Dynamic Image Fusion Performance Evaluation

Vladimir Petrović[*], Tim Cootes
Imaging Science
University of Manchester
Oxford Road, Manchester, M13 9PT, UK
{v.petrovic,t.cootes}@manchester.ac.uk

Rade Pavlović
Military Technical Academy
Pavla Juriši a Šturma 33
11000 Belgrade, Serbia

Abstract - This paper deals with the problem of objective evaluation of dynamic, multi-sensor image fusion. For this purpose an established static image fusion evaluation framework, based on gradient information preservation between the inputs and the fused image, is extended to deal with additional scene and object motion information present in multi-sensor sequences. In particular formulations for dynamic, multi-sensor information preservation models are proposed to provide space-time localised fusion performance estimates. Perceptual importance distribution models are derived to accommodate temporal data and provide a natural generalisation of localised performance estimates into both global and continuous dynamic fusion performance scores. The proposed system is described in detail and shown to exhibit better evaluation accuracy, robustness and sensitivity when compared to existing dynamic fusion metrics on an evaluation of several established image fusion algorithms applied to multi-sensor sequences from an array of dynamic fusion scenarios.

Keywords: Video fusion performance, objective fusion evaluation.

1 Introduction

Multiple sensor modalities offer additional information and improved robustness to a range of imaging applications but come at a price of a large increase in raw data that needs to be processed. This can often overwhelm human observers and machine vision systems entrusted with the task of extracting useful information from such imagery. Image fusion combines information from a number of multi-sensor images into a single fused one, making processing more efficient and display of information more reliable. A variety of fusion algorithms have been proposed [1-5] to tackle this problem.

Of particular interest is the application of image fusion in real-time multi-sensor imaging such as that used in military, civilian avionics, medical imaging and surveillance. In this context continuous (video) streams of up to 30 images from different sensors are fused into a single output fusion stream each second.

Most fusion algorithms proposed so far deal with still image fusion and explicitly aim at achieving optimal accuracy of representing spatial information from the inputs in the fused image. This does not preclude their use in video fusion as they can be applied to each multi-sensor frame independently but it also creates a number of specific problems. One is the temporal stability of the result where large fluctuations in appearance of the fused images can be caused by fusion parameters varying across the sequence. Furthermore, fusion accuracy usually comes at the expense of lower efficiency which is not a problem when still images are fused in an offline manner, but can directly preclude application of such algorithms to real-time fusion. For these reasons a number of specific video fusion algorithms have also been proposed to specifically deal with the issue of fusing moving pictures [4,5,16]. Together with existing still image fusion schemes these systems represent a considerable pool of possible algorithmic solutions for any video fusion application.

Consequently there is a need for robust methods of evaluating the results of dynamic image fusion and comparing the performance of various algorithms. An issue of further interest is that of adaptive fusion. Recently, schemes have emerged that rely on robust objective evaluation to adapt the parameters of the fusion algorithm to current conditions and inputs in order to achieve optimal fusion robustness [17]. This is particularly applicable to real-time video fusion where input conditions may change considerably over long periods while costs of parameter optimisation can be spread across a number of frames [17]. In this context robustness of performance evaluation is also critical and may not be sufficiently provided by existing still fusion evaluation metrics.

Performance evaluation of still image fusion has been studied relatively extensively in the past with a number of algorithms published in the literature [5-14]. They are clustered around a number of key ideas. The most natural approach is the concept of subjective fusion evaluation where representative audiences of observers are asked to perform some tasks with or simply view and evaluate fused imagery [6,7]. The main drawback of such trials however, is that they require complex display equipment and an organisation of an audience making them highly impractical.

More recently, objective fusion metrics that require no display equipment or audience have emerged [8-14]. They require no ground truth data and produce a single numerical score reflecting fusion performance based entirely on the analysis of the inputs and the fused image. They can be realised computationally in full making them suitable for demanding video fusion evaluation. One such evaluation approach is based on the Universal Image Quality Index [9,13] where local image statistics are used to define a similarity between all corresponding 8×8 blocks across input and fused images. Information theoretic measures based on global image statistics such as entropy and mutual information have also been considered within the context of fusion evaluation [10,12]. These metrics explicitly ignore local structure but despite this apparent shortcoming, when considering reasonable fusion algorithms (that aim to preserve spatial structure of the inputs), they can achieve high levels of evaluation accuracy [12].

