Defence Science Journal, Vol. 58, No. 1, January 2008, pp. 136-146

Fusion of Noisy Multi-sensor Imagery

Anima Mishra and Subrata Rakshit

Centre for Artificial Intelligence & Robotics, Bangalore–560 093

ABSTRACT

Interest in fusing multiple sensor data for both military and civil applications has been growing. Some of the important applications integrate image information from multiple sensors to aid in navigation guidance, object detection and recognition, medical diagnosis, data compression, etc. While, human beings may visually inspect various images and integrate information, it is of interest to develop algorithms that can fuse various input imagery to produce a composite image. Fusion of images from various sensor modalities is expected to produce an output that captures all the relevant information in the input. The standard multi-resolution-based edge fusion scheme has been reviewed in this paper. A theoretical framework is given for this edge fusion method by showing how edge fusion can be framed as information maximisation. However, the presence of noise complicates the situation. The framework developed is used to show that for noisy images, all edges no longer correspond to information. In this paper, various techniques have been presented for fusion of noisy multi-sensor images. These techniques are developed for a single resolution as well as using multi-resolution decomposition. Some of the techniques are based on modifying edge maps by filtering images, while others depend on alternate definition of information maps. Both these approaches can also be combined. Experiments show that the proposed algorithms work well for various kinds of noisy multi-sensor images.

Keywords: Image fusion, edge maps, information maps, noise filtering, multi-resolution, wavelets, Laplacian pyramids, multi-sensor images

1. INTRODUCTION

With the availability of multi-sensor data in many fields such as remote sensing, medical imaging, machine vision and military applications, sensor image fusion has emerged as a new and promising field of research. The current definition of multi-sensor image fusion is very broad and the fusion can take place at signal, pixel, feature, and symbol-levels. Image fusion provides the means to integrate complementary and redundant information from multiple images into a composite image more suitable for human visual perception, and computer processing such as segmentation, feature extraction, and target recognition. Integrating disparate information improves interpretation capabilities. This leads to more accurate analysis, increased utility, and more robust performance. Besides, the redundant information from images is encoded just once in the output. This results in a more efficient storage and dimensionality reduction in feature vectors.

The main issue in fusion of many types of images for visual display is content preservation. Important details from all the input images should be preserved in the output image, while ensuring

Received 20 April 2006, revised 01 March 2007

that the merging technique should not introduce any kind of artifacts, blur and spurious patterns. The edges in images, which correspond to object boundaries, are usually the key features in images. There are various scenarios when constituent input images can encode different edges for the same physical scene. These are: (i) variable depth of focus (each input image may have objects at different depths in sharp focus), (ii) variable spectral response (each image may be using a different part of the spectrum, eg, optical and IR), and (iii) variable time of capture (input images may constitute snapshots taken at different times capturing different objects in the same field of view). In each case, one can create a single image having all the desirable information.

Prior to fusion, all input images must be registered to the same field of view. Image registration[1,2] is the process of geometrically-transforming images of the same scene such that all common objects have identical positions within each image. In simple cases, translations alone may be sufficient. In other cases, one may have to determine an affine transform to take care of scale, rotation, and projection. For the most general case, one would have to determine the camera parameters, estimate the 3-D structure of the scene and then render registered versions of the images to the extent possible. Only the multi-sensor images to be fused are assumed to be registered.

The fusion of registered images for capturing all the sharp edges has been discussed in the literature[3-5]. These methods are based on decomposition of images into wavelet representations. Wavelet and pyramid[6] subbands highlight the edges present in an image at various scales. The Laplacian subbands can also be combined using variants of the 'choose the highest value' logic and the resultant subbands can be reconstructed to create an output image[7]. All the merging techniques efficiently fuse clean and noise-free images. However, these techniques cannot be directly applied to noisy images. A number of sources of noise are available. IR images are grainy due to thermal (shot) noise, and SAR images are grainy due to specular noise. Dual field scanning can cause motion-induced jitters at edges, and wireless transmission can also add spot-and-line noise. If the fusion process carefully preserves the edges due to these noise artifacts, the composite image will have very poor quality. The fusion of noisy images requires a careful analysis of the theoretical basis of conventional fusion algorithms. Based on the understanding gained, one can try to improvise methods for fusion of noisy images. This is the focus of this paper.

This paper reports the traditional methods of fusion. A connection is made between the edge-based fusion and the information in images. Next, one considers the possible modifications that can be made for noisy images. The first solution method uses conventional techniques to reduce the noisy fusion problem to the noise-free fusion problem. The second method uses a more unconventional approach – it redefines the measure of information for images and utilises this new measure for fusion. Finally, the results are shown for registered noisy input images and quantitative evaluation of perceptual significance of various noise patterns. All the methods assume registered images, exploit multi-resolution representations and focus on edges. In practice, specific applications may require representations for texture, colour or selected target shapes.

