieces are a camera position (with orientation) in 3D virtual space, and a projection that maps a sphere of view directions into a single image (planar representation).

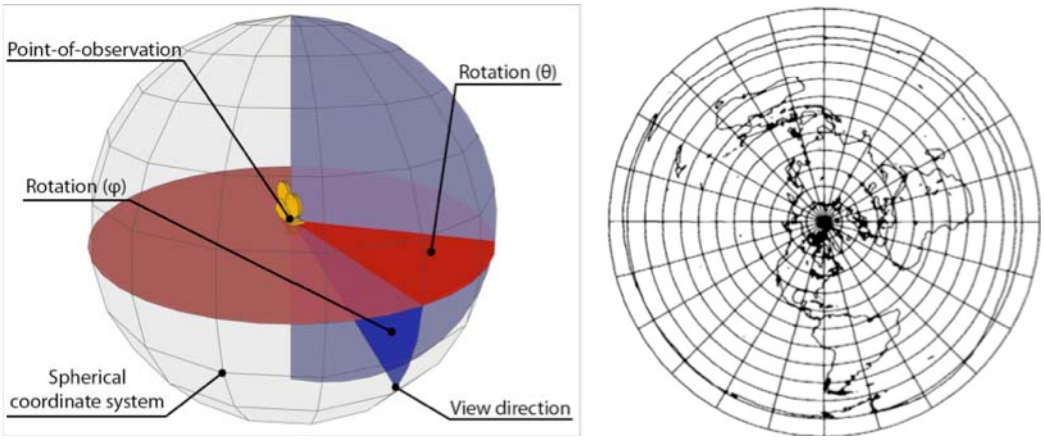

Figure 8: View Sphere Depicting the Spherical Relationship between a Point-of-Observation and all Potential View Directions *(left)* **and a Planar Image of the same Sphere Generated using a Azimuthal Equidistant Map** *(right)*

The planar image is the representation that connects the 3D virtual environment to the Artificial Attention algorithm/agent. Transforming the 3D environment into a planar image means that standard image processing software packages, such as OpenCV, can be used to perform feature extraction and image manipulation. The process of warping or transforming a sphere of view directions (360º in theta and 180º in phi) is called projection. Many projections exist, some have spatial analogies, and some do not. All are based on mathematics. The selection of a projection depends on the desired properties that will be conserved from the original spherical representation. Some examples of properties that can be conserved are distance, area, and parallel lines. Importantly, the transformation of a non-zero Gaussian surface, like a sphere, into a zero Gaussian surface will alter some properties. In the present case, the desired property of the 2D projection is that the image be readily understandable to an outside observer. To this end, an Azimuthal equidistant projection is used with the center pole aligned with the forward-looking view direction, shown in Figure 8. Using the Azimuthal equidistant projection in the 3D virtual environment results in the images shown in Figure 9. These illustrate the spatial relationships of forward (left) and backward (right) view directions for the projection from two different points of observation. The relaxation of conserved properties is possible because of the method used to sample from the 2D projection/image. The sampling or extracting of pixels will be described in detail during the attention process description.

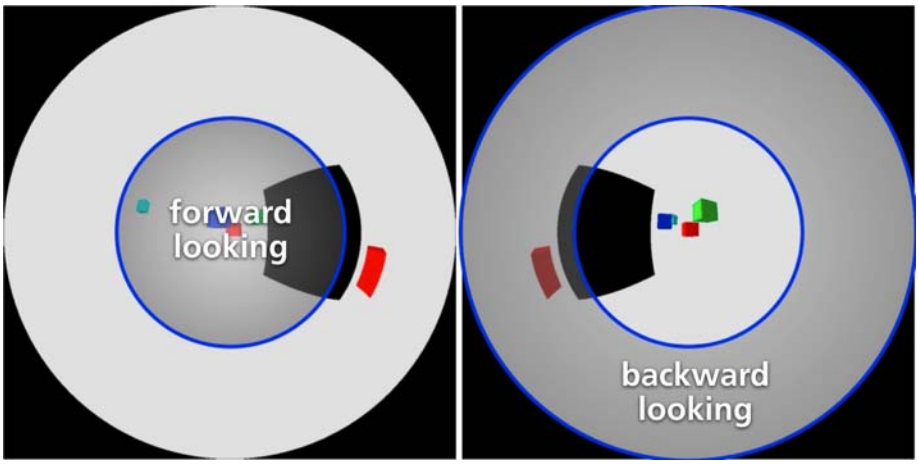

Figure 9: Azimuthal Equidistant Projection Captured from the Virtual Environment with Annotation Depicting the Forward Looking Hemisphere *(left)* **and the Backward Looking Hemisphere** *(right)*

The orientation for the projection is along a pole, where the pole is aligned with the view direction of the camera in the 3D virtual environment. The resulting equations for the Azimuthal equidistant projection with polar orientation are the following:

$$r = \sqrt{x_{pixel}^2 + y_{pixel}^2} \quad (1)$$

$$\Theta_{sphere} = \tan^{-1}\left(\frac{y_{pixel}}{x_{pixel}}\right) \quad (2)$$

$$\Phi_{sphere} = \frac{r \times FOV}{2} \quad (3)$$

$$x_{ray} = \sin(\Phi_{sphere}) \times \cos(\Theta_{sphere}) \quad (4)$$

$$y_{ray} = \sin(\Phi_{sphere}) \times \sin(\Theta_{sphere}) \quad (5)$$

$$z_{ray} = -\cos(\Phi_{sphere}) \quad (6)$$

where the field-of-view (FOV) is 360º for an entire sphere.

The projected image of the 3D environment is the input to the Artificial Attention algorithm. The algorithm will sample over the image by changing the view direction to the image. That is, the algorithm will select a view direction, theta and phi, and then transform this orientation into a set of pixels within the projected image. A set of samples over time defines the attention field for a single agent and is a useful frame of reference for representing how the Artificial Attention algorithm is performing. The next section is dedicated to describing this representation and some properties and caveats in using this representation to understand algorithm performance.

5.3 Sampling Representation

The Artificial Attention diagram presented in the previous section is a useful representation for understanding the structure of an Artificial Attention agent, however,

as a sampling process the diagram provides very little insight into how the algorithm will respond to a particular input, such as a sequence of events. Depicting a computational attention model as a sequence of images (i.e., a video) for a specific input is common for other computational models. Representing the Attention algorithm performance in this way relies on using the input image as a visual frame of reference. For the Artificial Attention algorithm, this means the output is based on an Azimuthal equidistant projection.

A sampling representation snapshot is shown on the right hand side of Figure 10 next to an input image on the left. Because the goal at this point is to describe the sampling representation, not the performance of the algorithm, a snapshot in time is shown that illustrates the properties and caveats of this representation. To support the purpose of the description, the snapshot occurs after a non-trivial amount of time has elapsed in the movements of the simulated objects.

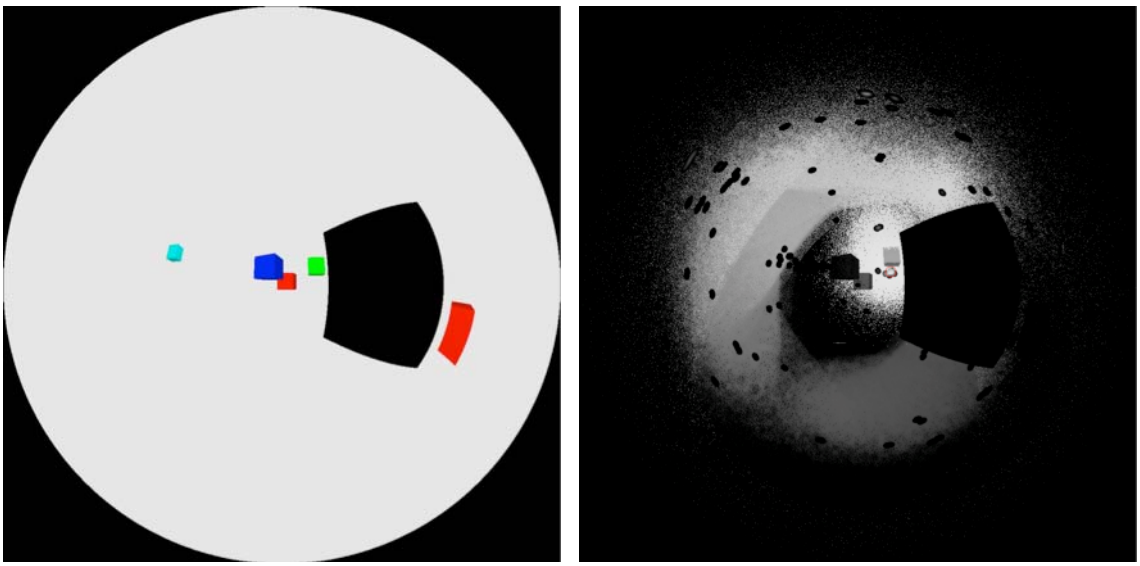

Figure 10: Test Environment Azimuthal Equidistant Projection *(left)* **and a Snapshot of the Attentional Field for an Artificial Attention Agent** *(right)*

Before properties of the sampling representation are described, a few cautionary points about the representation and its descriptive power. The sampling representation is "as if" you are watching the sampling behavior, or current field of awareness, of an attention agent, e.g., a person. A real attention agent is not "aware," just as you are not aware, of the extent and shape of your attention panorama. Instead, for you and most likely any attention agent, the panorama is wide, full, and complete; of course this is far from the truth from the point of view of neurobiology. Another and equally important aspect of the sampling representation is that this information is not usually accessible. We never see the attention field or sampling behavior of another attention agent in this global way.

