Region-Based Image Fusion Using Complex Wavelets

J. J. Lewis **R. J. O'Callaghan** **S. G. Nikolov** **D. R. Bull** **C. N. Canagarajah**
The Centre for Communications Research
University of Bristol
Bristol, BS8 1UB
UK
John.Lewis@bristol.ac.uk

Abstract – *This paper focuses on a novel region-based image fusion method which facilitates increased flexibility with the definition of a variety of fusion rules. A dual-tree complex wavelet transform (DT-CWT) is used to segment the features of the input images, either jointly or separately, to produce a region map. Characteristics of each region are calculated and a region-based approach is used to fuse the images, region-by-region, in the wavelet domain. This method gives results comparable to pixel-based fusion methods, but despite an increase in complexity, has a number of advantages over such methods. These include: the ability to use more intelligent semantic fusion rules; and for regions with certain properties to be attenuated or accentuated.*

Keywords: Image Fusion, Feature-Based Fusion, Region-Based Fusion, Segmentation.

1 Introduction

Image and video fusion is a specialisation of the more general topic of data fusion, dealing with image and video data [1]. There are a number of potential advantages of integrating the data from multiple sensors. These include [2]:

- Redundant information provided by a group of sensors can reduce overall uncertainty and increase accuracy of the integrated image;
- Complementary information from different sensors allows features in a scene to be perceived that would not be possible from individual sensors;
- More timely information is available as a group of sensors can collect information of a scene more quickly than a single sensor.

Image fusion is defined in [3] as the process by which several images, or some of their features are combined together to form a single image. The fusion process must satisfy the following requirements as described in [4]:

- Preserve all relevant information in the fused image;
- Suppress irrelevant parts of the image and noise;
- Minimise any artifacts or inconsistencies in the fused image.

Image fusion can be performed at four main levels [2]. These, sorted in ascending order of abstraction, are: signal; pixel; feature; and symbolic level. At pixel-level, images are combined by considering individual pixel values or small arbitrary regions of pixels in order to make the fusion decision. Pixel-based fusion is briefly discussed in Sec. 2, particularly focusing on fusion using complex wavelets.

One method of achieving feature-level fusion is with a region-based fusion scheme. An image is initially segmented in some way to produce a set of regions. Various properties of these regions can be calculated and used to determine which features from which images are included in the fused image. This has advantages over pixel-based methods as more intelligent semantic fusion rules can be considered based on actual features in the image, rather than on single or arbitrary groups of pixels. Sec. 3 gives an overview of our region fusion scheme.

The remainder of the paper is organised with results and discussion in Sec. 4 and the conclusion in Sec. 5. Throughout the paper, it is assumed that all input images are registered.

2 Review of Pixel Level Fusion

A number of pixel level fusion schemes exist. These range from simple averaging of the pixel values of registered images to more complex multi-resolution (MR) pyramid and wavelet methods (for example see [3, 3–12]). Transform fusion methods, defined in Eq. (1), generally involve transforming each of the registered input images $I_1, I_2, \ldots, I_N$ from normal image space into some other domain by applying an MR transform, ω. The transformed images are fused using some fusion rule, ϕ, and the fused image, F, reconstructed by performing the inverse transform, ω^{-1} [3].

$$F = \omega^{-1}(\phi(\omega(I_1), \omega(I_2), \ldots, \omega(I_N))) \qquad (1)$$

2.1 The Wavelet Transform

Wavelet transforms have been successfully used in many fusion schemes. A common wavelet analysis technique used for fusion is the discrete wavelet transform (DWT) [3, 7, 11, 13]. It has been found to have some advantages over pyramid schemes such as: increased directional information [7]; no blocking artifacts that often occur

in pyramid-fused images [7]; better signal-to-noise ratios than pyramid-based fusion [14]; improved perception over pyramid-based fused images, compared using human analysis [7, 14].

A major problem with the DWT is its shift variant nature caused by sub-sampling which occurs at each level. A small shift in the input signal results in a completely different distribution of energy between DWT coefficients at different scales [9]. A shift invariant DWT (SIDWT), described in [4], yields a very over-complete signal representation as there is no sub-sampling.

The Dual Tree Complex Wavelet Transform (DT-CWT) [9, 15, 16] is an over complete wavelet that provides both good shift invariance and directional selectivity over the DWT, although there is an increased memory and computational cost. Two fully decimated trees are produced, one for the odd samples and one for the even samples generated at the first level. The DT-CWT has reduced over completeness compared with the SIDWT, an increased directional sensitivity over the DWT and is able to distinguish between positive and negative orientations giving six distinct sub-bands at each level, the orientations of which are $\pm 15°, \pm 45°, \pm 75°$. The DT-CWT gives perfect reconstruction as the filters are chosen from a perfect reconstruction bi-orthogonal set [9]. It is applied to images by separable complex filtering in two dimensions. The bi-orthogonal Antonini and Q-shift filters used are described in [15]. The increased shift invariance and directional sensitivity mean that the DT-CWT gives improved fusion results over the DWT [10].

