

Multi-level Image Fusion

Vladimir Petrović

Orasys MV, 5 Bore Kostića, 11077 Novi Beograd, Serbia
and Imaging Science Biomedical Engineering, University of Manchester, M13 9PL, UK

ABSTRACT

Signal-level image fusion has in recent years established itself as a useful tool for dealing with vast amounts of image data obtained by disparate sensors. In many modern multisensor systems, fusion algorithms significantly reduce the amount of raw data that needs to be presented or processed without loss of information content as well as provide an effective way of information integration. One of the most useful and widespread applications of signal-level image fusion is for display purposes. Fused images provide the observer with a more reliable and more complete representation of the scene than would be obtained through single sensor display configurations. In recent years, a plethora of algorithms that deal with the problem of fusion for display has been proposed. However, almost all are based on relatively basic processing techniques and do not consider information from higher levels of abstraction. As some recent studies have shown this does not always satisfy the complex demands of a human observer and a more subjectively meaningful approach is required. This paper presents a fusion framework based on the idea that subjectively relevant fusion could be achieved if information at higher levels of abstraction such as image edges and image segment boundaries are used to guide the basic signal-level fusion process. Fusion of processed, higher level information to form a blueprint for fusion at signal level and fusion of information from multiple levels of extraction into a single fusion engine are both considered. When tested on two conventional signal-level fusion methodologies, such multi-level fusion structures eliminated undesirable effects such as fusion artefacts and loss of visually vital information that compromise their usefulness. Images produced by inclusion of multi-level information in the fusion process are clearer and of generally better quality than those obtained through conventional low-level fusion. This is verified through subjective evaluation and established objective fusion performance metrics.

Keywords: image fusion, signal-level fusion, feature-level fusion, multi-level fusion

1. INTRODUCTION

The emergence and relatively fast proliferation of multisensor arrays has established multisensor information fusion¹ and image fusion in particular as important areas of research. The availability of reliable and accurate imaging sensors at ever decreasing prices have combined to make the use of multisensor arrays a widespread practice in applications such as remote sensing, night vision, medical imaging and security and surveillance. Integration into systems of disparate sensors as independent information sources can improve their performance and add greatly to their robustness. Multisensor arrays cover broader portions of the spectrum, can offer higher sensitivity and resolution and are generally more reliable than single sensor suites, especially in adverse conditions². However, the introduction of additional sensors into an existing array, although generally advantageous, introduces a number of characteristic problems. One of the most significant is the problem of data overload. In practice, additional sensors can increase the amount of data produced by the array to the levels with which the processing power of the system cannot cope. Similarly, presenting human observers with more than one visual cue simultaneously causes confusion and diminishes their performance³.

Signal-level image fusion is an established tool for dealing with vast amounts of image data obtained by disparate sensors. Fusion algorithms provide for a significant reduction in the amount of raw data without loss of information content. Additionally, fusion provides effective information integration, which is not possible when processing single sensor outputs individually. This is of particular importance in one of the most useful and widespread applications of signal-level image fusion, for display purposes. Fused images provide the observer with a better and more complete representation of the scene than would be obtained through otherwise ineffective single sensor display configurations². In response to this, a plethora of algorithms that perform fusion for display have been proposed.

One of the earliest multisensor image fusion methods, proposed by Toet³, was based on multiresolution image analysis using contrast pyramids. Discrete Wavelet Transform was first used in image fusion purposes in a system by Li *et. al.*⁴,

and a thorough investigation into DWT and other related multiscale and multiresolution fusion methods is provided by Zhang and Blum in ¹³. Multiscale fusion for visual display was considered, with additional orientation sensitivity by Peli *et. al.* ⁵ and in an efficient real time framework by Petrović and Xydeas ⁶. More recently, a fusion system based on wavelet transform modulus maxima was proposed by Qu *et. al.* ⁷ and a multiscale method based on morphological towers was presented by Mukhopadhyay and Chanda ⁸ for fusion of medical images. However, almost all fusion methods presented so far are based on relatively basic processing techniques and do not consider subjectively relevant information from higher levels of abstraction. As some recent studies have shown this does not always satisfy the complex demands of a human observer ⁹⁻¹¹ and a more subjectively meaningful approach is required.