With these cautions in mind, there are at least two ways to understand the construction of the representation. The first interpretation of the representation, and probably more accurate to human attention, is as an active build-up of the environment surrounding the

attention agent. Samples taken from the environment are added to the representation, while a process of continual decay removes older samples. The decay is most visible in Figure 10 in the lower left region of the sampling representation. The use of the term "active" captures the continual movement of the sampling process to counteract the process of decay.

A second interpretation of the representation is as an active revealing of the environment surrounding the attention agent. The benefit of describing the representation as revealing is its similarity to the computational implementation. A mask, generated by the attention agent, specifies which parts of the environment are revealed and in what state of decay. Unfortunately the term 'revealing' gives an impression that the world is a given without the indication that work is required by an attention agent to construct and maintain a dynamic panorama.

Independent of the intuition used to gain an understanding of the sampling representation; there are several properties or features worth noting about the sampling representation.

1. *The result is a dynamic panorama; the attention field is neither fixed nor static*

   The interaction of an active sampling process that responds to features in the environment and an overall decay process is a panorama of awareness that is continually changing, size, shape, and extent. None of these descriptors of the panorama fixed, rather they change over time or snapshots.

2. *Structural features of attention awareness such as holes are visible*

   A well-known phenomenon in human attention is attention blindness, the absence of awareness of some property, feature, or activity in an environment. Attention blindness can occur over spatial areas, temporal scales, and functional scales. The representation of the Artificial Attention algorithm captures one sense of attention blindness, which is like a "hole in awareness." We mean this term in a quite literal sense for the active sampling process. For a dynamic panorama, like the one shown in Figure 10, there is a portion of the viewable field internal to the panorama where the environment is not sampled (or has not been sampled recently). Demonstrating a hole in awareness is a significant step forward in modeling attention and allows us to explore attention concepts such as pacing, which is a relationship between an attention sampling process and the speed of activities in an environment. We believe pacing will become a critical property of performance of an Artificial Attention system.

3. *The 'dot' shapes are sample locations that are remnants of generating the representation, not a consequence of the algorithm itself.*

   The precise construction of the representation is a consequence of the algorithm and functional requirements; in addition, the representation is also subject to consequences of implementation. In this case, the dots are a partial consequence of implementation arising from the removal of center process pixels from the surround process mask. The result is the surround sample mask maintains the negative of the center process samples. There is a small amount of utility to these "dots" as they

provide observability of the Artificial Attention agent view direction sampling history.

4. *Center and surround visual angles are based on human physiology*

   The visual angles for center and surround processes, along with the color and intensity information, were selected for this implementation of an Artificial Attention algorithm based on human physiology. This is important to note because of the appearance of the representation.

5. *The representation appears mostly grayscale, even though RGB color is updated for center sampling process*

   There is a significant difference between watching a sampling process and experiencing a sample process. This is very clear than when we compare the representation generated from the Artificial Attention algorithm and the world experience by a person. What a person experiences is not this mostly grayscale world (with holes). Instead, we experience a full and colorful world. From the physiology of the human eye you would not predict a person would see a world of many hues, but more likely a gray world with isolated moments of color. The representation of the Artificial Attention algorithm output supports the later interpretation. The world as seen should be mostly gray. The point behind this comparison is to be cautious when watching the sampling process of another agent; it is absolutely not the same as experiencing the sampling process.

Now that the test environment, map projection, and sampling representation have been described, a sufficient foundation exists to describe the implementation of an Artificial Attention agent algorithm. The description of the Artificial Attention agent algorithm will follow Figure 6 with supporting figures in the form of the sampling representation.

5.4 Algorithm Overview

The implemented Artificial Attention agent algorithm begins with two pieces of information, an initial view direction and an environment input image. These two pieces of information, along with several other process specific parameters (e.g., feature weightings and visual angle), are then used to instantiate two separate sampling processes, a center process and a surround process. Both processes first extract a sample from the environment using the view direction and process a specific visual angle. Because of the different visual angles of the two processes and the method of extraction, the two samples are different. The center sample is a narrow visual angle (~2º half angle) with full pixel resolution over the entire sample, and uses all three color channels (red, green, and blue). In contrast, the surround sample is a large visual angle (~60º half angle) with a Gaussian pixel resolution over the entire sample, and uses only intensity values (i.e., brightness or grayscale).

It is at this point that a new data structure called a sampling history is necessary. A sampling history exists per process and is similar to the sampling representation except the history only contains samples specific to the process with which it is associated. Before each process, the sampling history is updated with the latest sample from the

process, and a decay function is applied. After the decay, the extracted sample from each process is added to the process' sampling history. The sampling history is then passed to a set of feature detectors specific to the type of process. For example, the sampling history for the center process is passed to a corner detector and the sampling history for the surround process is passed to a motion detector. Each feature detector generates a feature map, an image frame-of-reference with the map cells indicating the presence or absence of the particular feature. Each feature map is normalized and then all feature maps for each process are linearly combined to create a single feature map for each process.

Generation of the center and surround process feature maps concludes the center and surround process computation for a single time step. At the Artificial Attention agent level two data structures from the center and surround processes are used. First the sampling histories for both processes are combined to create a single sampling representation that was defined in the previous section. Second, the feature maps from both processes are normalized and then linearly combined, just as the individual feature maps were normalized and combined at the process level. The result is a single feature map for the Artificial Attention agent. This feature map represents the potential locations that the Artificial Attention agent will orient to in the next iteration of the algorithm, the next time step. Interpreted this way, the feature map serves as a probability map or probability distribution. The next view direction is then selected probabilistically. The Attention Algorithm repeats by updating time and using the new view direction to update the center and surround processes to execute a new iteration.

The remainder of this section is dedicated to a more detailed description of the Artificial Attention agent, an attention process, and the behavior of an Artificial Attention agent based on different weightings of center and surround processes.

5.5 Artificial Attention Agent

The Artificial Attention agent serves several purposes. First, the agent data structure is the link between the environment and the attention processes. As the link, it manages the flow of time, view direction into the environment, the individual process execution, and the combining of separate attention processes into a single whole. The Artificial Attention agent data structure is therefore a critical component of the simulation as currently architected. Of these different aspects of the Artificial Attention agent, several are over simplifications of the current implementation. For instance, the flow of time is linear. In fact in the current implementation, the flow of time is defined by the loading of a new environment image. Another simplification is the process execution; currently the processes execute sequentially, first the center process, then the surround process. Although both the flow of time and process execution order are simplifications, we do not believe they limit the validity of the results.

Another critical detail of the center and surround temporal execution is that these two processes execute at different rates. For every execution of the center process, the surround process executes three times. So, a more accurate description of the execution of these two processes is center, surround, surround, surround, center, surround,

surround, surround, and so on. Currently the Artificial Attention agent manages these execution rates, although this may not be the simplest implementation. All of these simplifications are important to note for completeness and to highlight areas for future development.

The other aspects of the Artificial Attention agent are more richly developed because they are more critical components of building an Artificial Attention algorithm. The view direction is critical because it impacts the sample extraction. The view direction, and more importantly the change in view direction (i.e., reorientation), is a key aspect of any attention agent. We have found view direction to be a useful component across a broad range of contexts in human-sensor systems (Morison, 2010). A take-away from the present work is that view direction in computational models of attention is essential. If a computational model of attention does not explicitly define a view direction, the model should be carefully examined to understand how the model works around this constraint. The final important aspect of the Artificial Attention agent is the method of combining the output from the individual Artificial Attention processes (i.e., the center and surround processes). The current implementation uses a weighted linear summation.

The center and surround process feature maps are combined using a weighted linear sum. The linear combination is shown in Eqn. 7. The weighting provides a mathematical approach to emphasize either the center over the surround or vice versa. Each pixel in the Artificial Attention agent feature map (i.e., all possible view directions) is bounded between 0 and 1, which is enforced by the normalization and the weightings shown in Eqn. 8. Another restriction on the process weights is that they sum to one Eqn. 9 ensuring that the final agent feature map values (per pixel) are limited to the range 0 to 1, shown in Eqn. 10.

$$\mathrm{A} = \sum_{i=0}^{N} a_i \mathbf{M}_i \qquad (7)$$

$$0 < a_i < 1 \qquad (8)$$

$$\sum_{i=0}^{N} a_i = 1 \qquad (9)$$

$$\max(\mathrm{A}) = 1 \qquad (10)$$

where, A is the Artificial Attention agent feature map at time t, a_i is the process weighting, and $\mathbf{M}_i$ is the process feature map at time t.