2. EDGE-BASED FUSION

2.1 Basic Edge Fusion Methods

The basic idea of multi-resolution-based edge-based fusion methods can be summarised as follows:

- Create a multi-resolution representation of each input image (pyramid or wavelet using any filter order or basis). This creates bipolar, bandpass subband images at various resolutions. The edges in the input images show up as large non-zero values in these subbands.
- The sharper edges typically have larger non-zero values as the subbands encode differences from the local means. Thus, the more difficult problem of determining the edge sharpness translates into the more tractable problem of determining the pixel magnitudes. This is strictly true only for edges of various

sharpness between areas having the same intensity difference. When images having very different dynamic ranges need to be fused, different representations are used.

- The multi-resolution subbands of the output image are created pixel-by-pixel. For each pixel location, the corresponding pixels in the subbands of the various input are analysed and the maximum magnitude pixel is chosen for the output. The reconstruction of these new subbands gives an output image that combines the sharpest edges from all-input images.

Fusion results have been reported for both wavelets[3-5] and pyramids. The choice of the Laplacian pyramids[6,7] is recommended for two reasons: (i) the pyramid subbands are non-directional. The edges (at a given scale) are not broken up into various subbands and (ii) aliasing cancellation is not presumed in pyramid reconstruction. Thus, arbitrarily constructed pyramids can be reconstructed without the risk of introducing artifacts in the final image. In addition to the Laplacian pyramids, contrast decomposition-based ratio of low-pass (ROLP) pyramids[8-10] have also been used for edge-based fusion. The difference in dynamic ranges between the optical and the IR imagery leads to optical images always dominating the fused output for the Laplacian pyramids and wavelets. An ROLP is shown to give better results for optical and IR image fusion as perceptually important details in IR images, with a relatively high local luminance contrast, are preserved in the composite image.

The results of edge-based fusion using max rule with the Laplacian pyramids and ROLP pyramids are shown in Fig.1. The top row shows results of fusion for three optical images having different depths of focus. The bottom row shows fusion of optical and IR using both the Laplacian and the ROLP representations.

2.2 Relation to Information Fusion

The edge-based fusion methods outlined above are usually motivated from the functional point of view: Their justification is that these work. However, there are many cases where these do not work, eg, for noisy images, texture-dominated images, and for applications where user wants some specific feature to be fused, etc. To set reasonable expectations from fusion algorithms, as well as to devise extensions for some of the above cases, it is necessary to

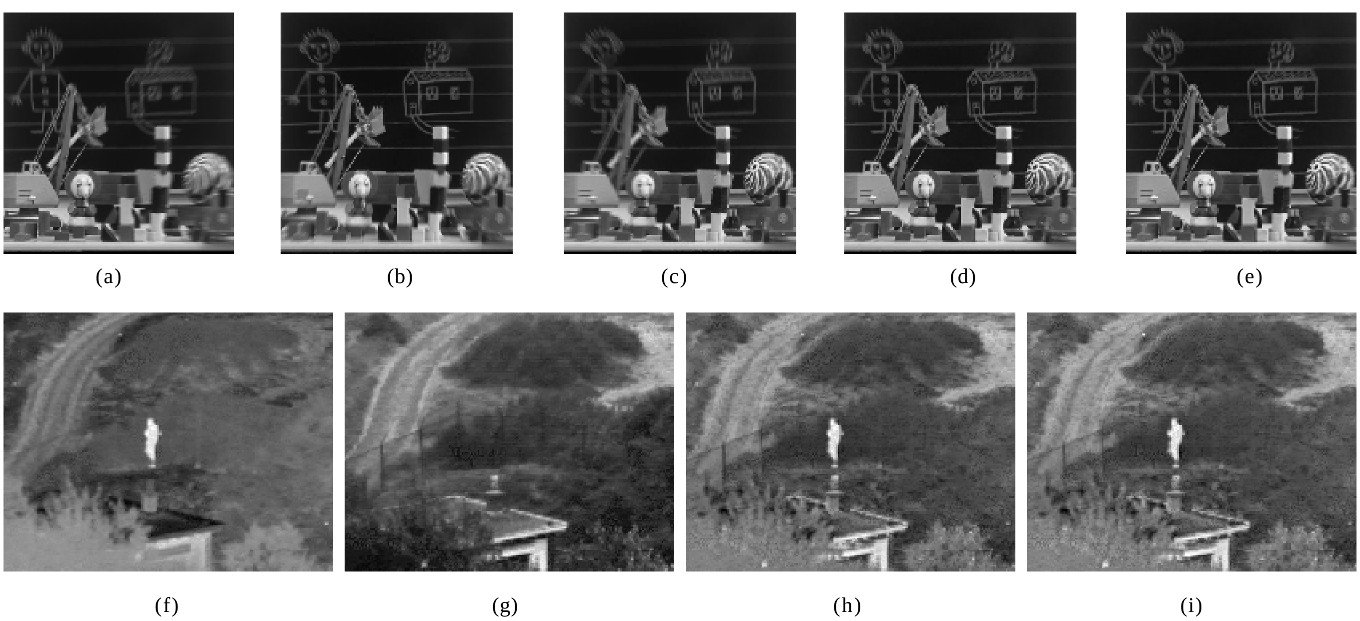


Figure 1. Three optical input images having different depths of focus are shown in figures (a), (b), and (c). The images are fused using (d) the Laplacian pyramids and (e) the ROLP pyramids. The bottom row shows the fusion of (f) IR image and (g) optical image using (h) the Laplacian pyramids, and (i) the ROLP pyramids.

relate the edge fusion methods to a theory of information in images.

