# Scanpath Analysis of Fused Multi-Sensor Images with Luminance Change: A Pilot Study

T.D. Dixon, J. Li, J.M. Noyes, T. Troscianko
Department of Experimental Psychology
University of Bristol
Bristol, UK
Timothy.Dixon@bristol.ac.uk

S.G. Nikolov, J. Lewis, E.F. Canga, D.R. Bull,
C.N. Canagarajah
Centre for Communications Research
University of Bristol
Bristol, UK
Stavri.Nikolov@bristol.ac.uk

*Abstract* - **Image fusion is the process of combining images of differing modalities, such as visible and infrared (IR) images. Significant work has recently been carried out comparing methods of fused image assessment, with findings strongly suggesting that a task-centred approach would be beneficial to the assessment process. The current paper reports a pilot study analysing eye movements of participants involved in four tasks. The first and second tasks involved tracking a human figure wearing camouflage clothing walking through thick undergrowth at light and dark luminance levels, whilst the third and fourth task required tracking an individual in a crowd, again at two luminance levels. Participants were shown the original visible and IR images individually, pixel-averaged, contrast pyramid, and dual-tree complex wavelet fused video sequences. They viewed each display and sequence three times to compare inter-subject scanpath variability. This paper describes the initial analysis of the eye-tracking data gathered from the pilot study. These were also compared with computational metric assessment of the image sequences.**

**Keywords:** Image Fusion, Video Fusion, Scanpath Analysis, Video Assessment, Eye-Tracking, Psychophysics

# 1 Introduction

The current paper presents a novel use of an eye-tracking paradigm to analyse participants' scanpaths across a range of fused video displays. Recent literature searches suggest that there has been to date no use of such a paradigm to assess appropriate fusion methods in any previous research. The current section introduces background research in this area, whilst Section 2 reports the current experimental method. Section 3 considers the results obtained, which are discussed in the final section.

## 1.1 Image and Video Fusion

Image fusion is the process of combining multiple images of varying modalities (e.g. Infrared [IR] and visible light radiation) to attain a composite that has the most 'useful' information for a given task. Whilst much academic research has recently focused on the methods for fusing static images, and for the assessment of such images, little work has been carried out with regard to video fusion. Loza and colleagues [1] reported that research into fused video computational metrics is exceptionally scarce. However, some larger commercial companies in the field do have their own in-house dedicated video fusion methods, although these are generally not accessible to the wider academic world.

Two methods of static image fusion that have recently been of interest are the Discrete Wavelet Transform (DWT) and the Dual-Tree Complex Wavelet Transform (DT-CWT). The shift-variant DWT method [2] is widely used, and is the most basic of the wavelet transform methods. These methods involve transforming the input images into the wavelet domain, with the wavelet coefficients processed and combined based on some fusion rule, and the inverse transform being carried out.

The DT-CWT [3] method is an alternative form of a DWT. This method has greater directional selectivity than the DWT, and unlike the DWT method is shift invariant with reduced over completeness. Thus, unlike the DWT, the DT-CWT can directionally select from positive as well as negative orientations, giving six sampling sub-bands at ±15°, ±45°, ±75°. Additionally, small shifts in the input images do not cause such great distortions in the energy distributions of the output wavelet coefficients. These advantages come at the cost of greater computational expense. This fusion scheme has been shown to produce better results than other DWT methods [4, 5], as well as other pyramid and averaging methods [6, 7], across a range of qualitative and quantitative assessments. In the current paper, a simple averaging scheme (AVE) was also used, for reference. This method is computationally very inexpensive, and simply involves averaging between pixel values of the input images.

All of these fusion methods can be applied quite simply to videos. One process involves taking each frame individually from a registered sequence of IR-visible video, and fusing each colour plane of the visible camera separately with the IR sequence. This can then provide a basic, colour-fused output. For the purposes of the current paper, this is the method that has been adopted.

## 1.2 Fused Image Assessment

Current work has begun to look at objective quantitative ways of human image assessment. Initiated by Toet and colleagues [8, 9], definite advances have been made in applying some form of task to the assessment process, and moving away from the ever-present subjective quality assessment. Furthermore, in recent findings [5-7], it has been shown that objective task results can differ significantly from subjective ratings. It is thus essential to choose a good task when assessing fused images or video sequences, and to go beyond simply applying a subjective rating to the fused outputs.

Objective task data can also be compared with the other ubiquitous method of fused image assessment: computational image quality metrics. The problem often found with basic metrics is that in the past they have been found not to correlate well with subjective ratings [1, 10]. As has been shown in work carried by Dixon and colleagues [5-7], metric comparison with objective human task results might overcome this explanatory gap. More recent metrics based on aspects of the human visual system might be able to overcome some of these shortcomings [1]. Two such metrics that have been applied to fused image analysis are Petrovic and Xydeas' metric [11], and Piella's Image Fusion Quality Index (IFIQ) [12].