At the agent level of the Artificial Attention algorithm, the key property for demonstrating feasibility of the Artificial Attention approach is that the center-surround process relationship has a meaningful influence on algorithm performance. The Artificial Attention approach to managing data overload requires that an attention agent's sensitivity to focusing versus reorienting can be controlled. Without this control, the utility of the approach is undermined. In addition to verifying feasibility, manipulation of the relative center surround weighting is an important global test to verify algorithm

design and implementation. If variation of center-surround process weighting fails to produce the anticipated behavior, the cause might rest in the implementation of the algorithm, not in the design of the algorithm itself. To vary the center-surround process weighting we vary the process feature weighting, a_i, of the Artificial Attention agent. In order to maximize the variation of the algorithms performance without presenting an overwhelming amount of data we selected two process-weighting conditions:

Condition 1: $a_{center} = 0.2,\ a_{surround} = 0.8$

Condition 2: $a_{center} = 0.8,\ a_{surround} = 0.2$

The contrast between these two center-surround weightings is illustrated using a sequence of output representation snapshots from the Artificial Attention algorithm, shown in Figures 11-14, using the two different weightings operating on the same input environment. The left column of Figures 11-14 is the input environment, the center column is condition 1 with 20% center weighting and 80% surround weighting, and the third column is condition 2 with 80% center weighting and 20% surround weighting.

In general, and as anticipated, the performance of the algorithm varied between the two conditions. However, the difference in performance in terms of the dynamic panorama generated was larger than expected. In terms of observability of objects, activities, and events in the environment, Condition 1 was significantly more useful than Condition 2. There are several reasons for this given the different performance characteristics. Condition 2 focused extremely well and in some sense for the objects and events in the environment, too well. Condition 2 because of this strong tendency to focus also tended to wander off when objects, activities, or events were not detected immediately. In Figures 11–14 the initial cube movements and the final cube movements were missed entirely being too far off into the periphery and too weak to grab the attention process. In contrast, Condition 2 tended to reorient quite quickly when even a weak signal in the far periphery was found. In some ways, for this world and in this context, sensitivity to reorienting is more important than focusing for keeping pace with activity in the environment. This balance between reorienting and focusing is also visible in the change in the dynamic panorama over the image sequence. The dynamic panorama of Condition 1 makes several transitions (~4) between narrow and wide visual fields. In contrast, the visual field of the dynamic panorama in Condition 2 makes only two transitions and one of these is potentially a consequence of initialization. No matter the environment or the view direction, the focusing condition is always tending to focus, and hence, more susceptible to missing objects, activities, and events.

Another important aspect of Condition 1 is that even though it is more sensitive to the surround and reorienting, when objects and activities are focused, tracking occurs. With no additional coding, the algorithm tends to track activities such as motion. Moreover, the algorithm continues to track on-going activities even when distracter events such as occlusions and dynamic occlusions occur. This behavior is extremely valuable as an existence proof that no additional information is strictly necessary to create a relatively

complicated behavior (e.g., tracking an object of interest). Although difficult to see in the sequence of images, tracking an object behind a dynamic occlusion occurs at the bottom of Figure 14.

Two additional interesting features of Figures 11–14 are the changes in sampling density between the two conditions over time and the interpretation of the two conditions and dynamic panoramas they generate. The sampling density is relatively easy to see between the two conditions because of the "dot" artifact. Comparing within conditions, Condition 1 illustrates a consistent shift in sampling density. A good example of this shift is Figure 12, where the algorithm sampling from a wide area and then begins to focus down. In contrast, Figure 7 for Condition 2 captures a "jump and go" sampling behavior with high sampling at each location; the diversity of sampling in Condition 1 is absent.

Across conditions there are also useful sampling density comparisons, for instance, the third row of Figure 6. Comparing the two conditions, both previously focused on the moving cubes through the scene, but the more recently sampling for Condition 1 sampled the entire back half of the view sphere, whereas Condition 2 is slowly sampling backward and consequently missing a large portion of the backward viewable field. Finally, as an outside observer watching the sampling behavior the two conditions result in two different interpretations. Condition 1 gives a sense that the algorithm is constantly trying to look for something new or not previously observed across the entire viewable field, but not at the expense of tracking specific activities. On the other hand, Condition 2 gives an impression that the algorithm focuses in different regions of the viewable field unnecessarily. As an outside observer, quite quickly it is apparent there is nothing in the locations where the algorithm spends significant time sampling. More importantly, given the representation it is obvious there is a large portion of the viewable field that is going un-sampled where important objects, activities, and events may be occurring.

The results from the two conditions compared in this section indicate that Artificial Attention is a feasible approach for managing the data overload problem. Moreover, the two conditions indicate that sensitivity to reorienting is an important quality of an Artificial Attention agent. The subsequent sections provide more detail of the underlying algorithm, however, no comparison testing was performed at these lower levels. Future work will examine how different feature weightings impact overall algorithm performance.

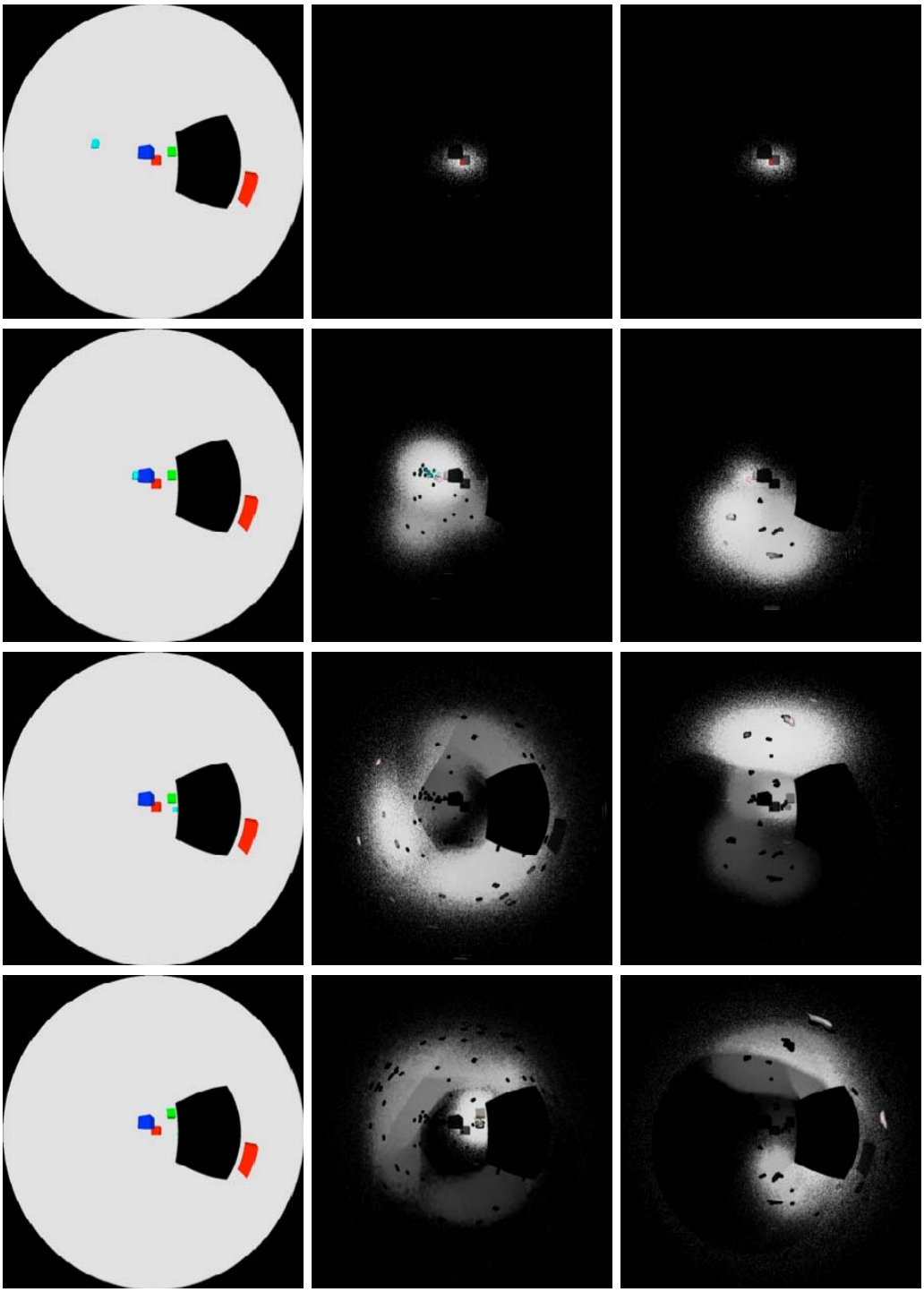

Figure 11: Comparison of Two Center-Surround Weighting Conditions
(The figure is composed of three columns, the environment (left), condition 1 (center), and condition 2 (right))

