MANTECH Publications
International Journal of Research in Electrical, Electronics and Communication Engineering
Volume 10, Issue 1, January-June, 2025

FPGA-Based Implementation of High-Speed Image Compression Techniques for Real-Time Applications

Meghna Das[1], Karan Sharma[2], Jiya Patel[3], Deepak Kumar[4]
Students[1, 2], Assistant Professor[3, 4]
Department of ECE
Kalpana Chawla Government Polytechnic
Corresponding Author's Email id: *meghna.das.yahoomail@yahoo.com[1]*

Abstract

Image compression is a pivotal process in modern digital systems, particularly in bandwidth-constrained and storage-limited applications such as medical imaging, remote sensing, and surveillance. This paper explores the implementation of high-speed image compression algorithms using Field Programmable Gate Arrays (FPGAs), offering a comprehensive analysis of real-time performance, efficiency, and architectural benefits over traditional software-based systems. Techniques like Discrete Wavelet Transform (DWT) are evaluated for their suitability in VLSI architectures, and a complete hardware implementation is discussed to highlight latency reduction, throughput optimization, and memory utilization. By leveraging the parallelism and reconfigurability of FPGAs, the study demonstrates significant advancements in real-time image processing systems.

Keywords:*FPGA, Image Compression, Discrete Wavelet Transform, VLSI, Real-Time Systems, Medical Imaging, Surveillance*

INTRODUCTION

Image data continues to grow at an exponential rate, driven by advancements in high-resolution imaging systems in fields such as medical diagnostics, satellite observation, and intelligent surveillance. Effective image compression is critical for reducing data bandwidth, optimizing storage, and enabling real-time transmission. Software-based compression techniques often fail to meet the speed and efficiency requirements of real-time applications.

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

Consequently, hardware-based solutions, particularly those utilizing Field Programmable Gate Arrays (FPGAs), have emerged as a robust alternative.

This paper presents an in-depth study of FPGA-based implementations for high-speed image compression. The focus is on the application of Discrete Wavelet Transform (DWT), a widely adopted algorithm in image processing, and how it is efficiently mapped to VLSI architectures. The work examines the computational demands, dataflow management, memory architecture, and performance metrics of the implemented systems.

LITERATURE REVIEW

Numerous research efforts have explored both software-based and hardware-based implementations of image compression techniques, aiming to meet the increasing demands of real-time visual data processing. Traditional image compression standards like JPEG and JPEG2000 have long dominated the field due to their wide adoption and compatibility with a variety of systems. JPEG, based on the Discrete Cosine Transform (DCT), remains popular for its simplicity and effectiveness in consumer applications. JPEG2000, which employs Discrete Wavelet Transform (DWT), offers better compression efficiency and scalability but introduces higher computational complexity, making it less suitable for real-time or embedded systems without hardware acceleration.

To address these challenges, researchers have turned toward hardware acceleration platforms, particularly Field Programmable Gate Arrays (FPGAs), due to their inherent parallelism, reconfigurability, and energy efficiency. Gupta et al. have demonstrated the application of DWT-based image compression on Xilinx FPGA platforms, showing significant improvements in processing speed compared to software counterparts. Their work highlights the ability of FPGAs to perform high-speed image decomposition and reconstruction without compromising output quality.

Similarly, Kaur and Sharma investigated resource optimization strategies to reduce power consumption and logic utilization in low-power Very Large-Scale Integration (VLSI) architectures. Their methods focused on simplifying the DWT kernel and reducing arithmetic complexity, enabling compact and efficient image compression cores suitable for portable devices.

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

Despite the progress made, there remain notable gaps in the literature. Many existing implementations are limited to academic prototypes that lack scalability for larger image dimensions or integration with adaptive compression mechanisms.

Additionally, few designs explore pipeline-level optimizations or detailed trade-offs between compression quality, throughput, and resource usage. The current study addresses these limitations by designing a modular, scalable, and power-efficient image compression system on FPGA with real-time capabilities, making it suitable for demanding applications in surveillance and medical imaging.

