Journal of Engineering and Applied Sciences 13 (23): 10133-10140, 2018
ISSN: 1816-949X

Optimization of 3D Fused Image Using Feature Matching Based on SIFT and SURF

Saad M. Darwish and Maha M. Ghoneim
Department of Information Technology, Institute of Graduate Studies and Research,
Alexandria University, 163 Horreya Avenue, El-Shatby, P.O. Box 832,
21526 Alexandria, Egypt

Abstract: Collecting information from sensors with different physical characteristics increases the understanding of our surroundings and can provide 3D view. Moreover, 3D color images provide geo-scientists, environment planners, mapping experts and military officers with easy-to-understand color images and useful height information which improves the interpretation of the environment of the areas of interest. This study proposes a novel method to optimize F-transform fused 3D images with feature matching technique based on Scale Invariant Feature Transform (SIFT) followed by Speed Up Robust Feature (SURF). Quantitative and visual results show that a more focused and cleared fused image is obtained after applying feature matching with SIFT followed by further refinement with SURF. The proposed method is robust and independent of scale, light intensity and orientation of camera. It shows that, F-transform is a promising 3-D multisensor image fusion algorithm that surpasses the previous approaches based on Hermite and wavelet transform due to its computational simplicity.

Key words: Image fusion, 3D Fusion system, F-transform, SURF, SIFT, wavelet

INTRODUCTION

Image fusion is a dominant computer vision technique that refers to the process of combining multiple images of a certain scene to produce a single more informative composite image. The output fused image should contain a more useful description of the scene than provided by any of the individual source images and have more useful information for human visual or machine perception (Dong *et al*., 2009). Lately, data fusion has been incorporated to different image processing techniques such as pattern recognition, visual enhancement, object detection and area surveillance (Dong *et al*., 2009).

Multiple sensor fusion method used for the 3D reconstruction of scene geometry remains a highly challenging problem. The main reason is the difficulty of finding multi-view correspondence in case of complex occlusions or visible sparse texture. Although, these difficulties could be partially mitigated by increasing the set of views or resolution of the images, intrinsic problems still remain such as the random noise incorporated by the use of Time of Flight (ToF) active sensors that use laser light to probe the subject (Kim *et al*., 2009).

In the literature, image fusion can be categorized generally into two main categories: single sensor image fusion system and multi-sensor image fusion system (Dong *et al*., 2009). Multi-sensor image fusion system surpasses the limitations of a single sensor vision system by combining the images from these sensors to form a composite image.

In recent years, several image fusion techniques have been proposed, they differ according to different mathematical fields: statistical methods (using aggregation operators, such as the Min Max method (Dong *et al*., 2009), estimation theory (Zaveri and Zaveri, 2010a, b), fuzzy methods (Zaveri and Zaveri, 2011; Kim *et al*., 2009), optimization methods (e.g., neural networks, genetic algorithms) and multi-scale decomposition methods, which incorporate various transforms, for example, discrete wavelet transforms.

Wavelets allow for the hierarchical decomposition of a signal or an image as they are a type of multi-resolution function approximation. As the majority of applications of a fusion scheme deal with features within the image not the actual pixels (Zaveri and Zaveri, 2010a, b), it will be more beneficial to focus on feature information in the fusion process. Region based fusion schemes initially transform pre-registered images using a wavelet

Corresponding Author: Saad M. Darwish, Department of Information Technology, Institute of Graduate Studies and Research, Alexandria University, 163Horreya Avenue, El-Shatby, P.O. Box 832, 21526 Alexandria, Egypt

transform. The transform coefficients are then used to deduce regions incorporating image features. These are then fused based on a simple region property such as average activity.

Similarly, to traditional transforms (Fourier and wavelet), the F-transform (an abbreviation for the fuzzy transform) performs a transformation of an original universe of functions into a universe of their "skeleton models" (vectors of F-transform components) in which further computation is easier. In this study, we show that the F-transform technique is an efficient method for 3-D multisensor image fusion that could overcome the various shortcomings encountered in the previous approaches based on Hermite and wavelet transform such as: their computational complexity; Spectral content of small objects often lost in the fused images; They are not shift invariants and consequently the fusion methods using.

DWT lead to unstable and flickering results. Moreover, the F-transform can be used as an effective means to enhance spatial resolution by enhancing the edges (Zaveri and Zaveri, 2010a, b). This method is computationally simple and can be applied in real time applications.