### *1.2.1 Petrovic and Xydeas Metric*

Petrovic and Xydeas [11] proposed a metric that measures the amount of edge information 'transferred' from the source image to the fused image, to give an estimation of the performance of the fusion algorithm. A Sobel edge operator calculates strength and orientation information of each pixel in the input and output images. These measurements are then used to estimate edge strength and orientation preservation values reflecting the perceptual importance of the corresponding edge elements within the input images. These maps are used to weight the estimates of the edge information, which gives the normalised summation of the performance metric ($Q_P^{AB/F}$). Note that in this method the visual information is associated with the edge information whilst the region information is ignored.

$$Q_P^{AB/F}=\frac{\sum_{n=1}^{N}\sum_{m=1}^{M}Q^{AF}(n,m)w_A(n,m)+Q^{BF}(n,m)w_B(n,m)}{\sum_{n=1}^{N}\sum_{m=1}^{M}w_A(n,m)+w_B(n,m)} \quad (1)$$

### *1.2.2 Piella Metric*

This image fusion quality index (IFQI) measures three different aspects: correlation, luminance distortion and contrast distortion. In order to apply this metric for image fusion evaluation, Piella and Heijmans [12] introduce salient information to reflect the relative importance of image A compared to image B within the window w.

$$Q_w(A,B,F)=\sum_{w\in W}c(w)(\lambda(w)Q(A,F\mid w)+(1-\lambda(w))Q(B,F\mid w)) \quad (2)$$

Finally, to account for the relevance of edge information, the same measure is computed with the 'edge images' instead of the (grey-scale) images A, B and F. As with the previous metric, this metric does not require a ground-truth or reference image.

## 1.3 The Eye-Tracking Paradigm

One alternative method of attaining data related to visual input is to record scanpaths with the use of eye-tracking technology. A broad range of research into scene perception analysing eye movements, and in particular fixations of the eye, has been carried out (e.g. [13]). Furthermore, whilst such research has focused on static image interpretation, psychologists are also experimenting with dynamic computer scenes and actual live action eye tracking [14].

When under ongoing cognitive and perceptual load, eye movements tend towards larger saccades, with critical questions revolving around where and how long an individual fixates a scene or scene-element [13]. However, a range of other eye movements can also occur under varying circumstances, including smooth pursuit, slow drift and stabilisation reflex [15]. The kinds of eye movements that are elicited by a particular task can thus reveal information about the underlying cognitive processes in action.

Investigation into eye movements has considered viewing strategies for people studying complex natural and computer-generated scenes. Individuals have been found to be able to grasp the 'gist' of a natural scene very quickly: within 100ms [16]. Other research has shown that more successful players of the game Tetris showed better-maintained two-fixation scanpaths with increased cognitive load over a period of time [17].

Studies considering smooth pursuit eye movements, that is, those steady movements associated with slow and even tracking, have also found significant variation. Wallace et al. [18] provided evidence for a computational model of smooth pursuit initiation based on direction and velocity of a tracked target, combined with the target's edge and feature information. It has also been found that the application of a secondary task whilst carrying out smooth pursuit tracking of a target can significantly degrade the performance of the pursuit [19]. Given the broad range of findings, it seems appropriate to apply this knowledge to the area of fused image assessment.

## 1.4 Our Approach

The current paper focuses on the combination of analysing gaze fixation data with the use of secondary tasks in two tracking scenarios. Whilst our previous work has been carried out on static images (e.g. [5-7]), the use of video stimuli also allows for the creation of more realistic scenarios, in which eye-tracking efficiency and accuracy may be of paramount importance.

# 2 Method

The current experimental design constitutes part of a larger study yet to be published.

## 2.1 Design

The current experiment manipulated display method across five levels: visible (Viz) light video, IR video, AVE fused, shift-variant DWT fused, and DT-CWT fused videos. In Session 1 half the participants viewed the videos in the order given, whilst half viewed in the reverse order. In the second session all participants viewed the sequences in the opposite order to that which they viewed in Session 1, and in Session 3 they viewed in the original order. This was done to help counterbalance for ordering effects.

The two tasks carried out by the participants were based around the video sequences 'Tropical 2.1_i' and 'Tropical 2.1_iii', captured in a larger data-gathering project, as described in [20] and Section 2.3. A pair of scenarios was shown to the participants with similar content, but with one of the pair having much greater atmospheric luminance than the other. As stated, this order was reversed in the second session.

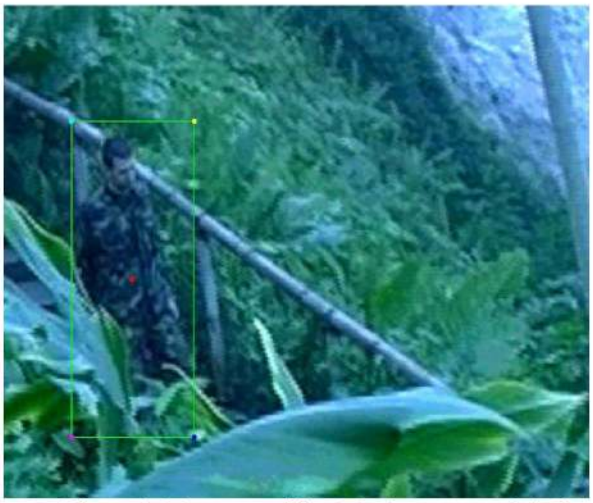


Figure 1: Target-marking Tool, sequence 2.1_i.