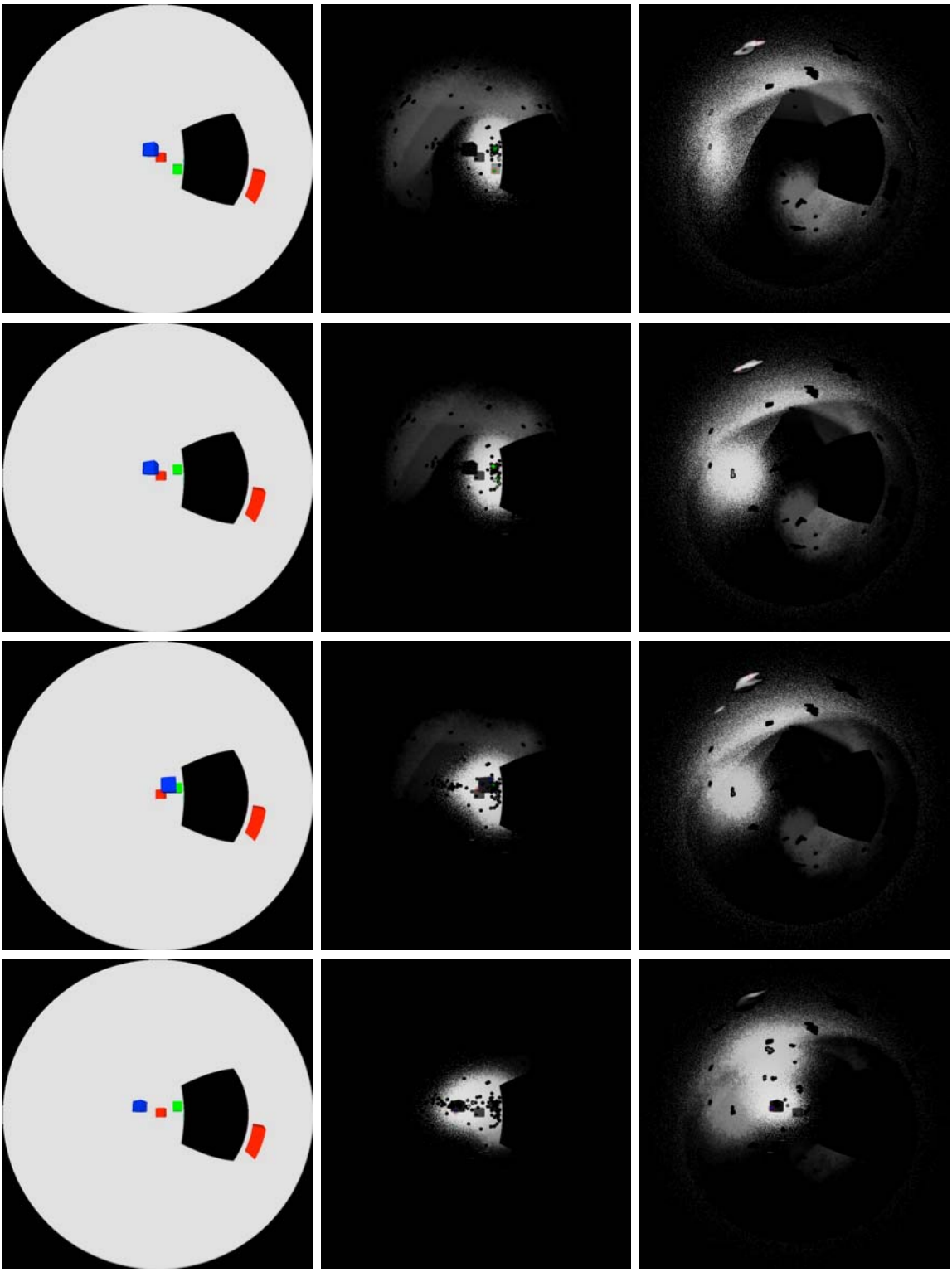

Figure 12: Continuation of Figure 6 Showing Two Different Center-Surround Weighting Conditions

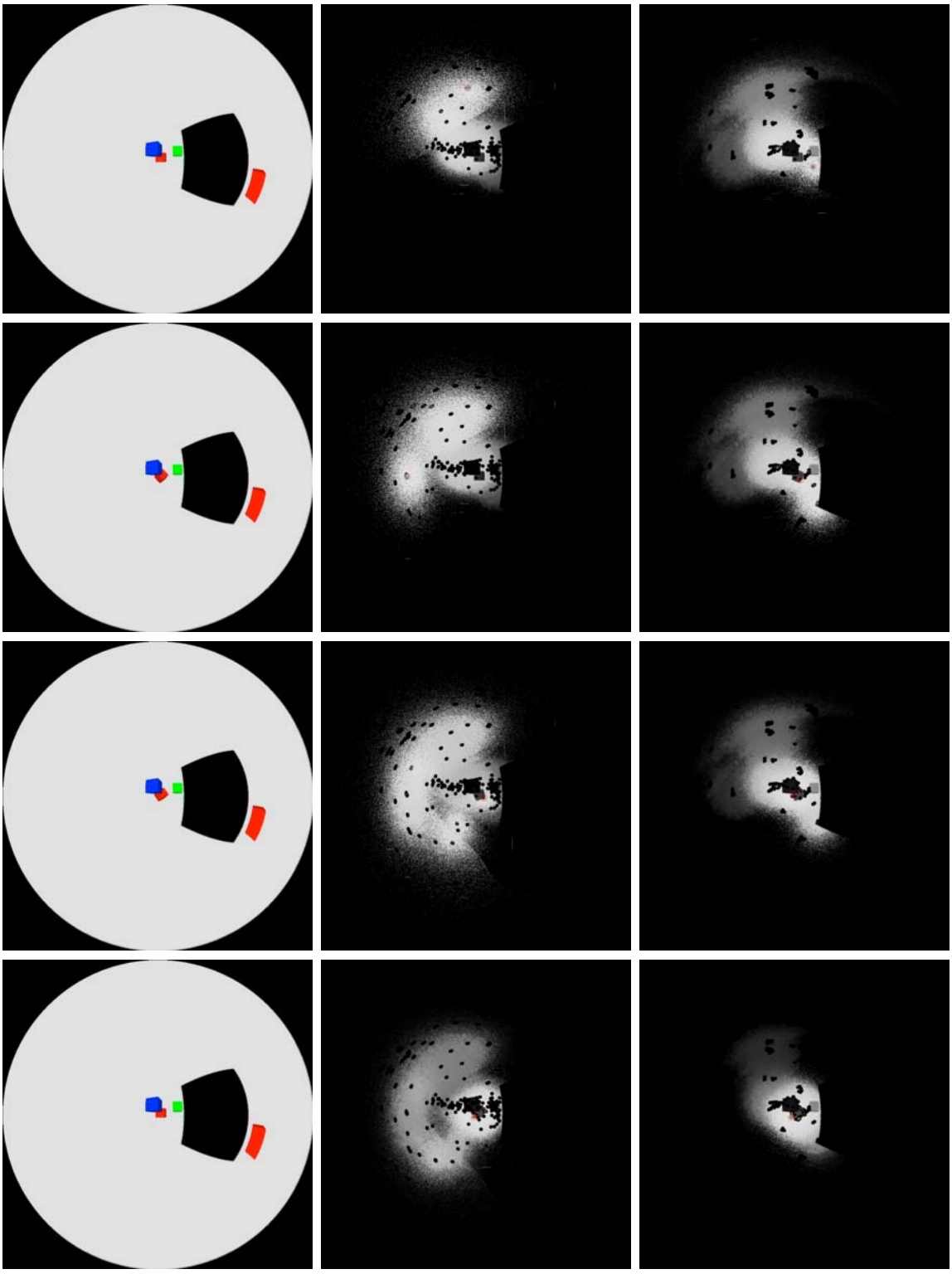

Figure 13: Continuation of Figure 6 Showing Two Different Center-Surround Weighting Conditions