Scale Invariant Feature Transform (SIFT) method is used to detect and match image feature point. SIFT features with stand image rotation, scaling, translation and change in illumination (Kaur and Agrawal, 2016). Its applications range from object recognition, image mosaic to localization of mobile robots. SIFT feature points extraction consists of following four steps:

Detection of extreme of scale-space: The extreme points of scale space are selected as candidate matching feature image points in SIFT algorithm. Image J(i, j) scale space is defined:

$$S(m, n, \sigma) = J(m, n) * G(m, n) \quad (1)$$

Where:

* = The two dimensional convolution
G (m, n) = A Gaussian function
σ = The standard deviation of normal Gaussian distribution. The extreme is detected in image convolution and Difference of Gaussian (DoG)

Localization of key points: Taylor expansion is constructed as DoG function in scale space. In this process, key-point candidates were produced in large number, out of which some were unstable. Then the detailed fit is performed for the adjacent data to compute scale accuracy, location and the ratio of the principal curvatures. The computed data is used to reject points with low contrast, high noise sensitivity and poor localization along an edge. Lowe mentioned that the location of the extremes to the accuracy of sub-pixels is done by fitting a three dimensional quadratic function to the laplacian based scale space. Taylor expansion based on the function of scale space D (p, q, σ) which is shifted for aligning the origin to the point of interest is used in this approach.

Orientation assignment: Each feature point is assigned a main direction with magnitude M(m, n) and gradient direction θ (m,n).

$$M(i, j) = [(S(m, n+1)-S(m, n-1))^2 + (S(m+1, n)-S(m-1, n))^2]^2$$

$$\theta(m, n) = \arctan\frac{S(m, n+1) - S(m, n-1)}{S(m+1, n) - S(m-1, n)} \quad (2)$$

Descriptor of key point: The area around the critical point of image is divided into blocks, histogram is calculated for every block and the vector with 128 dimensions is generated.

Speeded-Up Robust Features based algorithm (SURF) (Bay *et al*., 2006) which relies on scale space theory, gained popularity due to its computing speed. The detector is based on the Hessian matrix but uses a very basic approximation just as DoG is a very basic Laplacian-based detector. It relies on integral images to reduce the computation time and we therefore, call it the 'Fast-Hessian' detector. The descriptor, on the other hand, describes a distribution of Haar-wavelet responses within the interest point neighborhood. Again, integral images are exploited for speed. Moreover, only 64 dimensions are used reducing the time for feature computation and matching and increasing simultaneously the robustness. Also, a new indexing step is presented based on the sign of the Laplacian which increases not only the matching speed but also the robustness of the descriptor.

It generates a stack in order to restore the same resolution. The local maxima are estimated using Hessian matrix (H). The Hessian matrix of an image at any point $X = (x, y)^t$ is:

$$H(x, \sigma) = \begin{pmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{pmatrix} \quad (3)$$

where, $L_{xx}(x, \sigma)$ represents the convolution of middle point X with the Gaussian filter $\partial^2 g(\sigma)/\partial x^2$. To enhance the computing speed, the box filter approximation is taken instead of Gaussian filter. The determinant of Hessian matrix, ΔH can be reduced to:

$$\Delta H = D_{xx}D_{yy} - (wD_{xy})^2 \quad (4)$$

The response for each spot can be determined by assigning $\omega = 0.9$. A threshold is set for non-maxima suppression to detect the extreme points. The stable feature points are chosen by comparing with the neighboring values followed by the interpolation operation in scale space. Gaussian weighing coefficients are merged with Haar wavelet responses to extract the interest points. The Haar wavelet responses in vertical direction (dy) and in horizontal direction (dx) are summed up along with the absolute value of the response as:

$$V_{sub} = \left(\sum dx,\ \sum dy,\ \sum |dx|,\ \sum |dy|\right) \quad (5)$$

Literature review: Lately, many approaches for 3D image fusion have been introduced. By Yang *et al*. (2010), Qingxiong envisioned a conference room where depth sensors enabled to record in 3D the position and pose of users which enabled them to interact with digital media and contents shown on immersive displays. In this case, fusing different types of sensors was proposed such that relying on passive stereo in highly textured regions while using data from active depth sensors in featureless regions. A Time of Flight (ToF) active depth sensor specifies the signal strength received at each sensor pixel. This allows the computation of a ToF sensor confidence map in addition to a stereo confidence map which was also computed based on local image features. Both confidence maps are then introduced into the cost function which populates the 3D volume created by a plane-sweeping stereo matching algorithm.