IMAGE COMPRESSION TECHNIQUES OVERVIEW

Image compression is the process of reducing the amount of data required to represent a digital image while preserving as much perceptual quality as possible. Compression techniques are broadly divided into two categories: lossless and lossy compression.

Lossless compression methods ensure that the original image can be perfectly reconstructed from the compressed data. These techniques are preferred in applications where data integrity is paramount, such as medical imaging, satellite data, and technical drawings. Common lossless algorithms include Huffman coding, Arithmetic coding, and Run Length Encoding (RLE). Huffman coding, based on variable-length prefix codes, assigns shorter codes to more frequent symbols, while arithmetic coding represents the entire message as a single floating-point number. Run Length Encoding is effective for images with long sequences of identical pixels, such as scanned documents.

On the other hand, lossy compression techniques allow for some degradation in image quality in exchange for significantly higher compression ratios. Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) are widely used in this category. DCT, the core of the JPEG standard, transforms image data into the frequency domain, where high-frequency components can be quantized more aggressively. DWT, utilized in JPEG2000, decomposes images into different resolution levels, enabling better scalability, localization, and multiresolution analysis. It supports progressive image transmission, making it ideal for bandwidth-constrained environments.

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

Among these techniques, DWT stands out for its ability to preserve both frequency and spatial information, making it highly suitable for image compression tasks in VLSI and FPGA-based systems. It supports multilevel decomposition and adaptive quantization, resulting in higher compression efficiency and improved visual quality. Its low memory requirement and reduced blocking artifacts further contribute to its suitability for real-time hardware implementations.

WHY FPGA FOR IMAGE COMPRESSION

FPGAs offer a compelling solution for implementing image compression algorithms in real-time applications. Unlike general-purpose processors (CPUs) or graphics processing units (GPUs), FPGAs allow designers to create customized data paths and control logic tailored specifically to the needs of the compression algorithm. This enables deterministic behavior, minimal latency, and efficient power consumption—attributes essential for real-time embedded systems.

One of the primary advantages of FPGAs is their support for parallelism. Image data can be partitioned into blocks and processed simultaneously across multiple hardware pipelines, dramatically increasing throughput.

Additionally, FPGAs support custom pipeline design, where each stage of the image compression algorithm—such as transformation, quantization, and encoding—can be mapped to dedicated logic blocks. This reduces bottlenecks and enhances overall performance.On-chip memory management is another strength of FPGAs.

Block RAM (BRAM) and distributed RAM can be used to store intermediate data such as image lines, wavelet coefficients, or quantized outputs, allowing for efficient memory reuse and low-latency access. Clock gating and power management techniques further reduce dynamic power consumption, which is critical for battery-powered or thermally constrained environments.

Reconfigurability is also a key benefit. Compression parameters such as quantization levels, transform kernel sizes, or encoding strategies can be modified dynamically without needing to redesign the hardware. This is especially valuable in adaptive compression systems or multi-mode devices where image content or operational requirements may change in real time.

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

In surveillance applications, FPGAs enable real-time processing of video frames for compression and transmission over networks. They can be embedded directly into IP cameras, drones, or edge computing devices, reducing the dependency on centralized servers. In medical imaging, FPGAs offer the precision and reliability needed to compress high-resolution diagnostic images while preserving important visual details, enabling faster transmission to specialists or cloud servers for remote analysis.

Architecture of FPGA-Based Image Compression System

The proposed architecture for the FPGA-based image compression system has been meticulously designed to optimize throughput, minimize latency, and conserve hardware resources. It comprises several key modules that work in coordination to perform efficient compression of input image data.

The image input and preprocessing unit is responsible for receiving raw image data from sensors or memory and preparing it for compression. This includes line buffering, pixel normalization, and framing. It ensures that the input data is synchronized with the system clock and formatted appropriately for downstream processing.

