A Categorization of Multiscale-Decomposition-Based Image Fusion Schemes with a Performance Study for a Digital Camera Application

ZHONG ZHANG AND RICK S. BLUM, SENIOR MEMBER, IEEE

The objective of image fusion is to combine information from multiple images of the same scene. The result of image fusion is a single image which is more suitable for human and machine perception or further image-processing tasks. In this paper, a generic image fusion framework based on multiscale decomposition is studied. This framework provides freedom to choose different multiscale decomposition methods and different fusion rules. The framework includes all of the existing multiscale-decomposition-based fusion approaches we found in the literature which did not assume a statistical model for the source images. Different image fusion approaches are investigated based on this framework. Some evaluation measures are suggested and applied to compare the performance of these fusion schemes for a digital camera application. The comparisons indicate that our framework includes some new approaches which outperform the existing approaches for the cases we consider.

Keywords*—*Image fusion, multiscale decomposition, multisensor fusion.

I. INTRODUCTION

There has been a growing interest in the use of multiple sensors to increase the capabilities of intelligent machines and systems. Due to this, multisensor fusion has become an area of intense research activity in the past few years [1]–[5]. Multisensor fusion refers to the synergistic combination of different sources of sensory information into one representational format. The information to be fused may come from multiple sensors monitored over a common period of time or from a single sensor monitored over an extended time period. Many sensors produce images. We use the term image fusion to denote a process generating a single image which contains a more accurate description of the scene than any of the individual source images. This fused image should be more useful for human visual or machine perception. This type of image fusion is also called pixel-level multisensor fusion [6]. The different images to be fused can come from different sensors of the same basic type or they may come from different types of sensors. The sensors used for image fusion need to be accurately coaligned so that their images will be in spatial registration. In recent years, image fusion has become an important and useful technique for image analysis and computer vision [6]–[10].

A simple image fusion method is to take the average of the source images pixel by pixel. However, along with simplicity comes several undesired side effects including reduced contrast. In recent years, many researchers recognized that multiscale transforms are very useful for analyzing the information content of images for the purpose of fusion. Multiscale representation of a signal was first studied by Rosenfeld [11], Witkin [12], and others. Researchers such as Marr [13], Burt and Adelson [14], and Linderberg [15] established that multiscale information can be useful in a number of image processing applications. More recently, wavelet theory has emerged as a well developed yet rapidly expanding mathematical foundation for a class of multiscale representations. At the same time, some sophisticated image fusion approaches based on multiscale representations began to emerge and receive increased attention. Most of these approaches were based on combining the multiscale decompositions (MSD's) of the source images. Fig. 1 illustrates the block diagram of a generic image fusion scheme based on multiscale analysis. The basic idea is to perform a multiscale transform (MST) on each source image, then construct a composite multiscale representation from these. The fused image is obtained by taking an inverse multiscale transform (IMST).

The following examples illustrate the use of image fusion in some practical applications. Fig. 2 shows a pair of digital camera images. In one image, the focus is on the Pepsi can. In the other image, the focus is on the testing card. In

Manuscript received July 15, 1998; revised February 1, 1999. This work was supported by the Office of Naval Research (ONR) under Grant N00014-97-1-0774 and by the ONR/DoD MURI Program under Contract N00014-95-1-0501.

The authors are with the Electrical Engineering and Computer Science Department, Lehigh University, Bethlehem, PA 18015 USA.

Publisher Item Identifier S 0018-9219(99)05752-7.

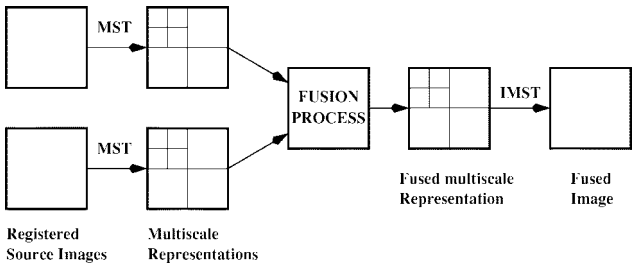


Fig. 1. Block diagram of a generic image fusion scheme.

(a) (b) (c)

Fig. 2. Fusion result for multifocus images: (a) image 1 (focus on left); (b) image 2 (focus on right); and (c) fused image (all focus).

(a) (b) (c)

Fig. 3. Fusion result for visual and MMW images: (a) image 1 (visual); (b) image 2 (MMW); and (c) fused image.

the fused image, the Pepsi can, the table, and the testing card are all in focus. Fusing images from different types of sensors is also of interest but is a considerably more complicated problem. Fig. 3 shows a pair of visual and 94-GHz millimeter-wave (MMW) images.[1] The visual image provides the outline and the appearance of the people while the MMW image shows the existence of a gun. From the fused image, there is considerable evidence to suspect that the person on the right has a concealed gun beneath his clothes. This fused image may be very helpful to a police officer, for example, who must respond promptly. The fused images in Figs. 2 and 3 were obtained using the method illustrated in Fig. 1, using techniques to be described in detail later in the paper. Concealed weapon detection is an increasingly important topic in the general area of law enforcement, and image fusion has been identified as a key technology to enable progress on this topic [16], [17]. Other application areas of image fusion include remote sensing [18], medical imaging [19], quality and defect inspection [20], drug surveillance, and intelligent robots [4].

The goal of this paper is to present a framework for classifying MSD image fusion schemes. The framework is applicable to all of the existing schemes we have encountered in an extensive literature search, and it naturally suggests some new schemes which have not yet been considered. These new schemes are found to outperform the existing schemes in many cases of interest. The framework allows one to relate different image fusion schemes and understand when each will provide better performance. The framework we present has not been discussed in the literature we encountered in our search, and this framework should encourage more directed research on the topic of image fusion. Illustrative examples are provided here under controlled conditions which allow meaningful performance comparisons of the various schemes. The performance measures used in this paper provide some quantitative comparison among different fusion schemes. Such comparisons have been lacking. While the performance measures used here may not be best for all image fusion applications, they are reasonable for fusing images for the digital camera application, which is the focus of this performance analysis section. Due to the difficult nature of fusing images from different sensor types, we avoid this in our analytical performance studies, which is one of the first treatments to quantitatively assess image fusion performance.

The framework involves five different aspects of image fusion. The most basic of these, for example, is the MSD method. In previous research, the most commonly used MSD methods for image fusion were the pyramid transform (PT) [21]–[27] and the discrete wavelet transform (DWT) [28]–[38]. In this paper, a type of shift-invariant DWT is also suggested for image fusion. This shift-invariant discrete wavelet transform uses an overcomplete version of the wavelet basis, which is called a discrete wavelet frame (DWF). This paper is one of the first to consider the DWF for image fusion. Fusion schemes based on DWF will be called DWF fusion schemes. In many cases, the DWF fusion schemes outperform the previous schemes. More detailed discussion of the existing work is delayed until we discuss our framework for MSD image fusion. Once the framework is presented, a more lucid discussion can be provided.

MSD-based fusion schemes provide much better performance than the simple methods studied previously. Their good performance appears to be due to the following facts.

- The human visual system is especially sensitive to local contrast changes, i.e., edges [23]. Apparently, rapid contrast changes contain extremely useful information for the human observer.
- MSD provides information on the magnitude of rapid contrast changes in the image.
- MSD provides both spatial and frequency domain localization.

Different combinations of MSD methods and fusion procedures yield different performance. A careful study of this issue has been lacking. In fact, there has been very little study comparing various MSD fusion algorithms, even at a basic level. Here we attempt to provide such a study. In our study, we wanted to focus mainly on fusing the MSD coefficients. For this reason, we consider only a few basic types of MSD methods in our comparisons. They are the Laplacian pyramid transform (LPT) method

[1] The source images were obtained from Thermotex Corporation.

