Integrated Computer-Aided Engineering 12 (2005) 135–146
IOS Press

A multiscale approach to pixel-level image fusion

A. Ben Hamza[a], Yun He[b], Hamid Krim[c,*] and Alan Willsky[d]
[a]*Concordia Institute for Information Systems Engineering, Concordia University, Montréal, Quebec, Canada H3G 1T7*
E-mail: hamza@ciise.concordia.ca
[b]*Tality Corporation, Cary, NC 27511, USA*
E-mail: yhe@tality.com
[c]*Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695-7914, USA*
E-mail: ahk@ncsu.edu
[d]*Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA*
E-mail: willsky@mit.edu

Abstract. Pixel level image fusion refers to the processing and synergistic combination of information gathered by various imaging sources to provide a better understanding of a scene. We formulate the image fusion as an optimization problem and propose an information theoretic approach in a multiscale framework to obtain its solution. A biorthogonal wavelet transform of each source image is first calculated, and a new Jensen-Rényi divergence-based fusion algorithm is developed to construct composite wavelet coefficients according to the measurement of the information patterns inherent in the source images. Experimental results on fusion of multi-sensor navigation images, multi-focus optical images, multi-modality medical images and multi-spectral remote sensing images are presented to illustrate the proposed fusion scheme.

1. Introduction

The rapid development and deployment of new imaging sensors underline the urgency for novel image processing techniques that can effectively fuse images from different sensors into a single composite for interpretation. Fusion typically begins with two or more registered images with different representations of the same scene. They may come from different viewing conditions, or even different sensors. Image fusion of multiple sensors in a vision system could significantly reduce human/machine error in detection and recognition of objects thanks to the inherent redundancy and extended coverage. For example, fusion of forward looking infrared (FLIR) and low light television images (LLTV) obtained by an airborne sensor platform would aid a pilot to navigate in poor weather conditions or darkness.

The actual fusion process can take place at different levels of information representation. A generic categorization is to consider a process at signal, pixel, or feature and symbolic levels [12]. We focus on the so-called pixel level fusion process, where a composite image has to be synthesized from several input images. Some generic requirements must be imposed on the fusion result. The fusion process should preserve all relevant information of the input imagery in the composite image (pattern conservation). In particular, that fusion scheme should not introduce any artifacts or inconsistencies which would distract a human observer or the following processing stages (i.e. causality).

Over the past two decades, a wide variety of pixel-level image fusion algorithms have been developed.

*Corresponding author.

ISSN 1069-2509/05/$17.00

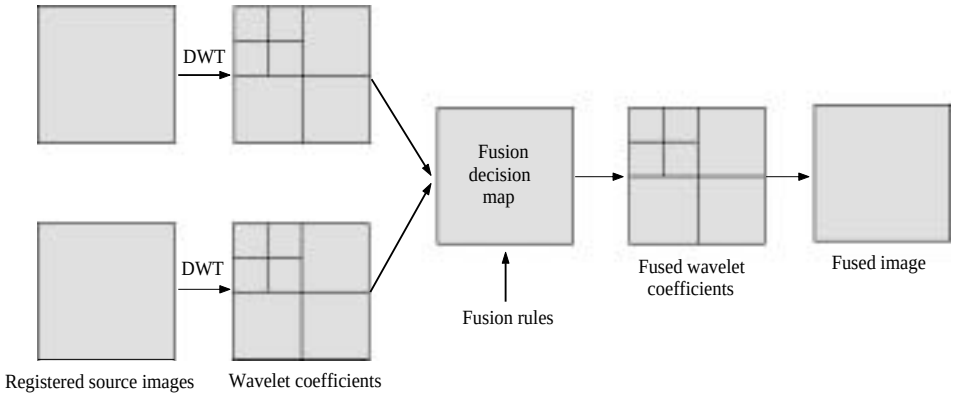

Fig. 1. A general framework for multiscale fusion with wavelet transform.

These techniques can be classified into linear superposition, logical filter [1], mathematical morphology [9], image algebra [5,15], artificial neural network [7], and simulated annealing methods [18]. All of these algorithms aimed at making fused image reveal new information concerning features that can not be perceived in individual sensor images. Some useful information has, however, been discarded since each fusion scheme tends more heavily to favor some attributes of the image over others. A detailed overview of these techniques is given in [12].

Inspired by the fact that the human visual system processes and analyzes image information at different scales, researchers have recently proposed a multiscale based fusion method and this has been widely accepted [4] as one of the most effective techniques for image fusion. A multiscale transform which may be in a form of a pyramid or wavelet transform, is first calculated for each input image, and a composite is then formed by selecting the coefficients from the multiscale representations of all source images. A fused image is finally recovered through an inverse transformation. In Burt's pioneering work [2], a Laplacian pyramid and a "choose max" rule is proposed as a model for binocular fusion in human stereo vision. For each coefficient in the pyramids of source images, the one with the maximum amplitude is copied to the composite pyramid which serves to reconstruct a fused image by its inverse pyramid transform. More recently, fusion within a gradient pyramid was shown to provide improved stability and noise immunity [3].

Wavelet theory has played a particularly important role in multiscale analysis (see [8,17] and the references therein). A general framework for multiscale fusion with wavelet transform is shown in Fig. 1. A number of papers have proposed fusion algorithms which are based on the orthogonal wavelet transform [10,19,16]. The wavelet transform success may be credited to certain advantages it offers over the Laplacian pyramid-based techniques. The wavelet bases may be chosen to be orthogonal, making the information gleaned at different resolution unique. The pyramid decomposition, on the other hand, contains redundancy between different scales. Furthermore, a wavelet image representation provides directional information in the high-low, low-high and high-high bands, while the pyramid representation fails to introduce any spatial orientation selectivity into the decomposition process. A major limitation in all recent wavelet-based fusion algorithms, however, is the absence of a good fusion criterion. Most existing selection rules are to a large extent similar to "choose max". This in turn induces a significant amount of high frequency noise introduced by a systematic and sudden inclusion of the fused maximal wavelet coefficient of a given source. This is particularly problematic, knowing the highly undesirable perception of high frequency noise by human vision.

