# Image Fusion Evaluation

Computer Vision Group[1*]

[1]Department of Computer Science and Artificial Intelligence, University of Granada, CITIC-UGR, 18071 Granada, Spain

[*]To whom correspondence should be addressed; E-mail: jags@decsai.ugr.es

**Firstly we propose to rank order images fused by different methods according to the computational attention value of their regions of interest. To this aim we compute for each of the fused images a multi-bitrate attention map, following a rational model of computational attention. From this attention map, we then calculate the average attention score within areas of interest for each bitrate while using a prioritization scheme. Here the prioritization protocol is used to simulate the basic cognitive process of visual information adquisition by human users.**

**Secondly we develop a novel axiomatic approach to rank order fused images using the important information visibility. It allows to determine if a particular fused image is more visually efficient than another competitor in a dataset. The visual efficiency is predicted by an image fusion evaluation score that satisfies a basic set of axioms; with the underlying assumption being that if a function satisfies all the desirable constraints it is expected to be more effective in real applications.**

# Introduction

The objective of image fusion is to represent relevant information from multiple individual images in a single image. The range of available image fusion techniques and systems is steadily increasing. Some fusion methods may represent important visual information more distinctively than others, thereby conveying it more efficiently to the human observer. Hence, there is a growing need for metrics to evaluate and compare the visual quality of fused imagery.

There is a large body of work existing now on the topic of objective evaluation of image fusion. A number of objective metrics exist of varying degrees of complexity and a host of different approaches (*3–7*).

In Reference (*1*) we present a novel approach to rank order fused images from a dataset using the important information visibility. It allows to determine if a particular fused image is more visually efficient than another competitor in a dataset. The visual efficiency is predicted by a normalized measure of computational attention within the regions of interest.

The Human Visual System (HVS) appears to employ a serial computational strategy to select locations of interest in the processing of massive amounts of incoming visual information with nearly real-time capacity of reaction (*8, 17, 24*). Thus the detection and analysis of visual objects seem to involve either covert shifts of attention or saccadic eye movements, and the image analysis and scene understanding may be performed by biological visual systems through a temporal serialization into smaller, localized analysis tasks (*12, 14*).

Following (*27*), Koch and Ullman (*17*) develops the idea of saliency map to accomplish preattentive selection by encoding the saliency of objects in the visual environment. Competition among neurons in this map gives rise to a single winning location that corresponds to the most salient object (i.e., the next target). Based on (*17*), Itti et al. (*12*) describes a preattentive selection mechanism based on the architecture of the primate visual system, in which 42 maps encoding intensity, orientation and grey in a center-surround fashion at a number of spatial scales can be combined into a single saliency map. A Winner-take-all algorithm is finally conducted in order to predict the gaze location.

One of the closest computational models to the local processing biological reality within the HVS (*18*) built a psychovisual space before the center-surround difference. Nevertheless recent results in visual attention (*9*) brought confirmations for a global integration of feature information all over the visual field which is possible thanks to the impressive neuronal network (*20*). Following the approach that attention may be due to global properties, Osberger and Maeder (*23*) used segmentation to separate the image into several homogeneous areas and five features were used in assigning a relative importance to the areas. Walker et al. (*28*) suggested that saliency may be related to the probability that a feature will be misclassified with any of the other features within an image or a database, while several others (*21, 22*) stated that object components saliency may be inversely proportional to their occurrence within the image. Otherwise, Itti and Baldi (*13*) proposed a probabilistic approach of surprise based on the Kullback-Leibler divergence between what was expected to happen and the actual observation. Mancas (*20*) introduced a rarity-based three-level attention model handling mono-dimensional signals as well as images or video sequences.

Anyway, Privitera and Stark (*25*) showed that different computational attention models were tuned for some kinds of images and often react very badly to other images, and thus, it should be very difficult to use only an attention model in all the applications. Hence, in (*10*) a set of axioms were proposed to prescribe constraints that seem to us imperative to acknowledge in the problem of allocating attention. The result is a multi-bitrate attention map which provides us with a computational attention score for each spatial location at high and low quality versions of the image reconstruction. The novelty of this map is that: (i) It allows distinct attention score for the same spatial location at different picture quality; (ii) it avoids certain forms of behavioral inconsistency in the absence of a priori knowledge about the locations of interest, which is a characteristic of rational systems; and (iii) a particular integration of feature information (e.g., grey, intensity, orientation) is not used in assigning attention scores to the points, therefore computational attention is not tuned for only certain images. In any case, the critical difference with respect to the approaches discussed above is that we will be able to predict exactly the behavior of the axiomatic solution according to its principles. (*10*) proved that the rational model of computational attention yields a high probability of correct classification with respect to a reference rank order given by the target distinctness measured by human observers.

## Comparative Visual Efficiency of Image Fusion Methods

Our work in Reference (*1*) presents a comparative visual efficiency analysis of input and fused images which are reconstructed at high and low fidelity using a transmission method (without region-dependent quality of encoding) to predict the prioritization of image information at different bitrates, (*15*). A high value of the normalized mean attention within the areas of interest at a reconstruction fidelity $\rho$ means that the prioritization protocol brings the attention onto the important regions to human subjects at this particular time of progressive transmission —corresponding to a bitrate $\rho$ of picture quality. Coull (*16*) showed the mutually beneficial effects of attention and timing, since attention is distributed in time as well as space. Based on rate-attention curves, as given by the normalized mean attention score within the areas of interest across bitrates, here we predict which are the images from a dataset more able to provide all salient information in the source images to the potential human operator.

In Reference (*1*) we present a comparative analysis of the fused images within high saliency regions which are automatically detected based on computational attention. This attention model will predict what regions involuntarily attract attention in an input image, since their visual information saliency. The novelty of this approach to image fusion evaluation is the use of rate-attention curves which are given by the normalized mean attention score within the areas of interest across bitrates while using a prioritization protocol, and where the detection of regions of interest is achieved either interactively through human user intervention or by the use of automated detection.

## Axiomatic Approach to Image Fusion Evaluation

A different approach to image fusion evaluation can be to first state some general principles that the solution of the problem must obey, and then derive the solution that satisfies exactly the principles.

Reference (*2*) proposes a novel axiomatic approach to rank order fused images using the important information visibility. It allows to determine if a particular fused image is more visually efficient than another competitor in a dataset. The visual efficiency is predicted by an image fusion evaluation score that satisfies a basic set of axioms; with the underlying assumption being that if a function satisfies all the desirable constraints it is expected to be more effective in real applications.

To this aim, we discuss the image representational model for fusion evaluation and review the computational attention model. We then present the basic axiomatic characterization. The first axiom states the basic case when both input and fused collections of high-saliency points contain only one location. The second axiom shows the reference change effect when we add a new high-saliency location to the collection of the input image. A third axiom says that adding a high-saliency location to the fused image collection should increase the score if the high-saliency location is in the input image collection which is used as reference in the evaluation of fused images. By the contrary, adding a high-saliency location to the fused image collection should decrease the score if the high-saliency location is not in the input image collection. The fourth axiom says that the impact of adding a high-saliency location to the fused image collection should decrease as we add more and more examples of the same high-saliency location which are detected at finer bitrates.

Reference (*2*) proposes an image fusion evaluation score and proves that it satisfies the basic axiomatic characterization. Several experimental results are also presented to analyze if this evaluation score agrees indeed with human observer performance, making the approach valuable for practical applications.

The critical difference with respect to the other approaches in the Literature is that we will be able to predict exactly the behavior of the axiomatic solution according to its principles. Thus, the axiomatic score function can be used to rank order the fused images according to the information visibility from the input scenes in such a way to verify certain forms of behavioral consistency like the novelty effect, the reference change effect, as well as the positive addition effect among others.

# Experimental Results

We can now perform a comparative visual efficiency analysis of input and fused images using the seven datasets given in Figs. 1–7. Based on the axiomatic score function given in (*2*), here we predict which are the fused images in a dataset more able to provide all salient information from the input images to the potential human operator.

A high value of the axiomatic score function $E(I_n^{HS}, F_m^{HS})$ defined in (*2*) means that the fused image $F_m^{HS}$ brings the attention onto the same high-saliency locations for the input image $I_n^{HS}$. It predicts a faster detection, using a particular fused image, of high-saliency areas from the input (e.g., visual, infrared, both) images; therefore the axiomatic score can be used to rank order the fused images according to the information visibility from the input scenes.

