Survey on Image Fusion: Hand Designed to Deep Learning Algorithms

Heena Patel and Kishor P. Upla, SVNIT, Surat, Gujarat, India

***Abstract*—Image fusion is the process of combining/integrating multiple images to generate the single image having all meaningful information. The input/source images are captured through various sensing devices under different parameter setting. It is impossible to focus on all information or all small objects in single image. Hence, Image fusion methods provide the composite image known as fused image with complementary information. Fused image should be more suitable for human as well as machine perception. Therefore, several methods have been developing to improve the quality of images. Traditional methods include spatial and transform domain based image fusion in which spatial domain includes the fusion methods with pixel, blocks or segmentation based processing. Whereas, transform domain utilizes the transform theories to transform the image in another domain rather than same domain of the input and performs fusion rule on transformed image. Spatial domain methods produce spatial and spectral distortion in the fused image whereas transform methods often perform inadequately when images obtained from different sensor modalities. Recently the concept of deep learning has enlarged to enhance the development in image processing and computer vision problems such as segmentation, classification, super-resolution, etc. Deep learning algorithms such as convolutional neural network (CNN), deep autoencoder (DAE), and deep belief networks (DBN) with different category of images such as multi-modal, multi-resolution, multi-temporal and multi-focus have been proposed for image fusion. Its applications include disease analysis, disaster assessment, providing a complete information for diagnosis, detection of change, etc.**

Keywords—CNN, DAE, DBN, deep learning, neural networks, spatial domain, transform domain

I. INTRODUCTION

The aim of fusion is to create a composite image from the multiple numbers of images which contains complementary details of the same image [1]. Different imaging devices are used to capture the source/input images or it can be possible to generate the source images using the single sensor having different parameter setting on an individual image.

Owing to this benefit, image fusion techniques demonstrate great impact in the number of applications that includes multiple images of the same scene. For example, to identify disease in the medical laboratory, it is essential to physicians to collect more images with the different modality such as images from the magnetic resonance (MR), computed tomography (CT), single photon emission computed tomography (SPECT), etc. In this circumstances, composite the meaningful information from the different sources may more helpful to decrease the effort to accomplish precise analysis. Another important application of fusion is included in digital photography to cover the specific range of camera under the different focal lens setting, remote sensing images are used to achieve high-resolution images, multi-temporal images can be used to detect the changes, etc.

Recently the concept of deep learning (DL) has enlarged to enhance the development in image processing and computer vision problems such as segmentation [2], classification [3], super-resolution [4], etc. Hence, an investigation in the field of fusion of images using the concept of deep learning has become the active hot topic. Deep learning algorithms such as convolutional neural network (CNN), deep autoencoder (DAE), deep belief networks (DBN) with various architectures have been proposed for image fusion. DBN is utilized to extract different specific multiple features from the source images and classify it according to their related features. deep features having very low redundancy can be extracted using DAE. Moreover, CNN is most recently used algorithm and more efficient for image fusion task. As it is known that CNN is used for feature extraction as well as classification problems. Hence, it overcomes the separate processing task of feature extraction and fusion rules. These algorithms are applied on different category of images to achieve the state of art performance.

The image fusion methods include several processing steps such as image transform, feature extraction and fusion rule. Researchers have been doing continuous investigation to improve the development of image fusion methods by introducing effective transform techniques, more robust feature extraction algorithms and more elaborative rules for the final level of fusion. Nevertheless, despite the substantial progress is achieved. In recent years, the field of image fusion is growing and it has become increasingly with some bottlenecks. In traditional methods of fusion, source image processing algorithms are designed manually. It is not hard to extract features of individual and implementation of fusion rules, but it is core issue for the large dataset. The issues of the feature extraction methods and fusion rules are carefully tackled in most recently proposed algorithms. These issues make the fusion method more and more complicated to achieve better performance. However, there are limitations of many factors which include computational cost and implementation difficulty. Therefore, it is somewhat difficult to manually implement an ideal design which fully solves the important issues of image fusion methods.

Moreover, the scarcity of effectively transform of the images is another great challenge in recent image fusion approach. As it is well known that transformation of images is essential to achieve the high quality image in fusion method. The improvement in fusion algorithm is achieved mostly with the image representation theories which make possible to further increase the progress of fusion task. Image representation theories include the different transform methods such as multi-scale transform (MST), sparse representation (SR), wavelet transform (WT), etc. However, still there exist number of defects when these popularized transform methods are used for fusion approach [5]. Therefore, it is necessary to investigate some new approaches for the image representation which are more robust and effective for the task of image fusion.

The objective evaluation is done when merged images quantitative evaluate the quality of the fused image. However, this is not an easy due to the reference images are not available in most image fusion tasks.

Objective metrics of the overall fusion of images can be divided into two groups: metrics, which are based on a fused image and which are based on both i.e. fused image and original images. The first group includes some simple image quality measures such as entropy, standard deviation and spatial frequency. The second group belongs to the issues of image fusion and it is grouped in four categories i.e. metrics based on image fusion, metrics based on information theory, metrics based on structural similarity of images and metrics based on human perception [1]. Evaluation of image fusion algorithms on remote sensing data includes spectral angle mapper (SAM), correlation coefficient (CC), root mean square error (RMSE), erreur relative globale adimensionnelle de synthese (ERGAS), Q4, etc. Spectral and spatial distortion index of panchromatic images are measured by quality with no reference (QNR) index. The evaluation metrics may be different for different applications of image fusion.

