An Advanced Image Fusion Algorithm Based on Wavelet Transform – Incorporation with PCA and Morphological Processing

Yufeng Zheng, Edward A. Essock and Bruce C. Hansen
Dept. of Psychological & Brain Sciences, University of Louisville, Louisville, KY 40292, USA

ABSTRACT

There are numerous applications for image fusion, some of which include medical imaging, remote sensing, nighttime operations and multi-spectral imaging. In general, the discrete wavelet transform (DWT) and various pyramids (such as Laplacian, ratio, contrast, gradient and morphological pyramids) are the most common and effective methods. For quantitative evaluation of the quality of fused imagery, the root mean square error (RMSE) is the most suitable measure of quality if there is a "ground truth" image available; otherwise, the entropy, spatial frequency or image quality index of the input images and the fused images can be calculated and compared. Here, after analyzing the pyramids' performance with the four measures mentioned, an advanced wavelet transform (*a*DWT) method that incorporates principal component analysis (PCA) and morphological processing into a regular DWT fusion algorithm is presented. Specifically, at each scale of the wavelet transformed images, a principle vector was derived from two input images and then applied to two of the images' approximation coefficients (i.e., they were fused by using the principal eigenvector). For the detail coefficients (i.e., three quarters of the coefficients), the larger absolute values were chosen and subjected to a neighborhood morphological processing procedure which served to verify the selected pixels by using a "filling" and "cleaning" operation (this operation filled or removed isolated pixels in a 3-by-3 local region). The fusion performance of the advanced DWT (*a*DWT) method proposed here was compared with six other common methods, and, based on the four quantitative measures, was found to perform the best when tested on the four input image types. Since the different image sources used here varied with respect to intensity, contrast, noise, and intrinsic characteristics, the *a*DWT is a promising image fusion procedure for inhomogeneous imagery.

Keywords: Image fusion, image pyramid, wavelet transform (DWT), principal component analysis (PCA), morphological processing.

1. INTRODUCTION

Along with the development of new imaging sensors arises the need for a meaningful combination of all employed imaging sources (i.e. image fusion). There are many applications of image fusion including medical imaging, remote sensing, nighttime operations and multi-spectral imaging. A general definition of image fusion has been stated as "the combination of two or more different images to form a new image by using a certain algorithm"[1]. The actual fusion process can take place at different levels of information representation. A generic categorization of these different levels can be sorted in ascending order of abstraction: pixel, feature, and decision level. This paper focuses on the so-called pixel-level fusion process, where a composite image has to be built of several (typically two) input images.

Image fusion is a tool that serves to combine multisource imagery by using advanced image processing techniques. Specifically, it aims at the integration of disparate and complementary data in order to enhance the information apparent in the images, as well as to increase the reliability of the interpretation. This leads to more accurate data[2] and increased utility[3,29]. In addition, it has been stated that fused data provides for robust operational performance such as increased confidence, reduced ambiguity, improved reliability and improved classification[3,30].

Currently most of the fusion techniques described in the literature are pixel-based. Many techniques are based on multiresolution processing which allows for a combination of edge information at different scales. A very popular

approach is given by the wavelet transform[4-7]. Use of pyramid-based fusion methods is described elsewhere[8-9]. The rule for combining the detail information is an important issue. The most common rule for fusion is to take the detail coefficient with highest energy (e.g. by simply choosing the highest absolute value in the DWT) from one of the bands (such as high-pass filtered bands). A general framework of image fusion can be found in Ref. [10].

In pixel-level image fusion, some general requirements[11] are imposed on the fusion result: (1) The fusion process should preserve all relevant information of the input imagery in the composite image (pattern conservation); (2) The fusion scheme should not introduce any artifacts or inconsistencies which would distract the human observer or subsequent processing stages; (3) The fusion process should be shift and rotation invariant (i.e. the fusion result should not depend on the location or orientation of an object in the input imagery, which is crucial to pattern recognition or object detection). However, in this paper little consideration is given to Requirement (3) that only becomes crucial to some particular applications of image fusion (e.g. target recognition and tracking). Instead, quantitative evaluation of the quality of fused imagery is considered most important for an objective comparison of the fusion algorithms' performances. For example, if there is a "ground truth" image available, the root mean square error (RMSE) is the most natural and suitable measure of quality, otherwise, the usual procedure is to calculate and compare the entropy, or a spatial frequency or image quality index of the input images and the fused images.

Based on an analysis of pyramids' performance with the four evaluation measures tested on four categories of imagery, an advanced wavelet transform (*a*DWT) method that incorporates principal component analysis (PCA) and morphological processing into a regular DWT fusion algorithm is presented in this paper. The subsequent sections of this paper are organized as follows. First, brief reviews of DWT and PCA techniques are given. Next is a full description of pyramid methods, the presented method (*a*DWT), and image quality evaluation methods. Lastly, the experimental results and discussion are described, followed by conclusions.

