

Hybrid Fused Displays: Between Pixel- and Region-Based Image Fusion

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

Abstract – *This paper describes a new class of fused image displays called hybrid fused displays. Such displays can be created by gradually transforming a fused image computed by one fusion method into another fused image computed by a different fusion method, either in the spatial domain or in another domain. By defining either global image-to-image or local region-to-region mapping(s) between the two key fused displays, sequences of hybrid fused displays can be generated. In this paper we propose a number of global and local inbetweening algorithms which transform pixel-based fused displays into various region-based fused displays. The main motivation behind this research is that while pixel-based fused displays generally provide better situational awareness, region-based fused displays, such as cut-and-paste displays or pseudo-coloured segmentation maps, often lead to better performance in tasks such as target detection and tracking. Having the ability to generate hybrid displays between these two classes is essential for being able to evaluate their perception, performance and usefulness for various tasks in different scenarios and applications.*

Keywords: Image Fusion, Segmentation, Pixel-Based Fusion, Region-Based Fusion, Hybrid Fused Displays, Inbetweening.

1 Introduction

Image fusion is the process by which several images coming from different sensors or modalities, or some of their features, are combined together to form a single fused image. Fused images are primarily used: (a) to be presented to human observers for viewing or interpretation; or (b) to be further processed by a computer using different image processing techniques. With a growing number and variety of image fusion algorithms, it is becoming increasingly important to be able to evaluate and compare them objectively in terms of both usefulness and performance. In the case where fused images are meant to be shown to a human for viewing or interpretation, the performance of a fusion algorithm must be measured as the degree to which it enhances the viewer’s ability to perform certain practical tasks [1]. The perception of the global structure of a scene to a large extent determines global scene recognition or situational awareness [2]. The perception of detail in a scene is relevant to tasks involving target detection and recognition, and visual search. In [2] the perception of global structure was tested by measuring the accuracy with which observers were able to distinguish a scene that was

presented right side up from one that was presented upside down, and perceive the position of the horizon. In the same paper the capability to discriminate fine detail was tested by extracting patches that display different objects of interest, e.g. building, vehicles, roads, humans, etc. and asking the observer to judge whether the presented image contains an exemplar of one of these object categories.

A few perceptual evaluation studies [1–3] have compared human performance for specific tasks, e.g. detection and localisation, for a number of fused displays, including pyramid fusion and colour fusion. In all of these human performance for such tasks was studied for key fused displays as well as for the input displays. Different fused displays, however, provide different benefits, e.g. pixel-based fused displays usually are good for situational awareness, while coloured segmented displays often lead to improved target detection and tracking, probably due to the improved speed and accuracy of information uptake [4] and the fewer fixations they require to locate colour coded targets [5]. Thus, one could argue that in different applications and scenarios, hybrid fused displays which are a mixture of some key fused displays may prove to be optimal. For complex tasks, such hybrid displays may provide better overall results in terms of human performance when compared to the key fused displays from which they are constructed. Hybrid displays are also invaluable tools in perceptual experiments that try to assess different fusion methods because they allow gradual transition between different fused images. The main goal of this paper is to propose techniques for the creation of such hybrid fused displays from pixel- and region-based fused images.

2 Pixel-Based Image Fusion

There are a number of pixel-based fusion schemes ranging from simply averaging the pixel values of registered images to more complex multi-resolution (MR) methods such as pyramid methods (for example see [6, 7]) or wavelet methods [8–10]. A useful review of MR fusion schemes is given in [11]. MR pixel-based fusion methods, shown in Eq. (1), generally involve transforming each of the registered input

images I_1, I_2, \dots, 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 is reconstructed by performing the inverse transform, ω^{-1} .

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

2.1 Image Fusion in the Wavelet Domain

A common wavelet family used for fusion is the discrete wavelet transform (DWT). The DWT has been used, for example, in [8] and has been found to have some advantages over pyramid schemes. A major problem with the DWT is its shift variant nature caused by the 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 [12]. A shift-invariant DWT (SIDWT) described in [9] is an extension of the DWT that yields an over-complete signal representation.

2.2 Pixel-Based Fusion Using DT-CWT

The Dual Tree Complex Wavelet Transform (DT-CWT) developed by Kingsbury [12] is an over complete wavelet transform that provides both good shift invariance and directional selectivity over the DWT, although, there is also an increased memory and computational cost compared with the DWT. The DT-CWT has reduced over completeness compared with the SIDWT. Effectively, the down sampling by 2 that occurs after the first level of the DWT is removed. Thus, 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 increased directional sensitivity over the DWT and is able to distinguish between positive and negative orientations giving six distinct subbands at each level, the orientations of which are $\pm 15^\circ, \pm 45^\circ, \pm 75^\circ$. The increased shift invariance and directional sensitivity of the DT-CWT lead to improved fusion results [10] over the DWT. The pixel-level fusion scheme used in this paper, employs the DT-CWT to obtain a MR decomposition of the input images. The wavelet coefficients are then combined, using some fusion rule or rules, to produce a single set of coefficients corresponding to the fused image. This process is shown in Fig. 1. A variety of fusion rules can be used to combine the wavelet coefficients of the input images. A simple and effective *choose maximum* scheme, where each of the wavelet coefficients is considered in turn is frequently used in image fusion. As wavelets tend to pick out the salient features of an image, e.g. edges and boundaries, the choose maximum scheme produces good results. This fusion rule is used to generate all pixel-based fused displays shown in the paper.