This paper presents some initial results of a fusion framework based on the idea that subjectively relevant fusion could be achieved if information at higher levels of abstraction such as image edges and image segment boundaries are used to guide the basic signal-level fusion process. Fusion of processed, higher level information to form a blueprint for fusion at signal level and of information from multiple levels of extraction into a single fusion engine are both considered. Such a multi-level approach is able to eliminate undesirable effects such as *fusion artefacts* and loss of vital visual information, generally improve the quality of the fused image and overall reliability of the fusion process.

The next section describes the general structure of image based information fusion and outlines two basic techniques for signal-level image fusion, which were tested within the multi-level fusion framework. Section 3 outlines the proposed multi-level image fusion system structure as well as methods for combining information from different levels of abstraction. Preliminary results of the proposed multi-level fusion systems are presented and discussed in section 4 while the paper is concluded in section 5.

2. IMAGE INFORMATION FUSION

Image based information fusion can be performed at three different levels of abstraction: signal, feature and decision level ¹. Universal fusion system structure by Dasarathy ¹² that illustrates them is shown in Figure 1. Main difference between the levels is in the amount of processing that is performed on the image prior to fusion and hence the format in which this information is fused and the type of fusion techniques applied. The information is captured from an observation of the scene by the sensors, which present it to the system in form of two digital image signals (Input Images). These images can be combined directly (signal-level fusion) into a fused image that represents the information present in the input images in a single signal. Alternatively, input images (and potentially the fused) can be processed (e.g. edge detection, segmentation) to extract information about the basic features present in them. This information is of a more descriptive nature and can be combined from all cues into a single feature description set (Fused Feature Set) by applying feature-level fusion techniques. This information then forms a basis for reaching decisions about (evaluating) the observed scene. Local decision makers produce probabilistic inferences about the scene from the feature sets provided by the lower level and these can be fused using decision level fusion techniques into a final evaluation (of the state) of the observed scene. This structure is important in the context of the concepts presented in this paper since it illustrates well the one directional flow of information that the proposed framework reverses in order to obtain a more reliable and visually acceptable fused image.

2.1 Signal-level Image Fusion

Multisensor signal-level image fusion is a process of combining several image signals produced by disparate sensors into a single, fused image. The goal is to display, in the fused image all the useful information presented in any number of multisensor inputs. Additionally, the fusion algorithm must ensure that no significant information loss occurs and no false information (fusion artefacts) is introduced into the fused image. In broad terms, all the features visible in the individual input images should be as clearly visible in the fused image.

Two specific signal-level image fusion schemes are examined in the context of multi-level image fusion. Discrete Wavelet Transform (DWT) fusion ^{4,13} represents a full multiresolution information integration framework that combines input features at a number of different scales and orientations independently. Its general structure is illustrated in Figure 2. The input images are transformed through the process of multiresolution image decomposition into a representation domain where different parts of the original image spectrum are separated into different sub-band signals. The final arrangement of these sub-band signals is called the *image pyramid*.