2. BRIEF REVIEWS OF DWT AND PCA

2.1 Discrete wavelet transform

Wavelet-based analysis of signals is an interesting, and relatively recent, tool. Similar to Fourier series analysis, where sinusoids are chosen as the basis functions, wavelet analysis is based on a decomposition of a signal using an *orthonormal* family of basis functions. Unlike a sine wave, a wavelet has its energy concentrated in time, or, as in the present application, space. Sinusoids are useful in analyzing periodic and time-invariant phenomena, while wavelets are well suited for the analysis of transient, time-varying signals. Accordingly, in the spatial domain (e.g., a 2D image) DWT analysis also gives the best performance in detecting discontinuities or subtle changes in gray level.

Suppose $f(x) \in L^2(\mathbf{R})$ (where $\mathbf{R}$ is the set of real numbers and $L^2(\mathbf{R})$ denotes the set of measurable, square-integrable one-dimensional functions) relative to the wavelet function $\psi(x)$ and scaling function $\varphi(x)$. A wavelet series expansion is similar in form to the well-known Fourier series expansion, in which it maps a function of a continuous variable into a sequence of coefficients. If the function being expanded is a sequence of numbers (e.g., samples of a continuous function $f(x)$), the resulting coefficients are called the *discrete wavelet transform* (DWT) of $f(x)$. The DWT transform pair is defined as following

$$W_\varphi(j_0,k) = \frac{1}{\sqrt{M}} \sum_x f(x) \varphi_{j_0,k}(x), \tag{1}$$

$$W_\psi(j,k) = \frac{1}{\sqrt{M}} \sum_x f(x) \psi_{j,k}(x), \text{for } j = j_0; \tag{2}$$

and

$$f(x) = \frac{1}{\sqrt{M}} \sum_k W_\varphi(j_0,k) \varphi_{j_0,k}(x) + \frac{1}{\sqrt{M}} \sum_{j=j_0}^{\infty} \sum_k W_\psi(j,k) \psi_{j,k}(x); \tag{3}$$

where $f(x)$, $\varphi_{j_0,k}(x)$ and $\psi_{j,k}(x)$ are functions of the discrete variable $x = 0, 1, 2, ..., M-1$. Normally, let $j_0 = 0$, and M (the length of the discrete samples of $f(x)$) is a power of 2 (i.e., $M = 2^J$) so that the summations are performed over $x = 0, 1, 2, ..., M-1$, $j = 0, 1, 2, ..., J-1$, and $k = 0, 1, 2, ..., 2^j - 1$. The transform itself is composed of M coefficients, the

minimum scale is 0, and the maximum scale is $J - 1$. The coefficients defined in Eqs. (1) and (2) are usually called *approximation* and *detail coefficients*, respectively. The process of computing these coefficients is referred to as DWT *analysis*. On the other hand, DWT *synthesis* (or inverse DWT) is defined by Eq. (3) to reconstruct *f(x)* with these coefficients. Finally, it should be noted that Eqs. (1) through (3) are valid for orthonormal bases and tight frames alone[12].

The *scaling* function is given by

$$\varphi_{j_0,k}(x) = 2^{j_0/2}\varphi(2^{j_0}x - k)\,, \tag{4}$$

for all j_0, $k \in \mathbf{Z}$ (where $\mathbf{Z}$ is the set of integers) and $\varphi_{j_0,k}(x) \in L^2(\mathbf{R})$. Here, k determines the position of $\varphi_{j_0,k}(x)$ along x-axis, j_0 determines $\varphi_{j_0,k}(x)$'s width and $2^{j_0/2}$ controls its height or amplitude.

The *wavelet* function is defined as

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^{j}x - k)\,. \tag{5}$$

The same notations are used here as those in Eq. (4). $\psi(x)$ is also referred to as the mother wavelet. Thus, wavelet basis functions are obtained from a single mother wavelet by translation and scaling. However, there is no single or universal mother wavelet function. The mother wavelet must simply satisfy a small set of conditions and is typically selected based on the domain of the signal or image processing problem. Almost all useful wavelet systems satisfy the multi-resolution condition[12]. This means that given an approximation of a signal *f(x)* using translations of a mother wavelet up to a selected scale, we can achieve a better approximation by using expansion signals with half the width and half the translation steps. This is conceptually similar to improving frequency resolution by doubling the number of harmonics (i.e., halving the fundamental harmonic) in a Fourier series expansion.

In practice, a mother wavelet is selected from 'ready-made' wavelets for a particular problem, and different wavelets $\psi(x)$ have different effects. For instance, Harr wavelets are suitable for representing a piecewise signal, and Daubechies wavelets are more suitable for compressing data. Here we chose the fourth order of Symlets (symmetric wavelets as shown in Fig. 1) to capture subtle gray level changes (e.g., edges in source images). Usually a wavelet with a shorter length is more sensitive to edge detection.

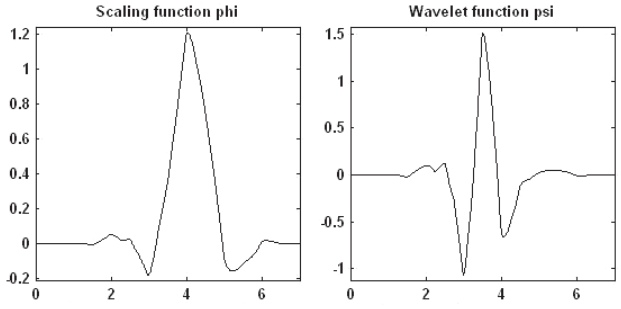


Fig. 1: The 4th order of Symlets — Scaling and Wavelet function.