3 Region-Based Image Fusion

A number of region-based fusion schemes have been proposed in the past, see for example [7, 13–16]. The

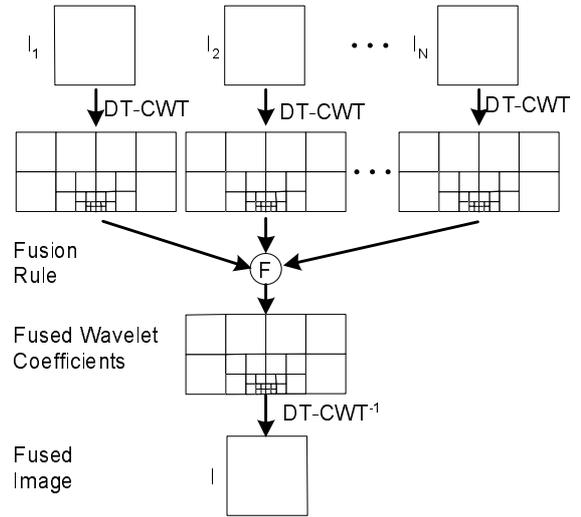


Fig. 1: Pixel-Based Image Fusion Using DT-CWT.

majority of these schemes follow a general pattern of initially transforming the pre-registered images using a multi-resolutional transform such as pyramids or wavelets. Regions representing image features are then extracted from the transform coefficients. The regions are then fused using some fusion rules based on the properties of a region such as its average energy. A new region-based image fusion method using a DT-CWT is proposed in [17]. Here we will briefly summarise the way this algorithm works.

3.1 Segmentation of Multi-Modality Images

The quality of the segmentation algorithm is of vital importance to the fusion process. The segmentation method used in this paper is an adapted combined morphological-spectral unsupervised image segmentation algorithm, which is described in [18]. The algorithm works in two stages. The first stage deals with both textured and non-textured features in a meaningful fashion. Texture information is extracted from the coefficients of the DT-CWT. This information is used to produce a perceptual gradient function whose watershed transform provides an initial segmentation. The second stage groups together primitive regions using a spectral clustering technique. Example surveillance IR and visible images are shown in Fig. 2 (a) and (b). The same images are pseudo-coloured in red and green, respectively, in Fig. 6 (a) and (b). Our 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. Secondly, the segmentation can be performed in two ways, either separately or jointly. For separate segmentation, each of the input images generates an independent segmentation map for each image.

$$S_1 = \sigma(I_1), \dots, S_n = \sigma(I_N) \quad (2)$$

The unimodal segmentation of the images in Fig. 6 (a) and (b) are shown in Fig. 6 (d) and (e). These segmentations

are displayed in colour by using the average intensity of each region as label and assigning the result to the red and green planes, respectively. Alternatively, all images could be segmented at the same time, such that a single *joint segmentation map* is produced for all of the input images.

$$S_{joint} = \sigma(I_1 \dots I_N) \quad (3)$$

The joint segmentation of the images in Fig. 6 (a) and (b) is shown in Fig. 6 (c). This kind of pseudo-colouring helps a lot in interpreting the fused image because it shows the contribution of each input modality. 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 all of the input images most efficiently. Compare the joint segmentation in 6 (c) with the union of unimodal segmentations in Fig. 6 (f). Particularly where different images have different regions or regions which appear as slightly different sizes in different modalities. This leads to problems dealing with the situation where regions partially overlap which could introduce artefacts if the overlapped region is incorrectly dealt with and will increase the time taken to fuse the images.

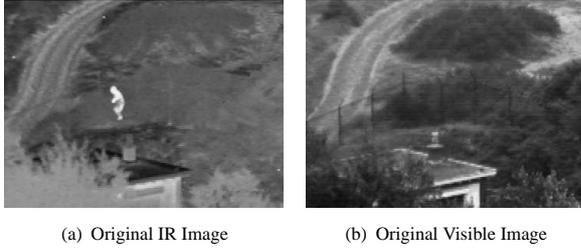


Fig. 2: Original IR and Visible Images Courtesy of Alexander Toet, TNO Human Factors Institute, The Netherlands.

3.2 Region-Based Fusion Using DT-CWT

The region-based fusion scheme proposed in [17] and used in this paper is carried out in a number of steps, as shown in Fig. 3. Initially, the N registered images I_1, I_2, \dots, I_N are transformed using ω , the DT-CWT.

$$[D_n, A_n] = \omega(I_n) \quad (4)$$

This gives a set of detail coefficients $D_{i,(\theta,l)}$ for each image $1 \dots N$, consisting of a group of six different subbands, θ at each level of decomposition, l . A_n is the approximation of the image I_n at the highest level. D_n is used to segment the images either jointly or separately, with the segmentation algorithm σ (already discussed in Sec. 3.1), giving the segmentation maps: S_{joint} or S_1, S_2, \dots, S_N ; or a list of all the T_n regions for each image: R_1, R_2, \dots, R_N ; where $R_n = \{r_{n,1}, r_{n,2}, \dots, r_{n,T_n}\}$, $n \in N$. This is down sampled by 2 to give a decimated segmentation map, for each level of the transform. A MR priority map, $P_n = \{p_{n,r_{n,1}}, p_{n,r_{n,2}}, \dots, p_{n,r_{n,T_n}}\}$, is then generated

for each region, $r_{n,t}$ in each image, n based on any of a variety of tests, such as entropy, variance, activity or a more abstract measure such as the proximity of two regions, the size of a region or a combination of tests. Regions are then either chosen or discarded based on this priority and the fusion rule, ϕ , giving the wavelet coefficients of the fused image. This is achieved by creating a mask, M which specifies which image each region should come from. A weighting mask can be used to accentuate or attenuate different regions. Finally, the fused image F is obtained by performing the inverse transform, ω^{-1} , on the fused weighted wavelet coefficients, $F = \omega^{-1}(D_F, A_F)$. For more details about this algorithm see [17].

