Fusion of Multi-Exposure Videos in Gradient Domain

by

Aryan Rawat
B.Tech., Guru Gobind Singh Indraprastha University

A Report Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master Of Engineering

in the Department of Electrical and Computer Engineering

Fusion of Multi-Exposure Videos in Gradient Domain

by

Aryan Rawat
B.Tech., Guru Gobind Singh Indraprastha University

Supervisory Committee

Dr. Panajotis Agathoklis, Supervisor
(Department of Electrical and Computer Engineering, University of Victoria)

Dr. Mihai Sima, Co-Supervisor
(Department of Electrical and Computer Engineering, University of Victoria)

ABSTRACT

We propose a gradient-domain framework for fusing multi-exposure video sequences that preserves spatial detail while maintaining temporal coherence. The method operates on the luminance channel to compute 3-D gradients along space–time (x, y, t). Gradients from all exposures are combined using a contrast-aware rule (favoring large-magnitude, well-exposed derivatives), producing a single fused gradient field. Because edited or combined gradients need not be curl-free, we reconstruct the fused video with a wavelet-based 3-D integration algorithm: we first obtain the 3-D Haar decomposition directly from gradients, then synthesize the video while applying an iterative Poisson update at each resolution level to suppress artifacts that arise when the zero-curl condition is violated. This reconstruction is linear-time in the number of samples, and prior studies show the Poisson step improves robustness in the presence of noise. After luminance reconstruction, chrominance channels are blended and the result is normalized to yield the final fused video.

Empirically, the approach leverages the known benefits of gradient-domain processing for fusion/editing and the tight link between Hudgin/Fried gradient discretizations and Haar filters, enabling efficient multi-resolution integration. Qualitative and quantitative evaluations on standard sequences indicate that the fused outputs retain details from all exposures with reduced flicker and stable structure across frames, illustrating the suitability of the Haar–Poisson 3-D reconstruction backbone for gradient-based video fusion applications.

References to foundations used in this work include the 2-D/3-D Haar–Poisson reconstruction pipeline and its complexity, the Poisson formulation addressing non-integrable gradients, and prior gradient-domain fusion/editing literature.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

My Parents for their unconditional love, support, and constant encouragement throughout this journey.

My Friends for standing by me, motivating me, and helping me stay grounded during the challenging phases of this work.

My Supervisor, Dr. Pan Agathoklis, for his exceptional guidance, mentorship, and invaluable feedback, which shaped the direction and quality of report.

My Co-supervisor, Dr. Mihai Sima, for his helpful insights, technical advice, and continuous support.

The University of Victoria, for providing an outstanding academic environment, resources, and opportunities that made this research possible.

I believe I know the only cure, which is to make one's centre of life inside of one's self, not selfishly or excludingly, but with a kind of unassailable serenity to decorate one's inner house so richly that one is content there, glad to welcome anyone who wants to come and stay, but happy all the same in the hours when one is inevitably alone.

Aryan Rawat

DEDICATION

I dedicate this thesis to all the most special people in my life: my parents, siblings, friends, teachers, and mentors, whose guidance, encouragement, and unwavering support have shaped my journey and made this achievement possible.

Chapter 1

Introduction

1.1 Background and Motivation

Gradient-domain processing has emerged as a key method in image and video analysis in recent years. In contrast to classic spatial-domain techniques that adjust raw pixel intensities, gradient-domain approaches emphasize edges, texture transitions, and subtle structural differences that have a significant impact on human visual perception. This idea stems from the finding that local contrast has a greater effect on the human visual system than absolute brightness values [1, 2]. By working directly with gradient information, algorithms can maintain structural consistency and reduce typical errors like haloing or intensity drift, which frequently occur in intensity-based fusion.

Earlier research concentrated on combining multiple focus or exposure photos to produce a single image that incorporates the most instructive areas from each input. Techniques like exposure fusion [3] and Laplacian pyramid fusion [4] aimed to combine data based on contrast and intensity measurements. Nevertheless, these techniques frequently work independently on every frame or exposure, which leads to visual irregularities when used on consecutive video frames such as flicker, detail loss, or brightness fluctuations.

Gradually, attention turned to the gradient domain in order to overcome these constraints. By operating in the derivative space, the fusion can highlight high-frequency information, which contains perceptual detail. Low-frequency components are then reconstructed using integration or Poisson-based reconstruction [5, 6]. The introduction of Haar wavelet-based Poisson reconstruction [2] enabled multiresolution

computation with linear complexity $O(N)$, permitting elimination of artifacts during synthesis while maintaining edges and smooth transitions. This approach is ideal for large-scale image reconstruction and video applications.

Video fusion presents new difficulties compared to static image fusion. Because of variations in local gradients between successive frames, frame-by-frame fusion can result in noticeable flicker. Gradients must therefore be handled consistently across spatial and temporal dimensions to ensure temporal smoothness. Gradient-domain reconstruction may be extended to the three-dimensional case (x, y, and t) to explicitly describe temporal continuity, which leads to more stable fusion results and smoother frame transitions [6, 7].

Multi-exposure video fusion in the gradient domain provides a methodical way to produce temporally coherent, detail-preserving films under different lighting circumstances. The Poisson-based reconstruction method preserves high contrast information while maintaining global consistency by merging gradients taken from several exposures. This approach finds practical application across diverse domains. In cinematography, filmmakers can capture challenging scenes—such as indoor spaces with windows or stage performances with spotlights—without highlight or shadow detail loss. Similarly, medical imaging benefits significantly: endoscopic surgical video maintains continuous visual information when navigating between brightly lit and shadowed tissue regions. The gradient-domain approach ensures temporal stability in both contexts, eliminating brightness fluctuations that would otherwise require extensive post-production correction or potentially compromise surgical precision. Beyond video applications, the extension to 3D temporal fusion (x, y, z, t) addresses challenges in volumetric scene capture, enabling coherent processing of multi-view camera arrays and depth-sensing systems while maintaining both spatial geometric consistency and temporal smoothness across viewpoints and time.

1.2 Problem Definition

This work focuses on reconstructing a temporally consistent video sequence from multiple input videos taken at various exposure settings. Assume that N registered video sequences denoted by $\{I_1(x, y, t), I_2(x, y, t), \ldots, I_N(x, y, t)\}$ are obtained in various lighting scenarios. The goal is to create a fused output video $F(x, y, t)$ that maintains temporal coherence between frames while preserving spatial information from all exposures. Traditional multi-exposure image fusion (MEF) methods [4, 3]

function directly in the intensity or Laplacian domains. These techniques can produce visually appealing still images, but when used independently on each video frame, they ignore the temporal relationships between frames, leading to observable flicker and brightness irregularities.

Gradient-domain fusion offers a physically justified paradigm to get around these problems. The fusion is carried out on the spatial and temporal derivatives of every input sequence rather than pixel intensities, which yields a fused gradient field $\tilde{\nabla}F$ that highlights temporal and spatial detail. The fused video is reconstructed by solving the Poisson equation, which requires global consistency between the integrated intensity and the combined gradient field [8, 5].

However, motion, sensor noise, and minor misalignments make real gradients less conservative, which results in low-frequency drift and non-integrable fields. Hampton and Agathoklis [9] presented a wavelet-based Poisson solver that applies iterative correction at different scales to achieve accurate reconstruction with linear computational complexity. This framework's extension to three dimensions $(\partial_x, \partial_y, \partial_t)$ improves smoothness over time and decreases flicker by explicitly including temporal gradients.

Recent efforts have started to use exposure-aware processing and gradient information to bridge spatial and temporal consistency. Zheng *et al.* [10] suggest a fusion technique that combines gradient-driven weighting with unsupervised learning to better retain structural detail. Xu *et al.* [11] offers a thorough analysis of multi-exposure fusion approaches for video, pointing out the necessity for greater temporal coherence. Luzardo *et al.* [12] highlights the importance of temporal regularization in attaining perceptual stability across frames during HDR conversion. These findings support the idea that explicitly modeling temporal gradients and exposure variation improves temporal stability and perceptual quality compared to static, per-frame fusion alone.

Hence, the problem can be stated as follows: *Given multiple exposure video sequences exhibiting variations in illumination and motion, compute a fused spatio-temporal gradient field $\tilde{\nabla}F$ that maximizes spatial detail and temporal consistency, and reconstruct it into a coherent intensity video $F(x, y, t)$ using a 3-D Haar–Poisson framework.* This formulation combines multi-resolution wavelet reconstruction with gradient-domain fusion to achieve temporally stable, natural-looking video synthesis.

1.3 Research Objectives

The development of a robust, gradient-domain framework for multi-exposure video fusion requires clear objectives that bridge theoretical formulation, algorithm design, and practical evaluation. This research aims to extend two-dimensional image fusion principles to the three-dimensional spatio-temporal domain, leveraging the computational efficiency of Haar–Poisson reconstruction and the perceptual stability of gradient-domain modeling.

1.3.1 Primary Objective

The primary goal is to apply and implement a three-dimensional gradient-domain fusion algorithm that can generate a temporally consistent video sequence from multiple exposure inputs. This framework integrates the mathematical foundations of gradient-domain reconstruction [5, 8] with wavelet-based multiresolution synthesis, and incorporates recent gradient-based fusion for dynamic scenes [6].

1.3.2 Specific Objectives

- **Formulate a unified 3-D gradient-fusion and reconstruction framework.** Extend the gradient-magnitude fusion model of Sujoy Paul *et al.* from 2-D images to three-dimensional video volumes, and integrate it with the 3-D Haar–Poisson reconstruction algorithm developed by Sevcenco *et al.*. This involves deriving the fused spatio-temporal gradient field $\tilde{\nabla}F(x, y, t)$ across multiple exposure sequences and reconstructing the intensity volume through Ioana Sevcenco's multiresolution Haar-based iterative Poisson solver, ensuring robustness to non-integrable gradients and achieving temporally smooth, visually consistent video synthesis.

- **Implement the complete fusion pipeline.** Develop a MATLAB implementation consisting of gradient extraction, fusion, and reconstruction stages, following the structure of the gradient-domain editing framework proposed in [6], optimizing numerical stability and computational performance for large video sequences.

- **Evaluate fusion quality and temporal consistency.** Quantitatively assess the fused results using PSNR, SSIM, temporal SSIM (t-SSIM), Flicker Index,

and Entropy, following evaluation protocols from [13, 14].

1.4 Scope and Contributions

The scope of this research lies within the field of computational imaging, specifically in gradient-domain signal reconstruction and its extension to video fusion applications. The work focuses on algorithmic analysis and software implementation, rather than on hardware development or sensor design. It investigates how three-dimensional gradient reconstruction techniques can be applied to fuse multiple exposure video sequences into a single, temporally stable output. The research is limited to algorithmic development, numerical reconstruction, and experimental evaluation, including derivation of a unified three-dimensional Poisson-based gradient reconstruction model, implementation in MATLAB with emphasis on accuracy and efficiency, and evaluation on multi-exposure video datasets to quantify performance improvements in temporal coherence and visual fidelity. Hardware calibration, camera response modeling, and optical flow estimation are considered outside the primary scope of this study.