2.2 Principle component analysis

Principal component analysis (PCA) is a mathematical procedure that transforms a number of potentially correlated variables into a smaller number of uncorrelated variables called *principal components*. The objective of PCA is to reduce the dimensionality (the number of variables) of the dataset while retaining most of the original variability in the data. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Thus, PCA is concerned with explaining the variance and covariance structure of a high-dimensional random vector through a few linear combinations of the original component variables.

A common way to find the principal components of a data set is by calculating the eigenvectors of the data covariance matrix[12]. These eigenvectors give the directions in which the data distribution is stretched most. The projections of the data on the eigenvectors are the principal components. The corresponding eigenvalues give an indication of the amount of information that the respective principal components represent. Principal components corresponding to large eigenvalues represent a large amount of information in the data set and thus tell us much about the relations between the data points.

3. THE METHODS FOR IMAGE FUSION AND EVALUATION

3.1 Pyramid methods

Image pyramids have been initially described for a multi-resolution image analysis and as a model for the binocular fusion in human vision[14-15]. An image pyramid can be described as collection of low- or band-pass copies of an original image in which both the band limit and sample density are reduced in regular steps. The basic strategy of image fusion based on pyramids is to use a feature selection rule to construct a fused pyramid representation from the pyramid representations of the original images. The composite image is obtained by taking the inverse pyramid transform. Several pyramid-based fusion schemes have been proposed in recent years, and are briefly described below.

(a) Laplacian pyramid

A set of band-pass copies of an image is referred to as the Laplacian pyramid due to the similarity to a Laplacian operator. Each level of the Laplacian pyramid is recursively constructed from its lower level by the following four basic steps: blurring (low-pass filtering); subsampling (reduce size); interpolation (expand); and differencing (to subtract two images pixel by pixel)[13]. In the Laplacian pyramid, the lowest level of the pyramid is constructed from the original image. The Laplacian pyramid was first introduced as a model for binocular fusion in human stereo vision[14-15], where the implementation used a Laplacian pyramid and a maximum selection rule at each point of the pyramid transform.

(b) Ratio of low-pass pyramid and Contrast pyramid

The ratio of low-pass (RoLP) or contrast pyramid, which was introduced by Toet[16-17], is very similar to a Laplacian pyramid. Originally the RoLP was explicitly intended for use by human observers. Every level of RoLp is the ratio of two successive levels of the Gaussian pyramid. In Refs. [16-19], a ratio of low pass pyramid and the maximum selection rule were used for visible-to-IR image fusion.

(c) Gradient pyramid

The gradient pyramid can be generated by applying gradient operators to each level of the Gaussian pyramid[20]. This produces, horizontal, vertical, and diagonal pyramid sets for each source in the Gaussian pyramid. Burt[21] proposed an image fusion scheme that was based on a gradient pyramid and an activity measure within a small window rather than just a single point. In Reference [22], Sims and Phllips demonstrated the qualitative and quantitative results of the above three image data fusion algorithms and their target signature variation.

(d) Morphological pyramid

A morphological pyramid can be constructed by the successive filtering of the original image with a sequence of nonlinear morphological operators (such as the open-close filter) and a specialized subsampling scheme[23]. The application of morphological pyramid to image fusion can be referenced in [24-25].

3.2 The DWT method and advanced DWT (*a*DWT)

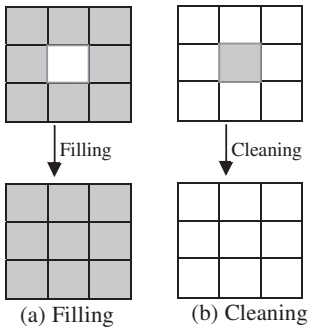
As with a pyramid method, the regular DWT method is also a multi-scale analysis method. In general, DWT coefficients from two input images' are fused (pixel-by-pixel) by choosing the average of the approximation coefficients at the highest transform scale, and the larger absolute value of the detail coefficients at each transform scale. Then an inverse DWT is performed to obtain the fused image. In the present method, we incorporate PCA and morphological processing into the fusion algorithm, which results in the *a*DWT. Specifically, at each DWT scale of a particular image, the DWT coefficients of a 2D image consist of four quarters: approximation, horizontal detail, vertical detail and diagonal detail. We apply PCA to two input images' approximation coefficients at the highest transform scale, i.e. fuse them by using the principal eigenvector (corresponding to the larger eigenvalue) derived from the two original images, as described in Eq. (6) below:

$$C_F = (a_1 \cdot C_A + a_2 \cdot C_B)/(a_1 + a_2), \tag{6}$$

where C_A and C_B are approximation coefficients transformed from input images A and B. C_F represents the fused coefficients; a_1 and a_2 are the elements of principal eigenvector, which were computed by analyzing the original input images (Note: not analyzing C_A and C_B because their sizes at the highest transform scale are too small to conduct an accurate result). In practice, we have to convert 2-D images into 1-D vectors by simply stacking each image column-by-

column so that the principle component can be computed as described in Section 2.2. Note that the denominator in Eq. (6) is used for normalization so that the fused image has the same energy distribution as original input images.