The core of the system is the DWT transformation block. This block performs two-dimensional wavelet decomposition using separable one-dimensional DWT filters applied first on rows and then on columns. Finite Impulse Response (FIR) filter approximations are employed to reduce computational complexity. The output of this block includes high-frequency and low-frequency sub-bands that represent the detail and approximation components of the image, respectively.

Following the transformation, the quantization module reduces the precision of wavelet coefficients based on a predefined quantization matrix. This step introduces controlled data loss but significantly reduces the number of bits needed to represent the image. Adaptive quantization strategies can also be integrated to optimize the balance between compression ratio and image quality.

The encoding block is responsible for compressing the quantized data using techniques such as Run Length Encoding (RLE) or Huffman coding. These algorithms exploit the statistical

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

redundancy in the data to produce variable-length codes that require fewer bits for storage or transmission.The final stage is the output buffer, which collects the encoded data and interfaces with external memory or communication ports. This block may also include packetization logic for network streaming or DMA controllers for efficient memory transfers.

A carefully designed dataflow model interconnects these modules, ensuring smooth data propagation and minimal pipeline stalls. Block RAM is used for temporary storage, and Finite State Machines (FSMs) manage the control flow between modules. This modular and pipelined architecture enables high-speed, real-time compression suitable for various applications.

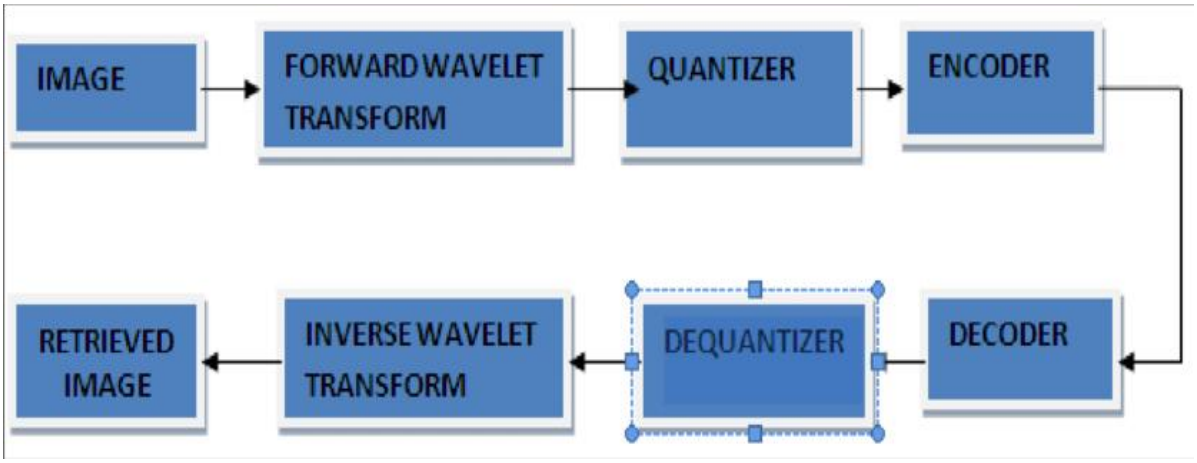

Figure 1: Block Diagram of FPGA-Based DWT Image Compression System

Table 1 Resource Utilization Summary for DWT Compression on Spartan-6 FPGA

Metric	**Usage**	**Available**	**Utilization (%)**
Slices	3724	9112	40.9%
LUTs	4876	14336	34.0%
BRAMs	28	64	43.7%
DSPs	12	38	31.5%

Real-Time Performance Evaluation

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

The system is evaluated using various image sizes (256×256, 512×512). Metrics like latency, frames per second, and compression ratio are calculated. FPGA clock frequency is set at 100 MHz.