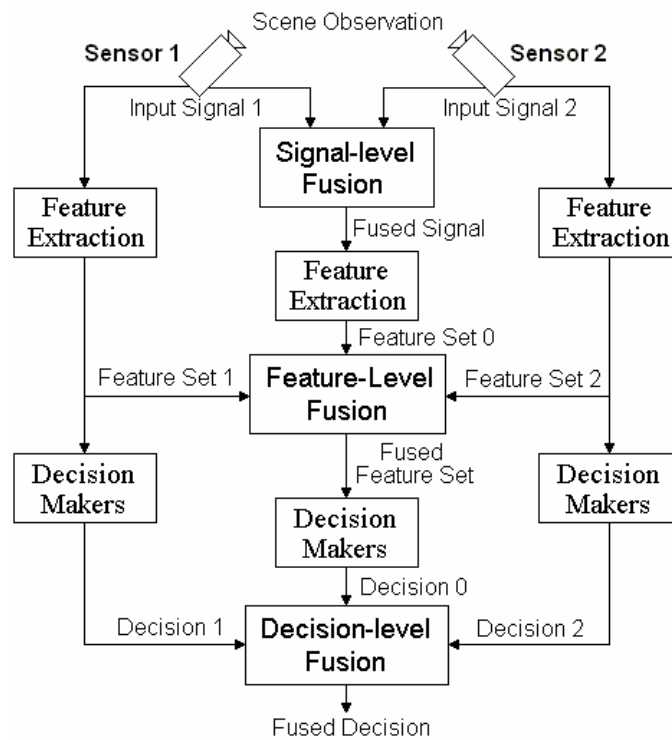


Figure 1: Universal Fusion System Architecture by Dasarathy

Fusion is performed in the pyramid domain by creating a fused pyramid using the information present in the input pyramids. This is usually referred to as the *pyramid fusion* process and can be performed in a number of different ways. The most successful approach is to use some form of *feature selection* that directly compares input pyramids coefficients on the basis of their importance and selects the one deemed more important for the fused pyramid¹³. Practically, selection maps are formed that indicate, at each fused pyramid pixel, which of the input pyramids is to be used as a source to copy the value from. These selection maps are the most optimal entry point for feature level information, explained in the following section. Finally, once a fused pyramid is completed it is input into the image reconstruction process, reverse of image decomposition, which produces the fused image.

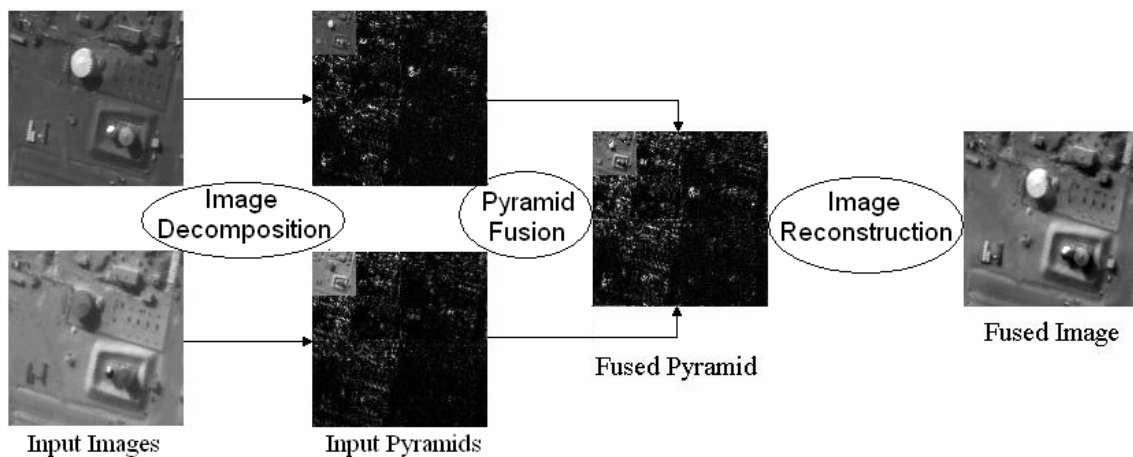


Figure 2: General Structure of a Multiresolution/Multiscale Image Fusion Mechanism

The second fusion scheme of interest is the efficient Bi-scale Multisensor Image Fusion (BMIF) proposed by Petrović and Xydeas⁶. Its basic structure is similar to DWT fusion in Figure 2, only here fusion is performed at only two levels of scale. The 2 sub-band signals are obtained by simple average filtering of input signals (giving the *background image* containing low-pass features) and subtraction of this signal from the original (giving the *foreground signal* containing high-pass features). There is no sub-sampling and there exists a direct spatial correspondence between the original image and the sub-bands. Furthermore, the foreground sub-bands are fused in the same way as DWT pyramids, through selection maps formed on the basis of the two input foreground signals⁶.

3. MULTI-LEVEL IMAGE FUSION

In order to obtain fused images that appear natural to the observer but still contain all the important information from the input images, a reliable method of identifying subjectively meaningful input image entities to be used in the process of signal-level image fusion must be found. Extraction of meaningful structure is precisely the aim of various image analysis methods employed at the feature-level of the universal fusion system architecture in Figure 1. Image fusion performance should therefore be improved by feeding back this information into the basic signal-level fusion process. Considering the structure in Figure 1, a structure of a multi-level process that uses feature-level information to fuse image signals is expressed in Figure 3. In this case, input signals are processed prior to fusion in order to identify meaningful structures they contain. These feature sets are then combined using feature-level fusion methods in order to obtain a clear, unified picture of what entities should be preserved in the fused images. This information is then fed back into the signal level fusion process, which produces the fused image according to its specifications. The most obvious difference between this and the universal fusion structure in Figure 1 is the feedback loop that reverses the flow of information from the feature back to the signal level.