Data collected included the raw gaze fixation data on the screen, as well as reaction accuracy from event data recorded from key presses whilst carrying out the tasks set. The eye fixation data was compared with pre-drawn 'target maps'. These were rectangular target boxes drawn around the soldier (target to be tracked), that were created using a toolbox which can be used to delineate rectangles throughout a sequence (see Figure 1). Targets were drawn at least every 15 frames where possible; when the tracking target was not visible for longer periods on the screen, estimations were made. Once the targets were drawn, the between-frame targets were calculated by interpolation and then visually inspected. Where necessary, these were individually readjusted. Thus, ground truth tracking data was created for the video sequences and used subsequently to evaluate human tracking performance.

## 2.2 Participants

Ten participants (5 females, 5 males) took part in the current study in exchange for monetary compensation,. Eight were naïve to the concepts and videos utilised. Ages ranged from 21 to 41 years (mean = 27.1, s.d. = 6.76). Participants were required to have normal or corrected-to-normal vision, and none had any history of colour vision problems.

## 2.3 Apparatus and Stimuli

A Tobii™ x50 remote eye tracker [21] was used to collect eye movement data. This is a table-mounted eye tracker that works at 50 Hz with an approximate accuracy of 0.5°. This was run using the ClearView 2.5.1 software package, on a 2.8 GHz Pentium IV dual processor PC, with 3 GB RAM, and twin SCSI hard drives. Stimuli were presented on a 19" flat screen CRT monitor running at 85 Hz, with screen resolution set to 800 by 600 pixels. Participants were required to use a chin-rest positioned 57cm from the monitor screen.

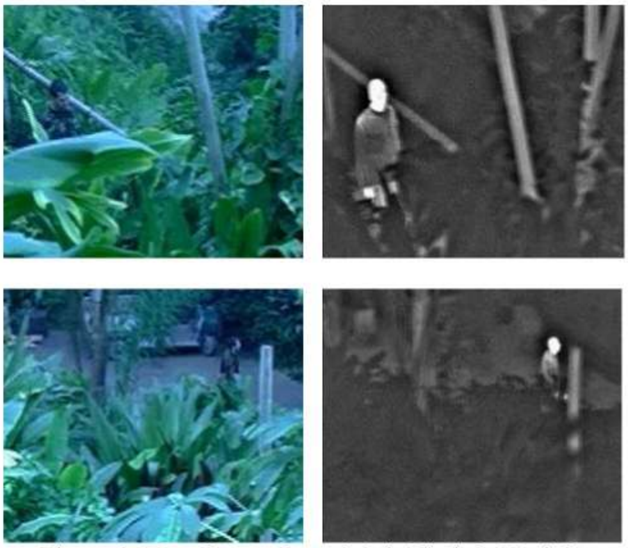


Figure 2: Two frames from 2.1_i, Viz left, IR right.

The two video sequences shown were part of a data-gathering project carried out at the Eden Project Biome in Cornwall, UK, and detailed in [20]. This project utilised an array of different sensors across two mornings and evenings filming a variety of scenarios. The two selected for the current paper were subclips from the 'Tropical Forest' collection sequences 2.1_i (High Luminance: HL) and 2.1_iii (Low Luminance: LL). These sequences both showed a 'soldier' (actor) dressed in camouflaged clothing walking down a pathway amongst foliage, through a clearing of trees and back across the way he came, as shown in Figures 2 and 3. The sequences were fused using the AVE, DWT, and CWT methods detailed in Section 1.1, yielding fused sequences as shown in Figure 4.

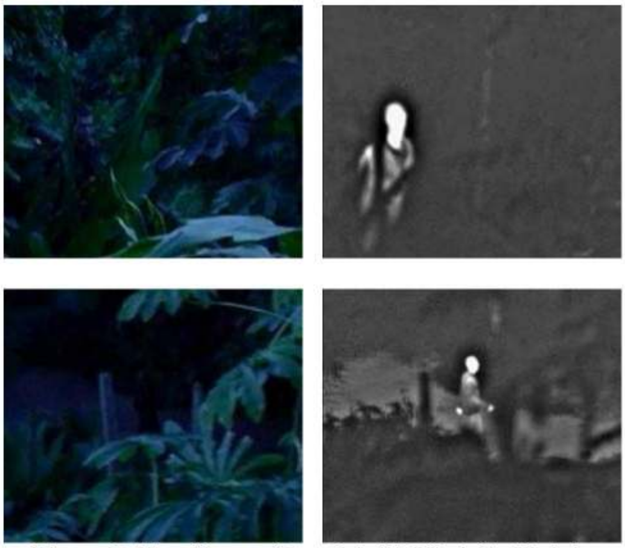
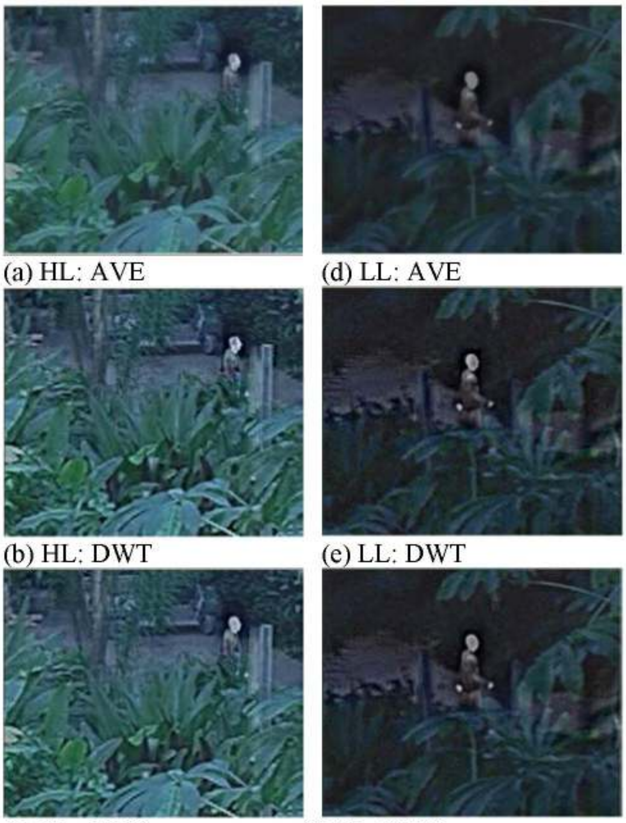


