S. Soon, A. Lugmayr, A. Woods, and T. Tan, "Understanding Head-Mounted Display FOV in Maritime Search and Rescue Object Detection," in 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), 2018, pp. 116-119

Understanding Head-Mounted Display FOV in Maritime Search and Rescue Object Detection

Susannah Soon
Curtin HIVE
Curtin University
Bentley, Australia
Susannah.Soon@curtin.edu.au

Artur Lugmayr
VisLab., Curtin Univ., Bentley, Australia & Aalto Univ., Helsinki, Finland
Artur.Lugmayr@curtin.edu.au
http://curtin.edu/vismedia

Andrew Woods
Curtin HIVE
Curtin University
Bentley, Australia
A.Woods@curtin.edu.au

Tele Tan
Dept. of Mechanical Engineering
Curtin University
Bentley, Australia
T.Tan@curtin.edu.au

***Abstract*— Object detection when viewing *Head Mounted Display (HMD)* imagery for maritime Search and Rescue (SAR) detection tasks poses many challenges, for example, objects are difficult to distinguish due to low contrast or low observability. We survey existing *Artificial Intelligence (AI)* image processing algorithms that improve object detection performance. We also examine central and peripheral vision *(HVS)* and their relation to *Field of View (FOV)* within the *Human Visual System* when viewing such images using HMDs. We present results from our user-study which simulates different maritime scenes used in object detection tasks. Users are tested viewing sample images with different visual features over different FOVs, to inform the development of an AI algorithm for object detection.**

Keywords—Head Mounted Display, FOV, human visual system, field of view, AI, central vision, peripheral vision, Human-Computer-Interaction, maritime search and rescue

I. Introduction

Maritime Search and Rescue (SAR) involves scanning an open water scene to achieve situational awareness and identification of objects of interest such as humans, vessels or landmarks. We investigate whether SAR tasks can be improved by viewing real-time still or video imagery captured via 360° panoramic cameras via HMDs. Increasing use of technology such as drones makes possible visual image capture of locations otherwise not accessible.

We use a camera system that performs automated capture of high resolution 8K real-time uncompressed 360° digital video. The camera system's daylight and lowlight sensors allow images to be captured in varying weather and environmental conditions. For example, in rain, fog, or night-time. Playback and magnification are available. Fig. 1 illustrates the different SAR imagery with varying *FOVs.*

Other display modalities including large-scale immersive displays and desktop monitors have been investigated by the authors [1]. Although, large-scale immersive displays provide good speed and accuracy results for object detection, their physical space requirements mean that the HMD modality is more suitable for operational scenarios.

To determine the HMD modality's effectiveness, we investigate visual acuity of the central and peripheral vision of the *Human Visual System (HVS)* by varying the horizontal *Field of View (FOV)* of the scene. We survey existing *Artificial Intelligence (AI)* image processing techniques used for object detection in maritime scenes.

This 'cognitive easy understandable big data' [2] application eases the cognitive load of SAR users where visibility, contrast and brightness are often low, and objects difficult to detect. We discuss the limitations of the human visual system when using HMDs, present a user study to examine different image presentations over different FOVs, and discuss how AI can improve object detection tasks.

II. Maritime Search and Rescue Imagery

Objects can be difficult to observe due varying SAR imagery quality. For example, scenes can be low contrast, low light (fog, cloud cover), night-time, display incorrect brightness (excess sun reflection off water), indistinguishable regions (white sky and snow), or high frequency objects (rain, water and waves). Objects may also be far away (e.g. 1-5 km), small and indistinguishable. The interpretation of the objects can require human verification.

Whilst the horizontal FOV of maritime SAR images is 360°, vertical FOV is limited as the camera array may operate in unstable conditions, where only a narrow vertical band of footage can be accurately stitched and cropped.

