Multiscale image fusion through guided filtering

Alexander Toet[1], Maarten A. Hogervorst
TNO, Soesterberg, The Netherlands

ABSTRACT

We introduce a multiscale image fusion scheme based on guided filtering. Guided filtering can effectively reduce noise while preserving detail boundaries. When applied in an iterative mode, guided filtering selectively eliminates small scale details while restoring larger scale edges. The proposed multi-scale image fusion scheme achieves optimal spatial consistency by using guided filtering both at the decomposition and at the recombination stage of the multiscale fusion process. First, size-selective iterative guided filtering is applied to decompose the source images into base and detail layers at multiple levels of resolution. Then, frequency-tuned filtering is used to compute saliency maps at successive levels of resolution. Next, at each resolution level a binary weighting map is obtained as the pixelwise maximum of corresponding source saliency maps. Guided filtering of the binary weighting maps with their corresponding source images as guidance images serves to reduce noise and to restore spatial consistency. The final fused image is obtained as the weighted recombination of the individual detail layers and the mean of the lowest resolution base layers. Application to multiband visual (intensified) and thermal infrared imagery demonstrates that the proposed method obtains state-of-the-art performance for the fusion of multispectral nightvision images. The method has a simple implementation and is computationally efficient.

Keywords: Guided filter, image fusion, saliency, intensified imagery, thermal imagery, nightvision, surveillance

1 INTRODUCTION

The increasing deployment and availability of co-registered multimodal imagery from different types of sensors has spurred the development of image fusion techniques in recent years[1]. The information provided by different sensors registering the same scene can either be (partially) redundant or complementary and may be corrupted with noise. Effective combinations of complementary and partially redundant multispectral imagery can therefore visualize information that is not directly evident from the individual input images. In general, image fusion aims to represent the visual information from any number of input images in a single composite (fused) image that is more informative than each of the input images alone, eliminating noise in the process while preventing both the loss of essential information and the introduction of artifacts. This requires the availability of filters that combine the extraction of relevant image details with noise reduction.

To date, a variety of image fusion algorithms have been proposed[1]. A popular class of algorithms are the multi-scale image fusion schemes, which decompose the source images into spatial primitives at multiple spatial scales, then integrate these primitives to form a new ('fused') multi-scale representation, and finally apply an inverse multi-scale transform to reconstruct the fused image[1]. However, most multi-scale transform-based image fusion methods introduce artifacts because their basic filters do not preserve spatial consistency[2]. This has led to the use of edge preserving filters to decompose source images into base and detail layers while preserving the edge information in the fusion process. Techniques that have been applied include weighted least squares filter[3], L_1 fidelity using L_0 gradient[4], L_0 gradient minimization[5], cross bilateral filter[6] and anisotropic diffusion[7]. However, most of these techniques are computationally expensive and tend to oversharpen edges, which makes them less suitable for application in multiscale schemes[8].

The recently introduced Guided Filter (GF[9]) is a computationally efficient, edge-preserving translation-variant operator based on a local linear model which avoids the drawbacks of bilateral filtering (BLF) and other previous approaches. When the input image also serves as the guidance image, the GF behaves like the edge preserving BLF. Hence, the GF can gracefully eliminate small details while recovering larger scale edges when applied in an iterative framework.

---

[1] lex.toet@tno.nl

In this paper we propose a multi-scale image fusion scheme, where iterative guided filtering is used to decompose the input images into base and detail layers at successive spatial scales, and guided filtering is used to construct the weight maps used in the recombination process.

The rest of this paper is organized as follows. Section 2 briefly discusses the principles of edge preserving filtering and introduces (iterative) guided filtering. Section 3 presents the proposed guided fusion based image fusion scheme. Section 4 presents the imagery and computational methods that were used to assess the performance of the new image fusion scheme. The results of the evaluation study are presented in Section 5. Finally, in Section 6 the results are discussed and some conclusions are presented.

2   EDGE PRESERVING FILTERING

In this section we briefly introduce the edge preserving bilateral and joint bilateral filters, show how they are related to the guided filter, and how the application of a guided filter in an iterative framework results in size selective filtering of small scale image details combined with the recovery of larger scale edges.

2.1   Bilateral filter

The bilateral filter is a non-linear filter that computes the output at each pixel as a Gaussian weighted average of their spatial and spectral distances. It prevents blurring across edges by assigning larger weights to pixels that are spatially close and have similar intensity values[10]. It uses a combination of (typically Gaussian) spatial and a range (intensity) filter kernels that perform a blurring in the spatial domain weighted by the local variation in the intensity domain. It combines a classic low-pass filter with an edge-stopping function that attenuates the filter kernel weights at locations where the intensity difference between pixels is large. Formally, at a given image location (pixel) i, the filtered output O_i is given by:

$$O_i = \frac{1}{K_i} \sum_{j \in \Omega} I_j \, f(\| i - j \|) \, g(\| I_i - I_j \|)$$ (1)

where

f is the *spatial* filter kernel (e.g., a Gaussian centered at *i*),

g is the *range* or intensity (edge-stopping) filter kernel (centered at the image value at *i*),

Ω is the spatial support of the kernel,

K_i is a normalizing factor (the sum of the $f \cdot g$ filter weights) .

Intensity edges are preserved since the bilateral filter decreases not only with the spatial distance but also with the intensity distance. Though the filter is efficient and effectively reduces noise while preserving edges in many situations, it has the undesirable property that it can reverse the intensity gradient near sharp edges[9].

In the joint (or cross) bilateral filter (JBLF) the range filter is applied to a second or guidance image G [11]:

$$O_i = \frac{1}{K_i} \sum_{j \in \Omega} I_j \cdot f(\| i - j \|) \cdot g(\| G_i - G_j \|)$$ (2)

The JBLF can prevent over- or under- blur near edges by using a related image G to guide the edge stopping range filter.

