Image fusion in a man-portable multi-sensor optoelectronic device based on an embedded system on module

Nguyen Quang Hiep[1*], Dang Van Tang[2], Pham Hoang Son[3]

[1]Le Quy Don Technical University, 236, Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam;
[2]Photo Electric and Electronic One Member Limited Liability Company, Phu Dien, Hanoi, Vietnam;
[3]Institute of Weapons, General Department of Defense Industry, Cau Dien, Phu Dien, Hanoi, Vietnam.
[*]Corresponding author: quanghiepktq@gmail.com

Received 17 Feb. 2026; Revised 10 Apr. 2026; Accepted 14 Apr. 2026; Published 25 Jun. 2026.
DOI: https://doi.org/10.54939/1859-1043.j.mst.112.2026.149-156

ABSTRACT

Man-portable optoelectronic devices combining visible (or/and low-light) and infrared (IR) sensors require real-time image fusion to enhance situational awareness under degraded visual conditions, while operating under strict size, weight, power consumption, and cost (SWaP-C) constraints. This paper presents the design and implementation of a real-time EO/IR image fusion system based on an embedded System-on-Module (SoM) architecture for a handheld multi-sensor platform. The proposed system supports synchronized acquisition from multiple visible and IR cameras and executes parallel fusion processing using a multi-core CPU with GPU/NPU acceleration. Experimental results demonstrate real-time performance with end-to-end processing latency below 40 ms, sustained frame rates of 25 fps for fused video output. The fused image exhibits better contrast, clearer details, and improved target differentiation compared to the original input images. The experiments indicate that SoM-based architectures can provide a balanced and practical solution for real-time multi-channel EO/IR image fusion in man-portable applications.

Keywords: Multi-sensor optoelectronic system; Image fusion; Embedded System-on-Module (SoM).

1. INTRODUCTION

Multi-sensor electro-optical (EO) and infrared (IR) imaging systems have become increasingly important in observation, surveillance, and targeting applications. By combining complementary information from visible and IR sensors, image fusion improves target detection, recognition, and situational awareness under degraded visual conditions such as low illumination, fog, smoke, and partial obscuration [1-6]. Consequently, EO/IR image fusion has been widely adopted in man-portable weapon sights, handheld reconnaissance devices, and compact ISR systems.

Man-portable platforms impose stringent size, weight, power, and cost (SWaP-C) constraints while requiring real-time operation and low processing latency. These systems must simultaneously support multi-sensor acquisition, image fusion, display rendering, and user interaction within limited computing and power budgets. Therefore, the selection of an embedded processing architecture is a critical design consideration.

Conventional EO/IR fusion platforms include FPGA, CPU, and GPU-based architectures. While FPGAs provide deterministic low-latency processing, they often increase development complexity and limit software flexibility [7, 8]. GPU-centric solutions offer strong parallel computing capability but may be less suitable for power-constrained handheld systems [9, 10]. As a result, achieving an effective balance among performance, flexibility, and SWaP-C remains a key challenge.

Modern System-on-Module (SoM) platforms integrate high-performance multi-core CPUs, GPUs, neural processing units (NPUs), and rich, high-speed sensor interfaces into a compact, power-efficient form factor [11]. Their flexibility, scalability, and ease of integration make them attractive for real-time handheld EO/IR imaging systems.

This paper presents a real-time EO/IR image fusion system based on an embedded SoM architecture. The proposed solution supports synchronized image acquisition, parallel fusion processing, and flexible hardware integration within a compact and power-efficient design. Experimental results demonstrate the feasibility of SoM-based architectures for real-time EO/IR image fusion under practical SWaP-C constraints.

2. IMAGE FUSION PLATFORM SELECTION

The selection of the processing platform is a critical design factor in real-time EO/IR image fusion systems, as it directly affects latency, throughput, power consumption, interface capability, and overall system complexity. Several embedded computing platforms, including FPGAs, NVIDIA Jetson modules, Raspberry Pi boards, and RK3588-based System-on-Module (SoM) solutions, were evaluated for handheld multi-sensor applications.

FPGAs provide deterministic real-time performance and high parallelism; however, they often require additional interface components and increase hardware and software development complexity. NVIDIA Jetson platforms offer strong GPU acceleration for AI applications but are less optimized for CPU-intensive fusion pipelines and low-power operation. Raspberry Pi platforms provide a low-cost solution but lack the processing capability required for computationally demanding multi-sensor image fusion tasks.