In this paper, and as we explain below, we apply a biorthogonal wavelet transform in carrying out a pixel level image fusion. It is well known that smooth biorthogonal wavelets of compact support may be either symmetric or antisymmetric, unlike orthogonal wavelets, with the exception of the Haar wavelet. Symmetric or antisymmetric wavelets are synthesized with perfect reconstruction filters with a linear phase. This is also a very desirable property for image fusion applications. In lieu of the "choose max" type of selection rule, we propose an information theoretic fusion scheme. To each pixel in a source image is associated a vector consisting of wavelet coefficients at that position across scales. This is in turn is used to reflect the "activity" of that pixel. We refer to these indicator vectors of all

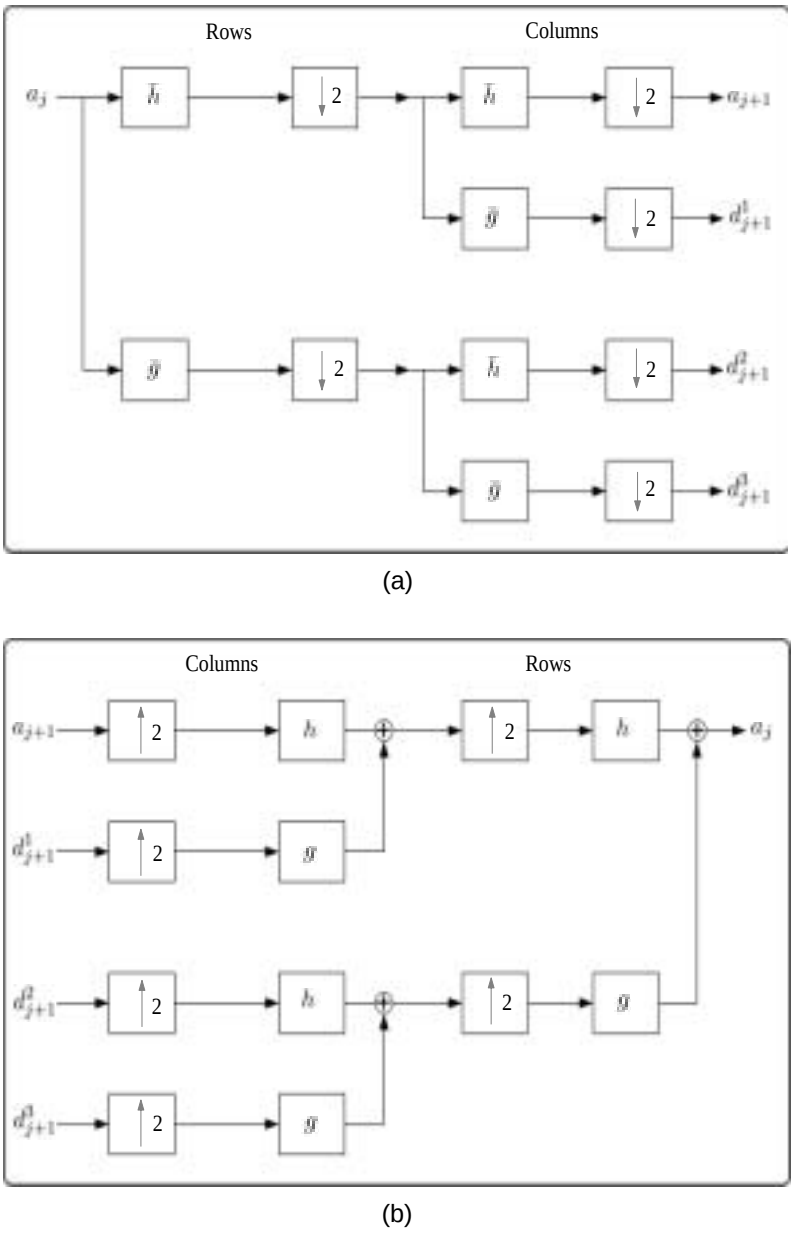

Fig. 2. A fast biorthogonal two-dimensional wavelet transform (a) and its inverse transform (b) implemented by perfect reconstruction filter banks.

the pixels in a source image as its *activity map*. A decision map is then obtained by applying an information theoretic divergence to measure all the source activity maps. To objectively perform a comparison between activity indicator vectors, we propose a recently developed Jensen-Rényi divergence divergence measure [6], which is defined in terms of Rényi entropy [14]. The wavelet coefficients of the fused image are in the end selected according to the decision map. Since all the fine to coarse scales are considered to assess the activity at a particular position within an image, our approach will clearly be more accurate, in the sense of selecting coefficients containing rich information than "select max" type of fusion schemes.

The remainder of this paper is organized as follows. In the next section and for completeness, we briefly review a biorthognal wavelet representation of an image. A concise formulation of the problem is given in

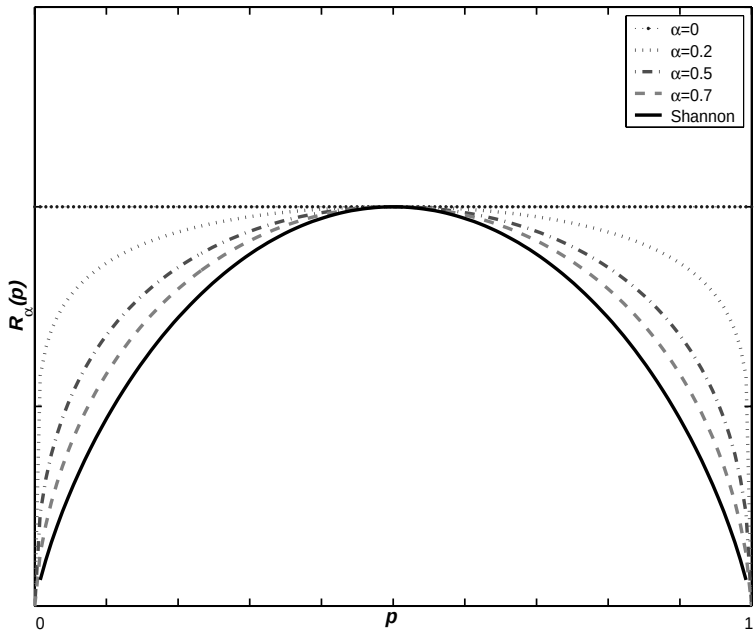

Fig. 3. Rényi's entropy.

Section 3. In Section 4, we propose an information theoretic fusion algorithm using the Jensen-Rényi divergence, and present some substantiating numerical experiments. We finally provide concluding remarks in Section 5.