Figure 3: Two frames from 2.1_iii, Viz left, IR right.

The sequences (frame size = 576 by 480 pixels) were displayed in the centre of the screen at 25 frames per second, with each individual sequence lasting 1711 frames (68 seconds). Each video file was compressed using the Microsoft Video 1 codec at best quality setting. All compression was applied after fusion taking place.



(a) HL: AVE (d) LL: AVE

(b) HL: DWT (e) LL: DWT

(c) HL: CWT (f) LL: CWT

Figure 4: Frames from HL and LL.

## 2.4 Procedure

Participants were asked to attend three sessions with each session consisting of the same experimental conditions. In Session 1, the first task (sequence HL) involved visually tracking a soldier walk down a path, as well as signalling by pressing the space bar when the soldier began and finished walking past a small electric vehicle present in the clearing, that is, one press for each incident (see Figure 4).

The second task (sequence LL) also involved tracking the soldier down the path. This time participants were asked to signal when the soldier was at the central point between two posts. Each sequence was shown five times, once in each different display, with the task remaining the same each time. Sixteen point ClearView calibration was carried out before each set of five films with calibration background luminance darker for LL to match the lower luminance in that sequence. In subsequent sessions the tasks and display orders were altered as described in Section 2.1.

## 2.5 Data Analysis

Raw eye fixation data were taken and compiled so that they could be compared with the target boxes previously created (see Figure 1). This is shown in Figure 5. Once it was known in which frame each recorded gaze point was located, a direct evaluation could be made with the target overlays. For each display modality in each task an accuracy ratio was calculated by dividing the number of gaze points correctly located inside the target map by the total number of gaze fixations recorded.

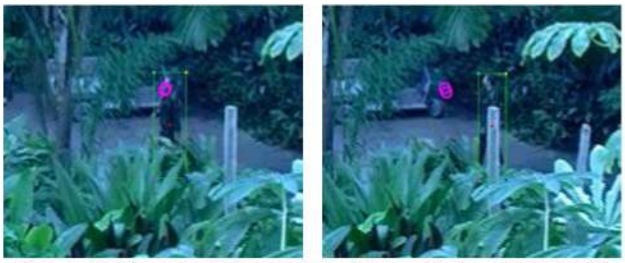


Figure 5: Gaze location comparison with target map. On the left is a hit, on the right a miss.

The ClearView program also supplied data on how valid each raw fixation was for each eye. This ranged from ‘0’ (definitely certain that a particular fixation belonged to a particular eye) to ‘3’ (very uncertain that a gaze point corresponds to an eye), with ‘4’ meaning that no eye was detected. In the current study, only fixation data with a validity of ‘0’ for both eyes was used. The eye fixation points of the two eyes were then averaged for every recorded pair of gaze fixations. This provided the most valid data, averaged to accommodate any variance caused by ‘drifting’ artefacts, which are usually inversely symmetrical.

The data for the task key presses was also matched to the number of frames in each sequence. Ground truth ‘correct’ frames were decided upon, which were then compared with the timing of the key presses in each sequence recorded. Timings using the ClearView timestamps were compared with the ground truth, with negative numbers entailing a key press before the chosen point, and positive numbers after.

It is planned that the current data will be analysed more extensively in the future, accounting for variations in pupil size, which can correlate with cognitive load.

# 3 Results

Initial analysis of the results revealed that one participant's gaze accuracy scores were significantly below the others, with scores in LL not reaching much above 0 in most conditions and across sessions. It was decided that the eye-tracking system had not been able to track this individual, verified by some poor calibration results. This is probably caused by the presence of a slight squint, and this participant's scores were not used in the rest of the analyses. As the two counterbalanced groups were now of differing sizes, initial tests were carried out to see if the unequal group sizes had a significant effect on the results: these were not significant so the groups remained as they were.

## 3.1 Scanpath Data Analysis

The HL eye location accuracy scores showed a general pattern of results indicating Viz and IR scored lowest, whilst AVE and DWT scored highest, as shown in Figure 6. These were analysed using a two-way repeated measures Analysis of Variance (ANOVA), with display modality as one factor and session as the second. This revealed a significant main effect of display (F(4, 32) = 6.50, p = 0.001), but not of session (F(2, 16) = 2.00, p > 0.05), and no interaction (F(8, 64) = 1.23, p>0.05). Post hoc testing using Tukey's Honestly Significant Difference (HSD) method revealed significant differences between Viz and all other display methods except IR (HSD = 0.0488, p = 0.05). In addition, DWT was significantly greater than IR, as well as AVE approaching significant difference, but interestingly, CWT was not significantly different to the IR condition. These results indicate that Viz and IR are performing worse than the three fusion schemes, with DWT performing slightly better than AVE in comparison to Viz and IR especially.