For the detail coefficients (the other three quarters of the coefficients) at each transform scale, the larger absolute values were selected, followed a neighborhood morphological processing, which served to verify the selected pixels by using a "filling" and "cleaning" operation (i.e., the operation fills or removes isolated pixels locally as shown in Fig. 2). Such an operation (similar to smoothing) can increase the consistence of coefficient selection thereby reducing the distortion in the fused image.



Note: Shaded pixels were chosen from image A, and white pixels were selected from image B.

Fig 2: Morphological processing (of a 3×3 region): Filling and Cleaning.

It is important to recognize that the number of DWT scales (J, defined in Section 2.1) that should be decomposed, depends on the input image size. However, the size of smallest transformed image should not be less than 2 by 2 pixels (i.e. $l_0 = 2$). Thus we had

$$J = \mathrm{int}(\log_2(\min(M,N)/l_0)), \tag{7}$$

where the image size is $M \times N$, l_0 represents for the minimal length of transformed image by DWT, *min* means taking the minimal value, and *int* stands for taking the integral part.

3.3 Image quality evaluation

As mentioned in the introduction, the general requirements of an image fusing process are that it should preserve all valid and useful pattern information from the source images, while at the same time it should not introduce artifacts that could interfere with subsequent analyses[15]. However, it is almost never possible to combine images without introducing some form of distortion. In the current body of literature, image fusion results tend to be evaluated visually by inspection or objectively[29,30]. Here we focus on quantitative measures (i.e., objective) that could be done automatically by computers. Four commonly used performance measures are listed below, some of which need a reference image while others do not. But it should be noted that the best criterion should be linked with the specific application.

(a) Root mean square error (RMSE)

The RSME between the reference image and the fused image is

$$\mathrm{RMSE} = \left(\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[I_R(i,j) - I_F(i,j)]^2}{M \times N} \right)^{\frac{1}{2}}, \tag{8}$$

where $I_R(i,j)$ and $I_F(i,j)$ are the image pixel values of the reference image (i.e. the ground truth image) and the fused image respectively. $M \times N$ is the image size.

(b) Entropy (E)

$$\mathrm{Entropy} = -\sum_{l=0}^{L-1} p(l)\log_2 p(l), \tag{9}$$

where $p(l)$ is the probability of gray level l (that can be computed by analyzing the image histogram), and the dynamic range of analyzed image is [0, L-1].

(c) Spatial frequency (SF)

Spatial frequency is used to measure the overall activity level of an image[26], which is defined as follows.

$$SF = \sqrt{(RF)^2 + (CF)^2}, \tag{10}$$

where *RF* and *CF* are row frequency and column frequency respectively.

$$RF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}[I(i,j)-I(i,j-1)]^2}\ , \tag{11a}$$

$$CF = \sqrt{\frac{1}{MN}\sum_{j=1}^{N}\sum_{i=2}^{M}[I(i,j)-I(i-1,j)]^2}\ . \tag{11b}$$

(d) Image quality index (IQI)

The image quality index was recently introduced by Wang and Bovik[27]. Given two sequences $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, let $\bar{x}$ denote the mean of x, and σ_x and σ_{xy} denote the variance of x and covariance of x and y, respectively. The global quality index of two vectors is defined as

$$Q_0(x,y) = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\bar{x}^2+\bar{y}^2)(\sigma_x^2+\sigma_y^2)}, \tag{12}$$

which can be decomposed as

$$Q_0(x,y) = \frac{\sigma_{xy}}{\sigma_x\sigma_y}\cdot\frac{2\bar{x}\bar{y}}{(\bar{x}^2+\bar{y}^2)}\cdot\frac{2\sigma_x\sigma_y}{(\sigma_x^2+\sigma_y^2)}\ . \tag{13}$$

Note that the first component in Eq. (13) is the correlation coefficient between x and y. This value is a measure for the similarity of the vectors x and y, and takes values between -1 and 1. Keep in mind that in this case (image quality evaluation), the values x_i, y_i are positive grey-scale values. The second component in Eq. (13) corresponds to the luminance distortion which has a dynamic range of [0, 1]. The third factor in Eq. (13) measures the contrast distortion and its range is also [0, 1]. In summary, $Q_0 \in [0, 1]$; and the maximum value $Q_0 = 1$ is achieved when x and y are identical.