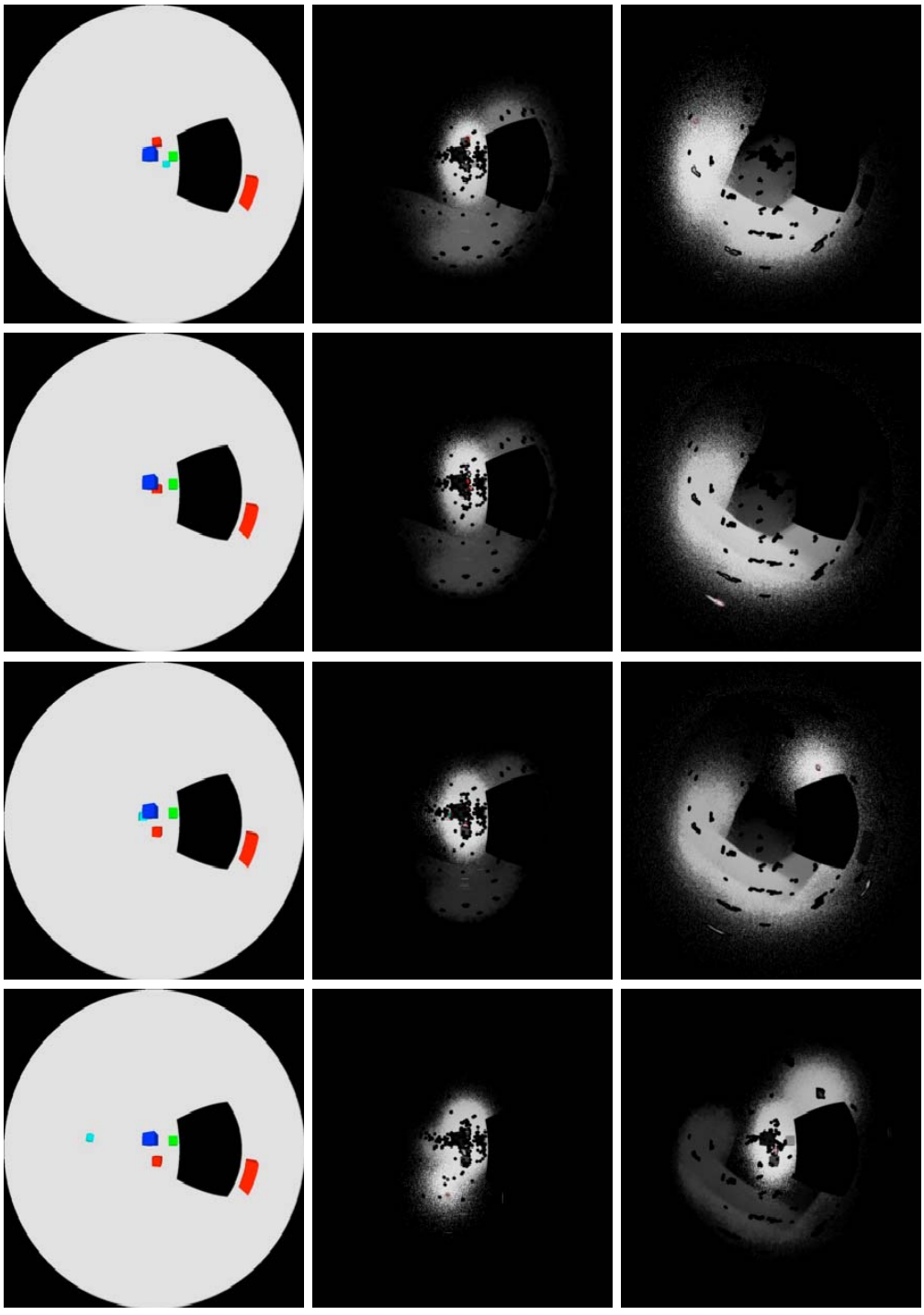

Figure 14: Continuation of Figure 6 Showing Two Different Center-Surround Weighting Conditions

5.6 Attention Process

The agent described in the previous section instantiates two separate sampling processes: a center process and a surround process. Both of these processes, and all processes in the current architecture, are based on the same structure. The section is organized by describing the generic flow of an Artificial Attention process with relevant differences between center and surround processes noted when applicable. The flow of an Artificial Attention process begins with a sample extraction. This is followed by feature detection on the sample, normalization of each feature map, and then combining all individual feature maps into a single process feature map.

Sample extraction from environment: An attention process begins by extracting a sample from the environment based on the view direction and the visual angle of the process. With these data and the known map projection, an ellipsoid region of pixels can be selected on the surface of a sphere, see Figure 15. Calculating the pixels based on a spherical surface is important. The appropriate pixels cannot be selected based on a simple closed circle in image space (i.e., map projection space). Another consideration in calculating the region on the surface of the sphere is the angular distance between two orientations (i.e., the view direction and any other pan-tilt orientation on the sphere). The current implementation uses the angle measured along the great circle defined by the two orientations.

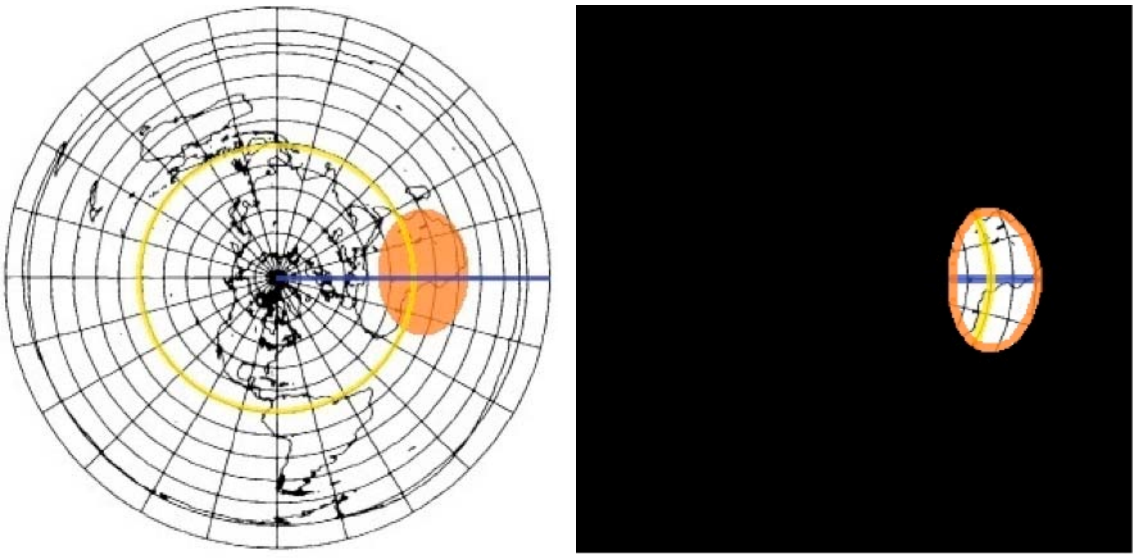

Figure 15: Extracted Sample of a Circular Region on the Surface of a Sphere
(The circular shape is transformed into an ellipsoid region in the Azimuthal Equidistant projection (left) and the resulting pixels extracted for a visual half angle of 30º with a view direction of theta = 90º and phi = 90º)

In addition to visual angle determining whether a pixel is part of a sample, a second criterion can also be used to reduce the number of pixels selected. The full resolution and reduced pixel extraction is shown in Figure 16. The pixel reduction is based on the

physiology of the human eye, specifically the reduction in number of receptors with increasing angular distance into the periphery. The reduced pixel on the right side of Figure 16 is implemented by treating a 2D Gaussian distribution as a probability distribution for pixel selection. The distribution is centered at the view direction and has a standard deviation of 1/8th of the visual angle of the sample. The goal was not to copy the physiology of the human retina, as this is not possible given the density of cones and rods in the eye, but to emphasize that the resolution in the periphery of the surround is neither the same resolution as the fovea, nor constant across the entire surround.

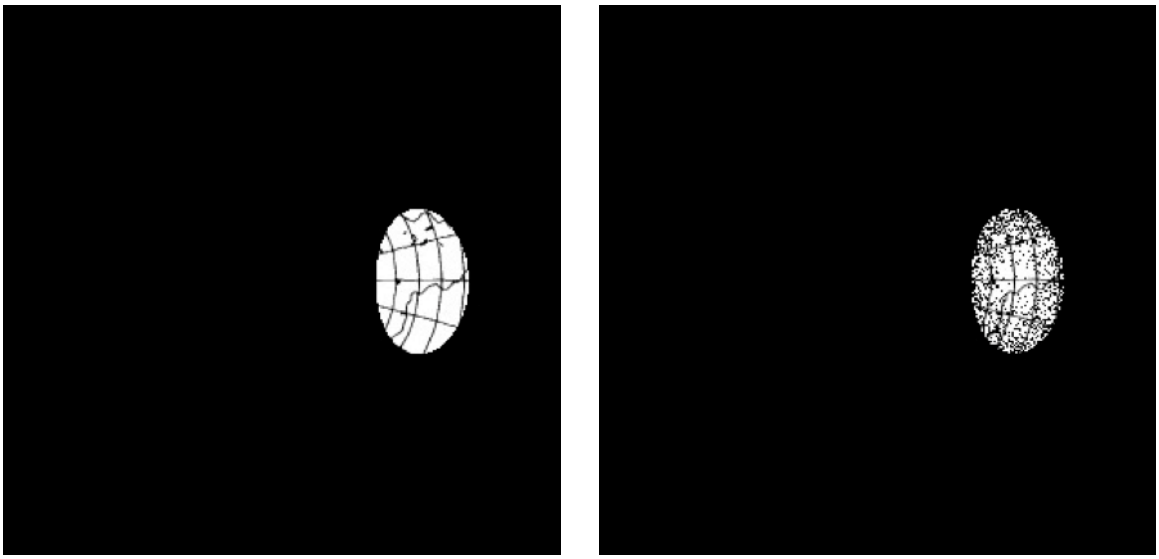

Figure 16: Extracted Sample can be Full Resolution *(left)* **or Pixels can be Removed** *(right) (The full resolution is used by the center process to simulate the high resolution fovea of the retina and the reduced resolution is used by the surround process to simulate the reduction in rod density with increasing angular distance into the periphery of the retina)*