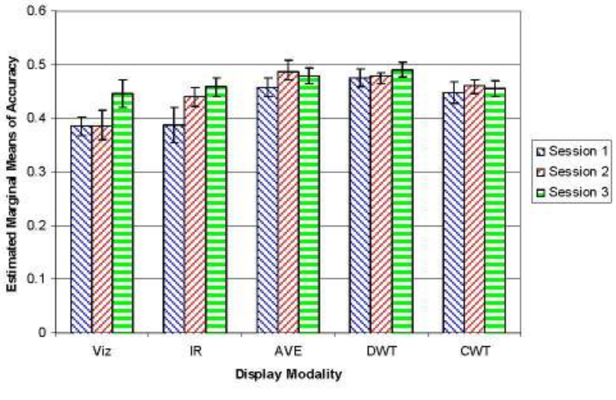


Figure 6: Means of High Luminance Accuracy.

Scores for LL (Figure 7) showed Viz to be much lower than the other methods, with AVE and DWT again apparently leading to best accuracy. ANOVA testing revealed a main effect of display method (F(4, 32) = 12.7, p < 0.001), but no main effect of session (F(2, 16) = 0.406, p>0.05) or an interaction (F(8, 64) = 0.936, p>0.05). Tukey HSD testing revealed significant differences between Viz and all of the other modalities individually, but no further differences between the modalities themselves (HSD = 0.0439, p = 0.01). This indicates that the Viz modality has performed much more poorly in the LL task, but the other methods perform equally well. It should be noted that the eye tracker could not scan one participant's left eye for Session 2, so right eye scores alone were used.

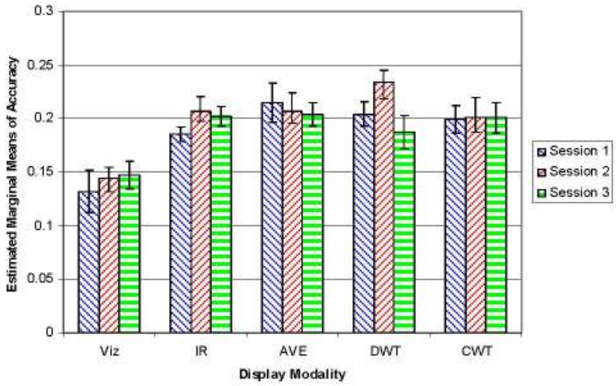


Figure 7: Means of Low Luminance Accuracy.

## 3.2 Task Results

The pattern of the HL task results revealed there was a trend towards signalling the beginning and end of target movement before the ideal timing, as shown by the prevalence of negative numbers in Figure 8. ANOVA testing revealed no significant main effects of modality (F(4, 32) = 1.28, p > 0.05), or session (F(2, 16) = 0.645, p>0.05), or an interaction (F(8, 64) = 0.465, p > 0.05). This shows no statistically meaningful difference between the conditions in this task. This is most probably due to one participant failing to complete the task in the IR and CWT conditions of Session 1, leading to much higher standard error than the other conditions (as shown in Figure 8).

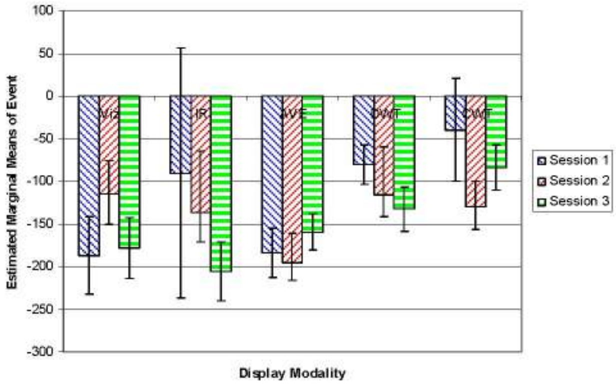


Figure 8: Means of High Luminance Reaction.

The task results for LL (Figure 9) indicated another trend towards reacting before the ideal time, except in Session 1 IR reactions. The repeated measures ANOVA revealed a significant main effect of display method (F(4, 32) = 3.78, p = 0.013), but not of session (F(2, 16) = 0.853, p > 0.05), or an interaction (F (8, 64) = 1.44, p > 0.05). Tukey testing revealed significant differences between Viz and IR (HSD = 147, p = 0.01), and between IR and DWT (HSD =117, p = 0.05). This indicates that there was the smallest latency in reaction time accuracy for IR, and the largest for Viz, with most people pressing the button much earlier than required in this condition, and somewhat more in the DWT condition.