Mutual information has also been the basis for the most significant sequence fusion evaluation metric proposed do far [5]. Its aim was to measure the effects of various decomposition approaches on the temporal stability and overall quality of a fused sequence of images. The metric is evaluated by considering differential entropies, *H*(.), of joint variables constructed from the inter-frame differences (IFDs) of input images S_1 and S_2 and the fused image F (1):

$$I((S_1,S_2);F)=H(S_1,S_2)+H(F)-H(S_1,S_2,F) \quad (1)$$

The entropies themselves are evaluated from joint probability densities constructed from the IFDs using the Parzen estimate (with a Gaussian kernel) and a random sub-sample of the data. This approach shares in the general shortcomings of a sparsely populated probability density that can be influenced heavily by outliers as well as in the fact that it only considers a part (some 25000 samples [5]) rather than all of the available data.

This paper extends an established static image fusion evaluation framework based on gradient information preservation between the inputs and the fused image to deal with additional scene and object motion information present in multi-sensor sequences. In particular novel formulations for dynamic, multi-sensor information preservation models are proposed to provide space-time localised fusion performance estimates. Additionally, perceptual importance distribution models are derived to accommodate temporal data and provide a natural generalisation of localised performance estimates into both global and continuous dynamic fusion performance scores. In the following section, the basic gradient based still fusion evaluation framework is considered. The proposed dynamic extension of the fusion evaluation framework is described in detail in Section 3. The proposed dynamic fusion metric is demonstrated on an evaluation of several established image fusion algorithms applied to multi-sensor sequences from an array of dynamic fusion scenarios. Its results are compared to the currently best dynamic fusion metric and established still fusion metrics applied on a frame-by-frame basis.

2 Gradient Based Fusion Evaluation

Gradient representation based objective image fusion performance metric $Q^{AB/F}$ [11,14] is based on the idea that a fusion algorithm that transfers input gradient information into the fused image more accurately performs better. For the fusion of input images A and B resulting in a fused image F, gradient strength **g** and orientation **α** (∈[0,π]) are extracted at each location (*n*,*m*) from each image using the Sobel operator. They are then used to define relative strength and orientation "change" factors G and A, between each input and the fused image:

$$G^{AF}_{n,m}=\left\{\begin{array}{ll}\dfrac{g^F_{n,m}}{g^A_{n,m}}, & if \quad g^A_{n,m}>g^F_{n,m} \\ \dfrac{g^A_{n,m}}{g^F_{n,m}}, & otherwise\end{array}\right\} \quad (2)$$

$$A^{AF}_{n,m}=2\pi^{-1}\left|\left|\alpha^A_{n,m}-\alpha^F_{n,m}\right|-\pi/2\right| \quad (3)$$

where M is 1 for $g^F>g^A$ and –1 otherwise. This effective reversal of the gradient change ensures that G^{AF} remains in the range [0,1]. Its significance is that any increase in gradient strength from the input into the fused image is considered as an artifact of the fusion process and results in G^{AF} < 1 (1 is perfect reconstruction). A spatial information preservation measure Qs^{AF} (originally denoted Q^{AF} in [11,14]) models information loss between *A* and *F* with respect to the gradient "change" parameters using sigmoid functions defined by constants *Γ*, κ_g, σ_g, κ_α, and σ_α:

$$Qs^{AF}_{n,m}=\frac{\Gamma}{\sqrt{(1+e^{k_g(G^{AF}_{n,m}-\sigma_g)})(1+e^{k_\alpha(A^{AF}_{n,m}-\sigma_\alpha)})}} \quad (4)$$

Overall fusion performance $Q^{AB/F}$ is evaluated as a sum of spatial information preservation estimates between each of the inputs and fused, Qs^{AF} and Qs^{BF}, weighted by local perceptual importance factors w^A and w^B usually defined as local gradient strength:

$$Q^{AB/F}=\frac{\sum_{\forall n,m}Qs^{AF}_{n,m}w^A_{n,m}+Qs^{BF}_{n,m}w^B_{n,m}}{\sum_{\forall n,m}w^A_{n,m}+w^B_{n,m}} \quad (5)$$

$Q^{AB/F}$ is in range [0,1] where 0 signifies complete loss of input information and $Q^{AB/F}=1$ indicates “ideal fusion” [11,14]. Optimal parameters for the information loss non-linearities were determined by optimising the results of the metric against results subjective evaluation trials with values of $[\kappa_g, \sigma_g, \kappa_\alpha, \sigma_\alpha] = [-11, 0.7, -24, 0.8]$ shown to give highest level of agreement between subjective and metric scores [7]. The final parameter Γ is found analytically from the other parameters so that Qs^{AF} is 1 when both G^{AF} and A^{AF} are also 1.