Firstly, Shannon's information theory is applied to the images. If one computes the entropy of pixel values in images, and thus estimates information in images, the result is very unsatisfactory. White noise emerges as the most informative image. A pixel having a rare (in that image) intensity value is not necessarily encode any rare artifact. No universal intensity histogram exhibts for the images (even natural images) that can be used to assign information value to pixel values. However, the histograms of subbands do exhibit such a universal histogram. The histograms of subbands have a peak around 0 (small magnitudes are the most common) and sharp falloff on either side (large magnitudes are rare). Information maps created based on image pixel intensities and subband pixel intensities are shown in Fig. 2 for a set of images. As can be seen, the information map based on subband pixel image intensities is a more appropriate map as it highlights regions (edges) that are considered perceptually significant. Thus for the case of good quality natural images, at least, one can use Shannon's entropy of subband images as a measure of information in images.

The multi-resolution-based edge fusion algorithm can now be cast as an information maximisation algorithm. For each image, subbands are being computed. For each pixel location in the subband, the input with the largest pixel value is the one that encodes the maximum information at that spatial location, at that frequency band. Selecting all the maximum values from the input subbands for the output subbands creates a maximally information encoding representation that is consistent with the input image set. The above analysis is based on the crucial assumption that the subband pixels with the larger values are encoding more information. The fusion method will fail to give good results as soon as that assumption is violated.

3. FUSION OF NOISY IMAGES

In noisy images, the number of larger valued pixels in the lower subbands increases to encode the fluctuations due to noise. It is no longer true that these large values are very rare, or that these are always encoding intensity variations that describe structurally significant features in the images. To rectify the situation, one can proceed in three different ways: (i) modify the maximum pixel rule, (ii) suppress noise/preprocess images before fusion, and (iii)

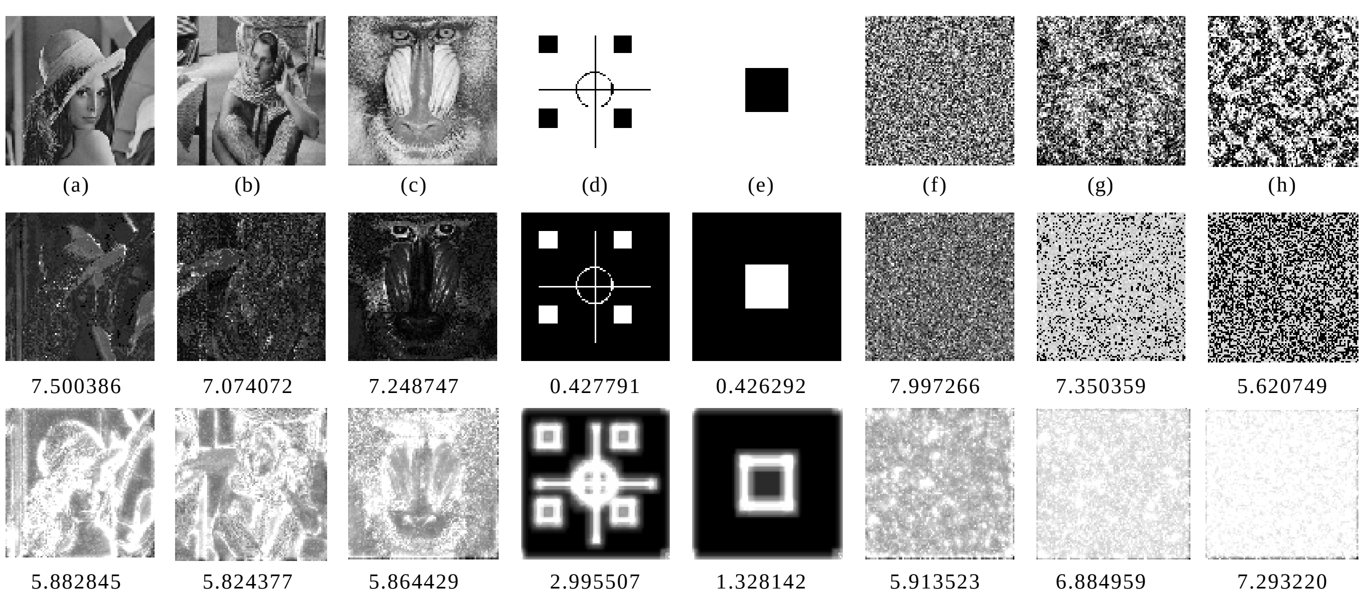


Figure 2. Top row represents a set of eight images, including natural images, binary patterns and textures. Middle row shows the information maps based on image pixel intensities and their frequencies of occurrence, and the bottom row shows the information maps based on subband pixel intensities and works better, at least for natural images (a), (b), and (c). A gross overestimation is for noise and texture images (f), (g), and (h).