2. Background on biorthogonal wavelet representation of images

Let $\{\varphi, \tilde{\varphi}\}$ and $\{\psi, \tilde{\psi}\}$ be two dual pairs of scaling functions and wavelets that generate biorthogonal wavelet basis of $\mathbf{L}^2(\mathbb{R}^2)$. For $\boldsymbol{t} = (t_1, t_2) \in \mathbb{R}^2$ and $\boldsymbol{n} = (n_1, n_2) \in \mathbb{Z}^2$, we write

$$\varphi^2_{j,\boldsymbol{n}}(\boldsymbol{t}) = \varphi_{j,n_1}(t_1)\varphi_{j,n_2}(t_2), \tag{1}$$

$$\begin{aligned}&\psi^1(\boldsymbol{t}) = \varphi(t_1)\psi(t_2),\ \psi^2(\boldsymbol{t}) = \psi(t_1)\varphi(t_2),\\ &\psi^3(\boldsymbol{t}) = \psi(t_1)\psi(t_2),\end{aligned} \tag{2}$$

$$\begin{aligned}&\tilde{\psi}^1(\boldsymbol{t}) = \tilde{\varphi}(t_1)\tilde{\psi}(t_2), \tilde{\psi}^2(\boldsymbol{t}) = \tilde{\psi}(t_1)\tilde{\varphi}(t_2),\\ &\tilde{\psi}^3(\boldsymbol{t}) = \tilde{\psi}(t_1)\tilde{\psi}(t_2).\end{aligned} \tag{3}$$

For $1 \leqslant k \leqslant 3$, we may write out

$$\psi^k_{j,\boldsymbol{n}}(\boldsymbol{t}) = \frac{1}{2^j}\psi^k\left(\frac{t_1 - 2^j n_1}{2^j}, \frac{t_2 - 2^j n_2}{2^j}\right) \tag{4}$$

and

$$\tilde{\psi}^k_{j,\boldsymbol{n}}(\boldsymbol{t}) = \frac{1}{2^j}\tilde{\psi}^k\left(\frac{t_1 - 2^j n_1}{2^j}, \frac{t_2 - 2^j n_2}{2^j}\right). \tag{5}$$

It is easy to verify [13] that

$$\{\psi^1_{j,\boldsymbol{n}}, \psi^2_{j,\boldsymbol{n}}, \psi^3_{j,\boldsymbol{n}}\}_{(j,\boldsymbol{n})\in\mathbb{Z}^3}$$

and

$$\{\tilde{\psi}^1_{j,\boldsymbol{n}}, \tilde{\psi}^2_{j,\boldsymbol{n}}, \tilde{\psi}^3_{j,\boldsymbol{n}}\}_{(j,\boldsymbol{n})\in\mathbb{Z}^3}$$

form biorthogonal bases of $\mathbf{L}^2(\mathbb{R}^2)$.

Any $f \in \mathbf{L}^2(\mathbb{R}^2)$ has two possible decompositions in these basis,

$$\begin{aligned}f &= \sum_j \sum_{\boldsymbol{n}} \sum_{k=1}^{3} < f, \psi^k_{j,\boldsymbol{n}} > \tilde{\psi}^k_{j,\boldsymbol{n}}\\ &= \sum_j \sum_{\boldsymbol{n}} \sum_{k=1}^{3} < f, \tilde{\psi}^k_{j,\boldsymbol{n}} > \psi^k_{j,\boldsymbol{n}}.\end{aligned} \tag{6}$$

Assuming we choose $\psi^k_{j,\boldsymbol{n}}$ as the analysis wavelet, at any scale 2^j, we denote the approximation coefficient by

$$a_j[\boldsymbol{n}] =< f, \varphi^2_{j,\boldsymbol{n}} >$$

and the wavelet coefficient by

$$d^k_j[\boldsymbol{n}] =< f, \psi^k_{j,\boldsymbol{n}} >,\ \ k = 1, 2, 3.$$

The three wavelets ψ^k extract image details at different scales and orientations. Over positive frequencies,

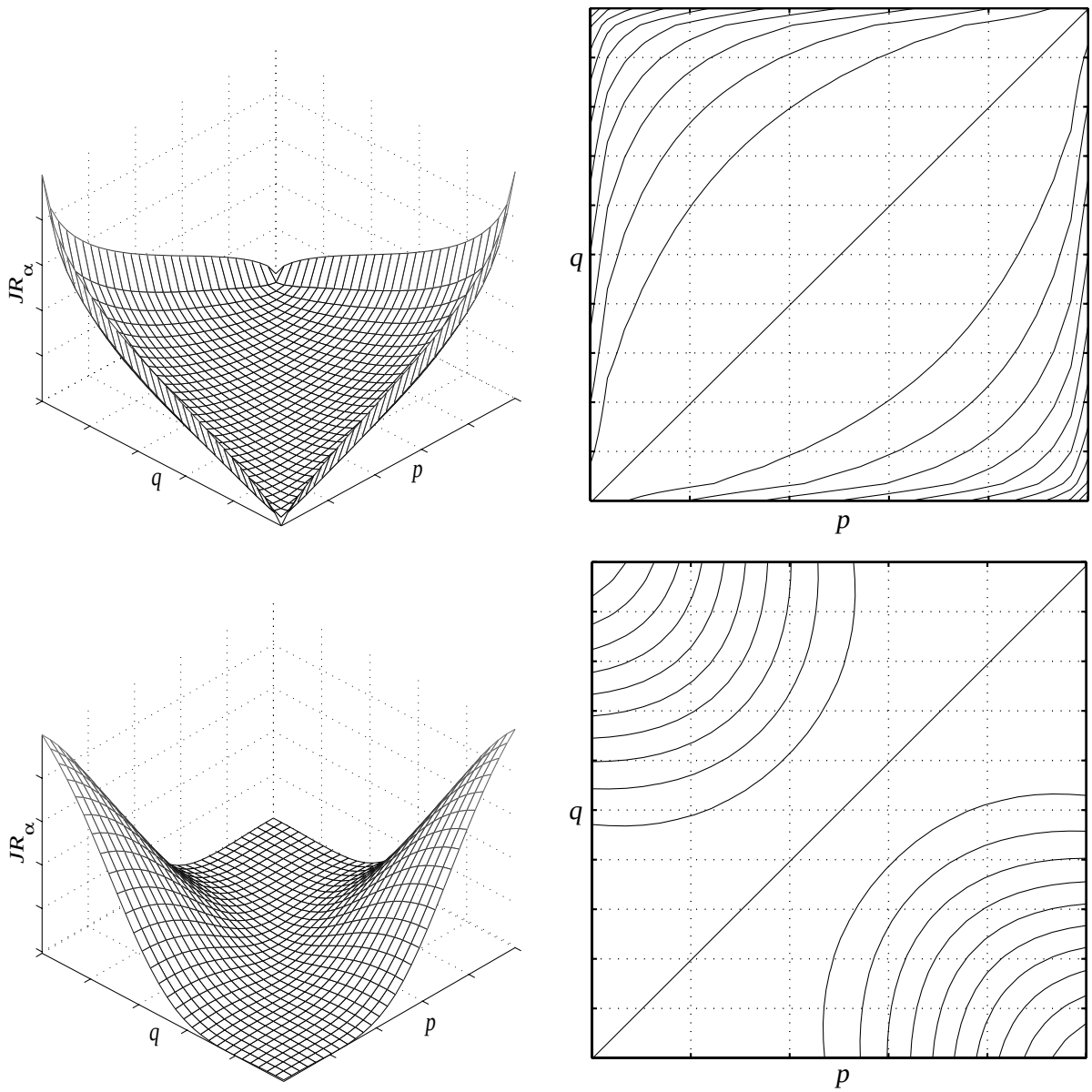

Fig. 4. 3D and contour plots of Jensen-Rényi divergence with equal weights. Top: $\alpha \in (0, 1)$. Bottom: $\alpha > 1$.