Piella and Heijmans[28] introduced a weighting procedure into Q_0 calculation, where S(A) denotes some saliency of image A. It should reflect the local relevance of image A that may depend on local variance, contrast, sharpness, or entropy. Given the local saliencies of two input images A and B, we compute a local weight λ indicating the relative importance of image A compared to image B: the larger λ, the more weight is given to image A. A typical choice for λ is

$$\lambda = S(\mathrm{A}) / [S(\mathrm{A}) + S(\mathrm{B})]. \tag{14}$$

Then, the fused image quality measure (i.e., the weighted image quality index) can be defined as

$$Q_w = \lambda Q_0(\mathrm{A},\mathrm{F}) + (1-\lambda)\, Q_0(\mathrm{B},\mathrm{F}). \tag{15}$$

Since image signals are generally non-stationary, it is more appropriate to measure the weighted image quality index Q_w over local regions (e.g. a slide window) and then combine the different results into a single measure. In our experiments, we chose S(A) = SF(A) (i.e., the spatial frequency of image A instead of using the local variance as Piella suggested).

4. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments are organized to apply seven image fusion algorithms (i.e., five pyramid-based and two DWT-based methods) with four image quality measures by using four types of imagery (already well registered and with 256 gray levels): (1) two simulated image pairs, (2) frequently used samples, (3) image intensified and infrared paired images (obtained from Army NVESD), and (4) near-infrared and infrared paired images (obtained in our lab). We apply each fusion method to the four imagery groups, and calculate their fusion performance with the four quantitative measures.

For imagery for which the ground truth image is available, RMSE is the suitable measure that indicates how close this fused image is to the reference image (or the ground truth). Since the potential artificial distortion during the fusion process can also increase the entropy or spatial frequency values of the fused image, the IQI is a relatively faithful

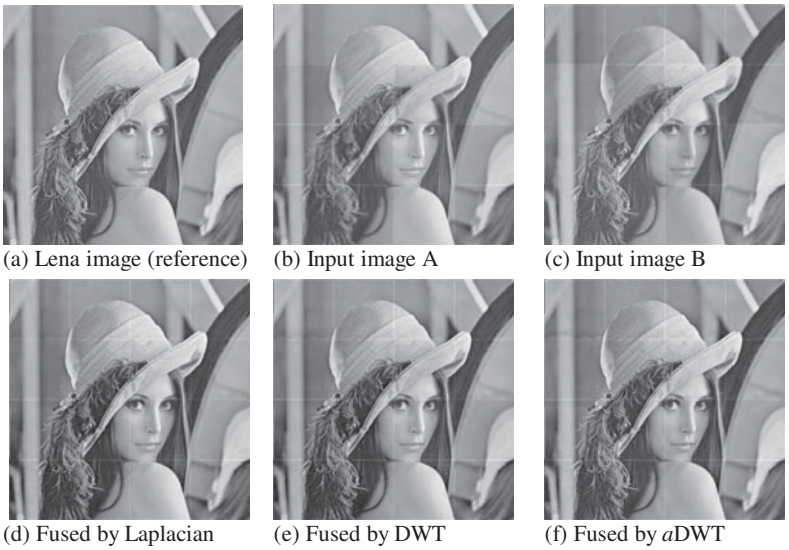
measure for those images without the ground truth image because the IQI gives a quantity (between 0 and 1) about how similar the fused image is to both input images.

In the *a*DWT method, the fourth order of "Symlets" were selected, but in the regular DWT method, the second order of "Daubechies" wavelets were used. *J*-level transforms (depends on image size, and computed by Eq. (7) given $l_0 = 4$) were analyzed with all pyramid and DWT methods.

4.1 Experimental results

(a) Two simulated off-focus image pairs (Lena, with ground truths)

Two input images were generated by filtering a given image (Lena image, the ground truth or reference image as shown in Fig. 3 (a)) with a 5×5 Gaussian window. The Lena image was divided into 16 blocks, 8 of which were randomly chosen to be blurred as one of the input images, referred as "image A". The other 8 blocks were blurred to form "image B". This resulted in Lena Imagery 1 (Lena-I1). A second set, Lena imagery 2 (Lena-I2), was obtained by further reducing 16 gray levels after Gaussian blurring. The orthogonal stripes, as shown in Fig. 3 (b) and (c), were caused by sliding the window along image block borders. It is more difficult to fuse Lena-I2 perfectly because the input images have different brightness and contrast. The image quality evaluation results of the fused images by different algorithms are given in Table 1. For Lena-I1, judged on RMSE, the best method is the contrast pyramid, however, *a*DWT did the best according to the measures of SF and IQI. For Lena-I2, the *a*DWT had the best performance with three measures expect for entropy. The Laplacian pyramid had the best performance based on the entropy measure. Only three fused images (same as following figure illustrations) were illustrated in Fig. 3 (d)-(e) because Laplacian pyramid was shown to be a better algorithm in other imagery, and the regular DWT method was compared to the *a*DWT. In fact, it is pretty hard to visually examine the difference between fused images, especially from the scaled hardcopy. (That is particularly true for Lena-I1 that was not displayed here.)



(a) Lena image (reference) (b) Input image A (c) Input image B

(d) Fused by Laplacian (e) Fused by DWT (f) Fused by *a*DWT

Fig. 3: Image fusion with the simulated imagery – Lena-I2: the original image size is 512 by 512 pixels. All images were scaled to fit in the paper space.

Table 1: Image quality evaluation results of the fused images tested on the two simulated image pairs – Lena-I1 and Lena-I2. The entropies and spatial frequencies of the reference image and input images are also given in the table.