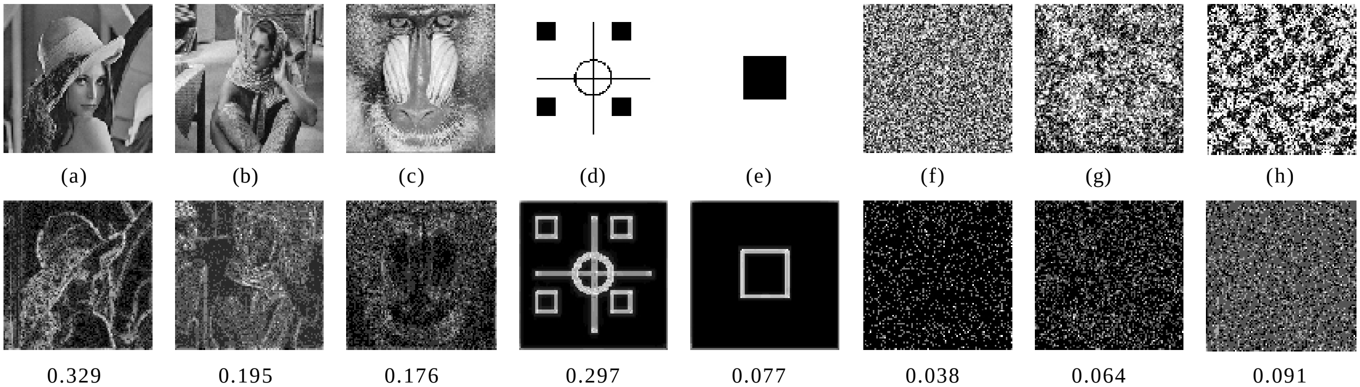


Figure 3. Top row shows a set of eight images, including natural images, binary patterns and textures. Bottom row shows the information maps based on MSOE (using inter-scale coherence model[11]). The scores are indicative of the relative information content of the various images. The coherence model used favours images with sparse but strong edges.

explicitly use information measures for weighting images.

3.1 Alternatives to Max Rule

The rule for selecting the pixel value based on input subbands need not only be the maximum rule. Many variations were tried, eg, mean, median, RMS, geometric mean, harmonic mean and exponential mean (log of the mean of exponentials). The rules that are biased towards the larger values (like RMS and exponential mean) give the best results. The motivation for using the rules other than the maximum rule is to have the output depending on all-input to some extent. This can be important if the objective is edge-evidence fusion rather than just selection of the best edge. As noise is unlikely to be coherent across the images, a noise-based edge in one image will lack support from the other images. In the presence of noise, edge-evidence fusion from an image cannot be believed without partial corroboration by other images. True edges, while being present more sharply in some input than others, will be present to some extent in all the images. The slight attenuation of the sharpest true edge is the price one pays for noise immunity. Figures 3 and 4 show the result for fusion using the RMS and exponential means. The input set is the images with different depths of focus.

Following are the various other fusion selection rules. If one denotes the input images by I_i, where $i = 0..(N\text{-}1)$ and the fused image by F.

- Arithmetic mean

$$F(x,y)=\frac{1}{N}\sum_{i=0}^{i=N-1} I_i(x,y)$$

- Max

$F(x,y)=\mathrm{Max}(I_i(x,y)), i=0..N\text{-}1$

Figure 4. Top row shows the image affected by random noise. Second row shows images after noise suppression after filtering. Bottom row shows (left to right) fusion of clean image, noisy images and filtered noisy images. The RMS fusion rule was used for all.

- Root mean square (RMS)

$$F(x,y)=\sqrt{\sum_{i=0}^{i=N-1}\left[I_i(x,y)^* I_i(x,y)\right]/N}$$

- Exponential mean

$$F(x,y)=\ln\left\{\frac{1}{N}\sum_{i=0}^{N-1}\exp\left[I_i(x,y)\right]\right\}$$

- Median

F(x,y)=Median [$I_i(x,y)$], i=0..N-1

- Geometric mean

$$F(x,y)=\sqrt[N]{I_i(x,y)},\ i=0..N-1$$

- Harmonic mean

$$F(x,y)=\frac{N}{\sum 1/I(x,y)}$$

These methods have various alternatives for fusion rule apart from the max rule. For fusion of noisy images, one has to deal with various combinations of fusion and filtering.