On the other hand, ToF depth sensors have many drawbacks as stated by Cahier *et al*. (2011), namely the limited resolution and accuracy and the high frame-to-frame noise they produce off reflective surfaces specifically.

The proposed scheme by Bind *et al*. (2013) utilizes feature based image mosaicing technique. The input images are first stitched together using Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF). Then the merging process is performed using Discrete Wavelet Transform (DWT) to extract the best features from the stitching results. Scale and rotational invariance property can be reached using SIFT.

The goal of this study is to show that higher quality 3D fused image can be obtained using the F-transform technique which has been proved to be a promising and efficient method for 3D image fusion. Unlike the wavelet transform which uses a single "mother wavelet" that determines all basic functions with different shapes, the F-transform performs a transformation of an original universe of functions into a universe of their "skeleton models" (vectors of F-transform components) in which further computation is easier. The resulted image is further refined using SIFT followed by SURF algorithms which have been proved to increase robustness in noisy environment as well as their rotational invariance.

MATERIALS AND METHODS

The main characteristic of the F-transform method is to maintain an acceptable quality in the reconstructed image even under strong compression rates. It was shown that the PSNR of the image compressed with the F-transform method gives PSNR values close to the PSNR obtained using the standard JPEG method under high compression rates (Di Martino *et al*., 2009; Darwish and Ghoneim, 2015). Furthermore, robustness against scale, rotational variation and noisy environment is reached by means of scale invariant feature transform. The following block diagram summarizes the proposed algorithm. The algorithm can be summarized as follows: image registration for the input images. Decompose input registered images $c_1, \ldots, c_K$ into F-transforms and error functions using the one-level decomposition. Apply the fusion operator to the respective F-transform components of the error functions e_i, $i \epsilon I$ and obtain the fused F-transform components of a new error function. Apply SIFT algorithm for the transformed images. Apply SURF algorithm for the transformed images. Reconstruct the fused image from the inverse F-transforms with the fused components of the new image and the fused components of the new error function.

Input images registration: Image registration is the process of overlaying images (two or more) of the same scene taken at different times from different viewpoints and/or by different sensors (Wyawahare *et al*., 2009). The registration geometrically aligns two images (the reference and sensed images). Figure 1 and 2 illustrates the four steps of image registration:

Feature detection: Salient and distinctive objects (closed-boundary regions, edges, contours, line intersections, corners, etc.) are manually or preferably, automatically detected. For further processing, these features can be represented by their point representatives (centers of gravity, line endings, distinctive points), known in the literature as Control Points (CPs).

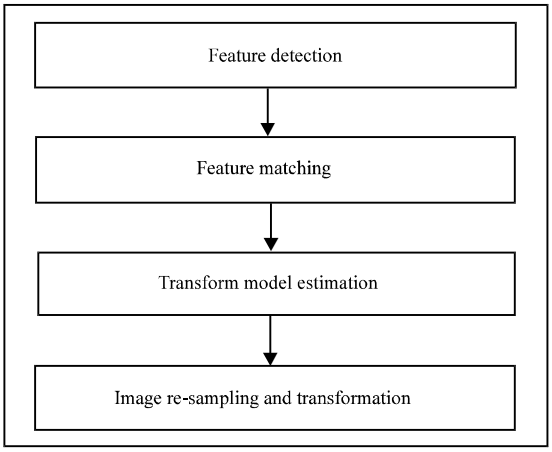

Fig.1: Block diagram of the proposed method

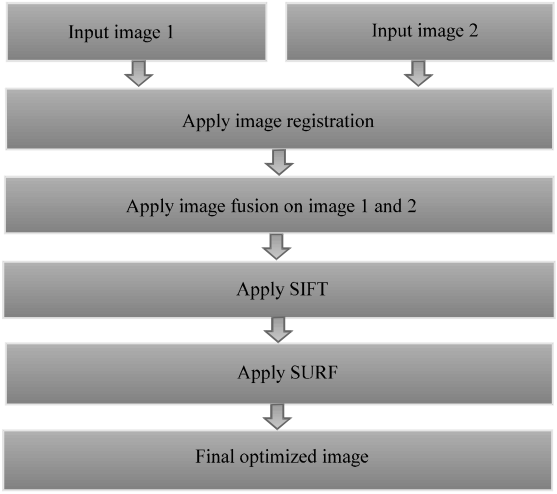

Fig. 2: Images registrations steps

Feature matching: In this step, the correspondence between the features detected in the sensed image and those detected in the reference image is established. Various feature descriptors and similarity measures along with spatial relationships among the features are used for that purpose.