II. LITERATURE REVIEW

Traditional image fusion methods include three step framework i.e. Image decomposition (feature extraction), fusion and reconstruction [6,7]. Various image fusion methods have been proposed and generally it is classified into transform domain and spatial domain. Spatial domain includes the fusion methods with pixel, blocks or segmentation based processing such as intensity-hue-saturation (IHS) based fusion, principal component analysis (PCA) based fusion, fusion based on arithmetic combinations, high pass filtering and total probability density fusion. Transform domain utilizes the transform theories to transform the image in another domain rather than same domain of the input and performs fusion rule on transformed image then after inverse transform is applied to get the final output image. Multi-scale transform (MST) theory is most widely used for the transform domain fusion. It includes pyramid based methods, wavelet transform based fusion and curvelet transform based fusion. Next sections describe the detail discussion of spatial domain, transform domain and deep learning based fusion.

A. *Spatial domain fusion*

Methods of spatial domain include simple fusion rules. The methods fall under spatial domain are listed below

1) Simple min/max/average rule based fusion
2) Intensity-hue-saturation (IHS) transform based fusion
3) Principal component analysis (PCA) based fusion
4) Arithmetic combinations

- Brovey transform (BT)
- Synthetic variable ratio technique (SVR)
- Ratio enhancement technique (RE)

5) High-pass filtering method (HPF)
6) Total probability density fusion

1) *Simple min/max/average rule based fusion:* This method performs very basic operation such as selection of pixel from source images, addition and averaging [6, 8]. It is not effective method but it works well when a single sensor is used which contains additive noise and it is also useful for the input images having high contrast. Method of maximum fusion rule selects the pixel having the maximum intensity from the corresponding input images. It is obtained by

$$f(i,j)=\sum_{i=1}^{n}\sum_{j=1}^{m}\max(I_1(i,j),I_2(i,j)) \quad (1)$$

Method of minimum fusion rule selects the pixel having the minimum intensity from the corresponding input images. It is obtained by

$$f(i,j)=\sum_{i=1}^{n}\sum_{j=1}^{m}\min(I_1(i,j),I_2(i,j)) \quad (2)$$

Method of averaging fusion rule performs the average operation on the pixels on the corresponding input images. It is obtained by

$$f(i,j)=\sum_{i=1}^{n}\sum_{j=1}^{m}(I_1(i,j)+I_2(i,j))/2 \quad (3)$$

Where, f represents the fused image, I_1 and I_2 are input images, m and n describe the size of particular image.

2) IHS transform based fusion: This method is useful for color image only. It is first introduced by Chavez and carper *et al.* [9] in 1991. Spectral i.e. (hue and saturation) and spatial i.e. intensity information are effectively separated using this method. Figure 1 shows the IHS based fusion. It first converts the input image into components of intensity, hue and saturation from the RGB image. Intensity component exhibits as a dominant component and defines total brightness of the color. A high resolution panchromatic image is substituted with the intensity component to increase the resolution of final image. Finally, band operation is performed to get the resultant fused image by converting IHS to RGB. Fused image contains spatial detail of panchromatic image having high resolution which is incorporated into it. It provides high spatial quality but it suffers from noise and artifacts. Due to the major limitation of involvement of only three bands, it also provides color distortion.

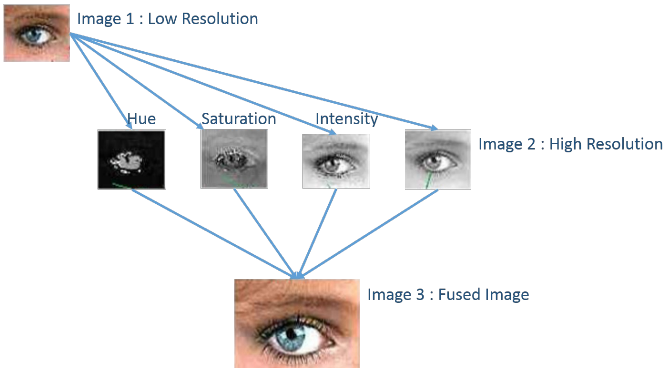

Figure 1: IHS based image fusion method

3) PCA based fusion: PCA is mostly used for data analysis to predict the model using principle axes theorem. Principle components can be obtained by finding the eigen values of correlation matrix of input data after normalizing it. Figure 2 shows the diagram of PCA based fusion method.