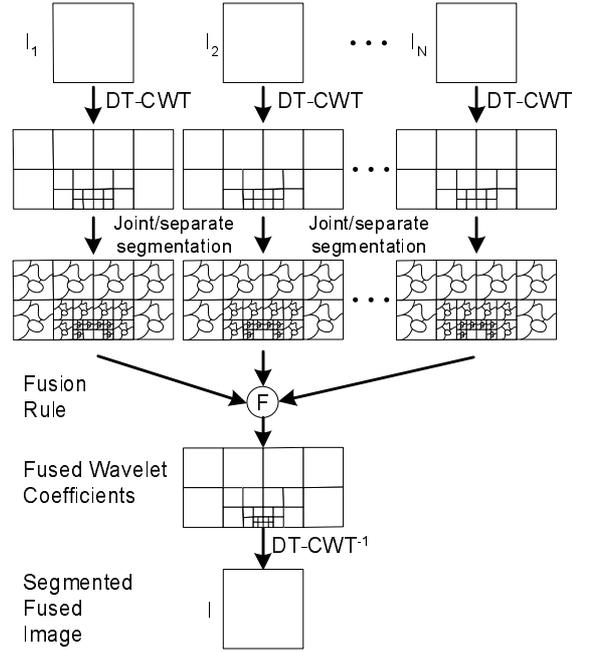


Fig. 3: Region-Based Image Fusion Using DT-CWT.

4 Comparison of Region- and Pixel-Based Fused Displays

There are a number of advantages and disadvantages of a region-based image fusion approach as compared to pixel-based image fusion:

- **Region-based fusion rules:** Rather than treating an image as a set of arbitrary pixels when fusing the image, rules are based on image regions;
- **Highlight regions:** Regions with certain features can be either accentuated or attenuated in the fused image;
- **Reduced Sensitivity to Noise:** Looking at semantic regions rather than at individual pixels can help overcome some of the problems with pixel-based fusion methods such as sensitivity to noise, blurring effects and mis-registration [15];

- **Poorer context:** Pixel-based fusion methods provide better contextual information and global scene recognition than region-based fusion methods;
- **Higher complexity:** Usually pixel-based fusion methods have lower computational complexity than region-based fusion methods.

5 Hybrid Fused Displays

If we have two different fused images or displays generated by two different fusion methods we may want to create a series of hybrid fused displays that are some kind of combination of the two 'key' fused displays. Such inbetweening techniques have been used extensively in the animation and film industries to create intermediate frames or images. Two such techniques that are commonly applied by animators are inbetweening and morphing. *Inbetweening* is the generation of intermediate transition positions from given start and end points or keyframes [19]. This technique is often used in animation, where a lead artist generates the beginning and end keyframes of a sequence (typically 1 second apart), a breakdown artist does the breakdowns (typically 4 frames apart), and an 'inbetweener' completes the rest. *Morphing* is a continuous deformation from one keyframe or 3-D model to another [19]. In 2-D it is generally performed by either distortion or deformation. In 3-D this is often achieved by approximating a surface with a triangular mesh that can then be continuously deformed. Creating a morph between two keyframes requires selection of corresponding key points, which makes it similar to image registration. Both techniques create inbetween frames or displays, thus gradually transforming an object in the first keyframe into the same or other object in the second keyframe. The construction of hybrid fused displays takes a similar approach, although in the case of registered images and joint segmentation the shape (boundary) of each corresponding region is the same in the two 'key' displays and only the interior of the regions changes gradually from one display to the other. In the context of this research we will use the term 'inbetweening' for the creation of a series of hybrid displays from two key fused images/displays. Here we consider two different types of hybrid fused displays: (a) those which use a global image-to-image mapping from one key display to the other; and (b) those which use a number of local region-to-region mappings. Several global and local inbetweening techniques are described in detail in the rest of this section.

5.1 Global Inbetweening

Let us have two fused displays F_1 and F_2 created by two different fusion methods from the registered input images I_1 and I_2 . We can define a global image-based transformation $T(F_1, F_2)$ which will gradually change the first key display F_1 into F_2 . This process will be called *global inbetweening*. Some possible transitions of this kind are illustrated in Fig. 4 where a labelled

joint segmentation map (Fig. 4 top-left), a cut-and-paste display (Fig. 4 middle-left) and a region-based fused display (Fig. 4 bottom-left) are transformed into a pixel-based fused display (Fig. 4 right) of the same IR and visible images. The transformation can be done in both directions depending on the type of hybrid displays we want to create. Three different hybrid displays computed by global inbetweening are shown in Figs. 7, 8 and 9.