Feature extraction from sample: A sample is a set of pixels that define a portion of a viewable field. The pixels, in isolation, are not meaningful to a computer algorithm. A large body of work called machine vision or computer vision investigates and develops techniques for extracting meaning from images, i.e., interpretations of patterns of pixels. Importantly, the techniques developed and the resulting meaning associated are based on many factors such as, the interests of the researcher, what is feasible (i.e., tractable, measurable, repeatable), and purposes of specific work. These techniques may rely on underlying human sensory or perceptual psychology, but not necessarily, and in the end it is safe to assume the techniques are completely different from human perception and attention processes. Despite this caveat, feature detectors are the basis for all computational models of attention. One must simply be careful about abstracting findings to a human attention system.

The implementation tests several generic feature detectors. We avoided using overly specific feature detectors like person detectors or shape descriptors to ensure we did not bias the work towards test environment specific factors, e.g., specific kinds of objects.

Some of these features are properties of the image itself, such as, intensity and color. Other features, such as corners and motion, are calculated using existing functionality within OpenCV, the computer graphics library. The features applied to a particular process are determined a priori and are different for the current center and surround processes. For the center process, the features extracted are the three color channels (red, green, and blue), corners, and recency visited. For the surround process, the features extracted are intensity, motion, corners, and recency visited. Independent of feature, the output from each feature extractor is a mask (sometimes referred to as a feature map), which uses an image or sample frame-of-reference to simplify later operations, like normalization and combining features together.

The current work uses absolute feature values, which is certainly an oversimplification. For future work, a more meaningful application of the feature detectors would be through a "deviation from typicality" approach. This type of approach would use build up a "typical" feature map (e.g., history of features) to identify when a new feature value deviates from established typicality or expectation. In many ways, a deviation from typicality approach would be more biologically plausible and account for other behavioral phenomena such as adaptation.

Normalization of sample: A complication of using multiple features is the process of combining features, which are, by necessity, dissimilar, i.e., not defined over the same range, and not defined over the same dimensions. Dissimilar features are necessary to increase the sensitivity of the attention process to a range of environments, objects, activities, and events. For instance, each color channel is defined over the range 0–255, whereas corners in the current implementation are defined as a two state discrete value 0 or 1, no corner and corner, respectively. Without normalizing or scaling these features to a common scale, the color channels would dominate the corner features of the extracted sample. Because of these dissimilar features, yet the need in the current implementation for a single feature representation for a process, a method of normalizing dissimilar features is necessary.

The normalization method is similar to the process feature map normalization in the previous section. The feature map is divided by the maximum value for the feature range. For the color channels the maximum value is 255 and for the corner features the maximum value is 1. With all features defined over a common range, the individual feature maps can be combined into a single process feature map.

Combining feature maps*:* After completing normalization of each process feature map, the maps must be combined to create a single feature representation for the process to be used by the Artificial Attention agent. As described previously in the Artificial Attention agent section, the process feature maps were combined using a linear weighted summation; other Artificial Attention processes use a similar approach. The features are combined using the linear summation shown in Eqn. 7 with the constraints imposed by Eqns. 8, 9, and 10. However, instead of the process weightings, a_i, the weightings refer to process specific feature weights.

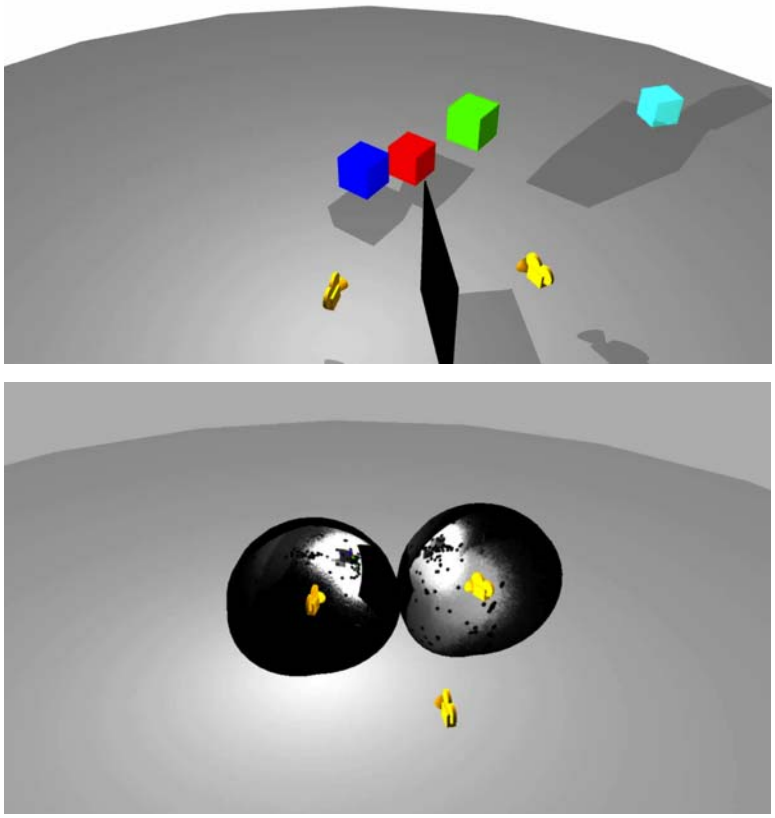

Figure 17: Test Environment Showing Two Points-of-Observation over which Two Separate Artificial Attention Processes are Operating *(top)*
(An attention space created by taking the output of two Artificial Attention processes operating at two different points-of-observation with a third point-of-observation that has a view of the output of the two attention processes)

The feature weights, like the process weightings, are another set of parameters in the Artificial Attention model. The possible distribution of weightings on individual features for a single process are many, however for demonstrating feasibility, the weightings for each feature are set with values that distribute the emphasis evenly across all features for the process. In the future, evaluation of the Artificial Attention algorithm performance with respect to feature weightings is necessary.

The Artificial Attention process is the maximum level of detail provided of the Artificial Attention algorithm. There are additional implementation details, however, the process level is likely sufficient for understanding the basic approach and capability of the algorithm. In the next section, instead of detail, we step back from details of the algorithm to present our current instantiation of Artificial Attention at layered sensing system scales. The current instantiation of Artificial Attention at layered sensing system scales is not the only instantiation possible, but it is a useful one.

6.0 ARTIFICIAL ATTENTION AT BEYOND-HUMAN-SCALES

This section is focused on applying the Artificial Attention algorithm described in the previous sections to a new scale. In the previous sections, the Artificial Attention algorithm was applied to a human scale, which we define as a single point-of-observation. Extending from this definition of human scale, a layered sensing system scale environment operates over multiple points-of-observation. Importantly, a layered sensing system scale environment is not a physical environment, but is the sampled environment created by another sampling process, one operating at a single point-of-observation. To test Artificial Attention at layered sensing system scales thus requires a physical environment with multiple points-of-observation over which sampling processes are operating. An environment that meets these requirements is shown in Figure 17 (top). This is actually the same test environment in Figure 7 with a different viewpoint that reveals the two cameras separated by an obscuring wall (i.e., the black plane). At these two camera positions, we have applied the Artificial Attention algorithm described previously. The result is a sequence of images of the kind shown in Figs. 11–14.

Using the images generated from the Artificial Attention agents operating at these two points-of-observation and a 3-dimensional virtual environment, we create an attention space in which we can instantiate a new point-of-observation. The virtual attention space is shown in Figure 17 (bottom). Depicted are the two original points-of-observation surrounded by spherical surfaces on which the sampling output from the attention processes is projected. The third camera at the bottom of the image is a new point-of-observation in the attention space that is able to observe the output from the first two points-of-observation. An Artificial Attention process operates on the input from this third point-of-observation. This is the simplest sense of an Artificial Attention process operating at layered sensing system scales.