2.2 Pixel Fusion with Complex Wavelets

The pixel-level fusion scheme used here, employs the DT-CWT to obtain a MR decomposition of the input images. The wavelet coefficients are then combined, using a *maximum-selection* fusion rule to produce a single set of coefficients corresponding to the fused image. This process is shown in Fig. 1.

The maximum-selection scheme selects the largest absolute wavelet coefficient at each location from the input images as the coefficient at that location in the fused image. As wavelets tend to pick out the salient features of an image, this scheme works well producing good results. More complex fusion rules have been developed such as [7], where the maximum absolute value in an arbitrary window is used as an activity measure of the central coefficient in the window. In [6] a decision map is created based on the activity of an arbitrary block around a central coefficient as well as the similarity between the two areas of an image.

3 The Region Level Fusion Scheme

The majority of applications of a fusion scheme are interested in features within the image, not in the actual pixels. Therefore, it seems reasonable to incorporate feature information into the fusion process [17]. There are a number of perceived advantages of this, including:

- **Intelligent fusion rules**: Fusion rules are based on combining groups of pixels which form the regions of an image. Thus, more useful tests for choosing the regions of a fused image, based on various properties of a region, can be implemented;
- **Highlight features**: Regions with certain properties can be either accentuated or attenuated in the fused image depending on a variety of the region's characteristics;
- **Reduced Sensitivity to Noise**: Processing semantic regions rather than at individual pixels or arbitrary regions can help overcome some of the problems with pixel-fusion methods such as sensitivity to noise, blurring effects and mis-registration [17];
- **Registration and Video Fusion**: The feature information extracted from the input images, could be used to aid the registration of the images. Region-video fusion schemes could use motion estimation to track the fused features, allowing the majority of frames to be quickly predicted from some fully fused frames.

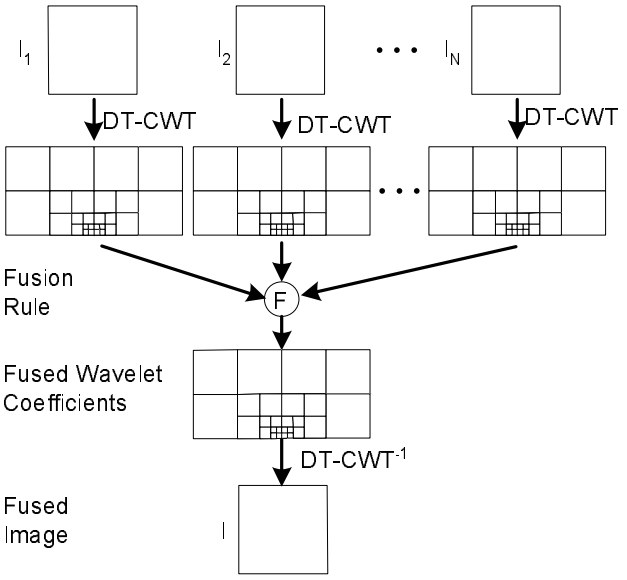


Fig. 1: The Pixel-Based Image Fusion Scheme Using the DT-CWT

A number of region-based fusion schemes have been proposed, for example, [6, 17–20]. These initially transform pre-registered images using an MR transform. Regions representing image features are then extracted from the transform coefficients. A grey-level clustering using a generalised pyramid linking method is used for segmentation in [6, 17, 20]. The regions are then fused based on a simple region property such as average activity. These methods do not take full advantage of the wealth of information that can be calculated for each region.

The novel region-based fusion scheme proposed in this paper is shown in Fig. 2. Initially, the N registered images $I_1, I_2, \ldots, I_N$ are transformed using ω, the DT-CWT.

$$[D_n, A_n] = \omega(I_n) \tag{2}$$

This gives a set of detail coefficients $D_{n,(\theta,l)}$ for each image I_n, consisting of a group of six different subbands,