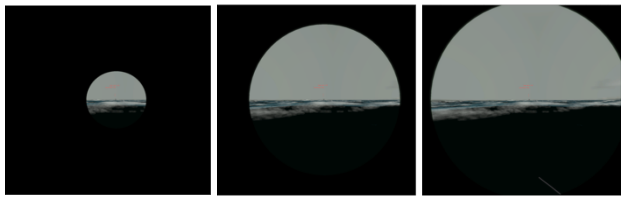


Figure 1. SAR imagery showing 30°, 70°, 85° FOVs.

III. Understanding the Human Visual System

A. Visual Acuity (Pixel Density and Retinal Resolution)

Visual acuity in HMDs can be measured in *Pixels per Degree (PPD)*. Retinal resolution is the pixel density where humans with normal vision cannot see any additional detail, generally 60 *Pixels per Degree (PPD)* (20/20 vision) [3]. Sensors capturing pixel densities greater than 60 PPD capture more information than can be visualised by the average HVS. HMDs offer sufficient spatial resolution to present imagery accurately, most in the range 10-15 PPD (horizontal monocular FOV) e.g. HTC Vive Pro[1] with 13PPD. Visual acuity can be affected by object brightness and speed, peripheral vision and eye movement. Retinal resolution is dependent on viewing distance between the eye and display, also affecting FOV.

B. Central and Peripheral Vision

The HVS provides a high visual acuity, wide FOV (~190°) comprising of central and peripheral vision [4]. High visual acuity central vision is achieved by cone cells in the retina's central fovea (1-2°). Cone cells detect colour, fine detail, and are responsive to stimuli. Peripheral vision uses rod cells which have low visual acuity, low-light

[1] HTC Vive, 2018. Available: https://www.vive.com/au

visibility, and reduced colour perception. Peripheral vision is effective at detecting motion as peripheral flickering [5].

In our experiment, we test central and peripheral vision by specifying the HMD's FOV. In the visual search task, peripheral vision is first used to analyse the scene, areas of interest are identified, and then central foveal visual analysis identifies the object [6][7]. The central foveal region performs spatio-temporal scanning sampling at five samples per second in high resolution. By varying the FOV in our experiment, we examine the capacity of central/peripheral vision in executing the object detection task. Object detection is also dependent on object size, shape, colour, frequency, and likelihood of occurrence.

In addition, we use edge detection algorithms to enhance visibility of features in the image, by looking to work such as Peli et. al. [8][9], who assist those with central vision loss (scotoma) by superimposing a magnified edge detected image over the top of the real-world, viewed via a see-through HMD. This increases the effective resolution of residual peripheral field. They also assist those with tunnel vision by presenting minified edge detected images to increase the horizontal span seen instantaneously.

C. *Field of View (FOV) in HMDs*

HMD's can render a wide horizontal FOV, e.g. the HTC Vive at 110°[2] (although in practice we measure FOV to be 90°). However, this FOV is smaller than a user's natural ~190° FOV, therefore, the user must turn their head to see the full virtual environment (i.e. *Field of Range (FOR))*. Larger FOV and FOR increases the perception of immersion [10]. However, if head motion latency is high, this will reduce immersivity and potentially induce *Virtual Reality (VR)* sickness [11].

We test users viewing scenes with 30°, 70°, 85° FOVs (Fig. 1). These FOVs are chosen based on recommendations in different domains under different conditions, e.g. Patterson et al. [12] suggests a minimum 40° FOV for object identification tasks, and Angel et al. [13] find wide FOV's are beneficial in target detection related tasks in military operations on land. They test day- and night-vision HMDs at 40°, 70° and 95° FOVs in an urban environment, at near (10-30m) and far (31-59m) target ranges. They find FOVs greater than 95° enabled detection tending towards 'normal vision' (~190° FOV) and showed significant improvement over a 40° FOV for targets that are far away and occluded. Ragan et. al. [14] also show that increased FOV allows more targets to be detected in their training task.