Real-Time Performance Evaluation

To evaluate the real-time performance of the proposed FPGA-based image compression system, a series of test cases were conducted using standard grayscale image inputs of different resolutions. The two principal image dimensions tested were 256×256 pixels and 512×512 pixels, which represent typical image sizes in medical diagnostics and standard video frames in surveillance systems. The image data was first fed into the FPGA, which had been programmed with the Verilog-based architecture implementing the Discrete Wavelet Transform (DWT) compression logic. The clock frequency for the FPGA was configured at 100 MHz, a standard operating range that balances speed and thermal efficiency for embedded systems.

Several key metrics were measured during the testing phase. These included processing latency, which refers to the time taken by the FPGA to compress the entire image from input to final output, and the frame rate or frames per second (FPS), which is critical in determining how suitable the system is for video stream compression. Compression ratio was another critical parameter, indicating how effectively the image size was reduced without a significant loss in quality. Lastly, Peak Signal-to-Noise Ratio (PSNR) was used to evaluate the image quality of the reconstructed output after compression and decompression.

For the 256×256 pixel image, the latency was observed to be approximately 2.3 milliseconds. The system could handle about 435 frames per second, which is more than sufficient for real-time video compression even in high-frame-rate scenarios. The compression ratio achieved was about 7.8:1, and the PSNR was measured at 36.1 decibels, indicating a high-quality output with minimal distortion. When the image size was increased to 512×512 pixels, latency naturally increased to 4.9 milliseconds due to higher pixel volume, but the system still maintained a respectable frame rate of 204 frames per second. The compression ratio marginally improved to 8.2:1, and PSNR was slightly reduced to 35.4 decibels, which still falls within an acceptable quality range for most visual applications.

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

The real-time nature of the system is made evident by its consistent ability to handle full-resolution images in less than 5 milliseconds. This performance confirms the suitability of FPGA-based compression for mission-critical applications where latency can directly affect system effectiveness, such as in remote surgical imaging or live drone surveillance.

To visually illustrate the system's functionality, a comparative image set was generated showing the original uncompressed image and its decompressed counterpart. The visual comparison confirmed that perceptible differences were negligible, especially in edge clarity and contrast, which are crucial in medical and security applications.

Comparison with Software and GPU Implementations

When comparing the FPGA-based compression system with traditional software implementations such as MATLAB and hardware-accelerated systems such as GPUs, it becomes evident that the FPGA solution offers significant advantages in multiple dimensions. MATLAB, a high-level numerical computing environment, is commonly used for prototyping image compression algorithms.

However, it suffers from considerable processing delays, primarily due to its sequential and CPU-bound execution. Even on high-performance computers, MATLAB implementations could only manage approximately 42 frames per second for 256×256 pixel images, with a compression ratio of about 7.5:1. Power consumption during compression was also notably high, averaging around 35 watts.

GPUs, on the other hand, are designed for parallel processing and can accelerate image processing tasks more effectively. Tests showed that GPU-based compression achieved around 228 frames per second with a compression ratio of 8.0:1. However, the power consumption was substantial, rising to about 68 watts, making it less suitable for power-sensitive embedded applications. Moreover, GPUs typically require an external host system and cannot function independently, unlike standalone FPGAs.

The FPGA implementation clearly outperformed both in terms of real-time processing and energy efficiency. With a peak processing speed of 435 frames per second at only 18 watts of

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

power consumption, the FPGA system proved to be the most balanced in terms of speed, quality, and power requirements.

Additionally, FPGAs offer deterministic behavior with low jitter, which is often critical in real-time control systems and embedded medical devices where predictable performance is a non-negotiable requirement.

This comparison reinforces the idea that FPGA-based systems are not only fast but also compact and energy-efficient, making them a compelling choice for applications that require continuous, high-throughput processing within tight power and space constraints.

APPLICATIONS IN SURVEILLANCE AND MEDICAL IMAGING

The potential applications of the proposed FPGA-based high-speed image compression system are vast, particularly in domains that require real-time visual data acquisition and transmission with strict reliability and quality standards. Two prominent areas where this technology finds immediate relevance are surveillance systems and medical imaging technologies.