2.1 Guided filtering

A guided image filter[9] is a translation-variant filter based on a local linear model. Guided image filtering involves an input image I , a guidance image G (which may be identical to the input image), and an output image O. The two filtering conditions are (i) that the local filter output is a linear transform of the guidance image G and (ii) as similar as possible to the input image I . The first condition implies that

$$O_i = a_k G_i + b_k \qquad \forall i \in \omega_k$$ (3)

where ω_k is a square window of size $(2r+1) \times (2r+1)$. The local linear model ensures that the output image O has an edge only at locations where the guidance image G also has one, because $\nabla O = a \nabla G$. The linear coefficients a_k and

b_k are constant in ω_k. They can be estimated by minimizing the squared difference between the output image O and the input image I (the second filtering condition) in the window ω_k, i.e. by minimizing the cost function E:

$$E(a_k, b_k) = \sum_{i \in \omega_k} \left((a_k G_i + b_k - I_i)^2 + \varepsilon\, a_k^2 \right) \tag{4}$$

where ε is a regularization parameter penalizing large a_k. The coefficients a_k and b_k can directly be solved by linear regression[9]. Since pixel i is contained in several different (overlapping) windows ω_k, the value of O_i in Equation (3) depends on the window over which it is calculated. This can be accounted for by averaging over all possible values of O_i:

$$O_i = \bar{a}_i\, G_i + \bar{b}_i \tag{5}$$

where $\bar{a}_i$ and $\bar{b}_i$ are the average coefficients of all windows overlapping i. Although the linear coefficients $(\bar{a}_i, \bar{b}_i)$ vary spatially, their gradients will be smaller than those of G near strong edges (since they are the output of a mean filter). As a result we have $\nabla O \approx \bar{a}\, \nabla G$, meaning that abrupt intensity changes in the guiding image G are still largely preserved in the output image O. Equation (5) defines the guided filter. When the input image also serves as the guidance image, the guided filter behaves like the edge preserving bilateral filter, with the parameters ε and the window size r having the same effects as respectively the range and the spatial variances of the bilateral filter. The guided filter is a computationally efficient, edge-preserving operator which avoids the gradient reversal artefacts of the bilateral filter.

2.2 Iterative guided filtering

The application of the joint bilateral filter (Equation (2)) in an iterative framework results in size selective filtering of small scale details combined with the recovery of larger scale edges[12]. In such a scheme the result G^{t+1} of the t-*th* iteration is obtained from the joint bilateral filtering of the input image I using the result G^t of the previous iteration step as the guidance image:

$$G^{t+1}_i = \frac{1}{K_i} \sum_{j \in \Omega} I_j \cdot f(\| i - j \|) \cdot g(\| G^t_i - G^t_j \|) \tag{6}$$

In this scheme details smaller than the Gaussian kernel of the bilateral filter are removed while the edges of the remaining details are iteratively restored. Hence, this scheme allows the selective elimination of small scale details while preserving the remaining image structure. Note that the initial guidance image G^1 can simply be a constant (e.g., zero) valued image since it updates to the Gaussian filtered input image in the first iteration step. Here we propose to replace the bilateral filter in this scheme by a guided filter to avoid any gradient reversal artefacts.

3 PROPOSED METHOD

The proposed multiscale decomposition fusion scheme consists of the following steps:

1. Iterative guided filtering is applied to decompose the source images into base layers (representing large scale variations) and detail layers (containing small scale variations).

2. Frequency-tuned filtering[13] is used to generate saliency maps for the source images.

3. Binary weighting maps are computed as the pixelwise maximum of the individual source saliency maps.

4. Guided filtering is applied to each binary weighting map with its corresponding source as the guidance image to reduce noise and to restore spatial consistency.

5. The fused image is computed as a weighted recombination of the individual source detail layers.

In a hierarchical framework steps 1-4 are performed at multiple spatial scales. In this paper we used a 4 level decomposition obtained by filtering at 3 levels of resolution.
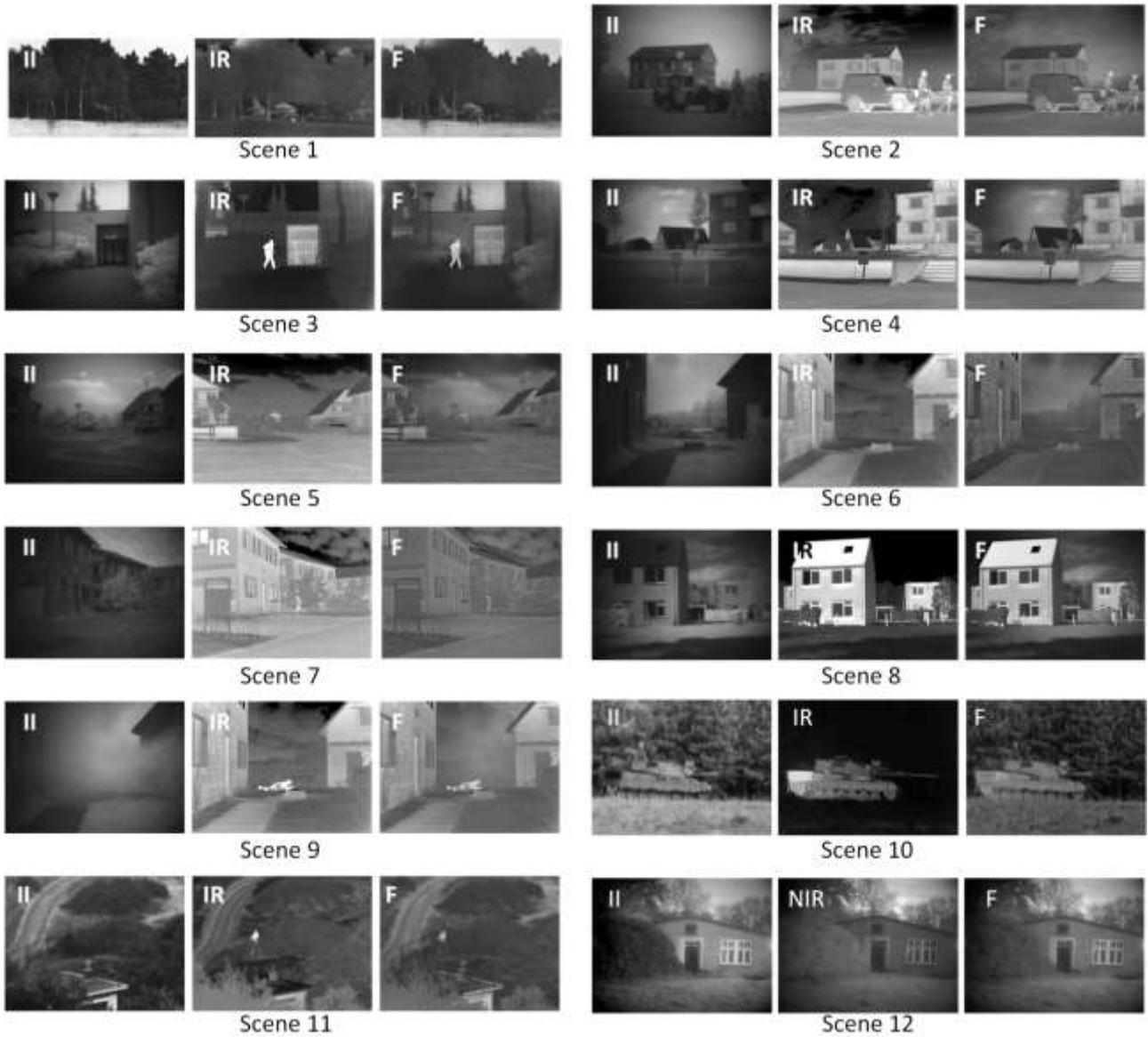
Figure 1. Original input and fused images for all 12 scenes. The intensified visual (II) , thermal infrared (IR) or near infrared (NIR: scene 12) source images together with the result of the proposed fusion scheme (F) for each of the 12 scenes used in this study.

