Depth Perception, Cueing, and Control

Barbara T. Sweet[*] and Mary K. Kaiser[†]
NASA Ames Research Center, Moffett Field, CA 94035

Humans rely on a variety of visual cues to inform them of the depth or range of a particular object or feature. Some cues are provided by physiological mechanisms, others from pictorial cues that are interpreted psychologically, and still others by the relative motions of objects or features induced by observer (or vehicle) motions. These cues provide different levels of information (ordinal, relative, absolute) and saliency depending upon depth, task, and interaction with other cues. Display technologies used for head-down and head-up displays, as well as out-the-window displays, have differing capabilities for providing depth cueing information to the observer/operator. In addition to technologies, display content and the source (camera/sensor versus computer rendering) provide varying degrees of cue information. Additionally, most displays create some degree of cue conflict. In this paper, visual depth cues and their interactions will be discussed, as well as display technology and content and related artifacts. Lastly, the role of depth cueing in performing closed-loop control tasks will be discussed.

I.   Introduction

Although popular culture tends to equate the perception of a three-dimensional (3-D) world with the use of stereo displays, the fact is that the human visual system derives depth and distance information from a variety of sources, and stereo vision (stereopsis) is not necessary and sometimes not sufficient for the creation of a robust and reliable depth map. Depth cues have been traditionally grouped by the underlying sources of information. Thus, this taxonomy differentiates between Primary (or Physiological) cues (which are derived from the physiological mechanisms of accommodation, convergence, and stereopsis) and Secondary (or Pictorial or Psychological) cues (which can be conveyed via pictorial images, as Renaissance artists discovered). More recently, the traditional taxonomy has been expanded to include the depth and motion information contained in optic flow, particularly motion parallax and radial expansion.

Beyond enumerating the various cues and their underlying sources, psychologists and simulator engineers have become increasingly interested in how humans integrate these sources of depth information to locomote and navigate through the environment. Thus, it has become increasingly important to understand what information each cue can provide within a particular operational context, and how useful that cue is in concert with other cues. We will discuss how humans' sensitivity to some cues diminish or increase with the distance to the target, while other cues' utility is constant across distance. Also, we will consider the nature of the information a cue can provide; some cues provide only ordinal information (e.g., "Object B is behind Object A"), others provide relative information (e.g., "Object B is twice as far away as Object A"), while still others can specify absolute information (e.g., "Object B is 2 units away and Object A is 1").

We then consider how depth perception is impacted when cues become impoverished, absent, or set in conflict with one another. While impoverishment and absence can occur in the natural environment (e.g., due to darkness or fog), cue conflict is most often the artifact of displays created by artists, psychologists, and display technologists such as simulator engineers. For as diligently as we strive to create high-fidelity simulators that recreate the perceptual experiences of the operational flight environment, limitations in current display and image generation technologies inevitably result in anomalous depth cues. Our challenge, then, is to understand the normative use of these cues in order to design simulation systems that minimize artifactual impact on control task performance, both in the simulator, and when transferring skills to the actual flight environment.

Let us begin, then, with a consideration of how we humans recreate our three-dimensional world from the light that enters our eyes.

---

[*] AST Man-Machine Systems, Human Systems Integration Division, MS 262-2, member.
[†] Research Psychologist, Human Systems Integration Division, MS 262-2, non-member.

II. Human Depth Perception

In popular culture, the perception of three-dimensional (3-D) space is often associated with stereovision. However, stereopsis is only one of many depth cues that inform our visual system. After all, the world does not collapse into a flat plane when we close one eye. Further, most people lacking functional stereopsis manage to lead safe, happy, and productive lives. In fact, some are unaware of their visual defect until they realize that they really are not getting their money's worth at "3-D" movie theaters.

A. Depth Cue Taxonomy

The traditional taxonomy of visual depth cues distinguishes between Primary (Physiological) and Secondary (Pictorial or Psychology) cues. These cues provide usable depth information in static situations, i.e. when there is no relative motion between the observer and the environment. Motion cues make up a third set of cues that require relative motion between the observer and the objects (or objects) being observed.

1. Primary Cues

The Primary depth cues are *accommodation*, *convergence*, and *stereopsis* (i.e., *binocular disparity*). These cues are considered "primary" as they result from direct "physiological signals" of the visual system. Accommodation cues come from the signals given to and received from the eye muscles that control the thinning and thickening of the lens as one focuses on more distal or close objects (Fig. 1). These signals consist of both efferent (transmitted from the brain to the muscle) and afferent (transmitted from the muscle to the brain) nerve impulses. Similarly, convergence cues come from the efferent and afferent signals of the ocular muscles that rotate our two eyes to focus on the same target (Fig. 2). The vergence angle of the eyes correlates to the distance from the observer to the target.