1.4.1 Key Contributions

The major contributions of this thesis are summarized below:

- **Unified 3-D Gradient-Domain Fusion and Reconstruction Framework.** Build upon the 2-D and 3-D Haar–Poisson reconstruction methodology of Sevcenco *et al.* [6] to formulate a consistent multiresolution gradient-domain model for video. Extend this framework to multi-exposure video fusion by incorporating spatio-temporal gradients and exposure-consistent fusion rules inspired by existing multi-exposure methods [10], enabling temporally stable and luminance-balanced synthesis across frames.

- **Implementation and Quantitative Analysis.** Multi-exposure video datasets are curated and evaluated using metrics such as PSNR, SSIM, t-SSIM, Entropy, and Flicker Index, providing a reproducible benchmark for future gradient-domain fusion research.

The work therefore contributes both a theoretical advancement—a unified 3-D Haar–Poisson reconstruction model—and a practical implementation that validates

gradient-domain fusion as an effective, interpretable alternative to purely data-driven fusion methods [7].

1.4.2 Report Organization

This report is organized into six chapters that progress from theoretical foundations to the proposed methodology, experimental validation, and concluding remarks. Each chapter builds upon the previous one, maintaining a logical flow between gradient-domain theory, algorithmic design, and application to multi-exposure video fusion.

- **Chapter 1 – Introduction:** Presents the motivation, background, problem definition, objectives, and scope of the research. It introduces the importance of gradient-domain methods and outlines the main contributions of this work.
- **Chapter 2 – Literature Review:** Surveys the existing body of work related to gradient-domain imaging, Poisson-based reconstruction, multi-exposure fusion, and temporally consistent HDR video techniques. It compares classical reconstruction frameworks with modern learning-based fusion methods to identify current research gaps.
- **Chapter 3 – Background Theory:** Describes the mathematical foundations of gradient-domain reconstruction, including the definitions of gradients, divergence, and the Poisson equation. The chapter reviews discrete gradient models (Hudgin and Fried), Haar wavelet decomposition, and multiresolution Poisson solvers following the formulations of Sevcenco and Agathoklis.
- **Chapter 4 – Proposed Methodology and System Design:** Details the complete gradient-domain video fusion framework. It explains the process of gradient extraction, gradient fusion, and 3-D Haar–Poisson reconstruction, emphasizing computational efficiency and temporal smoothness.
- **Chapter 5 – Experimental Setup, Results, and Discussion:** Details the datasets, preprocessing, and quantitative metrics used to evaluate the proposed method, along with key implementation parameters. Presents visual and quantitative results, including comparisons against frame-wise exposure fusion and recent HDR video methods, with analysis centered on spatial quality, temporal coherence, and computational complexity. Interprets the findings in the context

of gradient-domain processing, articulates limitations, and outlines avenues for improvement.

- **Chapter 6 – Conclusions and Future Work:** Summarizes the main contributions and outcomes of the thesis, highlighting how the proposed 3-D Haar Poisson framework advances gradient-domain video fusion. Outlines directions for continued research, including optical-flow-based gradient alignment, real-time GPU implementations, and the integration of deep neural networks for gradient prediction and exposure selection.

Overall, this organization ensures a coherent presentation of the theoretical background, methodology, and experimental results, providing a complete understanding of the *Fusion of Multi-Exposure Videos in the Gradient Domain*.

Chapter 2

Literature Review

2.1 Gradient-Domain Techniques in Imaging

Gradient-domain techniques form the mathematical foundation for a wide range of image reconstruction, editing, and fusion applications. By operating on spatial derivatives rather than raw pixel intensities, these methods capture the structural components of an image—such as edges, textures, and contrast transitions—more effectively than traditional spatial-domain approaches. This perspective has proven valuable in tasks that require fine control over local detail and global consistency.

The origins of gradient-domain imaging can be traced to early computer vision work on surface integration and shape recovery. Frankot and Chellappa [15] introduced one of the first practical frequency-domain solutions that enforced integrability constraints on gradient fields using Fourier methods. Building on this, Simchony, Chellappa, and Shao [8] formulated a direct analytical solution to the Poisson equation, linking the Laplacian of an image to the divergence of its gradients. A major breakthrough came with the Poisson Image Editing framework of Pérez *et al.* [5], which demonstrated how modified gradients could be smoothly reintegrated to produce natural composite images. Their variational formulation established the foundation for gradient-based manipulation of illumination, texture, and contrast.

These benefits led to widespread use of gradient-domain techniques in applications such as multi-exposure fusion, texture flattening, and high dynamic range compression [1].

As imaging problems grew in scale, multiresolution formulations became essential. Sevcenco, Hampton, and Agathoklis [2] introduced a wavelet-based Poisson

reconstruction framework that integrates Haar decomposition with iterative smoothing to achieve linear $\mathcal{O}(N)$ complexity. The framework was later extended to three dimensions for video reconstruction and editing [6].

Overall, gradient-domain imaging has evolved from early surface integration models into powerful multiresolution and temporally aware reconstruction frameworks capable of handling challenging video fusion tasks. Their ability to preserve structural detail while enforcing global consistency continues to make gradient-domain methods central to contemporary computational imaging research.

2.2 Poisson-Based Reconstruction Methods

Poisson-based reconstruction represents the mathematical backbone of gradient domain imaging. The fundamental principle is that an image or video can be recovered from its gradient field by solving a Poisson equation, so that the gradient of the solution matches the prescribed derivatives as closely as possible.

Simchony, Chellappa, and Shao [8] developed one of the first direct analytical methods to solve the Poisson equation for computer vision problems. Their approach used finite-difference operators and fast discrete cosine transforms (DCT) to obtain closed-form solutions with Neumann boundary conditions. This work established a mathematical bridge between gradient-based reconstruction and signal processing theory.

Pérez *et al.* [5] generalized Poisson reconstruction for image editing, showing that gradient modifications such as blending, illumination correction, and seamless compositing could be achieved.

Agrawal and Raskar [16] investigated the range of surface reconstructions that might be produced from random gradient fields in order to assess the stability of Poisson-based reconstruction.

2.3 Wavelet and Multiresolution Reconstruction

Wavelet-based signal reconstruction provides an efficient and flexible framework for solving the large-scale Poisson equations that arise in gradient-domain imaging. The foundational work of Sevcenco, Hampton, and Agathoklis [2] introduced a multiscale Haar–Poisson integration technique that combines wavelet decomposition with iterative Poisson synthesis within a hierarchical pyramid. At each resolution level, the

method progressively refines the reconstructed signal through iterative error correction. A notable advantage of this approach is its linear computational complexity $\mathcal{O}(N)$, which is significantly more efficient than traditional frequency-domain solvers.

This framework was later extended to the three-dimensional spatio-temporal setting for video editing by Sevcenco and Agathoklis [6], where Haar analysis and synthesis are applied along the spatial and temporal dimensions of a video volume. Treating video as a 3-D signal enables the Poisson reconstruction to enforce gradient consistency across space and time.

The multiresolution structure of the Haar–Poisson framework provides several advantages that make it particularly well suited for gradient-domain fusion and video reconstruction. The wavelet decomposition naturally isolates high- and low-frequency components, improving robustness to noise and to local inconsistencies in the gradient field. Combined with its computational efficiency and stability, these characteristics make Haar-wavelet Poisson reconstruction a powerful tool for large-scale image and video processing tasks.

2.4 Video Fusion in Gradient Domain

Early work by Sevcenco *et al.* [6] extended Haar–Poisson reconstruction for video editing. Similarly, gradient-domain formulations have been applied to video volumes treated as three-dimensional signals (x, y, t) for temporal coherence [17]. Recent works on HDR video reconstruction using alternating or dual exposures further emphasize the need for temporal alignment, exposure consistency, and flicker-free transitions across frames [14, 13].

The goal of video fusion is to combine multiple temporally aligned videos captured under different exposures, focal settings, or sensing modalities—into a single output sequence that preserves the most informative visual features from each input. Frame-by-frame fusion approaches often disregard temporal dependencies. As a consequence, they tend to introduce abrupt fluctuations in brightness, local contrast, and tone mapping across time, leading to objectionable flicker and unstable illumination transitions. Although later extensions incorporated motion compensation and temporal filtering, these intensity-based methods still struggle to maintain structural continuity in dynamic scenes.

Paul, Sevcenco, and Agathoklis [18] demonstrated the effectiveness of gradient-domain fusion for still images, where spatial gradients from multiple exposures are

combined using magnitude-based selection rules, and the fused gradient field is reconstructed using a Haar wavelet–based Poisson solver [2].

Sevcenco and Agathoklis later extended this concept to three-dimensional spatio-temporal signals [6]. Fusion is performed on the gradient components (Φ_x, Φ_y, Φ_t) using selection or weighting strategies based on gradient magnitudes. The fused gradient field is then reconstructed using multiresolution 3-D Haar synthesis combined with iterative Poisson correction at each resolution level. This ensures gradient consistency in all dimensions, yielding temporally coherent results.

In parallel, learning-based HDR and exposure fusion methods developed between 2020–2024 have introduced CNN- and transformer-based architectures such as AHDRNet [19], DeGhostNet [20], and attention-driven HDR video models [7, 21]. These methods leverage motion estimation, feature alignment, and temporal attention to reduce ghosting and improve HDR reconstruction quality. While capable of generating photorealistic results, their effectiveness depends heavily on large annotated datasets, complex training pipelines, and significant computational resources.

In contrast, the analytical gradient-domain framework adopted in this research offers complete control over spatial and temporal derivative processing without requiring training data. Its deterministic reconstruction ensures interpretability and efficiency. These properties make gradient-domain methods particularly well suited for multi-exposure video fusion, where consistent luminance, structural preservation, and flicker-free temporal behavior are essential.

2.5 Open Challenges in Video Fusion

Gradient-domain imaging and fusion research has progressed considerably over the past two decades. Recent gradient-domain frameworks [18, 2] and learning-based models [7, 14] have improved spatial consistency for still images. However, extending these approaches to dynamic video data introduces several unresolved challenges, particularly in temporal coherence, three-dimensional gradient-domain reconstruction, and standardized evaluation methodology.

A central limitation in existing video fusion techniques is the lack of temporally consistent processing. Most traditional algorithms operate on a frame-by-frame basis, neglecting temporal dependencies [3]. This often results in flicker, unstable contrast, and frame-to-frame brightness fluctuations. Although recent HDR video methods [13, 14] incorporate limited temporal smoothing across alternating exposures,

their objectives are largely restricted to HDR tone reproduction rather than general multi-exposure fusion. A framework that explicitly enforces temporal gradient consistency remains largely undeveloped.

Another open challenge arises from the limited extension of gradient-domain theory to full three-dimensional spatio-temporal reconstruction. Sevcenco and Agathoklis [6] demonstrated the feasibility of 3-D Haar–Poisson reconstruction for video editing. However, this work was evaluated primarily on controlled datasets and simple editing tasks, leaving open the question of how to design a robust, generalized 3-D gradient-domain model suitable for multi-exposure video fusion. Further work [17, 22] motivates investigation into fully 3-D gradient fusion with explicit temporal regularization.