$\hat{\varphi}$ and $\hat{\psi}$ have an energy mainly concentrated respectively on lower and higher frequencies.

For $\boldsymbol{\omega} = (\omega_1, \omega_2)$, the separable wavelet expressions imply that

$$\begin{aligned} &\hat{\psi}^1(\boldsymbol{\omega}) = \hat{\varphi}(\omega_1)\hat{\psi}(\omega_2), \hat{\psi}^2(\boldsymbol{\omega}) = \hat{\psi}(\omega_1)\hat{\varphi}(\omega_2), \\ &\hat{\psi}^3(\boldsymbol{\omega}) = \hat{\psi}(\omega_1)\hat{\psi}(\omega_2). \end{aligned} \tag{7}$$

Hence $|\hat{\psi}^1(\boldsymbol{\omega})|$ is larger at low horizontal frequencies ω_1 and high vertical frequencies ω_2, $|\hat{\psi}^2(\boldsymbol{\omega})|$ is larger at high horizontal frequencies ω_1 and low vertical frequencies ω_2, whereas $|\hat{\psi}^3(\boldsymbol{\omega})|$ is larger at both high horizontal frequencies ω_1 and high vertical frequencies ω_2. As a result, wavelet coefficients calculated with ψ^1 and ψ^2 are larger along edges which are respectively horizontal and vertical, and ψ^3 produces large coefficients at the corners.

The wavelet coefficients at scale 2^{j+1} are calculated from a_j with two dimensional separable convolutions and subsamplings. Let $\{h, g\}$ and $\{\tilde{h}, \tilde{g}\}$ be the perfect reconstruction filters associated with the biorthogonal wavelet $\{\psi, \tilde{\psi}\}$. For any pair of one-

dimensional filters $y[n]$ and $z[n]$, we write the product filter $yz[\boldsymbol{n}] = y[n_1]z[n_2]$, and denote $\bar{y}[n] = y[-n]$.

For $\boldsymbol{n} = (n_1, n_2)$,

$$a_{j+1}[\boldsymbol{n}] = a_j \star \bar{h}\bar{h}[\boldsymbol{n}], \tag{8}$$

$$d^1_{j+1}[\boldsymbol{n}] = a_j \star \bar{h}\bar{g}[\boldsymbol{n}], \tag{9}$$

$$d^2_{j+1}[\boldsymbol{n}] = a_j \star \bar{g}\bar{h}[\boldsymbol{n}], \tag{10}$$

$$d^3_{j+1}[\boldsymbol{n}] = a_j \star \bar{g}\bar{g}[\boldsymbol{n}]. \tag{11}$$

A separable two dimensional convolution may be factored into one-dimensional convolutions along rows and columns of the image. The factorization is illustrated in Fig. 2(a). The rows of a_j are first convolved with $\bar{h}$ and $\bar{g}$, and subsampled by 2. The columns of these two output images are then convolved respectively with $\bar{h}$ and $\bar{g}$ and subsampled to result in four subsampled images a_{j+1}, d^1_{j+1}, d^2_{j+1} and d^3_{j+1}.

We denote

$$\breve{y}[\boldsymbol{n}] = \breve{y}[n_1, n_2] = \begin{cases} y[k_1, k_2] \text{ if } (n_1, n_2) = (2k_1, 2k_2) \\ 0 \qquad\qquad \text{otherwise.} \end{cases}$$

The image is obtained by inserting a row of zeros and a column of zeros between pairs of consecutive rows and columns of $y[n_1, n_2]$. And a_j is recovered from the coarser scale approximation a_{j+1} and the wavelet coefficients d^1_{j+1}, d^2_{j+1} and d^3_{j+1} with two-dimensional separable convolutions

$$a_j[\boldsymbol{n}] = \breve{a}_{j+1} \star \tilde{h}\tilde{h}[\boldsymbol{n}] + \breve{d}^1_{j+1} \star \tilde{h}\tilde{g}[\boldsymbol{n}] + \breve{d}^2_{j+1} \star \tilde{g}\tilde{h}[\boldsymbol{n}] + \breve{d}^3_{j+1} \star \tilde{g}\tilde{g}[\boldsymbol{n}]. \tag{12}$$

These four convolutions can also be factored into six groups of one-dimensional convolutions along rows and columns, as illustrated in Fig. 2(b).

Let $a_J[\boldsymbol{n}]$ be a digital image whose pixel interval equals $2^J = N^{-1}$. We associate to $a_J[\boldsymbol{n}]$ a function $f(\boldsymbol{x}) \in \mathbb{R}^2$ approximated at the scale 2^J,

$$a_J[\boldsymbol{n}] = < f, \varphi^2_{J,\boldsymbol{n}} > \approx \frac{1}{N} f(N^{-1}n).$$

A biorthogonal wavelet image representation of a_J of depth $L - J$ is computed by iterating Eq. (12) for $J < j \leqslant L$:

$$\{d^1_j, d^2_j, d^3_j, a_L\}_{J<j\leqslant L} \tag{13}$$

The original digital image a_J is recovered from this wavelet representation by iterating the reconstruction Eq. (12) for $J < j \leqslant L$.

3. Pre-fusion processing and problem formulation

Let $f_1, f_2, \ldots, f_m : \mathbb{Z}^2 \to \mathbb{R}$ be m digital images of the same scene taken from different sensors. For the pixel level image fusion problem, we assume all the source images are registered so that the difference in resolution, coverage, treatment of a theme, characteristics of image acquisition methods are deferred to other sources. The goal of our fusion algorithm is to construct a composite image with all crucial information captured from all the source images are combined to effectively achieve a compressed version of all source image data. To this end, we exploit an information theoretic fusion approach based on a biorthogonal wavelet representation.

