Abstract— Image fusion is the technique of collecting and joining together the best visible and clear image parts to form the best resultant image. The images may be acquired from two different sources or at two different angles or the two different levels of focus. The earlier image fusion techniques implemented fixed levels of decomposition and this acted as the rigid rule for complete image. The image parts which are required being decomposed to lower or higher decomposition levels where improperly decomposed and produced non-optimized results. Whereas, in my proposed algorithm there is an implementation of the quantized frequency technique which optimize and decompose to a frequency quanta as required according to the pixel intensity and clarity. The decomposition levels are selected using frequency partitioning. The frequency partitioning provides selection levels of decomposition described in values from 0.0 to 1.0. This quantized system is flexible up to 0.001 values this certainly increases the CPU overhead but provides sharper results with detailed information signal.

Keywords— Image Fusion, Discrete Wavelet Transform (DWT), Frequency portioning, quantization.

I. INTRODUCTION

Image fusion defines the process of fusing visual information from a number of images into a single fused image. It is part of the much broader subject of multi-sensor information fusion, which has attracted a considerable amount of research attention in the last two decades.

Multi-sensor information fusion utilizes information obtained from a number of different sensors surveying an environment [8].

II. TYPES OF IMAGE FUSION TECHNIQUE

Image fusion methods can be broadly classified into two - spatial domain fusion and transform domain fusion. The fusion methods such as averaging, Brovey method [21], principal component analysis (PCA) [18] and HIS [4] based methods fall under spatial domain approaches. Another important spatial domain fusion method is the high pass filtering based technique. Here the high frequency details are injected into up sampled version of MS images.
The disadvantage of spatial domain approaches is that they produce spatial distortion in the fused image [2]. Spectral distortion becomes a negative factor while we go for further processing, such as classification problem. Spatial distortion can be very well handled by transform domain approaches on image fusion. The multi-resolution analysis has become a very useful tool for analysing remote sensing images.

The discrete wavelet transform [23] has become a very useful tool for fusion. Some other fusion methods are also there, such as Gaussian pyramid based [21], curve let transform based [21], etc. These methods show a better performance in spatial and spectral quality of the fused image compared to other spatial methods of fusion.

Image pyramids have been initially described for multi-resolution image analysis and as a model for the binocular fusion in human vision [5]. A generic image pyramid is a sequence of images where each image is constructed by low pass filtering and sub sampling from its predecessor. Due to sampling, the image size is halved in both spatial directions at each level of the decomposition process, thus leading to a multi-resolution signal representation.

The image pyramid helps to depict image representation in two pyramids: The smoothing pyramid (Gaussian Pyramid) containing the averaged pixel values, and the difference pyramid containing the pixel differences, i.e. the edges (Laplacian Pyramid)[1]. So the difference pyramid can be viewed as a multi-resolution edge representation of the input image.

The actual fusion process can be described by a generic multi-resolution fusion scheme which is applicable both to image pyramids and the Discrete Cosine Transformation (DCT) approach. Also computation of nonlinear pyramids, such as the ratio and contrast pyramid, where the multistage edge representation is computed by a pixel-by-pixel division of neighbouring resolutions.

In the first part of the technique the two images captured from different sensors called multilevel sensors [9] [10] is captured and saved. The two images loaded are processed to form the Gaussian pyramids. This is done primarily with the help of the Discrete Cosine Transform (DCT) [14][15] and Discrete Wavelet Transform (DWT) [11] [17]. My paper proposes a frequency portioning approach which is a fundamental alteration in the previous fixed level based approach. The fixed level based approach performed uniform and linear decomposition on each pixel of the image irrespective of the pixel intensity at that point whereas the frequency portioning approaches provide more precise level for each pixel.

This is done firstly by expressing the limits of the frequency \( f \), i.e. from 0.0 to 1.0 with intervals of 0.1 thereby creating 10 distinctive levels. After the declaration of the image frequency levels, the image is down sampled using a block size of 8x8, which is done using discrete cosine transform (DCT).

### 2.1 Discrete Cosine Transform (DCT)

Discrete cosine transform (DCT) coefficients are concentrated in the low frequency region; hence, it compacts more image values and edges contribute to high frequency coefficients.