IV. AI Techniques for Image Analysis

Computer vision image processing often employs AI in object detection and scene analysis, as e.g. neural network-based image analysis. Of interest are image processing algorithms used to enhance images where FOV is considered e.g. Everingham et. al.'s work [15] builds on traditional methods for image enhancement i.e. adaptive filtering edge detection used by Peli et al. (see [8], [16], [9]). Everingham et al. use image segmentation, feature extraction and neural net classifier to remove noise to enhance images in a scene viewed in HMDs, for low vision users (e.g. scotoma and tunnel vision). This provides semantic meaning in the scene e.g. assigning different colours to objects (road, footpath, sky or cars). Other features include making an object flash to alert the viewer, perhaps to signal 'danger'.

Basic edge detection enhances features of the scene but provides little information regarding the relevance of the object, furthermore, undesirable noise can be introduced. We are motivated by Everingham et al.'s application of semantic meaning to objects, and therefore employ edge detection in the maritime scene, segmenting the image so areas such as sea or sky can be meaningfully identified; and removing sea clutter and lesser useful textures (wave, clouds), so that objects can be detected more easily. Many more AI and computer vision algorithms are used in video analysis for maritime object detection in low observable scenes, refer to a survey by Moriera et al. [17].

V. User-Experience Study

We create a user experiment to investigate the impact of different FOVs on peripheral and central vision (based on Everingham et al. [15] and Peli et al. [8][9]); and examine different image presentations to understand which image features require attention to increase human performance.

A. *FOV Simulator Implementation*

We implement a FOV simulator by generating a series of 360° panoramic 8K resolution SAR test images allowing high visual acuity and providing a realistic simulation environment. The images are developed using the Unity[3] game engine. Fig. 2 shows sections of SAR imagery: (a) original maritime imagery showing environmental conditions (e.g. rough water) with an object located on the horizon; (b) edge detection (Canny filter) (features more visible); and (c) inverse edge detection (image segmentation of objects in scene (sea vs. sky), low contrast, and clutter and noise reduced).

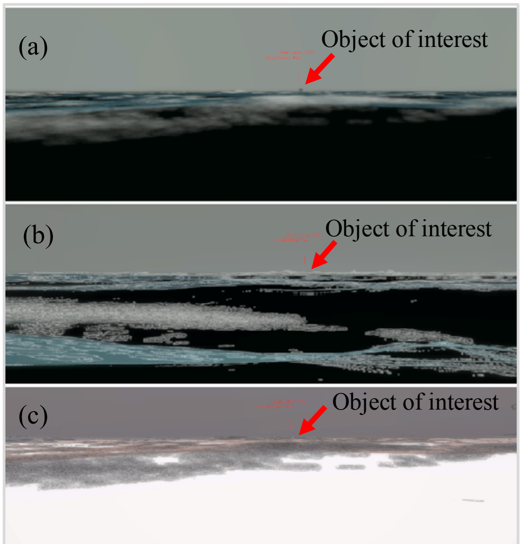


Figure 2. Three visual representations: (a) original; (b) edge detection (Canny filter) and (c) inverse edge detection.

[2] HTC Vive, 2018. Available: https://www.vive.com/au

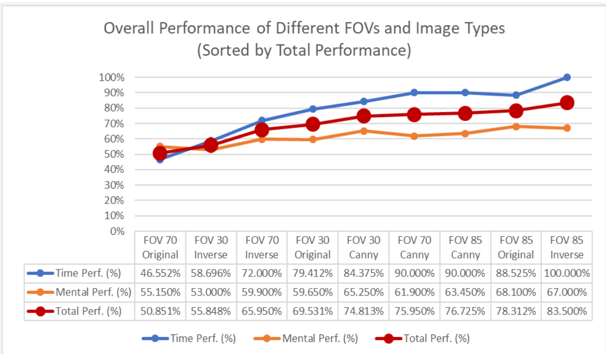
[3] Unity, 2018. Available: https://unity3d.com

Three different FOVs (30°, 70°, and 85°) are tested to understand the effects of the central and peripheral vision on the operator. Objects (4-5 pixels) are at a constant distance (e.g. 1km) in the scene with random bearings. Participants have 15s to detect the objects while standing and wearing an HMD. The HMD used is the HTC Vive Pro, allowing good visual acuity (13PPD).