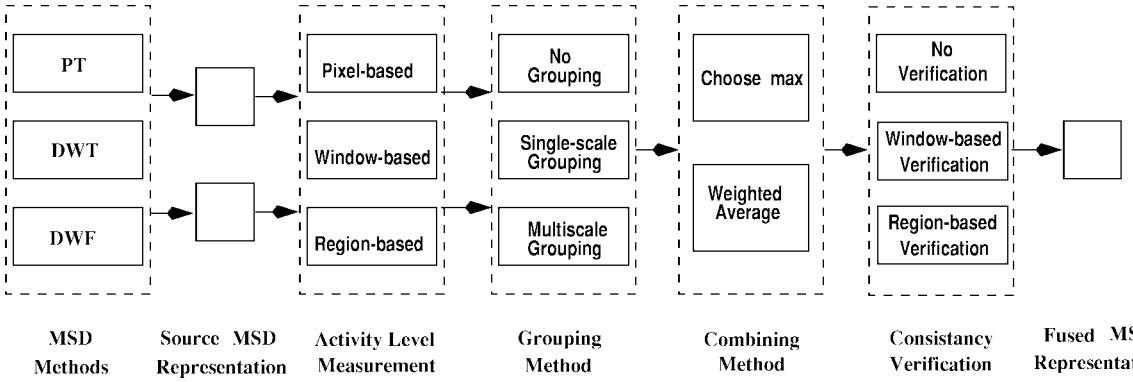


Fig. 4. The generic framework of image fusion schemes.

[14], the Daubechies's D8 orthonormal DWT [39], and the corresponding DWF. Studying these three methods should provide some general insight into how MSD methods perform and into the relative performance of pyramid, wavelet transform, and overcomplete wavelet transform approaches for image fusion. A more complete performance study of MSD methods, which might include a larger number of MSD methods, might be carried out at a later date. In our experiments, we apply two levels of MSD in every fusion scheme. We found that different levels of MSD appeared to provide different fusion performance, and so extensive study on the influence of decomposition level is another topic which should receive further study. In this paper, we only focus on MSD-based image fusion methods, where the fusion is performed at the pixel level. Other types of image fusion schemes, such as feature or decision fusion schemes, are not considered. In the image fusion approaches we studied, the fusion is performed without assuming any statistical model for the observed images. There are other approaches which focus on using stochastic models or the statistical properties of the observed images [40], [41]. These approaches are also not considered here. The performance measures we present in this paper are based on fusing images for the digital camera application, and so in our performance comparisons only visual images are considered. Fusion evaluation for schemes using other types of sensors is a more complicated issue which we will not address here. In all test cases we assume the source images to be in perfect registration. Misregistration could easily be the topic of another full paper, so we do not discuss it here for brevity.

The paper is organized as follows. Section II presents a generic framework for MSD-based image fusion schemes. Some frequently used MSD methods such as LPT, DWT, and DWF are briefly reviewed. Different alternatives in the fusion procedure are then described. Some existing image fusion schemes are also discussed within this generic framework. Section III presents some experimental results. Some performance measures are suggested and the different alternatives in the fusion procedure are compared using these performance measures. Some image fusion schemes which provide better performance than existing methods are uncovered. Section IV summarizes the paper.

II. A Generic Framework for Image Fusion Schemes

Fig. 4 illustrates a generic framework for image fusion schemes. At each sample position, we will make a decision on how the MSD representations of the source images should be used to construct the MSD representation of the fused image. This decision is based on a quantity we call the activity-level measurement. The activity-level measurement attempts to determine the quality of each source image. Some grouping methods and combining methods are also used to obtain the composite MSD representation of the fused image. A consistency verification procedure is then performed which incorporates the idea that a composite MSD coefficient is unlikely to be generated in a completely different manner from all its neighbors.

There are multiple alternatives in the procedures noted by the dashed boxes in Fig. 4. Different combinations of these alternatives lead to different fusion schemes. We will next discuss each procedure in Fig. 4 in more detail.

A. Multiscale Decomposition

A pyramid structure [14] is an efficient organization methodology for implementing multiscale representation and computation. A pyramid structure can be described as a collection of images at different scales which together represent the original image. Burt and Adelson considered an implementation of a pyramid structure called the Laplacian pyramid [14]. Each level of the Laplacian pyramid is recursively constructed from its lower level by the following four basic procedures: blurring (low-pass filtering); subsampling (reduce size); interpolation (expand in size); and differencing (to subtract two images pixel by pixel) in the order we have given. Thus a blurred and subsampled image is produced by the first two procedures at each decomposition level. These partial results, taken from different decomposition levels, can be used to construct a pyramid known as the Gaussian pyramid. In both the Laplacian and Gaussian pyramids, the lowest level of the pyramid is constructed from the original image.

In computing the Laplacian and Gaussian pyramids, the blurring is achieved using a convolution mask ω which should obey certain constraints [42]. Let $\mathbf{G}_k$ be the kth level of the Gaussian pyramid for the image $\mathbf{I}$. Then $\mathbf{G}_0 \equiv \mathbf{I}$ and

for $k > 0$

$$\mathbf{G}_k = [\omega * \mathbf{G}_{k-1}]_{\downarrow 2} \tag{1}$$

where $[\cdot]_{\downarrow 2}$ denotes downsampling of the signal by a factor of two [$G_{\downarrow 2}(i, j) = G(2i, 2j)$, $i, j = 0, \ldots, M/2-1$ for an $M \times M$ image]. The kth level of the Laplacian pyramid is defined as the weighted difference between successive levels of the Gaussian pyramid

$$\mathbf{L}_k = \mathbf{G}_k - 4\omega * [\mathbf{G}_{k+1}]_{\uparrow 2} \tag{2}$$

where $[\cdot]_{\uparrow 2}$ denotes upsampling [$G_{\uparrow 2}(2i, 2j) = G(i, j)$, $G_{\uparrow 2}(2i+1, 2j+1) = 0$, $i, j = 0, \ldots, M-1$ for an $M \times M$ image]. Here, convolution by w has the effect of interpolating the inserted zero samples.

An image can be reconstructed by the reverse procedure. Let $\hat{\mathbf{G}}$ be the recovered Gaussian pyramid. Reconstruction requires all levels of the Laplacian pyramid, as well as the top level of the Gaussian pyramid $\hat{\mathbf{G}}_N$. Thus the procedure is to set $\hat{\mathbf{G}}_N = \mathbf{G}_N$, and for $k < N$

$$\hat{\mathbf{G}}_k = \mathbf{L}_k + 4\omega * \left[\hat{\mathbf{G}}_{k+1}\right]_{\uparrow 2} \tag{3}$$

which can be used to compute $\hat{\mathbf{G}}_0$, the reconstructed version of the original image $\mathbf{G}_0$.

There are some alternative pyramid decomposition methods for image fusion, such as the gradient pyramid (GP) and the ratio-of-low-pass pyramid (RoLP). A GP for an image $\mathbf{I}$ can be obtained by applying a gradient operator to each level of the Gaussian pyramid representation. The image can be completely represented by a set of four such GP's, which represent horizontal, vertical, and the two diagonal directions [26]. The RoLP pyramid, which was introduced by Toet [23], is very similar to a Laplacian pyramid. Instead of using the difference of the successive Gaussian pyramid levels to form the Laplacian pyramid, the ratio of the successive low-pass filtered images is generated to form the so called RoLP pyramid.

