**УДК 681.78, 004.932**

# SPATIAL MATCHING OF INFRARED AND VISIBLE IMAGE COORDINATES BASED ON DEPTH ANYTHING V2 AI MODEL

**Yuandong Shao** (ITMO University)
**Scientific Supervisor – Researcher, Ph.D., Aleksandr S. Vasilev**
(ITMO University)

**Introduction.** In nowadays society, infrared and visible image fusion is widely implemented in many fields such as military reconnaissance, security monitoring, and automatic driving [1]. However, due to the difference in imaging principles, these two types of images differ in their spatial coordinate systems, leading to a difficult problem of matching and fusion between them. Traditional image matching methods frequently depend on the extraction and matching of feature points, but these methods are often difficult to achieve ideal results when dealing with infrared and visible light images due to the difficulty of extracting and matching feature points. In recent years, with the development of deep learning technology, image matching and fusion methods based on deep learning have gradually become a heating point of research. In this article, we study the special matching technique of infrared and visible image coordinates based on Depth Anything V2 AI model. Depth Anything V2 AI model is an advanced deep learning model, which is able to predict the depth information from a single image with high accuracy, and has a strong generalization ability and adaptability, and it can be widely used in the fields of three-dimensional reconstruction, object recognition and localization, scene understanding and other fields [2]. In terms of spatial coordinate point matching, the model is utilized to estimate the depth of infrared and visible light images respectively, to obtain the depth information of each pixel point in the image, and then the spatial coordinate points in the two images are corresponded by coordinate transformation and matching algorithm. The fusion of the matched images is carried out using the RFN-Nest network structure [3] for fusion, which combines the thermal radiation information in the infrared image with the detail information in the visible image to generate the fused image, which provides new ideas and methods for solving the matching and fusion problems of infrared and visible images.

**Main part.** Spatial coordinate point matching based on Depth Anything V2 AI model:

1) Image preprocessing: The IR and visible light images are first preprocessed, including image enhancement, noise removal, grayscale normalization and other operations, in order to improve the image quality and reduce the influence of external factors on the subsequent processing [4].
2) Feature extraction: Depth Anything V2 AI model is used to extract features from infrared and visible light images respectively. The model can automatically learn multi-dimensional information such as depth features and texture features in the image and represent them as feature vectors.
3) Feature Matching: Based on the extracted feature vectors, a feature matching algorithm is used to find similar feature point pairs in the infrared and visible images [4]. During the matching process, reliable matching pairs are filtered out by setting the matching threshold to improve the matching accuracy.
4) Coordinate conversion: For the successfully matched feature point pairs, according to their pixel coordinates in the respective images and the depth information output from the Depth Anything V2 AI model, coordinate conversion is performed by using internal and external parameters of the camera, and the image coordinates are converted to the coordinates under the unified spatial coordinate system, so as to realize the spatial coordinate point matching between the infrared and the visible light images.

When the spatial coordinate point matching is completed, the infrared and visible images will be fused:

1) Fusion Method Selection: Considering the requirements of the application scenario and the fused image, we choose the RFN - Nest structure as the image fusion method [3]. This structure can automatically learn and fuse the features of infrared and visible - light images, and has good performance in preserving image details and semantic information.
2) Fusion Algorithm Implementation: The RFN - Nest structure includes three parts: encoder,

residual fusion network (RFN), and decoder [3]. The encoder network extracts multi - scale deep features from the source images through max - pooling operations. The RFN is used to fuse the multi - modal deep features extracted at each scale. Shallow features retain more detail information, while deep features convey semantic information. Finally, the fused image is reconstructed by the nested connection decoder network, which fully utilizes the multi - scale structural features.

3) Fusion Result Optimization: After obtaining the preliminary fused image, we carry out optimization processing, such as contrast enhancement, color adjustment, edge sharpening, etc., to improve the visual effect and quality of the fused image. At the same time, we use some objective evaluation indicators, such as peak signal - to - noise ratio (PSNR) and structural similarity (SSIM), to evaluate the fusion results [5]. According to the evaluation results, we adjust and optimize the fusion algorithm and parameters.

**Conclusion.** In this paper, we proposed a method for spatial coordinate point matching and image fusion of infrared and visible - light images based on the Depth Anything V2 AI model and the RFN - Nest structure. Through image preprocessing, feature extraction, feature matching, and coordinate transformation, we achieved accurate spatial coordinate point matching of infrared and visible - light images. Then, we used the RFN - Nest structure for image fusion, which automatically learned and fused the features of the two types of images, and had good performance in preserving image details and semantic information. The proposed method has good application prospects in fields such as military reconnaissance, autonomous driving, security and surveillance, industrial inspection, and medical image diagnosis. In future work, we will further optimize the algorithm to improve the fusion effect and efficiency. At the same time, we will explore the application of the proposed method in more fields and scenarios, and carry out more in - depth research and exploration.

**References**:


1. Liu J., et al. MATCNN: Infrared and Visible Image Fusion Method Based on Multi-scale CNN with Attention Transformer // arXiv preprint. – 2025. – arXiv:2502.01959.
2. Yang L., et al. Depth Anything V2 // arXiv preprint. – 2024. – arXiv:2406.09414.
3. Li H., Wu X.-J., Kittler J. RFN-Nest: An end-to-end residual fusion network for infrared and visible images // Information Fusion. – 2021. – Vol. 73. – P. 72–86.
4. Luo Y., Luo Z. Infrared and visible image fusion: Methods, datasets, applications, and prospects // Applied Sciences. – 2023. – Vol. 13(19). – P. 10891.
5. Pérez-Delgado M.-L., Celebi M.E. A comparative study of color quantization methods using various image quality assessment indices // Multimedia Systems. – 2024. – Vol. 30(1). – P. 40.