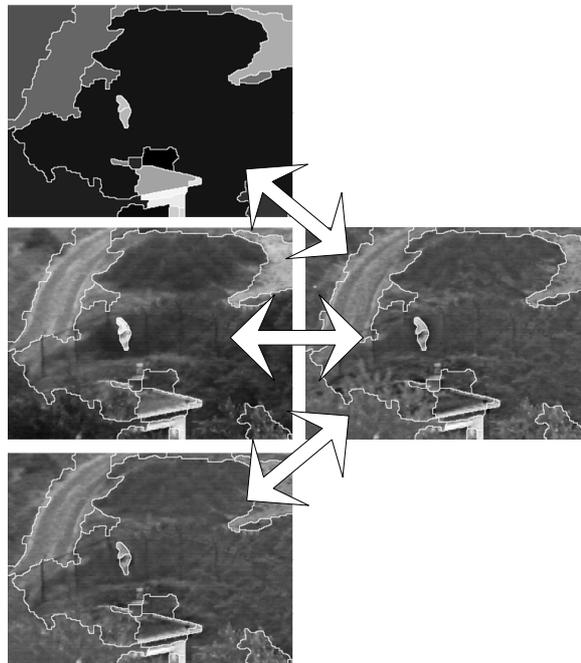


Fig. 4: Global Inbetweening.

The first example in Fig. 7 illustrates a transition between a pseudo-coloured joint segmentation map of the two input images (Fig. 7 (a)), where only two important regions are coloured (the 'road' and the human 'figure'), and a pixel-based fused display of the same input image computed by a maximum fusion rule in the wavelet domain using DT-CWT (Fig. 7 (f)). The hybrid displays are weighted averages of the two key displays treated as images, i.e. $H^i = w_1^i F_1 + w_2^i F_2$, where $w_2^i = (1 - w_1^i)$ and $i = 1, \dots, 6$. The weights (w_1^i, w_2^i) are given below for each hybrid display. In this gradual transition between a coloured segmentation map and a pixel-based fused display each consecutive hybrid display provides increased context to the 'figure' and 'road' regions. Thus more and more intensity and texture is 'injected' in each hybrid display. If we consider the opposite mapping, i.e. from a pixel-based fused display to a coloured segmentation map, this can be regarded as a process where some regions are progressively highlighted while the contextual information is decreased.

In the second example (see Fig. 8) the same transition is shown but this time the hybrid displays H^i are computed in the wavelet domain. After a DT-CWT is applied to the input images: (a) a joint segmentation map S_{joint} is calculated (only the two important regions are

preserved from it - the 'road' and the 'figure'); and (b) a pixel-based maximum fusion rule is applied to calculate a fused wavelet coefficient map. Then the low-pass wavelet coefficients A_n and the detail wavelet coefficients D_n of (a) and (b) are averaged using weighted averaging, i.e. $A_n^i = w_1^i A_{n,1} + w_2^i A_{n,2}$ and $D_n^i = w_1^i D_{n,1} + w_2^i D_{n,2}$, where $w_2^i = (1 - w_1^i)$ and $i = 1, \dots, 6$. An inverse DT-CWT is applied to compute the hybrid displays H^i . Note that the segmentation map in Fig. 8 (a) is a grey-scale analogue of the coloured segmentation map in Fig. 7 (a). Similarly to the previous figure, these hybrid displays show the process of gradually increasing the contextual intensity and texture information. Thus, the boundaries of the highlighted regions slowly disappear (Fig. 8 (d-f)).

Another example of a hybrid display using global inbetweening is given in Fig. 9. Here the mapping is between a 'cut-and-paste' fused display and a pixel-based fused display. The cut-and-paste display is a kind of region-based fused display where after segmentation and priority map calculation regions are selected from the input images and are cut and pasted to form the fused display. Again as in Fig. 7 the hybrid displays H^i are computed in the spatial domain as weighted averages of the key displays. These hybrid displays show a compromise between easier localisation and tracking of the figure, provided by the cut-and-paste display, and better scene structure integration and perception, provided by the pixel-based display.

5.2 Local Inbetweening

Let us again have two fused displays F_1 and F_2 created by two different fusion methods from the registered input images I_1 and I_2 . If one of the two key fused displays is a region-based one, i.e. the fused image is segmented in regions r_n , then another type of hybrid displays can also be constructed in which one or several regions r_n from F_1 can be mapped (gradually transformed) to their corresponding regions in F_2 , i.e. we can define a number of local transforms $t_k(r_{k,1}, r_{k,2})$. Such a process will be called *local region-to-region inbetweening* and is illustrated in Fig. 5.

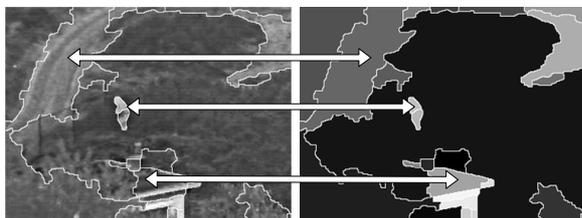


Fig. 5: Local Inbetweening.