The process of collecting imagery from the point-of-observation and executing the Artificial Attention process over the imagery is identical to the process described at human scales. The resulting data depicted in Figure 18 is presented in a manner similar to results from the Artificial Attention algorithm operating a human scale. The left column of Figure 18 shows the input imagery to the Artificial Attention algorithm, which uses the same Azimuthal equidistant projection. The center and right columns show the same center-surround conditions from the human scale results in Figs. 11–14. These conditions are:

Condition 1: $a_{center} = 0.2,\ a_{surround} = 0.8$

Condition 2: $a_{center} = 0.8,\ a_{surround} = 0.2$

The center column of Figure 18 is Condition 1 and the right column is Condition 2. An additional piece of information is necessary for this beyond human scale demonstration, which is the condition used for the Artificial Attention algorithms at human scale projected onto the spherical surfaces. For demonstrating feasibility we selected Condition 1 ($a_{center} = 0.2,\ a_{surround} = 0.8$), which observed more activities in the environment than Condition 2. The results are similar in nature to those from the

Artificial Attention process operating at human scale. In general, Condition 1 tends to reorient quickly to new objects, activities and events, whereas Condition 2 tends to wander off and focus in areas where there are no objects, activities, or events.

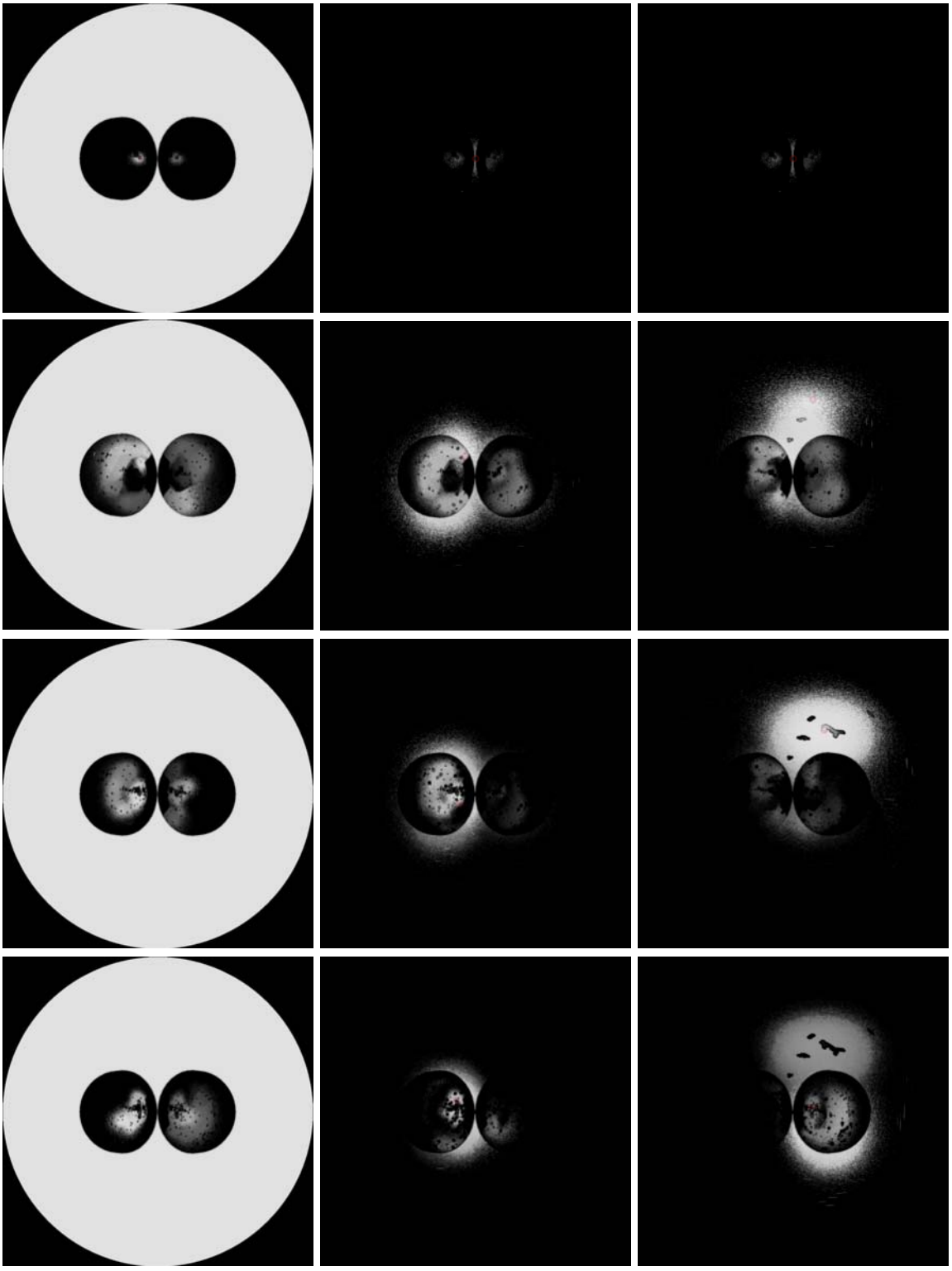

Figure 18: Output of the Artificial Attention Process Operating over the Output of Two Artificial Attention Processes Operating Two Different Points-of-Observation of the Same Environment

It is important to emphasize that what defines an object, activity, or event is now a function of a sampling process. The Artificial Attention algorithm at beyond human scale will respond to an object in the physical environment if sampled by the underlying Artificial Attention agent. However, the Artificial Attention process at layered sensing system scales will also respond to activities like the movement of the sampling process in the environment. This activity is not a physical activity in the physical environment but an attention response of the layered sensing system scale attention agent to the movement of another attention agent, i.e., a kind of joint attention. Restating this activity in another way, the layered sensing system scale attention process will look somewhere because another attention process reoriented to an object, activity, or event in the physical environment. The reorienting attention process is the activity or event of interest.

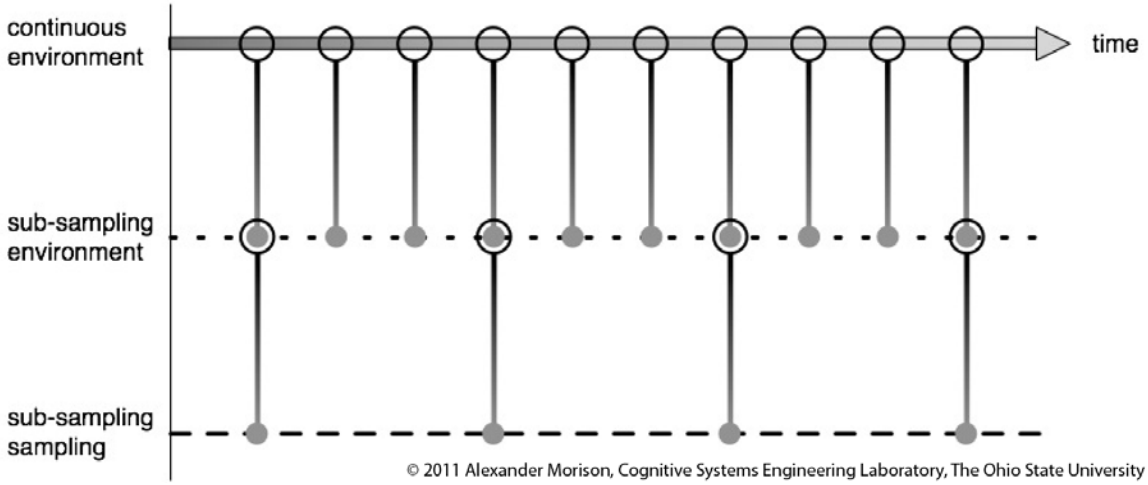

Figure 19: Temporal Sampling of the Artificial Attention Algorithm
(Current implementation of the algorithm reduces the length of the overall image sequence by 1/3rd each time an attention process runs over an image sequence)

Further work is necessary to capture properties of and metrics for Artificial Attention at layered sensing system scales. One step to achieve this goal is to lengthen the temporal span and increase the number of activities and events. The current implementation requires a longer temporal span because the surround process uses three environment images for each time step. The result is a reduction of $1/3^{rd}$ of the total sequence length making the total length approximately 4.4 seconds from a 40 second environment sequence. The process of down sampling due to the rate of execution of the surround process is illustrated in Figure 19. The temporal sampling is another sense of layered sensing system scales. At a human scale there are multiple temporal sampling rates that roughly correspond to saccades of the eye (faster sampling) and perceived movements of the eye (slower sampling). In the current implementation of a layered sensing system scale there is now another sense of sampling rate that emerges from the down sampling that occurs because Artificial Attention agents are nested, which nests attention spaces. The nesting creates a new sense of temporal scale that is slower than the sampling rate at a human scale.

7.0 DISCUSSION AND CONCLUSIONS

The present work on Artificial Attention has achieved several key goals, including demonstrating feasibility, identifying critical aspects of scaling computational models of attention to a layered sensing system, and developing a simulation environment for further development and testing of Artificial Attention algorithms. More specifically, we have:

- Created a computational algorithm connected to a general set of potential feature extractors;
- Instantiated tunable sampling processes;
- Instantiated the two process (center-surround) model for artificial attention;
- Demonstrated that center-surround weightings can influence sampling behavior meaningfully;
- Demonstrated that pacing is a key relationship between a sampling process and an environment.