Figure 1. Visual accommodation. *To shift the eye's focus from a distant target (top) to a close target (bottom), the eye muscles controlling lens thickness receive efferent signals. As they execute this neuromuscular command, afferent signals are returned to the brain.*

Unlike these first two primary depth cues, stereopsis information is not produced by neuromuscular signals of the eye. Rather, stereopsis capitalizes on the disparities in the visual images that the left and right eyes capture as they gaze upon a scene. These subtle difference in both between-object and within-object geometries allow our visual system to reconstruct the underlying spatial relationships from this binocular disparity.

Subjectively, stereopsis is a highly compelling depth cues (which is why most of us *do* feel we are getting our money's worth at 3-D movies). And while the process of stereopsis is strongly anchored, like accommodation and convergence, to neurological mechanisms, the informational basis of stereovision is strongly linked to the same disparity cues that underlie motion-based depth cues. We will discuss these motion-based cues later, but first will consider Secondary (or Pictorial or Psychological) depth cues.

2. Secondary Cues

As the reader may suspect, the traditional taxonomy of depth cues assigns a higher level of importance, or at least validity, to the Primary depth cues. This is largely due to the fact that, when the taxonomy was developed in the 19th century, artists had already learned how to create compelling "illusions" of depth on flat canvases. Because these cues seemed easy to artificially generate and manipulate, they were regarded as Secondary cues (easily exploited in Pictorial displays and considered open to Psychological interpretation). Today, perceptual psychologists recognize that these cues reflect critical geometric regularities (or *invariances*) of our natural spatial environment. The fact that
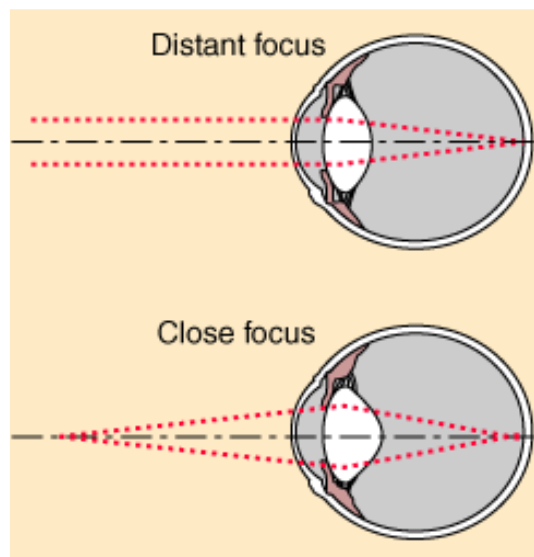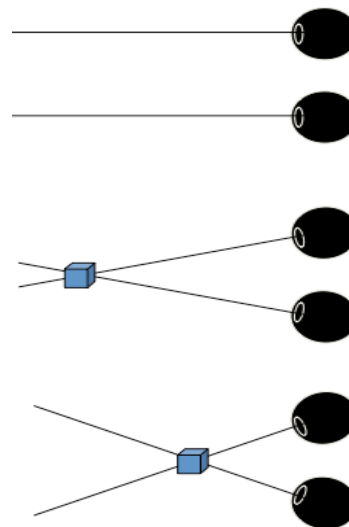

Figure 2. Convergence. *The vergence angle increases as the eyes shift their gaze from a distant target (top) to progressively closer targets (middle and bottom).*

American Institute of Aeronautics and Astronautics

they can be recreated via a pictorial display does not decrease their validity any more than the fact that stereo disparity can be recreated in a "3-D" display.

Image size is a compelling pictorial cue, especially when depicting objects of known, or assumed similar, size. Consider, for example, the line of airmen shown in Fig. 3a. As the men are assumed to be of similar heights (at least within the norms of the USAF), the man on the left appears farther away than the man on the right. The image size cue is most effective when in the context of (and consistent with) other linear perspective cues. Taken in isolation (as shown in Fig. 3b), image size becomes ambiguous. Further, if pitted against false perspective cues, as in the Ames Room demonstration (see Fig. 4), image size is interpreted to correlate with object size rather than distance, even when viewing familiar objects with known size.

The Ames Room demonstration is compelling because the room has been constructed so as to systematically distort the depth cues typically provided by *linear perspective*, i.e., that parallel lines converge at the horizon and height in the picture plane correlates with depth. Thus, in images of normal environments, roads and runways form trapezoids, and a nearer object's base is further from the horizon than the farther object's base (e.g., the airmen's boots in Fig. 3a).