3.2 Noise Suppression and Fusion

It is important to differentiate between the image fusion by max rule and the simple averaging. For the simple averaging process, explicit noise removal would not be needed. The averaging would itself attenuate noise, so long as noise was incoherent across frames and the signal was coherent. This is the logic underlying time averaging to improve the SNR while designing sensor systems. For true image fusion, one must be able to deal with situations where even the signal is not the same (coherent) among the input. This necessitates invocation of nonlinear rules like the max rule for edge fusion. However, when applied to noisy images, the max rule tends to amplify noise. The assumption is that large subband pixels are encoding relevant structures is rendered false due to noise. Once noise has been reduced, the assumption about large subband pixels being informative will again hold good and one can proceed with the edge-based fusion as before. Thus, the model of image fusion as information maximisation shows that for noisy input, it is better to do the noise removal first (for each input) prior to fusion. Noise removal methods can be employed to each input image independently. To the extent possible, noise may be removed or attenuated. This process will vary depending on the nature of the noise (Gaussian, salt, pepper, linear streaks, etc). Noise removal after the fusion of noisy input will not give good results. This is illustrated in Figs 4 and 5.

If the two images have very different contrasts, their subband pixel value distributions will be different as well and equally significant edges will not be

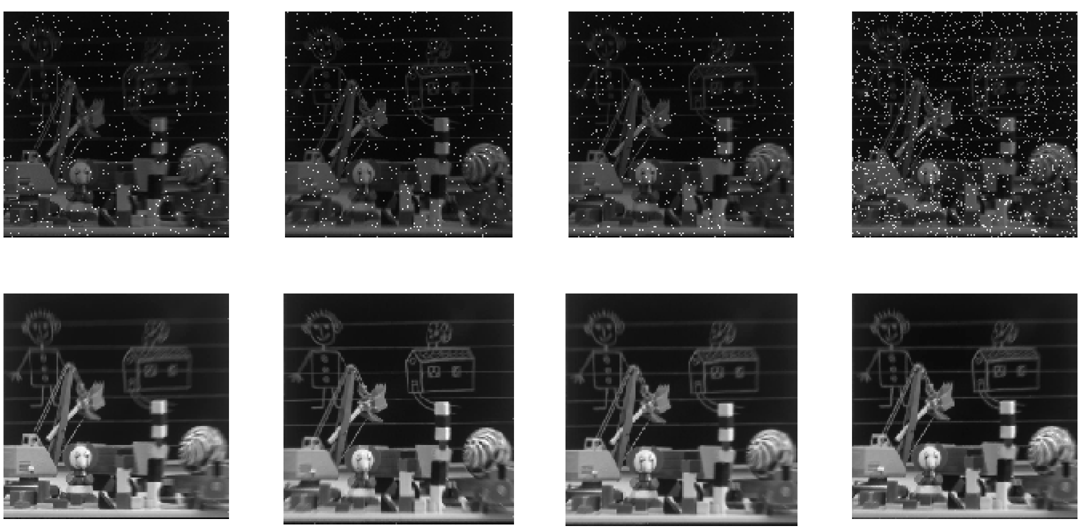


Figure 5. Top row shows the images affected by salt and pepper noise resulting in noisy-fused image. Bottom row shows the noisy images after median filtering and their fused output. The exponential mean rule was used for fusion of edges for all these images.

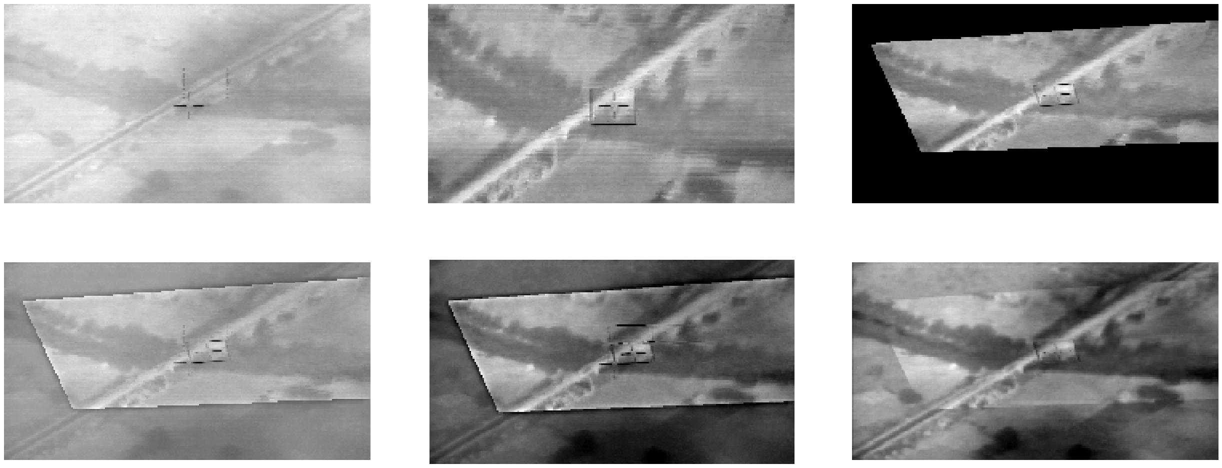


Figure 6. First row shows the optical image, unregistered IR image and registered IR image wrt optical image. Bottom-left image shows fusion of IR and optical images without noise removal, bottom-middle shows fusion followed by noise removal and bottom-right image shows fusion after noise removal and preprocessing.

encoded by equally large subband values. This equality has to be artificially established before the max pixel rule can be used for fusion. An example of this is shown in Fig. 6. Preprocessing the input clearly gives much better results. In this case, the preprocessing was customised for the two types of input to remove noise and equalise the dynamic ranges as well. The equalisation is needed to keep the assumption behind information maximisation valid.