Augmentation of $Q^{AB/F}$ with abstract information obtained from robust image segmentation [11] were shown to slightly improve on its robustness but require highly specialised implementations and are not considered within the context of dynamic fusion evaluation. Instead the original implementation of $Q^{AB/F}$ [14] is used in a frame-by-frame arrangement where each multi-sensor sequence frame is considered as an independent still image fusion. Performance score for entire sequence is obtained by averaging individual frame fusion performance scores.

3 Dynamic Fusion Evaluation

At a basic level evaluation of continuous image fusion can be viewed as an extension of the general still image fusion case where temporal information is implicitly transferred into the fused sequence by accurately fusing each frame individually. The simplest approach to evaluate a fused sequence would then be to apply existing still fusion metrics, such as MI [10] or $Q^{AB/F}$ to each frame individually, as described in previous section. This however, only takes into account spatial information and may result in sequences showing considerable instability in terms of significantly varying appearance between frames while still achieving a good performance score.

A dynamic, gradient based continuous fusion evaluation metric $DQ^{AB/F}$ is illustrated in Figure 1, explicitly evaluates the transfer of temporal information from the input sequences into the fused one. The evaluation is based on several (3) consecutive frames of all three sequences. The current frame from all three sequences is processed using the conventional $Q^{AB/F}$ approach to address preservation of input spatial information in the fused sequence. This yields a spatial information preservation estimate for each location (n,m) in the scene. Meanwhile previous and subsequent frames from each sequence are used in the evaluation of temporal gradient information that provides an estimate of preservation of input temporal information at each location in the fused sequence. Both sets of estimates are integrated into a single set of spatio-temporal information preservation estimates for each location in the scene. These estimates are then weighted with local perceptual importance coefficients to obtain a single performance score for the current frame. Individual frame scores are then simply averaged to obtain a single performance score for the whole sequence.

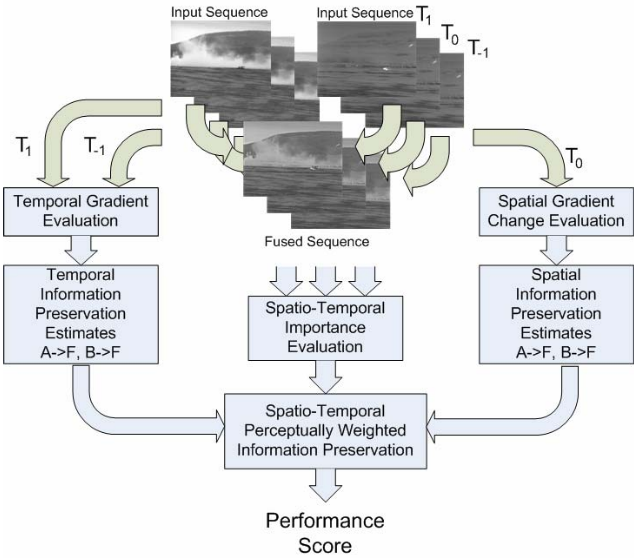
Input Sequence
Input Sequence T_1
T_0
T_{-1}
T_1
T_{-1}
T_0
Temporal Gradient Evaluation
Spatial Gradient Change Evaluation
Fused Sequence
Temporal Information Preservation Estimates A->F, B->F
Spatio-Temporal Importance Evaluation
Spatial Information Preservation Estimates A->F, B->F
Spatio-Temporal Perceptually Weighted Information Preservation
Performance Score

Figure 1: Architecture of the dynamic fusion evaluation metric $D^{AB/F}$

3.1 Temporal Information

Temporal information is recovered from each of the sequences using a temporal gradient operator, shown in Figure 2 a). This operator is analogous to the standard Sobel edge operator used to recover spatial information in the $Q^{AB/F}$ framework and shown for reference in 2 b). As we are only interested in recovering temporal information this operator is symmetrical in the spatial dimensions.