Example hybrid fused displays computed using local inbetweening are shown in Fig. 10. Here two regions (the 'road' and the 'figure') from a pixel-based fused display are transformed into the corresponding regions of a pseudo-coloured segmentation map display. The local

transformation is weighted averaging of the regions with two different functions defined for each of the region pairs, each being of the type $r^i = w_{m,1}^i r_{m,1} + w_{m,2}^i r_{m,2}$, where $w_{m,2}^i = (1 - w_{m,1}^i)$, $i = 1, \dots, 6$ and $m = 1, 2$. The weights ($w_{m,1}^i, w_{m,2}^i$) are given below each hybrid display, where $m = 1$ corresponds to the 'road' mapping and $m = 2$ corresponds to the 'figure' mapping. The main advantage of using local inbetweening over global inbetweening is that one could create hybrid fused displays where different regions in one key display transform to their corresponding regions in the other key display in different ways. In the example in Fig. 10 the transition for the 'figure' is done much more quickly than for the 'road'. This kind of hybrid displays give us the option to highlight different regions with varying degree and to provide different levels of contextual information in different parts/regions of the display.

5.3 A Tool to Create Hybrid Fused Displays

The algorithms used to create the hybrid fused displays described in this paper are implemented as part of a Image Fusion toolbox written in Matlab. The toolbox comprises numerous image fusion methods, e.g. spatial fusion, PCA fusion, pyramid and wavelet fusion. Unimodal and joint segmentation of the input images can also be done with the toolbox using a variety of segmentation algorithms. The user can then generate a number of hybrid displays by selecting the two 'key' fused displays, e.g. pixel-based and region-based, and then specifying the mapping between them, i.e. global or local, and its parameters. The hybrid displays are then either shown on the screen or saved to disk as a sequence of images in standard bitmap format.

6 Conclusions

Several new algorithms for creation of hybrid fused displays have been proposed in this paper. Hybrid displays between pixel-based fused images and various region-based fused images were generated by means of global (image-to-image) and local (region-to-region) inbetweening. Hybrid fused displays were constructed in both spatial and wavelet domains. These displays exhibit various degrees of detail, texture and colour and combine some of the benefits provided by different image fusion approaches. They also allow gradual transition between different fused images which can be very useful in perceptual experiments that try to assess fusion techniques.

Acknowledgements

This work has been partially funded by the UK MOD Data and Information Fusion Defence Technology Centre. The original visible and IR 'UN camp' image sequences are kindly supplied by Alexander Toet of the TNO Human Factors Research Institute, The Netherlands. These images are available online at www.imagefusion.org.

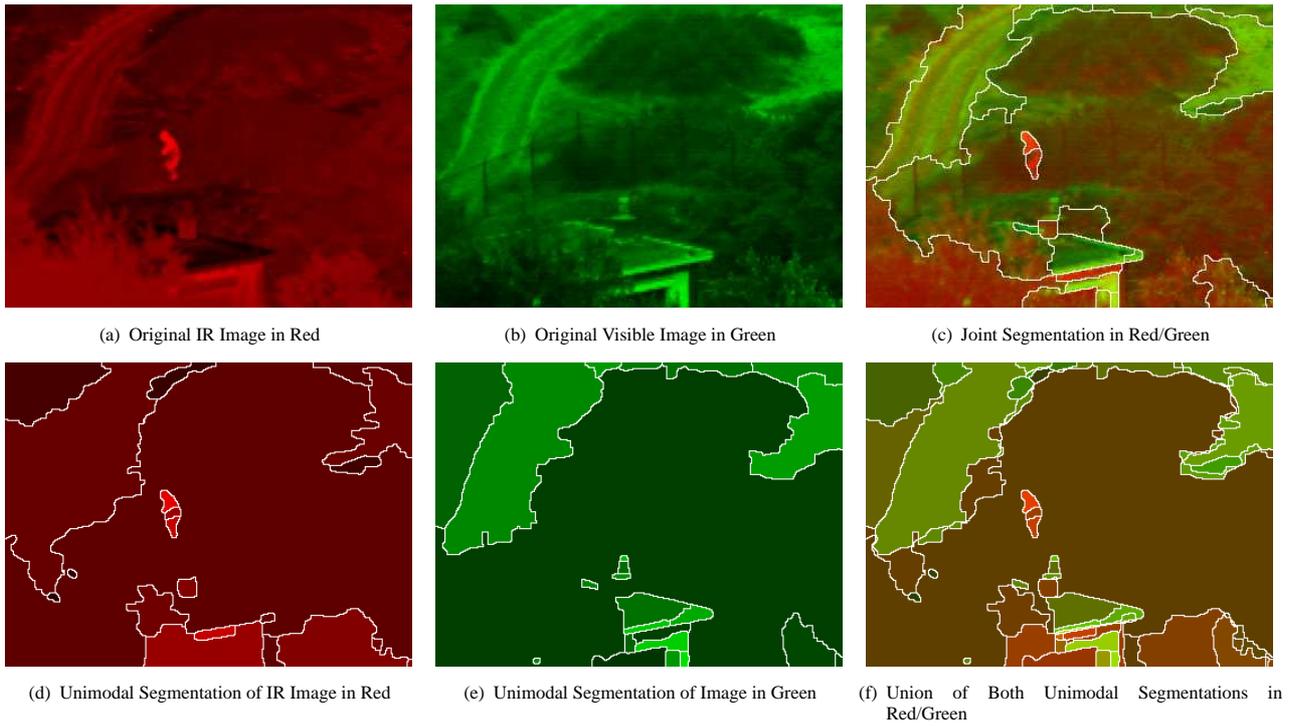


Fig. 6: Unimodal and Joint Segmentation of IR and Visible Images.