Definition 1. Let $Wf_i = \{d^1_i(j, \boldsymbol{n}), d^2_i(j, \boldsymbol{n}), d^3_i(j, \boldsymbol{n}), a_i(L, \boldsymbol{n})\}_{0<j\leqslant L, \boldsymbol{n}\in\mathbb{Z}^2}$ be a biorthogonal wavelet image representation of f_i as defined in Eq. (13). Without loss of generality, we set $J = 0$. For any $\boldsymbol{n} \in \mathbb{Z}^2$, an activity pattern vector is defined as

$$A_i(\boldsymbol{n}) = \left[\sum_{k=1}^{3} |d^k_i(1, 2^{L-1}\boldsymbol{n})|^2, \sum_{k=1}^{3} |d^k_i(2, 2^{L-2}\boldsymbol{n})|^2, \ldots, \sum_{k=1}^{3} |d^k_i(L, \boldsymbol{n})|^2 \right], \tag{14}$$

which is a $(1 \times L)$ vector of energy concentrated at pixel $f_i(2^L\boldsymbol{n})$ across scales. We refer to $\{A_i(\boldsymbol{n})\}_{\boldsymbol{n}\in\mathbb{Z}^2}$ as the activity map of source image f_i.

Activity maps highlight the inherent information pattern in source images. To proceed with a scale-driven fusion of the source wavelet coefficients, it is crucial to compare the activity patterns for every pixel. If for instance the activity patterns are different in some region, an averaging process of wavelet coefficients for a composite output is unlikely to be a good choice as it yields artifacts. On the other hand, if the activity patterns are similar in a region, an averaging procedure would increase information to the fused image on account of an enhanced contribution from different sources.

A reasonable measure for activity patterns should satisfy the following properties:

- It provides a quantitative difference measure between two or more activity patterns.
- It is nonnegative and symmetric.
- It vanishes (to zero) if and only if the activity patterns are exactly the same.
- It reaches the maximum value when activity patterns are degenerate distributions.

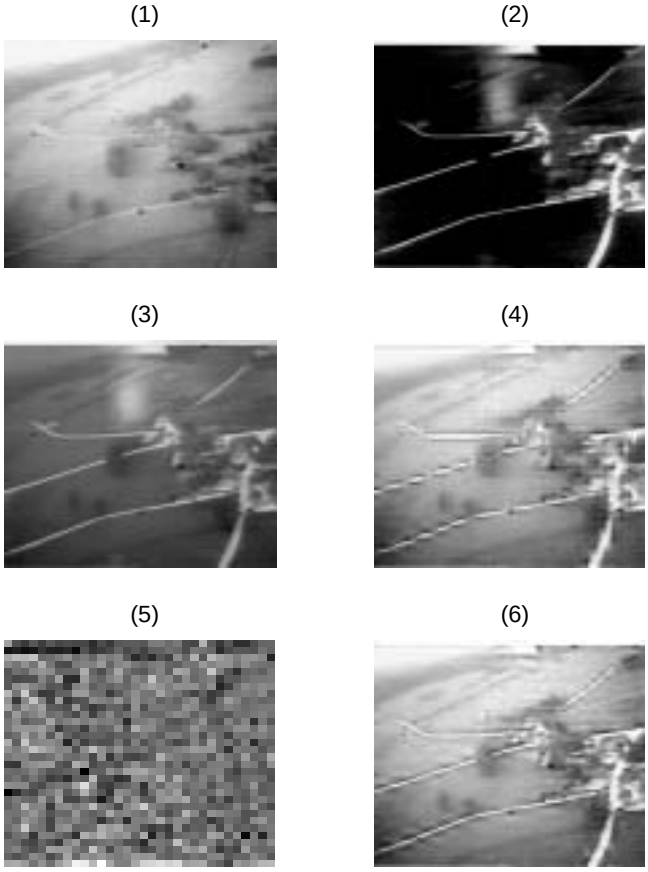

Fig. 5. Multi-sensor image fusion: (1) a low-light-television sensor image; (2) a forward-looking-infrared image; (3) fusion by averaging; (4) fusion by wavelet based maximum selection scheme; (5) a selection map; (6) fusion by the proposed information theoretic approach.

In the next section, we propose an information-theoretic measure called the Jensen-Rényi divergence, which satisfies the above requirements. A decision map is then generated by applying the Jensen-Rényi divergence to measure the coherence of source activity maps at the pixel level. We further segment the decision map into two regions, D_0 and D_1. The region D_0 is the set of pixels whose activity patterns are similar in all the source images, while the region D_1 is the set of pixels whose activity patterns are different. Our fusion technique is the solution to the following optimization problem.

$$Wf = \arg\min_{f\in\mathcal{F}}\Bigg\{\sum_{i=1}^{m}\Big(\sum_{j=1}^{L}\sum_{2^j\boldsymbol{n}\in D_0}|Wf(j,\boldsymbol{n}) - Wf_i(j,\boldsymbol{n})|^2\Big) - \sum_{j=1}^{L}\sum_{2^j\boldsymbol{n}\in D_1}|Wf(j,\boldsymbol{n})|^2\Bigg\}, \tag{15}$$

where $\mathcal{F}$ is the set of all the images f whose wavelet transform satisfies

$$\min(Wf_i(j,\boldsymbol{n})) \leqslant Wf(j,\boldsymbol{n}) \leqslant \max(Wf_i(j,\boldsymbol{n})),$$

for any $0 < j \leqslant L$ and $\boldsymbol{n} \in \mathbb{Z}^2$. This constraint ensures that the solution does not yield features beyond the scope of the source images (i.e. artifacts).

4. Information-theoretic image fusion

4.1. Jensen-Rényi divergence measure

Let $k \in \mathbb{N}$ and $X = \{x_1, x_2, \ldots, x_k\}$ be a finite set with a probability distribution $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$, i.e. $\sum_{j=1}^{k} p_j = 1$ and $p_j = P(X = x_j) \geqslant 0$, where $P(\cdot)$ denotes the probability.

