Objective image fusion performance measure

C.S. Xydeas and V. Petrović

A measure for objectively assessing the pixel level fusion performance is defined. The proposed metric reflects the quality of visual information obtained from the fusion of input images and can be used to compare the performance of different image fusion algorithms. Experimental results clearly indicate that this metric is perceptually meaningful.

Introduction: The recent availability of multisensor systems in key image application areas, such as remote and airborne sensing, has motivated researchers to work on image fusion in general and pixel level image fusion in particular. Thus a plethora of pixel level fusion algorithms have been developed [1 – 3] with different performance and complexity characteristics. Fusion performance is mainly assessed using informal subjective preference tests and, to date, little if any effort has been directed towards the development of objective image fusion performance metrics.

Given the input and single fused output images, in this Letter we address the problem of measuring fusion performance objectively. A performance measurement framework is defined which quantifies the fusion process and is subsequently used to compare the performance of different pixel level fusion systems. Furthermore, experimental results of this metric are shown to be in agreement with preference scores obtained from informal subjective tests. This clearly indicates that the proposed fusion measure is perceptually meaningful.

Fusion measure: The goal in pixel level image fusion is to combine and preserve in a single output image all the 'important' visual information that is present in a number of input images. Thus an objective fusion measure should (i) extract all the perceptually important information that exists in the input images and (ii) measure the ability of the fusion process to transfer as accurately as possible this information into the output image.

In this work we associate important visual information with the 'edge' information that is present in each pixel of an image. Note that this visual to edge information association is supported by human visual system [4] studies and is extensively used in image analysis and compression systems. Furthermore, by evaluating the amount of edge information that is transferred from input images to the fused image, a measure of fusion performance can be obtained.

Specifically, consider two input images A and B, and a resulting fused image F. Note that the following methodology can be easily applied to more than two input images. A Sobel edge operator is applied to yield the edge strength $g(n,m)$ and orientation $\alpha(n,m)$ information for each pixel $p(n,m)$, $1 \le n \le N$ and $1 \le m \le M$. Thus for an input image A

$$g_A(n,m) = \sqrt{s_A^x(n,m)^2 + s_A^y(n,m)^2} \tag{1}$$

$$\alpha_A(n,m) = \tan^{-1}\left(\frac{s_A^y(n,m)}{s_A^x(n,m)}\right) \tag{2}$$

where $s_A^x(n,m)$ and $s_A(n,m)$ are the output of the horizontal and vertical Sobel templates centred on pixel $p_A(n,m)$ and convolved with the corresponding pixels of image A.

The relative strength and orientation values of $G^{AF}(n,m)$ and $A^{AF}(n,m)$ of an input image A with respect to F are formed as

$$G^{AF}(n,m) = \begin{cases} \frac{g_F(n,m)}{g_A(n,m)} & \text{if } g_A(n,m) > g_F(n,m) \\ \frac{g_A(n,m)}{g_F(n,m)} & \text{otherwise} \end{cases} \tag{3}$$

$$A^{AF}(n,m) = 1 - \frac{|\alpha_A(n,m) - \alpha_F(n,m)|}{\pi/2} \tag{4}$$

These are used to derive the edge strength and orientation preservation values

$$Q_g^{AF}(n,m) = \frac{\Gamma_g}{1 + e^{\kappa_g(G^{AF}(n,m) - \sigma_g)}} \tag{5}$$

$$Q_\alpha^{AF}(n,m) = \frac{\Gamma_\alpha}{1 + e^{\kappa_\alpha(A^{AF}(n,m) - \sigma_\alpha)}} \tag{6}$$

$Q_g^{AF}(n,m)$ and $Q_\alpha^{AF}(n,m)$ model the perceptual loss of information in F, in terms of how well the strength and orientation values of a pixel $p(n,m)$ in A are represented in the fused image. The constants Γ_g, κ_g, σ_g and Γ_α, κ_α, σ_α determine the exact shape of the sigmoid functions used to form the edge strength and orientation preservation values, see eqns. 5 and 6. Edge information preservation values are then defined as

$$Q^{AF}(n,m) = Q_g^{AF}(n,m) Q_\alpha^{AF}(n,m) \tag{7}$$

with $0 \le Q^{AF}(n,m) \le 1$. A value of 0 corresponds to the complete loss of edge information, at location (n,m), as transferred from A into F. $Q^{AF}(n,m) = 1$ indicates 'fusion' from A to F with no loss of information.