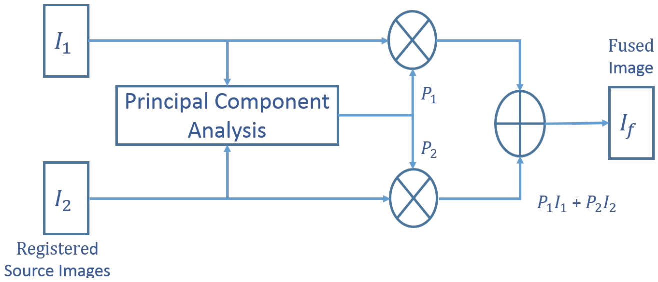

Figure 2: PCA based image fusion method

Finally, transformed variable values (principle components) are obtained from the corresponding input data. Then after input data is multiplied with principle components and fusion rule (max/average) is applied to get the high spatial quality image [9]. This method causes the distortion in spectral characteristics between the fused images and the original low resolution images. So it is highly criticized because the colors are not faithfully preserve in resulting image which found in the original images.

4) Arithmetic combinations based fusion: It is a simple and fast method. This method is good for Multisensory images. It provides superior resolution of MS images. It includes Brovey transform, Synthetic variable ratio technique and ratio enhancement technique.

Brovey Transform is developed by Gillespie *et al.* [8] in 1987. It performs RGB color transform so it is also known as color normalization transform method. Method includes simple combination of arithmetic operation such as addition, multiplication and division. Fusion is done by dividing the MS image by performing the addition of MS and PAN images. So it is calculated by the following equation

$$f(i,j) = \frac{MS_i}{MS_i + PAN} \tag{4}$$

Where, MS_i is the i^{th} band of multispectral image and PAN is the panchromatic image. It was implemented to increase the contrast visually in the high and low end of the histogram of particular images such as the images of water, shadow, urban area with high reflection, etc.

Synthetic variable ratio technique (SVR) is used to fuse multi-modal remote sensing images which cover large area [10]. It increases the fidelity of spectral information. The fusion is described as

$$f(i,j) = L \times \frac{PAN_{orig}}{PAN_{Hsyn}} \tag{5}$$

Where, PAN_{orig} is the original panchromatic image and L is obtained by transforming the original image into CIELab color space from RGB. PAN_{Hsyn} is calculated by

$$PAN_{Hsyn} = 0.480998 * B_1 + 0.790967 * B_2 + 0.966279 * B_3 + 0.655828 * B_4 \tag{6}$$

Where, B_1, B_2, B_3 and B_4 represent the gray values of R, G, B and IR bands of original MS image respectively.

However, this method is not much useful to overcome some disadvantages such as spectrum information interference from PAN images and spatial information interference from the MS images. Hence, the **ratio enhancement technique (RE)** [11] was introduced. The key benefit of RE method is to synthesize PAN images having low resolution which extracts the spatial information from the each band of MS image and spectral information from the PAN image. Due to synthesization, spectral distortion is removed in fusion process while PAN image preserves the spatial characteristic. The fusion is obtained by

$$f(i,j) = \frac{P_{hr} \times MS_{lr}^{i}}{P_{lr}^{i}} \tag{7}$$

P_{lr}^{i} is calculated by,

$$P_{lr}^{i} = P_{lf} + MS_{hf}^{i}. \tag{8}$$

Where, P_{hr} represents the high-resolution PAN image, MS_{lr}^{i} and P_{lr}^{i} are the i^{th} band of low-resolution MS image and PAN image respectively, P_{lf} is the low frequency component of the PAN image and MS_{hf}^{i} is the i^{th} band of high-frequency component of MS image.

5) High-pass filtering based fusion: To increase the potential of spatial resolution and obtain the qualitative features, it was introduced by Schowengerdt in 1980 for Landsat MSS dataset. Small convolution mask is applied on high-frequency spatial information to generate the image having high resolution. High spatial resolution image effectively reduces the low-frequency spectral information. Multispectral data is added to this filtered result and then it is divided by two to balance the values of brightness [12]. Fusion is done by performing following equation

$$f(i,j,k) = \frac{MS_{i,j,k} + HPF}{2} \tag{9}$$

Where, HPF is output of the filter and $MS_{i,j,k}$ represents i,j pixel of k^{th} band of MS image.

6) Total probability density based fusion: Motion of the object is the fundamental problem in computer vision application so the moving object detection and surveillance become difficult with numerical methods. Probability density based fusion is the statistical method which was introduced by Ingemar Cox in 1992 [13]. Figure 3 shows the diagram of probability density based fusion.

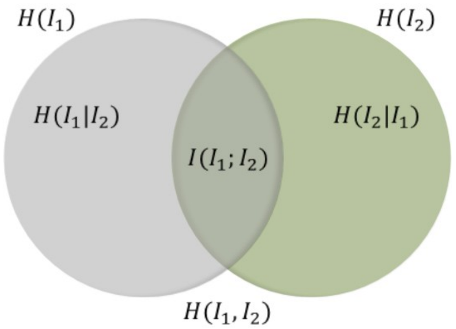

Figure 3: Probability based image fusion method

First it finds the conditional entropy $H(I_1 | I_2)$ or $H(I_2 | I_1)$ of individual objects and mutual information $I(I_1; I_2)$ is extracted of the multiple objects. Fusion is obtained by finding the joint probability $H(I_1, I_2)$ of the two objects/images. Method offers better performance for the real time application but it requires certain assumption to increase the capability. Hence, it becomes limitation of this method as all events are probabilistic. It is useful for geographical application only where some of the information is foreknown [14].