Among the evaluated platforms, the RK3588 SoM offers a balanced combination of computing performance, power efficiency, and integration flexibility [12]. Its octa-core CPU architecture supports multi-threaded image acquisition and fusion processing, while the integrated GPU and NPU enable graphics rendering and AI acceleration. The platform also provides multiple MIPI-CSI interfaces for simultaneous connection of visible and infrared cameras, together with diverse display and communication interfaces suitable for handheld devices.

In addition, the availability of Linux-based software support and the possibility of custom carrier-board design facilitate system integration and application-specific optimization. Therefore, the RK3588 SoM was selected as the processing platform for the proposed EO/IR image fusion system.

3. IMAGE FUSION ALGORITHM AND HARDWARE SYSTEM'S DEVELOPMENT

In practical implementations of image fusion software for optoelectronic systems, particularly man-portable multi-channel binoculars integrating visible, low-light, and thermal imaging channels, the selection of an appropriate image fusion algorithm must be carefully evaluated against multiple criteria, as outlined below:

***Processing efficiency (latency and resource utilization)*:** The software must satisfy real-time operational requirements, with a minimum processing throughput of approximately 25–30 frames per second (FPS) to support continuous observation. Accordingly, the selected algorithm should exhibit moderate computational complexity, avoid excessive CPU/GPU resource consumption, and be amenable to optimization on embedded hardware architectures such as ARM Cortex-A processors or system-on-chip (SoC) devices supporting hardware acceleration frameworks (e.g., OpenCL).

Fused image quality: The output image must preserve fine spatial details from higher-resolution channels, such as edges, textures, and object contours from the visible and low-light channels, while effectively highlighting salient thermal features corresponding to high- or low-temperature regions in the infrared channel. In addition, the fusion algorithm should minimize image degradation effects, including blurring, noise amplification, and local distortions arising from spectral mismatch or imperfect inter-channel synchronization.

Suitability for practical system integration: The selected algorithm should be well matched to real-world system constraints, including compatibility with pre-registered or frame-synchronized input images, support for simultaneous acquisition from multiple video streams, and efficient multi-threaded or multi-frame processing. Ease of integration into the overall software architecture and robustness in continuous operation are also key considerations for deployment in handheld optoelectronic systems.

Evaluating the image fusion algorithms against the above criteria shows that the Laplacian pyramid–based image fusion provides superior spatial detail preservation and local adaptability compared to PCA, which often degrades contrast due to its global linear projection. Relative to wavelet-based methods, it produces fewer artifacts and avoids shift sensitivity while maintaining lower computational complexity. Its simple, multi-scale structure enables efficient implementation and robust fusion of heterogeneous modalities, making it well-suited for real-time embedded EO/IR systems [13-14].

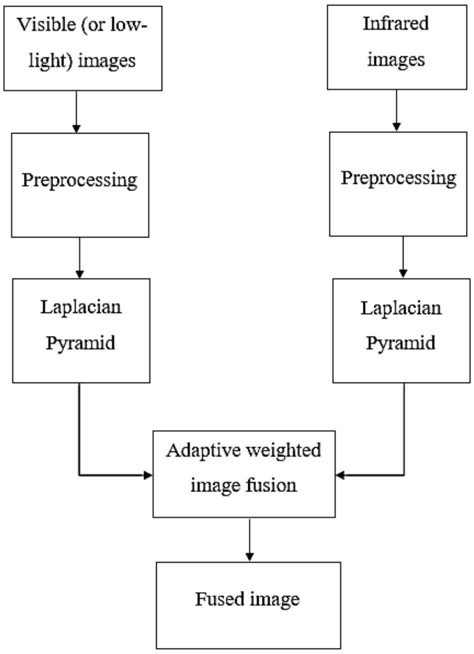


Figure 1. *The flow chart of the image fusion algorithm.*

The Laplacian pyramid algorithm (Figure 1) is implemented within a real-time processing pipeline tailored for the RK3588 embedded platform. Instead of a standalone algorithmic implementation, the fusion process is integrated into a multi-stage pipeline consisting of image preprocessing, geometric alignment, multi-scale decomposition, fusion, and reconstruction.