Texture gradients are typically cited as a separate cue to depth although, logically speaking, they are simply a case of image size. However, because the texture is continuous, the image gradient provides continuous depth scaling – an important cue feature. Further, because gradients are most often on the ground plane, this cue provides critical information for ground-referenced navigation and actions.

Another strong visual cue to depth is *occlusion*: closer objects occlude more distant objects along the line of sight. Pragmatically, partial occlusion is most useful, as some of the more distant object is still visible, yet is clearly behind the closer, occluding object. As with the Ames Room, it is possible to construct visual "illusions" that violate our interpretation of occlusion, but doing so requires the use of bizarrely contoured objects, unlike any the observer is likely to encounter outside of perceptual psychology demonstrations and clever works of art. Given the world we live in and the objects we typically see, occlusion-specified depth is an unambiguous cue.

Occlusion is also often complemented by another depth cue: *shadow*. When, for example, an illumination source (e.g., the sun) is behind the observer, a closer object will not only occlude part of an object; its shadow will fall on (or toward) the more distant object. Of course, shadows can provide an effective depth cue even in the absence of occlusion, as Fig. 5 demonstrates.

The final pictorial depth cue in the traditional taxonomy is *aerial*

a)

Figure 3. Image size. *When consistent with other linear perspective cues (a), image size is a strong cue to object depth. When viewed in isolation (b), image size become more ambiguous, even with objects of known, or assumed similar, size.*

Figure 4. The Ames Room demonstration. *False linear perspective created by the room's design (and specific to a single design eye-point), creates a compelling visual illusion.*
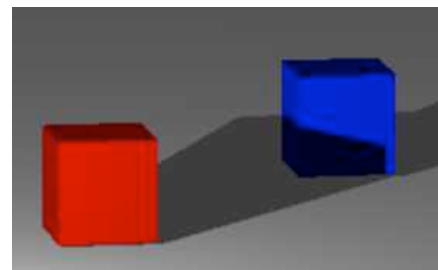

Figure 5. Shadows. *With the source of illumination coming from behind (and to the left) of the observer, the shadows reinforce other depth cues in specifying that the blue cube is further away than the red cube.*

(or atmospheric) perspective. This cue is limited to objects viewed at fairly large distances as it depends on the effect the air between the observer and objects has on image quality. In general, as distance increases, the contrast and hue saturation of an object's image will decrease. The impact of this attenuation will depend on the particulate and moisture content of the air. Aerial perspective is far more pronounced when viewing the Blue Ridge Mountains in Virginia than Superstition Mountains in Arizona (Fig. 6), and the effect is even more pronounced on a foggy day in San Francisco. Although the degree of atmospheric attenuation depends on atmospheric conditions, our visual systems expect *some* to occur. The complete lack of aerial perspective on the Moon was quite striking and made reliable distance estimation difficult.

3. Motion Cues

One limitation of the traditional taxonomy is that it delineates only static depth cues, that is, those available to the unmoving observer. Again, this reflects the theoretical biases of the era in which the taxonomy was developed. Of course, psychologists of the 19th century realized that human observers were capable of creating and perceiving motion. However, moving stimuli were considered a complication for the visual system rather than a unique source of information – motion was derived, not perceived. It was not until James Gibson's pioneering work that psychologist seriously considered the depth information provided by motion.[1,2]

This oversight is especially egregious if one considers how similar the most important motion-carried depth cue (motion parallax) is to one of the most compelling static depth cues: binocular disparity, the basis of stereovision. As we discussed, stereopsis emerges from the visual system processing the subtle disparities between the visual images sampled by the left and right eye. In a similar vein, motion parallax utilizes the image transformations resulting from observer motion. In principle, binocular disparity is equivalent to two "snapshots" sampled

Figure 6. Aerial (or atmospheric) perspective. *This cue is more useful in the Virginia Blue Ridge mountains (top) than in the Arizona desert (bottom). With minimal moisture and particulate matter in the air, there is little attenuation of image contrast and hue.*

from a lateral head motion. Of course, observer motion is not limited to this single degree of freedom; further the information feed from motion parallax is continuous. Thus, in terms of an informational analysis, motion parallax provides a far richer cue to depth than stereopsis, yet its utility was recognized (and empirically investigated) nearly a century later.[3]

Studies of depth-from-motion investigate object motion as well as observer motion. For example, when an object rotates, its shape (that is, within-object depth relations) is uniquely specified. Further, when an object moves at a given speed in the environment, its image speed correlates to its distance from the observer. Our focus, however, will be on the depth information generated by observer motion.

Although geometrically related, perceptual psychologists tend to differentiate between depth information resulting from movement perpendicular to the observer's line of sight (i.e., horizontal or vertical motion parallax) from the depth information revealed by movement along the line of sight (i.e., radial flow). Both are subclasses of the optic flow that is generated when observers move about their environment.