For each fused (CWT, DWT, PYR) image, we have that Fig. 13 displays the respective axiomatic score function across datasets (Set $\#1$ through Set $\#7$) in Figs. 1–7, using as reference

| Best Performance | **Visual** | **Infrared** | **Visual $\cup$ Infrared** |
|---|---|---|---|
| **Set 1** | CWT | CWT | CWT |
| **Set 2** | DWT | PYR | PYR |
| **Set 3** | CWT | CWT | CWT |
| **Set 4** | CWT | PYR | CWT |
| **Set 5** | CWT | PYR | PYR |
| **Set 6** | CWT | PYR | PYR |
| **Set 7** | DWT | PYR | DWT |

Table 1: Best Performance

the visual image (top), the infrared image (center), and both the visual and infrared image (bottom). For each dataset (Set $\#1$ through Set $\#7$) of input and fused images, Tables 1 and 2 summarize, respectively, the best and the worst performance as given by the axiomatic score function with the input image $I_n^{HS}$ being defined as the visual image, the infrared image, and the visual union infrared image.

The CWT fusion scheme yields the best performance in ten different comparisons (see Table 1); while the DWT fused images achieve the best performance in three comparisons and the PYR fused images in eight comparisons.

It appears that high-saliency locations from visual union infrared images are best detected using the CWT fusion scheme (see Table 1). Also it appears that high-saliency points from infrared images are best detected in the PYR fused images (see Table 1).

By the contrary we can see that the PYR fused images yield the worst performance in ten comparisons (see Table 2); while the CWT fused images yield the worst performance in two comparisons and the DWT fused images in nine comparisons.

It appears that high-saliency locations from visual images are worst detected using the PYR fusion scheme (see Table 2). Also it appears that high-saliency points from visual union infrared images are worst detected in the PYR and DWT fused images (see Table 2).

Table 3 summarizes the results for all the comparisons given in Tables 1 and 2 (see also

| Worst Performance | **Visual** | **Infrared** | **Visual $\cup$ Infrared** |
|---|---|---|---|
| **Set 1** | DWT | PYR | DWT |
| **Set 2** | CWT | DWT | DWT |
| **Set 3** | PYR | PYR | PYR |
| **Set 4** | PYR | DWT | PYR |
| **Set 5** | PYR | DWT | DWT |
| **Set 6** | PYR | DWT | DWT |
| **Set 7** | PYR | CWT | PYR |

Table 2: Worst Performance

| Best minus Worst | **Visual** | **Infrared** | **Visual $\cup$ Infrared** |
|---|---|---|---|
| **CWT** | 4 | 1 | 3 |
| **DWT** | 1 | -4 | -3 |
| **PYR** | -5 | 4 | 0 |

Table 3: Best minus Worst Performance

Fig. 13). From these results, it follows that the CWT fusion scheme appears to yield the best overall performance. These computational results largely agree with the order induced on the fused images by the detailedness (quality) of human segmentations (*26*).

In the literature, it has been found that a fusion scheme using a Wavelet Transform (like the CWT method) has advantages over the pyramid-based fusion such as increased directional information, no blocking artefacts that often occur in pyramid-fused images and improved perception compared using human analysis (*29,30*). The CWT provides both good shift invariance and directional selectivity over the DWT; and the increased shift invariance and directional sensitivity mean that the CWT gives improved fusion results over the DWT (*31*). These results agree with the comparative performance of the CWT, DWT and PYR schemes in the axiomatic score sense.

It would be of great interest if we compare the proposed method with other evaluation function. To this aim we now calculate the average attention score within the areas of interest, for each bitrate $\rho$, based on the attention map $\{A(P_i;\rho)\}_{P_i}$. The mean attention score within the areas of interest is normalized by dividing by the average attention achieved in the reconstruction at bitrate $\rho$. (*11*) provides a specification of the algorithm to compute the normalized attention score.

A high value of the mean attention within the areas of interest at reconstruction fidelity $\rho$ means that the coding method brings the attention onto the important regions using a bitrate $\rho$ of picture quality, which corresponds to a high saliency of the areas of interest. Hence it predicts a faster detection (due to the higher saliency) of the important areas in the input (visual and infrared) images; therefore the normalized attention score across bitrates can be used to rank order the important information visibility (see (*10*)).

For each one of the seven sets of input and fused images, in each row of Fig. 14 it is illustrated the rate-attention curves for the CWT, DWT and PYR fused images within automatically detected regions of interest for the visual image (left column) and for the infrared image (right column). The CWT, DWT and PYR fusion schemes yield the best performance in eight, five and one comparisons, respectively.

From these results, it follows that the CWT fusion scheme appears to yield the best overall performance in the rate-attention sense within high saliency regions for the visual image (left column) and for the infrared image (right column) which agrees with the comparative performance using the axiomatic score function proposed in this paper.

# References and Notes


1. Garcia, J.A., Rodriguez-Sanchez, R., Fdez-Valdivia, J., Toet, Lex, 2010. Comparative Visual Efficiency of Image Fusion Methods. Submitted to *Information Fusion*.

2. Garcia, J.A., Rodriguez-Sanchez, R., Fdez-Valdivia, J., Garrido, A., 2010. Axiomatic Approach to Image Fusion Evaluation. Submitted to *Information Sciences*.

