A Guided Filtering Fusion with SIFT for Multi-Focus Image Dataset

Firdous Qureshi[1] , Sriram Yadav[2], Avinash Sharma[3]

[1]M.Tech (CSE, Student), [2]Associate Professor (H.O.D), [3]Associate Professor
Department of CSE, MITS, Bhopal

Abstract

A fast and effective image fusion method is proposed for creating a highly informative fused image through merging multiple images. The proposed method GFF-SIFT is based on a two-scale decomposition of an image into a base layer containing large scale variations in intensity, and a detail layer capturing small scale details. SIFT method is used to find the feature descriptor of image dataset. A novel guided filtering and SIFT-based weighted average technique is proposed to make full use of spatial consistency for fusion of the base and detail layers individually. Experimental results demonstrate that the proposed method can obtain state-of-the-art performance for fusion of multi-focus images.

Keywords

Image fusion, Guided Filtering Fusion, Scale-Invariant Feature Transform, Weighted Average

1. Introduction

The development of information technologies makes it convenient for people to transmit mass data through Internet. However, it also provides vast opportunities for hackers to steal valuable information. Therefore, security becomes an important issue. Digital data hiding can hide sensitive information into multimedia for secure communications. Most multimedia data hiding techniques will distort the cover media in order to insert the additional information. But as they are in medical field or satellite images, a very small distortion matters a lot. Image embedding is considered as insertion of an image in an image in a way that prevents the observer knowing that the hidden image exists in the image. The image that isused to carry the embedded data is called host image. The embedded data is referred to as signature. Since the embedded data is in the form of an image, it is called as signature image. The signature image is often smaller than the host image. Image embedding is achieved by modifying the content of host image. When the embedding process is complete, the host image is slightly changed. The goal of image embedding is to ensure that embedded data can be recovered. The signature image is only recoverable by the owner who has the key to decoding the hidden data. Most multimedia data-embedding techniques modify and, hence, distort the host signal in order to Insert the additional information. The fusion right on the source images, which often have serious side effects such as reducing the contrast. People found that it would be better to perform the fusion in the transform domain. Pyramid transform appears to be very useful for this purpose. The basic idea is to construct the pyramid transform of the fused image from the pyramid transforms of the source images, then the fused image is obtained by taking inverse pyramid transform. Here are some major advantages of pyramid transform: It can provide information on the sharp contrast changes, and human visual system is especially sensitive to these sharp contrast changes. It can provide both spatial and frequency domain localization.

2. System Model

The basic system model of proposed method is strictly based on SIFT (Scale Invariant Feature Transform) and GFF (Guided Filtering Fusion) technique. SIFT Feature Descriptor detector has four main stages namely as Scale- space extrema detection, Key point localization, Orientation computation, Key point descriptor extraction [8]. In SIFT, we have to perform following process. Assume that, Take a 16 x16 window area that around interest node. Separate into a 4x4 grid of rectangular cells. Compute histogram of image gradient directions in each rectangular cell (8 bins each). 16 histograms x 8 orientations = 128 feature descriptors [8]
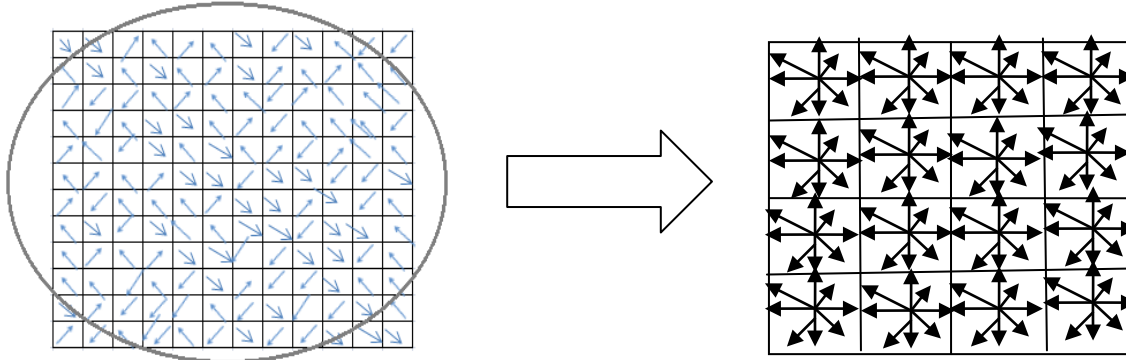