Shannon's entropy is defined as $H(\boldsymbol{p}) = -\sum_{j=1}^{k} p_j \log(p_j)$, and it is a measure of uncertainty, dispersion, information, and randomness. The maximum uncertainty or equivalently minimum information is achieved by the uniform distribution. Hence, we can think of entropy as a measure of uniformity of a probability dis-

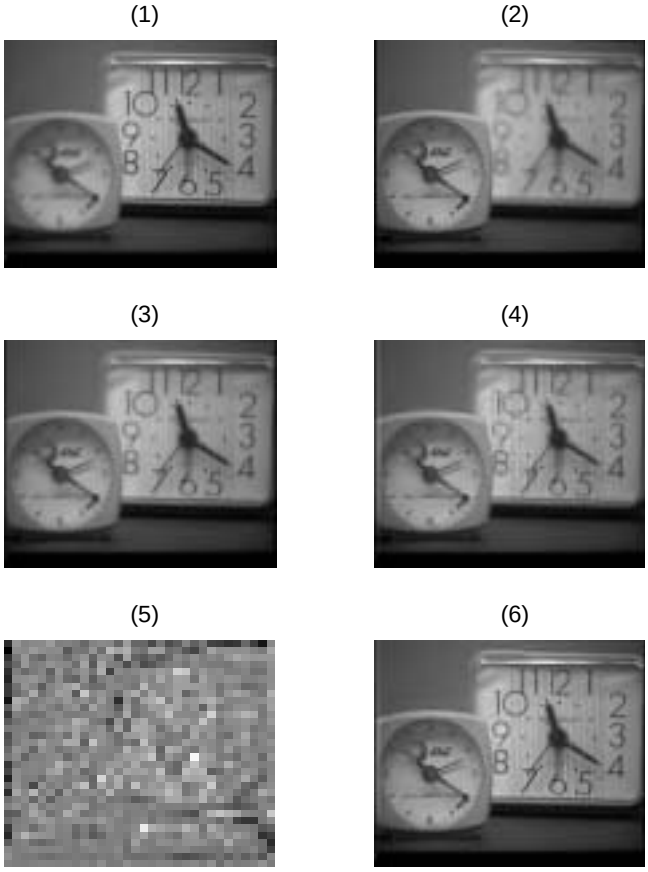

Fig. 6. Multi-focus image fusion: (1) an image focused on the larger clock; (2) an image focused on the smaller clock; (3) fusion by averaging; (4) fusion by wavelet based maximum selection scheme; (5) a selection map; (6) fusion by the proposed information theoretic approach.

tribution. Consequently, when uncertainty is higher, the degree of difficulty to predict the outcome of a draw from a probability distribution increases. A generalization of Shannon entropy is Rényi entropy [14] given by

$$R_\alpha(\boldsymbol{p}) = \frac{1}{1-\alpha}\log\sum_{j=1}^{k} p_j^\alpha, \quad \alpha \in (0,1)\cup(1,\infty). \tag{16}$$

For $\alpha > 1$, the Rényi entropy is neither concave nor convex.

For $\alpha \in (0,1)$, it is easy to see that Rényi entropy is concave, and tends to Shannon entropy $H(\boldsymbol{p})$ as $\alpha \to 1$. One may easily verify that R_α is a non-increasing function of α, and hence

$$R_\alpha(\boldsymbol{p}) \geqslant H(\boldsymbol{p}), \quad \forall \alpha \in (0,1). \tag{17}$$

When $\alpha \to 0$, Rényi entropy is equal to the logarithm of the cardinality of the set $\{j \in [1,k] : p_j > 0\}$.

Figure 3 depicts Rényi entropy for a Bernoulli distribution $\boldsymbol{p} = (p, 1-p)$, with different values of the parameter α. As illustrated in Fig. 3, the measure of uncertainty is at a minimum when Shannon entropy is used, and it increases as the parameter α decreases. Rényi entropy attains a maximum uncertainty when its exponential order α is equal to zero. Note that smaller values of α tend to emphasize probability tails.

Definition 2. Let $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ be n probability distributions. The Jensen-Rényi divergence is defined as

$$JR_\alpha^{\boldsymbol{\omega}}(\boldsymbol{p}_1,\ldots,\boldsymbol{p}_n) = R_\alpha\left(\sum_{i=1}^{n}\omega_i\boldsymbol{p}_i\right) - \sum_{i=1}^{n}\omega_i R_\alpha(\boldsymbol{p}_i),$$

where $R_\alpha(\boldsymbol{p})$ is Rényi's entropy, and $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$ be a weight vector such that $\sum_{i=1}^{n}\omega_i = 1$ and $\omega_i \geqslant 0$.

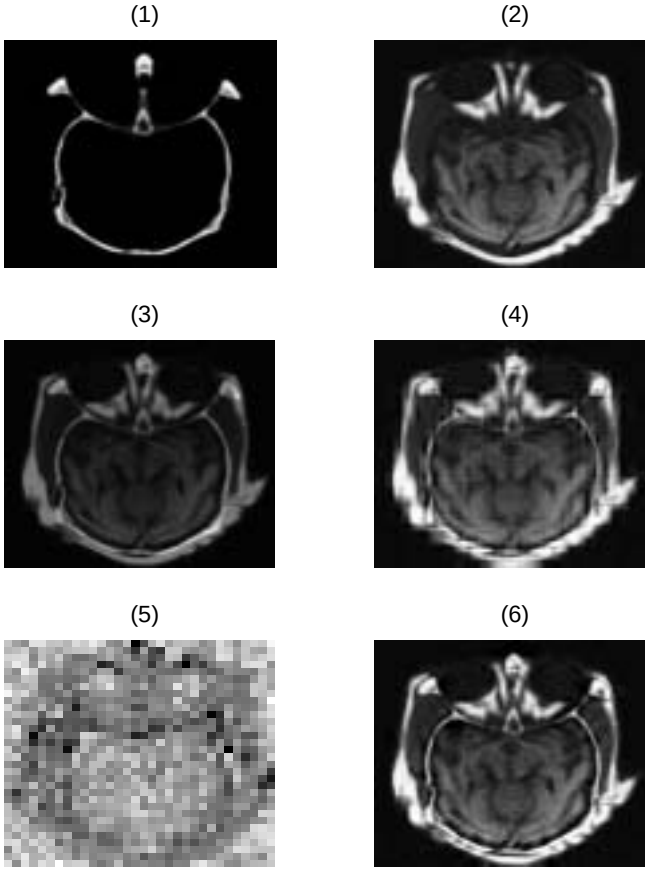

Fig. 7. Multi-modality image fusion: (1) a CT image; (2) a MRI image; (3) fusion by averaging; (4) fusion by wavelet based maximum selection scheme; (5) a selection map; (6) fusion by the proposed information theoretic approach.

Using the Jensen inequality, it is easy to check that the Jensen-Rényi divergence is nonnegative for $\alpha \in (0, 1)$. It is also symmetric and vanishes if and only if the probability distributions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ are equal, for all $\alpha > 0$.

Note that the Jensen-Shannon divergence [11] is a limiting case of the Jensen-Rényi divergence when $\alpha \to 1$.