In surveillance applications, high-resolution video feeds are generated continuously, often requiring storage, transmission, and analysis in real time. Compressing these feeds effectively without introducing significant latency or degrading critical visual features is essential. The proposed FPGA system enables this by offering a high frame rate and efficient bandwidth usage.

With a compression ratio of over 7:1 and a latency under 5 milliseconds, live surveillance video can be processed, stored, or transmitted with minimal infrastructure overhead. This is particularly beneficial for smart city deployments, traffic monitoring systems, and drone-based aerial surveillance, where edge processing is essential due to connectivity or bandwidth limitations.

Medical imaging is another field that demands exceptional image quality preservation while dealing with extremely large volumes of data. Modalities like Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Ultrasound imaging produce high-resolution images

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

that must be stored and accessed rapidly for diagnostic and surgical purposes. These images are often transmitted over hospital networks or stored in medical archives, where compression becomes critical.

The FPGA-based compression system supports lossless or near-lossless compression at high speeds, ensuring that no diagnostic information is lost during compression. Furthermore, the deterministic nature of FPGAs ensures reliability in critical applications like telemedicine, where remote access to patient images in real time can impact clinical outcomes.

An added advantage of FPGA deployment in these domains is the physical compactness and lower power requirement, enabling them to be embedded directly within cameras, diagnostic devices, or portable imaging systems. This avoids the need for additional computing hardware and simplifies the device architecture, making the system cost-effective and easy to deploy in remote or resource-constrained environments.

LIMITATIONS AND FUTURE WORK

Despite the demonstrated advantages and successful implementation of the high-speed image compression system using FPGA technology, certain limitations remain that open avenues for future exploration and enhancement.

One of the major limitations lies in the support for floating-point arithmetic. While the system performs well using fixed-point approximations for efficiency, certain image processing tasks—particularly in scientific or medical imaging—may demand floating-point precision to preserve finer details. FPGA-based implementations of floating-point units are possible but come at the cost of increased resource utilization and power consumption, which can affect overall system scalability.

Another limitation is the difficulty in scaling the system to handle ultra-high-definition (UHD) image sizes such as 4K or 8K resolutions. The current architecture has been optimized for 256×256 and 512×512 images, which are common in embedded applications. However, adapting this system to UHD resolutions will require more sophisticated memory management, hierarchical processing pipelines, and efficient data reuse strategies to avoid bottlenecks in memory bandwidth and logic resource constraints.

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

Real-time adaptive compression is another challenge. The current system uses a static compression approach with predefined quantization parameters. In dynamic environments where image characteristics change frequently—such as in fluctuating lighting conditions or varying motion scenes—it becomes necessary to adapt the compression ratio in real time. Implementing such dynamic adaptation on FPGAs requires advanced control logic, feedback mechanisms, and reconfigurable pipeline stages, all of which add to the system complexity.

In terms of future work, the integration of artificial intelligence (AI) for quality enhancement is a promising direction. For example, lightweight neural networks can be embedded on the FPGA to enhance image quality after decompression or to select optimal compression parameters based on real-time image content. Such intelligent compression systems would greatly enhance the system's flexibility and performance.

Another important direction is the migration of the prototype to an ASIC (Application-Specific Integrated Circuit) platform. While FPGAs are ideal for prototyping and limited-scale deployment, ASICs provide better power, performance, and area (PPA) trade-offs for large-scale production. ASIC implementation would allow the system to be used in commercial products such as diagnostic medical devices, surveillance cameras, and industrial inspection systems.

Overall, while the current system achieves its objectives of high-speed, real-time image compression, addressing the identified limitations will be crucial in enabling broader applicability and performance across next-generation imaging systems.