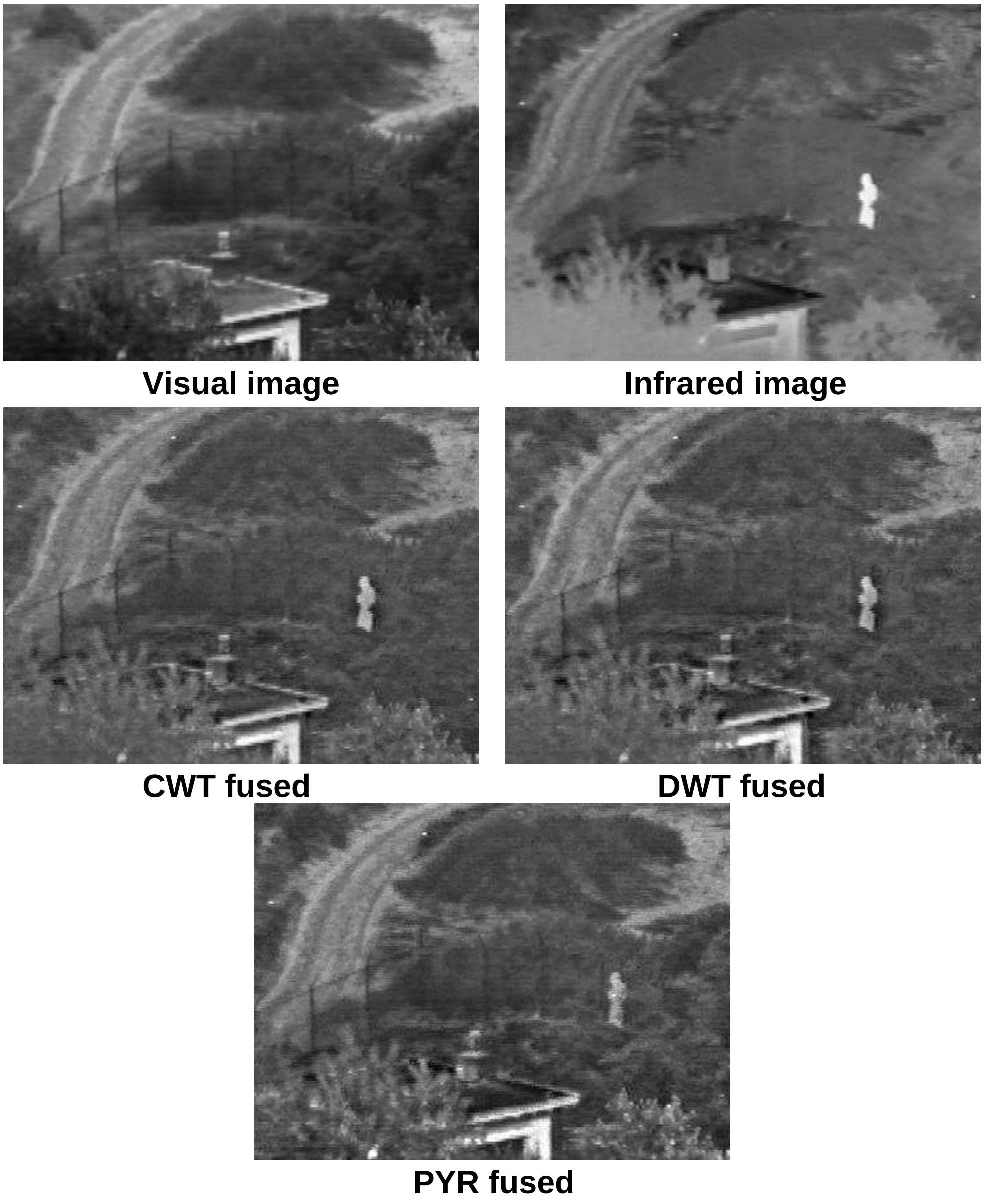


Figure 1: Set #1. Input and fused images

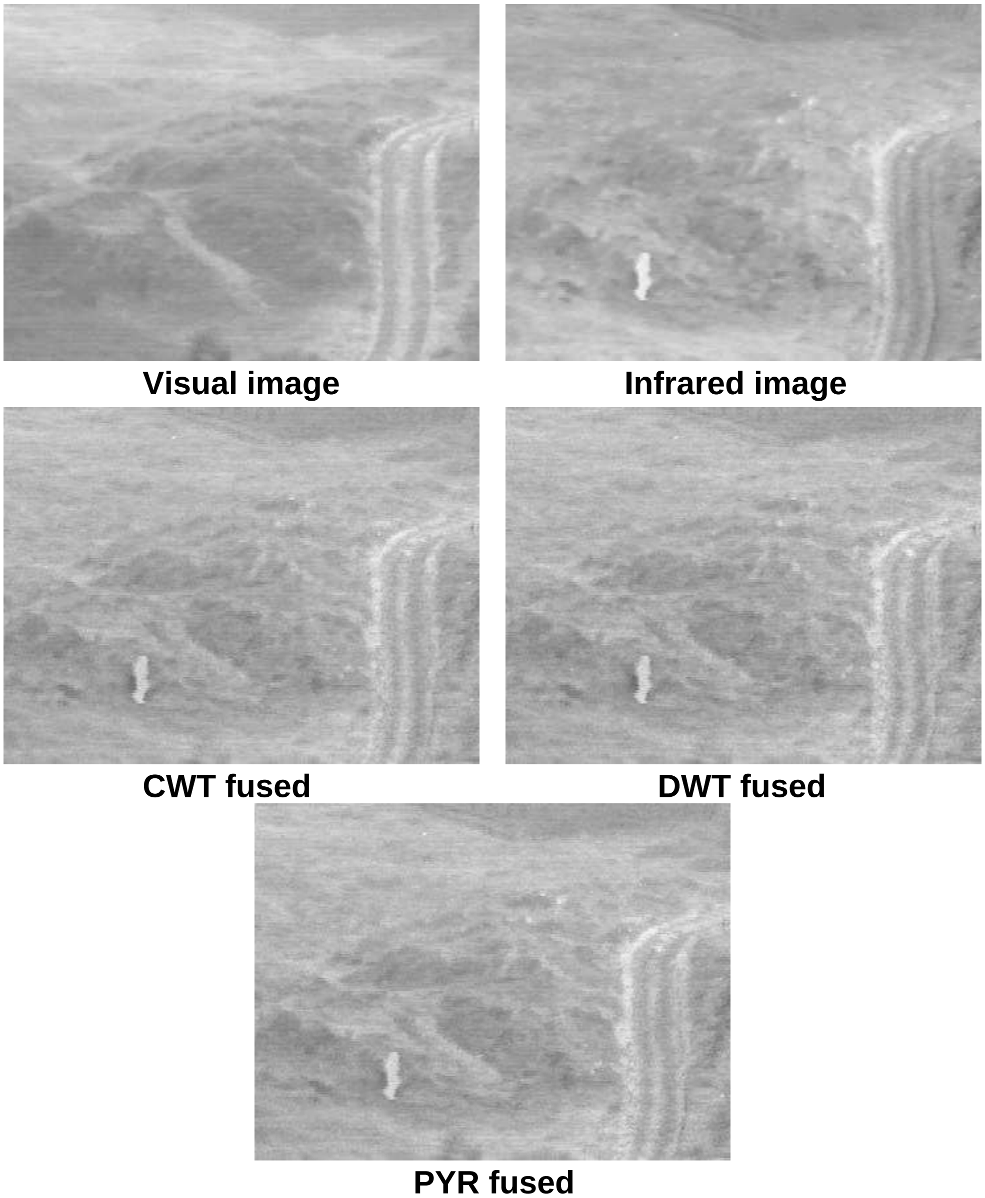


Figure 2: Set #2. Input and fused images

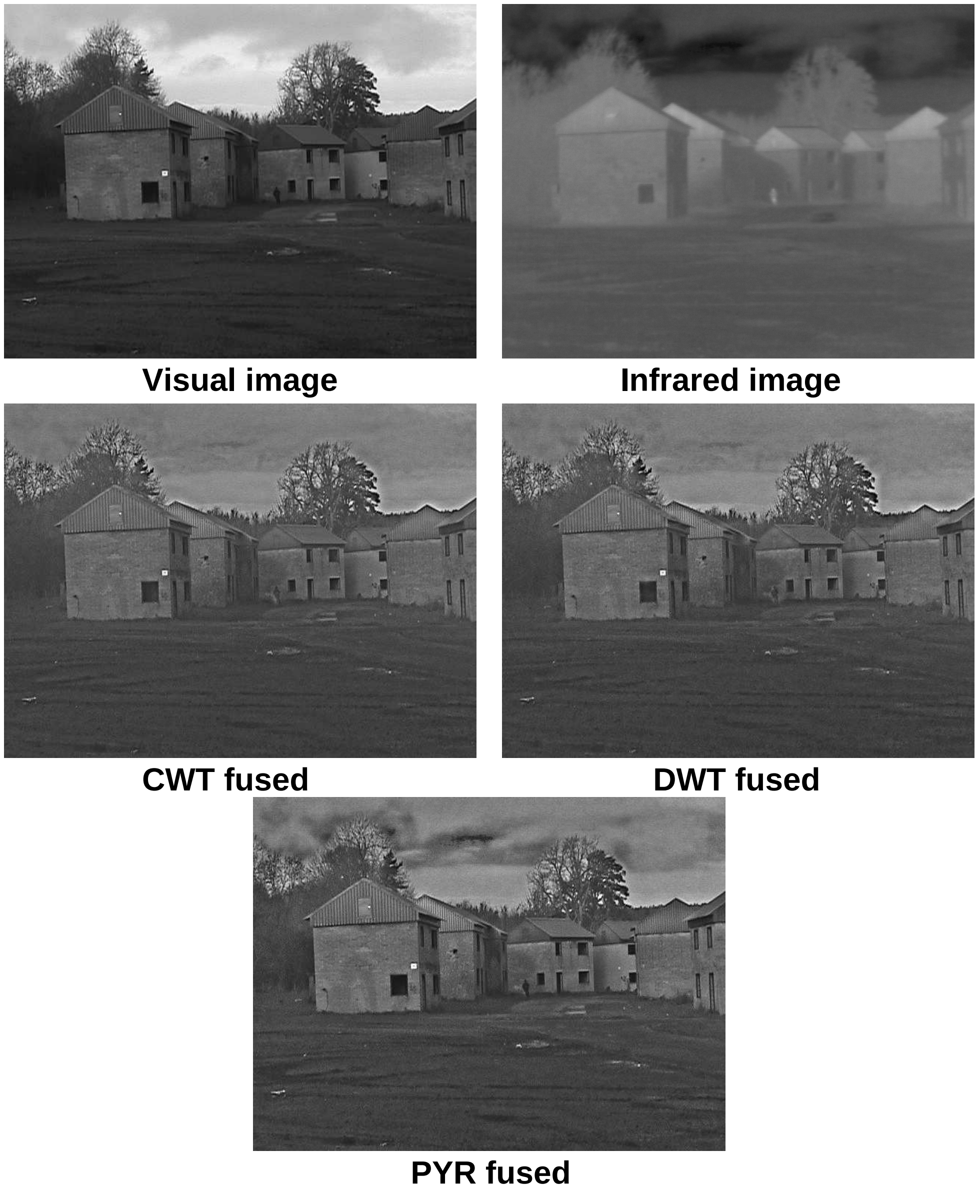


Figure 3: Set $\#3$. Input and fused images

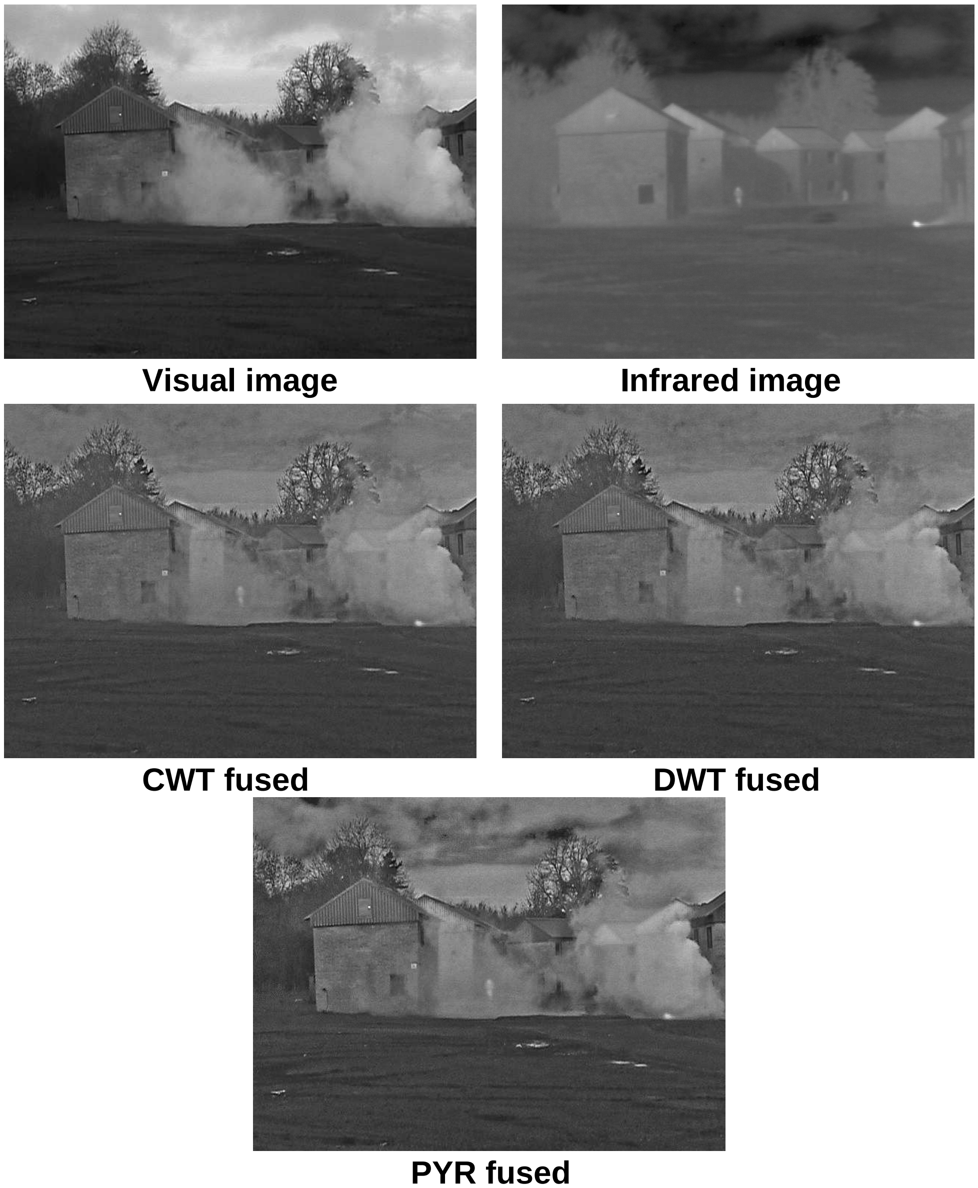


