IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 6, December, 2023, pp. 2731-2740
Journal homepage: http://iieta.org/journals/ts

Exploring the Application of Deep Learning in Multi-View Image Fusion in Complex Environments

Xiujuan Luo, Lili Shao*

School of Computer, Heze University, Heze 274015, China

Corresponding Author Email: Shaolili@hezeu.edu.cn

https://doi.org/10.18280/ts.400633

Received: 12 July 2023
Revised: 26 October 2023
Accepted: 5 November 2023
Available online: 30 December 2023

Keywords:
deep learning, multi-view image fusion, complex environments, moment of inertia axis method, morphological decomposition, attention feature integration

ABSTRACT

The advancement of technology has unveiled the immense potential of deep learning across various domains, notably in multi-view image fusion within complex environments. Multi-view image fusion aims to merge images from different perspectives to garner more comprehensive and detailed information. Despite this, challenges persist in such fusion under complex conditions, particularly when confronting significant variations in perspective and intricate lighting scenarios. Predominant deep learning approaches, reliant on extensive annotated data, grapple with high computational complexity when processing large-scale and high-dimensional image data, thus hindering real-time applicability. This exploration primarily focuses on two facets: multi-view image registration based on the moment of inertia axis method, and multi-view image fusion utilizing morphological decomposition and attention feature integration. The objective is to enhance the efficiency and effectiveness of multi-view image fusion in complex settings, propelling the practical advancement of deep learning technologies.

1. INTRODUCTION

With the advancement of technology, deep learning has demonstrated significant potential and advantages across various fields, especially in multi-view image fusion within complex environments, where its importance is self-evident [1, 2]. Multi-view image fusion, which involves the combination of images from different perspectives to obtain more comprehensive and detailed information, is a highly challenging task [3-6]. In complex environments, the considerable visual disparities between images from different perspectives render the fusion process difficult. In such contexts, the flexibility and powerful pattern recognition capabilities of deep learning become particularly crucial [7-11].

The study of multi-view image fusion not only contributes to advancing image processing technology but also holds significant theoretical value for enhancing the visual understanding capabilities of artificial intelligence systems [12-15]. For instance, in fields such as autonomous driving, drone reconnaissance, and 3D modeling, multi-view image fusion can provide more detailed and comprehensive visual information, thereby improving system performance and accuracy. Therefore, exploring the application of deep learning in multi-view image fusion within complex environments is of great value for advancing research and application in related fields [16, 17].

However, despite significant achievements of deep learning in many domains, including image processing, several issues and challenges remain in multi-view image fusion within complex environments. For example, existing deep learning methods often rely on large amounts of annotated data, which are challenging to obtain in practical applications [18-21]. Additionally, these methods struggle with high computational complexity when processing large-scale and high-dimensional image data, making it difficult to meet real-time requirements. Finally, existing fusion methods often fail to achieve satisfactory results when dealing with significant changes in perspective and complex lighting conditions [22-24].

This exploration focuses on two research areas: firstly, multi-view image registration based on the moment of inertia axis method, and secondly, multi-view image fusion using morphological decomposition and attention feature integration. The moment of inertia axis method, as a novel approach for image registration, effectively addresses the issue of multi-view image registration; meanwhile, the method based on morphological decomposition and attention feature integration demonstrates superior performance in extracting and integrating image features. Through the combination of these two methods, it is expected that a new approach will be proposed for effective multi-view image fusion in complex environments. This research not only aims to fill existing gaps in the field but also holds significant value in enhancing the efficiency and effectiveness of multi-view image fusion, as well as advancing the practical application of deep learning technologies.

2. MULTI-VIEW IMAGE REGISTRATION BASED ON MOMENTS AXIS

In the context of multi-view image fusion within complex

environments, significant visual disparities such as scale, rotation, and tilt may exist between images from different perspectives. These disparities can impact the effectiveness of image fusion. The method of moments axis is capable of computing the primary geometric characteristics of an image. By aligning images based on these geometric characteristics, visual disparities are either eliminated or reduced.

To achieve multi-view image registration in complex environments, geometric moments are utilized to capture the geometric properties of images. Specifically, geometric moments of different orders are computed based on the position and value of pixels. Assuming $d(z,t)$ represents the multi-view images to be registered, o and w are the order of geometric moments to be solved, wherein L_{ow} denotes the $(o+w)$-th order geometric moment, the following formula defines the geometric moment of multi-view images in complex environments:

$$L_{ow} = \int\int \left(z \hat{} o\right) * \left(t \hat{} w\right) d\left(z,t\right) dzdt \left(o, w = 0,1...\infty\right) \tag{1}$$

In this study, geometric moments are employed to describe and compare the shape characteristics of multi-view images, thereby enabling more accurate image registration and fusion. Each order of geometric moments represents a specific image characteristic. The zeroth and first-order moments are moment features of the image, assisting in describing the basic geometric attributes of the image. The zeroth-order moment, also known as the mass moment, can be interpreted as the area of the image or the sum of the pixels. In multi-view image fusion within complex environments, the zeroth-order moment is used to represent the overall scale of the image, i.e., the number of pixels contained in the image. The formula for calculating the zeroth-order moment L_{00} of the image $d(z,t)$ is given by:

$$L_{00} = \int\int d\left(z,t\right) dzdt \tag{2}$$

The first-order moment is used to calculate the centroid of the image, which is the average center of mass position. In multi-view images within complex environments, the first-order moment helps in locating the center position of the image, crucial for image alignment and registration. Assuming the first-order moment is represented by (L_{01}, L_{10}), the centroid of image $d(z, t)$ is $(\bar{z},\bar{t})$, and there is:

$$\bar{z} = \frac{L_{10}}{L_{00}}, \bar{t} = \frac{L_{01}}{L_{00}} \tag{3}$$

In multi-view image fusion scenarios within complex environments, central moments are defined as moments relative to the image centroid. They are calculated by subtracting the coordinate values of each pixel from the corresponding centroid coordinates in each dimension, then multiplying by the pixel values and summing. Central moments reflect the geometric characteristics of an image, such as shape, size, and orientation. By calculating and comparing the central moments of images from different perspectives, this study aims to understand and describe the shape characteristics of images, thereby achieving more accurate image registration and fusion. Such method based on central moments can effectively process multi-view images in complex environments, enhancing the effectiveness and efficiency of image fusion. The central moment I_{ow} of the original multi-view images can be obtained by shifting the coordinate origin to the image centroid as follows:

$$I_{ow} = \int\int \left[\left(z-\bar{z}\right) \hat{} o\right] * \left[\left(t-\bar{t}\right) \hat{} w\right] d\left(z,t\right) dzdt \tag{4}$$

The fundamental principle of multi-view image registration based on the method of moments axis involves using the geometric properties of images, including the centroid and orientation, to align them. Firstly, each image's first-order moment is calculated to find its centroid, the center of gravity position. The centroid, an average of all pixel positions, reflects the approximate location of the image. Further, the image's second-order moments are calculated to determine the principal axis direction, indicating the shape orientation. The angle between the image and the coordinate axes, calculated using second-order moments, reflects the image's rotation. Finally, by translating the image to the centroid position and rotating it according to the calculated angle, image registration can be achieved. This ensures that images from different perspectives align in spatial position and orientation, facilitating subsequent image fusion operations.