The signal energy due to smooth regions is compacted mostly into DC coefficients; hence edges in the spatial domain can only contribute energy to a small number of AC coefficients[3]. The 2D discrete cosine transform \( \mathbf{Z}(u,v) \) of an image or 2D signal \( \mathbf{z}(x,y) \) of size \( M \times N \) is defined as [10]:

\[
\mathbf{Z}(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \mathbf{z}(x,y) \cos \left( \frac{\pi (2x+1) u}{2M} \right) \cos \left( \frac{\pi (2y+1) v}{2N} \right), \quad 0 \leq u \leq M-1, \quad 0 \leq v \leq N-1
\]

Where \( \alpha(u) = \left\{ \begin{array}{ll} \frac{1}{\sqrt{M}}, & u = 0 \\ \frac{1}{\sqrt{M}}, & 1 \leq u \leq M-1 \end{array} \right. \) and \( \alpha(v) = \left\{ \begin{array}{ll} \frac{1}{\sqrt{N}}, & v = 0 \\ \frac{1}{\sqrt{N}}, & 1 \leq v \leq N-1 \end{array} \right. \) are discrete frequency variables \((x,y)\) pixel index. Similarly, the 2D inverse discrete cosine transform is defined as:

\[
\mathbf{z}(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v) \mathbf{Z}(u,v) \cos \left( \frac{\pi (2x+1) u}{2M} \right) \cos \left( \frac{\pi (2y+1) v}{2N} \right), \quad 0 \leq x \leq M-1, \quad 0 \leq y \leq N-1
\]

Image resize can be done in either spatial domain or DCT domain. Image resizing in spatial domain is computationally complex than transform domain. In DCT
domain, high frequency (HF) coefficients are truncated for down sampling and assuming HF coefficients to be zero for up sampling[6].

To down sample the image by a factor of two, the following procedure is carried out as shown below:

The image is divided by non-overlap blocks of size 8x8 and then transformed into DCT domain[]. The 4x4 low frequency (LF) coefficients out of each 8x8 DCT block \{I(i,m,n), 0 \leq m, n \leq 2, i=1,2, \ldots, 4\} are formed as result. A 4x4 IDCT is applied on the LF coefficients to get down sampling. In this way, four consecutive 8x8 blocks become four consecutive 4x4 blocks in spatial domain. This image in spatial domain can be down sampled by repeating the same procedure.

Both the images are down sampled to different frequency level, where decomposition level is selected as per the image pixel intensity of the image I1 and I2.

Let, there are two images (I1 and I2) to be fused. Pyramid construction is done for each image.

The decomposition level k is provided by the frequency f for 1st image given by:

\[ P_k \rightarrow \{z_k, l_0, l_1, \ldots, l_{k-1}\} \]

and similarly for of 2nd image

\[ P_k \rightarrow \{z_k, l_0, l_1, \ldots, l_{k-1}\} \]

Then the image fusion rule is as follows:

At kth level, \[ f_z = \frac{1}{2} z_k + \frac{1}{2} z_k \]

For k-1 to 0 levels

\[ f_z = \frac{1}{2} I_{k-1} + E(f_z) \]

Where

\[ f_z = \left\{ \begin{array}{ll}
1 & l_{k-1} \\
2 & l_{k-1} \leq |l_{k-1}| < |l_{k-1}| \\
2 & l_{k-1} = |l_{k-1}| \\
1 & l_{k-1} = |l_{k-1}| \\
\end{array} \right. \]

and the magnitude comparison is done on corresponding pixels.

Using Reduction Function [19] R: \[ Z_k = R(Z_{k-1}) \]

Using Expand Function [22] E:

\[ z_{k+1} = E(z_k) \]

\[ l_k = z_k - E(z_{k+1}) \]

where \( l_0, l_1, \ldots, l_{k-1} \) are image pyramids that contain band pass filtering images and keeping these records to utilize on reconstruction process and \( z_k \) is the coarser level image.

At coarser level

\[ \hat{z}_{k-1} = l_{k-1} + E(\hat{z}_k) \]

As a result, the pyramid \( l_{k-1} \rightarrow z_0 \) is the fused image.

III. PROPOSED TECHNIQUE

The Proposed Algorithm for image fusion using frequency partitioning

Step 1: Load the first image (I1)
Step 2: Load the second image (I2)
Step 3: Declare the frequency level f from 0.0 to 1.0 with intervals of 0.1 each.
Step 4: Down sample the images acquired in step 1 and 2 using DCT and by selecting frequency level f from step 3 according to the pixel intensity.
Step 5: The two decomposed images are fused using pyramid \( f_z = z_0 \)
Step 7: The fused image is up sampled using inverse DCT.
Step 8: Calculate parameters PSNR, MSE, SF and SD using the original image.