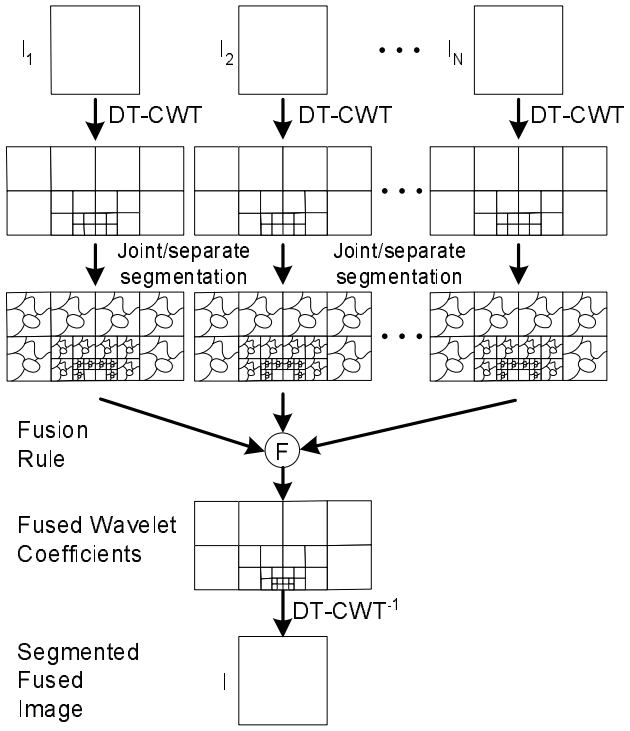


Fig. 2: The Region-Based Image Fusion Scheme Using the DT-CWT

θ, at each level of decomposition, l. A_n is the approximation of the image at the highest level. A combination of intensity information, I_n and textural information, D_n, is used to segment the images, either jointly or separately, with the segmentation algorithm ϕ, giving the segmentation maps: $S_1, S_2, \ldots, S_N$; or a list of all the T_n regions for each image: $R_1, R_2, \ldots, R_N$; where $R_n = \{r_{n,1}, r_{n,2}, \ldots, r_{n,T_n}\}$, $n \in N$. The map is down sampled by 2 to give a decimated segmentation map of the image at each level of the transform. When down sampling, priority is always given to a pixel from the smaller of the two regions. The segmentation algorithm is discussed in Sec. 3.1

A MR priority map,

$$P_n = \{p_{n,r_{n,1}}, p_{n,r_{n,2}}, \ldots, p_{n,r_{n,T_n}}\} \quad (3)$$

is then generated for each region, $r_{n,t}$ in each image, n based on any of a variety of tests. Calculating the priority is discussed further in Sec. 3.2. The priority generated for the input images shown in Figs. 5(a) and 5(b) are shown in Fig. 3. The priority of each region across all subbands and levels is the same. If priority is allowed to vary across subbands and levels, the features become blurred when the feature has a low energy in some of the subbands. This results in some coefficients which contribute to a feature in the original input image not being present in the fused image coefficients, causing distortion.

A feature, t_n, can be accentuated or attenuated by multiplying all of the wavelet coefficients corresponding to the region by a constant, w_{t_n}. A *weighting mask*, W_n, is created for each image, the same size as the wavelet coefficients, defining how much each coefficient should be multiplied by.

(a) For The IR Image (b) For The Visible Image

Fig. 3: Priorities Generated Using Entropy

Regions are then either chosen or discarded based on this priority and the fusion rule, ϕ, giving the wavelet coefficients of the fused image. A mask, M, is generated that specifies which image each region should come from. This is shown in Eq. (4). The mask is the same size as that of the wavelet coefficients of the fused image. An example of the mask for one subband at the second level is shown in Fig. 4. The algorithm always chooses the region with the maximum priority to determine which image each of the coefficients representing a region, t, should come from.

$$M_t = \phi(p_{1,t}, p_{2,t}, \ldots, p_{N,t}) \quad (4)$$

If $S_i \neq S_j$, a segmentation map, S_F, is created such that $S_F = S_1 \cup S_2 \cup \ldots \cup S_N$. Thus, where two regions $r_{i,p}$ and $r_{j,q}$ from image i and j overlap, both will be split into two regions, each with the same priority as the original.

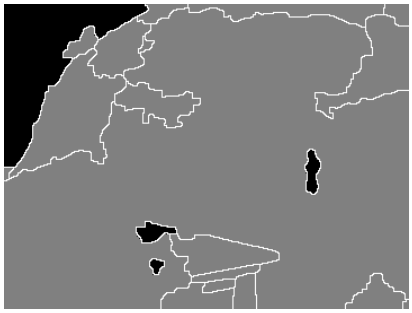


Fig. 4: The Image Mask: Black from IR, Gray from Visible

The fusion image is obtained by performing the inverse transform, ω^{-1}, on the fused weighted wavelet coefficients, $F = \omega^{-1}(D_F, A_F)$.

3.1 The Segmentation Algorithm

The quality of the segmentation algorithm is of vital importance to the fusion process. For the correct features to be present in the fused image, ideally, the segmentation algorithm should have the following properties:

- All required features are segmented as single separate regions. If a feature is missed, it may not be included in the fused image. If a feature is split into more than one region, each will be treated separately, possibly introducing artifacts into the fused image;

- As small a number of regions as possible should be created, as the time taken to compute the fused image increases with the number of regions.