Considering the practical scenarios of multi-view image fusion in complex environments, this study employs the method of moments axis based on image grayscale information for multi-view image registration. Assuming the image to be registered is denoted as $d(l, b)$, with the desired order of geometric moments represented by o and w, its $(o+w)$-th order geometric moment, l_{ow}, is defined as follows to align with the discrete characteristics of image grayscale values:

$$l_{ow} = \sum_{l=1}^{L}\sum_{b=1}^{B} l^{o} b^{w} d\left(l,b\right) \tag{5}$$

The centroid coordinates $(\bar{z},\bar{t})$ of the image $d(l,b)$ that conforms to the discrete characteristics of image grayscale values can be obtained through the following equations:

$$\bar{z} = \sum_{l=1}^{L}\sum_{b=1}^{B} l d\left(l,b\right) / \sum_{l=1}^{L}\sum_{b=1}^{B} d\left(l,b\right) \tag{6}$$

$$\bar{t} = \sum_{l=1}^{L}\sum_{b=1}^{B} n d\left(l,bn\right) / \sum_{l=1}^{L}\sum_{b=1}^{B} d\left(l,b\right) \tag{7}$$

The central moment ω_{ow} of the image $d(l, b)$, conforming to the discrete characteristics of image grayscale values, is calculated as:

$$\omega_{ow} = \sum_{l=1}^{L}\sum_{b=1}^{B} \left(l-\bar{z}\right)^{o} \left(b-\bar{t}\right)^{w} d\left(l,b\right) \tag{8}$$

Central moments, calculated relative to the image centroid, are obtained through second-order moments. These moments are derived by subtracting each pixel's coordinate value from the corresponding dimension's centroid coordinates, multiplying by pixel values, and summing. The axis, representing the primary direction of the image shape, is usually obtained from the image's second-order moments. In two-dimensional images, the direction of the principal axis can be determined by calculating the image's second-order central moments. Specifically, the direction of the principal axis is the

eigenvector of the covariance matrix formed by the second-order central moments, corresponding to the direction of the eigenvector associated with the largest eigenvalue. Similarly, the angle ϕ between the image and the coordinate system can also be calculated using second-order moments, as per the following formula:

$$\tan 2\phi = \frac{2\omega_{11}}{\omega_{20} - \omega_{02}} \tag{9}$$

The centroid serves as a reference point for image translation. Further, by comparing the centroids of images from different perspectives, their horizontal and vertical displacement amounts are calculated. The rotation angle of the image relative to the coordinate system can be determined through second-order moments. This rotation angle reflects the image's rotational state, aiding in determining the principal axis direction. Once the displacement amounts and rotation angle are determined, translation and rotation operations can be performed to adjust the image's position and orientation. This aligns the multi-view images in spatial position and orientation, achieving image registration. Assuming the centroid and the angle between the image and the coordinate system for the image to be registered are represented by z^-_o, z^-_o, ϕ^-_o, and for the reference image by z^-_E, z^-_E, ϕ^-_E, the horizontal and vertical displacement amounts and the rotation angle relative to the coordinate system for the image to be registered are denoted by Δz, Δt, $\Delta\phi$, respectively, as per the following calculation formulas:

$$\Delta z = \bar{z}_O - \bar{z}_E \tag{10}$$

$$\Delta t = \bar{t}_O - \bar{t}_E \tag{11}$$

$$\Delta\phi = \bar{\phi}_O - \bar{\phi}_E \tag{12}$$

3. MULTI-VIEW IMAGE FUSION BASED ON MORPHOLOGICAL DECOMPOSITION AND ATTENTION FEATURE INTEGRATION

In complex environments, multi-view image fusion encounters several challenges, such as significant changes in perspective and complex lighting conditions. Addressing these issues, this study proposes a method of multi-view image fusion based on morphological decomposition and attention feature integration. For multi-view image fusion in complex environments, morphological decomposition extracts shared features across different perspectives, as well as unique features of each perspective, providing richer and more detailed information for fusion. The attention feature integration method, an effective deep learning technique, automatically learns and extracts the most important features in images. By incorporating the attention mechanism, the model focuses on the most critical parts of the image during processing, thereby enhancing the accuracy of fusion. In the context of complex lighting conditions, attention feature integration automatically adjusts the model's focus, enabling it to identify the most vital features under varying lighting conditions, thus improving the effectiveness of fusion.

The constructed multi-view image fusion network structure comprises three modules: the MCA module, the feature extraction module, and the feature fusion module. The MCA module performs morphological component decomposition on multi-view images, subdividing them into structural and textural components. Structural components generally correspond to the main objects and contours in the image, while textural components relate to details and noise. This decomposition allows for better localization and extraction of key image features. Furthermore, the MCA module preliminarily fuses the structural and textural components of both types of images, effectively utilizing information from both perspectives. This fusion not only retains spatial and spectral information but also removes some redundancy and noise, crucial for enhancing the quality and efficiency of image fusion. The MCA module, by fully leveraging the spectral information and spatial detail present in both types of images, better facilitates image registration. This implies that the constructed model can understand and interpret multi-view images in a broader context, thus improving the accuracy and stability of image fusion. The MCA module primarily consists of a dual-branch feature extraction network, composed of two cascading convolutional subnetworks. The dual-branch structure allows the MCA module to simultaneously process image data captured from two different perspectives, with each convolutional subnetwork handling images from one perspective, enabling the extraction and learning of richer and deeper feature information. Figure 1 presents the framework diagram of the MCA module.

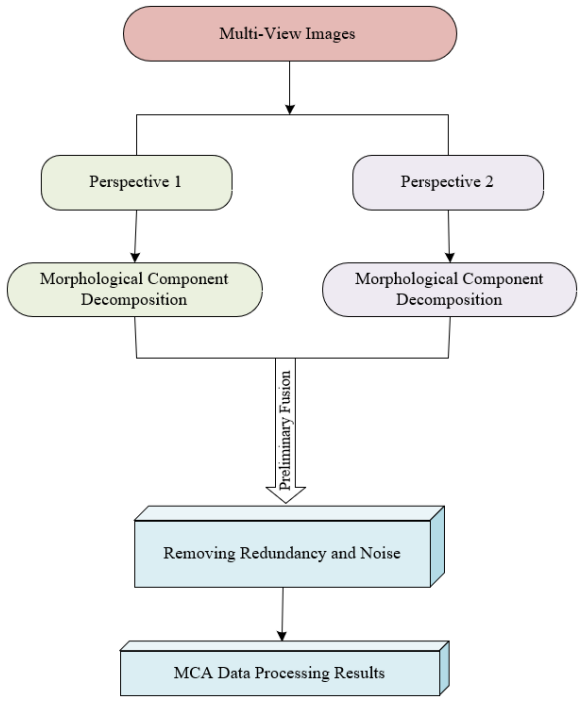

Figure 1. Framework of the MCA module

In the constructed multi-view image fusion network, the feature extraction network module utilizes a dual-branch cascading convolutional layer structure to extract spectral and spatial features processed by the MCA module. Spectral features reflect the spectral information of the image, such as color and brightness, while spatial features represent the spatial information, like shape and location. These features are crucial for understanding the content of the image. The design of the feature extraction network module, with its cascading convolutional layers, allows for capturing a broader range of contextual information while retaining local features. The dual-branch design of the feature extraction network module

enables simultaneous processing of images from two perspectives, providing strong support for subsequent image fusion.