B. *User Experiment Design*

In each experiment we measure i) object detection speed and accuracy; ii) mental effort (NASA-TLX); and iii) scene memory, over different FOVs with three different experimental scene representations (original, edge detection, inverse). To gain an initial understanding of HVS constraints and develop a user-testing methodology, we tested only five subjects (2 female/3 male aged between 20-50, mean age 37, possessing good visual acuity) to gain feedback on best scene representation techniques. We evaluate three FOVs, over three different image types, each over five random object bearings.

Mental effort is measured using the *NASA Task Load Index (NASA-TLX)* [18][19], on a scale of 1-10, with six demand components (mental, physical, temporal, performance, effort, and frustration). Many participants overestimated the self-perceived performance component of mental effort. To reduce this bias, we remove the performance component in the evaluation, and weight it lower in the overall evaluation. Firstly, we compare initial NASA-TLX results without adjusting weighting components (see section 'C'). For more meaningful average performance calculations, we weight performance and physical demand at 5% (compared to an original weighting of 22.5%) as these components reflect many participants' self-perceived high performance and low physical demand. Therefore, the NASA-TLX and time performance statistics are normalised where higher percentage values indicate higher or better performance, and overall performance for different FOVs and image types are comparable. Figs. 3 and 4 show average performances vs. FOV/image categories. The normalised NASA-TLX shows mental performance is inverse to mental effort.



	FOV 70 Original	FOV 30 Inverse	FOV 70 Inverse	FOV 30 Original	FOV 30 Canny	FOV 70 Canny	FOV 85 Canny	FOV 85 Original	FOV 85 Inverse
Time Perf. (%)	46.552%	58.696%	72.000%	79.412%	84.375%	90.000%	90.000%	88.525%	100.000%
Mental Perf. (%)	55.150%	53.000%	59.900%	59.650%	65.250%	61.900%	63.450%	68.100%	67.000%
Total Perf. (%)	50.851%	55.848%	65.950%	69.531%	74.813%	75.950%	76.725%	78.312%	83.500%

Figure 3. Avg. performances vs. FOV/image categories.

C. *Perceived Mental Effort (NASA-TLX)*

While comparing different image presentations between different FOVs (see Fig. 5), we conclude that the 85° FOV requires the least mental effort. Effort, frustration, physical and temporal demand are overall consistently lower than that observed in 30° FOV for all image types. We observe that different image types have a high variation in mental effort metrics, where image type is significant for a small 30° FOV, whereas lesser effect on mental effort is observed for wider FOVs.

The 70° FOV, sitting between 30° and 85° FOVs presents interesting insights e.g. alternative image presentations are less frustrating, and temporal demand is lower in comparison to the original imagery. Physical demand is similar to that of the 30° FOV, as participants are required to move more to identify the object.

We also compare mental effort for image types per FOV (see Fig. 5). Participants report the same perceived performance over different FOVs, while other factors vary. Lowest frustration levels, effort, temporal, physical, and mental demand are reported for 85° FOV. Edge detected images provide consistent metrics over the three FOV variations, and appear to outperform other image presentations (original or inverse). For inverse images, participants reported lower performance compared with the other two image types, whereas other metrics are equal or lower. For each image type, 85° FOV outperforms perceived mental effort in contrast to smaller FOVs. Interestingly the inverse image representations enable equal or lower frustration, mental demand, effort, temporal, and physical demand. 85° FOV inverse, appears the most effective viewpoint despite lower levels of reported perceived performance.

D. *Comparison of Total Time and Mental Performance*

Figs. 3 and 4 illustrate overall average performance of the study. Alternative image representations increase object detection time performance, thus decrease object detection times. Low mental performance (decreased mental effort) increase object detection times, while larger FOVs increase mental performance. Overall, original image representations only outperform other representations with FOVs>85°. Alternative image representations outperform original images when considering task completion times.