Unlike other entropy-based divergence measures such as the Kullback-Leibler divergence, the Jensen-Rényi divergence has the advantage of being symmetric and generalizable to any arbitrary number of probability distributions or data sets, with a the additional flexibility of assigning weights to these distributions (i.e. prior beliefs). Figure 4 shows three-dimensional representations and contour plots of the Jensen-Rényi divergence with equal weights between two Bernoulli distributions for $\alpha \in (0, 1)$ and also for $\alpha \in (1, \infty)$.

In addition to its convexity property, the Jensen-Rényi divergence is shown to be an adapted measure of disparity among probability distributions, as recently demonstrated in registering Inverse Synthetic Aperture Radar (ISAR) images [6].

4.2. Image fusion with Jensen-Rényi divergence

As noted earlier, our primary goal in image fusion is to integrate complementary information from multi-sensor data with resulting fused images offering richer features and a better human visual perception.

Let $f_1, f_2, \cdots, f_m : \mathbb{Z}^2 \to \mathbb{R}$ be m digital images generated by different sensors. Our information-theoretic fusion approach first calculates a biorthogonal wavelet image representation for each f_i, then a pixel level activity map $\{A_i(\boldsymbol{n})\}_{\boldsymbol{n} \in \mathbb{Z}^2}$ is formed, as described in Section 3. Denote by $\|\cdot\|$ the ℓ_1 norm, and for any $\boldsymbol{n} \in \mathbb{Z}^2$ we define a normalized activity pattern

$$\boldsymbol{p}_i(\boldsymbol{n}) = \begin{cases} A_i(\boldsymbol{n})/\|A_i(\boldsymbol{n})\| & \text{if } \|A_i(\boldsymbol{n})\| \neq 0 \\ \Delta_1 & \text{if } \|A_i(\boldsymbol{n})\| = 0, \end{cases}$$

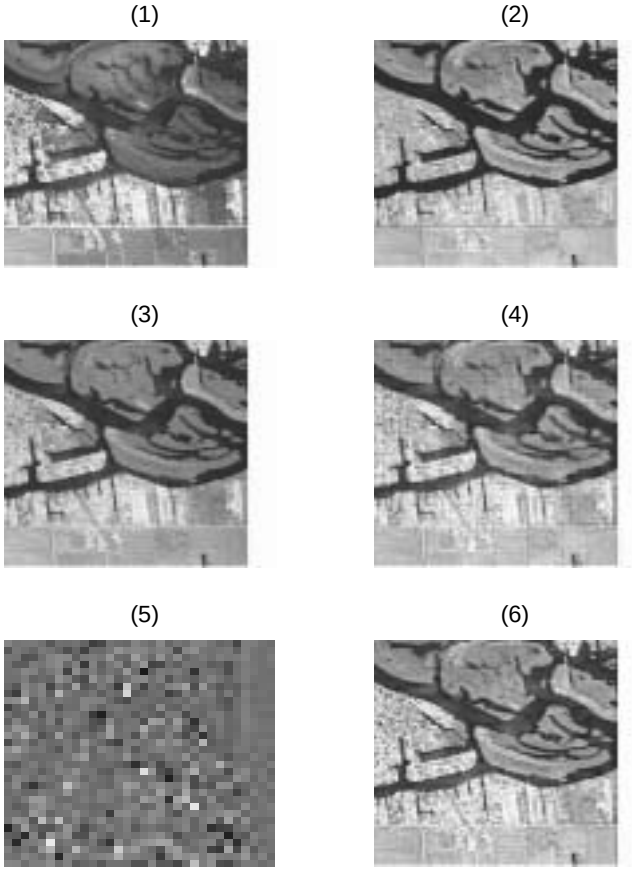

Fig. 8. Multi-spectral image fusion: (1) a high resolution remote sensing image; (2) a low resolution remote sensing image; (3) fusion by averaging; (4) fusion by wavelet based maximum selection scheme; (5) a selection map; (6) fusion by the proposed information theoretic approach.

where $\Delta_1 = [1, 0, \cdots, 0]$ is a $(1 \times L)$ degenerate distribution. To fuse the source wavelet coefficients, we compare the normalized activity patterns of all the source images in terms of Jensen-Rényi divergence, and create a selection map $\{S(\boldsymbol{n})\}_{\boldsymbol{n}\in\mathbb{Z}^2}$:

$$S(\boldsymbol{n}) = JR_{\alpha}^{\boldsymbol{\omega}}\left(\boldsymbol{p}_1(\boldsymbol{n}), \cdots, \boldsymbol{p}_n(\boldsymbol{n})\right). \tag{18}$$

The selection map is further segmented into two decision regions, D_0 and D_1. Setting T to be the mean value of a selection map, we write

$$D_0 = \{\boldsymbol{n} \in \mathbb{Z}^2 : S(\lfloor 2^{-L}\boldsymbol{n}\rfloor) < T\}$$

and

$$D_1 = \{\boldsymbol{n} \in \mathbb{Z}^2 : S(\lfloor 2^{-L}\boldsymbol{n}\rfloor) \geqslant T\}$$

where $\lfloor x \rfloor$ denotes the integer part of x.

Let f be a composite image with its wavelet coefficients

$$Wf = \{d_f^1(j,\boldsymbol{n}), d_F^2(j,\boldsymbol{n}), d_f^3(j,\boldsymbol{n}), a_f(L,\boldsymbol{n})\}_{0<j\leqslant L, \boldsymbol{n}\in\mathbb{Z}^2}$$

defined as

$$a_f^k(L,\boldsymbol{n}) = \begin{cases} (1/n)\sum_{i=1}^n a_i^k(L,\boldsymbol{n}) & \text{if } 2^L\boldsymbol{n} \in D_0 \\ \max(a_i^k(L,\boldsymbol{n})) & \text{if } 2^L\boldsymbol{n} \in D_1 \end{cases}$$

and for $k = 1, 2, 3$,

$$d_f^k(j,\boldsymbol{n}) = \begin{cases} (1/n)\sum_{i=1}^n d_i^k(j,\boldsymbol{n}) & \text{if } 2^j\boldsymbol{n} \in D_0 \\ \max(d_i^k(j,\boldsymbol{n})) & \text{if } 2^j\boldsymbol{n} \in D_1 \end{cases}$$

It can be verified that the fused image f is the solution to the optimization criteria Eq. (15).

In what follows, we present four examples, including multi-sensor navigation image fusion, multi-focus optical image fusion, multi-modality medical image fusion and multi-spectral remote sensing image fusion to illustrate the fusion scheme proposed above.