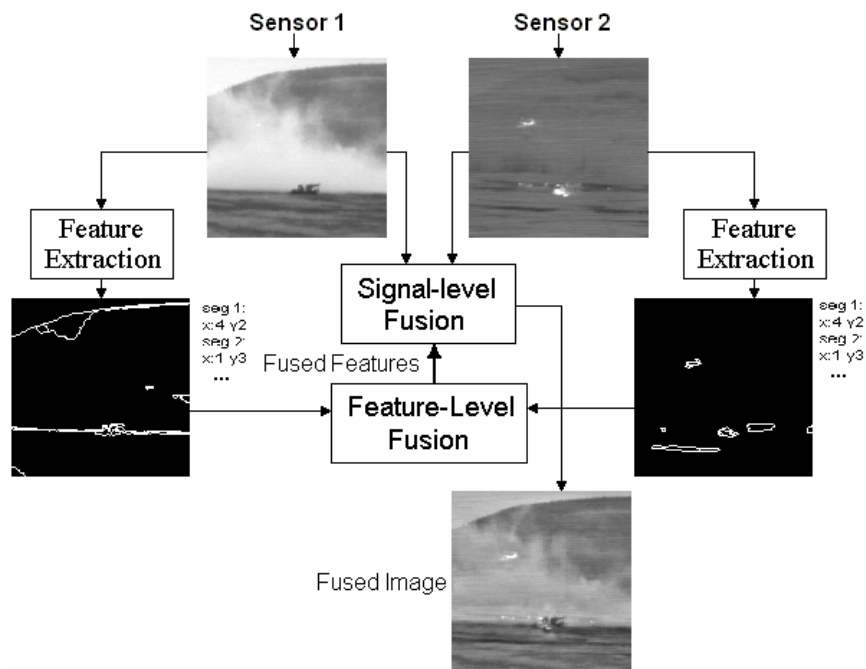


Figure 3: Multi-level image fusion structure

In the proposed multi-level fusion framework, feature level information used to guide the signal-level fusion process is in the form of edges and image segment boundaries. Image region boundaries are a basic element in the perception of image structure¹⁴. Feedback of these simple image descriptors as feature-level cues to the signal-level fusion process is also motivated by the view that the goal of image fusion can be expressed as representation, in the fused image, of all the edge information that is present in the input images^{15,16}. Furthermore, this form of feature level information is easily incorporated into the signal-level fusion process and the next two sub-sections describe how this information is first extracted from the input images, fused and then included into two basic signal-level fusion methodologies.

3.1 Feature extraction

Image boundaries and edges are changes in image intensity that capture human attention and transfer information about the observed scene. Their extraction for the purposes of the proposed multi-level fusion system can proceed according to two basic approaches: i) through direct edge detection which may not result in complete segment boundaries and ii) through the full image segmentation process which divides the image into a finite number of distinct regions with discretely defined boundaries. In this instance, only the former approach in the form of a widely accepted Canny edge detection method¹⁷ is used to demonstrate the impact of this type of information on the signal-level image fusion process. The feature extraction part of the process therefore, uses a single image input (e.g. A) to produce an edge map output E_A that indicates, in a binary way, a presence or a lack of, an edge at each pixel in the output image ($E_{A,n,m}=1$ edge at $A(n,m)$, $E_{A,n,m}=0$ no edge at $A(n,m)$). For a two input image fusion, edge map binary images are defined for both inputs A and B (E_A and E_B).