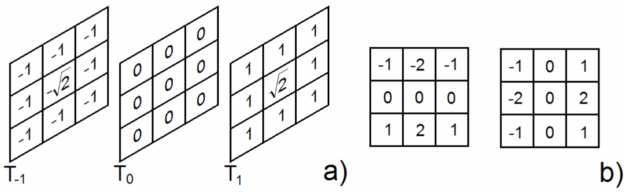

Figure: Gradient operators used to recover a) temporal and b) spatial gradients from the images

The temporal gradient ***gt*** is based on evaluating changes in image intensity at a particular location across time. For the current frame T_0, it is evaluated by filtering (convolution) of the previous and subsequent T_1 frames of the sequence by the S_{T-1} and S_{T+1} templates shown in Figure 2 a):

$$gt_{T_0} = \left| (I_{T_0-1} * S_{T-1}) + (I_{T_0+1} * S_{T+1}) \right| \qquad (6)$$

The additional temporal separation of the frames used in evaluation of ***gt*** add general noise robustness in comparison to the other alternative of using direct inter-frame differences, such as for example in [5]. An example of the results of this operator can be seen in Figure 3 where two consecutive frames from the infrared Dublin sequence (AIC Thermal/Visible Night-time Dataset collected by the University of Dublin) [18], are used to evaluate ***gt*** 3 c) and simple inter-frame differences 3 d). It is obvious that the ***gt*** image contains considerably less noise and a significantly better useful signal definition than the very noisy IFDs where useful signal is barely detectable.

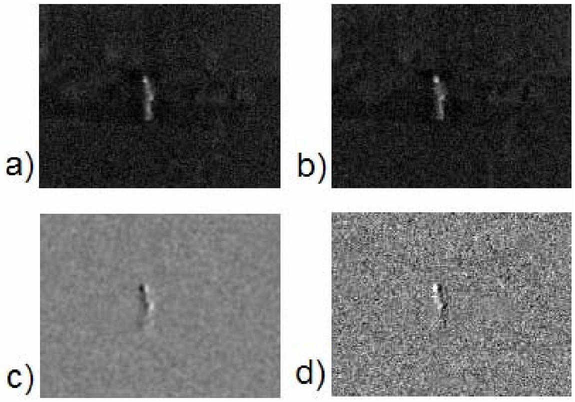

Figure 3: Consecutive frames from the Dublin sequence a) and b) and corresponding temporal gradients - c) temporal Sobel and d) inter-frame difference

Like with spatial gradient magnitude, a change in temporal gradient is evaluated between ***gt*** values of each of the input sequences and the fused sequence. Again we are only interested in preserving the magnitude of this parameter as sensor modalities with simply inverted intensity ranges (equal but opposite gradients) are considered to contain equivalent information. Temporal change parameter $T^{AF}{}_{n,m}$ is therefore treated in the same manner as spatial gradient magnitude. $T^{AF}{}_{n,m}$ is evaluated using equation (2) and $|\boldsymbol{gt}^{\{A,B,F\}}|$ as input. The temporal information preservation estimates are computed using the non-lonearity parameters used in evaluating the preservation of gradient magnitude:

$$Qt_{n,m}^{AF} = \frac{\Gamma}{\sqrt{1+e^{k_g(T_{n,m}^{AF}-\sigma_g)}}} \qquad (7)$$

3.2 Dynamic Importance

Motion in images is one of the main drivers of distribution of visual attention and hence perceptual importance. Humans are acutely sensitive to any temporal changes in a visual stimulus. Moreover this effect extends to peripheral (extra-foveal) vision meaning that observers notice temporal changes in a much larger area within a visual stimulus. Within the context of dynamic fusion evaluation this means that temporal change is a significant contributor to local perceptual importance assignment. In this preliminary study we have restricted our evaluation to a simple model of augmenting existing $Q^{AB/F}$ perceptual importance estimates $w^{A}{}_{n,m}$ (for image A) based on local contrast magnitude with the magnitude of the temporal gradient:

$$w_{n,m}^{A} = g_{n,m}^{A} + \left|gt_{n,m}^{A}\right| \qquad (8)$$

In this way larger perceptual importance is assigned to areas that exhibit movements of high contrast features. Such importance estimates are evaluated for all scene locations in both inputs, w^A and w^B.