Figure 1 shows the intensified visual (II) and thermal infrared (IR) or near infrared (NIR) images together with the results of the proposed image fusion scheme, for the 12 different scenes that were used in the present study. We will now discuss the proposed fusion scheme in more detail.

Consider two co-registered source images $X_0(x,y)$ and $Y_0(x,y)$. The proposed scheme then applies iterative guided filtering (IGF) to obtain progressively coarser image representations X_i and Y_i ($i > 0$):

$$\text{IGF}(X_i) = X_{i+1} \quad ; \quad i \in \{0,1,2\} \tag{7}$$

In this study the number of iteration steps is set to 4. By letting each lower resolution image serve as the base layer for the preceding higher resolution image the successive size-selective detail layers dX_i are simply obtained by subtraction as follows:

$$dX_i = X_i - X_{i+1} \quad ; \quad i \in \{0,1,2\} \tag{8}$$

Figure 2 shows the base and detail layers that are obtained this way for the tank scene (nr 10 in Figure 1). The edge-preserving properties of the iterative guided filter guarantee a graceful decomposition of the source images into details at different spatial scales. The filter size and regularization parameters used in this study are respectively set to $r_i = \{5, 10, 30\}$ and $\varepsilon_i = \{0.0001, 0.01, 0.1\}$ for $i = \{0,1,2\}$.

Visual saliency refers to the physical, bottom-up distinctness of image details[14]. It is a relative property that depends on the degree to which a detail is visually distinct from its background [15]. Since saliency quantifies the relative visual importance of image details saliency maps are frequently used in the weighted recombination phase of multiscale image fusion schemes [4, 16, 17]. Frequency tuned filtering computes bottom-up saliency as local multiscale luminance contrast[13]. The saliency map S for an image I is computed as

$$S(x, y) = \left\| \boldsymbol{I}_{\mu} - \boldsymbol{I}_f(x, y) \right\| \tag{9}$$

where

$\boldsymbol{I}_{\mu}$ is the arithmetic mean image feature vector,

$\boldsymbol{I}_f$ represents a Gaussian blurred version of the original image, using a 5x5 separable binomial kernel,

$\| \, \|$ is the L_2 norm (Euclidian distance), and

x, y are the pixel coordinates.

We compute saliency using frequency tuned filtering since a recent and extensive evaluation study comparing 13 state-of-the-art saliency models found that the output of this simple saliency model correlates more strongly with human visual perception than the output produced by any of the other available models[18].

In the proposed fusion scheme we first compute saliency maps S_{X_i} and S_{Y_i} for the individual source layers X_i and Y_i, $i \in \{0,1,2\}$. Binary weight maps BW_{X_i} and BW_{Y_i} are then computed by taking the pixelwise maximum of corresponding saliency maps S_{X_i} and S_{Y_i}:

$$BW_{X_i}(x, y) = \begin{cases} 1 & \text{if } S_{X_i}(x, y) > S_{Y_i}(x, y) \\ 0 & \text{otherwise} \end{cases}$$

$$\tag{10}$$

$$BW_{Y_i}(x, y) = \begin{cases} 1 & \text{if } S_{Y_i}(x, y) > S_{X_i}(x, y) \\ 0 & \text{otherwise} \end{cases}$$

The resulting binary weight maps are noisy and typically not well aligned with object boundaries, which may give rise to artefacts in the final fused image. Spatial consistency is therefore restored through guided filtering (GF) of these binary weight maps with the corresponding source layers as guidance images:

$$W_{X_i} = \text{GF}(BW_{X_i}, X_i)$$

$$\tag{11}$$

$$W_{Y_i} = \text{GF}(BW_{Y_i}, Y_i)$$

As noted before guided filtering combines noise reduction with edge preservation, while the output is locally approximately a scaled version of the guidance image. In the present scheme these properties are used to transform the binary weight maps into smooth continuous weight maps through guided filtering with the corresponding source images

as guidance images. Figure 3 illustrates the process of computing smoothed weight maps by guided filtering of the binary weight maps resulting from the pointwise maximum of the corresponding source layer saliency maps for the tank scene.