Structurally, the input to the feature extraction network module consists of the textural and structural components of the source images obtained through the MCA module. These two components contain key information of the image, which is essential for feature extraction. The feature extraction network module is composed of two subnetworks, each containing three consecutive convolutional modules. Each convolutional module consists of cascading 3×3 and 1×1 convolutional layers. The 3×3 convolutional layer extracts local spatial features, while the 1×1 convolutional layer captures global features. This cascading design enables the network to extract features at different levels, thus obtaining richer information. Each convolutional layer is followed by an LReLU activation function. The LReLU activation function adds non-linearity to the network, allowing it to learn and represent more complex features.

Assuming the outputs of the j-th convolutional layer in the two subnetworks of the feature extraction network are represented by $Y^j(j=1,2,3)$ and $V^j(j=1,2,3)$, with the channel index denoted by vg, and the convolution kernels of the j-th layer convolution in the two subnetworks are represented by $q^j_{Y(vg)}$ and $q^j_{V(vg)}$, the expressions for Y^j and V^j are as follows:

$$Y^j(u,k,vg) = M\ \mathrm{ReLU}\left(Y^{j-1}(u,k,vg) * q^j_{Y(vg)}\right) \tag{13}$$

$$V^j(u,k,vg) = M\ \mathrm{ReLU}\left(V^{j-1}(u,k,vg) * q^j_{V(vg)}\right) \tag{14}$$

Figure 2 gives a diagram showing the feature extraction and fusion network framework. The feature fusion network module is primarily utilized to generate multi-spectral images with high spatial resolution. Multi-spectral images, containing abundant spectral information, provide richer details than single spectral channels, thereby enhancing the quality of image fusion. Within the feature fusion network module, two types of attention mechanism blocks are designed: Dual Cascading Attention Mechanism (DCAM) and Effective Channel Attention (ECA). The DCAM focuses the model on important spatial information, while ECA guides the model towards significant spectral information. These attention mechanism blocks enable the model to better focus on and utilize critical information, generating high-quality fused images.

Structurally, the input to the feature fusion module consists of feature maps outputted by the feature extraction network, including textural and structural components. These feature maps first undergo weighted averaging before entering the initial 1×1 convolutional layer. The 1×1 convolutional layer effectively alters the depth of the feature maps without changing their spatial dimensions, providing richer information for subsequent operations. The output of the convolutional layer is concatenated with the first convolutional module of the dual-branch feature extraction network. This concatenation operation merges features along the depth dimension, allowing the network to consider information from different modules simultaneously. The concatenated feature map is then input into the second 1×1 convolutional layer, followed by the DCAM module. This mechanism enhances focus on important information within the feature map, improving the model's efficiency in utilizing key information. The output from the DCAM is concatenated with the second convolutional module of the dual-branch feature extraction network, followed by the third 1×1 convolutional layer. The concatenated feature map enters the ECA module, which captures inter-channel interaction information, aiding in extracting richer features. The output of ECA is concatenated for the last time with the third convolutional block of the feature extraction network, then output to the final 1×1 convolutional layer, resulting in the fused image.

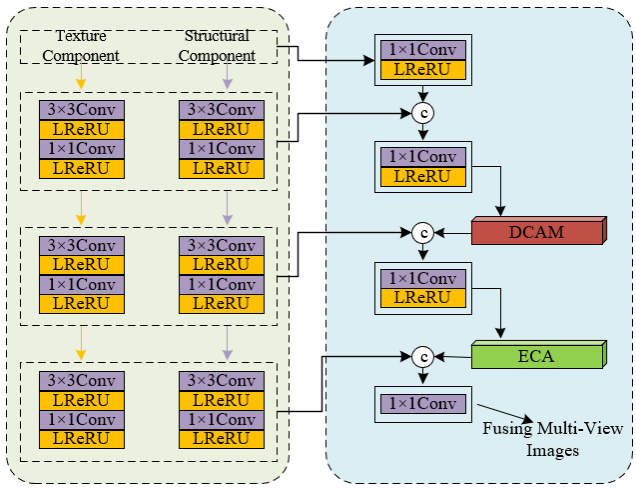

Figure 2. Feature extraction and fusion network framework diagram

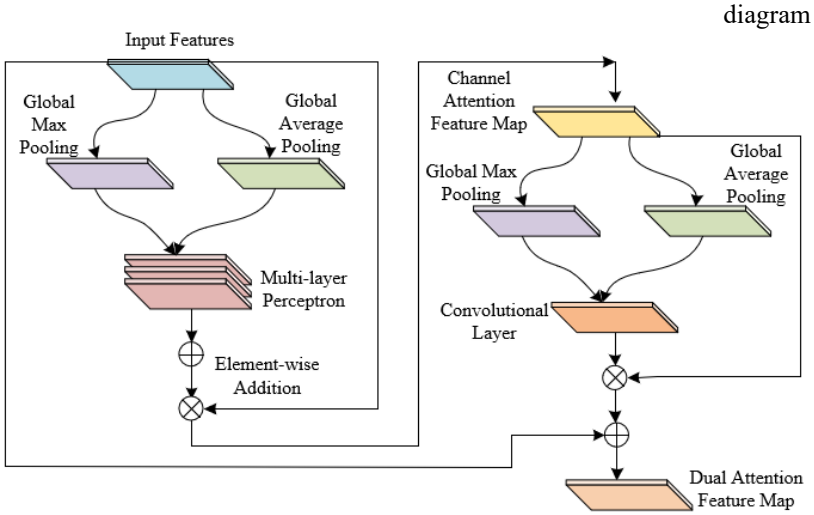

Figure 3. Principle of the DCAM

To focus on features significantly impacting the quality of the fused image, this paper utilizes the L2 norm as the loss function. This approach allows the model to prioritize important features during training, avoiding distraction by less significant ones. Additionally, the L2 norm offers faster convergence rates, crucial for training large-scale image fusion models. Let the number of training samples be represented by b, images from different perspectives by TX and LA, original images in the training set by U, and the fusion function outputted by the model by $D(TX,LA,\phi)$. The L2 loss is represented as:

$$\ell_1 = \frac{1}{b}\sum_{u=1}^{b} |D(TX, LA, \phi) - U|_1 \tag{15}$$

In multi-view image fusion, the information contained in images from different perspectives can vary significantly. Some information is vital for generating the fused image, while other information may be less important. Attention mechanisms assist the model in selecting and utilizing useful information while disregarding the unimportant, thereby enhancing the quality of the fused image. The DCAM, through spatial and channel attention mechanisms, generates attention feature maps in both spatial and channel dimensions, enabling the model to better utilize information in both dimensions. ECA emphasizes the importance of avoiding dimension reduction and inter-channel interaction when using spatial attention mechanisms, allowing the model to better capture the relationships between channels while maintaining the richness of spatial information.