Transform model estimation: The type and parameters of the so-called mapping functions, aligning the sensed image with the reference image are estimated. The parameters of the mapping functions are computed by means of the established feature correspondence.

Image re-sampling and transformation: The sensed image is transformed by means of the mapping functions. Image values in non-integer coordinates are computed by the appropriate interpolation technique (Zitova and Flusser, 2003).

SIFT algorithm: SIFT algorithm is based on feature spotting in scale space. The four major steps of this algorithm are: scale space detection (Kaur and Agrawal, 2016), preliminary confirm the key points, location and the scale as shown in Fig. 3. The middle point is compared with its neighborhood points to detect utmost points. Using Taylor expansion, the extreme points and location are carefully determined using the following quation:

$$D(x) = D + \frac{\partial D^T}{\partial x}x + \frac{1}{2}x^T\frac{\partial^2 D}{\partial x^2}x \quad (6)$$

By the help of key point neighborhoods, the gradient m (x, y) and the direction are estimated for an image L(x, y). Taking the gradient value and characteristic into consideration each sample points is added to the histogram. The direction for the feature points are estimated from the maximum peak values from the histogram.

Feature vectors (Kaur and Agrawal, 2016) are generated which is shown in Fig. 3. The arrow in each cell stands for gradient direction along with the amplitude of pixels. The seed point can be formed by aligning the unidirectional gradients followed by the normalization.

SURF algorithm: In this algorithm (Kaur and Agrawal, 2016), the detector is based on the Hessian matrix but uses a very basic approximation just as DoG is a very basic Laplacian-based detector. It relies on integral images to reduce the computation time and we therefore, call it the 'Fast-Hessian' detector. The descriptor, on the other hand, describes a distribution of Haar-wavelet responses within the interest point neighborhood. Again, integral images are exploited for speed. Moreover, only 64 dimensions are used, reducing the time for feature computation and matching and increasing simultaneously the robustness. Also, a new indexing step is presented based on the sign of the Laplacian which increases not only the matching speed but also the robustness of the descriptor.

F-transform: Generally speaking, the F-transform produces an image by a linear mapping from a set of ordinary continuous/discrete functions over a domain P onto a set of functions within a fuzzy partition of P. We assume that the reader is familiar with the notion of the fuzzy set and how is it represented. Below, we explain the F-transform in more detail. The explanation will be given, for example, of a discrete function that corresponds to the image u.

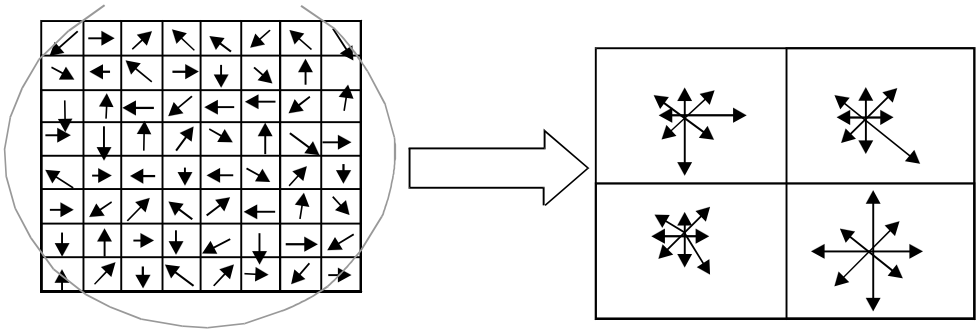

Fig. 3: Feature vector generation

Let u be represented by the discrete function u: P→R of two variables, where P = {(i, j) |i = 1, ... , N, j = 1, ... , M} is an N×M array of pixels and R is the set of reals. If (i, j)∈P is a pixel, then u (i, j) represents its intensity range. The F-transform of u corresponds u to the matrix $F_{nm}(u)$ of F-transform components:

$$F_{nm}(u) = \begin{pmatrix} F(u)_{11} \dots F(u)_{1m} \\ \vdots \quad \ddots \quad \vdots \\ F(u)_{n1} \dots F(u)_{nm} \end{pmatrix} \tag{7}$$

Each component $F[u]_{kl}$ is a local mean value of u over a support set of the respective fuzzy set $A_k \times B_l$. The latter is an element of a fuzzy partition of the Cartesian product of intervals [1, N]×[1, M]. Using the fact that a fuzzy partition of a Cartesian product is the Cartesian product of fuzzy partitions, we first introduce this notion for a single interval and then for a Cartesian product of intervals.