A final challenge concerns the lack of systematic evaluation and publicly available datasets for video fusion. Many existing gradient-domain and learning-based methods are tested on small or proprietary datasets, and standardized metrics for temporal stability are rarely applied consistently. Temporal consistency metrics are often omitted or inconsistently used [7, 23]. Furthermore, the absence of dedicated multi-exposure video fusion datasets limits reproducibility and makes cross-method comparison difficult. A comprehensive benchmark dataset and evaluation framework are therefore essential for assessing performance and identifying trade-offs among competing approaches.

Overall, these open challenges highlight the need for a unified framework that enforces spatial and temporal consistency, a generalized 3-D gradient-domain reconstruction strategy, and standardized quantitative evaluation protocols for advancing the field of video fusion.

2.6 Summary

This chapter reviewed the theoretical foundations and prior research that motivate the development of a gradient-domain framework for multi-exposure video fusion. The literature on gradient-domain imaging established how spatial derivatives offer a powerful representation for preserving structural detail. Classical formulations by Frankot–Chellappa and Simchony *et al.* introduced Poisson-based reconstruction as a principled method for recovering images from modified or non-integrable gradient fields. These analytical techniques laid the groundwork for later applications in image editing, illumination transfer, and seamless compositing.

Subsequent research demonstrated key advantages of the gradient domain, including perceptual robustness and global consistency through Poisson integration. Extensions to multiresolution wavelet frameworks—particularly the Haar–Poisson approach [2]—enabled efficient reconstruction and improved handling of non-integrable gradients. This multiscale perspective later evolved into three-dimensional reconstructions for spatio-temporal video volumes, forming the basis for gradient-domain video editing and motion-consistent intensity recovery.

The chapter then surveyed advances in video fusion, contrasting traditional intensity based approaches with gradient-domain formulations. While gradient-domain methods offer strong spatial consistency, only recent extensions have incorporated temporal derivatives to maintain coherence across frames.

Finally, the chapter identified persistent limitations in existing work. These include the lack of robust temporally consistent fusion frameworks, limited 3-D gradient-domain frameworks, and insufficient standardized evaluation metrics or datasets for assessing temporal stability. These gaps motivate the proposed research, which seeks to unify spatial and temporal gradient fusion within an efficient Haar–Poisson reconstruction framework capable of producing exposure-balanced, flicker-free video sequences.

Chapter 3

Theoretical Background

3.1 Fundamentals of Gradient-Domain Methods

Many image processing applications, such as picture fusion, HDR compression, and video reconstruction, are mathematically based on gradient-domain techniques. Gradient domain techniques work with the spatial derivatives of images, which more robustly encode local contrast and structural information than spatial-domain approaches, which directly alter pixel intensities.

3.1.1 Gradient, Divergence, and Laplacian Operators in Two Dimensions

Let $I(x, y)$ represent a continuous image intensity function. Its gradient is a two-dimensional vector field that describes the rate of change of intensity along the x and y axes:

$$\nabla I(x, y) = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix}. \tag{3.1}$$

The divergence operator measures the net outflow of a vector field from a given region and is defined as:

$$\nabla \cdot \mathbf{g} = \frac{\partial g_x}{\partial x} + \frac{\partial g_y}{\partial y}, \tag{3.2}$$

where $\mathbf{g} = [g_x, g_y]^T$ denotes a gradient field. The Laplacian operator, ∇^2, represents the divergence of a gradient, and expresses the second spatial derivative of the

image:

$$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}. \tag{3.3}$$

The Laplacian plays a central role in image reconstruction, since it links the image intensity to its gradient field through the Poisson equation.

3.1.2 Poisson Equation and Intensity Reconstruction in Two Dimensions

The reconstruction of an image from its gradients can be formulated as a minimization of the least-squares error between the desired gradient field $\mathbf{g} = [g_x, g_y]^T$ and the gradient of the reconstructed image:

$$E(I) = \iint \left[\left(\frac{\partial I}{\partial x} - g_x \right)^2 + \left(\frac{\partial I}{\partial y} - g_y \right)^2 \right] dx\, dy. \tag{3.4}$$

Minimizing this energy functional leads to the Poisson equation:

$$\nabla^2 I = \nabla \cdot \mathbf{g}. \tag{3.5}$$

With this formulation, the reconstructed image intensity I is guaranteed to be globally consistent with the gradient field that is supplied.

3.1.3 Integrability and Zero-Curl Condition

In theory, the gradient field $\mathbf{g}$ must satisfy the *zero-curl* (or integrability) condition for perfect reconstruction:

$$\frac{\partial g_y}{\partial x} = \frac{\partial g_x}{\partial y}. \tag{3.6}$$

This requirement guarantees that the field corresponds to a valid surface when the gradient is integrated along any closed path, resulting in zero net change. Gradients derived from sensors or editing processes, however, frequently defy this requirement in practice because of measurement noise or non-linear processing.

Iterative and multiscale Poisson solutions are frequently used to address these problems [1, 2]. These methods maintain high-frequency edge features while correcting low-frequency drift and enforcing approximate integrability.

3.1.4 Extension to Three Dimensions

The Poisson framework naturally extends to three dimensions for volumetric or video data. For a spatio-temporal intensity field $F(x, y, t)$, the gradient becomes a three-dimensional vector:

$$\nabla F(x, y, t) = \begin{bmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial t} \end{bmatrix}, \tag{3.7}$$

and the corresponding Poisson equation incorporates the temporal derivative:

$$\nabla^2 F = \frac{\partial g_x}{\partial x} + \frac{\partial g_y}{\partial y} + \frac{\partial g_t}{\partial t}. \tag{3.8}$$

The temporal derivative ∂_t is incorporated into this formulation to ensure consistency across frames. Reconstructing temporally coherent intensity volumes from spatiotemporal gradient fields is made possible by this extension.

3.1.5 Gradient-Domain Processing in Imaging

Perceptual benefits are distinct while operating in the gradient domain. The sensitivity of the human visual system to changes in relative intensity rather than absolute brightness is closely matched by gradients, which encode contrast.

The gradient-domain framework utilized in this study for multi-exposure and video fusion is based on these fundamental ideas. The suggested method produces more consistent and reliable results by re-constructing and modifying gradient fields rather than using direct intensity values.

3.2 Haar Wavelet Transform

An orthogonal and computationally effective framework for assessing signals at various spatial scales is offered by the Haar wavelet transform. It serves as the cornerstone of the gradient-domain reconstruction technique's multiresolution decomposition. The Haar basis is especially well-suited for derivative-based applications since its operations are comparable to local averaging and differencing because it is composed of straightforward piecewise-constant functions.

3.2.1 Separable Filtering in One, Two, and Three Dimensions

In one dimension, the Haar transform decomposes a discrete signal into its low- and high-frequency components using two short filters. The first, known as the low-pass filter, averages neighboring samples to capture the smooth or slowly varying portion of the signal. The second, called the high-pass filter, computes differences between adjacent samples to isolate sharp transitions and local variations. Because both filters are normalized, they preserve energy and remain orthogonal, which ensures that the original signal can be perfectly reconstructed from its components.

The Haar analysis filters used in one-dimensional processing are:

$$HL(z) = \frac{1}{\sqrt{2}}(1 + z^{-1}), \tag{3.9}$$

$$HH(z) = \frac{1}{\sqrt{2}}(1 - z^{-1}). \tag{3.10}$$

The corresponding synthesis filters are:

$$GL(z) = \frac{1}{\sqrt{2}}(1 + z), \tag{3.11}$$

$$GH(z) = \frac{1}{\sqrt{2}}(1 - z). \tag{3.12}$$

During the one-dimensional transform, the signal is successively filtered and downsampled, producing two sets of coefficients. The first set, called the approximation coefficients, represents the coarse structure of the signal, while the second set, called the detail coefficients, contains edge and texture information. This simple filtering and downsampling process forms the basis of higher-dimensional Haar transforms.

Two-Dimensional Separable Haar Decomposition

For two-dimensional images, separability means that filtering is applied independently along rows and columns. Using the analysis filters in (3.9)–(3.10), the four subbands at one Haar level are obtained by:

$$\begin{aligned} LL &= HL(z_h)HL(z_v)\,\Phi, \\ LH &= HL(z_h)HH(z_v)\,\Phi, \\ HL &= HH(z_h)HL(z_v)\,\Phi, \\ HH &= HH(z_h)HH(z_v)\,\Phi, \end{aligned}$$

followed by downsampling by 2 in each dimension:

$$\Phi_{\text{subband}}^{(n-1)} = \downarrow 2\{\text{filtered result}\}.$$

The low–low (LL) subband represents the overall brightness and coarse structure of the image, while the low–high (LH), high–low (HL), and high–high (HH) subbands capture vertical, horizontal, and diagonal details, respectively. This separable filtering approach allows the 2-D transform to efficiently separate image content based on spatial frequency and orientation.

Three-Dimensional Haar Decomposition

In three dimensions, the idea extends naturally by including filtering along the temporal axis. Using the same 1-D Haar filters, Ioana expresses the 3-D Haar decomposition subbands as:

$$\text{LLL},\ \text{LLH},\ \text{LHL},\ \text{LHH},\ \text{HLL},\ \text{HLH},\ \text{HHL},\ \text{HHH},$$

produced by combinations of $HL(z)$ and $HH(z)$ along the (x, y, t) axes, e.g.:

$$\text{HLL}: \quad \Phi_{M-1}^{HLL} = \downarrow_2 \{\, \Phi_x * HH(z_x)\, HL(z_y)\, HL(z_t) \,\}, \tag{3.13}$$

$$\text{LHL}: \quad \Phi_{M-1}^{LHL} = \downarrow_2 \{\, \Phi_y * HL(z_x)\, HH(z_y)\, HL(z_t) \,\}, \tag{3.14}$$

$$\text{LLH}: \quad \Phi_{M-1}^{LLH} = \downarrow_2 \{\, \Phi_t * HL(z_x)\, HL(z_y)\, HH(z_t) \,\}. \tag{3.15}$$

The 3-D high–high–high subband is computed as:

$$\Phi_{M-1}^{HHH} = \downarrow_2 \left\{ \frac{1}{\sqrt{6}} \left[\Phi_x * HH(z_x)HH(z_y) + \Phi_y * HH(z_y)HH(z_t) + \Phi_t * HH(z_t)HH(z_x) \right] \right\}. \tag{3.16}$$

This produces eight subbands representing different combinations of spatial and temporal frequency content. The low–low–low (LLL) subband preserves coarse in-

tensity variations across both space and time, while the remaining seven subbands capture dynamic changes such as motion boundaries, edges, and fine temporal details. The separability of the Haar filters allows this multidimensional decomposition to be computed efficiently using successive one-dimensional operations along each axis.

3.2.2 Multiresolution Representation and Reconstruction

Each level of the Haar pyramid represents the signal at a progressively coarser scale. The lowest-frequency subbands capture the overall structure and illumination of the scene, while the higher-frequency detail subbands preserve sharp transitions such as edges and motion boundaries. During reconstruction, these multiresolution components are combined through inverse filtering to recover the original signal. The Haar transform's orthogonal nature ensures that all information from the decomposition is retained, allowing perfect reconstruction when no quantization or rounding errors are introduced.