In general, one can use the alternatives to the max rule. The choice of the fusion strategy, multi-resolution (wavelet, Laplacian pyramid or contrast pyramid) and noise removal method can all be dependent on the nature of noise. Automating the choice of the above is an open problem. For specific situations, these options can be manually selected and set.

3.3 Fusion based on Information Map

Suppose for each noisy image at each pixel location one could have a measure of information being encoded. Then, one could directly formulate image fusion as a weighted average problem. This would not serve to combine information in an additive sense (three-blurred edges would not define a sharp edge) but would at least enable one to pick the best input for a location (like the max rule for noise-free images). The main difficulty in implementing this scheme is defining the right measure of local information in the images.

Shannon's information measure computes information based on probability of occurrence. If that measure is used to weigh the pixels, two problems still exist. Firstly, the theory is more suited to variables than signals, in as much as the measure depends on overall probability of occurrence rather than on specific patterns of occurrence. For images, this means that interchanging pixels within an image does not affect the measure – though such interchanges could completely alter the image itself. This drawback may be fixed by applying the entropy measure to multi-resolution representations of the image rather than the images themselves. Even for reasons of achieving similar distributions across (similar) the images, computing information based on subband pixel values is anyway preferred. Encoding spatial distribution now adds a second justification. The second problem with entropy is harder to circumvent. As entropy measures uncertainty, the measure, when applied for images, maximises for highly varying images that essentially look like white-noise. Application of entropy to subbands does not perceptibly alter this situation. The key problem here is that the human visual system (HVS) perceives variations as interesting only up to a certain point and then ignores these as noise. Information theory, having being developed to estimate the cost of communicating values of random variables, was never designed to mimic this behaviour of the HVS. This problem requires the development of a different model of (perceived) information in signals. One gives one such possible model that one has improvised for creating the information maps.

Shannon's entropy measures have been extended to second-order measures based on multi-resolution[11] that can more effectively identify information in images. The second-order entropy (SOE) of a variable, $H_2(x)$, is computed by treating the (sampled) *pdf* of x as a new random variable and computing the entropy of this derived variable. $H_2(.)$ penalises excessive variations and gives lower scores to noisy (random) variables.

3.4 Algorithm for Defining Local Information – the Multi-resolution Second-order Entropy

The composite method for computing image information involving multi-resolution second-order entropy (MSOE) for an image, I, as follows[11]:

(a) Given I, compute[1] multi-resolution subbands

$M(I) = L_0,\ L_1,\ L_2 \ldots\ldots L_{n-1},\ G_n$

(b) For each subband, compute the corresponding second-order entropy – $H_2(L_k)$

(c) Compute MSOE as a function of the subband second-order entropy's, one has:

$$MSOE\ (I) = F(H_2(L_0), H_2(L_1)\ldots\ldots\ldots H_2(G_n))$$

The functional form of $F()$ depends on the choice of HVS model or judgment. F can be a weighted sum (frequency-response model) favouring some scales or a projection operator (inter-scale coherence model) favouring strong sharp edges.

(d) Optionally, normalise the measure using entropy of L_0, one has:

$$(NMSOE)\ (I) = MSOE(I)\ /\ H(L_0)$$

where NMSOE is the normalised multi-resolution second-order entropy. The normalised measure is required for comparison between the images with different pixel sizes or for discounting low-amplitude high-frequency variations as noise.

(e) Consider the subband entropy maps as a pyramid and reconstruct it, consistent with F to generate the information map of image, I.

The multi-resolution second-order entropy (MSOE) can create information maps while incorporating various HVS models[11]. Maps created using these methods can be used to perform the subband fusion. The subband pixels are weighted by the information assigned to that location in the image by MSOE. The fusion rule can again be max rule or any of the other rules biased towards the larger values. The information map-based fusion makes explicit the assumptions being made about what constitutes information in an image. The information measures and information maps computed for the image set used in Fig. 2 is shown below in Fig. 3.

The multi-resolution second-order entropy MSOE-based information estimates can also be used to evaluate the results of fusion by various techniques. The numerical scores assigned to images can be used to decide which noise suppression or image fusion method is giving the better result. These scores are useful as an alternate to the subjective evaluation by human beings.