Figure 4: Set $\#4$. Input and fused images

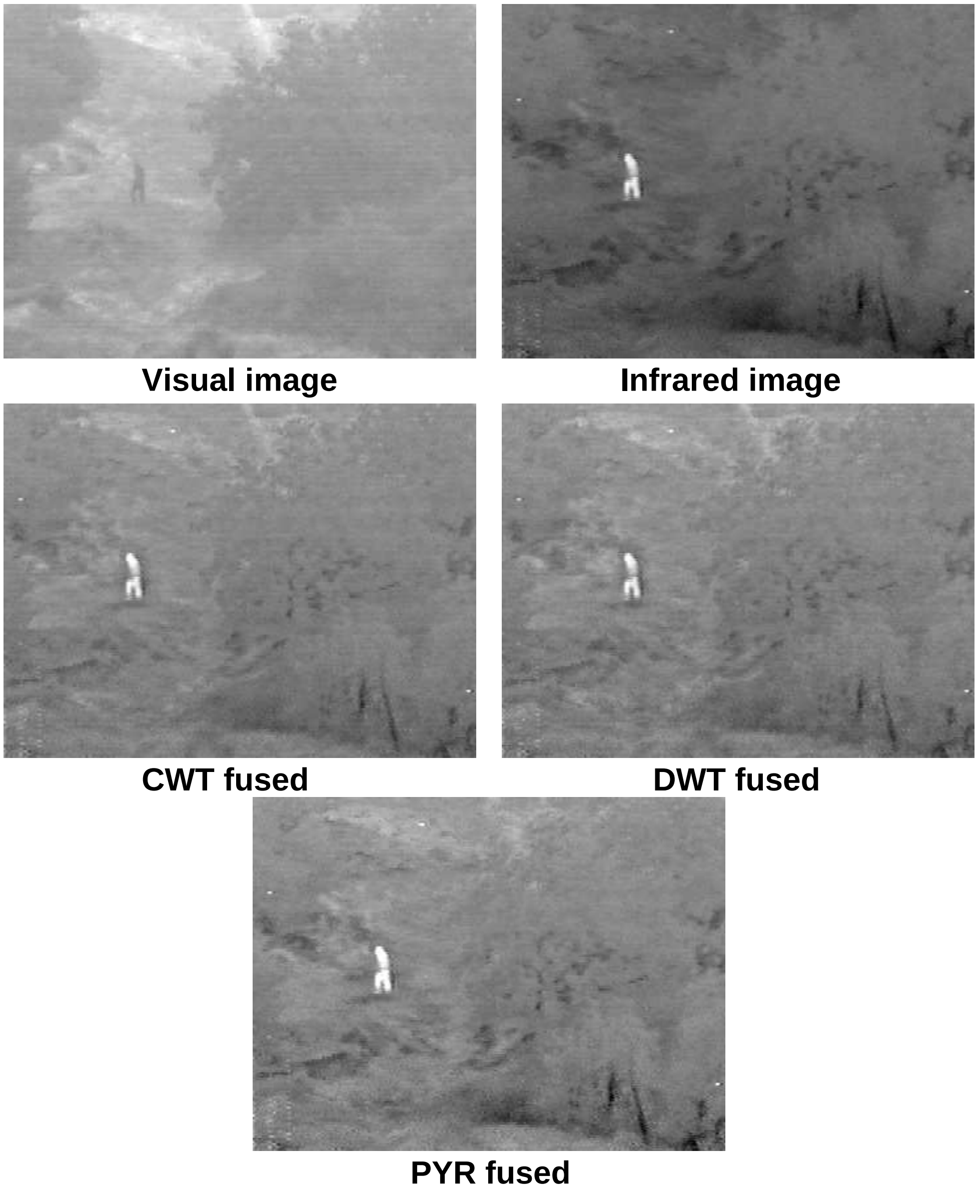


Figure 5: Set $\#5$. Input and fused images

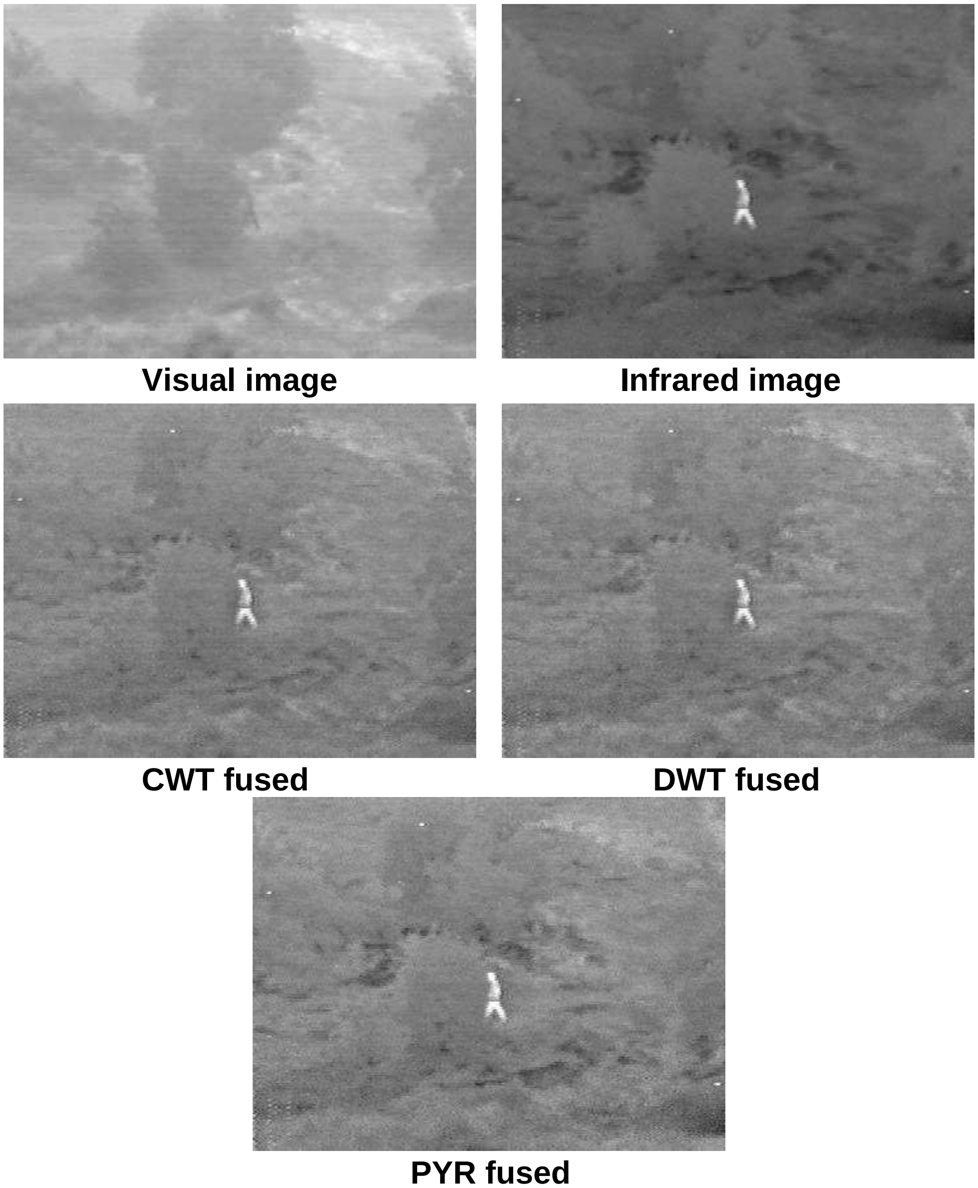


Figure 6: Set $\#6$. Input and fused images

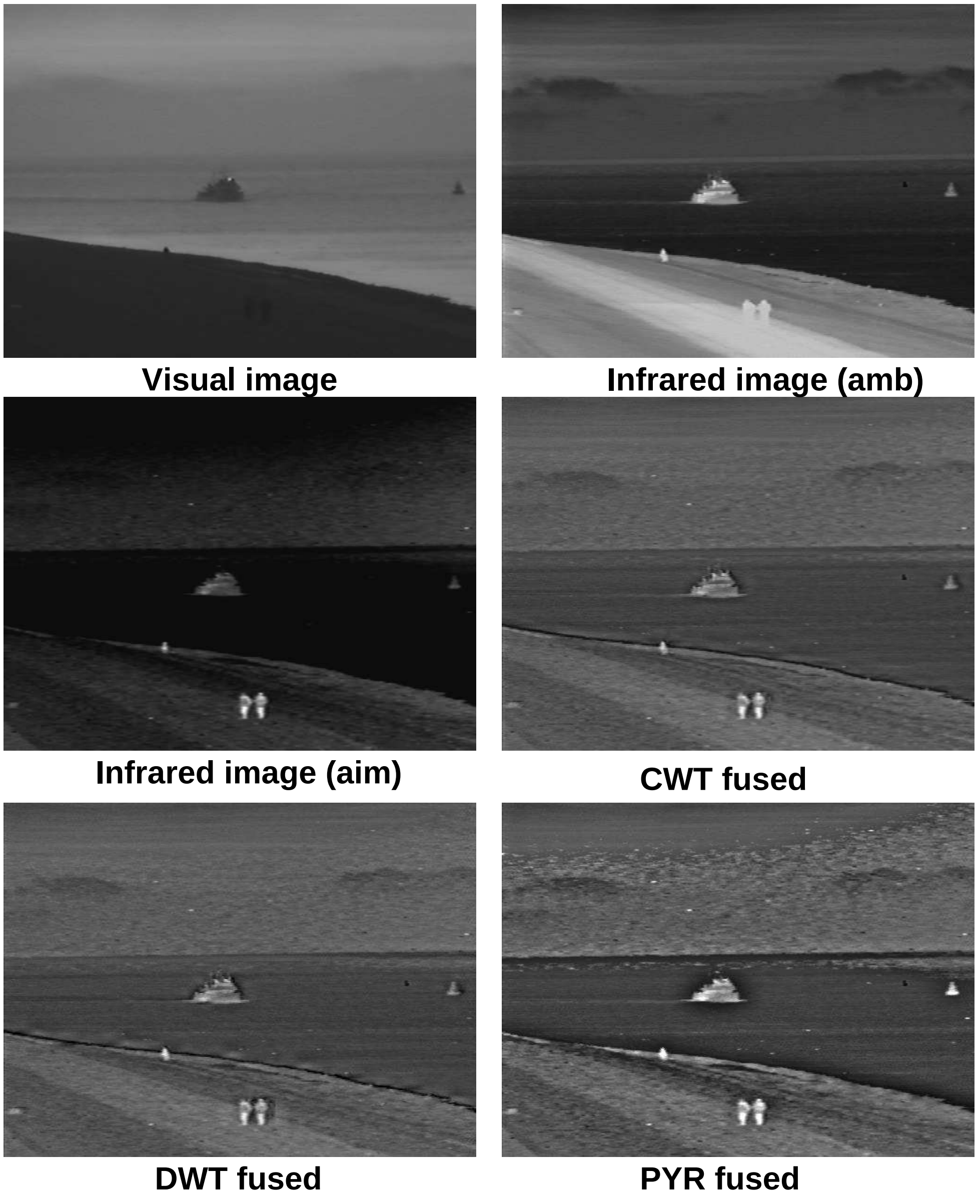


Figure 7: Set #7. Input and fused images