3.2 Feature level-fusion

The information in edge maps E_A and E_B that indicate the locations of important boundary locations in the input images are fused into two further boundary maps that indicate what boundaries the fusion process should preserve from each input image. Although more advanced methods have in the meantime become available¹⁸, reasonably successful fusion of these simple image feature information sets can be achieved, as the results indicate, using only simple pixel based rules. The simplest method is to feed image boundary maps E_A and E_B unchanged into the signal-level fusion process (referred to as OR fusion). However, due to the fact that input images will not be fully exclusive as they represent the same scene, it is likely that many of the boundaries will be present in both maps. As the primary goal of the fusion process is to ensure all the exclusive region boundaries from the input images are present in the fused image (common boundaries will be automatically included through signal level fusion) a different method of XOR fusion can be applied. In this case a boundary point in E_A or E_B is kept only if it is exclusive to its image. Otherwise, it is removed which leaves the pyramid fusion mechanism to resolve which of the two input boundaries is more significant. This is expressed for input A through equation (1) where XOR and AND represent the logical exclusive OR and AND operations on the input edge maps. Using the same terminology OR fusion described above becomes $E_A = X_A$. In both cases an equivalent process is applied to produce the fused exclusive edge map for input image B , X_B .

$$X_A = (E_A \text{ XOR } E_B) \text{ AND } E_A \quad (1)$$

3.3 Integrated signal-level image fusion

The two signal-level fusion methodologies investigated in this study, the DWT^{4,13} and BMIF⁶, were augmented with feature level information by modifying (fusing) the selection maps, that guide the process of HP sub-band (pyramid) fusion, with feature-level boundary information. In the DWT case, sub-band fusion takes place at a number of scales (resolutions) with feature level information being input at each separately. However, since there is no direct spatial correspondence between the raw image signal and high-frequency sub-bands produced from it, features extracted from the low-pass approximation of the image at the next level of resolution are used to guide the sub-band fusion process. Although this information may not necessarily correspond to the information present in the HP sub-bands, it is still meaningful as most realistic image structures produce significant features at a number of scales¹⁴. BMIF system on the other hand, exhibits direct correspondence between high-pass *foreground image* and the original image signal and spatial matching is automatic.

In both cases, the values of HP sub-band signals represent local saliency and direct selection maps are produced as in the case of direct signal-level fusion by selecting the coefficient with the larger absolute value for the fused image^{3,13}:

$$S_{n,m} = \begin{cases} 1, & \text{if } |A_{n,m}| > |B_{n,m}| \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $S_{n,m}$ is the selection for pixel at (n,m) and $A_{n,m}$ and $B_{n,m}$ are corresponding input pyramid coefficients ($S=1$ indicates that the fused pixel should come from A and 0 that is should come from B). Selection maps obtained in this way are fused with the fused feature-level maps for each input image through a simple logical OR operation. This produces two separate selection maps, one for each input sub-band, that indicate which of its pixels are to be passed on into the fused signal. Additionally, since in both sub-band domain types considered (DWT and BMIF) image

boundaries cause significant values in whole neighbourhoods of sub-band coefficients around their original position ^{4,6,13}, fused boundary maps are morphologically dilated by a simple square element in order to include all the necessary elements to reconstruct the boundary in the fused sub-band. The fused selection maps thus become:

$$S_{n,m}^A = S_{n,m} \mid \text{dilate}(X_A, k \times k)_{n,m} \quad (3)$$

$$S_{n,m}^B = (1 - S_{n,m}) \mid \text{dilate}(X_B, k \times k)_{n,m} \quad (4)$$

where \mid signifies a logical OR operation on binary maps and $\text{dilate}(I, k \times k)$ a morphological dilate operation over signal I with a square block $k \times k$ (a value of $k=3$ was used in our experiments). Practically, this means that every sub-band pixel on or around the desired input image boundary is included in the fused sub-band even if it would not otherwise be selected. Note that selection maps produced according to equations (3) and (4) are 1 for pixels that are to be kept from that particular input sub-band. The dilation operation followed by the OR operation also means that at some locations, both input sub-bands are to be considered as sources for the fused pyramid. In order to achieve this, the fused sub-band signal is finally produced from the input signals using a simple binary weighted sum, equation (5). In DWT fusion, this process is repeated for every pair of input sub-band signals, while in BMIF fusion it is performed only once, on the foreground images. Once sub-band (pyramid) fusion is completed, the respective image reconstruction processes produce the fused images.