Fused detail layers are then computed as the normalized weighted mean of the corresponding source detail layers:

$$dF_i = \frac{W_{X_i} \cdot dX_i + W_{Y_i} \cdot dY_i}{W_{X_i} + W_{Y_i}} \tag{12}$$

The fused image F is finally obtained by adding the fused detail layers to the average value of the lowest resolution source layers:

$$F = \frac{X_3 + Y_3}{2} + \sum_{i=0}^{2} dF_i \tag{13}$$

By using guided filtering both in the decomposition stage and in the recombination stage, this proposed fusion scheme optimally benefits from both the multi-scale edge-preserving characteristics (in the iterative framework) and the structure restoring capabilities (through guidance by the original source images) of the guided filter. The method is easy to implement and computationally efficient.

4   METHODS AND MATERIAL

This section presents the test imagery and computational metrics used to assess the performance of the proposed images fusion scheme in comparison to existing multiscale fusion schemes.

4.1   Test imagery

Figure 1 shows the intensified visual (II) , thermal infrared (IR) or near infrared (NIR: scene 12) source images together with the result of the proposed fusion scheme (F) for each of the 12 scenes used in this study. The 12 scenes are part of the TNO Image Fusion Dataset[19].

4.2   Multiscale fusion schemes used for comparison

In this study we compare the performance of our image fusion scheme with seven other popular image fusion methods based on multiscale decomposition including the Laplacian pyramid[20], the Ratio of Low-Pass pyramid[21], the contrast pyramid[22], the filter-subtract-decimate Laplacian pyramid[23], the gradient pyramid[24], the morphological pyramid[25], the discrete wavelet transform[26-29], and a shift invariant extension of the discrete wavelet transform[26]. To allow a straightforward comparison, the number of resolution levels is set to 4 in all methods, and simple averaging is used to compute the lowest resolution base level of the fused image representation. Figures 5-9 show the results of the proposed method together with the results of other seven fusion schemes for some of the scenes used in this study (scenes 2-5 and 10).

4.3   Objective evaluation metrics

Image fusion results can be evaluated using both subjective or objective measures. Subjective methods are based on psycho-visual testing and are typically expensive in terms of time, effort, and equipment required. Also, in most cases, there is only little difference among fusion results. This makes it difficult to subjectively perform the evaluation of fusion results. Therefore, many objective evaluation methods have been developed[30, 31]. However, so far, there is no universally accepted metric to objectively evaluate the image fusion results. In this paper, we use four frequently applied computational metrics to objectively evaluate and compare the performance of different image fusion methods. The metrics we use are Entropy, the Mean Structural Similarity Index (MSSIM), Normalized Mutual Information (NMI), and Normalized Feature Mutual Information (NFMI). These metrics will be briefly discussed in the following sections.
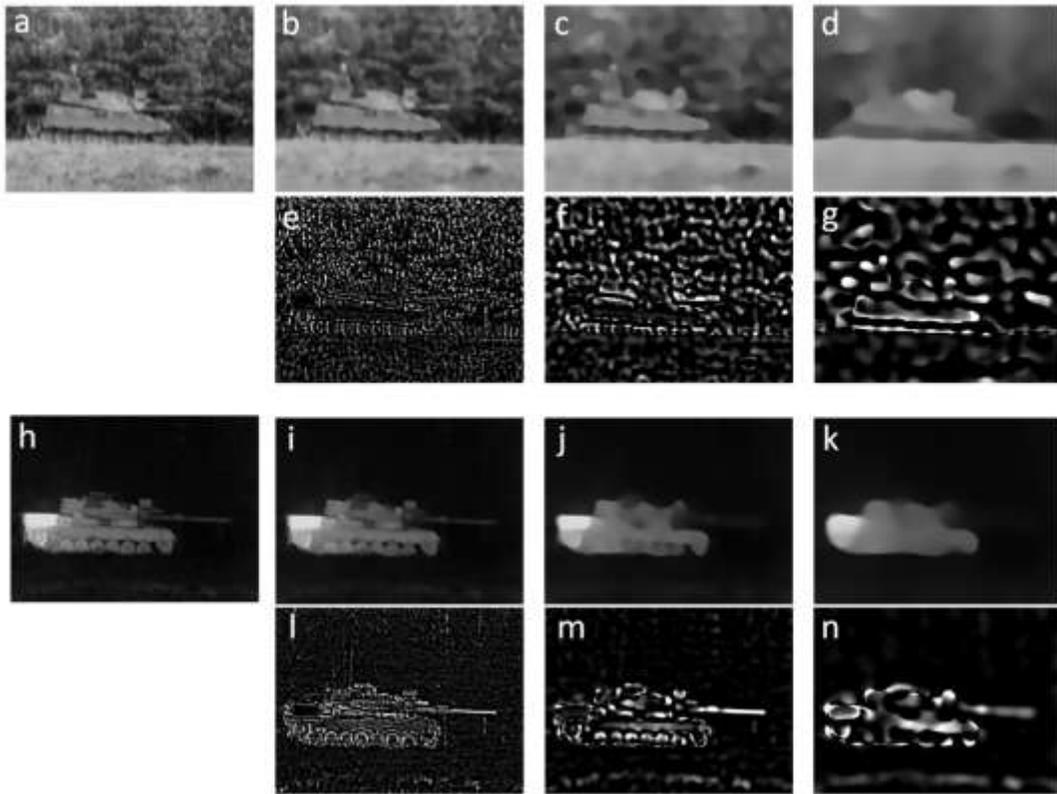
Figure 2. Base and detail layers for the tank scene. Original intensified visual (a) and thermal infrared (h) images for scene nr. 10, with their respective base (b-d and i-k) and detail (e-g and l-n) layers at successively lower levels of resolution.
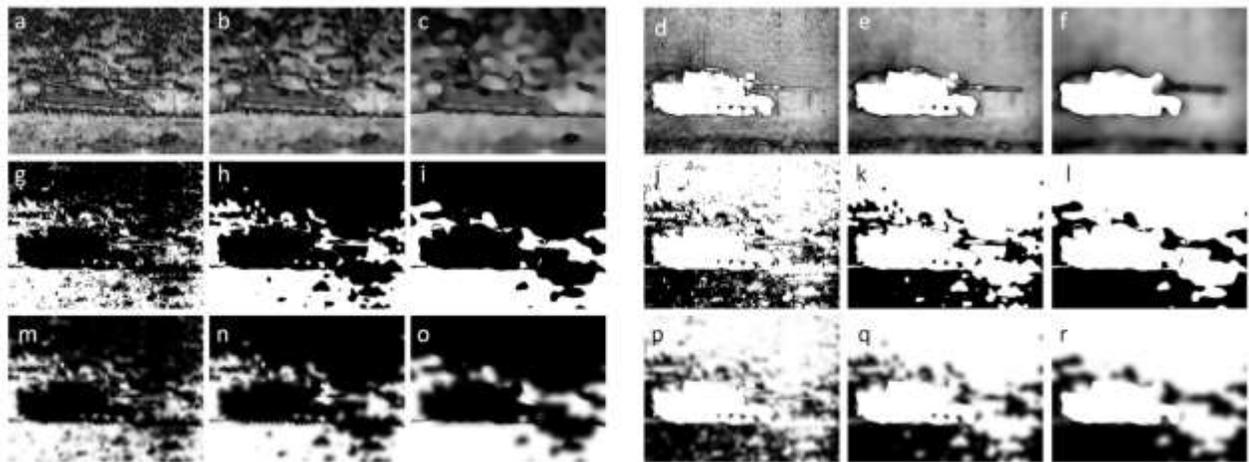