The wavelet representation introduced by Mallat [39], [43], [44] suggests that, for efficiency reasons, successive layers of the pyramid should include only the additional details, which are not already available at preceding levels. Taking this approach leads to using a set of filters to perform the wavelet transform. Fig. 5 illustrates the $i+1$th stage of the two-dimensional (2-D) DWT introduced in [43], where the particular pair of analysis filters h and g correspond to a particular type of wavelet used. LL^0 is the original image. The processing is recursively applied for each decomposition level. Fig. 5 shows the processing for the $i+1$ decomposition level. Thus, in the next decomposition level, LL^{i+1} is processed to produce LL^{i+2}, LH^{i+2}, HL^{i+2}, and HH^{i+2}.

It is well known that the DWT yields a shift variant signal representation. By this we mean that a simple integer shift of the input signal will usually result in a nontrivial modification of the DWT coefficients. This shift-variant property is introduced by the subsampling process illustrated in Fig. 5. Thus, an image fusion scheme based on the DWT will also be shift dependent, which is undesirable in practice (especially considering misregistration problems). One approach to solve this problem is introduced in [45]. The solution is related to the concept of the DWF [46], [47]. In contrast to the standard DWT, in each decomposition stage, DWF uses dilated analysis filters and drops the downsampling process. Fig. 6 illustrates the $i+1$th stage of 2-D DWF.

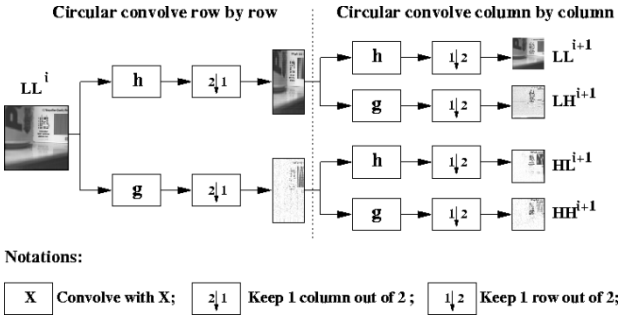


Fig. 5. One stage of 2-D DWT decomposition.

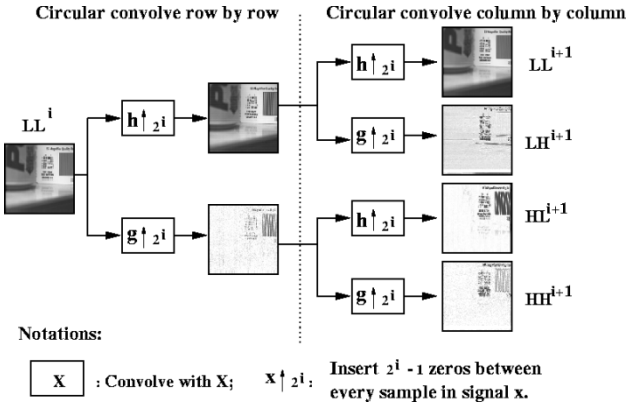


Fig. 6. One stage of 2-D DWF decomposition.

Figs. 5 and 6 show that after one stage of processing, an image is decomposed into four frequency bands: low–low (LL), low–high (LH), high–low (HL), and high–high (HH). Thus, a DWT or DWF with N decomposition levels will have $M = 3 * N + 1$ such frequency bands. The DWT will have a pyramid hierarchy. The sizes of frequency bands will decrease as the decomposition goes on. For the DWF, each frequency band will have the same size. Fig. 7 shows the 2-D structures of the above multiscale transforms with two decomposition levels. The dark blocks in different frequency bands correspond to the same group of pixels in the original image, which indicates the spatial localization of the transform. For a transform with K levels of decomposition, there is always only one low frequency band (G^K or LL^K in Fig. 7), the rest of bands are high-frequency bands in a given decomposition level, which provide detailed image information at different scales [43].

Besides the MSD analysis, another key issue in MSD-based image fusion is how to form the fused MSD representation from the MSD representations of the source images. We call the processing to achieve this goal a fusion rule. Some general alternatives for constructing a fusion rule are illustrated in Fig. 4. These include the choice of an activity-

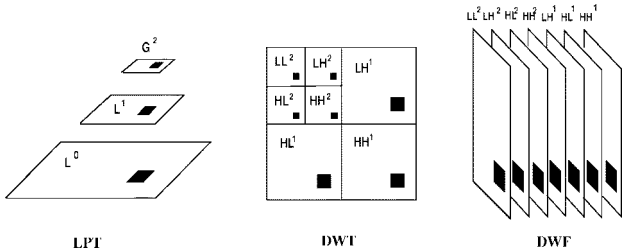


Fig. 7. MSD structures and grouping method.

level measurement, coefficient grouping method, coefficient combining method, and consistency verification method. When making fusion decisions, one common method is to select the MSD coefficient with the larger activity level. This makes the assumption that larger activity implies more information. Of course this may not always be true. An activity level may falsely judge interference as information. In some cases one may have knowledge of interference of a specific type (possibly producing high energy at specific frequencies) and it may be possible to consider this in the activity-level calculation.

To simplify the description of the different alternatives available in forming a fusion rule, we make an assumption that there are just two source images, $\mathbf{X}$ and $\mathbf{Y}$, and the fused image is $\mathbf{Z}$. We note that all the methods described in this paper can also be extended to cases with more than two source images. Generally for an image $\mathbf{I}$ we denote the MSD representation as $D_{\mathbf{I}}$ and the activity level as $A_{\mathbf{I}}$. Thus we shall encounter $D_{\mathbf{X}}$, $D_{\mathbf{Y}}$, $D_{\mathbf{Z}}$, $A_{\mathbf{X}}$, and $A_{\mathbf{Y}}$. Let $\vec{\mathbf{p}} = (m, n, k, l)$ indicate the index corresponding to a particular MSD coefficient, where m, n indicates the spatial position in a given frequency band, k the decomposition level, and l the frequency band of the MSD representation. The LPT does not use frequency bands so l is not used. Thus $D_{\mathbf{I}}(\vec{\mathbf{p}})$ and $A_{\mathbf{I}}(\vec{\mathbf{p}})$ are the MSD value and activity level of the corresponding coefficient, respectively.

B. Activity-Level Measurement

The activity level of an MSD coefficient reflects the local energy in the space spanned by the term in the expansion corresponding to this coefficient. There are three categories of methods to compute the activity level $A_{\mathbf{I}}(\vec{\mathbf{p}})$ at position $\vec{\mathbf{p}}$: coefficient-based measure; window-based measure; and region-based measure. The coefficient-based activity (CBA) measures consider each coefficient separately. The activity level is described by the absolute value or square of corresponding coefficient in the MSD representation. Here we use the absolute value

$$A_{\mathbf{I}}(\vec{\mathbf{p}}) = |D_{\mathbf{I}}(\vec{\mathbf{p}})|. \tag{4}$$

The window-based activity (WBA) measures employ a small (typically 3×3 or 5×5) window centered at the current coefficient position. We also have several alternatives here. One option is the weighted average method (WA-WBA)

$$A_{\mathbf{I}}(\vec{\mathbf{p}}) = \sum_{s\in\mathcal{S},\, t\in\mathcal{T}} \omega(s, t)|D_{\mathbf{I}}(m+s, n+t, k, l)| \tag{5}$$

where $\omega(s, t)$ is a weight and $\sum_{s\in\mathcal{S},\, t\in\mathcal{T}} \omega(s, t) = 1$, $\mathcal{S}$ and $\mathcal{T}$ are sets of horizontal and vertical indexes that describe the current window, the sums over s and t range over all samples in the window.

Another option is the rank filter method (RF-WBA). Define $\mathbf{Rank}(i)(\mathcal{Q})$ to pick the ith largest value in the set $\mathcal{Q}$. Then RF-WBA uses

$$A_{\mathbf{I}}(\vec{\mathbf{p}}) = \mathbf{Rank}(i)(\mathcal{Q}) \tag{6}$$

where $\mathcal{Q} = \{|D_{\mathbf{I}}(m+s, n+t, k, l)|\colon s \in \mathcal{S}, t \in \mathcal{T}\}$ and $\mathcal{S}$ and $\mathcal{T}$ are as defined in (5).