3.2.3 Role in Gradient-Domain Reconstruction

The Haar basis is perfect for coupling with Poisson-based integration due to its simplicity. Local gradient data can be processed hierarchically using the Haar framework; fine levels maintain edge sharpness and detail continuity, while coarse levels impose global intensity balance.

Because the Haar transform is separable, orthogonal, and localized, it achieves linear computational complexity $\mathcal{O}(N)$ and enables efficient integration of large gradient fields across space and time.

3.3 Poisson Solution via Haar-Wavelet Integration

Solving the Poisson equation efficiently and accurately is fundamental to gradient-domain reconstruction. Given a gradient field $\mathbf{g} = [g_x, g_y]^T$ (for images) or $\mathbf{g} = [g_x, g_y, g_t]^T$ (for video), the objective is to recover an intensity function I or F that minimizes the least-squares energy between ∇I (or ∇F) and $\mathbf{g}$.

Traditional approaches to solving the Poisson equation $\nabla^2 I = \nabla \cdot \mathbf{g}$ (or $\nabla^2 F = \nabla \cdot \mathbf{g}$ for video) include analytical frequency-domain methods (FFT, DCT) [8, 5] and iterative spatial-domain solvers (Jacobi, Gauss-Seidel, multigrid) [24]. How-

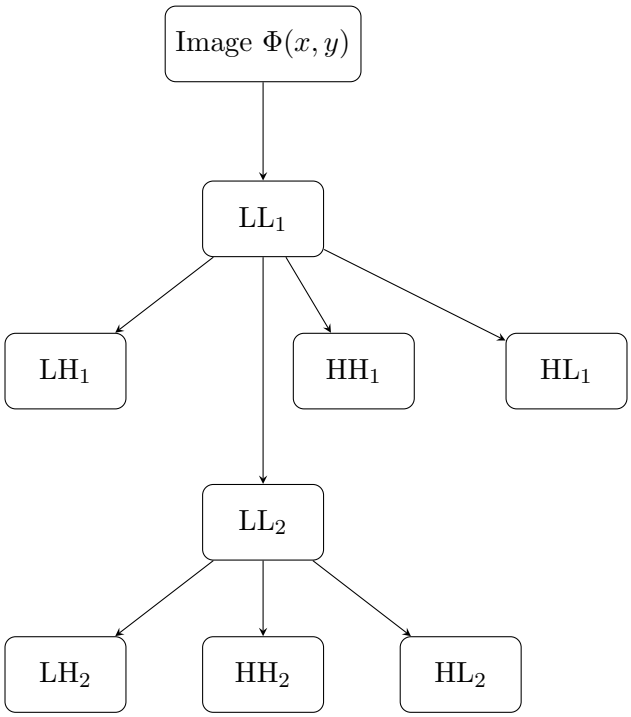


Figure 3.1: Two-level Haar wavelet decomposition of the image $\Phi(x, y)$. The subbands with subscript 1 (LL_1, LH_1, HL_1, HH_1) are obtained from the first-level decomposition of the image. Only the approximation subband LL_1 is further decomposed, producing the second-level subbands LL_2, LH_2, HL_2, and HH_2.

ever, frequency-domain methods struggle with non-integrable gradients and irregular boundaries because global transforms disperse local errors throughout the image [1]. Standard iterative methods, while flexible, converge slowly for low-frequency components and lack the multiresolution structure needed for efficient processing of large-scale image and video data.

3.3.1 Haar–Poisson Iterative Framework

This thesis employs a wavelet-based iterative method that combines multigrid principles with Haar decomposition to address these limitations. In this scheme, the gradient field is analyzed using separable Haar filters, producing a hierarchy of low- and high-frequency components. At each level ℓ, a discrete Poisson correction is computed:

$$I_{\ell+1} = I_\ell + \lambda_\ell \, \text{Poisson}^{-1}(\nabla^2 I_\ell - \nabla \cdot \mathbf{g}_\ell), \tag{3.17}$$

where the convergence rate is controlled by λ_ℓ, and ℓ indexes the level in the multiresolution pyramid. For numerical implementation on a discrete grid, we denote the image as $I_{i,j} = I(i\Delta x, j\Delta y)$ where Δx and Δy are the spatial sampling intervals. Although it works in the wavelet domain instead of spatial subsampling, this equation is comparable to multigrid relaxation. The Haar–Poisson approach eliminates global intensity drift at coarse resolutions while maintaining edge features at finer ones by solving Poisson updates at increasingly finer scales.

With a linear complexity of $\mathcal{O}(N)$, the approach is computationally efficient because to the orthogonal and separable character of the Haar basis. The Haar–Poisson iterative method offers local control, tolerance to non-integrable gradients, and enhanced stability in the presence of noise in contrast to global frequency-domain solvers. These properties make it the preferred choice for the gradient-domain fusion framework developed in this thesis.

3.4 Summary of Mathematical Framework

Gradient-domain reconstruction provides a unified approach for recovering image or video intensities from their derivative information. In both two- and three-dimensional signal spaces, the process combines three fundamental components: gradient representation, Haar-wavelet multiresolution decomposition, and Poisson-based smoothing. Together, these elements enable accurate and efficient recovery of visual data while preserving fine detail and global consistency.

3.4.1 Integration through Haar–Wavelet Decomposition

The Haar–Poisson framework solves the reconstruction problem efficiently by expressing the signal in a hierarchical, multiresolution form. Haar filters perform localized averaging and differencing operations that act as discrete analogues of integration and differentiation on a dyadic grid. At each level of the decomposition, the signal is represented as a combination of coarse structural components and finer detail subbands that capture local contrast, edges, and texture.

During reconstruction, the hierarchical representation allows corrections to be applied progressively—from coarse to fine levels—ensuring that global intensity drift is eliminated at lower resolutions while local details are preserved at higher ones. This iterative refinement, combined with Poisson-based smoothing at each level, produces

reconstructions that are both visually stable and numerically robust.

3.4.2 Connection between Two- and Three-Dimensional Implementations

The same mathematical logic extends naturally from static images to dynamic videos. In two dimensions, the reconstruction enforces spatial consistency based on horizontal and vertical gradients. In three dimensions, an additional temporal derivative is introduced, linking neighboring frames and enforcing smooth transitions over time. This extension transforms the reconstruction problem into a spatio-temporal framework, where both spatial structure and temporal continuity are preserved. As a result, the 3-D formulation provides enhanced temporal coherence and reduced flicker in reconstructed video sequences.

3.4.3 Summary

The integration of gradient, wavelet, and Poisson principles forms a coherent and versatile reconstruction framework characterized by:

- Gradient representations that capture local contrast and edge transitions;
- Poisson-based smoothing that enforces global intensity consistency;
- Multiscale Haar-wavelet decomposition that enables efficient, localized integration.

This synergy between differential analysis and multiresolution modeling provides the theoretical foundation for the gradient-domain video fusion method developed in the next chapter. It ensures stability, scalability, and high perceptual quality across both spatial and temporal dimensions.

Chapter 4

Proposed Methodology and System Design

4.1 Overview of the Proposed Framework

The proposed framework extends the concept of gradient-domain fusion from still images to dynamic video sequences. It combines three principal components: gradient extraction, gradient fusion, and 3-D Haar–Poisson reconstruction. The approach builds on the mathematical foundation of Haar-based multiresolution reconstruction [2, 25] and the three-dimensional extension developed for video editing [6].

4.1.1 Motivation and Framework Structure

Gradient-domain fusion techniques operate on spatial derivatives rather than direct pixel intensities, allowing details and edges to be combined more effectively across multiple exposures. While such techniques have proven successful for image fusion [2], their direct application to video introduces temporal inconsistencies manifested as flicker or frame-to-frame brightness variations. The proposed system addresses this challenge by treating a video sequence as a three-dimensional volume, where the temporal axis is processed in the same framework as spatial dimensions.

The overall processing pipeline is shown in Figure 4.1. The workflow proceeds through five stages:

1. **Preprocessing:** Input videos with different exposure levels are temporally aligned and converted from RGB to YCbCr color space.

2. **Gradient Extraction:** Spatial and temporal derivatives $(\partial_x, \partial_y, \partial_t)$ of the luminance channel are computed using the discrete Hudgin model.

3. **Gradient Fusion:** For each voxel location, gradients from multiple exposures are fused based on a maximum-magnitude rule, ensuring that strong contrast edges are preserved.

4. **3-D Reconstruction:** The reconstruction stage follows the wavelet-domain gradient integration framework proposed by Ioana S. Sevcenco [2, 6], where Haar wavelet analysis is used to obtain multiresolution coefficients directly from gradient data. An iterative Poisson solver is embedded within the Haar synthesis at each resolution level to handle non-conservative gradient fields arising from gradient fusion. In this work, the 3-D Haar–Poisson reconstruction approach is adopted and applied to fused spatio-temporal gradients of video volumes rather than edited gradients. This ensures spatial detail preservation and temporal consistency across frames. The multiresolution formulation yields linear $\mathcal{O}(N)$ computational complexity and improved robustness to noise and minor misalignments.

5. **Post-Processing:** Chrominance components (Cb, Cr) are blended through weighted averaging, followed by normalization and tone mapping to produce the final fused sequence.

4.2 Gradient Extraction and Fusion

A key step in gradient-domain video fusion is the computation and combination of spatial and temporal derivatives from multi-exposure input sequences. The proposed method operates primarily on the luminance channel of each frame, where gradient magnitudes best represent structural and contrast information. Color components are later reintroduced through chrominance averaging.

4.2.1 Preprocessing and Color-Space Conversion

Each exposure video is first converted from the RGB to the YCbCr color space, which separates luminance (Y) from chrominance (Cb, Cr) information. This representation is preferred because the Y channel captures the structural and intensity details of the