Figure 3. Computing smoothed weight maps by guided filtering of binary weight maps. Saliency maps at levels 0, 1 and 2 for respectively the intensified visual (a-c) and thermal infrared (d-f) images from Figure 2. Complementary binary weight maps for both image modalities (g-i and j-l) are obtained with a pointwise maximum operator at corresponding levels. Smooth continuous weight maps (m-o and p-r) are produced by guided filtering of the binary weight maps with their corresponding base layers as guidance images.
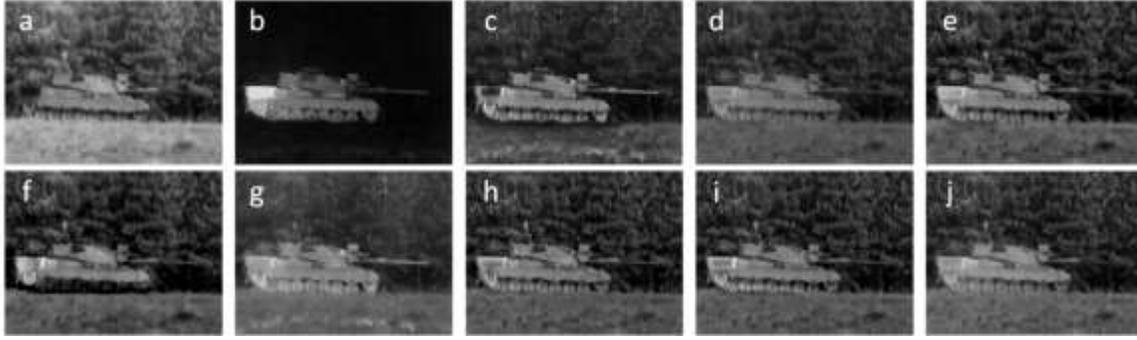
Figure 4. Comparison with existing multiscale fusion schemes. Original intensified visual (a) and thermal infrared (b) images for scene nr 10, and the fused results obtained with respectively a Contrast Pyramid (c), Gradient Pyramid (d), Laplace Pyramid (e), Morphological Pyramid (f), Ratio Pyramid (g), DWT (h), SIDWT (i), and the proposed method (j), for scene nr. 10.

Entropy

Entropy (E) is a measure of the information content in a fused image F. Entropy is defined as

$$E_F = -\sum_{i=0}^{L-1} P_F(i) \log P_F(i) \tag{14}$$

where $P_F(i)$ indicates the probability that a pixel in the fused image F has a gray value i, and the gray values range from 0 to L. The larger the entropy is, the more informative the fused image is. A fused image is more informative than either of its source images when its entropy is higher than the entropy of its source images.

Mean Structural Similarity Index

The Structural Similarity (SSIM[32]) index can be used to quantify the structural similarity between a source image A and a fused image F:

$$SSIM_{x,y} = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \cdot \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \cdot \tag{15}$$

where x and y represent local windows of size $M \times N$ in respectively A and F, and

$$\mu_x = \frac{1}{M \times N} \sum_{i=1}^{M}\sum_{j=1}^{N} x(i,j) \quad , \quad \mu_y = \frac{1}{M \times N} \sum_{i=1}^{M}\sum_{j=1}^{N} y(i,j) \tag{16}$$

$$\sigma_x^2 = \frac{1}{M \times N} \sum_{i=1}^{M}\sum_{j=1}^{N} (x(i,j)-\mu_x)^2 \quad , \quad \sigma_y^2 = \frac{1}{M \times N} \sum_{i=1}^{M}\sum_{j=1}^{N} (y(i,j)-\mu_y)^2 \tag{17}$$

$$\sigma_{xy}^2 = \frac{1}{M \times N} \sum_{i=1}^{M}\sum_{j=1}^{N} (x(i,j)-\mu_x)(y(i,j)-\mu_y) \tag{18}$$

By default, $C_1 = (0.01 \cdot L)^2$, $C2 = (0.03 \cdot L)^2$ and $C_3 = C_2 / 2$, where L is the maximal gray value. The SSIM is bounded between -1 and 1. It is typically computed over a sliding window to compare local patterns of pixel intensities that have been normalized for luminance and contrast. The Mean Structural Similarity (MSSIM) index quantifies the overall similarity between a source image A and a fused image F:

$$MSSIM_{A,F} = \frac{1}{N_w} \sum_{i=1}^{N_w} SSIM_{x_i,y_i} \tag{19}$$

where N_w represents the number of local windows of the image. An overall image fusion quality index can then be defined as the mean MSSIM values between each of the source images and the fused result:

$$MSSIM_F^{A,B} = \frac{MSSIM_{A,F} + MSSIM_{B,F}}{2} \tag{20}$$

$MSSIM_F^{A,B}$ ranges between -1 and 1.