Spatial domain methods are very simple and easy to implement but the disadvantage is that spatial and spectral distortion produce in the fused image and it becomes negative factor for problem such as classification. Some undesirable effects are produced such as blurry edges and contract reduction.

B. *Transform domain fusion*

Transform domain fusion approach overcomes the limitation of the spatial domain methods. It reduces the complexity and artifacts effects. In this method images are generally shifted in frequency domain. Multi-scale transform theory is used to perform the fusion in transform domain. It is one of the most widely used approach for the image fusion. Basically three methods are introduced in this domain namely pyramid based fusion [15], wavelet based fusion [16] and curvelet based fusion [17], Image fusion methods based on multi-scale transform theory are listed below [18]:

- Multi scale transform based fusion
  1) Pyramid method
     - Gaussian pyramid
     - Laplacian Pyramid
     - Gradient pyramid
     - Morphological pyramid
     - Ratio of low pass pyramid
     - Contrast pyramid
     - Filter-subtract-decimate pyramid
  2) Wavelet transforms
     - Discrete wavelet transforms (DWT)
     - Stationary wavelet transforms
     - Multi-wavelet transforms
     - Dual tree discrete wavelet transforms
     - Lifting wavelet transform
  3) Curvelet transforms

1) Pyramid based image fusion: Quality of captured images degrade due to the bad weather situations such as fog, ash, dust, smoke, etc. Therefore, image processing algorithms are used to improve the quality of images. Image pyramid algorithms represent the models for visual system of the human in which original image is decomposed in different levels as shown in Figure 4 Decomposed images form the pyramid structure and each image is obtained by applying the approach of specific pattern selection at different level. Fused image is obtained by integrating all levels of images. Finally, inverse pyramid transform is applied to obtain the resultant image with good quality.

Figure 4 shows the general process of pyramid based fusion. It includes various methods such as Gaussian pyramid, Laplacian Pyramid, gradient pyramid, morphological pyramid, ratio of low pass pyramid, contrast pyramid and filter-subtract-decimate pyramid [15]. These pyramid methods are used for different applications according to the constraint of a user that one may require visually informative image and other one may need only color details to get the qualitative image.

2) Wavelet based image fusion: Wavelet based methods decompose the images with multi-resolution and represent non-redundant information in the decomposed image.

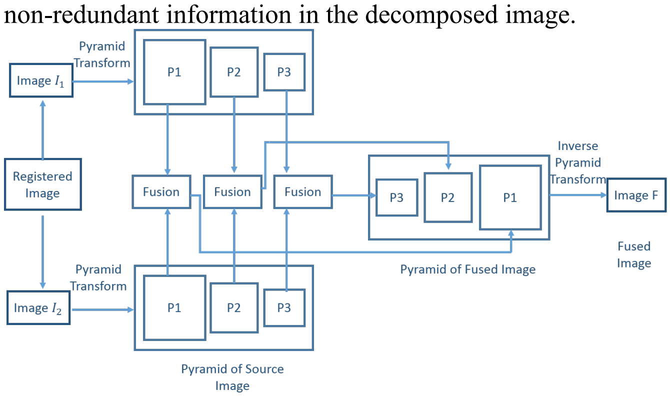

Figure 4: Pyramid based image fusion

It provides advantage over fourier transform because it gives better resolution in time domain as well as frequency domain. Whereas, Fourier transform brings up good resolution only in frequency domain. As shown in Figure 5 input images are decomposed into approximate and detailed components at specific level. Decomposed image includes approximate, horizontal detail, vertical detail and diagonal detail components. Each components of multiple images are integrated by applying fusion rule and finally resultant image is achieved by applying inverse wavelet transform.

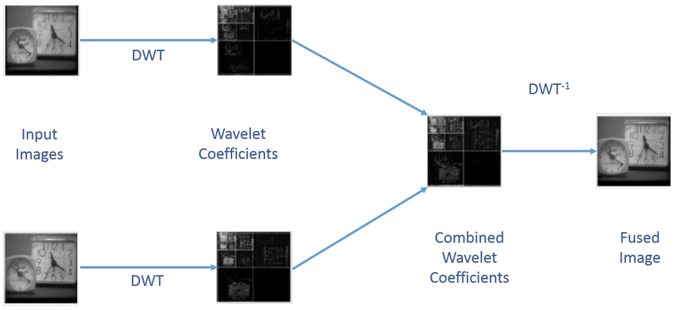

Figure 5: Wavelet based image fusion

Wavelet based mathod includes discrete wavelet transforms (DWT), stationary wavelet transforms, multi-wavelet transforms, dual tree discrete wavelet transforms and lifting wavelet transform. Shutao Li *et al.* in [16] investigated the decomposition level effects and performance on fusion using different wavelet methods. It performs better due to its shift-invariant property but the limitation is that it needs large memory and more time to perform shift invariant transform [18].

3) Curvelet based image fusion: Due to the limitation of wavelet transform to perform poor with the images having curve structure, the curvelet based image fusion was proposed by F. E. Ali *et al.* in [17]. Figure 6 shows the processing diagram of curvelet transform. First, the segmentation is done on the original image in three different bands using additive wavelet transform which approximates curved lines by straight lines. overlapping tiling is applied on sub-bands of Δ_1 and Δ_2. Here the aim of tiles is to avoid the edge effects. Then ridgelet transform is applied in radon domain which aims to detection of shape of curved objects. Curvelet transform is developed as a tool to utilize in graphical applications for representing the curve shapes. Fusion rule is applied to integrate the ridgelet transform of sub-bands Δ_1 and Δ_2 of multiple images. Finally, inverse curvelet transform is performed to obtain the resultant image.