As mentioned, the information carried in motion parallax is similar in content to that provided by binocular disparity. Both reveal depth information via changes in image distances of inter-object boundaries when the eye-point is displaced. However, binocular disparity contains only the inter-object depth information that can be extracted from static disparity between the two discrete images sampled by our left and right eyes. In contrast, motion parallax reflects continuous dynamic transformations of inter-object boundaries. Stereo images can be (and have been) created by sampling two "frames" of motion parallax (i.e., capturing a second image when the eye-point has moved a distance equal to the inter-ocular separation), but motion parallax captures a much larger (and continuous) range of eye-points.

4

The second component of optical flow is radial flow (or expansion). A forward-looking observer's locomotion creates image motion that radiates outward from their line of sight (i.e., their track vector). The velocity of an object's image depends on its distance from this focus of expansion and its distance from the observer. For example, if two objects are at the same initial distance from the observer, the image of the one further from the observer's track vector will have a greater velocity. Likewise, if two objects have the same initial angle relative to the observer's track vector but lie at different distances, the closer object will have the greater radial image velocity. Thus, a metric depth map can be recovered from the objects' radial flow rates relative to the objects' bearings (i.e., angles relative to the track vector). In fact, if the observer is moving at a constant velocity, the bearing angle divided by its rate of change can inform the observer of the amount of time until the object will pass his or her viewing plane.[4] (Unfortunately for collision avoidance, an object directly along the track vector will not have radial flow, so time-to-collision is specified only by the object's image expansion[5]– a far less salient source of information.)

We see, then, that image motions created by an observer's movement through the environment provide as rich of a source of depth information as the traditional primary and secondary cues. The next question to consider is how human observers utilize these cues to perceive a three-dimensional world. For while our taxonomy provides a useful description of the sources of depth information, it provides little insight into how observers utilize these cues to build their visual world. Understanding depth-cue usage requires us to consider other aspects of these information sources. Specifically, we will address three aspects: 1) how sensitive observers are to various depth cues; 2) the nature of the information conveyed by the various cues; and 3) how the demands of specific tasks will determine which depth cues are most useful and robust.

B. Depth Cue Usage

1. Sensitivity

In comparing binocular disparity and motion parallax, we noted that motion parallax holds two informational advantages. First, motion parallax captures the dynamic boundary-distance transformations rather than two static disparity samples. Second, and perhaps more significant, binocular disparity is constrained to a single level of disparity – that created by the distance between our two eyes (i.e., the inter-ocular distance or IOD). Simple geometry dictates that the disparity caused by our IOD decreases as the comparison objects become more distant. Thus, there are distance limits on our functional stereopsis. The exact limit depends on individuals' stereo-acuity, but generally stereopsis becomes useless for judging depth beyond ~10 m. Similarly, the neuromuscular information provided by accommodation and convergence falls off with distance. Conversely, aerial perspective is useful only at relatively large distances, while the utility of occlusion, image size, and shadowing are not impacted by distance.

Reference 6 provided a useful meta-analysis of the depth-cue psychophysical literature that plotted out which depth cues were useful at what distances. (By "useful" we mean that human observers are sensitive to the cues, i.e., that they are "supra-threshold" at these distances.) Reference 7 expanded on the analysis in Ref. 6 by introducing the notion that humans divide their spatial environment into

Figure 7. Sensitivity to depth cues. *Meta-analysis of psychophysical data in Ref. 6 maps out the difference in depth ($\Delta d/d$) required for depth cues to be supra-threshold as a function of distance from the observer. These sensitivities were then mapped into a three-space model in Ref. 7.*

three regions: 1) Personal Space, in which we tend to manipulate objects with our hands and perform fine motor tasks; 2) Action Space, in which we walk, run, and perform gross motor tasks; and 3) Vista Space, in which we orient to the larger environment and plan paths and routes. These regions were then superimposed onto refinements
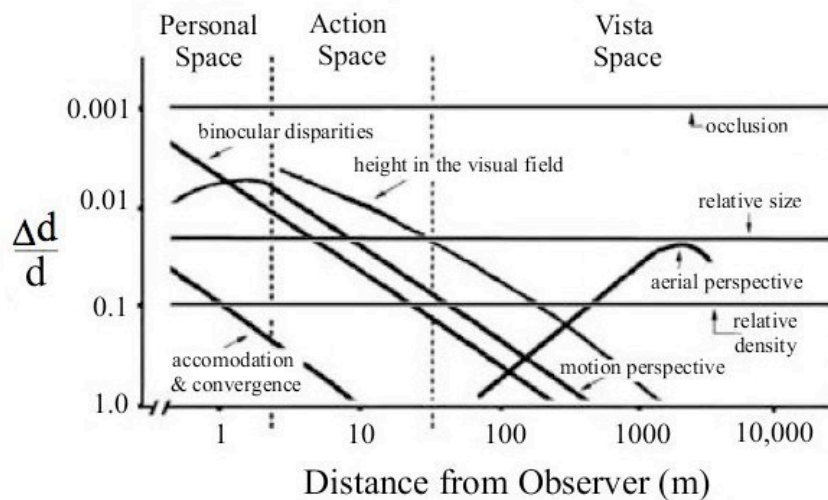
5