A popular choice is to use RF-WBA to pick the maximum value, i.e., let $i = 1$. In this way, a high activity value indicates the presence of a prominent feature in the local area. To reduce the influence of impulse noise, we may also let $i = 2$ or $i = 3$. In our experiments, we choose $i = 1$. One might also consider more general nonlinear processing of the coefficients in the window to compute activity that combines ranking and averaging [48].

The regions used in region-based activity (RBA) measurement are similar to windows with odd shapes. One RBA method was introduced in [38], where each region corresponds to an object or part of an object. In [38], a Canny edge detector [49] is applied to the low-frequency band (G^K or LL^K) of the K level MSD to produce an edge image. Region segmentation is performed on the same band using the edge information and a labeling algorithm. The output is a labeled image in which each different value represents a different region. Note that the low-frequency band of the MSD can be considered as a smoothed and subsampled version of the original image. Due to the spatial localization property, any region $\mathbf{R}^k$ in the low-frequency band has a corresponding group of coefficients in each high-frequency band. This is illustrated in Fig. 7. We define the group of coefficients in all high-frequency bands corresponding to region $\mathbf{R}^k$ as $\mathcal{C}(\mathbf{R}^k)$. The activity level of the region $\mathbf{R}^k$ in the image $\mathbf{I}$, $A_{\mathbf{I}}(\mathbf{R}^k)$, is computed as follows:

$$A_{\mathbf{I}}(\mathbf{R}^k) = \frac{1}{N_k} \sum_{\vec{\mathbf{p}}\in\mathcal{C}(\mathbf{R}^k)} A_{\mathbf{I}}(\vec{\mathbf{p}}) \tag{7}$$

where N_k is the total number of the coefficients in $\mathcal{C}(\mathbf{R}^k)$. $A_{\mathbf{I}}(\vec{\mathbf{p}})$ is obtained by CBA measurement. After all of the region activity levels are obtained, the activity level of each MSD coefficient is determined as follows. Notice that the spatial location of each coefficient in the MSD representation will be either on an edge or inside a region. For the coefficient whose spatial location is on an edge, its activity level will be measured by the CBA or the RF-WBA method. For the coefficient whose spatial location is in a region, it will take activity level of this region as its own activity measure. While we consider only (7) in our performance comparison, a slight generalization of (7) would compute activity over the region using a ranking or averaging as in (5) or (6) or some combination of these [48].

C. Coefficient Grouping Method

We notice that after the MSD process, each coefficient will have a set of corresponding coefficients in other frequency bands and other decomposition levels, as illustrated in Fig. 7 by the dark squares. All the shaded coefficients in Fig. 7 relate to the same group of pixels in the source image. For the majority of image fusion approaches, when determining the composite MSD representation, these coefficients are not associated with each other. We call these schemes no-grouping (NG) schemes. If the corresponding coefficients in the same decomposition scale are jointly constrained to take the same decision, we call this a single-scale grouping (SG) scheme. This is a more restrictive case. The most restrictive case is to consider all the corresponding MSD samples together and ensure that they are all fused the same way. We call this a multiscale grouping (MG) scheme. For the LPT-based image fusion, the NG and SG are the same since there is only one frequency band in each decomposition level.

D. Coefficient Combining Method

When combining the source MSD representations to produce the composite MSD representation, there are at least two alternatives. One is the choose-max (CM) scheme, which means just pick the coefficient with larger activity level and discard the other. If $\mathbf{Z}$ is the fused image, this can be described as $D_{\mathbf{Z}}(\vec{\mathbf{p}}) = D_{\mathbf{i}}(\vec{\mathbf{p}})$, where $\mathbf{i} = \mathbf{X}$ or $\mathbf{Y}$ depending on which source image satisfies

$$A_{\mathbf{i}}(\vec{\mathbf{p}}) = \max\left(A_{\mathbf{X}}(\vec{\mathbf{p}}),\, A_{\mathbf{Y}}(\vec{\mathbf{p}})\right). \tag{8}$$

Another combining scheme is the weighted average (WA) scheme. At each position $\vec{\mathbf{p}}$, the composite MSD can be obtained by

$$D_{\mathbf{Z}}(\vec{\mathbf{p}}) = \omega_{\mathbf{X}}(\vec{\mathbf{p}}) D_{\mathbf{X}}(\vec{\mathbf{p}}) + \omega_{\mathbf{Y}}(\vec{\mathbf{p}}) D_{\mathbf{Y}}(\vec{\mathbf{p}}). \tag{9}$$

The weights $\omega_{\mathbf{X}}$ and $\omega_{\mathbf{Y}}$ may depend on the activity levels of the source MSD coefficients and the similarity between the source images at the current position. Suppose $A_{\mathbf{X}}(\vec{\mathbf{p}}) < A_{\mathbf{Y}}(\vec{\mathbf{p}})$. One popular way to determine $\omega_{\mathbf{X}}$ and $\omega_{\mathbf{Y}}$ was suggested by Burt [26]. At first, a match measure $M_{\mathbf{XY}}(\vec{\mathbf{p}})$ is defined as a normalized correlation averaged over a neighborhood of $\vec{\mathbf{p}}$ as shown in (10) at the bottom of the page, where $\omega(\mathbf{s}, \mathbf{t})$, $A_{\mathbf{X}}$ and $A_{\mathbf{Y}}$ are as defined in (5). If $M_{\mathbf{XY}}$ is smaller than a threshold α, then $\omega_{\mathbf{X}} = 0$ and $\omega_{\mathbf{Y}} = 1$, else if $M_{\mathbf{XY}} \geq \alpha$ then

$$\omega_{\mathbf{X}} = \frac{1}{2} - \frac{1}{2}\left(\frac{1 - M_{\mathbf{XY}}}{1 - \alpha}\right) \quad \text{and} \quad \omega_{\mathbf{Y}} = 1 - \omega_{\mathbf{X}}. \tag{11}$$

We consider only these two schemes here since they are the simplest and appear most frequently in the literature. In general, nonlinear combining and ranking schemes could also be used.

Table 1
Some Frequently Used Abbreviations in This Paper

Abbreviation	Description
CBA	Coefficient-based activity level
CM	Choose-max coefficient combining
DWF	Discrete wavelet frame
DWT	Discrete wavelet thansform
LPT	Laplacian pyramid transform
MG	Multisacle gouping
NV	No consistency verification
PT	Pyramid transform
RBA	Region-based activity level
RBV	Region-based consistency verification
SG	Single-scale grouping
WA	Weighted average coefficient combining
WBA	Window-based activity measurement
RF-WBA	WBA using rank filter
WA-WBA	WBA using weighted-average
WBV	Window-based consistency verification

E. Consistency Verification

Consistency verification attempts to exploit the idea that it is very likely that a good fusion method will compute neighboring coefficients in the composite MSD in a similar manner. While it is easy to imagine applying consistency verification to other combining schemes, we applied this only to CM combining schemes in our tests. For CM combining, consistency verification is especially simple—it ensures that a composite MSD coefficient does not come from a different source image from most or all of its neighbors.

Li [37] applied consistency verification using a majority filter. Specifically if the center composite MSD coefficient comes from image $\mathbf{X}$ while the majority of the surrounding coefficients come from image $\mathbf{Y}$, the center sample is then changed to come from image $\mathbf{Y}$. In [37], 3×3 or 5×5 windows are used for the neighborhood. We call this method window-based verification (WBV).