An adapted version of the combined morphological-spectral unsupervised image segmentation algorithm is used, which is described in [21], enabling it to handle multi-modal images. The algorithm works in two stages. The first stage deals with both textured and non-textured regions in a meaningful fashion. The detail coefficients of the DT-CWT were used to process texture. The watershed transform of a combination of the perceptual gradient of the texture information and of the intensity information provides an initial segmentation. The second stage groups together primitive regions using a spectral clustering technique. A segmented IR and visible image is shown in Fig. 5.

The method can use either intensity information or textural information or both in order to obtain the segmentation map. This flexibility is useful for multi-modal fusion where some *a-priori* information of the sensor types is known. For example, IR images tend to lack textural information with most features having a similar intensity value throughout the region. Intuitively, an intensity only segmentation map should give better results than a texture-based segmentation.

Secondly, the segmentation can be performed either separately or jointly. For separate segmentation, each of the input images generates an independent segmentation map for each image.

$$S_1 = \sigma(I_1, D_1), \ldots, \; S_N = \sigma(I_N, D_N) \tag{5}$$

Alternatively, information from all images could be used to produce a joint segmentation map.

$$S_{joint} = \sigma(I_1 \ldots I_N, D_1 \ldots D_N) \tag{6}$$

A multiscale watershed segmentation algorithm is used to jointly segment multivalued images in [22]. In general, jointly segmented images work better for fusion. This is because the segmentation map will contain a minimum number of regions to represent all the features in the scene most efficiently. A problem can occur for separately segmented images, where different images have different features or features which appear as slightly different sizes in different modalities. Where regions partially overlap, if the overlapped region is incorrectly dealt with, artifacts will be introduced and the extra regions created to deal with the overlap will increase the time taken to fuse the images. However, if the information from the segmentation process is going to be used to register the images or if input images are completely different, it can be useful to separately segment the images. The effects of segmenting the images in different ways are shown in Fig. 5. In particular, the inefficient segmentation union of the two unimodal segmentation maps, which is necessary in order to fuse the images is show in 5(f).

3.2 Calculation of Priority and Fusion Rules

A measure of the average energy of the wavelet coefficients in a region is generally a good measure of the importance of a regions. In [6], a simple activity measure taking the absolute value of the wavelet coefficient is used. Thus, the priority of region r_{t_n} in image n, with size $|r_{t_n}|$ calculated with the detail coefficients $d_{n(\theta,l)}(x,y) \in D_{n(\theta,l)}$:

$$P(r_{t_n}) = \frac{1}{|r_{t_n}|} \sum_{\forall\theta,\forall l,(x,y)\in r_{t_n}} \left| d_{n(\theta,l)}(x,y) \right| \tag{7}$$

Alternatively, the variance of the wavelet coefficients over a region could be used as priority

$$P(r_{t_n}) = \frac{1}{|r_{t_n}|} \sum_{\forall\theta,\forall l,(x,y)\in r_{t_n}} (d_{n(\theta,l)}(x,y) - \bar{d}_{n(\theta,l)})^2 \tag{8}$$

where, $\bar{d}_{n(\theta,l)}$ is the average wavelet coefficient of the region r_{t_n}.

Thirdly, the entropy of the wavelet coefficients could be calculated and used as priority. The normalized Shannon entropy of a region is:

$$P(r_{t_n}) = \frac{1}{|r_{t_n}|} \sum_{\forall\theta,\forall l,(x,y)\in r_{t_n}} d^2_{n(\theta,l)}(x,y) \log d^2_{n(\theta,l)}(x,y) \tag{9}$$

with the convention $0 \log(0) = 0$

A variety of other methods could be used, possibly in combination with one or more of the above, to produce a priority map. The region's size, shape or position of the center of mass could also be used to contribute to the priority. For example, smaller regions could be given higher priorities than larger regions which are more likely to contain background information. Fig. 6 shows a Delaunay triangulation map. This shows how each of the regions are related to other surrounding regions in the image. This information could be used to generate priorities based on the position of two regions relative to each other. For example, in Fig. 6, the priority of the figure could be increased the closer the figure is to the house or road.