of the plots in Ref. 6 to determine which depth cues were likely to play prominent roles in each space (Fig. 7). This approach helped psychologists move beyond the notion of determining which cue (or cues) played *the* dominant role in depth perception. Instead, it leads toward a more nuanced understanding that our visual system is adaptive and dynamic, and will exploit those cues that are most functional in the space of current interest.

2. Information Characteristics

Another reason to reject the search for a dominance order among depth cues is the fact that, in addition to varying in strength and reliability at varying distances, different cues provide different *kinds* or *levels* of information. Occlusion, for example, is a highly reliable cue and functions equally well at near and far distances, but it provides only ordinal information. That is, it definitely specifies that Object B is behind Object A but, as an isolated cue, gives no information regarding how far behind Object B is. Other cues, such as texture gradient and image size, provide *relative* depth information (e.g., the distance from the observer to Object B is twice as large as the distance to Object A, but neither can specify whether those distances are 5 and 10 m or 20 and 40 m).

As we have discussed, optic flow can provide absolute depth information in a temporal domain (e.g., if constant velocity is maintained, the observer will pass Object A in 3 seconds and Object B in 6 seconds). And for many navigation and way-finding tasks, such temporal scaling is fully adequate for orchestrating control of our body's or vehicle's locomotion. In terms of providing absolute distance information, only the Primary cues (accommodation, convergence, and stereopsis) are available, and only because they are calibrated to our bodies (i.e., the IOD and neuromuscular signals). In principle, if one knows the velocity generating motion parallax, one can recover absolute depth and distances. Similarly, linear perspective cues can be scaled to eye-height that, if known, can be scaled to absolute values. (For many body-referenced actions, we seem to scale objects and distances in body-referenced metrics such as eye-height. For example, a steady gait moves us at two eye-heights per second, we can manage a step-up of one-third eye-height, and feel comfortable passing through a lateral gap of half an eye-height.)

Given the diverse nature of the information that the different depth cues provide, our visual system would benefit from leveraging these multiple sources. Reference 8 provides a useful discussion of the properties of different cue-fusion models, and introduces the notion of cue promotion (e.g., a cue-combination process that "promotes" a powerful but ordinal cue such as occlusion by utilizing relative or absolute depth information from another cue). In addition to providing a descriptive model of human depth perception, such cue combination research can suggest prescriptive algorithms for sensor fusion in artificial vision systems.

3. Task Demands

While the approaches of the above researchers provide important advances over traditional "which cue is dominant" studies, they still reflect an implicit assumption that the observer is a either a passive agent or, in the case of Cutting and Vishton, that the task one is performing in personal, action, or vista space has no influence on the utility of the various depth cues. This is hardly surprising given the fact that experimental paradigms in perceptual psychology have tended to record visual judgments (e.g., "which object is closer?"; "if Object A is 1 unit away, how far away is Object B?") rather than measure coupled perception-action performance.

We have performed a number of studies examining the use of depth cues in a visually guided closed-loop control task.[9] Our findings confirm both the notion that sensitivity influences cue usage (i.e., binocular disparity has a greater influence when the objects are closer) and that cues are combined rather than a dominant cue selected. But our results also showed that the control system's dynamics (i.e., whether participants controlled the velocity or the acceleration of the test object) influenced which depth cue proved more powerful. In brief, when participants controlled velocity, binocular disparity was dominant; when they controlled acceleration, relative size was dominant. This may reflect the fact that stereopsis, while providing a compelling depth cue, is a relatively slow visual process. It may not be able to provide information at a sufficient rate for an acceleration control task.

Our studies demonstrate both that the requirements of a task play a role in how people use the myriad depth cues available in our environment, and that our visual systems seamlessly adapt our cue combination to meet these task demands. Thus, human depth perception utilizes a number of primary, secondary, and motion depth cues in a manner well suited to the availability and utility of the cues as well as appropriate to the demands of the task.