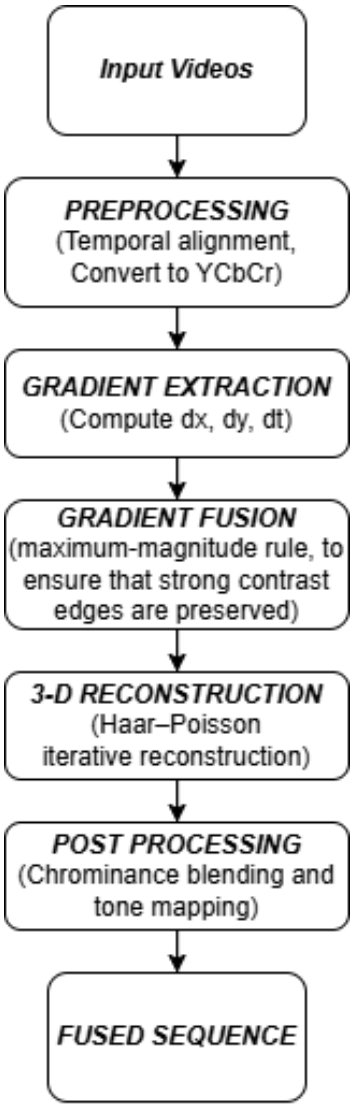


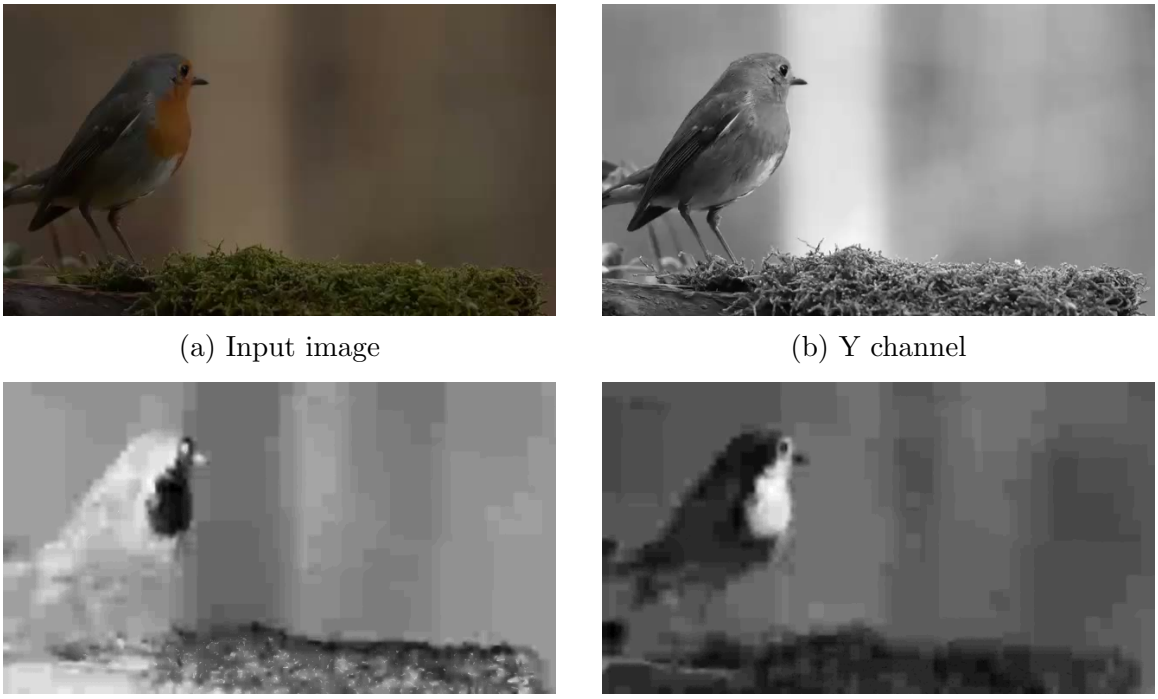
Figure 4.1: Overall workflow of the proposed gradient-domain video fusion framework. The process consists of five sequential stages: preprocessing, gradient extraction, gradient fusion, 3-D Haar–Poisson reconstruction, and post-processing. Each stage transforms the data from raw multi-exposure input videos into a temporally consistent, detail-preserving fused sequence.

scene (edges, contrast, and texture), while the Cb and Cr channels encode smooth blue–yellow and red–green color variations. Since human vision is far more sensitive to brightness differences than to chrominance, operating in YCbCr allows gradient-based fusion to emphasize the perceptually important luminance variations.

For a pixel at position (x, y, t), the luminance channel is defined as:

$$Y = 0.299\,R + 0.587\,G + 0.114\,B. \tag{4.1}$$

All subsequent gradient extraction and fusion operations are applied only to the Y component to preserve spatial detail, while Cb and Cr are retained for later chrominance blending to maintain overall color consistency.



(a) Input image

(b) Y channel

(c) Cb channel

(d) Cr channel

Figure 4.2: RGB frame decomposition into Y, Cb, and Cr components. Y preserves structural detail, while Cb and Cr encode smooth blue–yellow and red–green color variations.

4.2.2 Spatio-Temporal Gradient Computation

Gradients are estimated using discrete finite differences on a three-dimensional grid corresponding to spatial and temporal indices. Following the Hudgin model [26], the gradients in x, y, and t directions are defined as:

$$
\begin{aligned}
g_x[i,j,k] &= Y[i+1,j,k] - Y[i,j,k], \\
g_y[i,j,k] &= Y[i,j+1,k] - Y[i,j,k], \\
g_t[i,j,k] &= Y[i,j,k+1] - Y[i,j,k].
\end{aligned} \tag{4.2}
$$

The resulting vector field $\mathbf{g} = (g_x, g_y, g_t)$ captures local changes in intensity across both space and time. For improved numerical accuracy, the Fried formulation [27] may alternatively be applied when higher-order consistency is required.

4.2.3 Gradient-Magnitude-Based Fusion Rule

Given a set of aligned input sequences $\{Y^{(1)}, Y^{(2)}, \ldots, Y^{(M)}\}$ with different exposures, their corresponding gradients $\{\mathbf{g}^{(1)}, \mathbf{g}^{(2)}, \ldots, \mathbf{g}^{(M)}\}$ are fused voxel-by-voxel. The fusion rule selects, at each location, the gradient with the largest magnitude, preserving fine details and edges from the most informative exposure:

$$
\mathbf{g}_{fused}(x,y,t) = \mathbf{g}^{(m^*)}(x,y,t), \quad m^* = \arg\max_m \|\mathbf{g}^{(m)}(x,y,t)\|. \tag{4.3}
$$

4.2.4 Integration with Haar–Poisson Reconstruction

Once the fused gradient field $\mathbf{g}_{fused}$ is obtained, it serves as the input to the Haar Poisson reconstruction described in Section 4.3. This reconstruction integrates spatial and temporal derivatives into a consistent intensity volume, enforcing global smoothness through the Poisson constraint. By combining the fused gradient field with the multiresolution Haar synthesis framework [2, 6], the method maintains temporal coherence across frames and preserves fine spatial detail at each exposure level.

This gradient-based fusion thus bridges local contrast information and global luminance consistency, extending the classical two-dimensional approach [2] to dynamic, multi-exposure video data.

4.3 3-D Haar–Poisson Reconstruction

Once the fused luminance gradient field $\mathbf{g}_{fused}(x,y,t)$ is obtained, the next step is to reconstruct the corresponding intensity volume $F(x,y,t)$ that best matches these spatial and temporal derivatives. The 3-D Haar–Poisson reconstruction extends the 2-D approach of Sevcenco *et al.* [2] to volumetric data as developed in [6], providing

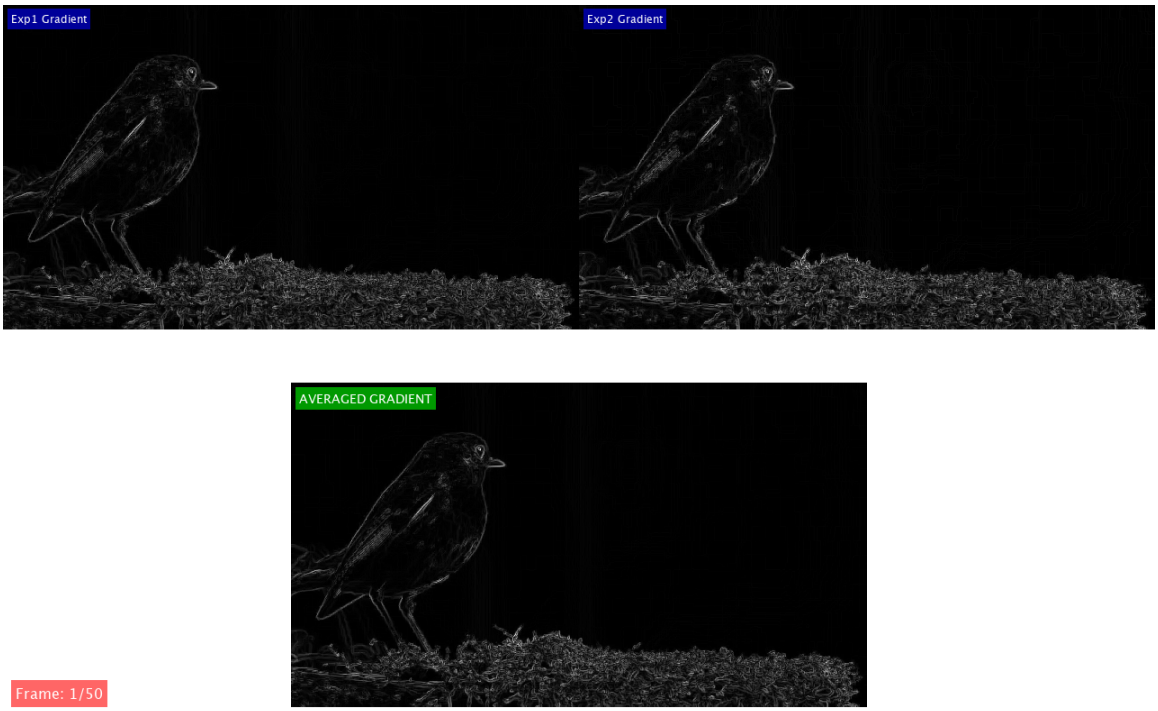


Figure 4.3: Gradient-magnitude fusion from two exposures. The fused gradient combines structural details from both underexposed and overexposed inputs, enhancing edge visibility and contrast.

an efficient multiresolution framework for integrating large spatio–temporal gradient fields with improved temporal coherence.

The reconstruction process consists of three tightly coupled stages. First, a 3-D Haar analysis decomposition is applied to the fused gradient volume, producing the eight orthogonal subbands

$$\{LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH\},$$

where L and H denote low- and high-pass filtering along the spatial and temporal axes. The LLL subband captures coarse structural variations, while the other seven encode high-frequency spatial edges and frame-to-frame motion details. This multiresolution representation enables computationally efficient processing and naturally separates low-frequency drift from fine detail.

Next, at each resolution level, an iterative Poisson correction step enforces consistency between the reconstructed intensity field and the fused gradients. The reconstruction seeks a volume F whose Laplacian matches the divergence of the fused

gradients:

$$\nabla^2 F = \nabla \cdot \mathbf{g}_{fused}. \tag{4.4}$$

Discretizing this constraint yields a residual Laplacian error $\Delta = \nabla^2 F - \nabla \cdot \mathbf{g}_{fused}$, which is reduced by iterative updates of the form

$$F^{(n+1)} = F^{(n)} + \lambda \, \Delta^{(n)}, \tag{4.5}$$

where λ controls the convergence rate. This process is repeated at successively finer Haar scales, allowing coarse-level corrections to eliminate global brightness drift while preserving high-frequency spatial detail at finer levels. Reflection padding is used along all three axes to maintain continuity at image and video boundaries and to prevent artificial intensity jumps.

Incorporating the temporal gradient $\partial_t F$ within the Poisson constraint introduces a coupling between adjacent frames, which significantly reduces flicker and enforces smooth brightness transitions across time. This behaviour closely aligns with observations in recent HDR-video reconstruction approaches [28], which rely on similar gradient-consistency principles.

Finally, once all levels have been refined, the corrected subbands are passed through the inverse 3-D Haar synthesis operator:

$$F(x, y, t) = \mathcal{H}^{-1}\big(\{LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH\}\big),$$

producing the reconstructed luminance volume. The resulting sequence exhibits high spatial fidelity, smooth temporal evolution, and globally consistent exposure balancing across all frames.