Normalized Mutual Information

Mutual Information (MI) measures the amount of information that two images have in common. It can be used to quantify the amount of information from a source image that is transferred to a fused image[33]. The mutual information MI_{AF} between a source image A and a fused image F is defined as:

$$MI_{A,F} = \sum_{i,j} P_{A,F}(i,j) \, \log \frac{P_{A,F}(i,j)}{P_A(i)P_F(j)} \tag{21}$$

where $P_A(i)$ and $P_F(j)$ are the probability density functions in the individual images, and $P_{AF}(i,j)$ is the joint probability density function.

The traditional mutual information metric is unstable and may bias the measure towards the source image with the highest entropy. This problem can be resolved by computing the normalized mutual information (NMI) as follows[34]:

$$NMI_F^{A,B} = \frac{MI_{A,F}}{H_A + H_F} + \frac{MI_{B,F}}{H_B + H_F} \tag{22}$$

where H_A, H_B and H_F are the marginal entropy of A, B and F, and $MI_{A,F}$ and $MI_{B,F}$ represent the mutual information between respectively the source image A and the fused image F and between the source image B and the fused image F. A higher value of NMI indicates that more information from the source images is transferred to the fused image. The NMI metric varies between 0 and 1.

Normalized Feature Mutual Information

The Feature Mutual Information (FMI) metric calculates the amount of image features that two images have in common [35, 36]. This method outperforms other metrics (e.g., E, NMI) in consistency with the subjective quality measures. Previously proposed MI-based image fusion quality metrics use histograms to compute the amount of information a source and fused image have in common[33, 37]. However, image histograms contain no information about local image structure and only provide statistical measures of the number of pixels in a specific gray-level. However, since meaningful image information is contained in visual features, image fusion quality measures should measure the extent to which these visual features are transferred into the fused image from each of the source images. The Feature Mutual Information (FMI) metric calculates the mutual information between image feature maps[35, 36]. A typical image feature map is for instance the gradient map, which contains information about the pixel neighborhoods, edge strength and directions, texture and contrast. Given two source images as A and B and their fused image as F, the FMI metric first extracts feature maps of the source and fused images using a feature extraction method (e.g., gradient). Then, the feature images A', B' and F' are normalized to create their marginal probability density functions $P_{A'}$, $P_{B'}$ and $P_{F'}$. The joint probability density functions $P_{A',F'}$ and $P_{B',F'}$ are then estimated from the marginal distributions. The algorithm is described in more detail elsewhere[36]. The FMI metric between a source image A and a fused image F is then given by

$$FMI_{A,F} = MI_{A',F'} = \sum_{i,j} P_{A',F'}(i,j) \, \log \frac{P_{A',F'}(i,j)}{P_{A'}(i)P_{F'}(j)} \tag{23}$$

and the normalized feature mutual information (FMI) can be computed as follows

$$FMI_F^{A,B} = \frac{MI_{A',F'}}{H_{A'} + H_{F'}} + \frac{MI_{B',F'}}{H_{B'} + H_{F'}} \tag{24}$$

In practice the FMI is computed locally over small corresponding windows between the source and the fused images and averaged over all windows covering the image plane[35].

5   RESULTS

Here we assess the performance of the proposed image fusion scheme on the intensified visual and thermal infrared images for each of the 12 selected scenes, using Entropy, the Mean Structural Similarity Index (MSSIM), Normalized Mutual Information (NMI), and Normalized Feature Mutual Information (NFMI) as the objective performance measures. We also compare the results of the proposed method with those of seven other popular multiscale fusion schemes.

Table 1 lists the entropy of the fused result for the proposed method (IGF) and all seven multiscale comparison methods (Contrast Pyramid, DWT, Gradient Pyramid, Laplace Pyramid, Morphological Pyramid, Ratio Pyramid, SIDWT). It appears that IGF produces a fused image with the highest entropy for 9 of the 12 test scenes. Note that a larger entropy implies more edge information, but it does not mean that the additional edges are indeed meaningful (they may result from over enhancement or noise). Therefore, we also need to consider structural information metrics.

Table 2 shows that IGF outperforms all other multiscale methods tested here in terms of MSSIM. This means that the mean overall structural similarity between both source images the fused image F is largest for the proposed method.

Table 3 shows that IGF also outperforms all other multiscale methods tested here in terms of NMI. This indicates that the proposed IGF fusion scheme transfers more information from the source images to the fused image than any of the other methods.

Table 4 shows that IGF also outperforms 10 of the 12 other multiscale methods tested here in terms of NFMI. IGF is only outperformed by SIDWT for scene 1 and by the Contrast Pyramid for scene 7. This implies that fused images produced by the proposed IGF scheme typically have a larger amount of image features in common with their source images than the results of most other fusion schemes.

Summarizing, the proposed IGF fusion scheme appears to outperform the other multiscale fusion methods investigated here in most of the conditions tested.