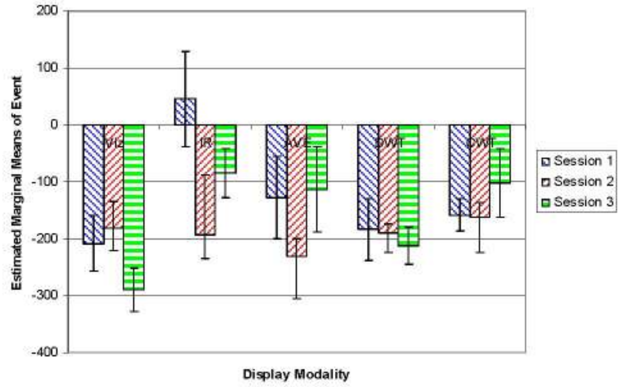


Figure 9: Means of Low Luminance Reaction.

## 3.3 Individual Comparison of Session

Due to the large variation in individual accuracy scores across sessions, it was decided to carry out separate analyses of participants' scores, collapsing across display modality for each participant. One-factor repeated measures ANOVAs of HL results revealed significant increase in accuracy across sessions for participant four (F(2, 8) = 16.3, p = 0.002; Figure 10), with two more participants close to significance. Tukey post-hoc testing on participant four's results revealed that the accuracy in Session 1 was significantly lower than Sessions 2 and 3, but no difference between Sessions 2 and 3 themselves (HSD = 0.0571, p = 0.05). This indicates an increase between Sessions 1 and 2, but not between the later sessions.

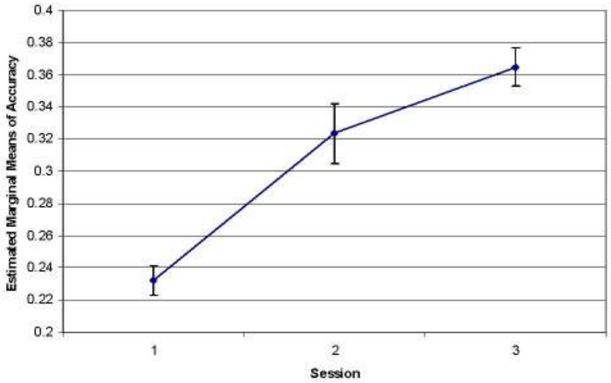


Figure 10: High Luminance Session Scores for Participant 4.

The individual analyses for LL revealed that participant one significantly increased in accuracy across sessions (F(2, 8) = 21.0, p = 0.001; Figure 11), and one other participant was approaching significance. Post-hoc testing also revealed significant differences between Session 1 and the other two, but not between Sessions 2 and 3 (HSD = 0.0571, p = 0.05). This again suggests any increase in accuracy will occur between Sessions 1 and 2.

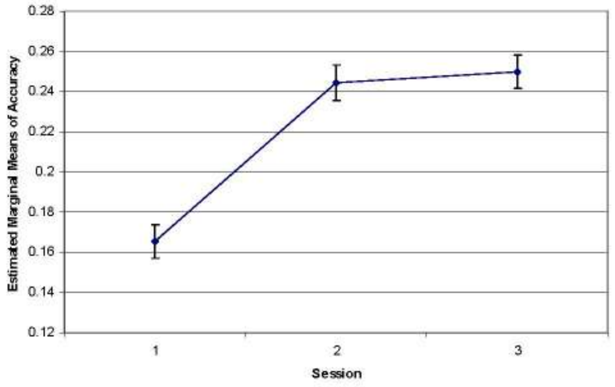


Figure 11: Low Luminance Session Scores for Participant 1.

Despite the two participants' results presented here showing a trend towards increasing accuracy over sessions, it should be noted that some of the results that were close to significance did not show this trend. In HL participant two (F(2, 8) = 6.46, p = 0.060), showed much worse accuracy in Session 2 than Sessions 1 and 3, as did participant eight (F(2, 8) = 3.77, p = 0.070) in LL. This indicates that there was not necessarily a regular pattern of accuracy increase across participants, as indeed is shown in Figures 6 and 7.

## 3.4 Metric Results

The metrics discussed were computed for the two sequences using 35 frames evenly distributed throughout each sequence: roughly every 49 frames. As can be seen in Figure 12, for sequence HL Piella's [16] metric rated the DWT and CWT methods equally highly and AVE much lower, with Petrovic's [14] metric following this trend.

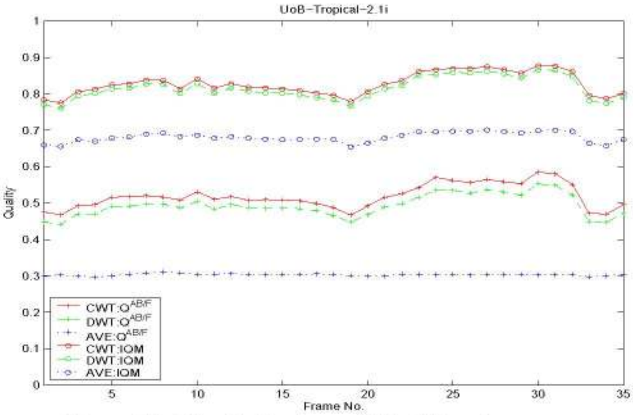


Figure 12: Metric Ratings of High Luminance.

The metric results for the LL sequence follow the trends of HL, as shown in Figure 13. CWT and DWT are again shown to be much better than the AVE fusion method. Interestingly, the AVE fusion method has performed well in the gaze accuracy ratings, but has not yielded such good results with the metric scores.