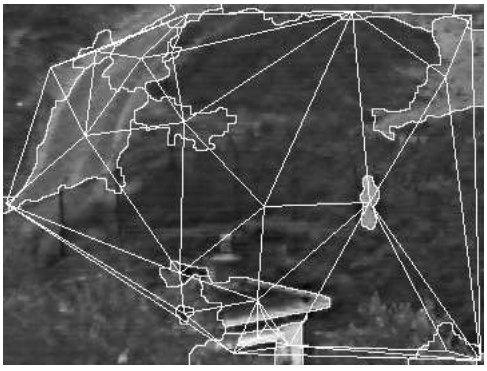


Fig. 6: The Fused Image with Segmentation Map and Triangulation Map Overlaid

Alternatively, this region information could be used to weight a region, increasing or decreasing its prominence in the image. A weighting system implemented to increase the weighting of a region dependent on its proximity to another region is described in Eq. (10). The weighting, w_i,

of a region, r_i of an image, with centre of mass at coordinates (x_i, y_i), is inversely proportional to the distance from a region r_j. The regions need not be from the same image. While the classification of regions as objects is beyond the scope of this research, a simple threshold can often be used to extract features such as people, animals and vehicles from infra-red images, for example, the figure from the IR image, Fig 5(a).

$$w_i \propto \frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \tag{10}$$

4 Results

Fig. 7 shows the fused result of the IR and visible images for both pixel and region-based methods. As can be observed, the region level scheme performs comparably with the pixel based fusion method. The input images have been jointly segmented and priority based on either entropy, variance or activity. Each of these methods have successfully chosen the majority of the background of the fused image from the visible image, while the important features in the IR image, namely the figure, have also been included. With these images, there appears to be very little differentiating the three methods of calculating priority.

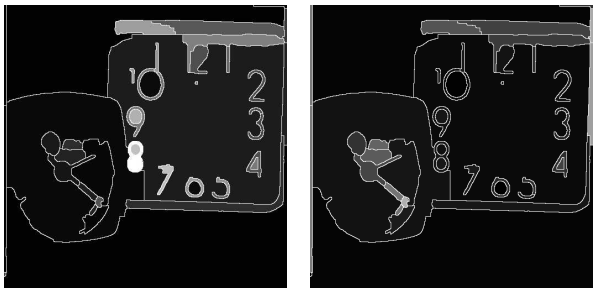
Fig. 7(e) shows the weighting mask generated by extracting all regions in the IR image above a certain threshold. This can be seen to successfully extract the feature representing the figure. The coefficients representing the figure are weighted by a factor 2.0, which accentuates the presence of the figure, as shown in Fig. 7(f).

Fig. 8 shows two multi-focus images. The clock images are jointly segmented and the priorities calculated using entropy are shown in the figure. The in focus parts of the image have higher priorities than the out of focus images, thus, the fused image shows both clocks in focus. The region-fused image is perceived slightly better than the pixel-fused image. The out of focus image is slightly larger than the in focus image. The pixel-fusion algorithm picks these out of focus pixels over some parts of the background, introducing artifacts.

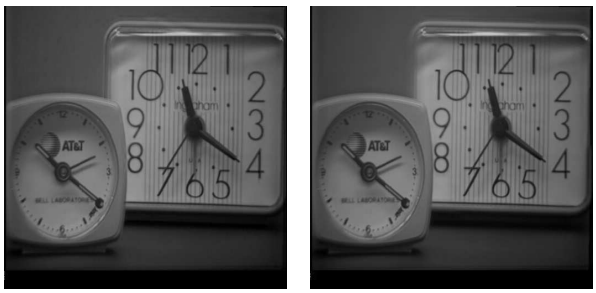
Fig. 9 shows three input images depicting some people by the sea, with a ship and buoy at sea. The joint segmentation map shown in this figure accurately segments the people on the beach as well as the boat and buoy. The pixel-fused image has some errors due to mis-registration of the input images, which are not visible in the region-fused image. However, some detail around the boat is missing in the region-fused image.

Fig. 10 shows images from a sequence of fused IR and visible images. The distance between the centre of masses of the region representing the road and the figure is calculated and the coefficients of the figure are weighted inversely proportional to the closeness to the road. For this experiment the road was manually selected and the figure detected by thresholding the IR image. The calculated distance (given in pixels) and the weighting applied to the figure is given for each image. These images are jointly segmented and fused using an entropy priority. The figure is seen to get brighter as he moves closer to the road. This

(a) Input Image A (b) Input Image B



(c) Entropy Priority Map for Clock A for the Jointly Segmented Input Images (d) Entropy Priority Map for Clock B for the Jointly Segmented Input Images



(e) Pixel-Based Fused Image (f) Region-Based Fused Image

Fig. 8: Fusion of Multifocus Clocks Image

task differs from that originally proposed in [23] for this image sequence, which was to determine the figure's position relative to the fence. However, this is a very difficult task and the task proposed is a worthwhile exercise showing how this algorithm could be applied.

These results were obtained with a Matlab implementation of the algorithm as part of an image fusion toolbox.