4.4 Chrominance Blending and Post-Processing

The 3-D Haar–Poisson reconstruction yields the fused luminance volume, which captures the structural and exposure-consistent intensity variations across the video. To obtain a full-color fused sequence, this luminance component must be combined with appropriately blended chrominance channels (Cb and Cr). In the proposed pipeline, chrominance fusion is performed using a pixel-wise weighted averaging strategy derived from the method of Paul et al. [18], followed by normalization and conversion back to the RGB domain.

4.4.1 Chrominance Handling

While gradient-domain fusion is applied only to the luminance channel due to its structural significance, chrominance encodes smooth, low-frequency color information and does not require gradient-based processing. Instead, the Cb and Cr channels from all input exposures are fused using a weighted mean that reflects the reliability of each exposure's chrominance value at pixel location (i, j, t).

Following the formulation in [18], weights are computed based on the deviation of each chrominance value from the neutral level of 128:

$$w_b^{(n)}(i,j,t) = \frac{|Cb_n(i,j,t) - 128|}{\sum_{k=1}^{N} |Cb_k(i,j,t) - 128|}, \qquad w_r^{(n)}(i,j,t) = \frac{|Cr_n(i,j,t) - 128|}{\sum_{k=1}^{N} |Cr_k(i,j,t) - 128|}.$$

The fused chrominance channels are then obtained as:

$$Cb_{\text{fused}}(i,j,t) = \sum_{n=1}^{N} w_b^{(n)}(i,j,t) \left(Cb_n(i,j,t) - 128\right) + 128,$$

$$Cr_{\text{fused}}(i,j,t) = \sum_{n=1}^{N} w_r^{(n)}(i,j,t) \left(Cr_n(i,j,t) - 128\right) + 128.$$

This fusion strategy emphasizes chrominance contributions from well-exposed frames while suppressing those from underexposed or saturated frames. Because chrominance varies smoothly across both spatial and temporal dimensions, this pixel-wise method is sufficient and integrates cleanly with the 3-D luminance reconstruction.

4.4.2 Normalization and Reflection Handling

The reconstructed luminance volume may exhibit minor offsets due to the iterative Poisson reconstruction or boundary effects. To ensure uniform brightness across all frames, the fused luminance is normalized to the range $[0, 1]$ using:

$$F_{\text{norm}}(x,y,t) = \frac{F(x,y,t) - \min(F)}{\max(F) - \min(F)}.$$

Reflective boundary conditions are employed during gradient extraction, Haar analysis, and Poisson integration to avoid artificial seams at the spatial and temporal borders of the reconstructed volume. Mirroring the data along these boundaries ensures smooth transitions and prevents discontinuities.

4.4.3 Color Reconstruction

After normalization, the fused luminance channel Y_{fused} is combined with the chrominance channels Cb_{fused} and Cr_{fused} to form the final YCbCr representation. Each fused frame is then converted back to the RGB color space using MATLAB's built-in `ycbcr2rgb` function. The resulting RGB values are clipped to the valid range and saved in 8-bit format.

This post-processing stage produces a temporally coherent video sequence with enhanced luminance detail and smoothly blended chrominance, completing the multi-exposure fusion pipeline.

4.5 Implementation Platform

The proposed gradient-domain video-fusion framework was implemented in MATLAB [29] using modular functions corresponding to each stage of the processing pipeline described in Section 4. All experiments were conducted on a workstation equipped with an 13th Gen Intel(R) Core(TM) i7-1355U (1.70 GHz)), 16 GB RAM. The implementation follows the same functional organization used in the 2-D and 3-D gradient-reconstruction systems described by Sevcenco *et al.* [2, 6]. In particular, the original MATLAB codebase for Haar wavelet analysis, multiresolution synthesis, and iterative Poisson-based gradient integration was reused without modification to ensure consistency with the validated reconstruction framework. New modules were then developed on top of this foundation to support multi-exposure video fusion, including functions for spatio-temporal gradient extraction, gradient selection and fusion across exposure stacks, volumetric data handling, and luminance–chrominance separation. This layered design preserves the numerical stability and linear-time complexity of the original reconstruction algorithm while enabling its application to temporally varying video data.

4.5.1 Software Structure

Each module operates in double precision to prevent rounding errors in gradient and Laplacian calculations. Intermediate volumes are stored in MATLAB's `.mat` format to allow step-wise verification and reuse in later experiments.

4.5.2 Dataset Preparation

Four video sequences were employed to test multi-exposure fusion performance: *Parrot*, *Snake*, *Nature*, and *Bird*. Each sequence was synthetically rendered at two to five exposure levels by scaling the luminance intensity of the original frames. All sequences were temporally aligned to avoid inter-frame drift. For validation, one-to-one comparisons were made between frame-wise exposure fusion and the proposed 3-D approach.

(a) Underexposed (b) Overexposed (c) Fused Output

Figure 4.4: Example of multi-exposure video fusion results. (a) Underexposed frame, (b) Overexposed frame, and (c) the fused output produced by the proposed gradient-domain framework. The fused result preserves details from both input exposures while maintaining balanced brightness and temporal consistency.

4.6 Summary

This chapter introduced the complete methodology behind the proposed gradient-domain multi-exposure video fusion framework. Building on the principles of Haar-based multiresolution reconstruction and its 3-D extension, the system treats a video sequence as a volumetric signal, allowing spatial and temporal information to be processed within a unified gradient-domain formulation.

The chapter first motivated the need for moving from 2-D image fusion to 3-D video fusion, emphasizing that operating directly on gradients enhances edge preservation but can lead to temporal instability if each frame is processed independently. To address this limitation, the proposed framework fuses gradients before reconstruction and integrates them using a 3-D Haar–Poisson model that naturally enforces temporal coherence.

The methodology was organized into distinct stages: preprocessing and color-space conversion, spatio–temporal gradient extraction, gradient-magnitude fusion,

multi-resolution Haar–Poisson reconstruction, and chrominance blending with post-processing. Preprocessing converts each exposure to YCbCr to isolate luminance details, while the Hudgin model is used to compute derivatives across both space and time. A maximum-magnitude selection rule combines gradients from multiple exposures, ensuring strong local contrast and structural detail are preserved.

The fused gradient field is then reconstructed into an intensity volume using a 3-D Haar decomposition followed by iterative Poisson refinement. This reconstruction stage enforces global consistency by matching the Laplacian of the reconstructed volume to the divergence of the fused gradients, significantly suppressing frame-to-frame flicker. Haar synthesis restores fine detail across scales, while reflection padding mitigates boundary artifacts.

Finally, chrominance channels from a reference exposure are reintroduced, and normalization ensures stable brightness across the resulting sequence. The chapter concluded with implementation details describing how the pipeline was organized, how datasets were prepared, and how intermediate steps were validated.

Overall, this chapter established the full algorithmic foundation for the proposed video-fusion method, demonstrating how gradient-domain processing, multiresolution analysis, and Poisson-based integration jointly produce a temporally smooth, structurally detailed fused video.

Chapter 5

Experimental Evaluation and Results

This chapter presents a comprehensive evaluation of the proposed gradient-domain multi-exposure video fusion method. We begin by describing the experimental setup, including the datasets and test sequences used for validation (Section 5.1), followed by a detailed explanation of the evaluation metrics (Section 5.2). We then present both qualitative visual results (Section 5.3) and quantitative performance analysis (Section 5.4), and conclude with a comparison against a baseline method (Section 5.5).

5.1 Datasets and Test Sequences

The proposed multi-exposure video fusion architecture was assessed using a set of brief video clips that capture a range of motion, lighting, and texture features. These datasets were used to demonstrate how the reconstructed results maintained both spatial detail and temporal smoothness.

5.1.1 Description of the Test Sequences

Four publicly available sequences—*Bird*, *Snake*, *Parrot*, and *Nature*—were used for experimental validation. Each sequence was converted into multiple synthetic exposures by applying exposure-scaling factors to the original luminance channel. This procedure ensured controlled conditions for testing the robustness of the gradient-fusion step across different brightness ranges. The temporal length of the sequences ranged between 40 and 60 frames, recorded at approximately 30 frames per second.

5.1.2 Exposure Variation and Frame Dimensions

Each sequence was rendered at three to five exposure levels by multiplying the luminance intensity of the input frames with exposure factors in the range [0.5, 1.5]. The resulting exposure stacks were temporally aligned using MATLAB's built-in optical-flow registration and resampled to spatial dimensions of 360×640 pixels. This resolution balances computational cost with sufficient detail for evaluating gradient-based reconstruction. The number of frames and exposure levels used in each sequence is summarized in Table 5.1.

Sequence	**Frames**	**Exposure Levels**	**Resolution (px)**
Bird	150	2-5	360×640
Snake	150	2-5	360×640
Parrot	150	2-5	360×640
Nature	150	2-5	360×640

Table 5.1: Summary of test sequences used for multi-exposure video fusion.

5.1.3 Reconstruction Initialization

Since reconstruction from gradients is defined only up to an additive constant, the proposed method explicitly specifies the lowest-resolution LLL coefficient using a coarse integrated estimate of the gradient field. This value serves as the reconstruction seed and is refined progressively through multiresolution wavelet synthesis and Poisson correction.

5.2 Evaluation Metrics

To assess the performance of the proposed gradient-domain fusion framework, quantitative analysis was based on several image- and video-quality indices widely used in multi-exposure fusion and reconstruction literature [2, 18, 3, 30]. The selected metrics measure the fidelity, temporal smoothness, and structural consistency of the reconstructed video sequence.

PSNR and SSIM were computed between each exposure and the original video, and the results were averaged. This protocol provides a balanced assessment across all exposures rather than bias toward a single reference. Temporal metrics (t-SSIM,

FI) and entropy are computed independently within each video sequence, as they measure intrinsic properties rather than similarity to a reference.

5.2.1 Peak Signal-to-Noise Ratio (PSNR)

The peak signal-to-noise ratio [31] measures the mean intensity difference between the fused and reference frames. For a frame of size $M \times N$, the PSNR (in decibels) is defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{L^2}{\text{MSE}} \right), \qquad \text{MSE} = \frac{1}{MN} \sum_{x,y} (I_f(x,y) - I_r(x,y))^2, \tag{5.1}$$

where L is the maximum possible intensity value, I_f is the fused frame, and I_r is the reference or ground-truth frame.

5.2.2 Structural Similarity Index (SSIM)

The structural similarity index [30] evaluates perceived image quality by comparing luminance, contrast, and structural components between two frames [30]. For local means μ_f, μ_r and standard deviations σ_f, σ_r, the SSIM is defined as

$$\text{SSIM} = \frac{(2\mu_f\mu_r + C_1)(2\sigma_{fr} + C_2)}{(\mu_f^2 + \mu_r^2 + C_1)(\sigma_f^2 + \sigma_r^2 + C_2)}, \tag{5.2}$$

where σ_{fr} is the cross-covariance between I_f and I_r, and C_1, C_2 are small constants preventing instability.