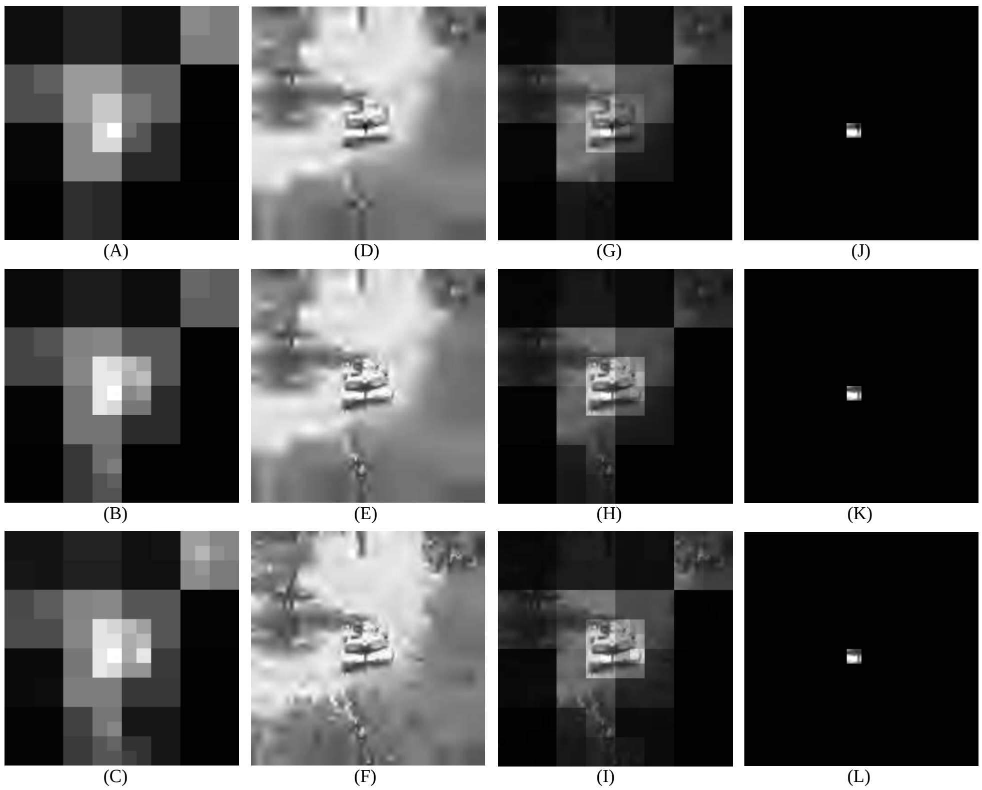


Figure 8: (A), (B), (C) attention scores for three different quality versions–given in (D), (E) and (F)–of a highly visible target; (G), (H), (I) Blending of the respective score maps and image reconstructions; (J), (K), (L) Most salient locations.