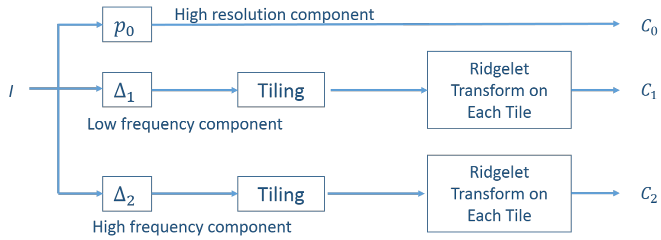

Figure 6: Processing diagram of curvelet transform

The fusion based on multi-scale transform techniques often performs inadequately when images obtained from different sensor modalities. Hence, the next section describes the recently proposed techniques which are based on biologically inspired algorithms i.e. deep learning.

C. Deep learning based image fusion

Traditional methods mostly include two crucial factors such as feature extraction and fusion rule in which algorithms are designed in manual manner and it is very difficult to process for large dataset. With tremendous growth of machine learning and deep learning, recently researchers have developed the image fusion algorithms by utilizing the concept of deep learning. It is demonstrated that DL is very useful to extract effective features and classification problems and many researchers have shown that DL algorithms performs better than any traditional methods. Using various imaging devices, images are categorized in four category such as multi-modal, multi-resolution, multi-focus and multi-temporal images. Here the main goal is to process these multiple category of images and obtain the productive image which may be more visual to human and more informative for machine. Various image fusion algorithms on multi category of images are discussed in next subsection.

1) Multi-modal image fusion: Multi-modality includes the images with different features such as images from moving object, various part of human body, geographic area, etc. It is difficult task to obtain useful features from the multi-modal images in single image. Various DL algorithms have been introduced to extract complicate features from the particular image and fuse it to further processing such as classification, object detection and recognition, etc.

In [19] Ke *et al.* proposed the algorithm using CNN, RBM and DBN architecture as shown in Figure 7. They have extracted features for the shallow and deep modality at low, middle and high level from the geographic images.

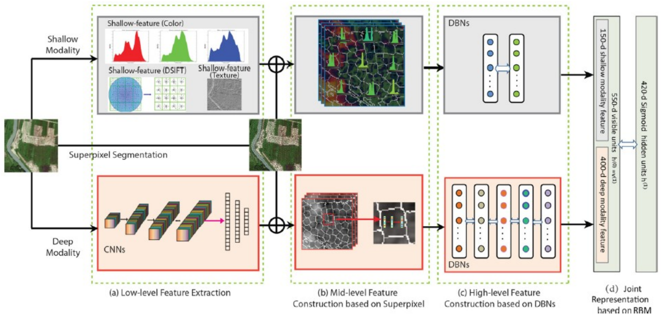

Figure 7: Multi-modal image fusion using deep learning [19] algorithms

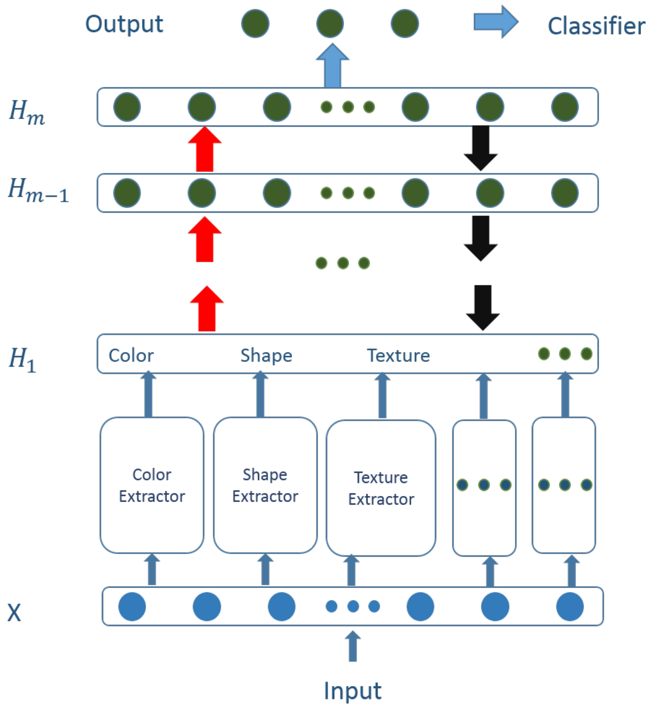

Figure 8: Multi-modal image fusion using deep autoencoder [20]

Shallow features include shift invariant feature detection (SIFT), color and local binary pattern (LBP) whereas deep features are extracted by using CNN. It is further processed to extract mid-level feature by averaging lower level feature maps. Finally high level features are extracted using DBN. Extracted features from shallow and deep modality are fused and processed through RBM to solve classification problem. Morover, Gang Ma *et al.* in [20] proposed the image fusion algorithm using DAE architecture. As shown in Figure 8 multiple features such as color, shape, texture, etc. are extracted using particular feature extraction tool. Then after it is processed through DAE to extract deep features and softmax classifier is used for final classification problem.