Let [1, N] = {x|≤x≤N} be an interval on the real line R, n≥2, a number of fuzzy sets in a fuzzy partition of [1, N] and h = N-1/n-1 the distance between nodes $x_1, \dots, x_n \in [1, N]$, where $x_1 = 1$, $x_k = x_1 + (k-1)h$, k = 1, ... , n.

Fuzzy sets $A_1, \dots, A_n$: [1, N]→[0, 1] establish a h-uniform fuzzy partition of [1, N] if the following requirements are fulfilled:

- For every k = 1, ... , n, $A_k(x) = 0$ if $x \in [1,N] \setminus [x_{k-1}, x_{k+1}]$, where $x_0 = x_1$, $x^{N+1} = x_N$
- For every k = 1, ... , n, A_k is continuous on $[x_{k-1}, x_{k+1}]$, where $x_0 = x_1$, $x^{N+1} = x_N$
- For every i = 1, ... , N, $\Sigma nk = 1\ A_k(i) = 1$
- For every k = 1, ... ,n, $\Sigma Ni = 1\ A_k(i) > 0$
- For every k = 2, ... , n-1, A_k is symmetrical with respect to the line $x = x_k$

The membership functions of the respective fuzzy sets in a fuzzy partition are called basic functions. The example of triangular basic functions $A_1, \dots, A_n$, n = 2 on the interval [1, N] is given below:

$$A_1(x) = \begin{cases} 1 - \dfrac{(x - x_1)}{h}, & x \in [x_1, x_2] \\ 0, \text{ otherwise} \end{cases}$$

$$A_k(x) = \begin{cases} \dfrac{(x - x_k)}{h}, & x \in [x_{k-1}, x_{k+1}] \\ 0, \text{ otherwise} \end{cases} \tag{8}$$

$$A_n(x) = \begin{cases} \dfrac{(x - x_{n-1})}{h}, & x \in [x_{n-1}, x_n] \\ 0, \text{ otherwise} \end{cases}$$

Note that, the shape (e.g., triangular or sinusoidal) of a basic function in a fuzzy partition is not predetermined and can be chosen according to additional requirements. We now introduce two extreme fuzzy partitions of [1, N] that will be used in the following.

Largest partition: The largest partition contains only one fuzzy set, A_1: [1, N]→[0, 1] such that for all x∈[1, N], $A_1(x) = 1$.

Finest partition: The finest partition is established by N fuzzy sets, $A_1, \dots, A_N$: [1, N]→[0, 1] such that for all k, l = 1, ... , N, k≠l, $A_k(x_k) = 1$ and $A_k(x_l) = 0$.

If fuzzy sets $A_1, \dots, A_n$ establish a fuzzy partition of [1, N] and $B_1, \dots, B_m$ do the same for [1, M], then the Cartesian product $\{A_1, \dots, A_n\} \times \{B_1, \dots, B_m\}$ of these fuzzy partitions is the set of all fuzzy sets $A_k \times B_l$, k = 1, ... , n, l = 1, ... , m. The membership function $A_k \times B_l$: [1, N]×[1, M]→[0, 1] is equal to the product $A_k \_E\, B_l$ of the respective membership functions. Fuzzy sets $A_k \times B_l$, k = 1, ... , n, l = 1, ... , m establish a fuzzy partition of the Cartesian product [1, N]×[1, M]. Let u: P→R and fuzzy sets $A_k \times B_l$, k = 1, ... , n, l = 1, ... , m, establish a fuzzy partition of [1, N]×[1, M]. The (direct) F-transform of u (with respect to the chosen partition) is an image of the mapping F[u]: $\{A_1, \dots, A_n\} \times \{B_1, \dots, B_m\} \to R$ defined by:

$$F[u](A_k \times B_l) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} u(i, j)A_k(i)B_l(j)}{\sum_{i=1}^{N}\sum_{j=1}^{M} A_k(i)B_l(j)} \quad (9)$$

where, k = 1, ... , n, l = 1, ... , m. The value $F[u](A_k \times B_l)$ is called an F-transform component of u and is denoted by $F[u]_{kl}$. The components $F[u]_{kl}$ can be arranged into the matrix representation as in Eq. 7 or into the vector representation as follows: $(F[u]_{11}, \ldots, F[u]_{1m}, \ldots, F[u]_{n1}, \ldots, F[u]_{nm})$.