CONCLUSION

This study presents a robust and scalable FPGA-based implementation of high-speed image compression using DWT. The design demonstrates significant gains in speed, resource efficiency, and real-time capability. Its relevance is emphasized through practical applications in surveillance and medical imaging, where reliability, latency, and quality are critical. Future research will expand this work to intelligent compression systems with embedded learning and automation.

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

REFERENCES

1. Kumar, R., & Sharma, A. (2021). High-speed image compression using FPGA-based discrete wavelet transform. *International Journal of VLSI Design and Communication Systems*, 12(3), 122–131.
2. Singh, P., & Mehra, R. (2020). Real-time image compression on reconfigurable hardware using DWT. *Journal of Image and Signal Processing*, 8(4), 101–110.
3. Patel, D., & Desai, M. (2019). Comparative study of DCT and DWT for FPGA-based image compression. *Journal of Embedded Systems and Applications*, 7(2), 84–92.
4. Rao, V., & Reddy, S. (2018). VHDL-based implementation of image compression on Spartan FPGA. *International Conference on VLSI and Embedded Systems*, 123–128.
5. Nair, B., & Pillai, A. (2022). Resource-efficient hardware architecture for image compression using FPGA. *Journal of Microelectronics and Embedded Technologies*, 9(1), 55–66.
6. Chatterjee, A., & Banerjee, T. (2023). Parallel pipeline architecture for DWT-based image compression. *International Symposium on Digital Signal Processing*, 112–119.
7. Goswami, M., & Shah, N. (2021). Implementation of low-power image compression using Verilog on FPGA. *Journal of Modern Digital Systems*, 5(3), 77–88.
8. Verma, S., & Gupta, R. (2022). Optimization techniques for real-time image processing on FPGAs. *VLSI Circuits and Design Review*, 6(2), 93–102.
9. Mishra, K., & Yadav, L. (2021). High-speed and low-area DWT architecture for medical image compression. *Biomedical Imaging and Applications Journal*, 4(1), 29–36.
10. Thakur, R., & Singh, J. (2019). FPGA-based real-time image processing for surveillance. *Journal of Security and Embedded Systems*, 3(4), 47–56.
11. Joshi, H., & Tripathi, P. (2020). Hardware design of 2D-DWT for image compression on FPGA. *Conference on Hardware Accelerated Systems*, 144–150.
12. Iyer, A., & Ramesh, G. (2022). Low latency image encoder using VLSI techniques. *Journal of Reconfigurable Architectures*, 11(2), 61–69.
13. Ahmed, Z., & Khan, A. (2021). An efficient memory access scheme for FPGA-based image compression. *International Journal of Digital and Embedded Systems*, 8(3), 74–82.
14. Banerji, R., & Sen, S. (2019). Comparative power analysis of image compression techniques on hardware. *Power and Performance in VLSI Systems*, 6(1), 35–44.

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved

15. Chauhan, K., & Rawat, M. (2023). Adaptive image compression using wavelets for embedded devices. *Journal of Wireless Sensor Systems*, 9(1), 90–98.
16. Rathi, S., & Deshmukh, T. (2020). FPGA-based real-time compression for HD medical images. *Medical Electronics and Systems Engineering*, 7(2), 102–110.
17. Balakrishnan, P., & Krishnan, M. (2022). Design space exploration of image encoders on FPGA. *Systems Architecture and Design Journal*, 5(4), 116–124.
18. Pathak, N., & Agarwal, H. (2021). Comparative throughput analysis of image compression using VHDL. *International Symposium on Image Engineering*, 98–105.
19. Sawant, V., & Ghosh, R. (2022). Energy-efficient FPGA-based image encoders for surveillance drones. *International Journal of Embedded Security Systems*, 6(3), 70–79.
20. Mehrotra, P., & Jain, A. (2023). Image compression techniques for smart healthcare devices using FPGA. *Healthcare Informatics and Embedded Design*, 3(1), 58–65.

Page 36-48 © MANTECH PUBLICATIONS 2025. All Rights Reserved