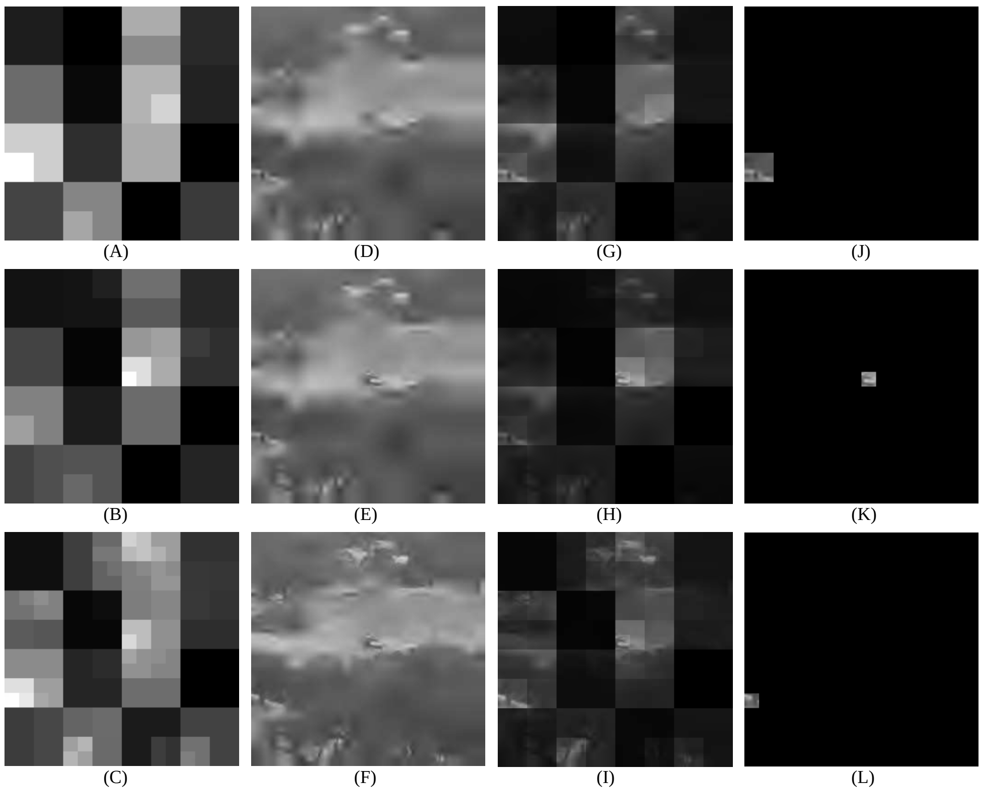


Figure 9: (A), (B), (C) illustrate attention scores for three quality versions, respectively (D), (E) and (F), of a military vehicle in low visibility conditions; (G), (H), (I) show blendings of the respective score maps and image reconstructions; (J), (K), (L) Most salient locations.

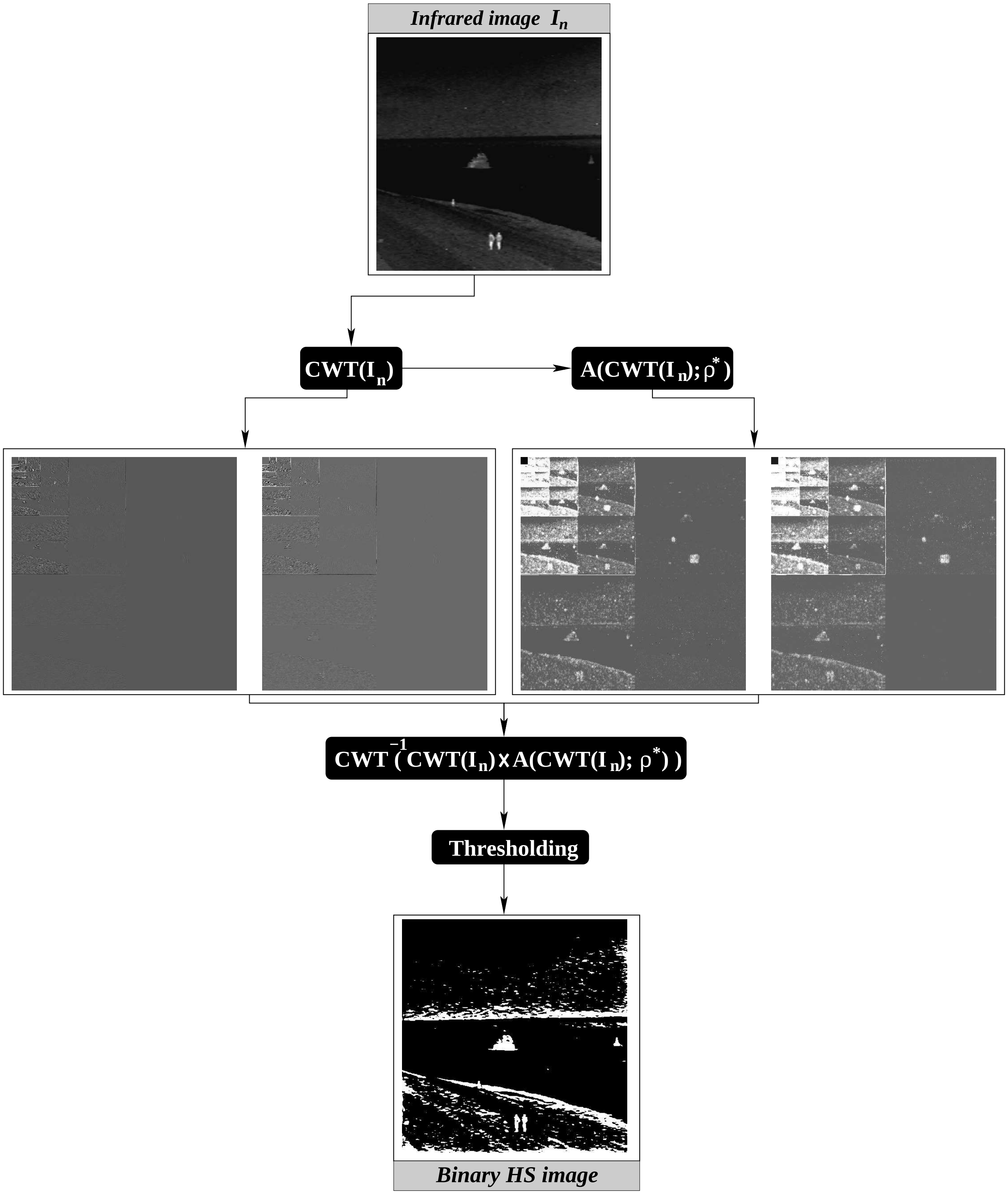


Figure 10: Automatic detection of high-saliency locations for $I_n$ = Infrared image; at a bitrate $\rho^* = 0.5$ bit/pixel.

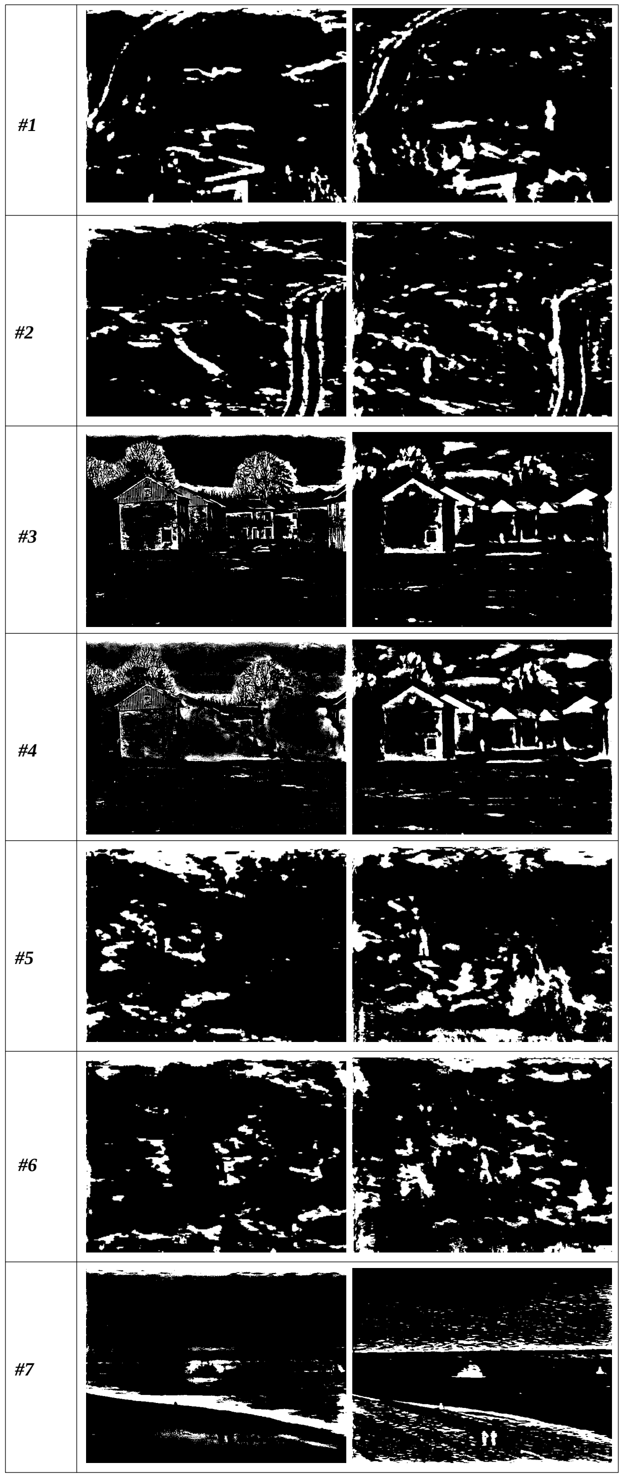


Figure 11: High-Saliency image (HS) automatically detected, at a bitrate $\rho^* = 0.5$ bit/pixel, for the visual image (left column) and the infrared image (right column) in each one of the seven sets of input images.