At the preprocessing stage, input frames from thermal and visible sensors are resized and spatially aligned. A center-fit scaling strategy combined with an affine transformation is applied using pre-calibrated translation parameters to compensate for sensor misalignment. This step ensures pixel-level correspondence between modalities before fusion.

Following alignment, Laplacian pyramids are constructed for each input image. The Gaussian pyramid is first generated through iterative downsampling, and adjacent levels are subtracted to obtain band-pass representations.

During the fusion stage, a pixel-wise adaptive weighting strategy is employed at each pyramid level. This adaptive scheme assigns higher weights to regions with stronger local responses, thereby preserving salient features such as edges and textures from both modalities.

After fusion, the final image is reconstructed through iterative upsampling and accumulation of the fused pyramid levels. Additional post-processing, including brightness adjustment and optional color mapping for thermal visualization, is applied before display.

To achieve real-time performance, the entire pipeline is executed in a dedicated processing thread, while image acquisition and synchronization are handled in parallel threads. This pipelined execution model enables overlap between stages and efficient utilization of the multi-core CPU. Computationally intensive operations, such as pyramid construction, resizing, and pixel-wise fusion, are offloaded to the GPU via OpenCL, reducing CPU load and ensuring stable throughput.

Owing to its localized operations and inherent parallelism, the Laplacian pyramid algorithm achieves stable real-time performance (approximately 25 fps) with deterministic latency, making it suitable for embedded multi-sensor fusion systems.

The development of an embedded hardware platform based on the RK3588 System-on-Module (SoM) follows a co-design approach integrating hardware and software components. The process begins with system-level requirement analysis, including sensor interfaces, display outputs, power constraints, and form factor considerations.

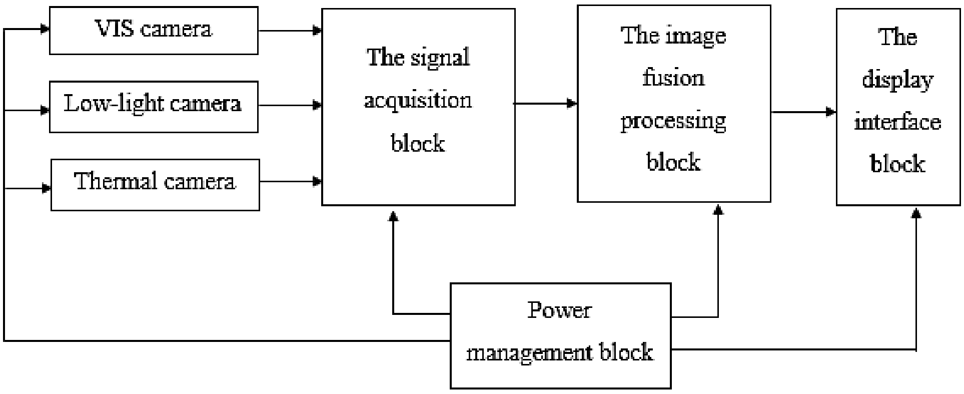


Figure 2. *Block diagram of the hardware architecture.*

A custom carrier board is designed to support the SoM and required peripherals. The hardware architecture is organized into four primary functional blocks: power management, signal acquisition, processing, and display interfacing (Figure 2).

The power management block provides multiple regulated voltage rails with proper sequencing and protection mechanisms to ensure stable operation in battery-powered environments. The signal acquisition block integrates high-speed interfaces to support multi-camera input, enabling real-time acquisition from visible, low-light, and thermal sensors. The processing block, centered on the RK3588 SoM, executes the image fusion pipeline using multi-core CPU resources, while leveraging GPU acceleration for graphics rendering. The display interface block supports multiple output standards (e.g., HDMI, eDP, MIPI-DSI), allowing flexible integration with embedded displays and graphical overlays.

From a hardware perspective, high-speed PCB design techniques are applied, including impedance-controlled routing, differential pair matching, and electromagnetic compatibility (EMC) considerations. Thermal management is also addressed to maintain system reliability under sustained workloads.

On the software side, an embedded Linux distribution (Buildroot) is customized to optimize boot time and resource usage. Device drivers and middleware are implemented to support synchronized image acquisition, multi-threaded processing, and system integration. The application layer incorporates image fusion algorithms, image processing libraries, and graphical user interfaces.

Overall, the RK3588 SoM combined with a tailored carrier board provides a compact, power-efficient, and scalable solution for real-time EO/IR image fusion in embedded systems (Figure 3).