$$F_{n,m} = S_{n,m}^A A_{n,m} + S_{n,m}^B B_{n,m} \quad (5)$$

4. RESULTS

The results of the proposed multi-level fusion system were evaluated through visual inspection (subjective) and objective evaluation performed using a slightly modified version of the objective image fusion performance metric described in ^{15,16}. The metric described by Xydeas and Petrović concentrates on the purest form of visual information preservation in the fused image and considers contrast amplification of input features as a distortion of the fusion system ^{15,16}. However, amplification does not necessarily represent distortion, particularly in fusion for visual display, as the original information already exists in the input images. A slightly modified version of the same metric that allows contrast amplification of existing input features in the fused image is therefore used to give a more realistic indication of performance of the proposed multi-level system in fusion for display. It should be noted that the modified metric is still capable of discerning between simple contrast amplification of the existing features and fusion artefacts (false information).

Fusion performance metric Q_{amp} , producing results in the range 0 (total failure) to 1 (ideal fusion), was used to evaluate performance of the proposed multi-scale fusion framework based on the DWT and BMIF fusion methodologies and their straightforward signal-level fusion realisations described in section 2. The measurements were performed on an input set comprising 166 different input image pairs, taken with a range of sensors including visual and IR range cameras, low-light instruments and hyperspectral scanners. Each fusion scheme fused each input pair, the fused images were evaluated and the scores recorded. Average Q_{amp} results over 166 input pairs, for each scheme are given in Table 1. The introduction of feature-level information into the fusion process makes a significant impact on the performance of both systems tested. In the case of DWT fusion, the score with XOR feature-level fusion increases from 0.6596 to 0.6796 and then on to 0.6827 for multi-level fusion using an OR combination of boundary maps. At the same time XOR based system performs better than the reference signal-level fusion system in 86% of the input pairs. This figure is even higher, 90% for the OR fusion based multi-level system.

| Fusion Scheme: | Signal-level fusion | Multi-level Fusion | |
|----------------|---------------------|--------------------|--------|
| | | XOR | OR |
| DWT | 0.6596 | 0.6796 | 0.6827 |
| BMIF | 0.5155 | 0.6723 | 0.6707 |

Table 1: Average Q_{amp} scores for different multi-level and signal level fusion schemes

BMIF based multi-level fusion exhibits and even bigger improvement, from $Q_{amp}=0.5155$ to staggering 0.6723, almost as good as the far more complex DWT fusion. This time however, XOR fusion outperforms OR (0.6707) fusion slightly. Multi-level BMIF fusion is better than direct signal-level fusion in 94% and 95% of the input pairs for XOR and OR feature-level fusion set-ups respectively.

The operation and performance of the presented multi-level fusion framework is best illustrated using a graphical example. Figure 4 shows fusion using multi-level and signal-level BMIF fusion. Input images shown along the top row, 4a and 4b result in exclusive edge maps, X_A and X_B , obtained using the procedure described in sections 3.1 and 3.2, 4c and 4d. These images illustrate well the exclusive nature of the information contained in the two input images. Also, the input images seem somewhat over segmented, however since the aim of the process is not to extract meaningful structure but merely guide the process of signal level fusion, this is even desirable.