Fusion: We now proceed with a detailed description of the simple F-transform-based image-fusion algorithm (SA). The fusion is performed between multiple input images from different sensors having multiple viewpoints, which results in a 3D image output. We assume that the image u is a discrete real function, u = u(x, y) defined on the N×M array of pixels. P = {(i, j) |i = 1, ... ,N, j = 1, ... ,M}, so that, u : P→R.

Moreover, let fuzzy sets $A_k \times B_l$, k = 1, ... , n, l = 1, ... , m, where, 0<n = N, 0<m = M establish a fuzzy partition of [1, N]×[1, M]. We begin with the following representation of u on P:

$$u(x, y) = u_{nm}(x, y)+e(x, y) \quad (10)$$

where, 0<n≤N, 0<m≤M:

$$e(x, y) = u(x, y) - u_{nm}(x, y), \forall(x, y)\in P \quad (11)$$

Where:

u_{nm} = The inverse F-transform of u
e = The respective residuum

If we replace e in Eq. 10 by its inverse F-transform eNM with respect to the finest partition of [1, N]×[1, M], the above representation can then be rewritten as follows:

$$u(x, y) = u_{nm}(x, y)+e_{NM}(x, y), \forall(x, y)\in P \quad (12)$$

We call Eq. 12 a one-level decomposition of u. If function u is smooth, then the error function eNM is small and the one-level decomposition Eq. 12 is sufficient for our fusion algorithm. However, images generally contain various types of degradation that disrupt their smoothness. As a result, the error function eNM in Eq. 12 is not negligible and the one-level decomposition is insufficient for our purpose. In this case, we continue with the decomposition of the error function e in Eq. 10. We decompose e into its inverse F-transform e_{nm} (with respect to a finer fuzzy partition of [1, N]×[1, M] with n':n<n_≤N and m':m<m'≤M basic functions, respectively) and a new error function e'. Thus, we obtain the second-level decomposition of u:

$$u(x, y) = u_{nm}(x, y)+e_{nm}(x, y)+e'(x, y),$$
$$e'(x, y) = e(x, y)-e_{nm}(x, y)\forall(x, y)\in P \quad (13)$$

In our research, we use the simple F-transform-based fusion algorithm (SA) which is based the one-level decomposition Eq. 12. The main role in fusion algorithms is played by the so-called fusion operator κ:RK→R, defined as follows:

$$x(x_1, \ldots, x_k) = x_p, \text{ if } |x_p| = \max(|x_1|, ..., |x_k|) \quad (14)$$

Assume that, we are given K≥2 input images $c_1, \ldots, c_K$ with various types of degradation. Our aim is to recognize undistorted parts in the given images and to fuse them into one image. In this section, we describe the algorithm for image fusion based on the one-level decomposition. Each input image c_i, i = 1, ... , K is assumed to be a discrete real function $c_i = c_i$ (x, y) defined on the N×M array of pixels. P = {(x, y) | x = 1, ... , N, y = 1, ... , M}, so that, ci : P→R. Moreover, the set [1, N]× [1, M] is assumed to be partitioned by fuzzy sets $A_k \times B_l$, where, k = 1, ... , n, l = 1, ... , m and 0<n≤N, 0<m≤M. Denote I = {1, 2, ... , K}.

Inverse F-transform: The inverse F-transform of u is a function on P which is represented by the following inversion Eq. 15:

$$u_{nm}(i, j) = \sum_{k=1}^{n}\sum_{j=1}^{m} F[u]_{kl}A_k(i)B_l(j) \quad (15)$$

where, i = 1, ... , N, j = 1, ... , M. It can be shown that, the inverse F-transform u_{n*m} approximates the original function u on the domain P.