Fig 1: Process for finding keypoint descriptor from an image using SIFT procedure

Guided filtering fusion is a technique of obtain fused image through weight map construction [1]. Basically, the basic steps of this technique is as follows

- Two Scale image decomposition
- Weight map construction
- Two Scale Image reconstruction

In first step, we have to consider two source multi-focus images, now find out base and detail layer using average filtering approach, these layers are further uses for weighted average of base and detail layer. In second step, Construct Weight map of an image on the basis of saliency map on source multi-focus image, then apply guided filter method, in this method source images are considered as guidance images and weight map images taken as input images [4]. In third step, calculate weighted average of output of first step base-detail layer of source image and output of second step base-detail layer of weight map through guided filtering method.
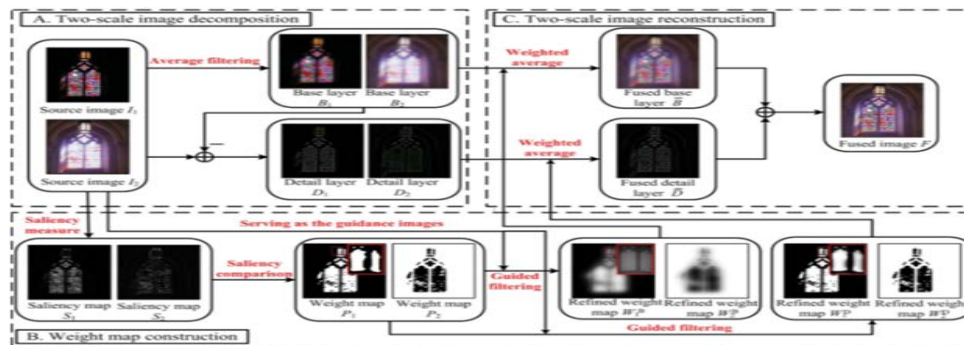

Fig 2: Process of obtain fused image using guided filtering fusion (GFF)

3. Previous Work

A large number of image fusion methods [4]–[7] have been proposed in literature. Among these methods, multi-scale image fusion [5] and data-driven image fusion [6] are very successful methods. They focus on different data representations, e.g., multi-scale co-efficient [5], [6], or data driven decomposition co-efficient [6], [7] and different image fusion rules to guide the fusion of co-efficient. The major advantage of these methods is that they can well preserve the details of different source images. However, these kinds of methods may produce brightness and color distortions since spatial consistency is not well considered in the fusion process. To make full use of spatial context, optimization based image fusion approaches, e.g., generalized random walks [3], and Markov random fields [5] based methods have been proposed. These methods focus on estimating spatially smooth and edge-aligned weights by solving an energy function and then fusing the source images by weighted average of pixel values. However, optimization based methods have a common limitation, i.e., inefficiency, since they require multiple iterations to find the global optimal solution. Moreover, another drawback is that global optimization based methods may over-smooth the resulting weights, which is not good for fusion.

SIFT descriptors, proposed by Lowe in [8], present distinctive invariant features of images that consist of location, scale, orientation, and feature vector. The scale and location of SIFT descriptors are determined by maxima and minima of difference-of-Gaussian images. One orientation is assigned to each SIFT descriptor according to the dominant direction of the local gradient histogram. The feature vector is a 128-dimension vector that characterizes a local region by gradient histogram in different directions. Since SIFT descriptors have a good interpretation of the response properties of complex neurons in the visual cortex [15] and an excellent practical performance, they have been extensively applied to object recognition, image retrieval, 3D reconstruction, annotation, watermarking, and so on.