Having $Q^{AF}(n,m)$ and $Q^{BF}(n,m)$ for $N \times M$ size images, a normalised weighted performance metric $Q_P^{AB/F}$ of a given fusion process P that operates on images A and B, and produces F is obtained as follows:

$$Q_P^{AB/F} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} Q^{AF}(n,m)w^A(n,m) + Q^{BF}(n,m)w^B(n,m)}{\sum_{i=1}^{N}\sum_{j=1}^{M}(w^A(i,j) + w^B(i,j))} \tag{8}$$

Note that the edge preservation values, $Q^{AF}(n,m)$ and $Q^{BF}(n,m)$, are weighted by $w^A(n,m)$ and $w^B(n,m)$, respectively. In general, edge preservation values which correspond to pixels with high edge strength should influence $Q_P^{AB/F}$ more than those of relatively low edge strength. Thus, $w^A(n,m) = [g_A(n,m)]^L$ and $w^B(n,m) = [g_B(n,m)]^L$ where L is a constant. Also note that $0 \le Q_P^{AB/F}(n,m) \le 1$.

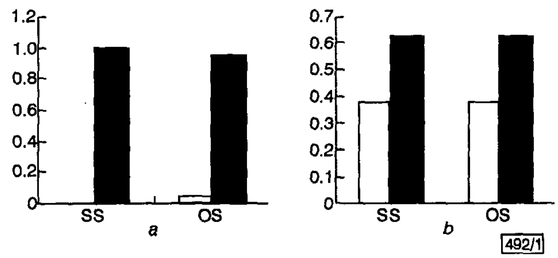


Fig. 1 *Hard decision subjective and objective performance assessment*

$L = 1$, $\Gamma_g = 0.9994$, $\kappa_g = -15$, $\sigma_g = 0.5$, $\Gamma_\alpha = 0.9879$, $\kappa_\alpha = -22$ and $\sigma_\alpha = 0.8$

a 10 input images and 11 subjects
☐ scheme 1
■ scheme 2
b 12 input images and 9 subjects
☐ scheme 1
■ scheme 3

Results: To establish the subjective relevance of the proposed methodology in assessing the performance of pixel level fusion systems, $Q_P^{AB/F}$ was estimated in relation to three fusion algorithms. The first, scheme 1, is a conventional multiresolution fusion system that employs an 'area' type subband pixel selection approach during pyramid fusion [3]. Scheme 2 employs the same conventional quadrature mirror filter (QMF) decomposition approach with an advanced cross-band selection technique for pyramid fusion [1]. Scheme 3 is a computationally efficient system based on a background/foreground decomposition and fusion process.

Thus informal subjective tests were performed using pairs of input images and the corresponding fused output images produced by two different fusion algorithms. Subjects were asked to vote in favour of one of the two systems or indicate that the algorithms perform equally well. A preference for a particular fused image assigned one point to the system used to produce it, whereas half a point was given to both systems in the case of equal preference. An average subjective score (SS) was therefore obtained for each fusion system, using the above hard decision process, when this test was applied over a set of K pairs of input images.

The same hard decision preference and corresponding point allocation scheme were also employed using the $Q_P^{AB/F}$ objective measure. That is, for a particular input pair

$$Q_I^{AB/F} > Q_{II}^{AB/F} \quad \text{1 point assigned to scheme 1}$$
$$Q_I^{AB/F} < Q_{II}^{AB/F} \quad \text{1 point assigned to scheme 2}$$
$$Q_I^{AB/F} = Q_{II}^{AB/F} \quad \tfrac{1}{2}\text{ point assigned to both scheme 1 and scheme 2}$$

This process yielded an average objective score (OS).

Fig. 1 shows the SS and OS values obtained from two experiments. The first assessed schemes 1 and 2 and involved $K = 10$ pairs of input images and 11 subjects, see Fig. 1*a*. The SS and OS values for schemes 1 and 3 obtained for $K = 12$ and nine subjects are shown in Fig. 1*b*. The pairs of input images used in these experiments are aerial visible and infra-red registered imagery [1, 2]. These were selected to represent a wide range of imaging content as captured using hyperspectral sensors. Note that the level of agreement between subjective and objective results is particularly high. At the same time, the size of the set of input pairs, K, used in these experiments is relatively small.

Conclusion: A novel objective pixel level image fusion assessment framework has been presented and has been used to compare the performance of different fusion algorithms. Preliminary experiments show that the hard decision fusion performance assessment obtained using the proposed objective measure agrees remarkably well with that obtained from informal subjective tests.

Acknowledgments: The authors gratefully acknowledge the financial and technical support provided throughout the project by the BAe Military Aircraft and Aerostructures Mission Systems Group.