In our own research, we employed the region-based verification (RBV) approach. An edge image and a region image are first generated as described in Section II-B. We distinguish the MSD coefficients as edge samples and region samples based on their corresponding position in the edge and region images. For region samples we apply the majority filter over the corresponding region. For the edge samples, we apply a constraint to the majority filter. The decision on the center sample will not be changed unless this decision is different from that of all the neighboring samples.

F. Existing Schemes in the Framework

We have now discussed each procedure in the image fusion framework illustrated in Fig. 4. Table 1 summarizes the abbreviations we introduced in early sections, which will be frequently used in the following sections. Table 2 illustrates

$$M_{\mathbf{XY}}(\vec{\mathbf{p}}) = \frac{2 \sum_{s \in \mathcal{S},\, t \in \mathcal{T}} \omega(s, t) D_{\mathbf{X}}(m+s, n+t, k, l) D_{\mathbf{Y}}(m+s, n+t, k, l)}{A_{\mathbf{X}}^2(\vec{\mathbf{p}}) + A_{\mathbf{Y}}^2(\vec{\mathbf{p}})} \tag{10}$$

Table 2
Some Existing Fusion Schemes and Their Specification in the Image Fusion Framework

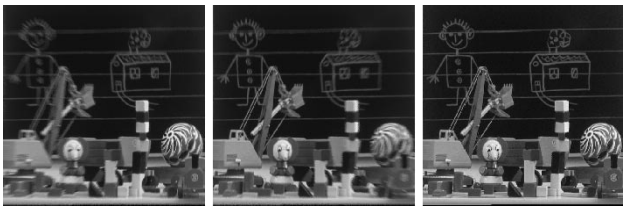
Reference	MSD Method	Activity Measure	Grouping	Combining	Verification
[21, 22]	PT (LP)	CBA	NG	CM	NV
[23, 24, 25]	PT (RoLP)	CBA	NG	CM	NV
[26, 27]	PT (GP)	WA-WBA	NG	WA	NV
[28, 19, 30 ,32]	DWT	CBA	NG	CM	NV
[31, 35]	DWT	CBA	NG	CM	WBV
[33, 36]	DWT	WA-WBA	NG	WA	NV
[34]	DWT	WA-WBA	SG	CM	NV
[37]	DWT	RF-WBA	NG	CM	WBV
[38]	DWT	RBA	MG	CM	WBV

some existing image fusion schemes and shows how they can be described by our framework. Some new approaches can also be derived from this framework. This can be achieved by choosing different alternatives for some procedures. In the following section, we study performance of different fusion schemes based on the generic framework.

III. Performance Comparison for Different Fusion Schemes

In the past, image fusion results were usually evaluated visually. Quantitatively assessing the performance in practical applications is a complicated issue because the ideal composite images is normally unknown. One way to achieve this in a controlled setting is to generate pairs of distorted source images from a known test image, then compare the fused image and the original test image. This approach is reasonable to judge performance for cases where the source images come from the same type of sensors, which is our focus here. In our tests we consider only visual images, which one might argue are the most prevalent and basic image type. The tests are particularly relevant for the digital camera application we discussed in Section I. We designed the following method for our experiments.

From a test image, two out-of-focus images are created by radial blurring. The blurring is taken along a radial line, centered at a point we selected. This centering point will be the clearest part of the image, just like the focus point of a camera. Thus, we can take blurred images with different focus points as the source images. We take the original image as the reference image. To accomplish the blurring, a Gaussian smoothing kernel was used as a convolution mask. The form of the filter was changed as a function of the spatial location where mask is applied. In particular, the standard deviation of the smoothing kernel was increased in proportion to the distance between the current pixel and the selected centering point. Fig. 8 shows an example. Sixteen different reference images were employed in our tests. These images are shown in Fig. 9. In our results we present average performance over the 16 test cases.



(a) (b) (c)

Fig. 8. Radial blurred images and the reference image: (a) image 1 (focus on right); (b) image 2 (focus on left); and (c) reference image (all focus).

Fig. 9. The 16 test images.

A. Evaluation Criteria

Three evaluation criteria were used.

- The root mean square error (RMSE) between the reference image $\mathbf{R}$ and the fused image $\mathbf{Z}$ (i, j denote pixel location)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}[\mathbf{R}(i,j)-\mathbf{Z}(i,j)]^2}{N^2}}. \quad (12)$$

- The mutual information (MI) between the reference image $\mathbf{R}$ and the fused image $\mathbf{Z}$

$$\text{MI}(\mathbf{R};\mathbf{Z}) = \sum_{i_1=1}^{L}\sum_{i_2=1}^{L} h_{\mathbf{R},\mathbf{Z}}(i_1,i_2)\log_2\frac{h_{\mathbf{R},\mathbf{Z}}(i_1,i_2)}{h_{\mathbf{R}}(i_1)h_{\mathbf{Z}}(i_2)} \quad (13)$$

where $h_{\mathbf{R},\mathbf{Z}}$ indicates the normalized joint graylevel histogram of images $\mathbf{R}$ and $\mathbf{Z}$, $h_{\mathbf{R}}$, $h_{\mathbf{Z}}$ are the normalized marginal histograms of the two images, and L is the number of graylevels.

- The percentage of correct decisions

$$P_c = \frac{N_c}{N_t}\times 100 \quad (14)$$

where N_c is the number of correct decisions and N_t is the total number of decisions. For CM combining, N_t is the size of MSD representation. For the WA combining, both N_c and N_t are counted based only on those cases when $\omega_{\mathbf{X}}$ in (9) is equal to zero or one. In our examples, $\omega_{\mathbf{X}}$ was zero or one very often (half the time or more).

Table 3
Performance of the Schemes Using Different MSD Methods

Combinations		MSD	$\overline{RMSE}$	$\overline{MI}$	$\overline{P_c}(\%)$
1	CBA +	LPT	4.09	3.79	55.9
	NG+	DWT	4.99	3.36	54.5
	CM+NV	DWF	4.39	3.59	55.9
2	WA-WBA+	LPT	4.35	3.94	59.7
	NG+	DWT	4.77	3.46	57.9
	WA+NV	DWF	4.41	3.79	61.2
3	RF-WBA+	LPT	2.89	5.23	71.8
	MG+	DWT	3.55	4.46	72.3
	CM+NV	DWF	3.08	4.87	74.7
4	RBA+	LPT	2.88	4.95	75.2
	MG+	DWT	2.72	5.02	78.3
	CM+WBV	DWF	2.67	5.21	81.2

Table 4
The Performance of the Schemes Using Different Activity-Level Measures

Combinations		Activity level	$\overline{RMSE}$	$\overline{MI}$	$\overline{P_c}(\%)$
1	LPT+	CBA	4.03	3.87	56.7
	NG+	RF-WBA	3.77	4.03	62.9
	WA+	WA-WBA	4.35	3.94	59.7
	NV	RBA	3.39	4.52	65.9
2	DWT+	CBA	4.72	3.63	62.7
	MG+	RF-WBA	3.55	4.46	72.3
	CM+	WA-WBA	3.43	4.51	70.8
	NV	RBA	2.72	5.02	78.3
3	DWF+	CBA	3.17	4.20	69.6
	MG+	RF-WBA	2.82	5.14	79.3
	CM+	WA-WBA	2.98	5.17	78.6
	WBV	RBA	2.67	5.21	81.2