4. Proposed Methodology

The proposed method of this paper is GFF-SIFT (Guided Filtering based Fusion with Scale Invariant Feature Transform). Recently, edge-preserving filters [2], [3] have been an active research topic in image processing. Edge-preserving smoothing filters such as guided filter [4], weighted least squares [5], and bilateral filter [7] can avoid ringing artifacts since they will not blur strong edges in the decomposition process. Among them, the guided filter is a recently proposed edge-preserving filter, and the computing time of which is independent of the filter size. Furthermore, the guided filter is based on a local linear model, making it qualified for other applications such as image matting, up-sampling and colorization [6]. In this paper, the SIFT method applied on image for finding the keypoint descriptor [8]. Images of one scene may be taken from different viewpoints or may suffer transformations such as rotation, noise etc [2]-[3]. So it is likely that two images of the same scene will be different. The task of finding similarity correspondences between two images of the same scene or object has thus become a challenging problem in a number of vision applications. Such applications range from image registration, camera calibration, object recognition, scene localization in navigation systems, image retrieval based search engines etc. For image matching, extraction of such information (i.e. features) is required from the images which can provide reliable matching between different viewpoints of the same image. Feature detection occurs within an image and seeks to describe only those parts of that image where we can get unique information or signatures (i.e. feature descriptors). During training, feature descriptors are extracted from sample images and stored. In classification, feature descriptors of a query image are then matched with all trained image features and the trained image giving maximum correspondence is considered the best match. Feature descriptor matching can be based on distances such as Euclidean, Mahalanobis or distance ratios. For detect feature descriptors, we use SIFT method [6]. SIFT Feature Descriptor detector has four main stages namely as

Scale- space extrema detection

Key point localization

Orientation computation

Key point descriptor extraction.

Now apply guided filter for image fusion. In theory, the guided filter assumes that the filtering output O is a linear transformation of the guidance image I in a local window ω_k centered at pixel k.
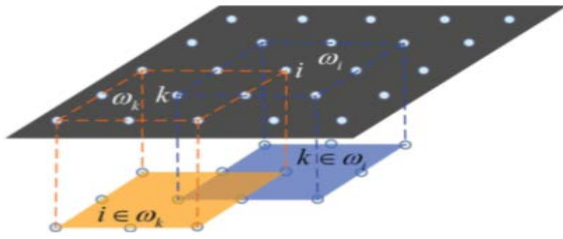

Fig 3: illustration of window choice

$$O_i = a_k I_i + b_k \quad \forall i \in \omega_k \qquad (1)$$

Where, ω_k is a tile window of size (2r+1)×(2r+1). The linear value coefficients a_k and b_k are constant value in ω_k and can be calculated by reducing the squared difference between the output picture O and the input picture P.

$$E(a_k, b_k) = \sum_{i \in \omega_k} ((a_k I_i + b_k - P_i)^2 + e a^2_k) \qquad (2)$$

Where, e is a regularization argument supply by the user. The coefficients value a_k and b_k can be directly solved by linear regression value [6] as follows:

$$a_k = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_k} I_i P_i - \mu_k \overline{P_k}}{\delta_k + e} \qquad (3)$$

$$b_k = \overline{P_k} - a_k \mu_k \qquad (4)$$

where μ_k and δ_k are the mean and variance value of I in ω_k respectively, $|\omega|$ is the no. of pixels in ω_k, and P_k is the mean of P in ω_k. Next, the result image can be calculated as per equation to (1). As shown in Fig. 2, all local value windows

centered at pixel k in the window ω_i that will contain pixel i. So, the cost of O_i in (1) will change when it is computed in different windows ω_k. To solve this question, all the possible costs of coefficients a_k and b_k are first averaged. Then, the filtering output is estimated as follows:

$$O_i = \overline{a}_i I_i + \overline{b}_i \qquad (5)$$

Where $\overline{a}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} a_k, \overline{b}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} b_k$. In this chapter $G_{r.e}(P, I)$ is used to represent the guided filtering operation, where r and e are the parameters which decide the filter size and blur degree of the guided filter, respectively. Moreover, P and I refer to the input image and guidance image, respectively. Furthermore, when the input is a color image, the filtering output can be obtained by conducting the guided filtering on the red, green, and blue channels of the input image, respectively. And when the guidance image I is a color image, the guided filter should be extended by the following steps. First, equation (1) is rewritten as follows:

$$O_i = a_k^T I_i + b_k, \forall i \in \omega_k \qquad (6)$$

Where, a_k is a 3×1 coefficient vector and I_i is a 3×1 color vector. Then, similar to (3)–(5), the output of guided filtering can be calculated as follows:

$$a_k = (\sum_k + eU)(\frac{1}{|\omega|}) \sum_{i \in \omega_k} I_i p_i - \mu_k \overline{P_k} \qquad (7)$$

$$b_k = \overline{p_k} - a_k^T \mu_k \qquad (8)$$

$$O_i = \overline{a}_i^T I_i + \overline{b}_i \qquad (9)$$

where Σ_k is the 3×3 covariance matrix of I in ω_k, and U is the 3×3 identity matrix.

In following figure the main processes of the proposed guided filtering based fusion method (GFF). First, an average filter is utilized to get the two-scale representations. Then, the base and detail layers are fused

through using a guided filtering based weighted average method.

A. Two-Scale Image Decomposition

As shown in following figure, the source images are first decomposed into two-scale representations by average filtering. The lower layer of each given image is obtained in following way:

$$B_n = I_n * Z \qquad (10)$$

Where, I_n is the nth source image, Z is the average value of filter, and the size of the filter is conventionally fixed. Once the lower layer is found, the detail layer can be found by subtracting the lower layer from the source image.

$$D_n = I_n - B_n. \qquad (11)$$

The two-scale decomposition step concerns at separating each source image into a lower layer consisting the large-scale variations in intensity level and an another layer containing the small - scale details.

B. Weight Map Construction with Guided Filtering

The weight map is constructed as follows as per above figure. First, Laplacian filtering is applied to each given image to obtain the high-pass image H_n.

$$H_n = I_n * L \qquad (12)$$

Where, L is a 3×3 Laplacian filter matrix. Then, the local average of the value of H_n is used to design the saliency maps S_n.

$$S_n = |H_n| * g_{rg,\sigma g} \qquad (13)$$

g is a Gaussian low-pass filter, and the parameters r_g and σ_g are set to 5. The measured saliency maps support good characterization level of the saliency of detail information. Next, the saliency values are compared to determine the weight maps as follows:

$$P_n^k = \{_0^1 \; 1 \text{ if } S_n^k = \max(S_1^k, S_2^k, ..., S_N^k) \text{ , 0 otherwise}$$

$$(14)$$

Where N is number of given images, S_n^k is the saliency value of the pixel k in the nth image. However, the weight value obtain above are usually noisy and not aligned with object boundaries, which may produce artifacts to the fused image. Using spatial consistency is an effective way to solve this problem. Spatial consistency means that if two adjacent pixels have similar brightness or color, they will tend to have similar weights. A popular spatial consistency based fusion approach is formulating an energy function, where the pixels saliencies are encoded in the function and edge aligned weights are enforced by regularization terms, e.g., a smoothness term. This energy function can be then minimized globally to obtain the desired weight maps. However, the optimization based methods are often relatively inefficient. An interesting alternative to optimization based methods is proposed. Guided image filtering is performed on each weight map P_n with the corresponding source image I_n serving as the guidance image.

$$W_n^B = G_{r1,e1}(P_n, I_n)$$
$$W_n^D = G_{r2,e2}(P_n, I_n)$$

$$(15), (16)$$

Where, r_1, e_1, r_2, and r_2 are the parameters of the guided filter, W_n^B and W_n^D are the resulting weight maps of the base and detail layers. Finally, the values of the N weight maps are normalized such that they sum to one at each pixel k.