The current work has demonstrated the feasibility of scaling an Artificial Attention algorithm to operate at the scale of a layered sensing system. We have also shown that creating a scalable Artificial Attention algorithm can still operate at a single point of observation (i.e., a human scale). In the results presented, the scaling of the Artificial Attention algorithm includes both spatial and temporal dimensions. However, the spatial and temporal dimensions are not the only areas of development relevant to expanding human-sensor systems. Other areas of development include technology advances such as sensor modality. An appreciation of the expanding dimensions and scales in human-sensor systems increases the need for new computational models of attention that can operate over these multiple areas of development.

In addition to the present work demonstrating the feasibility of developing an Artificial Attention algorithm to operate at the scale of a layered sensing system, the present work also identified several key aspects of such an algorithm. First, an active sampling process must sample at a rate that keeps pace with activities and events in the environment or space. At these new scales the environment or space may be physical or conceptual, like an attention space. Human attention, as an instance of a successful sampling process, is not a single uniform process, but is more accurately described by at least two separate but interdependent sampling processes. These active processes function as a center process tracking activities and objects and a surround process looking for new events in new places, respectively. Importantly, pace is not a descriptor of a sampling process (like an attention process) or an environment, but is a descriptor of the relationship between the sampling process and an environment. So a pacing measure must describe the match between an attention algorithm, composed of two or more active sampling processes, and an environment, composed of objects, activities, and events. The pacing parameter will be a complex measure of the performance of any attention system (e.g. of situation awareness) and any computational model proposed for attention. The pacing parameter will likely be the most descriptive measure of the performance of any attention system.

Another important aspect confirmed by the initial results is the relationship between center and surround processes. Previous research has established that the influence of each process on the performance of any attentional system varies depending on context and top down goals. Although the feasibility demonstration does not examine this variation in detail, we were surprised by the differential performance in observing and tracking activities exhibited by the algorithm under different weightings. The apparently better performance with greater weighting on the surround reorienting process differs from basic intuition and is a critical initial finding that requires more detailed examination.

Another important aspect of Artificial Attention, which this testing confirms is the dynamic panorama parameter. Woods and Sarter (2010) identified this concept as critical to Artificial Attention and missing from current models. The implementation and testing showed that the dynamic panorama parameter emerges naturally and meaningfully from the interaction between a sampling process and a viewable field that is only partially observable at any given moment. Importantly, these sampling process dynamics exist over time and space. This means the panorama or attention field is not fixed. The parameter of interest then is a measure of the dynamics of panorama extent given top down priorities/goals and the level of activity and change in the environment. The extent parameter can capture both the shape of the panorama (spatial dimension) and the change of the panorama shape over time. Also note that the tests showed that Artificial Attention models can behave such that holes in awareness occur or other problems such as tunnel vision arise. Showing that Artificial Attention algorithms can exhibit such behaviors suggest additional lines of research. For example, further work could identify conditions that are likely to lead to poor sampling, and new results on what creates and sustains attentional pacing could lead to Artificial Attention systems that demonstrate better than human performance (in quality, persistence, and scale).

Artificial Attention is a new frontier in the efforts to escape form data overload and to take advantage of the opportunity created by advances in ubiquitous sensing. It provides a unique research direction to develop new technologies. The simulation approach developed in this work creates a research test bed that can be utilized much further to investigate the performance of Artificial Attention systems, run repeated trials, test parameter settings, and develop new metrics such as pacing and panorama extent.

8.0 RECOMMENDATIONS

The successful demonstration of scaling Artificial Attention to a layered sensing system leads to several practical and research related recommendations to make future progress. Thus far development of the Artificial Attention model has used a functional engineering paradigm, that is, the initial model development is based on the basic structure of attention, and additional complexity is added in to produce the desired functions of the model. Consequently, the model is as simple as possible to produce the desired behavior. Continuation of the functional engineering paradigm is likely to produce a model that combines simplicity to support understanding and functionality to produce desired results. The additional progress in engineering development will require development in metrics like the pacing parameter. Another recommendation is the development and validation of metrics for assessing the performance of computational models of attention operating at scales of a layered sensing system. Previous measures, such as human eye tracks, are potentially relevant, but at layered sensing system scales this is not likely to be useful because the visual field will be highly dependent on the mechanism for navigating a representation, like a wide-area image. Overall, the present work has not demonstrated the impact of the Artificial Attention model on the data overload problem. Performing this assessment is another recommendation from the present work.

In addition to practical recommendations related to performance of the algorithm for supporting exploration of data at layered sensing system scales, there are also significant recommendations for advancing Artificial Attention as a research program. A first recommendation is to examine in finer detail the relationship between center and surround processes and the associated impact on sampling performance. The weighting between center and surround, and how this weighting changes over time, is potentially the most significant parameter for tuning algorithm performance. Indeed, the weighting between center and surround is a potential factor in influencing the top-down component of human attention. This is another research recommendation. Another recommendation is to understand how the dynamic panorama illustrated in this work contributes to, or is a representation of, holes in awareness. That is, can the holes in awareness provide an alternative explanation of attention tunneling? Finally, we recommend including energetics in future development. Energetics captures the cost associated with an attention process. Attention is not free, but requires significant resources and cannot be maintained indefinitely. A model of attention must include some notion of energetics to defend against the notion of infinite resources that dominates computation in general.

In summary, our recommendations are to continue work on the following:

- A functional engineering paradigm
- Pacing as a measure of Attention Algorithm performance
- The overall impact of Artificial Attention on data overload
- Pursue fundamental research questions:
  - What is the relationship between center and surround processes?

- How can top-down attention effects (i.e., endogenous attention) be computationally model
- How do holes in awareness emerge (changing shape of panorama extent)?
- What is the interaction between attention and energetics?

9.0 REFERENCES

Broadbent, D. E. (1958). *Perception and Communication.* Oxford University Press.

Christoffersen, K., Woods, D. D. and Blike, G. T. (2007). Discovering the Events Expert Practitioners Extract from Dynamic Data Streams: The mUMP Technique. *Cognition, Technology, and Work*, 9, 81-98.

Frintrop, S., Klodt, M., & Rome, E. (2007). A Real-Time Visual Attention System Using Integral Images. *Proceedings of the 5th International Conference on Computer Vision (ICVS 2007)*.

Itti, L., & Baldi, P. (2006). Bayesian Surprise Attracts Human Attention. *Advances in Neural Information Processing Systems*, *18*, 547.

Itti, L., & Koch, C. (2001). Computational Modeling of Visual Attention. *Nature Reviews: Neuroscience*, *2*, 194–203.

Koch, C., & Ullman, S. (1984). Selecting One Among The Many: A Simple Network Implementing Shifts in Selective Visual Attention.

Le Meur, O., Le Callet, P., Barba, D., & Thoreau, D. (2006). A Coherent Computational Approach to Model Bottom-Up Visual Attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 802–817.

Morison A. (2010). Perspective Control: Technology to Solve the Multiple Feeds Problem in Sensor Systems. Unpublished Doctoral Dissertation, Ohio State University.

Morison A. and Woods, D.D. (in press). Human-Robot Interaction as Extended Perception. In R. R. Hoffman, P. A. Hancock, R. Parasuraman, J. L. Szalma and M. Scerbo (eds.), *Cambridge Handbook of Applied Perception Research*.

Posner, M. I. (1971). Components of Attention. *Psychological Review, 78*(5), 391–408.

Treisman, A., & Gormican, S. (1988). Feature Analysis in Early Vision: Evidence from Search Asymmetries. *Psychological Review*, *95*(1), 15.

Tsotsos, J. K., Culhane, S. M., Kei Wai, W. Y., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling Visual Attention Via Selective Tuning. *Artificial intelligence, 78*(1-2), 507-545.

Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., & Zheng, S. (2008). Attention-Situation Awareness (A-SA) Model of Pilot Error. *Human Performance Modeling in Aviation*, 213–239.

Woods, D.D. and Hollnagel, E. (2006). *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. Boca Raton FL: Taylor & Francis.

Woods, D. D., & Sarter, N. B. (2010). Capturing the Dynamics of Attention Control from Individual to Distributed Systems: The Shape of Models to Come. *Theoretical Issues in Ergonomics Science, 11*(1), 7–28.

Woods, D.D., Patterson, E.S., and Roth, E.M. (2002). Can we ever Escape from Data Overload? A Cognitive Systems Diagnosis. *Cognition, Technology, and Work*, 4(1): 22-36.

Woods, D. D., Tittle, J., Feil, M. and Roesler, A. (2004). Envisioning Human-Robot Coordination for Future Operations. *IEEE SMC Part C*, 34(2), 210-218.

Yarbus, A. L. (1967). *Eye Movements and Vision*. Plenum Press.

LIST OF ACRONYMS

2D	2 Dimensional
3D	3 Dimensional
A-SA	Attention-Situation Awareness
RGB	Red Green Blue
USAF	United States Air Force