Figure 3 illustrates the principle of the DCAM. This mechanism takes a feature map D of size $G*Q*V$ as input. Through global average pooling and global max pooling, the average and maximum values of each channel within the feature map are computed. These processes enable the model to capture global information of each channel, determining the importance of each channel to aggregate the spatial information of the feature mapping. The new feature vectors obtained after pooling are then processed by a multi-layer perceptron (MLP). By incorporating the MLP, the model learns nonlinear relationships between channels, thus acquiring more accurate channel weights. This approach allows the model to enhance the efficiency of channel information utilization, based on the global information of the channels, thereby improving the effectiveness of image fusion. Assuming the *Sigmoid* activation function is denoted as $\sigma(\cdot)$, global average pooling and max pooling as $X^V_{AV}(\cdot)$ and $X^V_{MAX}(\cdot)$ respectively, and the weights of the MLP as Q_1 and Q_2, the output vector is obtained by element-wise addition, expressed as follows:

$$D_V = \sigma\begin{pmatrix} Q_2\left(\ \mathrm{ReLU}\left(Q_1 X^C_{AV}\right)\right) \\ +Q_2\left(\ \mathrm{ReLU}\left(Q_1 X^V_{MAX}\right)\right) \end{pmatrix} \tag{16}$$

The channel attention mechanism aids the model in identifying important channels. To enable the network to more flexibly capture contextual information, the spatial attention mechanism builds upon the channel attention mechanism, further assisting the model in identifying crucial spatial positions within these important channels. This is represented as:

$$D_A = \sigma\left(d^{7\times7}\left(X^S_{AV}, X^S_{MAX}\right)\right) \tag{17}$$

As the model needs to consider information in both the channel and spatial dimensions, and these dimensions are interrelated, it concatenates the information from both dimensions to produce the final attention feature map. This method optimizes the model in both dimensions, thereby generating fused images with higher quality:

$$D_{VA} = D_V \cdot D_A + D \tag{18}$$

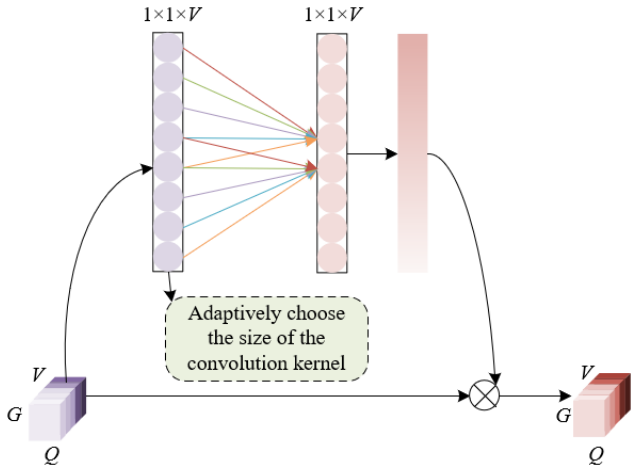

Figure 4. Principle of the adaptive channel attention mechanism

Since the optimal convolutional dimension may vary under different circumstances, for images containing extensive global information, larger convolutional dimensions may be required to capture this global information. Conversely, for images with abundant local information, smaller convolutional dimensions may be needed to capture these local details. The adaptive channel attention mechanism determines convolutional dimensions adaptively and alters the ratio between convolution kernel size and the number of channels to enhance the effectiveness of multi-view image fusion. Figure 4 presents the principle of the adaptive channel attention mechanism. Assuming J can only take odd values, denoted as $|\cdot|_{ODD}$, the formula for defining the kernel size J given the number of channels V is as follows:

$$J = \psi(V) = \left|\frac{\log_2^V}{e} + \frac{n}{e}\right|_{ODD} \tag{19}$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

From the data provided in Table 1, it is observed that as the grayscale levels decrease from 256 to 2 levels, the alignment of multi-view image registration exhibits varying degrees of reduction. For high-resolution 256×256 images, as the grayscale levels decrease from 256 to 2 levels, the alignment reduces from 11.258 to 2.014, showing the most significant decrease in alignment at the highest resolution when grayscale is reduced. For medium-resolution 128×128 and 64×64 images, the trend of alignment reduction is similar to that of the 256×256 images, but the magnitude of the decrease is less. This may indicate that medium-resolution images have certain robustness to reductions in grayscale levels. For low-resolution 32×32 and 16×16 images, especially at lower grayscale levels such as 8 and 4 levels, the alignment decreases

less, suggesting that low-resolution images are less affected by grayscale levels. These data lead to the conclusion that the moment of inertia axis method maintains high alignment for high-resolution images across different grayscale levels, demonstrating its effectiveness in multi-view image registration at high resolutions. As the grayscale levels decrease, the alignment reduces, but even at lower grayscale levels, such as 32 and 16, the method still maintains relatively good alignment.

Table 1. Alignment of multi-view images at different grayscale levels

Grayscale Levels	256 Levels	128 Levels	64 Levels	32 Levels	16 Levels	8 Levels	4 Levels	2 Levels
256×256	11.258	11.897	11.701	11.012	8.202	5.231	2.034	2.014
128×128	11.236	11.885	11.235	11.112	8.423	5.487	2.568	2.115
64×64	11.141	11.587	11.895	11.324	8.421	5.224	3.478	2.003
32×32	11.232	11.326	11.320	9.356	7.235	5.856	2.301	2.215
16×16	22.369	15.698	11.548	8.234	8.523	5.239	2.652	2.014