Haidong Shao *et al.* in [21] developed the deep feature fusion algorithm for fault diagnosis from rotating machine. As it is known that moving object almost contains noisy data, Hence, DAE along with contractive autoencoder (CAE) is used to extract deep features from the moving machine. DAE is specially used to reduce the background noise from the image and CAE increases the robustness of feature learning process [22]. In that, a new concept called locality preservation projection (LPP) is introduced to fuse the extracted features [23]. Extracted features from the i^{th} and j^{th} images are fused by following equation

$$f_i = A^T y_i \tag{10}$$

Where, y_i represents the extracted features from the i^{th} image. A^T is the transformation matrix. Main goal is to find the A such that high dimensional data y can be project onto the low dimensional space of f. It can be determine by minimizing the following objective function

$$\min \sum_{i,j=1}^{n} (f_i - f_j)^2 W_{ij} \tag{11}$$

With

$$W_{ij} = \begin{cases} \exp(\| y_i - y_j \|^2), & y_j \in N_k(y_i) \text{ or } y_i \in N_k(y_j) \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Where, W_{ij} is a weight matrix and $y_j \in N_k(y_i)$ represents that y_j is the k nearest neighbor of y_i. Optimization problem can be solve by finding the eigen values λ which represent the eigen vector as transformation matrix. Eigen values can be determine by generalize the objective function as following of f. It can be determine by minimizing the following objective function

$$\frac{1}{2}\min \sum_{i,j=1}^{n} (f_i - f_j)^2 W_{ij} = \frac{1}{2}\min \sum_{i,j=1}^{n} (A^T y_i - A^T y_j)^2 W_{ij}$$
$$= A^T Y(\phi - W)Y^T A = A^T YLY^T A \tag{13}$$

Where, $\phi_{ii} = \sum_{j=1}^{n} W_{ij}$ represents the diagonal matrix and $L = \phi - W$ is a Laplacian matrix. Finally, A is obtained by solving the generalized eigenvalue problem,

$$YLY^T A = \lambda YDY^T A. \tag{14}$$

Where, λ represent the eigenvalue and A is corresponding eigen vector. So, the final low dimensional fused vector can be obtained by

$$F = A^T Y. \tag{15}$$

Fused vector F contains local meaningful information of the original deep features.

2) *Multi-temporal image fusion:* Multi-temporal images are most widely used for detecting the change in many fields such as change of land surface, area growth, dynamics in forest and vegetation, disaster monitoring and many changes can be detected at local, global and regional scale [24,25]. These images are obtained through remote sensor devices. Unchangeable natural climates that make lots of problems and damages to the life of people, environment and economy. Hence, after suffering from natural disaster, it is necessity to get expeditious, reliable and precise damage information for counseling the activities. However, the quality obtained from the remote sensors is not better as it should be necessary to recognize the changes. Therefore, the image fusion methods have been proposed to improve the quality of images which are obtained at different time. George P. Petropoulos *et al.* [26] in 2013 introduced machine learning approach to detect the change of surface mining activity and reclamation based on a multi-temporal Landsat TM imagery. M. Vakalopoulou *et al.* in 2015 [27] proposed the image fusion algorithm with deep learning features to detect the building from very high resolution multispectral data. They have used seven layer CNN using ImageNet architecture. Gonzalo Farias *et al.* [28] proposed the feature extraction algorithm using deep learning. In which two layers of DAE is used to extract the features from the temporal images and SVM is used as a classification problem.

3) *Multi-resolution image fusion:* Remote sensing devices provide the information with multi-resolution which include PAN, multi-spectral and hyperspectral images. Usually it should be contained all the spatial information from PAN images and all the spectral information from multi and hyperspectral images in single fused image. Various methods have been proposed recently in which Frosti Palsson *et al.* in [29] introduced CNN algorithm to fuse multi-spectral and hyperspectral image. First, they used PCA to reduce the dimensionality of the original image and then three layer CNN is applied. Hence, it reduces the complexity of the CNNs architecture and it makes the system more robust with less computational time. In [30], Yushi chen *et al.* proposed the deep fusion of multi-spectral, hyperspectral and LiDAR(light detection and ranging) images using two layer CNN only. LiDAR images contain elevation details of the surface corresponding to sensor arrangement. Souleyman chaib *et al.* in [31] have used the VGGNet architecture of CNN with 19 layers to extract features from the very high resolution images. Then they applied Discriminant component analysis (DCA) to extract only meaningful features and finally SVM is used as a classifier to classify the object.

4) *Multi-focus image fusion:* It is dominant field of image fusion technique. It is difficult to obtain all the objects with clarity in single image. Multi-focus image fusion method addresses multi-object in a image with productive information by fusing the different images of same scene taken with different focus parameter setting. Its aim to direct mapping between the captured focus image and source image. Various spatial and transform based fusion methods have been developed to get the qualitative image but as it is known that manually designed algorithms for feature extraction and fusion rule do not consider all important factors in account. So, the concept of deep learning is introduced to get all-in-focus image.