4. RESULTS AND DISCUSSION

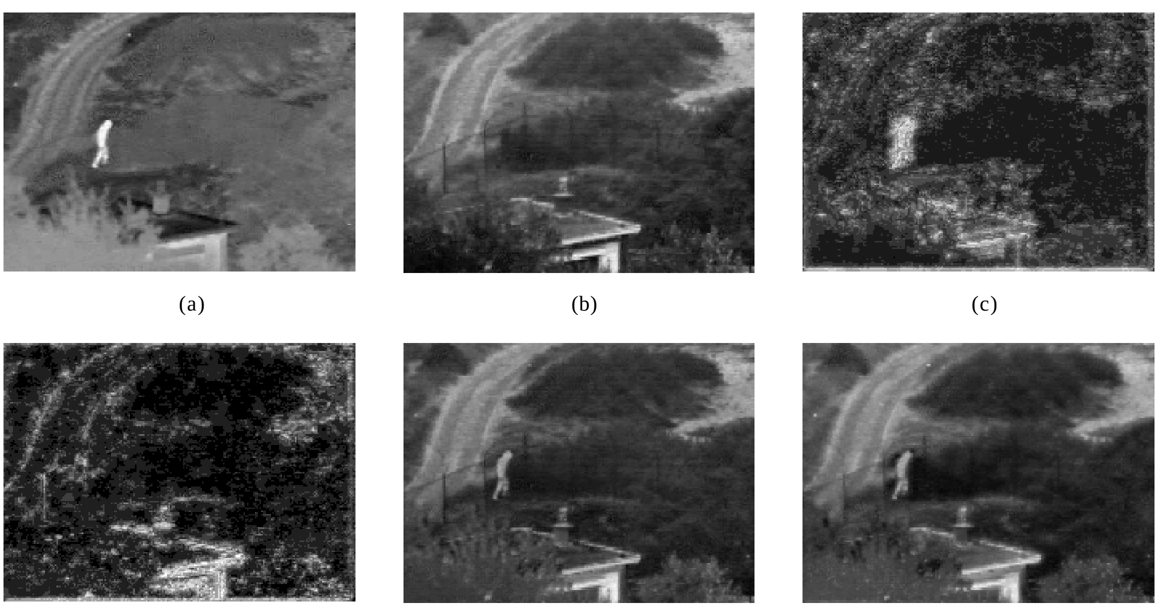
The methods for fusion described above have been applied on various images, and the results are presented. Perfect registration is a prerequisite for image fusion. Here, it is assumed that input images are registered perfectly.

Consider the three images in Fig. 4 (top row). These are the noisy versions of images shown in Fig. 1 (top left). The second row shows noise suppressed versions. For suppressing bright noise against dark background, a geometric mean-based filter was used. The third row shows the results of fusing the clean images, the noisy images and the noise-suppressed images. The fusion of the clean images is clearly the best. The fusion of the noisy images ends up amplifying the noise, making de-noising more difficult. The fusion of the noise-reduced images shows a marked improvement. All these images were fused using the RMS fusion rule.

In Fig. 5, one illustrates the fusion results for a different types of noise (salt and pepper) and a different fusion logic (exponential mean). Though fusion of noisy images results in a focused image, it shows noise enhancement as well. The fusion result of median-filtered images is closer to the fused result for noise-free images. The filtering process degrades the sharpness of objects but the overall improvement due to better noise rejection makes this the better technique for fusion of noisy images.

In Fig. 6, two input images are shown in the top row, the leftmost being an optical image and the next being an IR image. These images are low in contrast, unregistered and corrupted by jitter and line noise. The rightmost image in the top row is the registered IR image. The bottom row shows various fusion results. The left image shows fusion without noise removal and preprocessing. This fused image has low contrast and is corrupted by noise. The middle image shows result of fusion followed by the noise removal and contrast enhancement. The noise removal process affects the fused information, thus deteriorating the information gathered in the resulting image. Since the original images were already low in contrast, the enhancement methods tend to emphasise the edges in the registered image. The result of fusion applied after noise removal and contrast equalisation is shown in the last image. This fused image is better as compared to previous two in terms of contrast, noise and sharpness.

To show the use of information maps in fusion, one takes two images as shown in the first row of Fig. 7. One is an IR image and the other one is an optical image. Infrared image shows three bright spots and one human. As the information in IR image is directly coded as intensity (bright = hot), the pixel intensities rather than edge magnitude can be considered as information. The basic fusion is performed as an intensity addition [Fig. 7 (e)]. The optical image contains scene objects like trees, fence, roads, etc. The second row shows information maps of these two images. Brighter regions in the information map contribute more towards image quality and hence are used in the fusion selection criteria at various resolution levels. The ability of proposed measure is tested by ranking the images using the image quality measure[11].



(a) (b) (c)

(d) (e) (f)

Figure 7. Information maps in fusion: (a) IR image, (b) optical image, (c) information map of IR image, (d) information map of optical image, (e) intensity-based fusion, and (f) information map-based fusion.

This measure assigns a numerical score to the fused output. Table 1 shows qualitative measure of input images in Fig. 7 and their fused results. These measures quantitatively depict quality of the images. It can be observed that fusion method based on information map achieves higher score than the others. The same can be verified subjectively. This shows that fusion using information map works better, both qualitatively and quantitatively.