The motivation of the proposed weight construction method is as follows. According to (1), (3) and (4), it can be seen that if the local variance at a position i is very small which means that the pixel is in a flat area of the guidance image, then a_k will become close to 0 and the filtering output O will equal to $\overline{P_k}$ i.e., the average of adjacent input pixels. In contrast, if the local variance of pixel i is very

large which means that the pixel i is in an edge area, then a_k will become far from zero. As demonstrated in [1], $\nabla O \approx \overline{a} \nabla I$ will become true, which means that only the weights in one side of the edge will be averaged. In both situations, those pixels with similar color or brightness tend to have similar weights. This is exactly the principle of spatial consistency.

Furthermore, as shown in above figure, the base layers look spatially smooth and thus the corresponding weights also should be spatially smooth. Otherwise, artificial edges may be produced. In contrast, sharp and edge-aligned weights are preferred for fusing the detail layers since details may be lost when the weights are over-smoothed. Therefore, a large filter size and a large blur degree are preferred for fusing the base layers, while a small filter size and a small blur degree are preferred for the detail layers.

C. Two-Scale Image Reconstruction

Two-scale image reconstruction consists of the following two steps. First, the base and detail layers of different source images are fused together by weighted averaging

$$\overline{B} = \sum_{n=1}^{N} W_n^B B_n \qquad (17)$$

$$\overline{D} = \sum_{n=1}^{N} W_n^D D_n \qquad (18)$$

Then, the fused image F is obtained by combining the fused base layer $\overline{B}$ and the fused detail layer $\overline{D}$

$$F = \overline{B} + \overline{D} \qquad (19)$$

GFF (Guided Filtering based Fusion) technique is based on generic SIFT method, which is used for finding the keypoint descriptor. Guided Filtering is used for find fused image from image descriptor. This method increases the accuracy of pixel ratio of fused image. Hence, we are explains the proposed GFF-SIFT method. The potential of the GFF-SIFT is illustrated on a 3D object recognition task using the Coil database. The images are either represents by a matrix of their pixel values (bitmap representation) or by a color histogram. In both cases, the proposed system requires feature extraction and performs recognition on images regarded as points of a space of high dimension. The feature extraction is perform by SIFT scheme. We also purpose an extension of the basic color histogram which keeps more about the information contained in the images.

The Algorithm of proposed method is explained below:

[fusimg] = GFF_SIFT(image1, image2)

Step 1: Consider the two multi-focus source images with same resolution.

Step 2: Now we find out the SIFT descriptors of each source image of cell array for images of image dataset. SIFT method perform the following sequence of steps for find the keypoint descriptors for texture feature.

Scale-Space Extreme Detection: The initial step of evaluation finds total all scale-space and different image area in image dataset nodes. It is completely apply effectively by using a Difference-of-Gaussian (DoG) mapping to represents potential interest keypoints of feature descriptors which are scale invariant and orientation in image dataset nodes.

Keypoints Localization: All candidate area of image in selected ROI (Region of Interest), a detailed prototype is fit to analyze keypoints area and its scale-space. Keypoints of image area in image ROI are chooses basis on calculate of existing stability.

Orientation Assignment: One or more orientations task are applied to each keypoints area based on local image data nodes gradient directions. Each and every future image operations are implemented on image keypoint dataset which has been transformed relative to the applied orientation, scale, and location for each feature descriptor,

hence providing invariance to these transformations in image data nodes.

Keypoints Descriptor: The local image gradients value are measured at the choose scale-space in the Region of Interest (ROI) around all keypoints in image dataset points. These are transformed into a presentation that permits for significant levels of local shape, location and orientation and changes in illumination of image dataset points.

Step 3: Above step are perform in repeated form, then all the descriptor of images are store, Now apply Guided Filtering method for obtaining fused images.

Step 4: In Guided Filtering, apply two-scale image decomposition using average filter for find base and detail layer of each source image from image database.