17 November 1999
Electronics Letters Online No: 20000267
DOI: 10.1049/el:20000267

C.S. Xydeas (*Department of Communication Systems, University of Lancaster, Bailrigg, Lancaster, LA1 4YR, United Kingdom*)

V. Petrović (*Department of Electrical Engineering, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom*)

E-mail: vladimir@spec16.ee.man.ac.uk

References

1 PETROVIĆ, V., and XYDEAS, C.: 'Multiresolution image fusion using cross band feature selection', *Proc. SPIE*, 1999, **3719**, pp. 319–326
2 PETROVIĆ, V., and XYDEAS, C.: 'Computationally efficient pixel-level image fusion'. Proc. Eurofusion99, October 1999, pp. 177–184
3 LI, H., MUNJANATH, S., and MITRA, S.: 'Multisensor image fusion using the wavelet transform', *Graphical Models Image Proc*, 1995, **57**, (3), pp. 235–245
4 HANDEE, W., and WELLS, P.: 'The perception of visual information' (Springer, New York, 1997)

Reference block updating when tracking with block matching algorithm

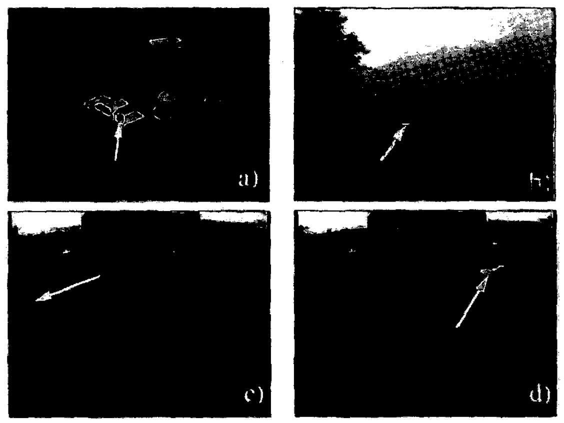
A.M. Peacock, S. Matsunaga, D. Renshaw, J. Hannah and A. Murray

An important problem when using the block matching algorithm to track objects is how to update the reference block to take account of the changing target appearance. A variety of update strategies are reported and compared.

Introduction: The block matching algorithm (BMA) is a popular correlation-based approach to motion estimation [1] and tracking [2]. In this approach, the motion of a block of pixels, termed a reference block, is estimated by looking for the most similar block of pixels in subsequent frames. Since the appearance of tracked objects can change over time, the reference block must be updated to take account of these changes. In this Letter we report a number of strategies for updating the reference block and discuss the relative performance of these strategies when applied to real data sequences. This work forms part of a larger investigation into analogue chip design for BMA based tracking, and so issues relating to this implementation are also discussed.

Adaptive block matching: In the BMA [1], it is assumed that the appearance of a block of pixels remains constant over time and motion. This is a reasonable assumption given high frame rates and short time periods. The motion of a block of pixels is estimated by looking for the most similar block in subsequent frames according to some similarity measure. When the BMA is used for tracking, the reference block must be altered to take account of changes in the appearance of the target object. We term this extended algorithm the adaptive block matching algorithm (ABMA).

There are a number of choices in the ABMA including the block size, the block shape, the similarity measure, selection of the initial block and the reference block update strategy. The search is typically constrained to a window the size of which must also be chosen. Errors in tracking can have numerous causes including camera noise and inappropriate choice of these parameters. We found that sub-pixel motion could also cause the tracking to be lost. The problem is due small changes in object appearance being insufficient to cause the tracking position to follow the motion, meanwhile the reference block is updated and so slowly comes to no longer resemble the target. We call this rounding error. We are concerned only with the choice of reference block update method, as the other parameters have been sufficiently investigated previously [2 – 4].



197/1

Fig. 1 *Example frames from track sequences (not to same scale) with arrows indicating tracked points*

a Hamburg taxi
b Car following
c Car park A (pedestrian)
d Car park B (car)

Test data: The results from four test data sets are presented; these are the well known Hamburg taxi sequence, two sequences from a car park security camera, and a sequence taken from one car following another. The Hamburg taxi sequence is too short to be informative, but is included to assist the reader in replicating the results. Since in this Letter we only consider the reference block update strategy, the sequences were chosen to have none of the missing or unintelligible frames which can appear in real data sequences.

In the car following data set, both the target and the camera are in motion and so there are many small quick movements due to uneven roads. Car park sequence A shows a pedestrian walking, and B shows a car moving behind a fence. The car park data set is particularly noisy and so no single reference block is ever a particularly good representation of the target. Fig. 1 shows typical frames from the data sets, with arrows indicating the tracking points. In all these three data sets, the appearance of the target varies considerably over the sequence.