We now turn to the issue of how we can use modern technology (i.e., image generators and displays) to recreate critical depth cues for human-in-the-loop flight simulation. We also consider cases in which limitations in these technologies are likely to produce ambiguous or conflicting depth cues, and the possible impact on operator performance.

III. Displays

It is the display that affects the Primary depth cues (*accommodation*, *convergence*, and *binocular disparity*) as well as *depth-from-motion* resulting from observer head motion. Specific display technologies are capable of

providing the information for depth perception resulting from these cues to varying degrees. A summary of various display types, and their capability of producing depth cues, is presented in Table 1. The discussion will focus on specific display types, and display functions that can be implemented on any of the display types.

Table 1. Depth cueing capabilities as a function of display type

Cue	Display Type			Display Function			
	Real-Image	Collimated	HMD	No Stereo	Stereo	Not Head-Tracked	Head-Tracked
Accommodation	Fixed (screen distance)	Fixed (collimation distance)	Fixed (function of optics)				
Convergence				Accommodation Distance	Correct		
Stereopsis				None	Correct		
Motion (resulting from head motion)						None	Correct

C. Display Types
1. Real-Image
In a real-image display, the image is viewed directly from a display surface, such as a projection screen, or direct-viewing of a display. In these cases, the viewer is focused directly on the screen or display surface, and both lens accommodation and convergence will be based on the screen distance. Distance to the screen or display varies significantly in the simulation community, from front-projected large-radius dome displays, to rear-projected faceted screens viewed from a short distance (as in Ref. 10), to direct viewing of an illuminating display surface. Significant distortion can occur when the display is not viewed from the design eye reference point.

2. Collimated
Collimated displays are viewed through optical elements that collimate, or make parallel, the incoming light rays. A perfectly collimated display will optically behave as if the display elements are infinitely far away. In practice, most collimated displays are not designed for optical infinity, but for an approximately 40-foot distance. Both lens accommodation and convergence are consistent with this optical distance.

Collimated displays are often credited with providing "eye relief" to the pilot (i.e., creating convergence/accommodation consistent w/ "far" out-the-window viewing), but they provide another characteristic that is potentially more important in many applications – image stability. Naturally viewed objects that are distant remain stable in the image relative to translational movements of the observer. For example, if a pilot is referencing the horizon to maintain pitch and roll attitudes, any movement of the horizon in the pilot's viewpoint is accurately interpreted as vehicle motion. This will remain true in spite of any translational motion of the pilot's head. However, in a real-image display with relatively close display surfaces, small movements of the pilot's head will confound interpretation of changes in horizon position; it could have moved as a result of the pilot head motion. This was thought to be a factor in an experiment conducted in Ref. 11, which compared hovering performance with a real-image and collimated display. The pilot's control of attitude rates and translational velocities was superior with the collimated display.

3. Helmet-mounted Displays
A third method for image presentation is head- or helmet-mounted. Given the proximity of the operator's eyes to the display(s), some sort of accommodative relief is required. Hence, many helmet-mounted displays provide collimation. Also, because most helmet-mounted displays present independent screen (or screens) to each eye, it's relatively simple to provide separate images directly to the two eyes, thus providing stereo capability. One potential limitation of helmet-mounted stereo displays, however, is the achievable disparity threshold due to resolution of the displays. At this time of this report, typical helmet-mounted displays can achieve resolutions in the range of 1.6[‡] to 4.0[§] amin/pix. Projected displays, on the other hand, can provide higher levels of resolution by adjusting the image size and distance of the viewer to achieve a smaller amin/pixel.

[‡] Rockwell Collins SimEyeSX45, 36 h x 29 v deg field of view, 1280 x 1024 pixels
[§] Sensics piSight, 179 h x 60 v deg field of view, 2600 x 1200 pixels

D. Display Functions

1. Stereo

In this paper, the term "stereo display" is used to classify any type of display that is capable of producing *binocular disparity*. There are many different methods available to provide disparity in displays, and much recent development has occurred aimed at increasing the consumer market for 3-D displays and content[**]. The utilization of 3D display technologies is quite different between the aerospace simulation community and the consumer entertainment industry. As a result, many display technologies currently used and developed for consumer market are not ideal for aerospace applications.

In order to create binocular disparity, it is necessary to provide different images to the two eyes, the disparity between features in the images based on the depth of the imaged features. Some commonly used methods in the simulation community include projected displays, which provide separate images through temporal multiplexing (active stereo); projected displays, which provide separate images through spatially-aligned displays (passive stereo), and helmet-mounted displays.