3.3 Dynamic Fusion Performance DQ$^{AB/F}$

Dynamic fusion performance is evaluated by combining the spatial information preservation estimates obtained from the current frame, equation (4), and temporal information preservation estimates obtained from previous and subsequent frames, to equation (7), into a single spatio-temporal information preservation estimate for each scene location:

$$Q_{n,m}^{AF} = \sqrt{Qs_{n,m}^{AF} + Qt_{n,m}^{AF}} \qquad (9)$$

These estimates are then combined into a single performance score for the current frame $Q^{AB/F}(T)$:

$$Q^{AB/F}(T) = \frac{\sum_{\forall n,m} Q_{n,m}^{AF} w_{n,m}^{A} + Q_{n,m}^{BF} w_{n,m}^{B}}{\sum_{\forall n,m} w_{n,m}^{A} + w_{n,m}^{B}} \qquad (10)$$

In contrast to conventional $Q^{AB/F}$, $Q^{AB/F}(T)$ also incorporates the estimates of the preservation of temporal information centred at frame T. Finally, a single fusion performance score for the entire sequence is obtained by simply averaging all the individual frame performance scores over all N_T frames:

$$DQ^{AB/F} = \frac{1}{N_T}\sum_{T=1}^{N_T} Q^{AB/F}(T) \qquad (11)$$

Note that DQ is expected to have on average lower values than still image $Q^{AB/F}$ due to the additional multiplicative factor of temporal information preservation in (9) that is in the range 0 to 1.