5 Conclusions

This paper has demonstrated that comparable results can be achieved with region and pixel-based methods. While region-based fusion methods are generally more complex, there are a number of advantages of such schemes over pixel-based fusion. The advantages demonstrated here include novel more intelligent fusion rules and the ability to treat regions differently depending on a variety of properties such as average activity, size and position relative to other regions in the scene. Future work should include a thorough objective and subjective testing of the effects of using either joint or separate multi-modal image segmentation and the effects of using only intensity or textural information or both on image fusion.

Acknowledgements

This work has been partially funded by the UK MOD Data and Information Fusion Defence Technology Center. The original IR and visible images are kindly supplied by Alexander Toet of the TNO Human Factors Research Institute and the multifocus clocks images by O. Rockinger. These images are available online at `www.imagefusion.org`.

References

[1] H. Maître and I. Bloch. Image fusion. *Vistas in Astronomy*, 41(3):329–335, 1997.

[2] M. Abidi and R. Gonzalez. *Data Fusion in Robotics and Machine Intelligence*. Academic Press, USA, 1992.

[3] S. Nikolov. Image fusion: A survey of methods, applications, systems and interfaces. Technical Report UoB-SYNERGY-TR02, University of Bristol, UK, August 1998.

[4] O. Rockinger. Pixel-level fusion of image sequences using wavelet frames. In *Proceedings of the 16th Leeds Applied Shape Research workshop*. Leeds university Press, 1996.

[5] A. Toet. Hierarchical image fusion. *Machine Vision and Applications*, 3:1–11, 1990.

[6] G. Piella. A general framework for multiresolution image fusion: from pixels to regions. *Information Fusion*, 4:259–280, 2003.

[7] H. Li, S. Manjunath, and S. Mitra. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3):235–245, 1995.

[8] O. Rockinger. Image sequence fusion using a shift-invariant wavelet transform. In *Proceedings of the IEEE International Conference on Image Processing*, volume III, pages 288–291, 1997.

[9] N. Kingsbury. The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters. In *IEEE Digital Signal Processing Workshop*, volume 86, 1998.

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

[11] P. Hill. *Wavelet Based Texture Analysis and Segmentation for Image Retrieval and Fusion*. PhD thesis, Department of Electrical and Electronic Engineering, University of Bristol, UK, 2002.

[12] Z. Zhang and R. Blum. A categorization and study of multiscale-decomposition-based image fusion schemes. *Proceedings of the IEEE*, pages 1315–1328, 1999.

[13] L. Chipman, T. Orr, and L. Graham. Wavelets and image fusion. In *Wavelet Applications in Signal and Image Processing III*, volume 2569, pages 208–219, 1995.

[14] T. Wilson, S. Rogers, and M. Kabrisky. Perceptual based hyperspectral image fusion using multi-spectral analysis. *Optical Engineering*, 34(11):3154–3164, 1995.

[15] N. Kingsbury. The dual-tree complex wavelet transform with improved orthogonality and symmetry properties. In *Proceedings IEEE Conference on Image Processing*, Vancouver, 2000.

[16] N. Kingsbury. Image processing with complex wavelets. In B. Silverman and J. Vassilicos, editors, *Wavelets: The key to intermittent Information*, pages 165–185. Oxford University Press, USA, 1999.

[17] G. Piella. A region-based multiresolution image fusion algorthim. In *ISIF Fusion 2002 conference*, Annapolis, July 2002.

[18] Z. Zhang and R. Blum. Region-based image fusion scheme for concealed weapon detection. In *Proceedings of the 31st Annual Conference on Information Sciences and Systems*, March 1997.