Step 5: Now apply weight map construction, obtain saliency map from source images, now compare saliency map and feature descriptor corresponding images, then find out weight map of base and detail layer of each source image from image database.

Step 6: Now image reconstruction on the base layer and detail layer of each source image. Hence, the reconstruct image is taken as fused image.

5. Experimental Step/Results

A. Experimental Setup:

Experiments are performed on multi-focus image databases, which contains 8 pairs of images including aerial images, outdoor images (natural, industrial) and indoor images (with different focus points and exposure settings), Fig. 4 shows 8 pairs of images of the multi-focus database.

The proposed guided filtering based fusion with scale invariant feature transform method (GFF-SIFT) is compared with guided filtering based fusion (GFF) [1].

Fig 4: Multi-focus image database composed by eight pairs of multi-focus images.

C. Results and Analysis

Image Quality is a characteristic of an image that measures the perceived image degradation (typically, compared to an ideal or perfect image). Image Quality assessment methods can be broadly classified into two categories: Full Reference Methods (FR) and No Reference Method (NR). In FR, the quality of an image is measure in comparison with a reference

image which is assumed to be perfect in quality. NR methods do not employ a reference image.

The results of proposed method are determines on the basis of seven quality metrics. In the following graph, x-axis shows the image index and y-axis shows the determine image quality metrics. Image Index value is actually represents the fused image of considered 8 pair of images. For example, (in Fig 4) image 1 and image 2 are fused as image index value 1, image 3 and image 4 are fused as image index value 2, in a same way image 15 and image 16 are fused as image index value 8 respectively. Results of proposed method (GFF-SIFT) is compare with Guided Filtering based Fusion (GFF) method. Analysis of comparison is performed in terms of following quality metrics.

(a) Mean Squared Error

Mean square error is a measure of image quality index. The large value of mean square means that image is a poor quality. Mean square error between the reference image and the fused image is

$$MSE = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(A_{ij} - B_{ij} \right)^2$$

.......(20)

In fig 5, MSE value of GFF is quite high than proposed method GFF-SIFT, hence fused image quality of GFF-SIFT is better than GFF method.
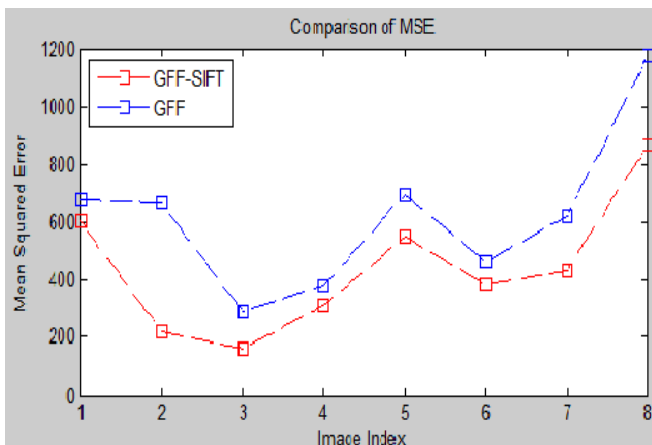

Fig 5: MSE Comparison between GFF-SIFT and GFF

(b) Peak Signal to Noise Ratio:

The ratio between maximum possible powers of the signal to the power of the corrupting noise that creates distortion of image. The peak signal to noise ratio can be represented as

$$PSNR(db) = 20\log \frac{255\sqrt{3mn}}{\sum_{i=1}^{m} \sum_{j=1}^{n} (A_{ij} - B_{ij})^2}$$