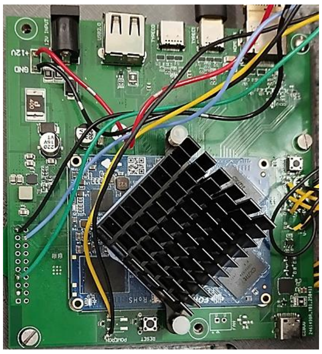
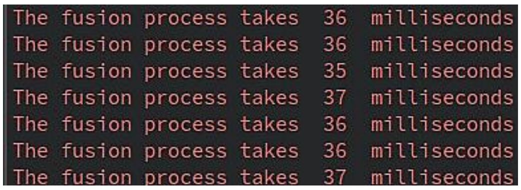


Figure 3. *Developed a board based on the RK3588 SoM.*

4. EXPERIMENTAL RESULTS AND DISCUSSION

To highlight the effeciency of image fusion implemented in the aforementioned embedded SoM platform, the experiments were conducted under various typical and considerred observation conditions [8]: when the target (a man) was obscured by smoke in daytime (Figure 5), when the target was concealed behind vegetation under low-light observation conditions (Figure 6), and when a strong light source was present in low-light observation (Figure 7).

The fusion latency was measured using a high-resolution timer (QElapsedTimer) placed around the fusion pipeline. This approach captures the end-to-end processing time of the fusion stage, excluding image acquisition and display overhead, thereby providing an accurate evaluation of the computational cost of the algorithm (Figure 4).



```
The fusion process takes  36  milliseconds
The fusion process takes  36  milliseconds
The fusion process takes  35  milliseconds
The fusion process takes  37  milliseconds
The fusion process takes  36  milliseconds
The fusion process takes  36  milliseconds
The fusion process takes  37  milliseconds
```

Figure 4. *Time cost of the fusion algorithm.*

The processing speed of the embedded system is reflected in the frame rate achieved during image fusion. The results show that, in both the fusion mode of the daylight channel with the thermal imaging channel and the fusion mode of the low-light (high-sensitivity) sensor channel with the thermal imaging channel, the frame rate substantially reaches 25 Hz.

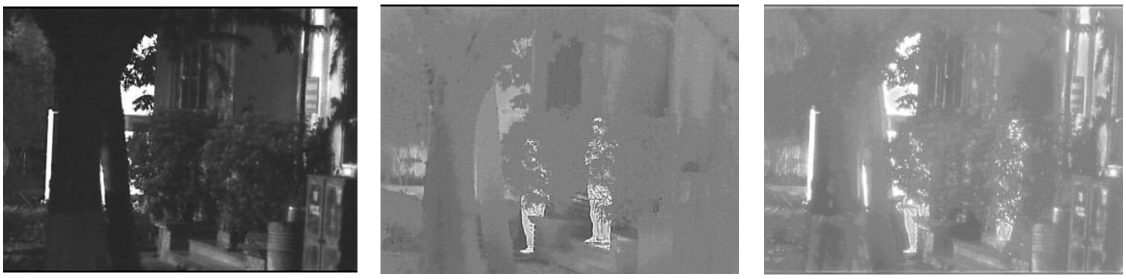
The quality of the fused image is evaluated through common objective metrics used to evaluate image quality, such as Entropy (EN) based on information theory; Average Gradient (AG) based on image features; Mutual Information (MI) based on information transfer from source images to the fused image; and Visual Information Fidelity for Fusion (VIFF) based on visual perception [9, 15]. The value of these metrics is shown in Tables 1-3.

a) Visible image. *b) Thermal image.* *c) Fused image.*

Figure 5. *The target was obscured by smoke during the day.*

Table 1. *Image quality analysis when the target was obscured by smoke in daytime.*

Metric	Visible image	Thermal image	Fused image
EN	7,2649	7,1964	7,2774
AG	1,9257	2,8694	2,1969
MI			2,41
VIFF			0,63



a) Low-light image. *b) Thermal image.* *c) Fused image.*

Figure 6. *The target was concealed behind vegetation under low-light observation conditions*

Table 2. *Image quality analysis when the target was concealed behind vegetation.*

Metric	Low-light image	Thermal image	Fused image
EN	6,1025	5,6028	6,2396
AG	3,2822	2,1891	2,8431
MI			2,18
VIFF			0,59

a) Low-light image. *b) Thermal image.* *c) Fused image.*

Figure 7. *The strong light source was present in the low-light observation.*

Table 3. *Image quality analysis when a strong light source is present at night time.*