Table 5
Performance of the Schemes Using Different Grouping Decision Methods

Combinations		Grouping	$\overline{RMSE}$	$\overline{MI}$	$\overline{P_c}(\%)$
1	LPT+	NG	4.09	3.79	55.9
	CBA+	SG	4.09	3.79	55.9
	WA+NV	MG	3.47	4.52	61.7
2	DWT+	NG	4.81	3.43	57.3
	RF-WBA+	SG	4.56	3.53	63.9
	CM+WBV	MG	3.42	4.71	74.2
3	DWT+	NG	4.77	3.46	57.9
	WA-WBA+	SG	4.61	3.55	61.3
	WA+NV	MG	3.46	4.49	70.5
4	DWF+	NG	4.39	3.59	55.9
	CBA+	SG	4.09	3.78	57.7
	CM+NV	MG	3.35	4.01	67.1

Table 6
Performance of the Schemes Using Different Combining Methods

Combinations		Comb.	$\overline{RMSE}$	$\overline{MI}$	$\overline{P_c}(\%)$
1	LPT+NG+NV	CM	4.01	3.91	60.5
	+WA-WBA	WA	4.35	3.94	59.7
2	DWT+MG+NV	CM	3.55	4.46	72.3
	+RF-WBA	WA	3.63	4.41	71.1
3	DWT+MG+NV	CM	3.43	4.51	70.8
	+WA-WBA	WA	3.46	4.49	70.5
4	DWF+MG+NV	CM	3.35	4.01	67.1
	+CBA	WA	3.41	4.14	68.4

Table 7
Performance of the Schemes Using Different Consistency Verification Methods

Combinations		Verif.	$\overline{RMSE}$	$\overline{MI}$	$\overline{P_c}(\%)$
1	LPT+	NV	4.01	3.91	60.5
	WA-WBA+	WBV	3.92	3.98	60.9
	NG+CM	RBV	3.17	4.88	67.8
2	DWT+	NV	3.55	4.46	72.3
	RF-WBA+	WBV	3.42	4.71	74.2
	MG+CM	RBV	2.69	5.25	78.8
3	DWF+	NV	3.08	4.87	74.7
	RF-WBA+	WBV	2.82	5.14	79.3
	MG+CM	RBV	2.58	5.22	81.2

B. Experimental Results

We compared the different alternatives for each procedure of the generic framework described in Fig. 4, and the results are shown by Tables 3–7. In these tables, the first column shows the combinations of alternatives for the procedures other than the one under consideration, and the second column lists the different alternatives for the current procedure. Columns 3–5 show the performance using the criteria we introduced in Section III-A. All the abbreviations are described in Table 1.

1) Comparing Schemes Using Different MSD Methods: To compare the performance of schemes using different MSD methods, we considered some different combinations of alternatives for the other fusion procedures for each of the LPT, DWT, and DWF MSD methods. The results are shown in Table 3. Recall that all the results in this section show the performance average over the 16 images in Fig. 9.

From the results in Table 3, we observe that different MSD methods appear to provide different image fusion performance. Without considering spatial and scale grouping and consistency verification, LPT-based schemes provide best performance in terms of $\overline{\text{RMSE}}$ and $\overline{\text{MI}}$. This is interesting since one could argue that the $\overline{P_c}$ criteria is not as informative as $\overline{\text{RMSE}}$ and $\overline{\text{MI}}$. In particular, $\overline{P_c}$ always weights errors equally regardless of their importance. In fact some errors may cause unnoticeable difference in the fused image. When spatial and scale grouping and consistency verification are considered, we see from Table 3 that the LPT schemes are no longer best. These observations appear to be related to the good spatial localization of the LPT approach in the cases we considered. By this we mean that, in the LPT case, each pixel in the reconstructed image is significantly affected by only a few coefficients, so that an error in a given coefficient has less of an effect on the fused image. In the DWT and DWF, an error in a given coefficient has a greater effect on the final image, but grouping and consistency verification generally make these errors much less likely. We see from Table 3 that the DWF and DWT-based schemes appear to outperform the LPT-based schemes when spatial grouping, scale grouping, and consistency verification are used.

2) Comparing Schemes Using Different Activity-Level Measures: To compare the performance of schemes using different activity-level measurements, we tested CBA, RF-WBA, WA-WBA, and RBA measurements with different combinations of alternatives for the other fusion procedures. The results are shown in Table 4. From these results, we observe that applying spatial grouping in computing the activity-level measurement appears to improve fusion

Table 8
Some Existing Image Fusion Schemes and Their Performance

Schemes	Combinations of fusion method	References	$\overline{RMSE}$	$\overline{MI}$	$\overline{P_c}(\%)$
1	LPT+CBA+NG+CM+NV	[21, 22]	4.09	3.79	55.9
2	DWT+CBA+NG+CM+NV	[28, 29, 30, 32]	5.04	3.33	54.7
3	DWT+CBA+NG+CM+WBV	[31, 35]	4.83	3.45	56.7
4	DWT+WA-WBA+NG+WA+NV	[33, 36]	4.77	3.46	57.9
5	DWT+WA-WBA+SG+CM+NV	[34]	4.65	3.52	60.6
6	DWT+RF-WBA+NG+CM+WBV	[37]	4.81	3.43	57.3
7	DWT+RBA+MG+CM+WBV	[38]	2.72	5.02	78.3

Table 9
Some New Image Fusion Schemes and Their Performance

MSD	Activity Measure	Grouping	Combining	Verification	$\overline{RMSE}$	$\overline{MI}$	$\overline{P_c}(\%)$
LPT	RF-WBA	MG	CM	RBV	2.74	5.18	77.8
DWT	RF-WBA	MG	CM	RBV	2.69	5.25	78.8
DWF	RF-WBA	MG	CM	RBV	2.58	5.22	81.2

performance. RBA measurement generally appears to be the best method. RF-WBA measurement provides similar performance with WA-WBA measurement. We note that the test images in Fig. 9 are not very noisy. If noise is added, the relative performance of the RF-WBA and WA-WBA schemes may depend on the characteristics of the noise. Based on [48] WA-WBA may be better for Gaussian noise and RF-WBA may be better for impulsive noise.

3) Comparing Schemes Using Different Coefficient Grouping Methods: To compare the performance of schemes using different coefficient grouping methods, we tested NG, SG, and MG decision methods with different combinations of alternatives for the other fusion procedures. The results are shown in Table 5. From the results, we observe that MG always provides fusion performance improvement. The improvement apparently comes from the fact that significant image features tend to be stable with respect to variations in scale. These features tend to have a nonzero lifetime in scale space [15]. Thus, when comparing the corresponding image features in multiple source images, considering the components in all scales together provides a more robust fusion strategy.

4) Comparing Schemes Using Different Coefficient Combining Methods: To compare the performance of schemes using different coefficient combining methods, we tested the CM decision methods and the WA decision methods with different combinations of alternatives for the other fusion procedures. The results are shown in Table 6. From these results, neither the CM nor the WA decision methods are always better and the difference in performance is small. This is consistent with a fact we stated in (14), that weights of zero and one are often chosen. We believe that the small observed difference in performance with CM and WA is due to the lack of noise in the original images in Fig. 9. If noise is added, these findings may be different.

5) Comparing Different Consistency Verification Methods: To compare the performance of schemes using different consistency verification methods, we considered using the different combinations of alternatives for the other fusion procedures. We tested these combinations with no verification (NV), with WBV and with RBV, respectively. The results are shown in Table 7. These results illustrate that consistency verification is helpful for improving the fusion performance. RBV outperformed WBV and NV in all cases.