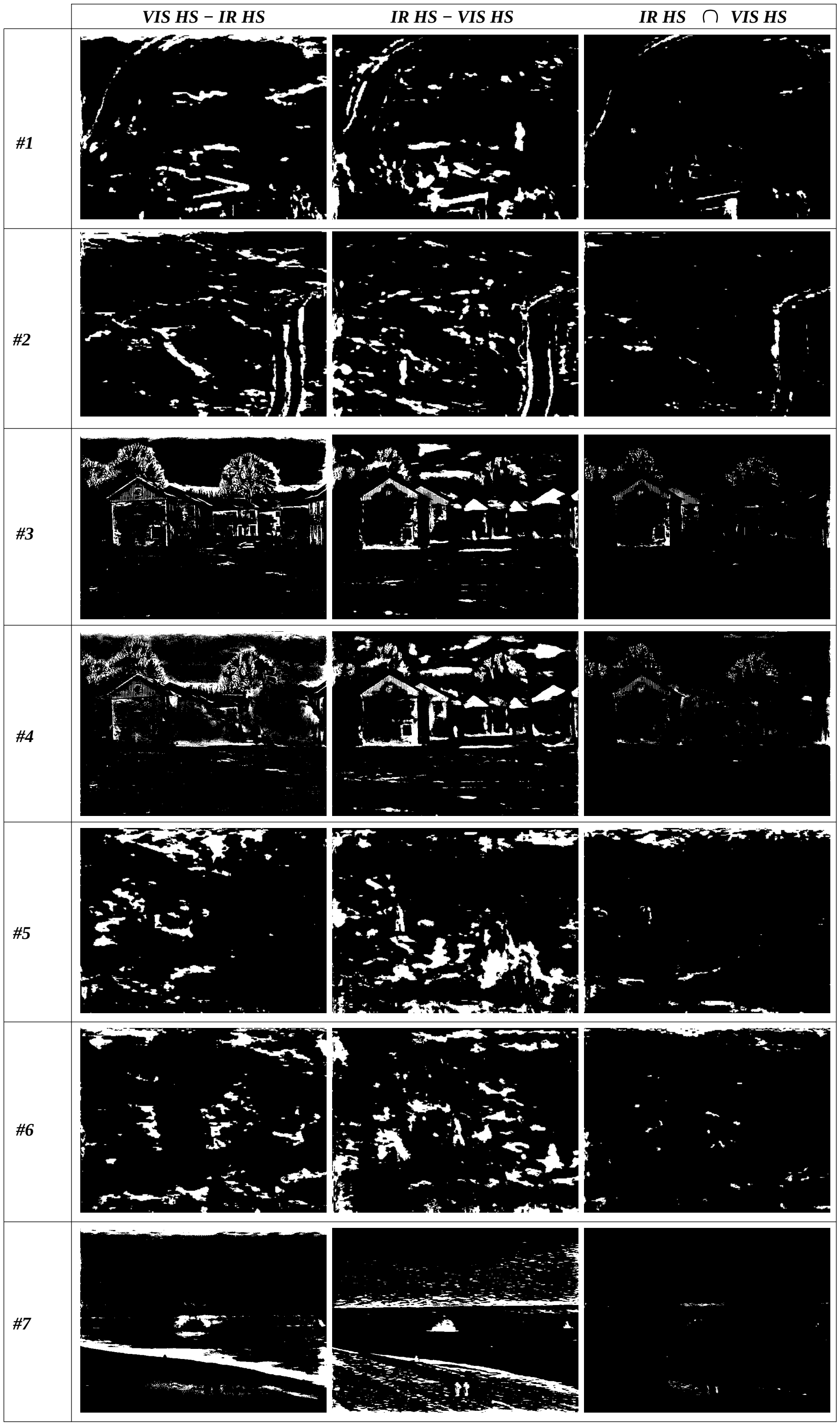


Figure 12: (left column) Visual HS minus Infrared HS; (middle column) Infrared HS minus Visual HS; (right column) Infrared HS intersect Visual HS.

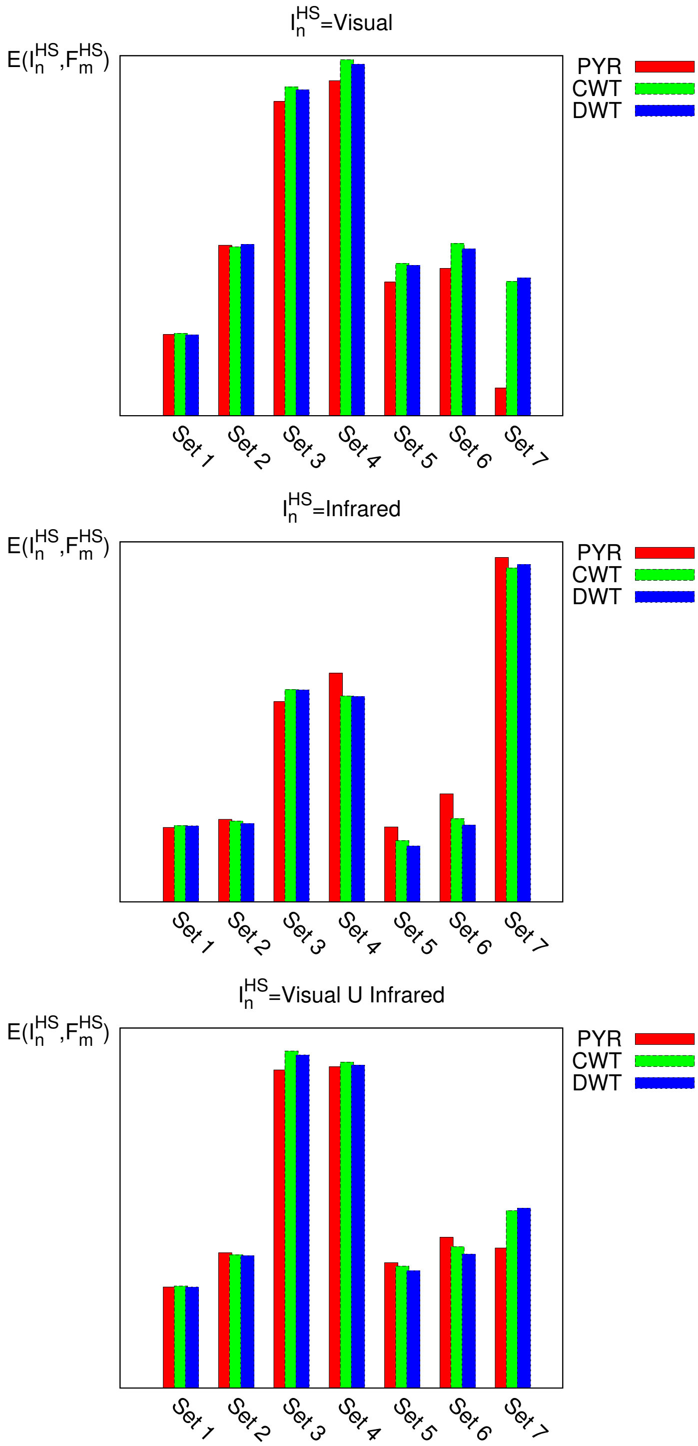


Figure 13: For each fused (CWT, DWT, PYR) image, axiomatic score across datasets (Set #1 through Set #7) using as reference the visual image (top), the infrared image (center), and both the visual and infrared image (bottom).