Imagery	Ref.	Input A	Input B	Eval.	Laplacian	Gradient	Contrast	Ratio	Morpho.	DWT	*a*DWT
Lena-I1	7.31	7.29	7.29	Entropy	**7.46**	7.2	7.44	7.38	7.38	7.42	7.45
	147.77	89.53	116.55	SF	145.59	121.98	145.9	92.68	126.03	142.19	**151.71**
				IQI	0.96	0.93	0.96	0.92	0.93	0.96	**0.96**
				RMSE	3.03	22.27	**1.53**	5.91	8.79	2.06	2.75
Lena-I2	7.31	7.36	7.33	Entropy	**7.56**	7.34	7.47	7.35	7.42	7.53	7.52
	147.77	92.34	119.76	SF	152.18	125.92	116.14	81.28	121.62	150.93	**159.17**
				IQI	0.95	0.94	0.94	0.9	0.92	0.94	**0.95**
				RMSE	11.31	15.17	14.85	13.41	8.51	12.47	**6.74**

The entropy is not a reliable measure that can be demonstrated due to the following issues: the entropy of a reference image is smaller than that of input images for Lena-I2; and its assessed results do not coincide with that of RMSE. On the other hand, the IQI is very trustworthy as it is consistent with RMSE's measure.

(b) Frequently used samples

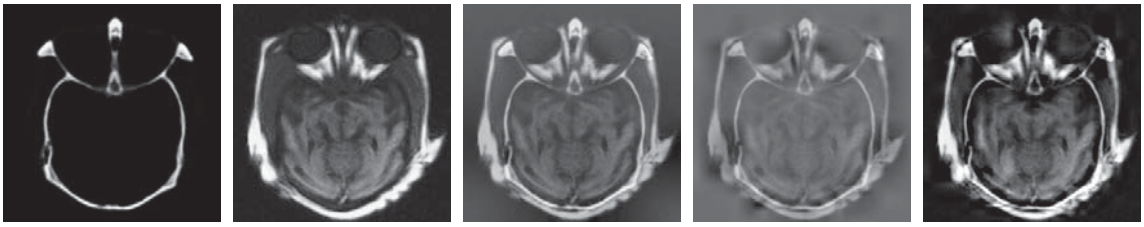
Three frequently used samples were tested and included: two clocks obtained with two different focal planes; two different types of medical imagery, CT and MRI; and a pair of remote sensing images (infrared & low-light "visible" sensors). The quantitative assessments of fused images are listed in Table 2. The *a*DWT's performance was almost the best according to most of the metrics. Judged on the measure of IQI, the *a*DWT's performance is a little bit worse than Lanlacian's when tested on remote sensing imagery only, but it is still much better than a regular DWT's result. Three selected fused images were given from Fig. 4 through Fig. 6. Perceptually, the *a*DWT fused images appear to look the best too. Note that there was no further image processing imposed on the fused images.

Table 2: Image quality evaluation results of the fused images tested on three frequently used samples.

Imagery	Input A	Input B	Eval.	Laplacian	Gradient	Contrast	Ratio	Morpho.	DWT	*a*DWT
Clocks	6.98	6.92	Entropy	7.12	7.22	7.23	7.29	7.03	7.26	**7.42**
	60.41	37.60	SF	50.8	40.11	73.83	38.03	45.47	63.8	**84.03**
			IQI	0.81	0.8	0.77	0.73	0.77	0.86	**0.87**
Medical	1.71	5.58	Entropy	**6.73**	6.64	1.47	3.04	5.57	6.2	6.67
	169.26	163.54	SF	135.53	136.96	70.7	7.13	130.64	61.84	**288.02**
			IQI	0.77	0.77	0.32	0.05	0.63	0.58	**0.77**
Remote	5.45	5.44	Entropy	**7.48**	7.46	7.29	7.16	7.15	7.23	7.03
	680.06	899.54	SF	631.09	618.09	602.7	272.32	399.26	403.28	**1182.89**
			IQI	**0.81**	0.78	0.73	0.61	0.7	0.74	0.8

(a) Input image A (b) Input image B (c) Fused by Laplacian (d) Fused by DWT (e) Fused by *a*DWT

Fig. 4: Image fusion with off-focus images: 512 by 512 pixels.



(a) Image A (CT) (b) Image B (MRI) (c) Fused by Laplacian (d) Fused by DWT (e) Fused by *a*DWT

Fig. 5: Image fusion with medical images: 256 by 256 pixels.

(a) Image A (infrared) (b) Image B (low-light) (c) Fused by Laplacian (d) Fused by DWT (e) Fused by *a*DWT

Fig. 6: Image fusion with remote sensing images: 512 by 512 pixels.