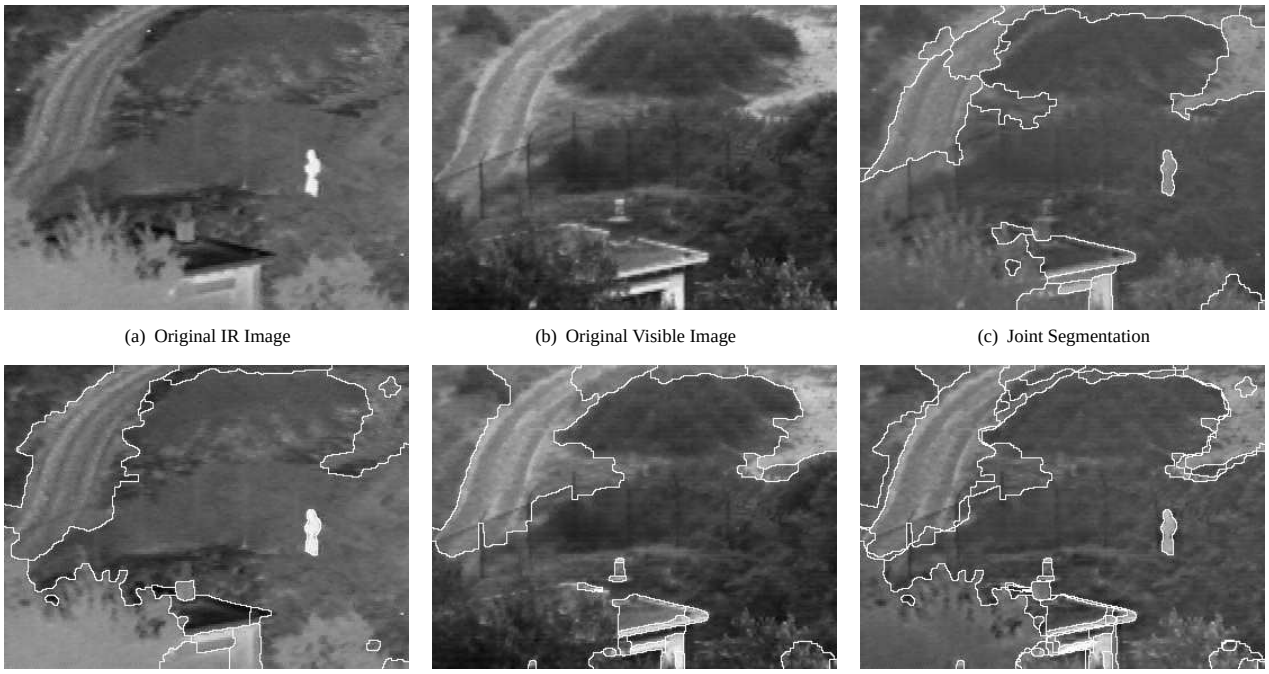
[19] B. Matuszewski, L.-K. Shark, and M. Varley. Region-based wavelet fusion of ultrasonic, radiographic and shearographyc non-destructive testing images. In *Proceedings of the 15th World Conference on Non-Destructive Testing*, Rome, October 2000.

[20] G. Piella and H. Heijmans. Multiresolution image fusion guided by a multimodal segmentation. In *Proceedings of Advanced Concepts of Intelligent Systems*, Ghent, Belgium, September 2002.

[21] R. O'Callaghan and D. Bull. Combined morphological-spectral unsupervised image segmentation. In Submission to The IEEE Transactions on Image Processing, 2003.

[22] P. Scheunders and J. Sijbers. Multiscale watershed segmentation of multivalued images. In *International Conference on Pattern Recognition*, Quebec, 2002.

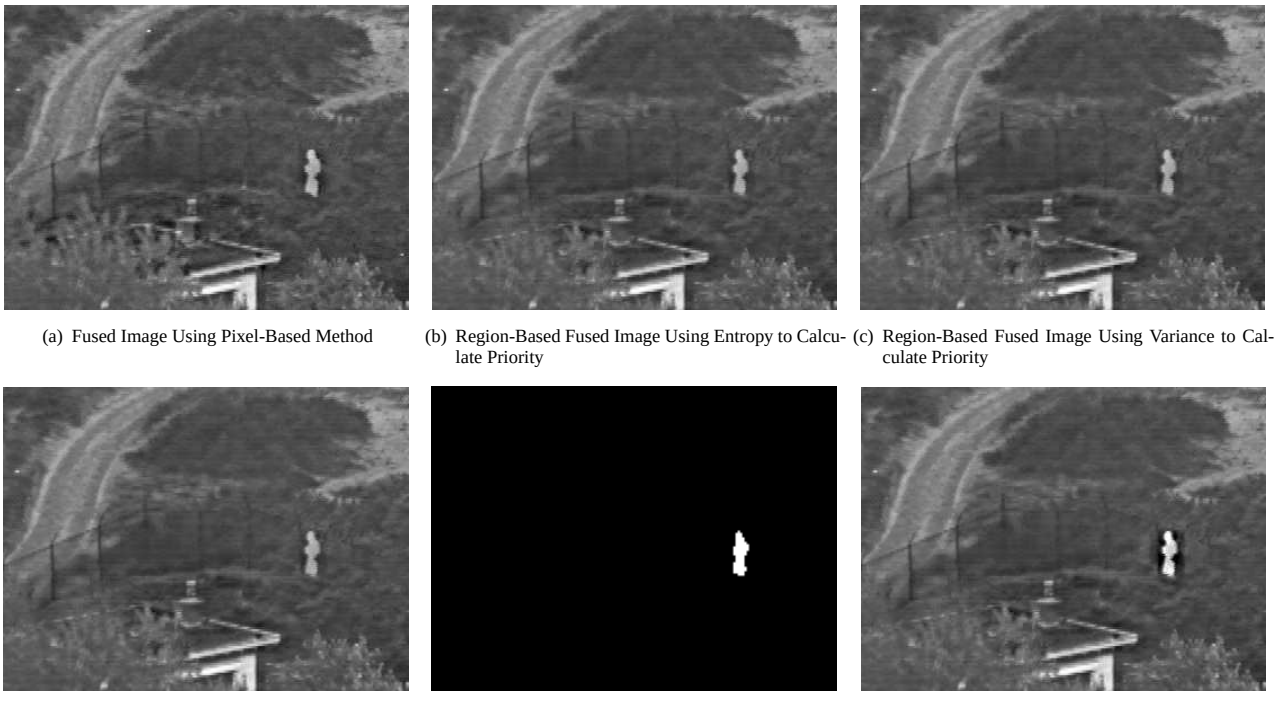
[23] A. Toet, J. IJspeert, A. Waxman, and M. Aguilar. Fusion of visible and thermal imagery improves situational awareness. *Displays*, 18:85–95, 1997.