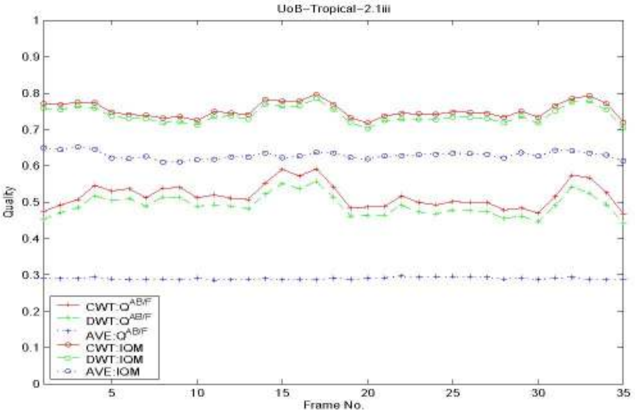


Figure 13: Metric Ratings of Low Luminance.

## 4 Discussion

The current pilot study has produced some very interesting initial findings, utilising a novel approach to the increasingly important problem of fused image assessment. This combined with task and metric results, has created the starting point for a new field of assessment possibilities.

The first point of interest is the finding that the AVE and DWT fusion methods were found to perform best in the HL tracking task. From a subjective point, the DWT appeared to create a sequence that was much noisier than the CWT method. However, it did delineate the edges of the soldier well, possibly allowing for a better match between one dimensional edge information and structural information, as predicted by [18]. Moreover, the computationally inexpensive AVE method has performed well, and might thus provide a 'cheap and dirty' method of quick video fusion when such a fused output is required. This is despite the AVE method reducing contrast in the fused image output. Critically, all of the fusion methods performed significantly better than the inputs (with the exception of CWT over IR), highlighting the advantages of using a fused sequence even when luminance levels are high, and the anticipated additional advantage of having IR information might be low. Thus far, only total hit/miss scores for the whole video have been counted and presented. Splitting the sequence into several parts (which differ in content) may reveal more significant differences between input and fused displays.

Secondly, the improvement seen across the three sessions in the individual participant results provides some evidence to show that people are learning more optimal looking strategies across time. As was found by Underwood [17], scanning strategies can vary, with more successful people (at the game Tetris) retaining certain eye movements, whilst less successful stop using such movements. In the current experiment, participants were able to learn the path of the soldier over explicit retrials of the tracking task. A more detailed analysis of the scanpath behaviour of the participants than is available here might yield a similar retention of certain strategies in the more accurate participants. However, as stated in Section 3.3, this trend only held for a few of the participants over the sessions, and therefore little theoretical claim can be made.

The lack of significant result for the HL sequence task might indicate several things. First, as the luminance was high in this sequence, the task was only a certain amount harder in the IR condition, where the (cold) vehicle did not show up as well. Thus participants may have found the experiment simply too easy, although the significant trend towards responding pre-emptively does seem to indicate that there is more occurring here. One other possibility is that the presence of the second task is affecting participant performance to some degree. In effect, this is the reverse of [19], which found that the addition of a secondary task to smooth pursuit tracking would affect the *tracking*. However, it is hard to make any stronger claims without a more exact task involved.

The results for LL suggest that when luminance is low, any method of attaining additional information regarding the target location will significantly improve upon a visible light camera alone. When combined with the task results, it becomes clear that the visible light display is seriously deprived in the amount of information it can provide to participants. Interestingly, the IR modality allowed most accurate button presses in the task, which is somewhat surprising as the IR image did not on its own supply much information about the location of the two posts, which were much more clearly visible in Viz. It might thus be that the IR display left participants in a state of uncertainty, which cancelled out to some degree the pre-emptive reactions seen in the other conditions.

The metric results presented a somewhat different picture to that shown in the tracking and task accuracy findings. CWT and DWT outperform AVE in both sequences, results which were not found in the tracking data. What is more interesting is what results would have been found if some subjective ratings task had been carried out. From the perceived quality of the three fusion methods, it could easily be posited that the DWT method would score lower than the AVE method in a subjective quality assessment, suggesting that the 'image quality' metrics might not even be modelling 'quality' per se. Importantly, both the metric results found, and the hypothetical subjective rating experiment present a different pattern of results to those yielded by the objective human tasks used.

The final issue of interest raised in the current paper is the general efficacy of an eye tracking image assessment paradigm. It is clear that such an approach can lead to significant results, and as far as pilot studies go, this one can be deemed a success. What is of more importance is how the findings could be improved upon in the future. Whilst the Tobii eye tracking system is easy to use, it lacks the much greater scan rates of some of the head-mounted systems. In addition, the current findings suggest that even though a relatively reliable calibration can be made with this system, this does not necessarily mean that the system will be able to accurately read eye movements in every case. Additionally, whilst the results were stable enough to provide statistical significance, they appear somewhat lower than might be expected. However, this can be accounted for by the large camera movement and the soldier also out being of sight for periods of time leading to incorrect tracking. As the tracking data used was verified against the original videos produced by the ClearView program, we are confident that the scores are an accurate representation of participants' scanpaths. It is anticipated that further studies will be able to corroborate the current findings.

One other issue that may have affected the results adversely (although the presence of statistically significant results suggests not too much so) is the methods of counterbalancing used. In the current study there may have been some ordering effects, caused by the participants only viewing the sequences in two arrangements. Ideally, all conditions of the independent variable should be presented in every order, but this would entail a sample cohort of at least 120 people. However, what is probably of greater significance is whether participants have viewed the inputs first, or the fused images. By strategically counterbalancing the input images on the one hand and the fused images on the other, a suitably sophisticated viewing order can be created.