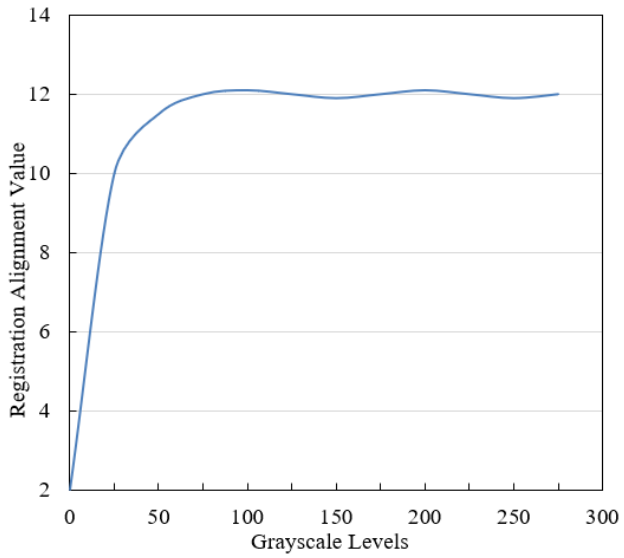

Figure 5. Change of alignment value with grayscale levels

Figure 5 depicts the change in registration alignment with varying grayscale levels. From the lowest 0 level to 300 levels, the alignment number gradually increases from 2 to 12.1, then slightly decreases to 11.9 at 300 levels. As the grayscale levels increase from 0 to 100, the alignment number significantly increases from 2 to 11.5. This indicates that at very low grayscale levels (i.e., low image contrast), the moment of inertia axis method significantly improves image registration alignment. As grayscale levels increase from 100 to 200, the alignment number gradually increases and stabilizes between 12 and 12.1. This stable alignment indicates that the moment of inertia axis method effectively maintains consistency and precision in image registration at medium grayscale levels. When grayscale levels exceed 200, the alignment number remains between 12 and 12.1, with a slight decrease to 11.9 at 300 levels. This slight decrease may be due to noise or other factors affecting registration accuracy at high grayscale levels. Overall, however, the alignment number remains relatively high. Therefore, the moment of inertia axis method is effective for multi-view image registration across the entire grayscale level range, maintaining high alignment. Particularly, a significant increase in alignment is observed as grayscale levels rise from low to medium, indicating good adaptability of the method to image grayscale and its ability to handle images of different contrasts. At high grayscale levels, there is a slight decrease in alignment, but it still remains high, demonstrating the robustness of the moment of inertia axis method. Hence, the multi-view image registration method based on the moment of inertia axis is effective and can maintain high alignment across different grayscale levels, which is significant for practical applications involving multi-view images.

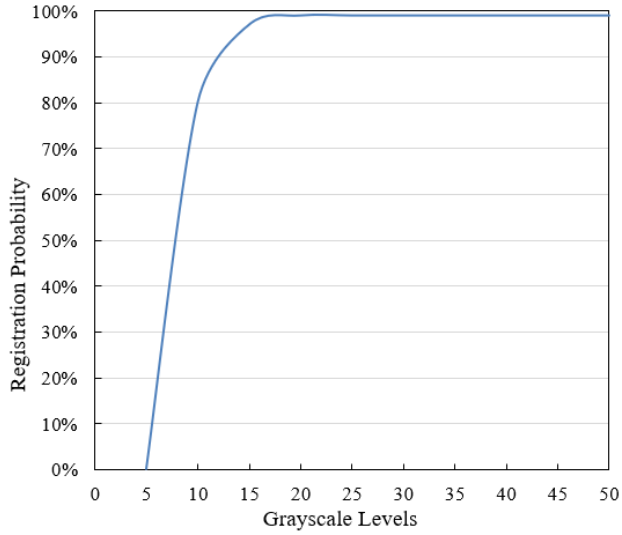

Figure 6. Registration probability variation with grayscale levels

Figure 6 presents the relationship between registration probability and grayscale levels. Within the grayscale range of 0 to 50 levels, the registration probability rapidly climbs from 0.0% to 99.0%. At the extreme low end of grayscale levels (0 level), the registration probability is 0.0%, implying that registration is nearly impossible without any grayscale information. This is expected, as registration requires distinguishable features between images. With a slight increase in grayscale levels, the registration probability sharply rises to 80.0%, indicating that even at very low grayscale levels, as long as basic contrast exists, the moment of inertia axis method can achieve high probability of registration. At medium grayscale levels (10 to 15 levels), the registration probability quickly reaches between 97.0% and 99.0%, demonstrating that this method can achieve very high registration accuracy even at relatively low grayscale levels. Once the grayscale levels reach 20 or above, the registration probability stabilizes at 99.0%, suggesting that beyond a certain level, increasing grayscale levels does not significantly improve the registration probability. This shows the high robustness of the moment of inertia axis method once a certain grayscale level is reached. Overall, the moment of inertia axis method for multi-view image registration exhibits high efficiency and robustness across different grayscale levels. Particularly, at very low grayscale levels, the method quickly increases registration probability, demonstrating a strong capability in registering low-contrast images.

The data provided in Table 2 illustrates the variation in registration probability and time with increasing grayscale levels. As the grayscale levels rise from 5 to 11, the registration probability significantly improves, starting from 91.25% and steadily climbing to 100%. There is a slight dip to 97.58% at grayscale level 10, but it then recovers and reaches

100% at level 11. This indicates that the moment of inertia axis method becomes more accurate in image registration with the increase of grayscale information. The registration time slightly increases from 0.724 seconds to 0.829 seconds, with a minimal increase over the process. This suggests that, although an increase in grayscale levels might lead to higher computational complexity due to more grayscale values being processed, the increase in registration time is very slight, indicating that the moment of inertia axis method maintains high computational efficiency relative to the improvement in registration accuracy. The moment of inertia axis method achieves high registration probability even at lower grayscale levels and continues to increase with rising grayscale levels, ultimately achieving perfect registration.

Table 2. Registration performance indicators with grayscale levels

Grayscale Levels	5	6	7	8	9	10	11
Registration Probability	91.25%	94.56%	96.84%	98.73%	98.99%	97.58%	100%
Registration Time	0.724*s*	0.748*s*	0.749*s*	0.789*s*	0.823*s*	0.817*s*	0.829*s*

Table 3. Comparison of image quality metrics for different multi-view image fusion methods

Fusion Method	*EN*	*SF*	*AG*	*CEN*	*IC*	*JE*
Laplacian Pyramid Fusion Algorithm	2.564	8.785	1.489	0.071	2.457	5.784
Wavelet Transform Fusion Algorithm	2.639	9.361	1.491	0.112	2.412	5.826
PCA Fusion Algorithm	2.647	12.142	2.124	0.057	2.639	6.415
The proposed method	2.742	13.268	2.268	0.043	3.241	6.528

Table 3 lists the comparisons of image quality metrics for several image fusion methods. According to the data, this study's method achieved the highest entropy value EN (2.742), indicating the highest information content in the fused images, implying that this method enhances the information content during fusion. Similarly, this method scored the highest in spatial frequency SF (13.268), suggesting the fused images possess optimal clarity and texture edge detail. On the average gradient (AG) metric, the proposed method (2.268) outperformed other methods, meaning the fused images are visually sharper with better image contrast. This method had the lowest cross-entropy value CEN (0.043), indicating high pixel-level similarity between the fused images and reference images, preserving the quality of the original perspectives. In terms of information capacity (IC), the proposed method scored the highest (3.241), representing the fused images' ability to present richer information. The highest joint entropy (JE) score (6.528) also belongs to the proposed method, indicating its superior effectiveness in merging edge and detail information. Overall, the multi-view image fusion method based on morphological decomposition and attention feature integration outperforms other listed methods across multiple quality metrics. These results demonstrate the proposed method's effectiveness in enhancing the fused images' information content, clarity, contrast, and overall quality. Particularly noteworthy are the high scores in spatial frequency and average gradient, indicating a significant advantage in preserving edge and texture information. Therefore, the proposed method is highly effective in processing multi-view image fusion, providing high-quality fused images for applications in related fields.

Analyzing the AG index for different multi-view image fusion methods as depicted in Figure 7, it is first noted that each model's performance varies across different sample numbers. In the context of image fusion, a higher AG typically means the fused image retains more edge information and detailed features. The data shows that the proposed model performs well, exhibiting consistency and strength above most other models, even though it is not the highest in certain samples. This indicates that the proposed method, based on morphological decomposition and attention feature integration, is also effective in preserving edge and detail information in images. Especially when compared to models without registration, the registration step is crucial for improving fusion quality.