(a) Original IR Image (b) Original Visible Image (c) Joint Segmentation

(d) Unimodal Segmentation of IR Image (e) Unimodal Segmentation of Visible Image (f) Union of Both Unimodal Segmentations

Fig. 5: Segmentation of IR and Visible Image



(a) Fused Image Using Pixel-Based Method (b) Region-Based Fused Image Using Entropy to Calculate Priority (c) Region-Based Fused Image Using Variance to Calculate Priority

(d) Region-Based Fused Image Using Activity to Calculate Priority (e) Mask Used as Multiplier on IR Coefficients (f) Region-Fused Image with Masked IR Coefficients Multiplied by a Factor of 2.0

Fig. 7: Fusion of IR and Visible Image

(a) 2-3μm Frequency Band IR Input Image

(b) 8-12μm Frequency Band IR Input Image

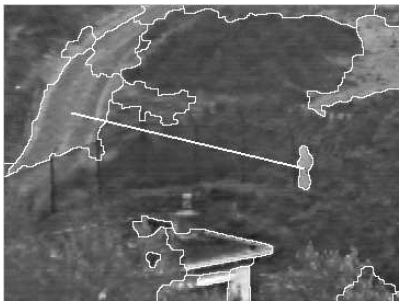
(c) Visible Range CCD Image

(d) Segmentation Map Generated by Jointly Segmenting All Input Images

(e) Pixel-Based Fused Image

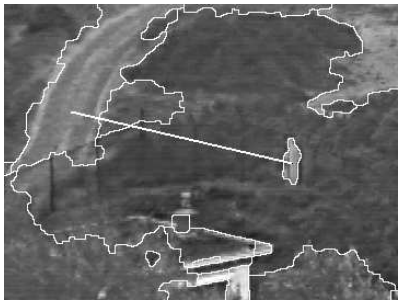
(f) Region-Based Fused Image

Fig. 9: Fusion of Two IR Images and a CCD Image

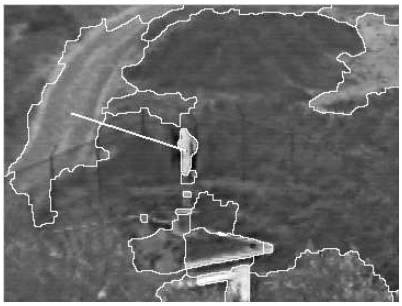


(a) Image No. 11. Dist = 220; Weighting = 1.10

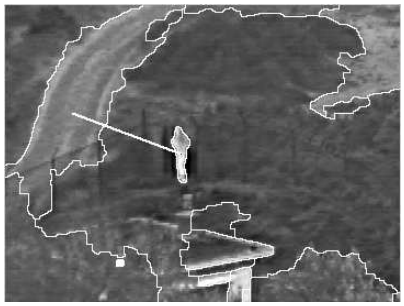
(b) Image No. 13. Dist = 213; Weighting = 1.13



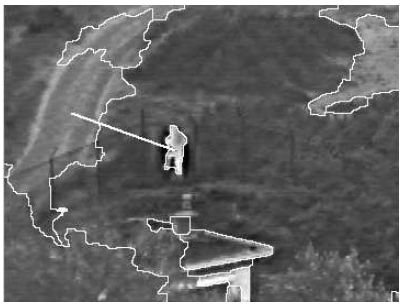
(c) Image No. 18. Dist = 207; Weighting = 1.16



(d) Image No. 19. Dist = 111; Weighting = 1.65



(e) Image No. 24. Dist = 105; Weighting = 1.67



(f) Image No. 26. Dist = 101; Weighting = 1.69

Fig. 10: Figure in IR Image Highlighted Depending on Closeness to the Road