The current study has shown that the use of an eye tracking paradigm can lead to a new and highly relevant method of assessing the value of fused images. The results obtained suggest that in certain task scenarios, simpler fusion methods can outperform more complex ones. Future work is planned as detailed, including more complex scanpath analysis accounting for scanpath similarity and using attention maps. Further, applying more complex tasks might allow for additional task-based response data to have more impact.

### Acknowledgements

This work has been funded by the UK MOD Data and Information Fusion Defence Technology Center. Scanpath data will be made public at: www.Scanpaths.org, whilst videos used in this study are available at: www.ImageFusion.org.

## References

[1] A. Loza, T. D. Dixon, E. F. Canga, S. G. Nikolov, D. R. Bull, C. N. Canagarajah, J. M. Noyes, and T. Troscianko. Methods of fused image analysis and assessment. In *Proceedings of the Advanced Study Institute Conference (NATO-ASI 2005): Multisensor Data and Information Processing for Rapid and Robust Situation and Threat Assessment*, Bulgaria, May, 2005.

[2] G. Simone, A. Farina, F. Morabito, A. Serpico and L. Burzzone. Image fusion techniques for remote sensing applications. *Information Fusion*, 3:3-15, 2002.

[3] Kingsbury, N. Image processing with complex wavelets. In *Wavelets: The Key to Intermittent Information*, B. Silverman and J. Vassilicos, Eds. Oxford University Press, USA, 165–185, 1999.

[4] S.G. Nikolov, P. Hill, D.B. Bull, and C.N. Canagarajah. Wavelets for image fusion. In A. Petrosian and F. Meyer, editors, *Wavelets in Signal and Image Analysis*, Computational Imaging and Vision Series, 213–244. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.

[5] T.D. Dixon, E.F. Canga, J.M. Noyes, T. Troscianko, D.R. Bull, and C.N. Canagarajah. Assessment of fused images: Objective, subjective and computational methods. *Perception,* 35, 3:422, 2006.

[6] T.D. Dixon, E.F. Canga, J.M. Noyes, T. Troscianko, S.G. Nikolov, D.R. Bull and C.N. Canagarajah. Methods for the assessment of fused images. *ACM Transactions on Applied Perception,* in press, 2006.

[7] E.F. Canga, T.D. Dixon, S.G. Nikolov, C.N. Canagarajah, D.R. Bull, J.M. Noyes and T. Troscianko. Characterisation of Image Fusion Quality Metrics for Surveillance Applications over Bandlimited Channels. *Proceedings of the Eighth International Conference on Information Fusion,* 231 – 321, US, July 2005.

[8] A. Toet, and E.M. Franken. Perceptual evaluation of different image fusion schemes. *Displays* 24, 1:25 – 37, February. 2003.

[9] A. Toet., J. IJspeert, A. Waxman and M. Aguilar. Fusion of visible and thermal imagery improves situational awareness. Displays, 18:85–95, 1997.

[10] A. Eskicioglu. Quality measurement for monochorome compressed images in the past 25 years. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* 1907 – 1910, 2000.

[11] V. Petrovic and C. Xydeas. On the effects of sensor noise in pixel-level image fusion performance. *Proceedings of the Third International Conference on Information Fusion,* 2:14 – 19, July 2000.

[12] G. Piella and H. A Heijmans. A new quality metric for image fusion. *International Conference on Image Processing, ICIP, Barcelona*, 2003.

[13] J. M. Henderson and A. Hollingworth. High-level scene perception. *Annual Review of Psychology,* 50:243-271, 1999.

[14] M. Land and S. Ferneaux. 1997. The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London B,* 352:1231-1259, 1997.

[15] R.H.S. Carpenter. *Movements of the Eye*. Pion, London, 1977.

[16] G. Underwood. Eye fixations on pictures of natural scenes: Getting the gist and identifying the components. In G. Underwood (Ed.) *Cognitive Processes in Eye Guidance*, Oxford University Press, New York, 2005.

[17] J. Underwood. Novice and expert performance with a dynamic control task: Scanpaths during a computer game. In G. Underwood (Ed.) *Cognitive Processes in Eye Guidance*, Oxford University Press, New York, 2005.

[18] J.M. Wallace, L.S. Stone and G.S. Masson. Object motion computation for the initiation of smooth pursuit eye movements in humans. *Journal of Neurophysiology*, 93:2279-2293, 2005.

[19] S.B. Hutton and D. Tegally. The effects of dividing attention on smooth pursuit eye tracking. *Experimental Brain Research,* 163:306-313, 2005.

[20] J. J. Lewis, S. G. Nikolov, A. Loza, E. Fernandez Canga, N. Cvejic, J. Li, A. Cardinali, C. N. Canagarajah, D. R. Bull, T. Riley, D. Hickman and M. I. Smith. The Eden Project multi-sensor data set. Available online at: http://www.ablen.com/hosting/imagefusion/resources/bibliographies/reports/TR-UoB-WS-Eden-Project-Data-Set.pdf, 2006.

[21] Tobii Technology: http://www.tobii.se/