Figure 8 provides the root mean square error index for different multi-view image fusion methods. The figure indicates that the performance of the proposed model is at a higher level. Although it does not reach the highest value, it performs better than some other models (such as the Laplacian Pyramid model, Wavelet Transform model, PCA model) in certain samples. This suggests that the proposed model has competitive effectiveness in terms of fusion error, especially when registration or morphological component analysis is not ideal.

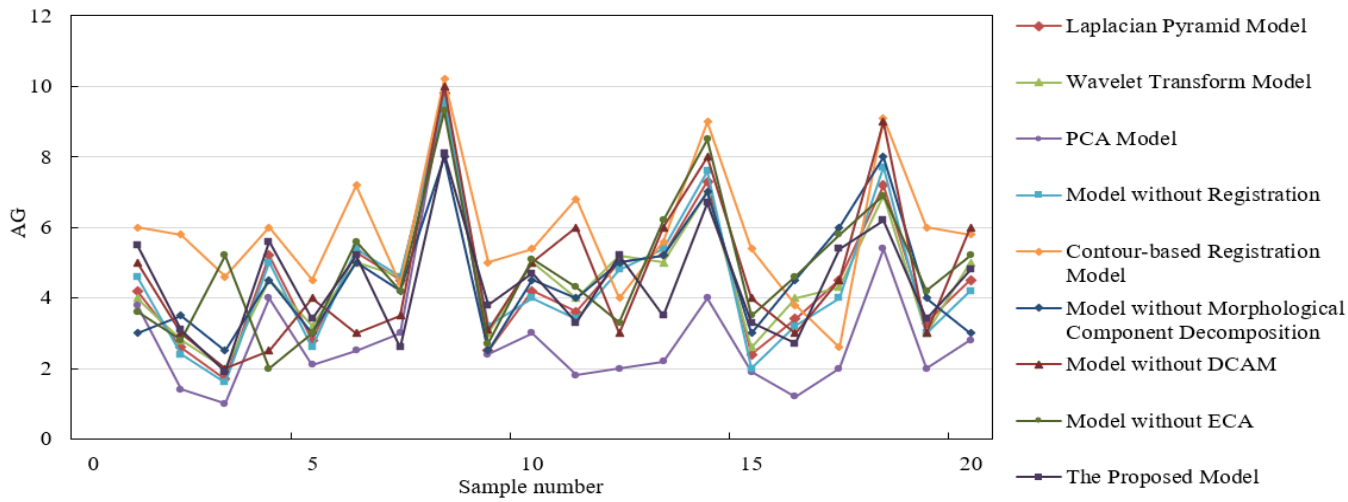

Figure 7. AG index comparison for different multi-view image fusion methods

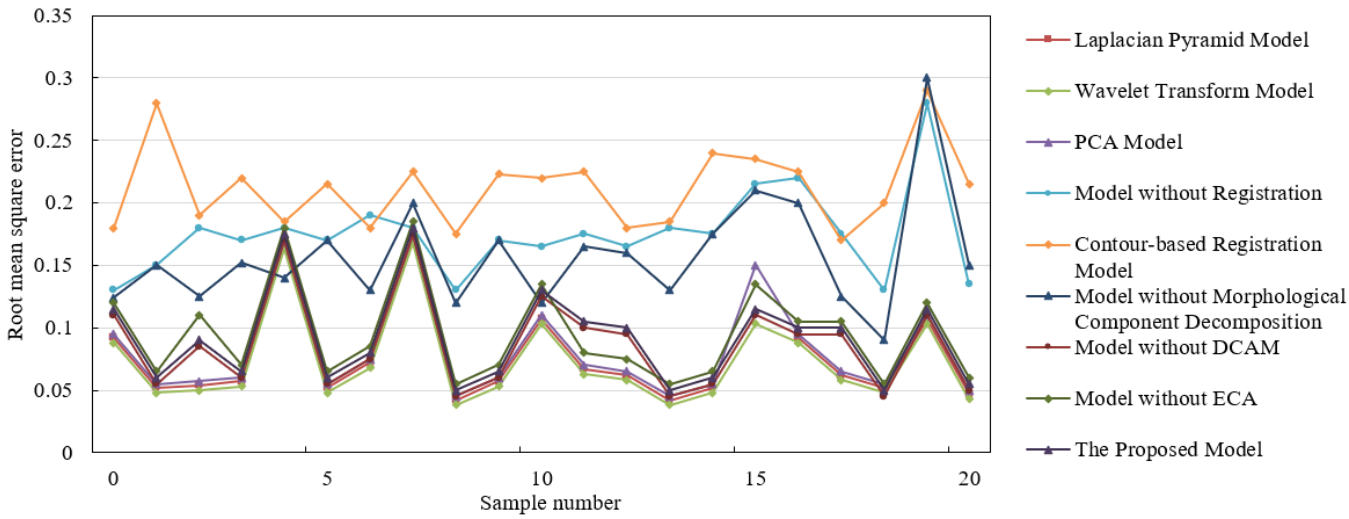

Figure 8. Root mean square error comparison of different multi-view image fusion methods

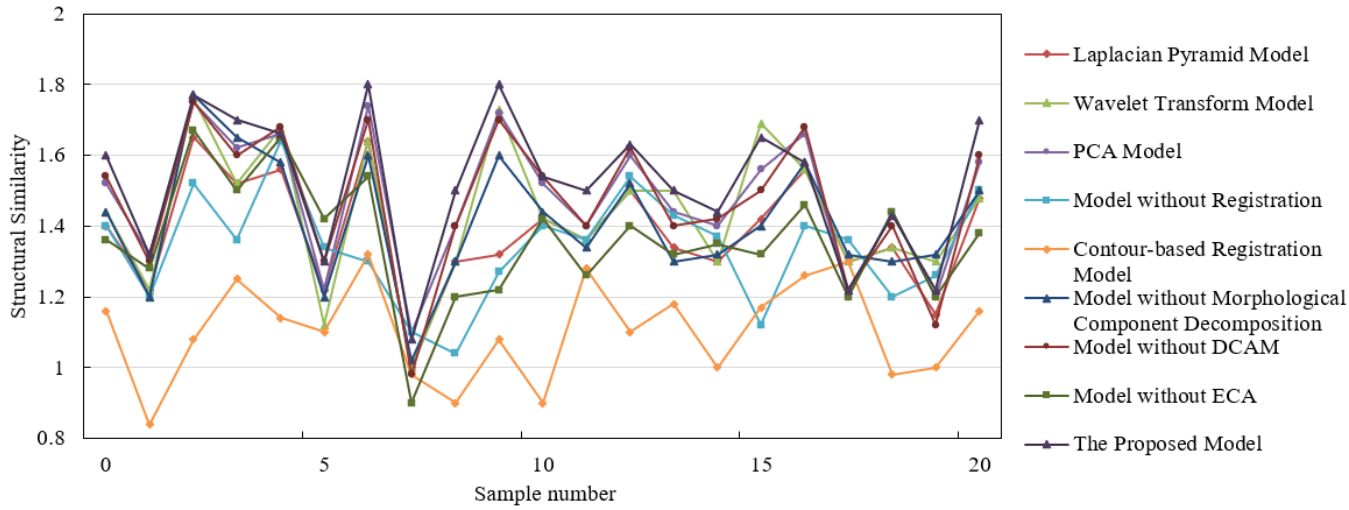

Figure 9. Structural similarity index comparison of different multi-view image fusion methods