Table 1. Ranking of fused images (Fig. 7)

Images	MSOE weighted measure	Normalised MSOE measure
(e)	0.343	0.184
(f)	0.367	0.191

Combination of noise suppression and information map-based fusion in case of noisy input images is shown in Fig. 8, where an infrared image affected by salt and pepper noise is shown. The optical image is noise-free. Here, fusion results using evidence edge-based fusion with median filtering is shown for comparison with results using information map of median filtered images. Table 2 shows qualitative measure of these images. The score suggests better quality of information map-based fused result, which can be verified subjectively.

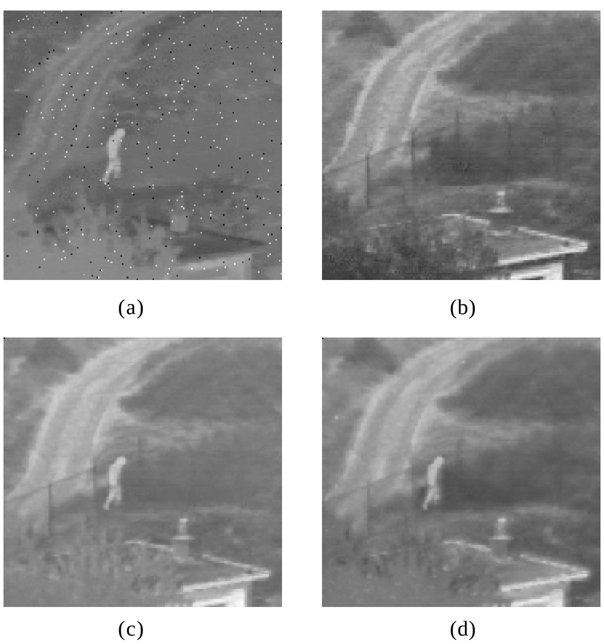


Figure 8. Combination of noise suppression and information map-based fusion: (a) IR image affected by spike noise but a human is visible, (b) optical image showing fence and other objects like trees and road clearly, (c) intensity fusion with median filtering, and (d) fusion using information map with median filtering.

Table 2. Ranking of fused images (Fig. 8)

Images	MSOE measure
(c)	0.339
(d)	0.350

5. CONCLUSIONS

In this paper, one has addressed the problem of fusion of noisy multi-sensor images. The standard methods for image fusion were put in the context of an information maximisation problem. Based on this, the extensions needed for noisy image fusion were developed. Three different approaches were presented to address three different concerns. In practice, any combination of the three may be used for getting the best results. A key contribution is the development of information maps for noisy images that do not require noiseless reference images. Fusion based on information map as decision map works well for all noisy images. This method preserves contrast details as well as edge details of the input images in the fused image. The methods are general and can be used on any kind of images, provided the objective is fusion and preservation of edges. Textures features cannot be handled at present.

ACKNOWLEDGEMENTS

The authors thank Shri N. Sitaram, Director, Centre for Artificial Intelligence & Robotics (CAIR), Bangalore, for his support and encouragement. They are also thankful to AIEL Div, ADE for UAV images. They also thank Shri M.R. Kesheorey, Dr G. Athithan and Dr N. Ramamurthy for their valuable suggestions.

REFERENCES

1. Zitova, Barbara & Flusser, Jan. Image registration methods: A survey. *Image Vision Compu.*, 2003, **21**, 977-1000.

2. Li, Hui; Manjunath, B.S., & Mitra, Sanjit K. A contour-based approach to multi-sensor image

registration. *IEEE Trans. Image Processing*, 1995, **4**, (3).

3. Li, Hui; Manjunath, B.S. & Mitra, Sanjit K. Multi-sensor image fusion using the wavelet transform. *GMIP.*, 1995, **57**, 235-45.

4. Koren, Laine. & Taylor. Image fusion using steerable dyadic wavelet transform. *ICIP.*, **C95**, 232-35.

5. Chipman, L.J.; Orr, T.M. & Graham, L.N. Wavelet and image fusion. *ICIP*., **C95**, 248-51.

6. Burt, Peter J. & Adelson, Edward H. The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.*, 1983, **31**, (4), 532-40.

7. Burt, Peter J. & Adelson, Edward H. Merging images through pattern decompositon. Applications of digital image processing. *SPIE Proceedings*, 1986, 173-81.

8. Toet, A. Image fusion by a ratio of low-pass pyramid. *Patt. Recog. Lett.*, 1989, **9**, 245-53.

9. Toet, A.; Ruyven, L.J. Van & Valeton, J.M. Merging thermal and visual images by a contrast pyramid. *Optical Engineering*, 1989, **28** (7), 789-92.

10. Yin, Zhang & Malcolm, Andrew A. Thermal and visual image processing and fusion. http://www.simtech.a-star.edu.sg/research/TechnicalReports/TR0630.pdf.

11. Rakshit, Subrata. & Mishra, Anima. Estimation of structural information content in image. *In* Proceedings of ACCV06 LNCS. 2006, **3851.** pp. 265-75.