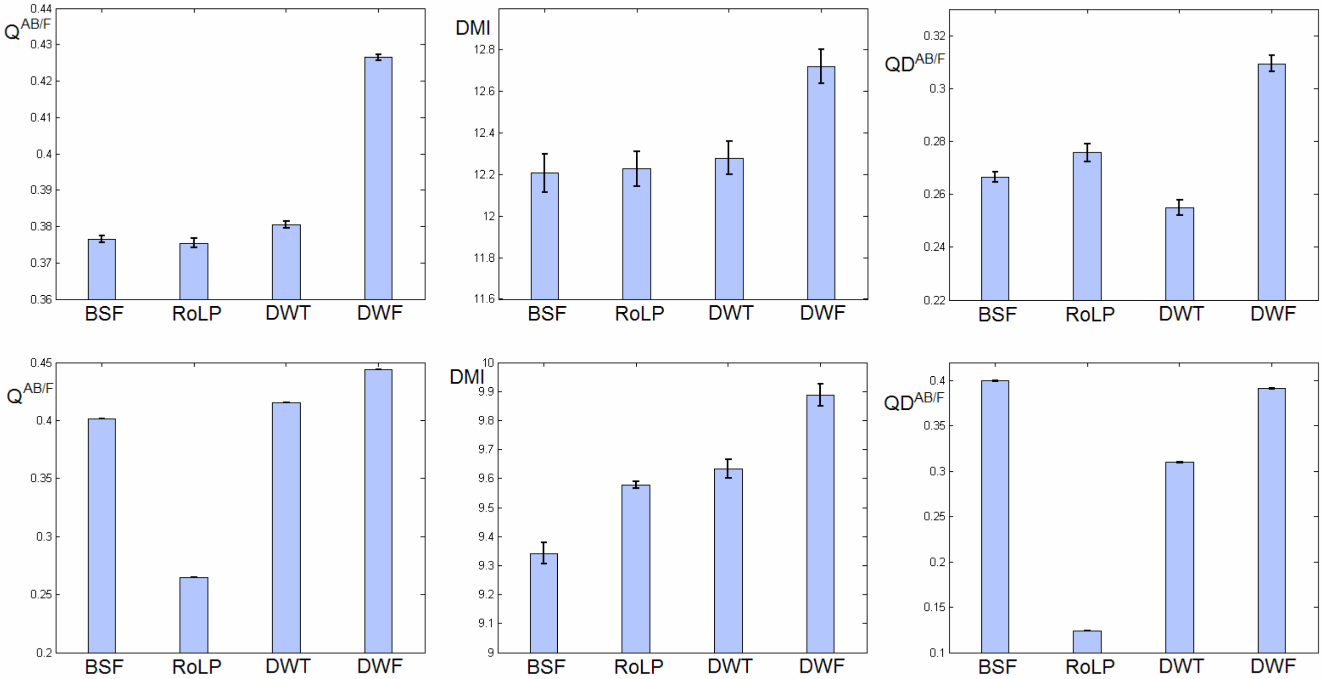

Figure 4: Performance of four fusion schemes on two well known multi-sensor sequences, evaluated using frame-by-frame $Q^{AB/F}$, dynamic mutual information DMI, and the proposed $DQ^{AB/F}$ metrics

4 Results

Proposed dynamic fusion evaluation metric $DQ^{AB/F}$ was tested on a representative sample of multi-sensor image sequences including both urban and natural scenes [15]. Sequences were fused using four representative fusion schemes, that include fast, dynamic bi-scale fusion (BSF) [4], Ratio of Low-Pass (RoLP) or contrast pyramid fusion [3], non-redundant wavelet fusion (DWT) [1] and redundant wavelet frame fusion (DWF) identified in [5] as the optimal dynamic fusion approach. The evaluation was compared against the original $Q^{AB/F}$ metric applied in a frame-by-frame arrangement and the dynamic mutual information (DMI) video fusion metric proposed in [5].

Figure 4 shows mean performance scores along with standard errors obtained by the three metrics on four fusion schemes fusing two representative multi-sensor sequences, the well known MS01 [15] and Dublin sequence [19]. From the results it can be seen that for the MS01 sequence both the frame-by-frame $Q^{AB/F}$ and dynamic mutual information DMI metrics are unable to discern between the performances of three of the schemes, BSF, RoLP and DWT, with enough certainty (outside 2 standard errors range).

The proposed $DQ^{AB/F}$ metric meanwhile provides unambiguous scores and rankings of the three schemes. More importantly it also identifies an important temporal stability weakness of the DWT approach, clearly missed by the other two metrics. Figure 5 illustrates this on four consecutive frames of the MS01 sequence.

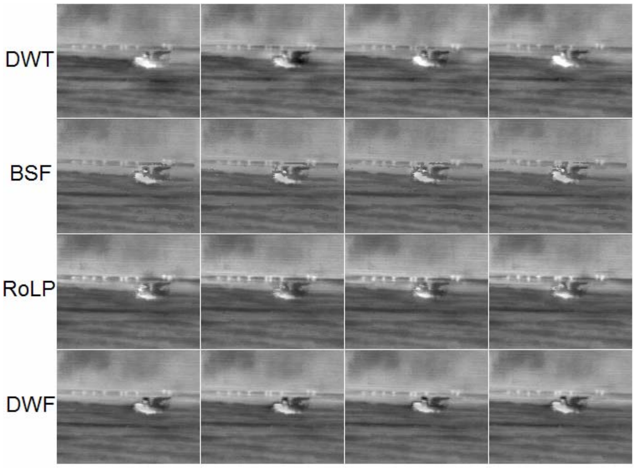

Figure 5: Example fused frames from the MS01 sequence illustrating temporal instability of DWT fusion, top row

Unstable pyramid coefficient selection causes significant fluctuations in object illumination across the frames of the fused sequence, note the appearance of the vehicle in the centre of the image. This effect, also observed in [5], creates an impression of varying object appearance where no such effects exist in the input sequences. Other schemes produce more stable results with DWF correctly identified by all metrics as providing best overall performance.

The proposed $DQ^{AB/F}$ metric also correctly identifies the poor performance of the RoLP scheme in fusing the noisy Dublin sequence. Low SNR infrared input, see Figure 3, adversely effects the local contrast estimates used in the RoLP pyramid fusion in such a way as to amplify the noise. This results in extremely unpleasant speckle noise in the fused sequence. Figure 6 illustrates this effect, where human figures clearly discernable in the IR and visual band inputs are almost invisible under the noise in the RoLP fused sequence. BSF fused images on the other hand are clear and unambiguous. DMI metric which operates on global statistics is not sensitive enough to evaluate such localised effects and RoLP fusion is scored better than BSF.