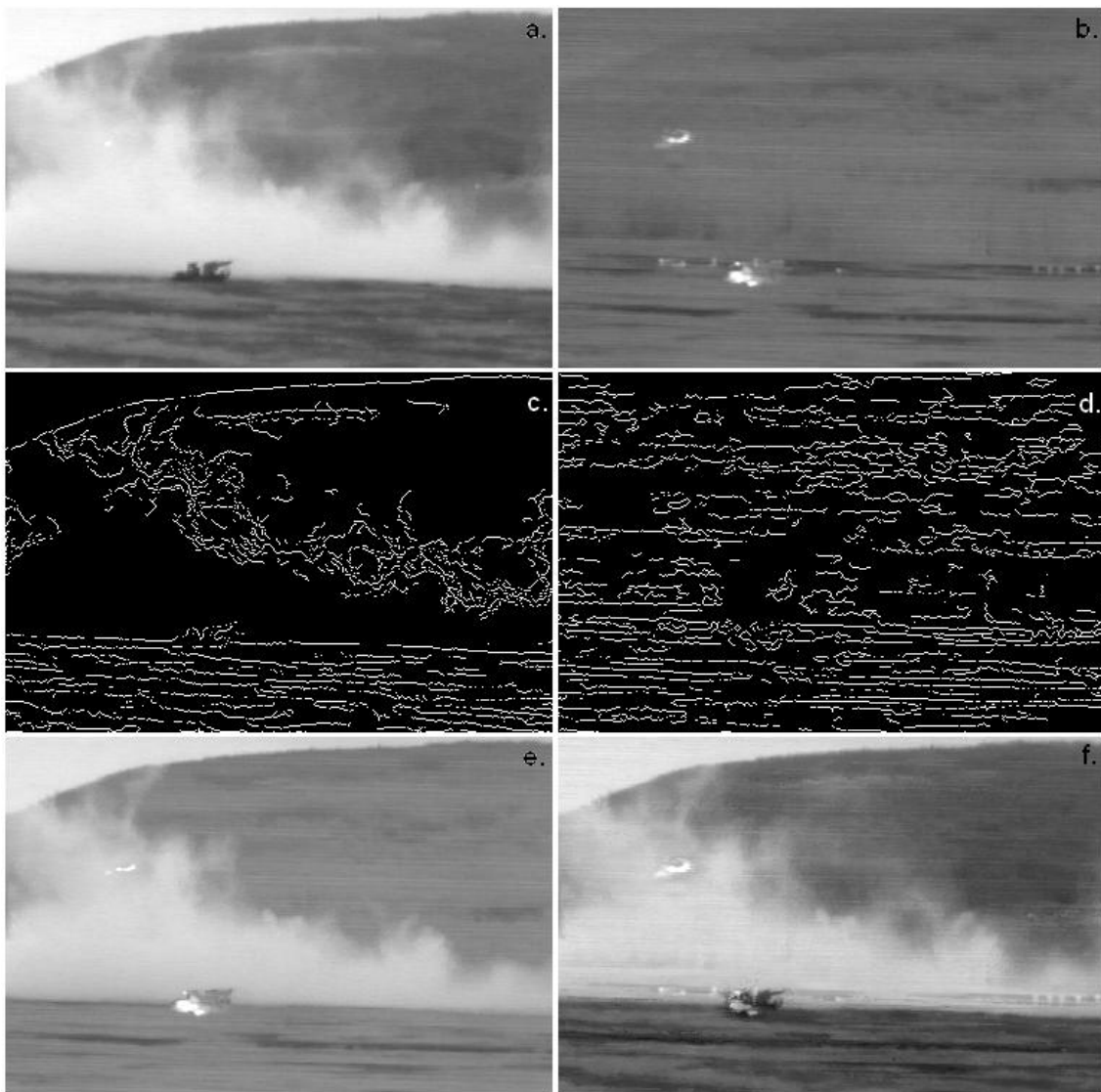


Figure 4: Input images a) and b), exclusive image boundaries c) and d) and fused images, using direct BMIF fusion e) and multi-level BMIF fusion implemented with XOR feature level fusion f)

The fused images produced with direct BMIF fusion and multi-level (XOR) BMIF fusion are displayed in 4e and 4f respectively. Inclusion of feature-level information into the sub-band fusion process obviously greatly increases the

reliability of the BMIF fusion scheme, as the Q_{amp} scores in Table 1 suggest. This is most obvious in the appearance of human figures in the smoke screen in the fused image 4f whereas they are not present in 4e. Even though direct BMIF fusion performs reasonably well with the other objects in the input images, its contrast and general clarity is inferior to that of the multi-level fused image.

Similarly, introduction of feature level information into DWT fusion also improves performance. A pair of input images shown in Figure 5a and 5b was fused with DWT signal-level and multi-level (XOR) fusion set-ups to produce fused images in 5c and 5d respectively. Once again multi-level fusion achieves a considerable improvement in clarity and general contrast of the fused image. A significant reduction in the presence of fusion artefacts (characteristic of DWT fusion) is also noticeable.