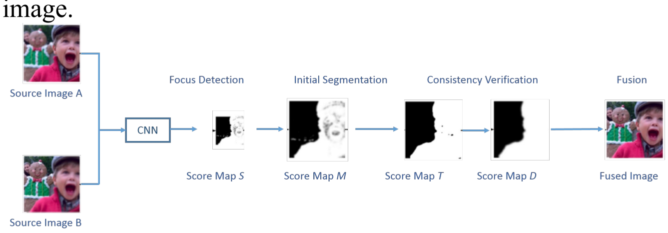

Figure 9: Multi-focus image fusion using DCNN [32]

Recently Yu Liu *et al.* in [32] have proposed the DCNN to fuse the multi-focus images. They have used three layer CNN and after then focus detection, initial segmentation and consistency verification is performed. Finally, image fusion is done by applying weighted average rule. Chaoben Du and Shesheng Gao have also introduced fusion through CNN using image segmentation in [33]. Processing algorithm is same in both papers but they have just changed stride and padding size of the convolutional layer. Table II summarizes the advantages and disadvantages of various image fusion methods. Table I shows the comparison in terms of accuracy of various deep learning algorithms. Here, it is observed that deep learning algorithm provides better accuracy and reduces the computational cost when dimensionality reduction algorithm is applied first and then it is processed through deep learning algorithms.

Table I: Comparison of various deep learning methods with accuracy

Database	Algorithms	Accuracy(%)
Multi-focus images	CNN+IS+FR [32]	98.65
	CNN+IS+FR [33]	97.00
Multi-resolution images	PCA+CNN [29]	98-99
	CNN+DCA+SVM [30]	97.42
	CNN+BN [31]	98.61
Multi-temporal images	DAE+SVM [26]	98.60
	SVM+RBF [27]	90.00
	CNN+SVM [28]	92.00
Multi-modal images	FE+ DAE+SC [19]	80.00
	FE+CNN+DBM+RBM [20]	82.08
	DAE+CAE+LPP [21]	95.00

Table II: Analysis with advantages and disadvantages of different methods [8, 14, 34]

	Methods	Advantages	Disadvantages
1	Simple Average	Simplest, works well for single	Reduced contrast

		sensor images containing additive noise, good for high contrast input images	
2	Simple Max/Min	Simple, highly focused output image obtained as compared to average method	Blurring, it has higher pixel intensity but it does not mean more content
3	Arithmetic combination	Simple and fast method, good for Multisensory images, provides superior visual and high resolution MS image	This method ignores the requirement of high quality synthesis of spectral information and causes spectral Distortion
4	IHS	It is a simple method to merge the images attributes, provides high spatial quality	The major limitation that only three bands are involved. It suffers from artifacts and noise which tends to higher contrast, causes color distortion.
5	PCA	Simple, fused images have high spatial quality, prevents certain features from dominating the image because of their large digital numbers	This method is highly criticized because of the distortion of the spectral characteristic between the fused images and the original low resolution images. Resulting image does not preserve faithfully the colors found in the original images
6	Probability density based	Offers real time better capabilities of tracking under certain conditions	All of the events as probabilities, it requires prior information or assumption for false measurement
7	Transform based	It has least spectral distortion, also provides better SNR than pixel based approach	In this method final fused image have a less spatial resolution
8	Deep learning based	Works as a Feature extractor as well as Classifier, reduce the complexity of the multiple algorithms, higher accuracy	It has more complex architecture, require massive computing power

III. Conclusion

In this paper, various image fusion methods have been discussed which provide the composite image known as fused image, provides the complementary information. Fused image may more suitable for human as well as machine perception for further processing. It includes spatial domain, transform domain and deep learning based image fusion methods. Architectures of deep learning are essential part of the artificial intelligence and it is proven to be more powerful in terms of speed and memory requirement compared to other feature extractor and classifier algorithms. Moreover, CNN performs as a deep feature extractor as well as a classifier so there is no any requirement to implement individual algorithm for different task of fusion. Therefore, it reduces the complexity of algorithms. Image fusion using deep learning algorithms provide more efficiency and accuracy in terms of computational processing speed and quality of image respectively.

IV. Future Scope

As deep learning has become the recently growing area of the research, there are lots of scope for the research. Deep learning algorithms provide superior quality on the fused image. Fusion image provides more meaningful information which is useful for further processing such as classification, object detection and recognition problems, etc. The recent deep learning algorithms have more complex architecture, they require large data set and massive computing power for training whereas manually collected labeled dataset requires huge amount of human efforts. Thus, first it is desired to reduce the dimensionality of input images and it processes through deep learning algorithms. At testing time, these deep models are high memory demanding and time-consuming, which makes them not suitable to be deployed on mobile platforms that have limited resources. It is important to investigate how to reduce the complexity and obtain fast-to-execute models without loss of accuracy.

References

[1] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," information Fusion, vol. 33, pp. 100–112, 2017.

[2] X. Zheng, H. Chen, and T. Xu, "Deep learning for chinese word segmentation and pos tagging." pp. 647–657, 2013.