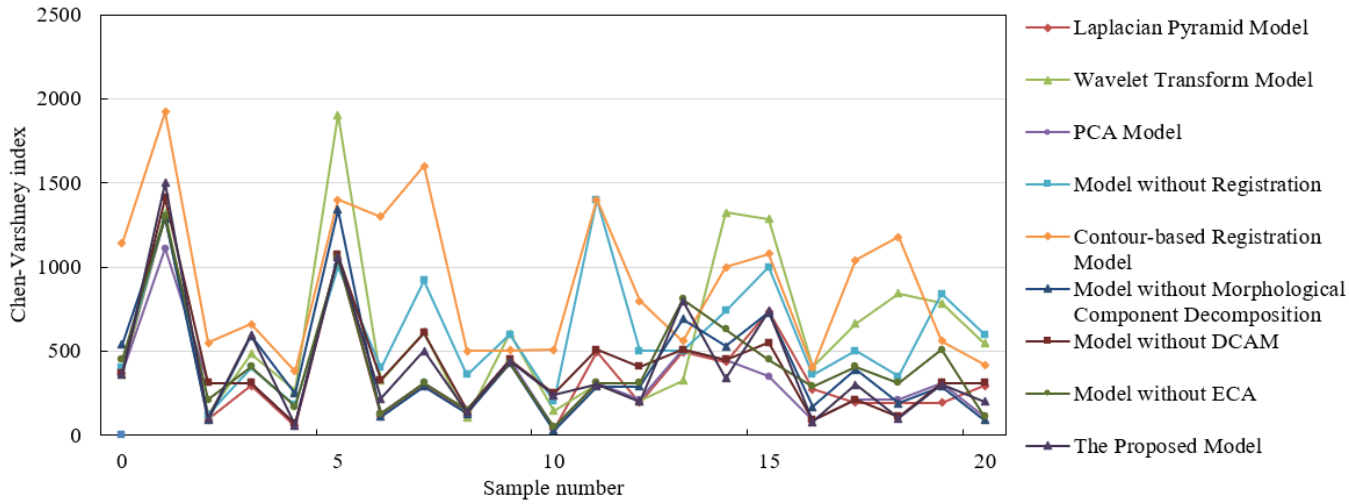

Figure 10. *Chen-Varshney* index comparison of different multi-view image fusion methods

The data presented in Figure 9 shows the structural similarity index of various multi-view image fusion methods across different sample numbers. It is evident that the proposed model displays the highest or near-highest index values on multiple sample numbers, particularly for sample numbers 0, 10, 15, 20, etc. This indicates that this model has superior structural similarity performance on these samples compared to other models. Figure 10 shows the comparison of the *Chen-Varshney* image quality index for outputs from different multi-view image fusion methods. The *Chen-Varshney* index is a metric used to assess the quality of image fusion, commonly employed to compare the performance of

different fusion techniques. Under this index, lower values typically indicate higher image quality, as it quantifies the error between the fused image and the reference image. The figure reveals that the proposed model has lower index values on most samples, especially on samples 0, 10, 20, etc., where the index values are at a lower level, indicating high-quality fused images with minimal differences from the reference images. Through comparative analysis, it can be concluded that the proposed model generally demonstrates consistent low *Chen-Varshney* index values, indicating its ability to provide fused images closer to the reference images across multiple samples. This shows the effectiveness of the method based on morphological decomposition and attention feature integration, as it maintains high fusion quality across different image contents and scenes.

Overall, the multi-view image fusion method based on morphological decomposition and attention feature fusion, as proposed in this study, exhibits effectiveness in maintaining image quality across multiple samples. This method may have advantages in preserving image details, reducing fusion artifacts, and enhancing the overall quality of the fused images. However, a more comprehensive assessment would require testing on a broader dataset and different types of image content and comparing with other advanced image fusion techniques.

Table 4. Comparative performance of different multi-view image fusion methods across viewpoint changes

Experimental Data	Algorithm	*SSIM*	*NMI*	Fusion Accuracy
First Set of Experimental Data (Different Perspectives)	*Laplacian Pyramid* Fusion Algorithm	88.84%	84.28%	85.26%
	Wavelet Transform Fusion Algorithm	85.13%	87.62%	88.36%
	PCA Fusion Algorithm	87.85%	84.25%	87.45%
	The proposed method	95.23%	94.27%	91.23%
Second Set of Experimental Data (Same Perspective)	*Laplacian Pyramid* Fusion Algorithm	83.12%	84.23%	91.52%
	Wavelet Transform Fusion Algorithm	83.56%	83.74%	91.46%
	PCA Fusion Algorithm	85.34%	81.52%	92.33%
	The proposed method	92.78%	91.35%	94.52%

Table 4 shows the performance comparison of different multi-view image fusion methods under perspective changes, including three image quality metrics: Structural Similarity (SSIM), Normalized Mutual Information (NMI), and Fusion Accuracy. For the two sets of experimental data provided (the first set involving fusion of views from different perspectives, and the second set from the same perspective), the following analysis can be made. In the first set of experimental data (different perspectives), this study's method shows the highest performance in all three metrics, with SSIM at 95.23%, NMI at 94.27%, and Fusion Accuracy at 91.23%. These figures are significantly higher than other comparative algorithms, indicating that this method maintains structural integrity and effectively preserves shared information between images, while maintaining high fusion accuracy when processing fusion of different angle views. In the second set of experimental data (same perspective), this study's method still performs the best, with SSIM, NMI, and Fusion Accuracy at 92.78%, 91.35%, and 94.52%, respectively. This further confirms the advantage of the proposed method in maintaining image visual quality and fusion accuracy.

5. CONCLUSION

The research work in this study is divided into two main parts. The first part, multi-view image registration based on the moment of inertia axis method, explores how to use this method to improve the accuracy of multi-view image registration. The moment of inertia axis method finds the principal axis of the image by calculating its geometric moments, thereby achieving effective registration of images from different viewpoints. This method is particularly useful in multi-view scenarios as it considers the global shape attributes of images and does not rely on local feature matching, thus showing good stability and accuracy even in cases of occlusions or inconsistent image quality. The second part, multi-view image fusion based on morphological decomposition and attention feature fusion, focuses on image fusion techniques. Morphological decomposition involves breaking down the image into different morphological components, helping to distinguish and extract different levels of image features. Attention feature fusion refers to assigning different weights to various features during the fusion process, enhancing important features and suppressing less significant information. Combining these two strategies effectively improves the quality of image fusion, particularly in maintaining image structure and information sharing.

Combining these two methods, this study has achieved significant results in the complex environment of multi-view image fusion. Experimental results demonstrate that the fusion method proposed in this study outperforms other commonly used image fusion methods, such as the Laplacian Pyramid fusion algorithm, Wavelet Transform fusion algorithm, and PCA fusion algorithm, in key performance indicators like SSIM, NMI, and Fusion Accuracy. These results not only confirm the effectiveness of this method in fusing views from different and same angles but also show its adaptability and stability.