Figure 5: Input images a) and b), signal-level c) and multi-level d) DWT image fusion

5. CONCLUSIONS

This paper presented some early results of an investigation into a very promising multi-level approach to image fusion. The proposed framework is based on the synergy of information from different levels of abstraction, namely signal and

feature level, in order to achieve an improved performance of image fusion for display. Feature level information in the form of image boundary information is combined at feature level using basic fusion rules and fed-back to guide the process of signal-level image fusion. This introduction of subjectively meaningful structure into the signal level image fusion results in a significant improvement in reliability of the fusion process (increases robustness), reduces the loss of information and eliminates the presence of fusion artefacts. This is particularly significant in the case of a very simple and efficient bi-scale fusion method⁶ whose performance gains in reliability to the levels of far more complex fusion schemes.

However, the feature-level fusion rules used here are only some of the simplest of the numerous feature-level fusion tools available and further research should involve a thorough investigation of this particular toolbox. Additionally, good results of the BMIF fusion scheme in this type of framework indicate a promise of a reliable yet computationally efficient fusion image system.

ACKNOWLEDGEMENTS

This project was supported by Orasys Machine Vision, Belgrade, Serbia and Imaging Science Biomedical Engineering, University of Manchester, UK. The TNO Human Factors Research Institute of the Netherlands and Defence Research Establishment Valcartier of Canada are gratefully acknowledged for some of the imagery used in the project.

REFERENCES

1. D Hall, J Llinas ed., *Handbook of Multisensor Data Fusion*, CRC Press, 2001
2. D Piccione, W Krebs, P Warren, R Driggers, "Electro-optic Sensors to Aid Tower Air Traffic Controllers", Proceedings of the Human Factors and Ergonomics Society, 46th Annual Meeting, 2002, pp 51-55
3. A Toet, "Hierarchical Image Fusion", *Machine Vision and Applications*, Vol. 3 (1990), pp 3-11
4. H Li, B Munjanath, S Mitra, "Multisensor Image Fusion Using the Wavelet Transform", *Graphical Models and Image Processing*, Vol. 57(3), 1995, pp 235-245
5. T Peli, E Peli, K Ellis, R Stahl, "Multi-Spectral Image Fusion for Visual Display", *Proc. SPIE*, Vol. 3719, 1999, pp 359-368
6. V Petrović, C Xydeas, "Computationally Efficient Pixel-level Image Fusion", *Proceedings of Eurofusion99*, Stratford-upon-Avon, October 1999, pp177-184
7. G Qu, D Zhang, P Yan, "Medical image fusion by wavelet transform modulus maxima", *Optics Express*, Vol. 9, No. 4, 2001, pp 184-190
8. S Mukhopadhyay, B Chanda, "Fusion of 2D grayscale images using multiscale morphology", *Pattern Recognition*, Vol. 34 (2001), pp 1939-1049
9. Toet A, Schoumans N, Ijspeert J, "Perceptual Evaluation of Different Nighttime Imaging Modalities", *Proc. Fusion2000*, Paris, 2000, pp TuD3-17 – TuD3-23
10. Toet A, Ijspeert J, "Perceptual evaluation of different image fusion schemes", *Proc. SPIE*, 2001, pp 436-441
11. Steele P, Perconti P, "Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage", *Proc. SPIE*, Vol. 3062, 1997, pp 88-100
12. B Dasarthy, "Universal Fusion System Architecture", on belur.tripod.com, September 1999
13. Z Zhang, R Blum, "A Categorization of Multiscale-Decomposition-Based Image Fusion Schemes with a Performance Study for a Digital Camera Application", *Proceedings of the IEEE*, Vol. 87(8), 1999, pp1315-1326

14. D Marr, *Vision*, W.H.Freeman, San Francisco, 1982
15. C Xydeas, V Petrović, "Objective Image Fusion Performance Measure", *Electronic Letters*, Vol. 36, No.4, February 2000, pp 308-309
16. C Xydeas, V Petrović, "Objective Pixel-level Image Fusion Performance Measure", *Proc. of SPIE*, Vol. 4051, April 2000, pp 89-99
17. J Canny, "A Computational Approach to Edge Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 1986, pp 679-698
18. L Yiyao, Y Venkatesh, C Ko, "A knowledge-based neural network for fusing edge maps of multi-sensor images", *Information Fusion*, Vol. 2 (2001), pp 121-133