Considering overall average performance (see Fig. 4), edge detection outperforms original and inverse representations in terms of mental performance, but alternative representations clearly outperform original image representations in terms of time performance. FOVs larger than 85° increase both, time and mental performance, but smaller FOVs decrease both. In other words, higher mental effort leads to shorter task times.

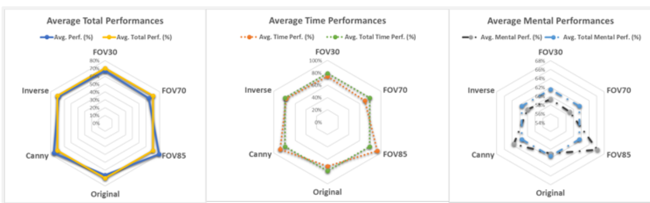


Figure 4. Performance Comparison.

E. *General Observations*

85° FOV inverse appears the best image type and FOV in terms of required mental effort. The inverse representation has low contrast, less clutter and destruction in the scene. A FOV of 85° allows the user to explore a larger part of the scene at once through eye-gaze, without physical movement. Thus, head rotation can be performed in intervals, instead of continuously as e.g. in the case with smaller FOVs. This indicates that peripheral vision is used to scan the scene, and then high acuity central vision isolates the object. The larger the FOV, the more the peripheral vision can be employed to assist object detection. Smaller FOVs constantly employ central vision, and peripheral vision is of little assistance.

Inverse image representations are lower contrast, thus less distracting, with fewer dominant visual objects in the foreground. This means that the user can focus on the horizon line, where the object is likely to be. Less clutter implies lower mental load, as the clutter in the scene is removed, similar to the effect of background subtraction e.g. used in motion detection. Edge detected image representations increase clutter in the scene, and higher contrasts, thus increasing the mental load. Original image representations suffer, as there is much confusion around waves, undefined objects, and reduce focus.

It is also important to note that subjects utilised different search strategies – some used continuous scanning, others rotated their heads and scanned the whole scene, then rotated their head again by the angle of the FOV to scan the newly presented scene. Nevertheless, higher mental effort leads to better performances in form of shorter task times.

Whilst subjects may believe original image presentations enable them to perform better while requiring lower mental effort, our results clearly show the opposite.

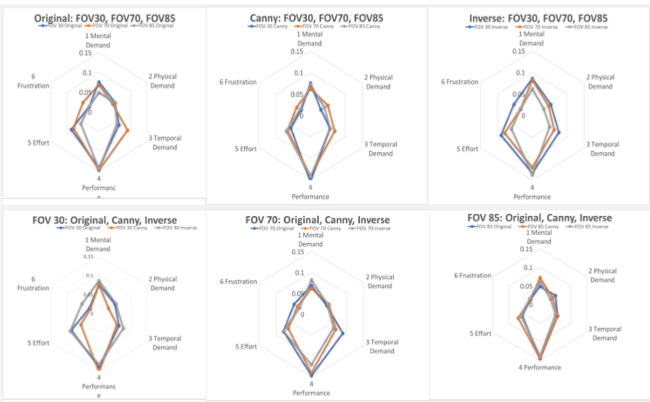


Figure 5. FOVs vs. Image Types.

VI. Conclusion and Future Work

We investigate the limitations of the HVS for maritime SAR imagery in HMDs and present some early results evaluating different FOV and image representations. In future work we will focus on developing advanced AI image processing algorithms for our existing software simulation and extend the user-study towards a larger audience. We are developing a video-based prototype and will examine the effect of motion in imagery as we extend our experiments from images to videos. We will employ new test methods including eye-tracking to understand scanning patterns and saccades, and utilise biofeedback devices [20].

Acknowledgment

We thank Richard Palmer, Ryan Day and Jesse Helliwell for programming assistance.

References

[1] R. Palmer and T. Tan, "Comparision of Simulated Submarine Target Detection Performance Across Three Display Modes." Curtin University, 2018.