Performance evaluation: The output image quality evaluation process consists of objective as well as subjective performance evaluation. In this study, Root Mean Square Error (RMSE), Mutual Information (MI), Normalized Absolute Error (NAE) and Standard Deviation (SD) are used to evaluate the quality of the final images.

RMSE as quality measure: The (RMSE) depicts the deviations between the reference image pixel value R(i, j) and the fused image pixel value F(i, j) where, I, j denote pixel location. It is computed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}[R(i, j)-F(i, j)]^2}{m\times n}} \quad (16)$$

Table 1: Absolute Error (NAE)

Parameters	SIFT	SURF	Final fused image
RMSE(dB)	41.6936	41.962	42.415
MI	1.209	1.264	1.465
EME	8.561	6.332	9.457
NAE	0.147	0.143	0.132

where, m×n is the input image size. The more the RMSE, the better the quality of the reconstructed image.

Mutual Information (MI): It measures the asymmetry between the two desired images and the fluctuation from its mean value MI for two images M (i, j) and N (i, j) can be expressed as:

$$MI = H(M)+H(N)-H(M, N) \quad (17)$$

where, H(M) is the entropy of image M(i, j), H(N) is the entropy of image N(i, j) and H(M, N) is the joint entropy of image M(i, j) and N(i, j).

Enhancement performance measure (EME): It is a quantitative method to measure the image enhancement. In terms of entropy it can be defined with the help of an image which is divided into k_1, k_2 blocks $W_{k,l}$ (i, j):

$$EME = \min_{\varnothing\in\{\varnothing\}}\left(EME(\varnothing)\right) = \min_{\varnothing\in\{\varnothing\}}\left(\frac{1}{k_1k_2}\right)\sum\sum 20\log\frac{1^{w}_{max;\,k,\,l(\varnothing)}-1^{w}_{min;\,k,\,l(\varnothing)}}{1^{w}_{max;\,k,\,l(\varnothing)}+1^{w}_{min;\,k,\,l(\varnothing)}} \quad (18)$$

where, $I^{w}_{max;\,k,\,l}$ and $I^{w}_{mix;\,k,\,l}$ are maximum and minimum of image X (n1, n2).

Normalized Absolute Error (NAE): The normalized absolute error can be used as a metric for image quality metric and is formulated as:

$$NAE = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|A_{i,\,j}-B_{i,\,j}\right|}{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|A_{i,\,j}\right|} \quad (19)$$

where, A-perfect image and B-fused image to be assessed. The objective evaluation of the fused image is depicted in Table 1.

RESULTS AND DISCUSSION

In the proposed techniques two three dimensional rotational images are captured at rotational angle of 10°. After their fusion, the images are processed through scale invariant feature transform and speeded-up robust features algorithms separately in a parallel process.

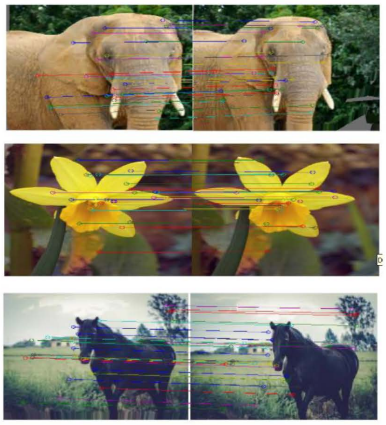

Fig. 4: Result of SIFT algorithm on fused image

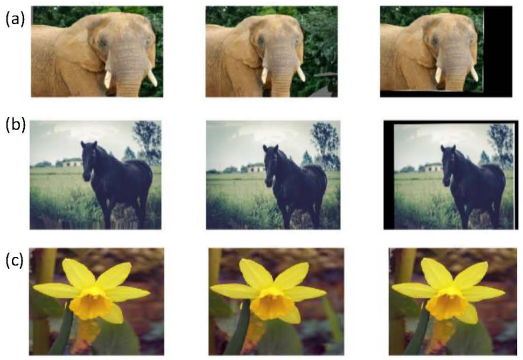

Fig. 5: Algorithm results: a, b) Input images and c) Final enhanced image with SIFT and SURF

Figure 4 and 5, respectively show the response of SIFT algorithm on the fused image and the final fused image after SIFT and SURF. The generated image is shown to be of high contrast, robust towards noise as well as illumination variation. The fused panoramic image is depicted in Fig. 4. SURF algorithm has the distinctive property of illumination invariance along with good scale and rotational invariance property whereas SIFT is more effective algorithm for scale and rotational image stitching. But it cannot cope up with illumination variation. Therefore, the resultant image proves superior as compared to the SIFT as well as SURF algorithms in terms of (RMSE), Mutual Information (MI), Measure of Enhancement (EME) and Normalized Table 1 performance analysis.