After the fusion of the images, the resultant is upsamplied to original size. Up sampling the image by a factor of two can be done by reversing the down sampling procedure. The image to be up sampled by using a factor of two is divided into 4x4 blocks. Four consecutive 4x4 blocks are transformed into DCT domain as shown in figure below. These are treated as Low Frequency (LF) coefficients.
and used as the LF components in the 8x8 blocks and the reaming High Frequency (HF) coefficients are assumed to be zero.

Then consecutive 8x8 blocks in DCT domain are converted into spatial domain by applying 8x8 IDCT. This procedure is also called expand function [22].

Now the image is analysed for the parameters PFE, SD, SF and PSNR [7] with reference to the original image. The parameter results have been depicted in the next chapter.

IV. QUALITY EVALUATION METRICS

Qualitative measures are used for accurate and meaningful assessment of fused images.

4.1 Peak Signal to Noise Ratio (PSNR)

PSNR will be high when the fused and reference images are alike. Higher value means better fusion. It is computed as:

$$PSNR = 10 \log_{10} \left( \frac{L^2}{RMSE} \right)$$

Where L is the number of gray levels in the image.

4.2 Standard Deviation (SD)

Important index to weight the information of image, it reflects the deviation degree of values relative to mean of image. The greater the SD, more dispersive the gray grade distribution is. Standard deviation would be more efficient in the absence of noise [30]. An image with high contrast would have a high standard deviation. It is calculated using the formula

$$SD = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (I_i (i,j) - \mu)^2}$$

4.3 Spatial Frequency (SF)

SF indicates the overall activity level in the fused image. The spatial frequency for the fused image $I_f$ of dimension MxN is defined as follows:

Row frequency:

$$RF = \sqrt{\frac{1}{MN} \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} [I_f(i,j) - I_f(i+1,j)]^2}$$

Column Frequency:

$$CF = \sqrt{\frac{1}{MN} \sum_{j=1}^{N-1} \sum_{i=1}^{M-1} [I_f(i,j) - I_f(i,j+1)]^2}$$

Spatial Frequency:

$$SF = \sqrt{RF^2 + CF^2}$$

4.4 Percentage Fit Error (PFE)

$$PFE = \frac{\text{norm}(I_r - I_f)}{\text{norm}(I_r)} \times 100$$

where norm is the operator to compute the largest singular value. It is computed as the norm of the difference between the corresponding pixels of reference and fused images to the norm of the reference image.

Table 1 Results

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Image Name</th>
<th>PSNR</th>
<th>SF</th>
<th>SD</th>
<th>PFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Plane</td>
<td>43.567547</td>
<td>15.556524</td>
<td>48.972577</td>
<td>49.872577</td>
</tr>
<tr>
<td>2.</td>
<td>Earth</td>
<td>43.115404</td>
<td>57.056126</td>
<td>53.541137</td>
<td>25.156901</td>
</tr>
<tr>
<td>3.</td>
<td>City</td>
<td>43.260216</td>
<td>14.763783</td>
<td>33.427962</td>
<td>18.147885</td>
</tr>
<tr>
<td>5.</td>
<td>Toy</td>
<td>42.561000</td>
<td>11.838051</td>
<td>49.427425</td>
<td>31.144479</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

In this paper, I implemented image fusion with the variable quantization approach. The frequency partitioning is done with the help of frequency factor f, which can be sub divided into the smaller quanta i.e. the frequency can be distributed from 0.0 to 1 having quanta of 0.1 each. With this frequency factor, images from multiple sensors with different focus or multi focus image scan be fused together. The variable quantization of the frequency for each pixel based on the estimation induced less amount of noise that proved to be the high efficiency factor of the image fusion technique.

REFERENCES

[1] A.Umamaheshvari1,K.Thanushkodi. “IMAGE FUSION TECHNIQUES” ECE Department, Dr.N.G.P. Institute of Technology Coimbatore, Tamilnadu, India july 2010
[2] Bo Huang, Associate Member, IEEE, Huihui Song, Hengbin Cui, Jigen Peng, and Zongben Xu. “Spatial and Spectral Image Fusion Using Sparse Matrix Factorization”, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.
image fusion using self-similarity and depth information

Jamal Saeedi, Karim Faez. "Infrared and visible image fusion using fuzzy logic and population-based optimization." Electrical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), 424, Hafez Ave., Tehran, Iran 2012.


Shih-Gu Huang. "Wavelet for Image Fusion." Graduate Institute of Communication Engineering & Department of Electrical Engineering, National Taiwan University.