[2] A. Lugmayr, B. Stockleben, C. Scheib, and M. Mailaparampil, "Cognitive Big Data. Survey and Review on Big Data Research and its Implications: What is Really 'New'? Cognitive Big Data!," *Journal of Knowledge Management (JMM)*, vol. 21, no. 1, 2017.

[3] P. K. Kaiser, "Prospective Evaluation of Visual Acuity Assessment: A Comparison of Snellen Versus ETDRS Charts in Clinical Practice (An AOS Thesis)," no. 107, pp. 311–324, 2009.

[4] E. B. Werner and C. G. Rossi, *Manual of Visual Fields*. Churchill Livingstone, 1991.

[5] H. E. Bedell and C. A. Johnson, "The Effect of Flicker on Foveal and Peripheral Thresholds for Oscillatory Motion," *Vision Research*, vol. 35, no. 15, pp. 2179–2189, 1995.

[6] C. J H Ludwig, J. Rhys Davies, and M. P Eckstein, "Foveal analysis and peripheral selection during active visual sampling," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, 2014.

[7] E. Akbas and M. P. Eckstein, "Object detection through search with a foveated visual system," *PLOS Computational Biology*, vol. 13, no. 10, pp. 1–28, 2017.

[8] E. Peli, R. B Goldstein, G. Young, C. Trempe, and S. M Buzney, "Image enhancement for the visually impaired: Simulations and experimental results," *Investigative ophthalmology & visual science*, vol. 32, pp. 2337–50, 1991.

[9] E. Peli, G. Luo, A. Bowers, and N. Rensing, "Applications of Augmented Vision Head-Mounted Systems in Vision Rehabilitation," *Journal of the Society for Information Display*, vol. 15, no. 12, pp. 1037-1045, 2007.

[10] A. Sowndararajan, R. Wang, and D. A. Bowman, "Quantifying the benefits of immersion for procedural training," in *Proceedings of the 2008 Workshop on Immersvie Projection Technologies/Emerging Display Technologies*, 2008.

[11] D. Mestre, "CAVE vs. Head-Mounted-Displays: Ongoing Thoughts," in *Proceedings of the IS&T International Symposium on Electronic Imaging, The Engineering Reality of Virtual Reality*, 2017.

[12] R. Patterson, M. D. Winterbottom, and B. J. Pierce, "Perceptual Issues in the Use of Head-Mounted Visual Displays," *Human Factors*, vol. 48, no. 3, pp. 555–573, 2006.

[13] H. A. Angel, M. Gaughan, V. P. G. S, and S. Boyne, "Examination of the effect of field of view on urban target detection." DRCC Toronto, 2005.

[14] E. D. Ragan, D. A. Bowman, R. Kopper, C. Stinson, S. Scerbo, and R. P. McMahan, "Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task," *IEEE Transactions on Visualization & Computer Graphics*, vol. 21, no. 7, pp. 794–807, Jul. 2015.

[15] M. R Everingham, B. T. Thomas, and T. Troscianko, "Head-Mounted Mobility Aid for Low Vision Using Scene Classification Techniques," *International Journal of Virtual Reality*, vol. 3, no. 4, pp. 1–10, 1999.

[16] E. Peli, "Augmented Vision for Central Scotoma and Peripheral Field Loss," in *International Conference on Low Vision*, 1999.

[17] R. Moreira, N. Ebecken, A. S. Alves, F. Livernet, and A. Campillo-Navetti, "A survey on video detection and tracking of maritime vessels," *International Journal of Research and Reviews in Applied Sciences*, vol. 20, no. 1, 2014.

[18] "NASA Task Load Index (NASA-TLX)." NASA. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000214 88.pdf

[19] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advanced Psychology*, vol. 52, 1988, pp. 139–183.

[20] A. Lugmayr and S. Bender, "Free UX Testing Tool: The LudoVico UX Machine for Physiological Sensor Data Recording, Analysis, and Visualization for User Experience Design Experiments," in *Proceedings of the SEACHI 2016 on Smart Cities for Better Living with HCI and UX*, 2016, pp. 36–41. http://doi.acm.org/10.1145/2898365.2899801