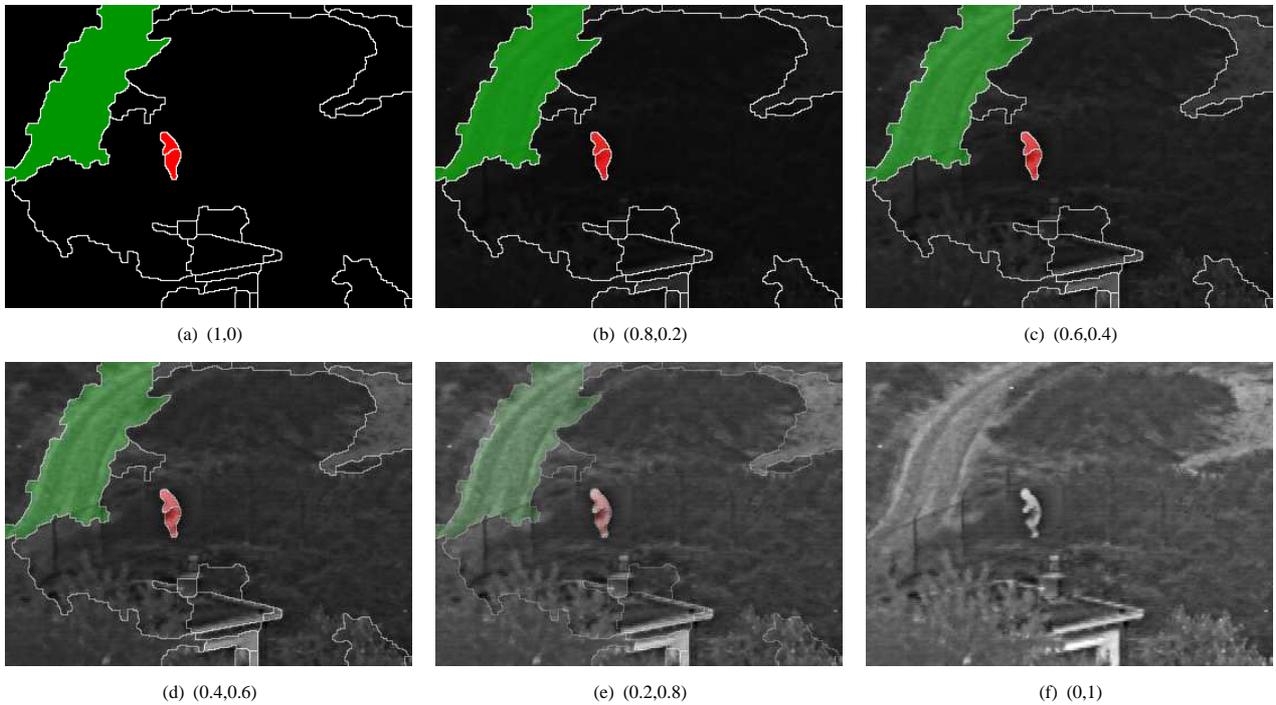


Fig. 7: Global Inbetweening in the Spatial Domain: from Coloured Segmentation Map to Pixel-Based Fused Display. The weights (w_1^i, w_2^i) are given below for each hybrid display.