Table 1. Entropy values for each of the methods tested and for all 12 scenes.

scene nr.	Contrast	DWT	Gradient	Laplace	Morph	Ratio	SIDWT	IGF
1	6.4818	6.4617	6.1931	6.5935	**6.6943**	6.5233	6.4406	6.5126
2	6.7744	6.6731	6.5873	6.7268	6.9835	6.7268	6.7075	**7.4233**
3	6.4340	6.5704	6.4965	6.6401	6.7032	6.6946	6.5878	**6.8589**
4	6.8367	6.8284	6.6756	7.0041	7.0906	6.7313	6.8547	**7.2491**
5	6.7549	6.6642	6.5582	6.7624	6.8618	6.5129	6.6813	**7.1177**
6	6.3753	6.3705	6.2430	6.5049	6.7608	6.2281	6.4116	**6.9044**
7	6.7470	6.3709	6.1890	6.5106	6.7445	6.3458	6.3817	**6.7869**
8	6.3229	7.3503	7.2935	7.3794	7.3501	7.4873	7.3406	**7.4891**
9	6.4903	6.4677	6.3513	6.5816	6.7295	6.3306	6.4753	**6.7796**
10	6.9627	7.0131	6.8390	7.1073	7.0530	7.0118	7.0224	**7.2782**
11	6.5442	6.4554	6.2110	6.5555	**6.8051**	6.4053	6.4572	6.2907
12	7.3335	7.3744	7.3379	7.3907	**7.4251**	7.3486	7.3746	7.3568

Table 2. MSSIM values for each of the methods tested and for all 12 scenes.

scene nr.	Contrast	DWT	Gradient	Laplace	Morph	Ratio	SIDWT	IGF
1	0.7851	0.7975	0.8326	0.8050	0.7321	0.8054	0.8114	**0.8381**
2	0.6018	0.6798	0.7130	0.6406	0.6203	0.6406	0.6935	**0.7213**
3	0.7206	0.7493	0.7849	0.7555	0.6882	0.7468	0.7629	**0.7932**
4	0.6401	0.6790	0.7162	0.6875	0.6155	0.6668	0.6949	**0.7184**
5	0.5856	0.6649	0.6938	0.6695	0.6250	0.6270	0.6769	**0.7038**
6	0.5689	0.6448	0.6755	0.6516	0.5961	0.6099	0.6598	**0.6921**
7	0.3939	0.5742	0.5994	0.5809	0.5320	0.4490	0.5889	**0.6344**
8	0.6474	0.6272	0.6630	0.6392	0.5791	0.6291	0.6463	**0.6940**
9	0.6224	0.6883	0.7224	0.6955	0.6445	0.6718	0.7089	**0.7405**
10	0.3913	0.5410	0.5715	0.5430	0.4899	0.4331	0.5513	**0.5961**
11	0.7174	0.7307	0.7754	0.7439	0.6559	0.7419	0.7539	**0.7908**
12	0.7945	0.8116	0.8466	0.8227	0.7815	0.8106	0.8365	**0.8646**

Table 3. NMI values for each of the methods tested and for all 12 scenes.

scene nr.	Contrast	DWT	Gradient	Laplace	Morph	Ratio	SIDWT	IGF
1	0.1534	0.1692	0.2052	0.1647	0.1699	0.1791	0.1796	**0.2818**
2	0.0989	0.0948	0.1158	0.0897	0.1028	0.0897	0.1028	**0.2994**
3	0.0898	0.1222	0.1493	0.1252	0.1171	0.1320	0.1280	**0.2231**
4	0.1102	0.1097	0.1322	0.1189	0.1169	0.1046	0.1177	**0.2294**
5	0.1236	0.1170	0.1379	0.1252	0.1318	0.1186	0.1251	**0.2166**
6	0.0857	0.0943	0.1162	0.0969	0.1068	0.0902	0.0980	**0.2229**
7	0.0697	0.0711	0.0839	0.0809	0.0888	0.0616	0.0781	**0.2147**
8	0.2192	0.1825	0.2198	0.1832	0.1884	0.2130	0.2021	**0.3090**
9	0.0692	0.0679	0.0781	0.0747	0.0790	0.0690	0.0731	**0.2013**
10	0.1375	0.1643	0.2043	0.1780	0.1761	0.1662	0.1760	**0.2962**
11	0.1055	0.1043	0.1177	0.1100	0.1047	0.1179	0.1115	**0.1646**
12	0.2572	0.2511	0.2746	0.2602	0.2438	0.2660	0.2649	**0.2987**

Table 4. NFMI values for each of the methods tested and for all 12 scenes.

scene nr.	Contrast	DWT	Gradient	Laplace	Morph	Ratio	SIDWT	IGF
1	0.4064	0.3812	0.3933	0.3888	0.3252	0.3498	**0.4084**	0.4008
2	0.4354	0.3876	0.4001	0.3493	0.3432	0.3493	0.4075	**0.4383**
3	0.4076	0.4081	0.4175	0.4138	0.3758	0.3552	0.4330	**0.4454**
4	0.4017	0.3913	0.4066	0.4051	0.3655	0.3497	0.4205	**0.4490**
5	0.4304	0.3971	0.4101	0.4081	0.3758	0.3497	0.4229	**0.4580**
6	0.4299	0.4074	0.4203	0.4164	0.3832	0.3570	0.4295	**0.4609**
7	**0.5050**	0.4383	0.4439	0.4357	0.3942	0.3779	0.4469	0.4286
8	0.4305	0.4074	0.4097	0.4113	0.3806	0.3553	0.4273	**0.4325**
9	0.4351	0.3959	0.4105	0.3995	0.3658	0.3539	0.4130	**0.4370**
10	0.4439	0.4251	0.4263	0.4268	0.3863	0.3465	0.4513	**0.5045**
11	0.3882	0.3798	0.3987	0.3804	0.3131	0.3453	0.4068	**0.4206**
12	0.4051	0.3725	0.3973	0.3820	0.3449	0.3635	0.4111	**0.4257**

6 DISCUSSION AND CONCLUSIONS

We propose a multiscale image fusion scheme based on guided filtering. Iterative guided filtering is used to decompose the source images into base and detail layers. Initial binary weighting maps are computed as the pixelwise maximum of the individual source saliency maps, obtained from frequency tuned filtering. Spatially consistent and smooth weighting maps are then obtained through guided filtering of the binary weighting maps with their corresponding source layers as guidance images. Saliency weighted recombination of the individual source detail layers and the mean of the lowest resolution source layers finally yields the fused image. The proposed multi-scale image fusion scheme achieves spatial consistency by using guided filtering both at the decomposition and at the recombination stage of the multiscale fusion process. Application to multiband visual (intensified) and thermal infrared imagery demonstrates that the proposed method obtains state-of-the-art performance for the fusion of multispectral nightvision images. The method has a simple implementation and is computationally efficient.