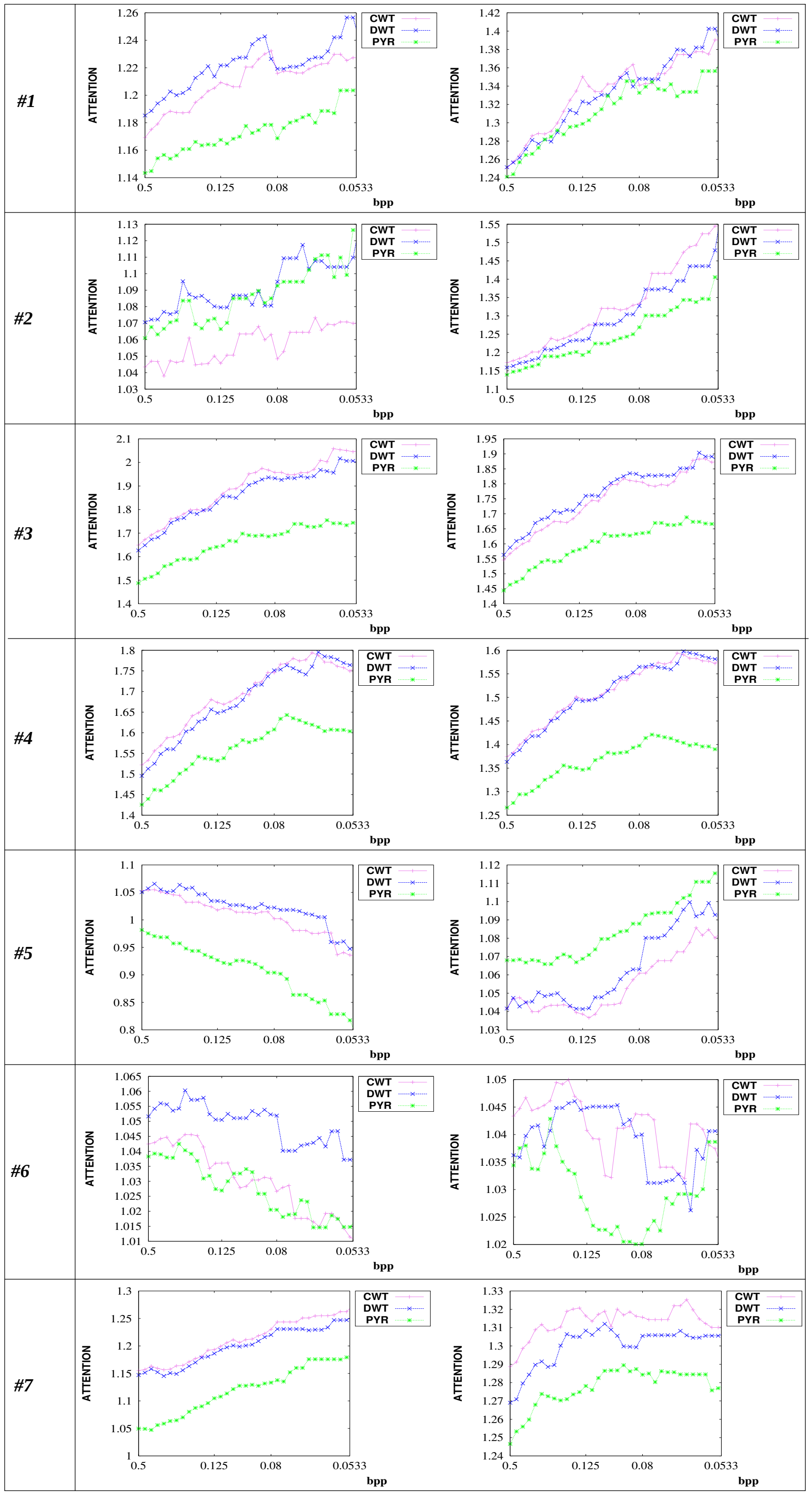


Figure 14: Rate-attention curves for the CWT, DWT and PYR fused images within automatically detected regions of interest for the visual image (left column) and the infrared image (right column).

3. Y. Zheng, E. A. Essock, B. C. Hansen and A. M. Haun, A new metric based on extended spatial frequency and its application to DWT based fusion algorithms. Information Fusion, Special Issue on Image fusion: Advances in the state of the art, 8(2):177-192, April 2007
4. H. Chen and P. K. Varshney, A human perception inspired quality metric for image fusion based on regional information. Information Fusion, Special Issue on Image fusion: Advances in the state of the art, 8(2):193-207, April 2007
5. V Petrovic, C Xydeas, ”Objective Evaluation of Signal-level Image Fusion Performance”, Optical Engineering, SPIE, Vol 44(8), 087003, 2005
6. G Qu, D Zhang, P Yan, Information measure for performance of image fusion, Electronics Letters, Vol. 38(7), pp 313-315
7. G Piella, H Heijmans, A New Quality Metric for Image Fusion, Proceedings International Conference on Image Processing, 2003, Vol. 3, pp 173-176
8. Bergen, J.R., Julesz, B., 1983. Parallel versus serial processing in rapid pattern discrimination. Nature 303, 696-698.
9. Boynton, G.M., 2005. Attention and visual perception. Current Opinion in Neurobiology 15, 465-469.
10. Garcia, J.A., Rodriguez-Sanchez, R., Fdez-Valdivia, J., 2010. Axiomatic Approach to Computational Attention. Pattern Recognition. Vol. 43(4), pp. 1618-1630, doi:10.1016/j.patcog.2009.09.027
11. Garcia, J.A., Rodriguez-Sanchez, R., Fdez-Valdivia, J., Martinez-Baena, J., 2010. Information Visibility Using Transmission Methods. Pattern Recognition Letters. Vol. 31(7), pp. 609-618, doi:10.1016/j.patrec.2009.12.001

12. Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Patt. Anal. Machi. Intell. 20, 1254-9.

13. Itti, L., Baldi, P., 2006. Bayesian surprise attracts human attention. Advances in Neural Information Processing Systems. MIT Press, Cambridge, MA, 1-9.

14. James, W., 1950. The Principles of Psychology. Dover Publications.

15. http://www.kakadusoftware.com/

16. Jennifer T. Coull, 2004. fMRI studies of temporal attention: allocation attention within, or towards, time. Cognitive Brain Research 21, 216-226.

17. Koch, C., Ullman, S., 1985. Shifts in selective visual attention:towards the underlying neural circuitery. Hum. Neurobiol. 4, 219-227.

18. Le Meur, O., Le Callet, P., Barba, D., 2006. A coherent computational approach to model the bottom-up visual attention. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5), 802-817.

19. Lewis, J.J., Nikolov, S.G., Canagarajah, C.N., Bull, D.R., Toet, A., 2006. Uni-modal versus joint segmentation for region-based image fusion, in: Proceedings of the Ninth International Conference on Information Fusion, International Society of Information Fusion, Mountain View, CA, 2006.

20. M. Mancas 2007. Computational Attention: Towards attentive computers. Presses universitaires de Louvain, ISBN: 978-2-87463-099-6, Belgium, 1-267.

21. Mudge, T.N., Turney, J.L., Voltz, R.A., 1987. Automatic generation of salient features for the recognition of partially occluded parts. Robotica 5, 117-127.

22. Oliva, A., Torralba, A., Castelhano, M.S., Henderson, J.M., 2003. Top-down control of visual attention in object detection. Proc. IEEE International Conference Image Processing 1, 253-256.

23. Osberger, W., Maeder, A.J., 1998. Automatic identification of perceptually important regions in an image using a model of the Human Visual System . The 14th International Conference on Pattern Recognition. Brisbane, Australia.

24. Posner, M.I., Cohen, Y., Rafal, R.D., 1982. Neural systems control of spatial orienting. Philos. Trans. R. Soc. Lond. B Biol. Sci. 298, 187-198.

25. Privitera, C.M., Stark, L., 2000. Algorithms for defining visual regions of interest: Comparison with eye fixations. IEEE Trans. on Pattern Analysis and Machine Intelligence 22, 970-982.

26. Toet, A., Hogervorst, M.A., Nikolov, S.G., Lewis, J.J., Dixon, T.D., Bull, D.R., Canagarajah, C.N., 2009. Towards cognitive image fusion, Information Fusion, doi: 10.1016/j.inffus.2009.06.008.

27. Treisman, A., Gelade, G., 1980. A feature integration theory of attention. Cognitive Psychology 12(1), 97-136.

28. Walker, K.N., Cootes, T.F., Taylor, C.J., 1998. Locating salient object features. Proc. British Machine Vision Conference. (Eds: P.H. Lewis and M.S. Nixon) Vol. 2, 557-566.

29. Li, H., Manjunath, S., Mitra, S., 1995. Multisensor image fusion using the wavelet transform. Graphical Models and Image Processing 57(3), 235-245.

30. Wilson, T., Rogers, S., Kabrisky, M., 1995. Perceptual based hyperspectral image fusion using multi-spectral analysis. Optical Engineering 34(11), 3154-3164.

31. Nikolov, S.G., Hill, P.R., Bull, D.R., Canagarajah, C.N., 2001. Wavelets for image fusion, in: Wavelets in Signal and Image Analysis, Computational Imaging And Vision Series, Kluwer Academic Publishers, Dordrecht, The Netherlands, 213-244.

32. Rohaly, A.M., Ahumada, A.J., and Watson, A.B., 1997. Object detection in natural backgrounds predicted by discrimination performance and models. Vision Research, 37 (23), 3225-3235.

33. Kingsbury, N., 2000. The dual-tree complex wavelet transform with improved orthogonality and symmetry properties, in: Proceedings IEEE Conference on Image Processing, Vancouver.

34. Rockinger, O. and Fechner, T, (1998). Pixel-level image fusion: The case of image sequences, in: Proc. SPIE, Vol. 3374, pp. 378-388. Examples on-line in http://www.metapix.de

35. Zeeuw, P.M., Piella G., Heijmans, H.J. (2004). A Matlab Toolbox for Image Fusion (MATIFUS). CWI. Probability, Networks and Algorithms, No. E-0424, pp.1-19. Software avalaible in: http://www.imagefusion.org/software.html

36. Rodriguez-Sanchez, R., Garca J.A., Fdez-Valdivia. J., (2010). From Computational Attention to Image Fusion. Image and Vision Computing.

37. Mitchell, H.B. (2010). Image Fusion Theories, Techniques and Applications. Springer Berlin Heidelberg.

**Acknowledgments.** This paper was sponsored by the Spanish Board for Science and Technology (MICINN) under grant TIN2010-15157.