C. Comparing Some Existing Fusion Schemes and Some Proposed Approaches

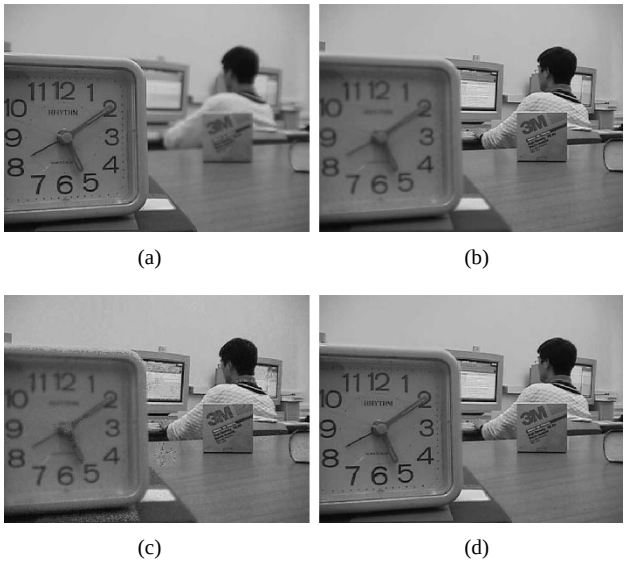
Table 8 shows some fusion schemes which have been studied by previous researchers using LPT or DWT MSD methods. These appear to be the best existing methods we found in the literature. These schemes all fit into the framework discussed in Section II. Table 8 also shows the experimental results obtained for these schemes. The approach suggested in [38] provides much better performance than the other schemes in Table 8.

Based on the fusion framework illustrated in Fig. 4, we propose some new image fusion schemes that appear to outperform the schemes in Table 8 for the test cases we considered. These approaches and their fusion performance are shown in Table 9.

All of the new schemes in Table 9 generally provide approximately equal or better performance than the previous methods listed in Table 8. The new schemes which employ the DWT and DWF are strictly better than all the schemes in Table 8. Complexity and memory requirement should also be considered. The computational complexity depends mainly on the MSD method used. Let N refer to the size of the image and L to the number of the decomposition levels used. For the Laplacian pyramid and the DWF decomposition, the complexity is $O(N \log N)$. For DWT, the complexity is $O(N)$. LPT-based schemes will need a little more storage memory than DWT-based schemes, and DWF-based schemes will use roughly 3L times more storage memory than that of DWT-based schemes.

D. More Image Fusion Examples

While this study has focused mainly on performance analysis using synthetically generated source images, we have also studied realistically generated images for a number of interesting applications (see Figs. 2 and 3). Here it is much more difficult to compare various algorithms, so instead we present results for the best algorithms identified



(a) (b) (c) (d)

Fig. 10. Digital camera image fusion: (a) image 1 (focus on the clock); (b) image 2 (focus on the student); (c) auto-focus image; and (d) fused image (all focus).

(a) (b) (c)

Fig. 11. Fusion result for motion blurred images: (a) tank is motion blurred; (b) background is not clear; and (c) fused image.

(a) (b) (c)

Fig. 12. Fusion result for images with occlusion: (a) clouds block the head; (b) clouds block the back; and (c) fused image.

in Section III-C for a few applications. As discussed in Section III we have particular interest in the fusion of multifocus digital camera images. Inexpensive cameras may have difficulty obtaining images which are in focus everywhere in some situations. Image fusion approaches have already been shown to be useful to avoid such problems [50]. Reference [50] discusses algorithms to control both focus and camera positions to generate an image sequence which is used to obtain a fused composite image. The image fusion approaches we studied here can work for any source images available—they do not rely on controlled camera motion. Fig. 10 shows an example. Fig. 10(a) and (b) are two original images, taken by a digital camera, which have different focus. Fig. 10(c) is the image obtained using the autofocus function of the camera. Fig. 10(d) is the result using the above image fusion approach. In all of our experiments of this type, image fusion using the best algorithms identified in Table 9 performs better than the autofocus of the camera.

In Figs. 11–12, we illustrate deblurring and cloud removal as two other applications of image fusion. In Fig. 11(a), the tank is motion blurred, and in Fig. 11(b) the background in the middle image is not in focus, while the fused image [Fig. 11(c)] has much better quality. In Fig. 12, some parts of the fighter plane body are blocked by the clouds in both of the source images, while the fused image provides a much clearer picture of the fighter.

IV. Discussions

In this paper, some MSD-based image fusion approaches have been studied within a generic image fusion framework. The study was focused on how to use the MSD data of the source images to produce a fused MSD representation which should be more informative to the observer (human or computer). Studies of this type have been lacking in the previous research. Experiments show that region-based fusion and multiscale grouping can almost always improve the image fusion performance for any of the MSD methods we considered.

The framework we suggest in this paper can not only describe the existing image fusion schemes, but it also introduces some new approaches. Such new approaches have been shown to outperform previous approaches.

There are some natural extensions to the research in this paper which became clear as this research progressed.

A. MSD Methods

In this paper, we just discussed some basic MSD methods for image fusion, including LPT, Daubechies's D8 orthonormal DWT, and the corresponding DWF. We found that among these basic MSD methods, DWF appears to provide the best performance at the cost of higher computational and storage expenses. An extensive study of other existing or emerging MSD methods for the purpose of image fusion should be a topic of future research.

B. Decomposition Levels

In our study, we found the number of decomposition levels used can influence image fusion performance. However, using more decomposition levels will not necessarily produce better results. Methods for choosing the appropriate number of decomposition level should receive further attention.

C. Activity Measure

In this paper, we use the absolute value of MSD coefficients as the basic activity measure. There are other ways to calculate activity, some of which may involve incorporating *a priori* information. It would be desirable to search for the activity level measures that most accurately reflect the relative importance of each MSD coefficient.

D. Combining Method

There are some combining methods that we have not considered here. Some of these would be interesting to study. For example, one might require that all coefficients from the same frequency band come from the same source image.

E. Extra Information

Our research employed the assumption that we do not have any further information about the source images other than the pixel values. If prior knowledge is available, the activity measure, combining method, and consistency verification can all use such information to improve fusion performance. Using prior information in the fusion is an important issue that should be studied.

F. Performance Measure

The methods used for testing fusion performance in Section III focus on digital camera applications. They may not be best for other image fusion applications. Designing procedures for fusion performance assessment for other applications is another topic of interest.

The research discussed in this paper was aimed at producing better overall images. We assume that a more informative description of the scene will necessarily improve the performance of further image processing tasks. However, we acknowledge that if one knew the goal of the overall processing, one might be able to tune the fusion toward this specific goal, which might be, for example, locating a particular object.

In this paper, we are promoting MSD-based pixel-level image fusion. Note that we are not advocating this type of fusion at the exclusion of other types of fusion. We believe that different types of fusion can often be applied in conjunction with one another and that further research on this topic is worthwhile and promising.

Another problem which has not been discussed here is image registration. How to align the source images prior to image fusion is also an important issue. In most previous studies, the images to be fused are assumed to be registered. In practice, registration is difficult to achieve and further study on registration and robust image fusion techniques appear to be justified.

REFERENCES

[1] P. K. Varshney, "Scanning the special issue on data fusion," *Proc. IEEE*, vol. 85, pp. 3–5, Jan. 1997.

[2] D. L. Hall, *Mathematical Techniques in Multisensor Data Fusion.* Boston, MA: Artech House, 1992.

[3] J. Llinas and E. Walts, *Multisensor Data Fusion.* Boston, MA: Artech House, 1990.

[4] M. A. Abidi and R. C. Gonzalez, *Data Fusion in Robotics and Machine Intelligence.* New York: Academic, 1992.

[5] J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing System.* Norwell, MA: Kluwer, 1990.

[6] R. C. Luo and M. G. Kay, *Multisensor Integration and Fusion for Intelligent Machines and Systems.* Norwood, NJ: Ablex, 1995.

[7] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, pp. 6–23, Jan. 1997.

[8] P. K. Varshney, "Multisensor data fusion," *Electron. Commun. Eng. J.*, vol. 9, no. 6, pp. 245–253, Dec. 1997.