Metric	Low-light image	Thermal image	Fused image
EN	5,9443	5,8335	6,1886
AG	0,8925	2,1559	2,3375
MI			2,26
VIFF			0,61

Based on the quantitative evaluation of fusion performance listed in Tables 1-3, it is observed that:

+ The fused image achieves the highest entropy among all inputs, indicating that the proposed method effectively preserves and slightly enhances the overall information content. This confirms that complementary features from both visible (or low-light) and thermal modalities are successfully integrated.

+ The AG of the fused image is higher than that of the visible image but lower than that of the thermal image in a smoke-obscured scene. In a low-light scene with vegetation, the AG of the fused image is higher than that of the thermal image but lower than that of the low-light image. In a low-light scene with a strong light source, the AG of the fused image is highest. This suggests that the fusion process improves spatial detail compared to the one channel while introducing some smoothing relative to other images. Such behavior is typical in real-time fusion systems where noise suppression and stability are prioritized.

+ The MI value is not less than 2.18 bits, confirming a high degree of statistical dependency between the fused image and the input modalities, indicating effective information transfer.

+ Additionally, the VIFF value is in the range of 0,59 to 0,63, demonstrating that the fused image retains a substantial portion of perceptual information from the source images and preserves most of the visually relevant information despite severe observation conditions.

Thus, after implementing the image fusion based on the embedded SoM platform, the following results were achieved:

+ The developed board successfully fuses two different image sources (thermal imaging and low-light or visible imaging) with a frame rate of 25 Hz.

+ The fused image retains the same resolution as the original images while incorporating useful information from both channels. It exhibited better contrast, clearer details, and improved target differentiation compared to the original images. This enhances target detection and recognition, especially in complex observation conditions.

+ The embedded SoM platform demonstrated its ability to synchronize multi-camera acquisition, parallel processing, and flexible fusion algorithms while meeting the SWaP-C constraints inherent to handheld devices, enabling real-time image fusion without significant latency.

5. CONCLUSIONS

Thus, in this paper, a real-time EO/IR image fusion system based on an embedded SoM architecture for a handheld multi-sensor device was presented. The proposed embedded SoM platform supports synchronized multi-camera acquisition, parallel processing, and flexible fusion algorithms while meeting the SWaP-C constraints of handheld systems, enabling real-time image fusion with high frame rate. The resulting fused image exhibits better contrast, clearer details, and improved target differentiation compared to the original images. The slight observed smoothing suggests that the current algorithm is conservative, prioritizing stability and noise suppression over aggressive detail enhancement. Handling more input sources requires higher-performance hardware, while fused image quality can be improved using deep learning methods. The experiments indicate that SoM-based architectures can provide a balanced and practical solution for real-time multi-channel EO/IR image fusion in man-portable applications.

REFERENCES

[1]. Zhang, J., *"Multi-Source Remote Sensing Data Fusion: Status and Trends"*, International Journal of Image and Data Fusion, Vol. 1, pp. 5–24, (2010).

[2]. Li, H., Qi, X., and Xie, W., *"Fast Infrared and Visible Image Fusion with Structural Decomposition"*, Knowledge-Based Systems, Vol. 204, p. 106182, (2020).

[3]. Zhang, S., Huang, F., Liu, B., Li, G., Chen, Y., Chen, Y., Zhou, B., and Wu, D., *"A Multi-Modal Image Fusion Framework Based on Guided Filter and Sparse Representation"*, Optics and Lasers in Engineering, Vol. 137, p. 106354, (2021).

[4]. N. A. Tuấn, *"Ứng dụng thuật toán tách biên hình ảnh Canny cho thiết bị quan sát trộn ảnh quang học"*, Tạp chí Nghiên cứu KH&CN quân sự, Số Đặc san FEE, tr. 273–280, (2019) (in Vietnamese).

[5]. T. V. Ha et al., *"Block Design of Edge Detection and Color Blending of Thermal Video on the Pynq-Z1 Kit Using Vivado HLS for Image Fusion Applications"*, Journal of Military Science and Technology, No. 77, pp. 128–136, (2022).