..(21) Where

A- fused image, B – perfect image, i – pixel pow index, j – pixel column index, M, N – Number of rows and columns respectively.
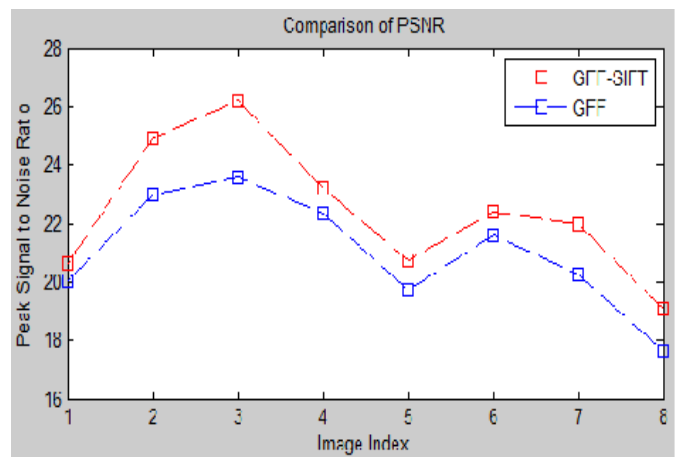

Fig 6: PSNR Comparison between GFF-SIFT and GFF

In fig 6, PSNR value of GFF is quite high than proposed method GFF-SIFT, hence distortion of fused image of GFF-SIFT is less than GFF method.

(c) Normalized Cross Correlation:

Normalized cross correlation is a measure of similarity of two waveforms as a function of the time lag applied to one of them. The cross correlation is similar in nature to the convolution of two functions.

$$NCC = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(A_{ij}.B_{ij})}{A_{ij}^2}$$
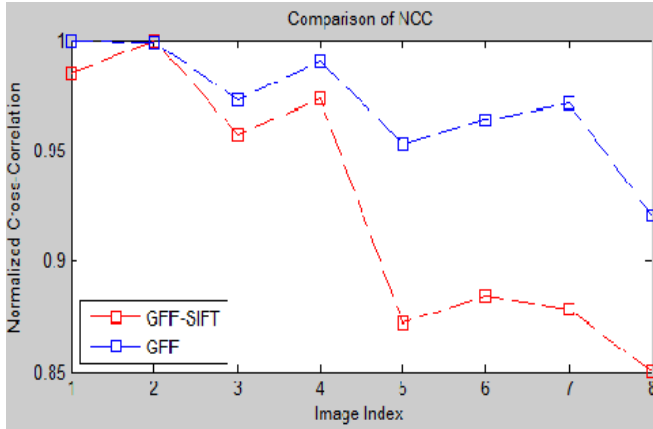
.... (22)

Fig 7: NCC Comparison between GFF-SIFT and GFF

In fig 7, NCC value of GFF is quite low than proposed method GFF-SIFT. Hence, In GFF-SIFT, similarity of fused image and source images is more closes as compare than GFF method.

(d) Maximum Difference:

Difference between any two pixels, such that the larger pixel appear after the smallest pixel. The large value of maximum difference means that image is poor in quality.

$$MD = Max\left(\left|A_{ij} - B_{ij}\right|\right)_{,i=1,2,....,m;\ j=1,2,....,n}\ ...\ (23)$$
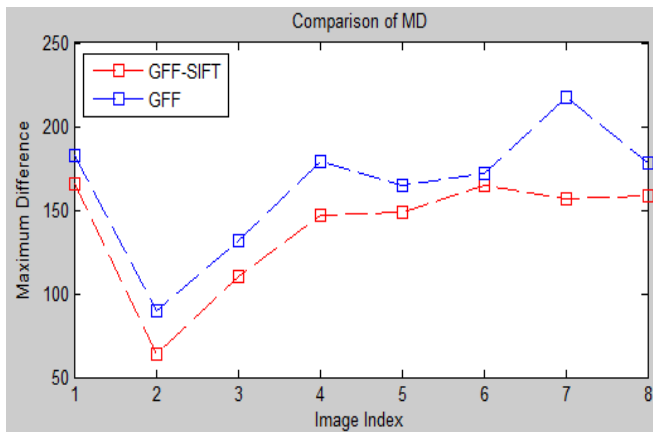

Fig 8: MD Comparison between GFF-SIFT and GFF

In fig. 8, Guided Filtering based Fusion (GFF) perform large value of MD as compare than GFF-SIFT (Guided Filtering based Fusion with Scale Invariant Feature Transform). Hence,

fused image quality is poor for GFF as compare than GFF-SIFT.