[3] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," IEEE Journal of Selected topics in applied earth observations and remote sensing, vol. 7, no. 6, pp. 2094–2107, 2014.

[4] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," pp. 184–199, 2014.

[5] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," Information Fusion, vol. 42, pp. 158–173, 2017.

[6] S. Li, B. Yang, and J. Hu, "Performance comparison of different multiresolution transforms for image fusion," Information Fusion, vol. 12, no. 2, pp. 74–84, 2011.

[7] A. Kaur, R. Kaur, M. Mahajan, and S. Singh, "A comparative study of multi scale transform based fusion techniques," International journal of computer science and network, vol. 5, no. 3, pp. 158–173, 2016.

[8] C. Morris and R. S. Rajesh, "Survey of spatial domain image fusion techniques," International Journal of Advanced Research in Computer Science Engineering and Information Technology, vol. 2, no. 3, pp. 249–254, April 2014.

[9] R. Princess, M, Kumar, Suresh, and R. B. M, "Comprehensive and comparative study of different image fusion techniques," vol. 3297, pp. 2320–3765, 10 2007.

[10] S. Chen, J. Luo, Z. Shen, G. Luo, and C. Zhu, "Using synthetic variable ratio method to fuse multi-source remotely sensed images based on sensor spectral response," vol. 2, pp. 1084–1087, July 2008.

[11] X. Li, Q. Xu, F. Gao, and L. Hu, "Pansharpening based on an improved ratio enhancement," pp. 1100–1103, July 2015.

[12] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," IEEE transactions on geoscience and remote sensing, vol. 43, no. 6, pp. 1391–1402, 2005.

[13] F. Castanedo, "A review of data fusion techniques," The Scientific World Journal, vol. 2013, 2013.

[14] I. J. Cox, "A review of statistical data association techniques for motion correspondence," International Journal of Computer Vision, vol. 10, no. 1, pp. 53–66, 1993.

[15] Sahu, Chhamman, Sahur, and R. Kumar, "Pyramid based image fusion," International Journal of Engg. Computer Science, vol. 3, no. 8, pp. 7890–7894, 2014.

[16] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniquesan introduction, review and comparison," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 62, no. 4, pp. 249–263, 2007.

[17] F. E. Ali, I. M. El-Dokany, A. A. Saad, and F. E. A. El-Samie, "Fusion of mr and ct images using the curvelet transform," pp. 1–8, March 2008.

[18] N. Paramanandham, K. Rajendiran, D. Narayanan, I. V. S., and M. Anand, "An efficient multi transform based fusion for multi focus images," pp. 0984–0988, April 2015.

[19] K. Li, C. Zou, S. Bu, Y. Liang, J. Zhang, and M. Gong, "Multi-modal feature fusion for geographic image annotation," Pattern Recognition, vol. 73, pp. 1–14, 2017.

[20] G. Ma, X. Yang, B. Zhang, and Z. Shi, "Multi-feature fusion deep networks," Neurocomputing, vol. 218, pp. 164–171, 2016.

[21] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," Knowledge-Based Systems, vol. 119, pp. 200–220, 2017.

[22] F.-q. Chen, Y. Wu, G.-d. Zhao, J.-m. Zhang, M. Zhu, and J. Bai, "Contractive de-noising auto-encoder," pp. 776–781, 2014.

[23] [Online]. Available: http://papers.nips.cc/paper/2359-localitypreserving-projections.pdf

[24] P. Du, S. Liu, J. Xia, and Y. Zhao, "Information fusion techniques for change detection from multi-temporal remote sensing images," Information Fusion, vol. 14, no. 1, pp. 19–27, 2013.

[25] Jain and Sanjay, "Multi temporal image fusion of earthquake satellite images," vol. 3, 01 2012.

[26] G. P. Petropoulos, P. Partsinevelos, and Z. Mitraka, "Change detection of surface mining activity and reclamation based on a machine learning approach of multi-temporal landsat tm imagery," Geocarto International, vol. 28, no. 4, pp. 323–342, 2013.

[27] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1873–1876, July 2015.

[28] G. Farias, S. Dormido-Canto, J. Vega, G. Ratt´a, H. Vargas, G. Hermosilla, L. Alfaro, and A. Valencia, "Automatic feature extraction in large fusion databases by using deep learning approach," Fusion Engineering and Design, vol. 112, pp. 979–983, 2016.

[29] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 639–643, May 2017.

[30] Y. Chen, C. Li, P. Ghamisi, X. Jia, and Y. Gu, "Deep fusion of remote sensing data for accurate classification," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 8, pp. 1253–1257, Aug 2017.

[31] S. Chaib, H. Liu, Y. Gu, and H. Yao, "Deep feature fusion for vhr remote sensing scene classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 8, pp. 4775–4784, Aug 2017.

[32] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," Information Fusion, vol. 36, pp. 191–207, 2017.

[33] C. Du and S. Gao, "Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network," IEEE Access, vol. 5, pp. 15 750–15 761, 2017.

[34] C. Gupta and P. Gupta, "Article: A study and evaluation of transform domain based image fusion techniques for visual sensor networks," International Journal of Computer Applications, vol. 116, no. 8, pp. 26–30, April 2015.