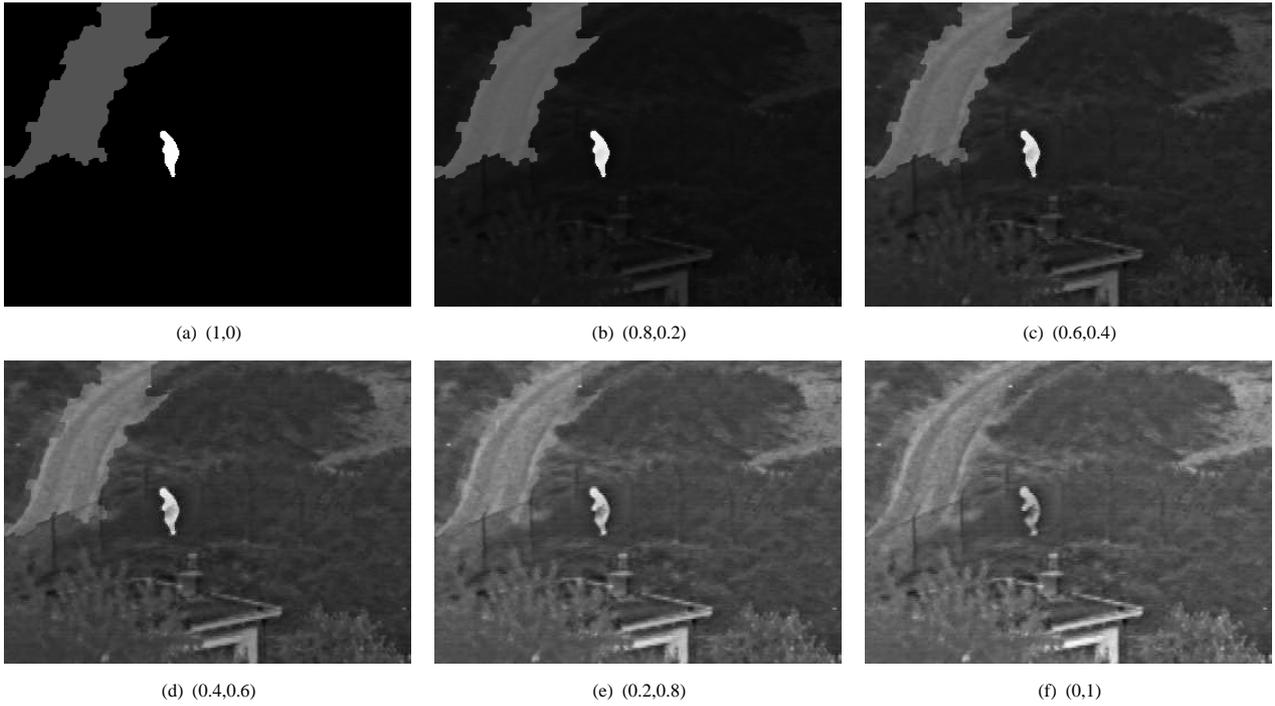


Fig. 8: Global Inbetweening in the Wavelet Domain: from Grey-Scale Segmentation Map to Pixel-Based Fused Display. The weights (w_1^i, w_2^i) are given below for each hybrid display.

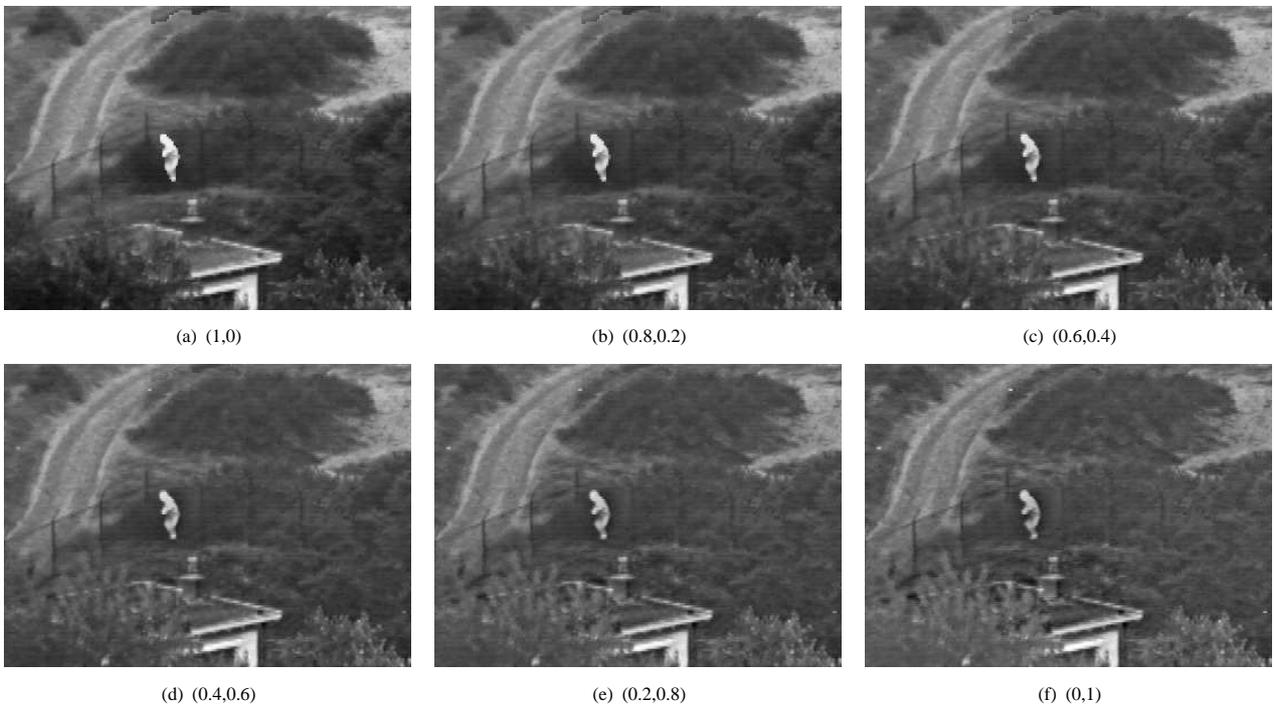


Fig. 9: Global Inbetweening in the Spatial Domain: from Cut-and-Paste Display to Pixel-Based Fused Display. The weights (w_1^i, w_2^i) are given below for each hybrid display.