(e) Structural Content:

The structural content measure used to compare two images in a number of small image patches the images have in common. The patches to be compared are chosen using 2D continuous wavelet which acts as a low level corner detector. The large value of structural content SC means that image is poor quality

$$SC = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}A_{ij}^{2}}{\sum_{i=1}^{m}\sum_{j=1}^{n}B_{ij}^{2}}\ \ ....\ (24)$$
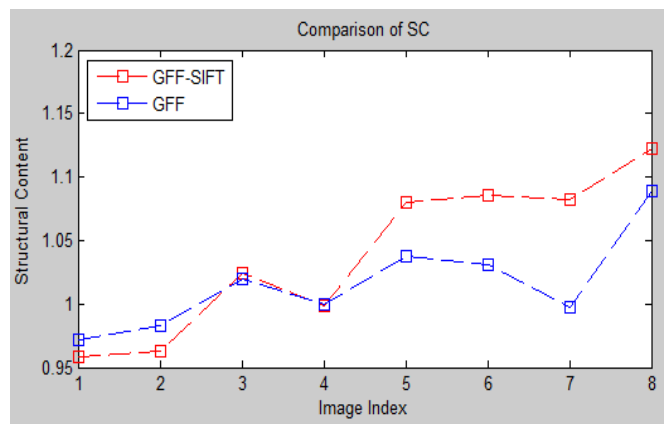

Fig 9: SC Comparison between GFF-SIFT and GFF

In fig. 9, the value of SC for image index 1 and 2 are low in GFF-SIFT as compare than GFF. Hence fused image quality of GFF-SIFT is better than GFF (for image index 1 and 2). The value of SC for image index 3 and 4 are quite same for GFF-SIFT and GFF. Hence fused image quality of GFF-SIFT is similar with GFF (for image index 3 and 4). The value of SC for image index 5, 6, 7 and 8 are high in GFF-SIFT as compare than GFF. Hence fused image quality of GFF-SIFT is poor than GFF (for image index 5, 6, 7 and 8). Therefore, the indoor fused image has better quality but outdoor fused image has poor quality for GFF-SIFT method as compare than GFF method.

6. Conclusion and Future Work

We have presented a novel image fusion method based on guided filtering with scale invariant feature transform. The proposed method utilizes the average filter to get the two-scale representations, which is simple and effective. The weight map construct on the basis of saliency map and feature descriptor of source images. More importantly, the guided filter is used in a novel way to make full use of the strong correlations between neighborhood pixels for weight optimization. Experiments show that the proposed method can well preserve the original and complementary information of multiple input images. Encouragingly, the proposed method is very robust to image registration. Furthermore, the proposed method is computationally efficient, making it quite qualified for real applications. At last, how to improve the performance of the proposed method by adaptively choosing the parameters of the guided filter can be further researched.

6. References

[1] Shutao Li, Xudong Kang, and Jianwen Hu, "Image Fusion with Guided Filtering," IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 7, JULY 2013

[2] D. Socolinsky and L. Wolff, "Multispectral image visualization through first-order fusion," IEEE Trans. Image Process., vol. 11, no. 8, pp. 923–931, Aug. 2002.

[3] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," IEEE Trans. Image Process., vol. 20, no. 12, pp. 3634–3646, Dec. 2011.

[4] S. Li, J. Kwok, I. Tsang, and Y. Wang, "Fusing images with different focuses using support vector machines," IEEE Trans. Neural Netw., vol. 15, no. 6, pp. 1555–1561, Nov. 2004.

[5] G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," Pattern Recognit., vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[6] D. Looney and D. Mandic, "Multiscale image fusion using complex extensions of EMD," IEEE Trans. Signal Process., vol. 57, no. 4, pp. 1626–1630, Apr. 2009.