2. Head-tracked

Most simulation applications do not provide *depth-from-motion* cues for head-related motion (vehicle-related depth-from-motion is provided through the image content transforming to reflect vehicular motion). However, it is possible to provide this cue if the position and orientation of the head is measured and used to modify the computer-generated image. Although an attractive concept, it is difficult to achieve satisfactory results. Studies have shown that humans are highly sensitive to the visual artifacts resulting from time delays and/or noise in the head position/orientation measurement.[12]

IV. Image Content/Image Generators

Real-time computer generated scene content can provide the Secondary depth cues, and *depth-from-motion* cues related to vehicle motion. *Image size*, *linear perspective*, and *occlusion* are all generated correctly if the appropriate features are enabled in the hardware and software (such as depth mapping to provide occlusion). Depth-from-motion is likewise correctly depicted for vehicular motions.

Computer generated scenes do not typically provide correct *texture gradient* cues. Although initial usage of computer generated scene content in simulators was limited to large, plain, polygonal features, highly-realistic visual textures are available in current systems. Photorealistic textures are commonly used, often to provide realistic ground detail. Although use of this technology can create very compelling scenes, there are details regarding the implementation that can create a different experience from natural viewing of texture in a scene. As indicated previously, depth information from viewing a texture gradient is a special case of relative size cues with a regularly repeating feature. Concrete, asphalt, grassy or rocky vistas are all examples of surfaces that provide depth information from a texture gradient. However, a naturally-viewed texture will typically have features at eye-limiting resolutions at all locations in the image. Although an observer can typically expect to see more fine detail in the closest surfaces and objects in a naturally viewed scene, computer generated scenes frequently do not provide this characteristics. Practical limits on the amount of texture information available in a computer database limit the amount of scene detail that can be shown, and this is most visible at the closest approach to surfaces. Methods to improve the near-field scene detail include mapping additional layers of finer texture (i.e., level-of-detail rendering) and implementing textures that are computed in real time to increase level of detail. Advanced graphics

Figure 8. Simulated aerial refueling scene. *The shadow of the boom on the aircraft provides an important relative depth cue for the boom operator. Image generated by MetaVR's Virtual Reality Scene Generator^TM (VRSG^TM).*
http://www.metavr.com/technology/aerialrefueling.html

[**] The 3D@Home consortium is a good source of educational material and recommendations regarding 3-D technologies and content, but is specifically targeted at entertainment applications. www.3dathome.com.

methods, often implemented or accelerated in hardware in state-of-the-art graphics cards, are frequently introduced in video games long in advance of implementation in the aerospace-related visual simulation realm.

In the current state-of-the-art of the industry, it is possible to incorporate *shadows*, but is not always done because of the level of complexity. Typically, shadows are incorporated in a simulation on a case-by-case basis when they are a necessary component for training or realism. Figure 8 demonstrates one such application, in which shadows provide important depth cues for both the pilot and boom operator in an aerial refueling mission.

Photorealistic textures (from satellite or aerial photography) often have shadows cast by trees, clouds, and buildings that are a part of the photograph. This can cause a non-realistic shadow, or even a conflict between computer-generated shadows and shadows present in the photographic texture content.

Aerial perspective/atmospheric cueing is available in many image generation systems. Recent developments in pixel shading algorithms have allowed the development of sophisticated atmospheric cueing, including volumetric, non-homogenous clouds and stratified fog layers.

Although not yet adopted in the aerospace visual simulation community, as specialized graphics hardware and computational power grow, it is increasingly possible to do real-time ray tracing that can provide more realistic textures, shadows, 'soft' shadows, and atmospheric cues.

V. Effective Use of Simulator Resources

As discussed previously, the majority of the secondary cues are provided in any image generator. Veridical texture mapping, particularly in the near-field of objects or environmental features with large extent in depth, is still not easily achieved (if at all), but is one of many pictorial cues which are typically present to inform depth judgments. Providing global real-time shadowing is not achievable, but when particular shadows are important cues (such as in aerial refueling) they can be provided with sufficient programming and processing resources.

Greater system complexity is required to support stereopsis and depth-from-motion perception. Stereopsis requires a stereo display, requiring more processing resources (or less scene complexity) as well as specialized hardware. Active stereo requires eyewear and an emitter to time the opening and closing of the shutter glasses, as well as a higher level of processing power to generate two eye views rather than one. Passive stereo requires double the projectors as well as double the processing to generate the images. Polarizing filters on the projectors, and polarizing eyewear, is also required. The third option, a helmet-mounted display is a similarly complex technology.

Head-tracking to provide depth-from-motion cues is a technically complex task. Part of the complexity arises from the need to calibrate the system to provide accurate position and orientation information. Additionally, effective use of head-tracking to produce head-related motion parallax requires extremely low latencies. Observers are capable of detecting latencies as low as 7 ms, and generally latencies need to be below 30 ms for the observer to interact naturally with the visual environment. As latency increases, observers begin to slow their head motions in order to minimize the disorienting affects of misregistration, which rather defeats the purpose of providing head-related scene motion.