4.3. Multi-sensor image fusion

To help helicopter pilots navigate under poor visibility conditions, such as fog or heavy rain, helicopters are equipped with several imaging sensors, which are

accessible to the pilot through a helmet mounted display. A typical sensor suite includes both a low-light-television (LLTV) sensor and a thermal imaging forward-looking-infrared (FLIR) sensor. In the current configuration, the pilot can only choose one of the two sensors to watch on his display. Sample LLTV and FLIR images are shown in Figs 5(1) and 5(2) respectively. A possible improvement is to combine both imaging sources into a single fused image.

Image fusion by standard techniques such as pixel averaging and multiscale based maximum selection scheme are shown in Figs 3(3) and 3(4) respectively. Note that the pixel averaging method has a "muddy" appearance. This is due primarily to the fact that averaging results in reduced contrast areas for all the patterns that appear in only one source. On the other hand, the maximum selection scheme produces some mosaic like artifacts due to the high frequency noise introduced by sudden switches between two sets of source wavelet coefficients. Image fusion with our proposed multiscale information-theoretic approach is illustrated in Fig. 5(6). As may be seen, all the significant features from both sources are retained in the fused image without the artifacts of previous approaches.

4.4. Multi-focus image fusion

Due to the limited depth-of-focus of optical lenses, it is often not possible to get an image which contains all relevant objects 'in focus'. One possibility to overcome this problem is to take several pictures with different focus points and to combine them into a single frame which finally includes the focused regions of all input images. Figure 6 illustrates our multiscale information-theoretic fusion approach. For comparison purpose, fusion by pixel averaging and multiscale based maximum selection scheme are shown in Figs 6(3) and 6(4).

4.5. Multi-modality image fusion

With the development of new imaging methods and increasing challenges in medical applications, arises the demand for meaningful and spatial correct combination of all available image datasets. Examples for imaging devices include computer tomography (CT), magnetic resonance imaging (MRI) or the newer positron emission tomography (PET). Our multiscale information-theoretic approach is illustrated in Fig. 7(6). For comparison purpose, fusion by pixel averaging and multiscale based maximum selection scheme are shown in Figs 7(3) and 7(4).

4.6. Multi-spectral image fusion

Image fusion is often of interest in Remote sensing applications: modern spectral sensors include up to several hundred spectral bands which may be either processed and visualized individually, or may be fused into a single image, depending on the image analysis task. Image fusion of two bands from a multispectral sensor with our multiscale information-theoretic approach is illustrated in Fig. 8. For comparison purpose, fusion by pixel averaging and multiscale based maximum selection scheme are shown in Figs 8(3) and 8(4).

5. Conclusion

In this paper, we proposed a new multiscale image fusion algorithm aimed at integrating complementary information from multi-sensor data so that the fused images are enhanced and more human perception-friendly. We formulate the image fusion as an optimization problem whose solution is achieved by the proposed method.

As a first step, a biorthogonal wavelet transform of each source image is calculated to yield a scale space representation. Compared to most orthogonal wavelets, biorthogonal wavelets may be synthesized by perfect reconstruction filters with a linear phase. This is a desirable property for image fusion applications.

Unlike the "choose max" type of selection rules, our proposed technique is based on an information theoretic fusion measure. Since all the scales, from fine to coarse, are considered in assessing the activity at a particular position within an image, our approach is more accurate, as the selected source coefficients yield richer information.

We have successfully tested the new technique on fusion of multi-sensor (low-light-television and forward-looking-infrared), multi-focus, multi-modality (CT and MRI), and multi-spectral images. The presented algorithm clearly outperforms current wavelet based fusion methods in preserving significant features from a variety of sources without the common shortcomings such as artifacts.

References

[1] P. Ajjimarangsee and T.L. Huntsberger, *Neural network model for fusion of visible and infrared sensor outputs*, Proc. SPIE, 1988, pp. 153–160.

[2] P.J. Burt, *The Pyramid as a Structure for Efficient Computation. Multiresolution image processing and analysis*, A. Rosenfeld, ed., Springer-Verlag, New York, 1984.
[3] P.J. Burt, *A gradient pyramid basis for pattern selective fusion*, Proc. of the Society for Information Display Conference, 1992.
[4] P.J. Burt and R.J. Kolczynski, *Enhanced image capture through fusion*, Proc. 4th Intl. Conference on Computer Vision, 1993, pp. 173–182.
[5] S. Chen, *Stochastic image algebra for multi-sensor fusion and spatial reasoning: a neural approach*, Proc. SPIE, 1989, pp. 146–154.
[6] Y. He, A. Ben Hamza and H. Krim, A generalized divergence measure for robust image registration, *IEEE Trans. on Signal Processing* (51) (2003), 1211–1220.
[7] T.L. Huntsberger, *Data fusion: A neural networks implementation. Data fusion in robotics and machine intelligence*, M. Abidi and R. Gonzalez, eds, Academic Press, San Diego, 1992.
[8] H. Krim, W. Willinger, A. Juditski and D. Tse, Special issue on multiscale statistical signal analysis and its applications, *IEEE Trans. Infor. Theory* (45) (1999), 825–1062.
[9] J.S. Lee, *Multiple sensor fusion based on morphological processing*, Proc. SPIE, 1988, pp. 94–100.
[10] H. Li, B.S. Manjunath and S.K. Mitra, Multisensor image fusion using the wavelet transform, *Graphical Models and Image Processing* (57) (1995), 235–245.
[11] J. Lin, Divergence measures based on the shannon entropy, *IEEE Trans. Information Theory* (37) (1991), 145–151.
[12] R. Luo and M. Kay, *Data fusion and sensor integration: state of the art in 1990s. Data fusion in robotics and machine intelligence*, M. Abidi and R. Gonzalez, eds, Academic Press, San Diego, 1992.
[13] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, Academic Press Limited, London, UK, 1997.
[14] A. Rényi, *On measures of entropy and information*, Selected Papers of Alfred Renyi, 1976, pp. 525–580.
[15] G.X. Ritter, J.N. Wilson and J.L. Davidson, *Image algebra application to multi-sensor and multi-data image manipulation*, Proc. SPIE, 1988, pp. 2–7.
[16] O. Rockinger, *Pixel level fusion of image sequences using wavelet frames*, Proc. 16th Leeds Annual Statistical Research Workshop, 1996, pp. 149–154.
[17] A.S. Willsky, Multiresolution markov models for signal and image processing, *Proceedings of the IEEE* (90) (2002), 1396–1458.
[18] W.A. Wright, A markov random field approach to data fusion and color segmentation, *Image Vision Comp* (7) (1989), 144–150.
[19] D.A. Yocky, Artifacts in wavelet image merging, *Optical Eng* (35) (1996), 2094–2101.