(c) Image intensified and infrared paired images (Army NVESD imagery)

Twenty-eight pairs of image intensified (II) and infrared (IR) realistic night-vision images taken outdoors were used here. The source images were moderately noisy duo to collection conditions. Therefore, to suppress the noise, a 2D median filter (with a 3 by 3 slide window) was first applied to all images (i.e. preprocessing) before any fusion operation. This is a common practice in fusing night time imagery; moreover denoised images were used for all algorithms. The filtered images are not shown here, but this filtering operation wasn't detrimental to the fused image quality (see Fig. 7 (c-e)). The complete assessments of the seven algorithms along with the three evaluation measures are listed in Table 3. Here the means of performances over all images were used. Although the image contents vary considerably, the averaged values still give a sense of an algorithm's performance. The IQI values are always between [0,1] no matter what the analyzed images were. Judging from the IQI values, the Laplacian pyramid did the best, but the *a*DWT was better than DWT. However, the *a*DWT was the best according to the other two metrics. One typical pair of images is illustrated in Fig. 7. The *a*DWT's result appears to look the best as well.

(a) Image A (II) (b) Image B (IR) (c) Fused by Laplacian (d) Fused by DWT (e) Fused by *a*DWT

Fig. 7: Image fusion with Army NVESD imagery: 220 by 220 pixels.

(d) Near infrared and infrared paired images (Lab imagery)

Here a total of 50 pairs of near infrared and infrared images that were captured both indoors and outdoors using our own facilities (Indigo Systems, Inc., Phoenix cameras). For the same reason as stated above, a 2D median filter was first used to remove noise. The mean performance is given in Table 3. For this imagery, the *a*DWT gave the best results with all three measures. One sample was drawn and displayed in Fig. 8. From the observation of the fused image, the *a*DWT shown the best.

Table 3: Averaged image quality evaluation results of the fused images tested on Navy imagery and Lab imagery.

Imagery	Input A	Input B	Eval.	Laplacian	Gradient	Contrast	Ratio	Morpho.	DWT	*a*DWT
Army (28×2)	5.65	6.46	Entropy	7.16	7.14	7.07	6.85	7.12	7.06	**7.27**
	104.21	123.15	SF	148.12	151	159.17	93.73	152.32	140.41	**208.37**
			IQI	**0.73**	0.72	0.62	0.5	0.68	0.69	0.71
Lab (50×2)	5.59	5.94	Entropy	7	7.08	6.64	6.51	6.86	6.9	**7.08**
	89.59	243.59	SF	166.95	163.57	186.37	86.75	155.73	140.75	**300.31**
			IQI	0.74	0.72	0.55	0.46	0.68	0.69	**0.78**

(a) Image A (near IR) (b) Image B (IR) (c) Fused by Laplacian (d) Fused by DWT (e) Fused by *a*DWT

Fig. 8: Image fusion with Lab imagery: 220 by 220 pixels.

4.2 Discussion

In the current experiments, we found that a lower order of wavelets (with shorter length) usually gave a more sharpened fused image, however it often resulted in a larger RMSE or increased distortion, hence the fourth order of Symlets was chosen. Similarly, a smaller l_0 in Eq. (7) (e.g. l_0 = 2, means higher level DWT decomposition) also resulted in more sharpened images, so we let l_0 = 4 in our experiments to suppress any possible distortion. As to the image quality evaluation, the entropy values are not a reliable measure as mentioned in the previous section. The experimental results also reveal that spatial frequency is more reliable than entropy (although currently they both can not distinguish the useful information from the distortion). We will discuss this particular problem and develop a new enhanced measure based on the current SF definition in a subsequent paper. Like RMSE, the IQI measures how similar the fused image is to both input images, which can exclude the distortion effect from the measured result. Consequently, the IQI is the most reliable measure among others for imagery without a ground truth image. Our experimental results also demonstrated this point. For example, in Table 1, the highest IQI values correspond to the lowest RMSE values.

Overall, the *a*DWT did the best in fusing inhomogeneous imagery. For off-focus images both of which have similar brightness and contrast, the ratio or contrast pyramid methods could also yield better fusion. For night vision imagery where brightness and contrast are very different between the two input images, Laplacian or gradient pyramid methods usually produced a very good fused image.

5. CONCLUSION

We presented a new image fusion method, advanced DWT (*a*DWT), which incorporated PCA (principle component analysis) and morphological processing into a regular DWT fusion procedure. We compared the image fusion performance of six common methods (five pyramid methods and a regular DWT method) and our novel method based on four important quantitative measures – the root mean square error (RMSE), the entropy, the spatial frequency and the image quality index. Overall, across the four different kinds of imagery, the *a*DWT performed the best. Different image sources vary considerably in their intensities, contrast, noise, and intrinsic characteristics; therefore a big challenge for a fusion algorithm is to perform well across a variety of image sources, thus *a*DWT is a very promising method to meet this goal.