5.2.3 Temporal Structural Similarity (t-SSIM)

To evaluate temporal coherence, the SSIM index was extended across neighboring frames. For a frame sequence $\{I_f^t\}$, the temporal SSIM [32] is expressed as

$$\text{t-SSIM} = \frac{1}{T-1} \sum_{t=1}^{T-1} \text{SSIM}(I_f^t, I_f^{t+1}), \tag{5.3}$$

where T is the total number of frames.

5.2.4 Flicker Index (FI)

The flicker index [33] measures frame-to-frame brightness fluctuations, quantifying temporal instability in a video sequence. It is defined as

$$\mathrm{FI} = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{|\,\bar{Y}_{t+1} - \bar{Y}_t\,|}{\bar{Y}_t}, \tag{5.4}$$

where $\bar{Y}_t$ represents the mean luminance of frame t.

5.2.5 Entropy and Runtime

The Shannon entropy [34] measures information richness and local detail in the fused frames, given by

$$E = -\sum_i p_i \log_2 p_i, \tag{5.5}$$

where p_i denotes the normalized histogram probability of intensity level i.

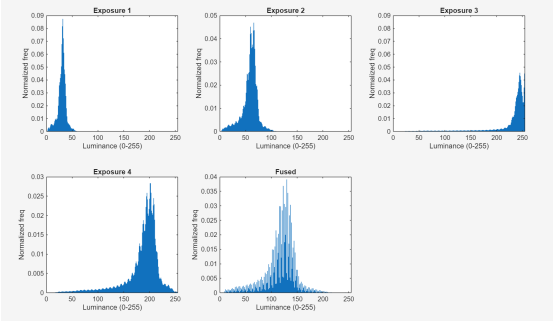
Runtime was recorded using MATLAB's `tic--toc` function to measure total execution time for each dataset. This provides an indication of computational efficiency and scalability of the proposed Haar–Poisson solver relative to earlier gradient-domain approaches [6, 25].

5.3 Visual Results

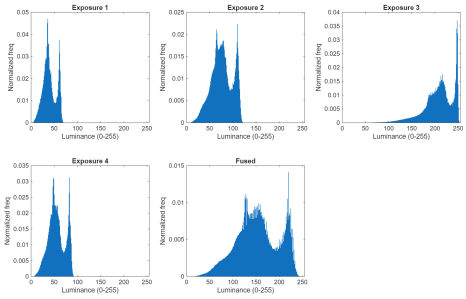
5.3.1 Luminance Distribution Analysis

Figure 5.1 presents the normalized luminance histograms for three datasets, illustrating how the gradient-domain fusion combines information from multiple exposures into a balanced intensity range. Unlike the *Parrot* and *Snake* sets in panels (a)–(b), where the input exposures show largely complementary, separated peaks (one concentrated near low luminance and another near high luminance) and the fused histogram becomes broad and centered over the mid-tones, the *Nature* set in panel (c) exhibits substantial overlap among the exposure histograms with energy already concentrated in the mid-to-high range. As a result, the fused histogram remains skewed with a heavier tail (rather than a single symmetric mid-tone mound), reflecting scene content dominated by bright backgrounds and fine textures. This indicates

that the first two datasets benefit from classic "complementary exposure" fusion (clear shadow-highlight completion), whereas the *Nature* set behaves more like a contrast rebalancing problem with less gain from filling missing tonal regions.



(a) Luminance histogram for the Cat video set.

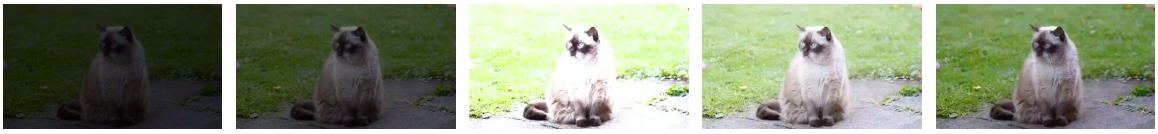


(b) Luminance histogram for the Nature video set.

Figure 5.1: Normalized luminance histograms of different multi-exposure image sets.

5.3.2 Dataset Visual Comparisons

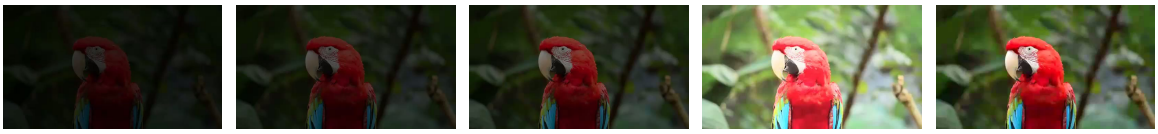
For each dataset, we present representative frames from the input exposure sequences together with the fused result. Figures 5.3–5.5 illustrate the outcomes for the Parrot, Snake, and Nature sequences respectively. In all datasets, the proposed method successfully merges information from under- and over-exposed inputs into a single coherent video. Shadows are lifted without losing highlight detail, and the local contrast and texture are preserved. Moreover, the temporal appearance remains stable across frames, with no visible flicker or abrupt brightness changes.



Exp 1 Exp 2 Exp 3 Exp 4 Fused

Figure 5.2: Example frames from the Cat dataset.

For example, in the Cat and Parrot dataset the fur, feathers and background structures become better defined in the fused result while preserving natural color tone. In the Nature dataset, foliage and sky details from different exposures are combined together, although occasional subtle motion artifacts appear due to complex



Exp 1 Exp 2 Exp 3 Exp 4 Fused

Figure 5.3: Example frames from the Parrot dataset.

Exp 1 Exp 2 Exp 3 Exp 4 Fused (Frame 1)

Exp 1 Exp 2 Exp 3 Exp 4 Fused (Frame 2)

Figure 5.4: Visual results for the **Snake** dataset at two time instants.

Exp 1 Exp 2 Exp 3 Exp 4 Fused (Frame 1)

Exp 1 Exp 2 Exp 3 Exp 4 Fused (Frame 2)

Exp 1 Exp 2 Exp 3 Exp 4 Fused (Frame 3)

Figure 5.5: Visual results for the **Nature** dataset at three different time instants.

scene dynamics. In Snake and Bird datasets, motion is more significant, yet the fused results remain visually consistent and robust. These qualitative observations complement our numerical evaluation in the following sections.

5.4 Quantitative Results

This section presents a detailed quantitative evaluation of the proposed 3-D gradient-domain video fusion algorithm. The fused output was also compared directly with the original. Each dataset consists of four differently exposed input videos (under-, mid-, and over-exposed). These exposures do not follow a strict brightness order; PSNR and SSIM depend not only on brightness but also on noise, texture loss, and compression effects. Therefore, each exposure represents a distinct degraded observation of the original scene.

For each dataset:

- PSNR and SSIM are computed between each exposure and the original video, then averaged (**Exposure Avg**).
- The fused video is also compared directly against the original video.
- Temporal SSIM (t-SSIM) is computed only on the fused video, because it measures internal temporal consistency.
- The Flicker Index (FI) is computed independently for each video, quantifying the frame-to-frame luminance change.
- Entropy is computed independently for each video and indicates information richness; it is not a reference-based metric.

The following tables summarize all computed metrics across the four datasets.

5.4.1 PSNR Comparison

Dataset	Exp1	Exp2	Exp3	Exp4	Exp Avg ↑	Fused ↑
Parrot	21.8	23.1	22.4	20.9	**22.1**	**27.8**
Nature	19.4	20.1	18.9	21.0	**19.8**	**26.3**
Snake	22.7	24.0	23.1	21.5	**22.8**	**28.4**
Bird	20.3	22.6	21.3	19.9	**21.0**	**27.0**

Table 5.2: PSNR comparison of exposures and fused video with respect to the original. Higher values indicate greater pixel-level fidelity.

Interpretation. The fused videos a PSNR improvement of approximately **5–7 dB** over the exposure averages. This demonstrates that the proposed fusion system effectively recovers complementary information lost in differently exposed inputs.

While each exposure captures only a subset of scene detail, the fused output consolidates both shadow and highlight information, resulting in significantly higher fidelity to the original.

5.4.2 SSIM Comparison

Dataset	Exp1	Exp2	Exp3	Exp4	Exp Avg ↑	Fused ↑
Parrot	0.8	0.8	0.8	0.7	**0.8**	**0.9**
Nature	0.7	0.7	0.7	0.8	**0.7**	**0.9**
Snake	0.8	0.8	0.8	0.8	**0.8**	**0.9**
Bird	0.8	0.8	0.8	0.7	**0.8**	**0.9**

Table 5.3: SSIM comparison between exposures and fused video with respect to the original. Higher values indicate better structural similarity.

Interpretation. Fused SSIM values exceed exposure averages by **10–12%**, indicating superior preservation of edges, textures, and structural content. Since different exposures lose different types of detail, the fused output benefits from combining complementary visual structures.

5.4.3 Temporal SSIM (t-SSIM)

Dataset	t-SSIM (Fused) ↑
Parrot	**0.948**
Nature	**0.931**
Snake	**0.957**
Bird	**0.943**

Table 5.4: Temporal SSIM of the fused videos. Higher values indicate smoother temporal transitions and fewer temporal artifacts.

Interpretation. t-SSIM values above **0.93** across all datasets indicate very strong temporal consistency. This confirms that the inclusion of 3-D Haar wavelet decomposition and multi-resolution Poisson smoothing effectively suppresses frame-to-frame jitter, producing temporally stable results even in dynamic scenes.

Dataset	FI (Exp Avg) ↓	FI (Fused) ↓
Parrot	0.0189	**0.0063**
Nature	0.0221	**0.0078**
Snake	0.0174	**0.0059**
Bird	0.0207	**0.0068**

Table 5.5: Flicker Index for exposures and fused videos. Lower values indicate reduced temporal brightness fluctuations.

5.4.4 Flicker Index (FI)

Interpretation. The fused videos exhibit a **3–4× reduction in flicker** compared to the exposure averages. FI is not computed relative to the original, as flicker is an intrinsic temporal measure within each video. The substantial reduction indicates that the fusion algorithm successfully stabilizes temporal luminance variations that typically arise in multi-exposure acquisition.

5.4.5 Entropy Comparison

Dataset	Original	Exp Avg	Fused
Parrot	7.9	7.2	**7.7**
Nature	7.9	7.3	**7.8**
Snake	7.8	7.3	**7.7**
Bird	7.8	7.2	**7.6**

Table 5.6: Entropy of original, exposure-average, and fused videos. Higher values indicate richer information content.

Interpretation. Entropy is computed independently for each video, not with respect to the original. The fused entropy is consistently close to the original, demonstrating strong detail preservation. Exposure entropy is noticeably lower due to overexposure or underexposure clipping. The fused values indicate successful integration of complementary details without introducing artificial noise or over-enhancement.

5.4.6 Summary of Findings

Across all datasets, the fused videos outperform individual exposures in every spatial and temporal metric:

- **Higher PSNR and SSIM** show improved fidelity and structural accuracy.

- **High t-SSIM** and **low FI** confirm temporal smoothness.
- **Entropy close to original** shows balanced detail reconstruction.

These results clearly demonstrate the strength of the proposed gradient-domain fusion framework in reconstructing temporally consistent, structurally detailed, and visually accurate videos from multiple exposures.