CONCLUSION

This study focuses on the application of SIFT and SURF algorithms to the previously F-transformed fused image. After a brief introduction to the theory of F-transform, detailed description of the fusion

algorithm was given. The proposed approach can be successfully applied in cases when input images are available as multi-sensor input images. Both subjective and objective results show that the proposed scheme outperforms other methods based on the wavelet transform. The F-transform has been proved to be an efficient model for the representation of signals. The input images are passed through two robust stitching algorithms, i.e., SIFT and SURF. The scale Invariant feature transform is invariant towards scale and rotational variation. It is also robust towards noisy environment. Speeded-Up robust features algorithm has very similar properties as SIFT. However, it has the properties of illumination invariance and good computational speed. Therefore, the fusion result of these two efficient algorithms gives rise to a panoramic image which carries all the properties of both.

The performance evaluation of proposed technique is done in terms of RMSE, MI, EME and NAE. The proposed method shows superior results as compared to both SIFT and SURF alone.

REFERENCES

Bay, H., T. Tuytelaars and L. Van Gool, 2006. Surf: Speeded up robust features. Proceedings of the 9th European Conference on Computer Vision, May 7-13, 2006, Springer, Berlin, Germany, ISBN:978- 3-540-33833-8, pp: 404-417.

Bind, V.S., P.R. Muduli and U.C. Pati, 2013. A robust technique for feature-based image mosaicing using image fusion. Intl. J. Adv. Comput. Res., 3: 263-268.

Cahier, L., T. Takahashi, T. Ogata and H. Okuno, 2011. Time-of-flight camera based probabilistic polygonal mesh mapping. Inf. Proc. Soc. Jpn., 2: 431-432.

Darwish, S.M. and M.M. Ghoneim, 2015. A novel 3D image fusion approach using F-transform. Proceedings of the 2015 25th International Conference on Computer Theory and Applications ICCTA, October 24-26, 2015, ICCTA, Alexandria, Egypt, pp: 191-196.

Di Martino, F., V. Loia and S. Sessa, 2009. Direct and inverse fuzzy transforms for coding/decoding color images in YUV space. J. Uncertain Syst., 3: 11-30.

Dong, J., D. Zhuang, Y. Huang and J. Fu, 2009. Advances in multi-sensor data fusion: Algorithms and applications. Sens., 9: 7771-7784.

Kaur, G. and P. Agrawal, 2016. Optimisation of image fusion using feature matching based on SIFT and RANSAC. Indian J. Sci. Technol., 9: 1-7.

Kim, Y.M., C. Theobalt, J. Diebel, J. Kosecka and B. Miscusik *et al*., 2009. Multi-view image and tof sensor fusion for dense 3D reconstruction. Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), September 27-October 4, 2009, IEEE, Kyoto, Japan, ISBN:978-1-4244-4442-7, pp: 1542-1549.

Wyawahare, M.V., P.M. Patil and H.K. Abhyankar, 2009. Image registration techniques: An overview. Int. J. Signal Process. Image Process. Pattern Recogn., 2: 11-28.

Yang, Q., K.H. Tan, B. Culbertson and J. Apostolopoulos, 2010. Fusion of active and passive sensors for fast 3d capture. Proceedings of the 2010 IEEE International Workshop on Multimedia Signal Processing (MMSP), October 4-6, 2010, IEEE, Saint Malo, France, ISBN:978-1-4244-8110-1, pp: 69-74.

Zaveri, T. and M. Zaveri, 2010a. A novel two step region based multifocus image fusion method. Intl. J. Comput. Electr. Eng., 2: 86-91.

Zaveri, T. and M. Zaveri, 2011. A novel region based multimodality image fusion method. J. Pattern Recognit. Res., 2: 140-153.

Zaveri, T.H. and M.A. Zaveri, 2010b. Novel hybrid multispectral image fusion method using fuzzy logic. Intl. J. Comput. Inf. Syst. Ind. Manage. Appl., 2: 096-103.

Zitova, B. and J. Flusser, 2003. Image registration methods: A survey. Image Vision Comput., 21: 977-1000.