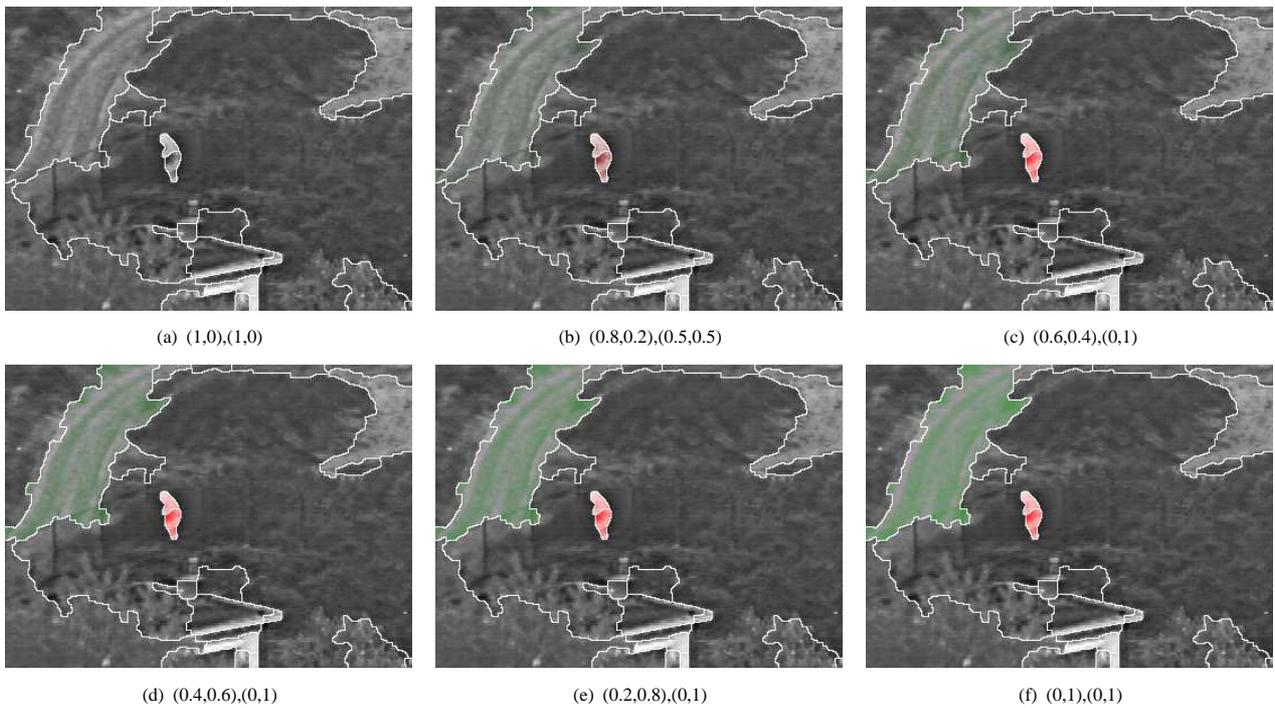


Fig. 10: Local Inbetweening in the Spatial Domain: from Pixel-Based Fused Display to Coloured Segmentation Map with Different Tempo of Change for 'Road' and 'Figure' Regions. The weights $(w_{m,1}^i, w_{m,2}^i)$ are given below each hybrid display, where $m = 1$ corresponds to the 'road' mapping and $m = 2$ corresponds to the 'figure' mapping.

References

- [1] A. Toet, J. K. IJspeert, A. M. Waxman, and M. Aguilar. Fusion of visible and thermal imagery improves situational awareness. *Displays*, 24:85–95, 1997.
- [2] A. Toet and E. M. Franken. Perceptual evaluation of different image fusion schemes. *Displays*, 24:25–37, 2003.
- [3] A. Toet, N. Schoumans, and J. K. IJspeert. Perceptual evaluation of different nighttime imaging modalities. In *3rd International Conference on Information Fusion (Fusion 2000)*, Paris, France, 10-13 July, volume I, pages TuD3–17. International Society of Information Fusion (ISIF), 2000.
- [4] R. E. Christ. Review and analysis of colour coding for visual displays. *Human Factors*, 17:542–570, 1975.
- [5] P. K. Hughes and D. J. Creed. Eye movement behaviour viewing colour-coded and monochrome avionic displays. *Ergonomics*, 37:1871–1884, 1994.
- [6] A. Toet. Hierarchical image fusion. *Machine Vision and Applications*, 3:1–11, 1990.
- [7] G. Piella. A general framework for multiresolution image fusion: from pixels to regions. *Information Fusion*, 4:259–280, 2003.
- [8] H. Li, S. Manjunath, and S. Mitra. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3):235–245, 1995.
- [9] 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.
- [10] S. G. Nikolov, P. Hill, D. R. 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, pages 213–244. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.
- [11] Z. Zhang and R. Blum. A categorization and study of multiscale-decomposition-based image fusion schemes. *Proceedings of the IEEE*, pages 1315–1328, 1999.
- [12] 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.
- [13] 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.
- [14] B. Matuszewski, L.-K. Shark, and M. Varley. Region-based wavelet fusion of ultrasonic, radiographic and shearographic non-destructive testing images. In *Proceedings of the 15th World Conference on Non-Destructive Testing*, Rome, October 2000.
- [15] G. Piella. A region-based multiresolution image fusion algorithm. In *ISIF Fusion 2002 conference*, pages 1557–1564, Annapolis, USA, July 2002.
- [16] G. Piella and H. Heijmans. Multiresolution image fusion guided by a multimodal segmentation. In *Proceedings of Advanced Concepts of Intelligent Systems*, pages 175–182, Ghent, Belgium, September 2002.
- [17] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and C. N. Canagarajah. Region-based image fusion using complex wavelets. In *7th International Conference on Information Fusion (Fusion 2004)*, Stockholm, Sweden, 28 June - 1 July. International Society of Information Fusion (ISIF), 2004.
- [18] R. J. O'Callaghan and D. R. Bull. Combined morphological-spectral unsupervised image segmentation. *IEEE Transactions on Image Processing* (submitted), 2003.
- [19] University of Edinburgh. Edinburgh Online Graphics Dictionary. <http://www.dai.ed.ac.uk/homes/rbf/grdict.htm>, 2003.