In summary, the multi-view image registration method based on the moment of inertia axis proposed in this study effectively addresses registration issues between different viewpoints, enhancing accuracy and robustness in the fusion process. The multi-view image fusion method using morphological decomposition and attention feature fusion exhibits outstanding performance in extracting and merging image features, especially in maintaining image structure and enhancing information sharing. The combination of these two methods provides an effective new approach for multi-view image fusion in complex environments. Experimental results show that this method achieves better results than existing technologies in multiple key performance indicators.

AUTHOR CONTRIBUTIONS

The first author, Xiujuan Luo, and the second author, Lili Shao, contributed equally to the paper.

REFERENCES

[1] Birkfellner, W., Figl, M., Furtado, H., Renner, A., Hatamikia, S., Hummel, J. (2018). Multi-modality imaging: A software fusion and image-guided therapy perspective. Frontiers in Physics, 6: 66. https://doi.org/10.3389/fphy.2018.00066

[2] Liu, G. (2021). Image perspective restoration considering multi-granularity distortion correction algorithm. Journal of Physics, 1744(3): 032074. https://doi.org/10.1088/1742-6596/1744/3/032074

[3] Abas, A.I., Baykan, N.A. (2021). Multi-focus image fusion with multi-scale transform optimized by metaheuristic algorithms. Traitement du Signal, 38(2): 247-259. https://doi.org/10.18280/ts.380201

[4] Ren, D., He, T., Dong, H. (2022). Joint cross-consistency learning and multi-feature fusion for person re-identification. Sensors, 22(23): 9387. https://doi.org/10.3390/s22239387

[5] Panguluri, S.K., Mohan, L. (2021). A DWT based novel multimodal image fusion method. Traitement du Signal, 38(3): 607-617. https://doi.org/10.18280/ts.380308

[6] Sun, X., Tian, Y., Lu, W., Wang, P., Niu, R., Yu, H., Fu, K. (2023). From single-to multi-modal remote sensing imagery interpretation: A survey and taxonomy. Science China Information Sciences, 66(4): 140301. https://doi.org/10.1007/s11432-022-3588-0

[7] Ma, T., Xiao, W. (2020). Multi-perspective identification method based on kernel canonical correlation analysis. In 2020 Chinese Automation Congress (CAC), Shanghai, China, pp. 1819-1824. https://doi.org/10.1109/CAC51589.2020.9326578

[8] Du, J., Huang, X., Xing, M., Zhang, T. (2023). Improved 3D semantic segmentation model based on RGB image and LiDAR point cloud fusion for Automantic driving. International Journal of Automotive Technology, 24(3): 787-797. https://doi.org/10.1007/s12239-023-0065-y

[9] Jakob, P., Madan, M., Schmid-Schirling, T., Valada, A. (2021). Multi-perspective anomaly detection. Sensors, 21(16): 5311. https://doi.org/10.3390/s21165311

[10] Mo, C., Li, Y., Zheng, L., Ren, Y., Wang, K., Li, Y., Xiong, Z. (2016). Obstacles detection based on millimetre-wave radar and image fusion techniques. In IET Conference Publications, Chongqing, China, no. CP697. https://doi.org/10.1049/cp.2016.1155

[11] Wang, X., Feng, S. (2019). Multi-perspective gait recognition based on classifier fusion. IET Image Processing, 13(11): 1885-1891. https://doi.org/10.1049/iet-ipr.2018.6566

[12] Wang, W., He, W., Lei, L., Guo, G. (2014). Polluted and perspective deformation DataMatrix code accurate locating based on multi-features fusion. Chinese Journal of Electronics, 23(3): 550-556. https://doi.org/10.1002/cta.1881

[13] Zhu, Z., Liang, H., Li, Y., Qi, G. (2022). A method for quality evaluation of multi-exposure fusion images with multi-scale gradient magnitude. In Proceedings of 2021 Chinese Intelligent Systems Conference: Volume II, China, pp. 121-129. https://doi.org/10.1007/978-981-16-6324-6_13

[14] Liu, Y., Shi, Y., Mu, F., Cheng, J., Chen, X. (2022). Glioma segmentation-oriented multi-modal MR image fusion with adversarial learning. IEEE/CAA Journal of Automatica Sinica, 9(8): 1528-1531. https://doi.org/10.1109/JAS.2022.105770

[15] Meng, Y.B., Chen, X.R., Liu, G.H., Xu, S.J., Li, T.Y. (2023). High and low density multi-dimension perspective multivariate information fusion crowd counting method. Control and Decision, 38(1): 181-189.

[16] Jia, B., Xu, J., Xing, H., Wu, P. (2022). Remote sensing image fusion based on morphological convolutional neural networks with information entropy for optimal scale. Sensors, 22(19): 7339. https://doi.org/10.3390/s22197339

[17] Gogu, L.B., Kumer, S.A. (2021). Multifocus image fusion using TE-CSR technique. In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, pp. 1737-1741. https://doi.org/10.1109/ICAC3N53548.2021.9725408

[18] Zhou, T., Cheng, Q., Lu, H., Li, Q., Zhang, X., Qiu, S. (2023). Deep learning methods for medical image fusion: A review. Computers in Biology and Medicine, 160: 106959. https://doi.org/10.1016/j.compbiomed.2023.106959

[19] El-Shafai, W., Ghandour, C., El-Rabaie, S. (2023). Improving traditional method used for medical image fusion by deep learning approach-based convolution neural network. Journal of Optics (India), 52(4): 2253-2263. https://doi.org/10.1007/s12596-023-01123-y

[20] Zhang, X., Demiris, Y. (2023). Visible and infrared image fusion using deep learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(8): 10535-10554. https://doi.org/10.1109/TPAMI.2023.3261282

[21] Chuang, L.Z., Chen, Y., Chung, Y., Wu, L., Tien, T. (2022). Bragg-region recognition of high-frequency radar spectra based on deep learning and image fusion processing. International Journal of Remote Sensing, 43(18): 6766-6782. https://doi.org/10.1080/01431161.2022.2145581

[22] Obayya, M., Saeed, M.K., Alruwais, N., Alotaibi, S.S., Assiri, M., Salama, A.S. (2023). Hybrid metaheuristics with deep learning-based fusion model for biomedical image analysis. IEEE Access, 11: 117149-117158. https://doi.org/10.1109/ACCESS.2023.3326369

[23] Yuan, J., Shi, Z., Chen, S. (2021). Feature fusion in deep-learning semantic image segmentation: A survey. In International Summit Smart City 360°, Springer International Publishing, Virtual Event, pp. 284-292. https://doi.org/10.1007/978-3-031-06371-8_18

[24] Maseleno, A., Kavitha, D., Ashok, K., Al Ansari, M.S., Satheesh, N., Reddy, R.V.K. (2023). An ensemble learning approach for multi-modal medical image fusion using deep convolutional neural networks. International Journal of Advanced Computer Science and Applications, 14(8): 758-769. https://doi.org/10.14569/IJACSA.2023.0140884