ACKNOWLEDGEMENT

REFERENCES

[1] S. Li, X. Kang, L. Fang *et al.*, "Pixel-level image fusion: A survey of the state of the art", Information Fusion, 33, 100-112 (2017).

[2] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering", IEEE Transactions on Image Processing, 22(7), 2864-2875 (2013).

[3] J. Yong, and W. Minghui, "Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter", IET Image Processing, 8(3), 183-190 (2014).

[4] G. Cui, H. Feng, Z. Xu *et al.*, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition", Optics Communications, 341, 199-209 (2015).

[5] J. Zhao, H. Feng, Z. Xu *et al.*, "Detail enhanced multi-source fusion using visual weight map extraction based on multi scale edge preserving decomposition", Optics Communications, 287, 45-52 (2013).

[6] B. K. S. Kumar, "Image fusion based on pixel significance using cross bilateral filter", Signal, Image and Video Processing, 9(5), 1193-1204 (2013).

[7] D. P. Bavirisetti, and R. Dhuli, "Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform", IEEE Sensors Journal, 16(1), 203-209 (2016).

[8] Z. Farbman, R. Fattal, D. Lischinski *et al.*, "Edge-preserving decompositions for multi-scale tone and detail manipulation", ACM Transactions on Graphics, 27(3 - Article No. 67), 1 (2008).

[9] K. He, J. Sun, and X. Tang, "Guided image filtering", IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(6), 1397-1409 (2013).

[10] C. Tomasi, and R. Manduchi, "Bilateral filtering for gray and color images". 839-846.

[11] G. Petschnigg, M. Agrawala, H. Hoppe *et al.*, "Digital photography with flash and no-flash image pairs". 664-672.

[12] Q. Zhang, X. Shen, L. Xu *et al.*, "Rolling guidance filter". Part III, 815-830.

[13] R. Achanta, S. Hemami, F. Estrada *et al.*, "Frequency-tuned salient region detection". 1597-1604.

[14] J. H. Fecteau, and D. P. Munoz, "Salience, relevance, and firing: A priority map for target selection", Trends in Cognitive Sciences, 10(8), 382-390 (2006).

[15] A. H. Wertheim, "Visual conspicuity: A new simple standard, its reliability, validity and applicability", Ergonomics, 53(3), 421-442 (2010).

[16] D. P. Bavirisetti, and R. Dhuli, "Two-scale image fusion of visible and infrared images using saliency detection", Infrared Physics & Technology, 76, 52-64 (2016).

[17] W. Gan, X. Wu, W. Wu *et al.*, "Infrared and visible image fusion with the use of multi-scale edge-preserving decomposition and guided image filter", Infrared Physics & Technology, 72, 37-51 (2015).

[18] A. Toet, "Computational versus psychophysical image saliency: a comparative evaluation study", IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(11), 2131-2146 (2011).

[19] A. Toet, "TNO Image fusion dataset", available from http://dx.doi.org/10.6084/m9.figshare.1008029 (2014).

[20] P. J. Burt, and E. H. Adelson, "The Laplacian pyramid as a compact image code", IEEE Transactions on Communications, 31(4), 532-540 (1983).

[21] A. Toet, "Image fusion by a ratio of low-pass pyramid", Pattern Recognition Letters, 9(4), 245-253 (1989).

[22] A. Toet, L. J. van Ruyven, and J. M. Valeton, "Merging thermal and visual images by a contrast pyramid", Optical Engineering, 28(7), 789-792 (1989).

[23] P. J. Burt, "Smart sensing with a pyramid vision machine", Proceedings IEEE, 76(8), 1006-1015 (1988).

[24] P. J. Burt, "A gradient pyramid basis for pattern-selective image fusion". 467-470.

[25] A. Toet, "A morphological pyramidal image decomposition", Pattern Recognition Letters, 9(4), 255-261 (1989).

[26] G. P. Lemeshewsky, "Multispectral multisensor image fusion using wavelet transforms". SPIE-3716, 214-222.

[27] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform", Computer Vision, Graphics and Image Processing: Graphical Models and Image Processing, 57(3), 235-245 (1995).

[28] S. Li, J. T. Kwok, and Y. Wang, "Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images", Information Fusion, 3(1), 17-23 (2002).

[29] P. Scheunders, and S. De Backer, "Fusion and merging of multispectral images using multiscale fundamental forms", Journal of the Optical Society of America A, 18(10), 2468-2477 (2001).

[30] S. Li, Z. Li, and J. Gong, "Multivariate statistical analysis of measures for assessing the quality of image fusion", International Journal of Image and Data Fusion, 1(1), 47-66 (2010).

[31] Z. Liu, E. P. Blasch, Z. Xue et al., "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study", IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(1), 94-109 (2012).

[32] Z. Wang, A. C. Bovik, H. R. Sheikh et al., "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, 13(4), 600-612 (2004).

[33] G. H. Qu, D. L. Zhang, and P. F. Yan, "Information measure for performance of image fusion", Electronics Letters, 38(7), 313-315 (2002).

[34] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on "Information measure for performance of image fusion"", Electronics Letters, 44(18), 1066-1067 (2008).

[35] M. Haghighat, and M. A. Razian, "Fast-FMI: Non-reference image fusion metric". 1-3.

[36] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features", Computers & Electrical Engineering, 37(5), 744-756 (2011).

[37] N. Cvejic, C. N. Canagarajah, and D. R. Bull, "Image fusion metric based on mutual information and Tsallis entropy", Electronics Letters, 42(11), 626-627 (2006).