In general the DQAB/F metric shows a much more robust evaluation demonstrated by the considerably lower standard error range for an equivalent number of evaluation frames compared with the DMI metric. It also demonstrates superior sensitivity demonstrated in its ability to rank the performance of different fusion schemes confidently, compared to both frame-by-frame $Q^{AB/F}$ and DMI metrics.

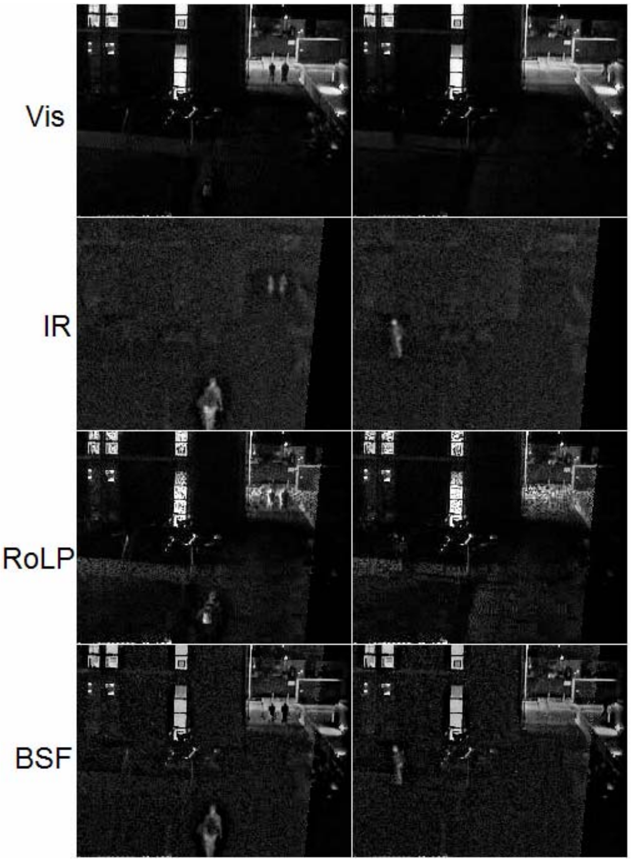

Figure 6: Examples from the Dublin sequence, first two rows are two frames of the visual and infrared inputs below them are corresponding RoLP and BSF fused frames

Numerical performance scores for these two as well as two further multi-sensor sequences (Octec and Uncamp [15]) are provided in Table 1. Once again the $DQ^{AB/F}$ is shown to produce unambiguous results and identifies the temporal weakness of the DWT approach across all sequences.

Table 1: Numerical scores of DQAB/F compared to still image QAB/F and DMI over four representative multi-sensor image sequences

MS01

Scheme:	BSF	RoLP	DWT	DWF
$Q^{AB/F}$	0.401	0.265	0.415	0.444
DMI	9.34	9.58	9.63	9.89
$DQ^{AB/F}$	0.266	0.276	0.255	0.309

Dublin

Scheme:	BSF	RoLP	DWT	DWF
$Q^{AB/F}$	0.377	0.375	0.38	0.426
DMI	12.2	12.2	12.3	12.7
$DQ^{AB/F}$	0.4	0.124	0.31	0.391

Octec

Scheme:	BSF	RoLP	DWT	DWF
$Q^{AB/F}$	0.348	0.345	0.346	0.386
DMI	10.1	9.91	10.4	10.4
$DQ^{AB/F}$	0.316	0.344	0.303	0.359

Uncamp

Scheme:	BSF	RoLP	DWT	DWF
$Q^{AB/F}$	0.307	0.346	0.3	0.361
DMI	10.1	10.5	10.6	10.9
$DQ^{AB/F}$	0.26	0.257	0.212	0.291

5 Conclusion

This paper presented a novel metric for evaluation of dynamic image fusion performance, based on estimation of spatio-temporal gradient information preservation. Temporal information is extracted from the image sequences using a robust temporal gradient filtering approach, shown to capture relevant temporal information even in low SNR conditions. An existing biological response model is used to determine the preservation of temporal information from the input sequences in the resulting fused sequence. Spatio-temporal preservation estimates are then obtained by integrating the temporal with conventionally evaluated spatial gradient preservation estimates. Performance for each frame is obtained by integrating local preservation estimates using local spatio-temporal importance estimates, and integrated over the entire sequence. Results of evaluation of a number of established fusion algorithms on a representative set of multi-sensor sequences have shown that the proposed $DQ^{AB/F}$ metric exhibits improved accuracy, robustness and sensitivity compared to existing state-of-the-art video fusion evaluation metrics as well as established still fusion metrics applied on a frame-by-frame basis.