REFERENCES

1. Genderen, J. L. VAN, and Pohl, C., "Image fusion: Issues, techniques and applications", *Intelligent Image Fusion, Proceedings EARSeL Workshop*, edited by J. L. van Genderen and V. Cappellini, pp. 18-26, Strasbourg, France, September 1994.
2. Keys, L. D., Schmidt, N. J., and Phillips, B. E., "A prototype example of sensor fusion used for a siting analysis", *Technical Papers 1990, ACSM-ASPRS Annual Convention, Image Processing and Remote Sensing*, **4**, 238-249.
3. Rogers, R. H., and Wood, L., "The history and status of merging multiple sensor data: an overview", *Technical Papers 1990, ACSM-ASPRS Annual Convention, Image Processing and Remote Sensing*, **4**, 352-360.
4. 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, 1995.
5. T. Pu and G. Ni, "Contrast-based image fusion using the discrete wavelet transform," *Optical Engineering*, vol. 39

no. 8, pp. 2075–2082, 2000.
6. D.A. Yocky, "Image merging and data fusion by means of the discrete two-dimensional wavelet transform," *J. Opt. Soc. Am. A*, vol. 12, no. 9, pp. 1834–1841, 1995.
7. J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala, and R. Arbiol, "Image fusion with additive multiresolution wavelet decomposition; applications to spot+landsat images," *J. Opt. Soc. Am. A*, vol. 16, pp. 467–474, 1999.
8. F. Jahard, D.A. Fish, A.A. Rio, C.P. Thompson, "Far/near infrared adapted pyramid-based fusion for automotive night vision," *Sixth International Conference on Image Processing and its Applications,* pp. 886–890, 1997.
9. B. Ajazzi, L. Alparone, S. Baronti, and R. Carla, "Assessment of pyramid-based multisensor image data fusion," *Image and Signal Processing for Remote Sensing*, S.B. Serpico, Ed., vol. **4**, pp. 237–248, 1998.
10. C. POHL and J. L. VAN GENDEREN, "Review article: Multisensor image fusion in remote sensing: concepts, methods and applications", *Int. J. Remote Sensing*, Vol. 19, No. 5, 823-854, 1998.
11. Rockinger, O.: "Pixel level fusion of image sequences using wavelet frames", *Proceedings of the 16th Leeds Applied Shape Research Workshop*, Leeds University Press, pp. 149-154, 1996.
12. Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing (Second Edition)*, Prentice Hall, Upper Saddle River, NJ, 2002.
13. P. J. Burt and E. Adelson, "The Laplacian Pyramid As a Compact Image Code", *IEEE Trans. Communications*, Vol. Com-31, No. 4, pp. 532-540, April 1983.
14. P. J. Burt, "The pyramid as structure for efficient computation", *Multiresolution Image Processing and Analysis* (A. Rosenfeld, ed.), pp.6-35, Springer-Verlag, New York/Berlin, 1984.
15. P. J. Burt and E. H. Adelson, "Merging images through pattern decomposition", *Proceedings of SPIE*, vol. 575, pp.173-182, 1985.
16. A. Toet, "Image fusion by a ratio of low pass pyramid", *Pattern Recognition Letters*, vol.9, no.4, pp.245-253, 1989.
17. A. Toet, L. J. Van Ruyven, and J. M. Valeton, "Merging thermal and visual images by contrast pyramid", *Optical Engineering*, vol.28, no.7, pp.789-792, 1989.
18. A. Toet, "Hierarchical image fusion", *Machine Vision and Applications*, vol.3, no.1, pp.1-11, 1990.
19. A. Toet, "Multiscale contrast enhancement with application to image fusion", *Optical Engineering*, vol.31, no.5, pp.1026-1031, 1992.
20. P. J. Burt, "A gradient pyramid basis for pattern-selective image fusion", *Society for Information Display Digest of Technical Papers*, vol.16, pp.467-470, 1985.
21. P. J. Burt and R. J. Lolczynski, "Enhanced image capture through fusion", *Proceedings of the 4th International Conference on Computer Vision*, pp.173-182, Berlin, Germany, 1993.
22. S. Richard, F. Sims, and M. A Phillips, "Target signature consistency of image data fusion alternatives", *Optical Engineering*, vol.36, no.3, pp.743-754, 1997.
23. A. Toet, "A morphological pyramid image decomposition", *Pattern Recognition Letters*, vol.9, no.4, pp.255-261, 1989.
24. G. K. Matsopoulos, S. Marshall, and J. Brunt, "Multiresolution morphological fusion of MR and CT images of the human brain", *Proceedings of IEE - Vision, Image and Signal Processing*, vol.141, no.3, pp.137-142, 1994.
25. G. K. Matsopoulos and S. Marshall, "Application of morphological pyramids: fusion of MR and CT phantoms", *Journal of Visual Communication and Image Representation*, vol.6, no.2, pp.196-207, 1995.
26. Eskicioglu, A.M., Fisher, P.S, "Image quality measure and their performance", *IEEE Trans. Comm.*, Vol. 43, No. 12, 2959-2965, 1995.
27. Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, March 2002.
28. Gemma Piella and Henk Heijmans, "A New Quality Metric for Image Fusion", *2003 International Conference on Image Processing*, Barcelona, Spain, Sept. 14, 2003.
29. Edward A. Essock, Michael J. Sinai, Jason S. McCarley, William K. Krebs, and J. Kevin DeFord, "Perceptual Ability with Real-World Nighttime Scenes: Image-Intensified, Infrared, and Fused-Color Imagery", *Human Factors*, Vol. 41, No. 3, pp 438-452, September 1999.
30. Robert R. Hoffman, Arthur B. Markman, *Interpreting Remote Sensing Imagery: Human Factors*, Lewis Publishers, Boca Raton, Florida, February 2001.