[9] J. K. Aggarwal, *Multisensor Fusion for Computer Vision.* Berlin, Germany: Springer-Verlag, 1993.

[10] L. A. Klein, *Sensor and Data Fusion Concepts and Applications*. San Jose, CA: SPIE, 1993.

[11] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.*, vol. C-20, pp. 562–569, May 1971.

[12] A. Witkin and J. M. Tenenbaum, "On the role of structure in vision," in *Human and Machine Vision.* New York: Academic, 1983, pp. 481–544.

[13] D. Marr, *Vision.* San Francisco, CA: W. H. Freeman, 1982.

[14] P. J. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, pp. 532–540, Apr. 1983.

[15] T. Lindeberg, *Scale-Space Theory in Computer Vision.* Norwell, MA: Kluwer, 1994.

[16] D. D. Ferris, Jr., R. W. McMillan, N. C. Currie, M. C. Wicks, and A. Slamani, "Sensors for military special operations and law enforcement applications," in *Proc. SPIE*, vol. 3062, 1997, pp. 173–180.

[17] M. A. Slamani, L. Ramac, M. Uner, P. K. Varshney, D. D. Weiner, M. Alford, D. D. Ferris, Jr., and V. Vannicola, "Enhancement and fusion of data for concealed weapons detection," in *Proc. SPIE*, vol. 3068, 1997, pp. 8–19.

[18] M. M. Daniel and A. S. Willsky, "A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing," *Proc. IEEE*, vol. 85, pp. 164–180, Jan. 1997.

[19] D. Hill, P. Edwards, and D. Hawkes, "Fusing medical images," *Image Processing*, vol. 6, no. 2, pp. 22–24, 1994.

[20] J. M. Reed and S. Hutchinson, "Image fusion and subpixel parameter estimation for automated optical inspection of electronic components," *IEEE Trans. Ind. Electron.*, vol. 43, pp. 346–354, June 1996.

[21] P. J. Burt, "The pyramid as structure for efficient computation," in *Multiresolution Image Processing and Analysis.* Berlin, Germany: Springer-Verlag, 1984, pp. 6–35.

[22] A. Akerman, III, "Pyramid techniques for multisensor fusion," in *Proc. SPIE*, vol. 1828, 1992, pp. 124–131.

[23] A. Toet, L. J. van Ruyven, and J. M. Valeton, "Merging thermal and visual images by a contrast pyramid," *Opt. Eng.*, vol. 28, no. 7, pp. 789–792, July 1989.

[24] A. Toet, "Hierarchical image fusion," *Mach. Vision Applicat.*, vol. 3, pp. 1–11, Mar. 1990.

[25] ——, "Multiscale contrast enhancement with application to image fusion," *Opt. Eng.*, vol. 31, no. 5, pp. 1026–1031, 1992.

[26] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Proc. 4th Int. Conf. Computer Vision*, Berlin, Germany, May 1993, pp. 173–182.

[27] Y. T. Zhou, "Multi-sensor image fusion," in *Proc. IEEE Int. Conf. Image Processing, IEEE'94,* Austin, TX, pp. 193–197.

[28] T. Huntsberger and B. Jawerth, "Wavelet based sensor fusion," in *Proc. SPIE*, vol. 2059, 1993, pp. 488–498.

[29] T. Ranchin, L. Wald, and M. Mangolini, "Efficient data fusion using wavelet transform: The case of spot satellite images," in *Proc. SPIE*, vol. 2934, 1993, pp. 171–178.

[30] I. Koren, A. Laine, and F. Taylor, "Image fusion using steerable dyadic wavelet transform," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Washington, DC, 1995, pp. 232–235.

[31] L. J. Chipman, T. M. Orr, and L. N. Graham, "Wavelets and image fusion," in *Proc. SPIE*, vol. 2569, 1995, pp. 208–219.

[32] C. Lejeune, "Wavelet transform for infrared application," in *Proc. SPIE*, vol. 2552, 1995, pp. 313–324.

[33] T. A. Wilson, S. K. Rogers, and L. R. Myers, "Perceptual-based hyperspectral image fusion using multiresolution analysis," *Opt. Eng.*, vol. 34, no. 11, pp. 3154–3164, 1995.

[34] X. Jiang, L. Zhou, and Z. Gao, "Multispectral image fusion using wavelet transform," in *Proc. SPIE*, vol. 2898, 1996, pp. 35–42.

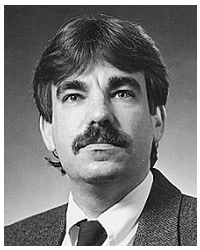
[35] L. Peytavin, "Cross-sensor resolution enhancement of hyperspectral images using wavelet decomposition," in *Proc. SPIE*, vol. 2758, 1996, pp. 193–197.

[36] M. K. Uner, L. C. Ramac, and P. K. Varshney, "Concealed weapon detection: An image fusion approach," in *Proc. SPIE*, vol. 2942, 1997, pp. 123–132.

[37] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models and Image Processing*, vol. 57, no. 3, pp. 235–245, May 1995.
[38] Z. Zhang and R. S. Blum, "A region-based image fusion scheme for concealed weapon detection," in *Proc. 31st Annu. Conf. Information Sciences and Systems*, Baltimore, MD, Mar. 97, pp. 168–173.
[39] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. Pure and Applied Math.*, vol. 41, pp. 909–996, Nov. 1988.
[40] K. C. Chou, A. S. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 464–478, Mar. 1994.
[41] H. C. Hsin and C. C. Li, "Wavelet-based filtering in scale space for data fusion," in *Proc. SPIE*, vol. 2569, 1995, pp. 701–712.
[42] H. Wechsler, *Computational Vision.* New York: Academic, 1990.
[43] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, pp. 674–693, July 1989.
[44] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Processing*, vol. 40, pp. 2207–2232, Sept. 1992.
[45] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Processing*, vol. 4, pp. 1549–1560, Nov. 1995.
[46] R. M. Young, *An Introduction to Nonharmonic Fourier Series.* New York: Academic, 1980.
[47] I. Daubechies, "Ten lectures on wavelets," SIAM, Tech. Rep./CBMS-NSF Lecture Notes no. 61, 1992.
[48] P. P. Gandhi and S. A. Kassam, "Design and performance of combination filters for signal restoration," *IEEE Trans. Signal Processing*, vol. 39, pp. 1524–1540, July 1991.
[49] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 679–698, Nov. 1986.
[50] W. B. Seales and S. Dutta, "Everywhere-in-focus image fusion using controllable cameras," in *Proc. SPIE*, vol. 2905, 1996, pp. 227–234.

Zhong Zhang received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 1993. He is currently pursuing the Ph.D. degree in electrical engineering and computer science at Lehigh University, Bethlehem, PA.

His research interests include multisensor data fusion, image fusion, computer vision, and pattern recognition.



Rick S. Blum (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Pennsylvania State University, University Park, in 1984 and the M.S. and Ph.D degrees in electrical engineering from the University of Pennsylvania, Philadelphia, in 1987 and 1991, respectively.

From 1984 to 1991, he was a Member of Technical Staff at General Electric (GE) Aerospace, Valley Forge, PA, and he graduated from GE's Advanced Course in engineering. Since 1991, he has been with the Electrical Engineering and Computer Science Department, Lehigh University, Bethlehem, PA, where he is currently an Associate Professor and holds a Class of 1961 Professorship. His research interests include signal detection and estimation and related topics in the areas of signal processing and communications. He holds a patent for a parallel signal and image processor architecture.

Dr. Blum is a member of Eta Kappa Nu and Sigma Xi. He was awarded an ONR Young Investigator Award in 1997 and an NSF Research Initiation Award in 1992. He is currently an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE COMMUNICATIONS LETTERS.