Further work on the dynamic fusion evaluation will include exploration of different methods of determining integrated spatio-temporal gradients to obtain a more compact description of spatio-temporal information. Development of more accurate spatio-temporal perceptual importance distribution algorithms that model more accurately the complex nature of human attention distribution will also be pursued. This should include methods for uneven distribution of perceptual importance across different frames in the same sequence as well as more accurate spatial distributions of dynamic perceptual importance. A general method for including colour information in the process of fusion evaluation will also be explored to address the growing trend [16] of integrating colour visual band with greyscale non-visible band imagery.

Acknowledgements

The authors would like to acknowledge the TNO Human Factors in the Netherlands, Defence Research Authority Valcartier Canada, University of Dublin and Octec Ltd. for the use of their imagery in the work reported here.

References

[1] S Nikolov, P Hill, D Bull, C Canagarajah, Wavelets for image fusion, in Wavelets in signal and image analysis, A. Petrosian and F. Meyer (Eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001, pp 213-244

[2] Z Zhang, R Blum, A Categorization of Multi-scale-Decomposition-Based Image Fusion Schemes with a Performance Study for a Digital Camera Application, Proceedings of the IEEE, Vol. 87(8), 1999, pp1315-1326

[3] A Toet, Hierarchical Image Fusion, J. Machine Vision and Applications, Vol. 3, 1990, pp 3-11

[4] V Petrović, C Xydeas, Computationally Efficient Pixel-level Image Fusion, Proceedings of Eurofusion99, Stratford-upon-Avon, October 1999, pp177-184

[5] O Rockinger, T Fechner, Pixel-Level Image Fusion: The Case of Image Sequences, Proceedings SPIE, Vol. 3374, 1998, pp 378-388

[6] A Toet, E Franken, Perceptual Evaluation of different image fusion schemes, Displays, Vol. 24, 2003, pp 25-37

[7] V Petrović, *Subjective tests for image fusion evaluation and objective metric validation*, Information Fusion, Vol. 8(2), 208-216, Elsevier, 2007

[8] C Xydeas, V Petrović, *Objective Image Fusion Performance Measure*, Electronics Letters, Vol. 36(4), 308-309, IEE, 2000

[9] G Piella, H Heijmans, *A New Quality Metric for Image Fusion*, Proceedings International Conference on Image Processing, Vol. 3, 173-176, IEEE, 2003

[10] G Qu, D Zhang, P Yan, *Information measure for performance of image fusion*, Electronics Letters, Vol. 38(7), 313-315, IEE, 2002

[11] V Tsagaris, V Anastassopoulos, *Global measure for assessing image fusion methods*, Optical Engineering, Vol. 45, SPIE, 2006

[12] V Petrović, T Cootes, Information Representation for Image Fusion Evaluation, *Proceedings of Fusion 2006*, Florence, ISIF, July 2006

[13] N Cvejić, D Bull, C Canegarajah, *A New Metric for Multimodal Image Sensor Fusion*, Electronics Letters, Vol. 43(2), 95-96, IEE, 2007

[14] V Petrović, C Xydeas, "Objective Evaluation of Signal-level Image Fusion Performance", *Optical Engineering*, SPIE, Vol 44(8), 087003, 2005

[15] The Online Resource for Research in Image Fusion, *www.imagefusion.org*, 2005

[16] Li J, Nikolov S, Benton C, Scott-Samuel, Motion-Based Video Fusion Using Optical Flow Information, Fusion 2006, Florence, ISIF 2006

[17] V Petrović, T. Cootes, *Objectively adaptive image fusion*, Information Fusion, Vol. 8(2), Elsevier, 2007, 168-176

[18] C Ó Conaire, N O'Connor, E Cooke, A Smeaton, Comparison of fusion methods for thermo-visual surveillance tracking. International Conference on Information Fusion ISIF 2006