While stereo displays will provide veridical convergence cues as well as binocular disparity, no display provides correct accommodation information across the entire depth range. This is perhaps not a concern given that both convergence and accommodation have been shown to be relatively weak cues relative to other depth cues (Fig. 7). When stereo is used, care must be taken to match the accommodation/convergence as much as possible; poor accommodation/convergence matching is has been demonstrated to produce visual fatigue as well as image distortion[13].

Whether to expend additional resources to produce binocular disparity and/or depth-from-motion cues needs to be considered from a cost-benefit ratio. As shown in Fig. 7, these cues are thought to be most effective in the previously defined 'Personal Space' and 'Action Space', up to 10 meters. However, in operational communities, effective stereopsis is considered to be beneficial at much larger distances, up to 200 m. Flight operations thought to benefit from stereopsis include taxi, takeoff, landing, formation flight, and aerial refueling. In fact, the military vision standards for stereopsis in pilots is extremely high; the Air Force requires discrimination of 25 arc-seconds of disparity, the Navy and Army require 40 arc-seconds. The value of providing stereopsis and/or depth-from-motion cues for operational training has yet to be established.

VI. Conclusion

The human visual system is sophisticated and opportunistic in its use of depth information. It is able to integrate a number of cues, taking into account each cue's reliability and applicability for the current operational task. And while our visual system is accustomed to adapting to impoverished or degraded depth cues, it is only in synthetic displays (from early cave drawings to high-tech flight simulators) that these cues are placed in conflict. As

designers, we need to be aware of the limitations of our technologies with regards to producing veridical depth cues, and strive to provide those that are necessary and sufficient to perform the simulated operations. Our failure in this domain will create inferior simulator systems, which can lead to incomplete training or even negative transfer of training. Our success will create effective simulator systems that allow pilots to learn and explore operational scenarios in a safe and controlled environment, and ensure the efficient investment of hardware and software resources.

Acknowledgments

References

[1]Gibson, J. J., *Perception of the Visual World*, Houghton Mifflin, Boston, 1950.

[2]Gibson, J. J., *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, 1979.

[3]Cutting, J. E., *Perception with an Eye for Motion*, MIT Press, Cambridge, MA, 1986.

[4]Kaiser, M. K., and Mowafy, L., "Optical Specification of Time-to-Passage: Observers' Sensitivity to Global Tau," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 19, No. 5, 1993, pp. 1028-1040.

[5]Lee, D. N., "A Theory of Visual Control of Braking Based on Information about Time-to-Collision," *Perception*, Vol. 5, No. 4, 1976, pp. 437-459.

[6]Nagata, S. "How to Reinforce Perception of Depth in Single Two-Dimensional Pictures," *Pictorial Communication in Virtual and Real Environments*, edited by S. Ellis, M. Kaiser, and A. Grunwald, Taylor & Francis, London, 1991, pp. 527-545.

[7]Cutting, J. E., and Vishton, P. M. "Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information About Depth," *Handbook of Perception and Cognition: Perception of Space and Motion*, edited by W. Epstein and S. Rogers, Academic Press, San Diego, CA, 1995, pp. 69-117.

[8]Landy, M. S., Maloney, L. T., Johnston, E. B., and Young, M., "Measurement and Modeling of Depth Cue Combination: In Defense of Weak Fusion," *Vision Research*, Vol. 35, No. 3, 1995, pp. 389-412.

[9]Sweet, B. T., and Kaiser, M. K., "Integration of Size and Binocular Disparity Visual Cues in Manual Depth-Control Tasks," *AIAA Modeling and Simulation Technologies Conference and Exhibit*, Keystone, CO, 2006.

[10]Wright, D. R., Best, L. G., and Peppler, P. W., "M2DART: A Real-Image Simulator Visual Display System," U.S. Air Force Research Lab., AFRL-HE-AZ-TR-1998-0097, Mesa, AZ, Nov. 1998.

[11]Chung, W., Sweet, B., Kaiser, M., and Lewis, E., "Visual Cueing Effects Investigation for a Hover Task," *Proceedings of AIAA Modeling and Simulation Technologies Conference*, Austin, TX, 2003.

[12]Ellis, S. R., Mania, K., Adelstein, B. D., and Hill, M. I., "Generalizability of Latency Detection in a Variety of Virtual Environments," *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*, 2004, pp. 2632-2636.

[13]Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S., "Vergence-Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue," *Journal of Vision*, Vol. 8, No. 3, 2008, pp. 1-30.