[6]. V. N. Le et al., *"A Night Vision Color Image Fusion Method Based on Statistical Color Transform Technology in YUV Color Space"*, Journal of Military Science and Technology, Vol. 93, No. 93, pp. 114–120, (2024).

[7]. A. Boutros and V. Betz, *"FPGA Architecture: Principles and Progression"*, IEEE Circuits and Systems Magazine, Vol. 21, No. 2, pp. 4–29, (2021).

[8]. Ling Zhang, Xuefei Yang, Zhenlong Wan, Dingxin Cao, and Yingcheng Lin, *"A Real-Time FPGA Implementation of Infrared and Visible Image Fusion Using Guided Filter and Saliency Detection"*, Sensors, Vol. 22, p. 8487, (2022).

[9]. Seung-Hun Yoo et al., *"Accelerating Multi-Scale Image Fusion Algorithms Using CUDA"*, IEEE International Conference of Soft Computing and Pattern Recognition, (2009).

[10]. Changqi Sun, Cong Zhang, and Naixue Xiong, *"Infrared and Visible Image Fusion Techniques Based on Deep Learning: A Review"*, Electronics, Vol. 9, p. 2162, (2020).

[11]. Tao Yang, Xinyu Zhang, Qingbo Xu, Shuangxi Zhang, and Tong Wang, *"An Embedded-GPU-Based Scheme for Real-Time Imaging Processing of Unmanned Aerial Vehicle Borne Video Synthetic Aperture Radar"*, Remote Sensing, Vol. 16, No. 1, p. 191, (2024).

[12]. *"Rockchip RK3588 Datasheet"*, Rockchip Electronics Co., Ltd., (2023).

[13]. Song, Y., Gao, K., Ni, G., and Lu, R., *"Implementation of Real-Time Laplacian Pyramid Image Fusion Processing Based on FPGA"*, Proceedings of the Electronic Imaging and Multimedia Technology V, Beijing, China, (2007).

[14]. Liangliang Li, Yan Shi, Ming Lv, Zhenhong Jia, Minqin Liu, Xiaobin Zhao, Xueyu Zhang, and Hongbing Ma, *"Infrared and Visible Image Fusion via Sparse Representation and Guided Filtering in Laplacian Pyramid Domain"*, Remote Sensing, Vol. 16, p. 3804, (2024).

[15]. Jiayi Ma, Yong Ma, and Chang Li, *"Infrared and Visible Image Fusion Methods and Applications: A Survey"*, Information Fusion, (2018).

TÓM TẮT

Trộn ảnh trong thiết bị quang điện tử đa kênh cầm tay trên nền hệ thống nhúng

Các thiết bị quang điện tử cầm tay đa kênh bao gồm các cảm biến ảnh vùng nhìn thấy (hoặc/và độ nhạy cao) và hồng ngoại (IR) đòi hỏi kỹ thuật trộn ảnh thời gian thực nhằm nâng cao khả năng nhận thức tình huống trong các điều kiện quan sát suy giảm, đồng thời phải đáp ứng các ràng buộc nghiêm ngặt về kích thước, khối lượng, tiêu thụ năng lượng và chi phí. Bài báo này trình bày thiết kế và triển khai một hệ thống trộn ảnh EO/IR thời gian thực dựa trên kiến trúc System-on-Module (SoM) nhúng dành cho một nền tảng đa cảm biến cầm tay. Hệ thống được đề xuất thu nhận đồng bộ dữ liệu từ nhiều camera ở các vùng phổ khác nhau, thực hiện xử lý trộn ảnh song song sử dụng CPU đa lõi kết hợp tăng tốc bằng GPU/NPU. Kết quả thực nghiệm cho thấy hệ thống đạt hiệu năng thời gian thực, với độ trễ xử lý đầu-cuối nhỏ hơn 40 ms và tốc độ khung hình ổn định 25 fps đối với video ảnh trộn đầu ra. Ảnh trộn thể hiện độ tương phản tốt hơn, chi tiết rõ ràng hơn và khả năng phân biệt mục tiêu được cải thiện so với các ảnh đầu vào ban đầu. Như vậy, kiến trúc SoM có thể cung cấp một giải pháp cân bằng và thực tiễn cho việc trộn ảnh EO/IR đa kênh thời gian thực trong các ứng dụng quang điện tử cầm tay

Từ khoá: Hệ thống quang điện tử đa kênh cầm tay; Trộn ảnh; System-on-Module.