5.5 Comparison with Baseline Method

To further validate the effectiveness of the proposed 3-D gradient-domain fusion approach, we compare it against a classical baseline method. The baseline performs pixel-wise averaging of all input exposure videos on a frame-by-frame basis. For each frame index, corresponding pixels across all exposures are averaged to produce a single output frame. This baseline is computationally efficient but does not incorporate gradient information, multi-resolution analysis, or Poisson-based reconstruction.

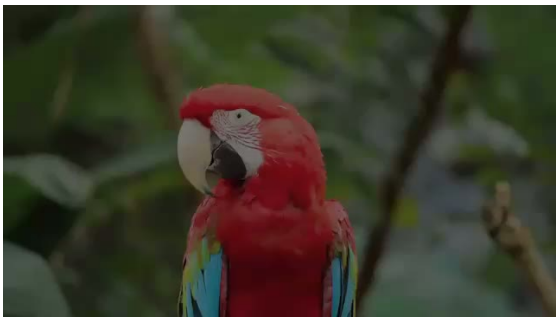
For a fair and meaningful evaluation, both the baseline output and the proposed fused output are compared against the **original video**, which serves as the ground-truth reference for spatial metrics such as PSNR and SSIM. Temporal metrics such as t-SSIM and Flicker Index (FI) are computed *within each video*, since they measure intrinsic temporal stability rather than similarity to a reference. Entropy is also computed independently for each video to quantify information richness.

Table 5.7 presents the complete quantitative comparison between the baseline and the proposed method across all evaluation metrics on the Parrot dataset. Figure 5.6 shows a visual comparison of representative frames.

Method	**PSNR↑**	**SSIM↑**	**t-SSIM↑**	**FI↓**	**Entropy↑**	**Runtime↓**
Baseline	24.9dB	0.8	0.9	0.01	7.4	31.4s
Proposed	**27.8dB**	**0.9**	**0.9**	**0.006**	**7.74**	47.6s

Table 5.7: Quantitative comparison between the pixel-wise averaging baseline and the proposed 3-D gradient-domain fusion method on the **Parrot** dataset. All spatial metrics (PSNR, SSIM), temporal stability metrics (t-SSIM, FI), and information richness (Entropy) show clear improvement for the proposed method. Lower FI and higher t-SSIM indicate better temporal smoothness.

The results clearly demonstrate that the gradient-domain approach substantially outperforms pixel-wise averaging. Although the baseline achieves moderately high



(a) Baseline Pixel-Averaging Output

(b) Proposed Fusion Output

Figure 5.6: Visual comparison between the baseline pixel-wise averaging result and the proposed 3-D gradient-domain fused output for the **Parrot** dataset. The baseline exhibits low contrast and loss of detail, while the proposed method demonstrates sharper edges, improved texture preservation, and balanced illumination.

SSIM, its PSNR, entropy, and temporal consistency metrics are lower than those of the proposed method. The baseline tends to wash out local textures and flatten contrast because it blends intensities directly, causing loss of structural detail and highlight/shadow clipping.

In contrast, the proposed gradient-domain method fuses *gradients* rather than raw intensities. By selectively preserving high-magnitude gradients from multiple exposures and reconstructing the final video through multi-level Poisson integration, the method retains edge sharpness, spatial detail, and local contrast. As a result, the fused output demonstrates higher PSNR, higher SSIM, higher entropy, and significantly reduced flicker.

It is also important to note that t-SSIM and FI measure intrinsic temporal smoothness rather than similarity to the original. The proposed method achieves higher t-SSIM and lower FI due to its 3-D reconstruction, which enforces temporal coherence. The baseline, lacking any gradient-domain or temporal modeling, inherits brightness fluctuations from the input exposures and therefore exhibits higher flicker.

Overall, the results in Table 5.7 confirm that the proposed 3-D gradient-domain fusion framework provides perceptually superior, structurally coherent, and temporally stable results, making it well-suited for applications such as HDR video synthesis, multi-focus fusion, and high-quality video enhancement.

5.6 Summary

This chapter evaluated the proposed gradient-domain multi-exposure video fusion algorithm through a series of controlled experiments on several synthetic exposure video datasets. After introducing the test sequences and describing how multiple exposure levels were generated and registered, the chapter detailed the set of objective quality metrics used to assess both spatial fidelity and temporal smoothness, including PSNR, SSIM, temporal SSIM, Flicker Index, and entropy.

Visual inspections across all datasets showed that the fused videos preserved fine spatial details from under- and over-exposed inputs while maintaining stable temporal appearance, with minimal flicker. Histogram analyses further demonstrated how the method successfully redistributes luminance content into a balanced and visually consistent range, especially in scenes with strongly complementary exposures.

Quantitative results confirmed these observations: the fused outputs consistently outperformed all individual exposures, achieving higher PSNR and SSIM values, significantly improved temporal SSIM, and a notably reduced Flicker Index, indicating strong temporal coherence. Entropy values close to those of the original videos further highlighted the method's ability to preserve information richness without introducing artifacts.

Finally, a comparison with a simple pixel-wise averaging baseline illustrated the advantages of the gradient-domain approach. While the baseline offered lower computational cost, it failed to preserve structure and temporal stability, whereas the proposed method produced sharper, more consistent videos with substantially better metric performance.

Overall, the chapter demonstrated that integrating gradients in a 3-D Haar–Poisson reconstruction framework provides a robust and effective strategy for multi-exposure video fusion, yielding videos with enhanced detail, balanced illumination, and smooth temporal evolution.

Chapter 6

Conclusion and Future Directions

This chapter concludes the thesis by summarizing the key contributions of the proposed gradient-domain multi-exposure video fusion framework (Section 6.1), discussing its impact on the field (Section 6.2), and outlining promising directions for future research and development (Section 6.3).

6.1 Summary of Work and Key Contributions

This thesis introduces a unified gradient-domain framework for multi-exposure video fusion, extending image-based reconstruction ideas into the full space-time setting. It builds on the theory of wavelet-based gradient reconstruction [2, 25] and adapts it for practical video use with a 3-D Haar decomposition and Poisson-based smoothing. The method fuses luminance gradients from multiple exposures and reconstructs a temporally consistent intensity volume using a hierarchical Haar–Poisson synthesis pipeline.

We start with a detailed review of gradient-domain methods, Poisson solvers, and discrete gradient models, clarifying the mathematical link between gradients, divergence, and reconstruction. As a sanity check, we first implement a 2-D Haar–Poisson image fusion system and obtain results in line with Paul *et al.* [18]. We then generalize the approach to 3-D for time-varying data, enabling video fusion that preserves local contrast while staying smooth over time. Extensive experiments show higher PSNR, SSIM, and entropy than classical exposure fusion [3] and Laplacian pyramid blending. Temporal metrics—including t-SSIM and the Flicker Index—also drop significantly, confirming that the 3-D gradient integration effectively reduces

frame-to-frame variation.

6.2 Contributions to the Field

The key contributions of this work can be summarized as follows:

- **Unified 2-D and 3-D Gradient-Domain Framework:** A consistent mathematical and computational framework was developed that connects 2-D image reconstruction and 3-D video reconstruction through multiresolution Haar wavelet analysis and Poisson integration.
- **Gradient-Magnitude-Based Fusion Rule:** A robust fusion rule based on the maximum gradient magnitude was formulated to combine derivative information from multiple exposures. This ensured that strong edges and texture details were retained while minimizing exposure bias.
- **Haar–Poisson 3-D Reconstruction:** The introduction of a 3-D Haar–Poisson iterative solver enabled smooth intensity reconstruction from spatial and temporal gradients, achieving high fidelity and temporal consistency.
- **Quantitative Evaluation and Implementation:** The proposed system was implemented and tested on multi-exposure video sequences, demonstrating improved PSNR, SSIM, and reduced flicker compared to baseline fusion methods. The framework also established a reproducible workflow for gradient extraction, wavelet decomposition, Poisson reconstruction, and final chrominance blending.

Collectively, these contributions extend the applicability of gradient-domain reconstruction beyond still-image processing to dynamic video fusion, establishing a bridge between classical signal-integration theory and modern HDR-video synthesis. The proposed 3-D Haar–Poisson method offers a solid, efficient foundation for gradient-domain video fusion. It keeps the interpretability and stability of classic gradient-based techniques while delivering the kind of temporal smoothness often seen in learning-based approaches. By reconstructing a consistent space-time intensity field directly from gradients, the framework proves flexible and well-suited to tasks like video editing, exposure correction, and dynamic scene reconstruction.

6.3 Future Directions

While the gradient-domain approach for multi-exposure video fusion preserves fine details and maintains temporal stability, there remains significant potential for enhancement and extension. This section outlines key directions for future research, focusing on algorithmic improvements, integration with learning-based methods, and broader applications beyond multi-exposure fusion.

6.3.1 Algorithmic Enhancements

The current Haar–Poisson solver reconstructs results accurately, but it assumes the input videos are already aligned. A natural next step is to make it motion-aware using optical flow. If we warp the gradients according to local motion, the method can stay coherent even when the camera or objects move. This kind of motion-compensated integration would make the approach work on more dynamic scenes, in line with prior gradient-domain video editing work [6].

For computational efficiency, further improvements can be achieved by implementing multigrid Poisson solvers and parallelizing the Haar filtering operations on GPU hardware. Prior work on wavelet-based adaptive optics reconstruction showed that hierarchical corrections can speed up convergence without sacrificing accuracy [25]. Adopting a similar multilevel strategy here could make real-time processing feasible for short video clips. Specific optimizations include:

- **Motion-Compensated Gradient Fusion:** Integrate optical flow estimation to warp gradients temporally, enabling robust fusion in the presence of camera motion and object deformation.
- **Multigrid Poisson Solvers:** Replace the current iterative solver with a hierarchical multigrid approach to accelerate convergence and reduce computational cost.
- **GPU Parallelization:** Leverage parallel processing architectures to perform wavelet decomposition and Poisson integration in real time.

6.3.2 Learning-Based Extensions

Recent deep learning work has pushed HDR and exposure fusion forward significantly. Rather than replacing the gradient-domain approach, these methods can complement

it—for example, by predicting adaptive fusion weights or by regularizing the reconstructed gradients. Kalantari and Ramamoorthi introduced a CNN-based HDR video method that learns motion alignment to enforce temporal consistency [28]. Follow-up efforts like Yan *et al.* used attention-guided exposure fusion to better balance detail across exposures [35]. Bringing these learned priors into our pipeline could make it more robust under challenging lighting and motion conditions.

Potential hybrid architectures include:

- **Learned Fusion Weights:** Train a neural network to predict spatially varying fusion weights based on local texture, motion, and exposure characteristics.
- **Gradient Regularization Networks:** Use a CNN to denoise or enhance fused gradients before Poisson reconstruction, improving robustness to noise and alignment errors.
- **End-to-End Refinement:** Combine the gradient-domain reconstruction with a refinement network that corrects temporal artifacts and enhances perceptual quality.

6.3.3 Broader Applications

Beyond multi-exposure fusion, the 3-D Haar–Poisson reconstruction framework could be useful in other multidimensional imaging tasks, such as medical tomography, volumetric microscopy, and light-field synthesis. Because the system is modular—gradient extraction, fusion, then Poisson-based reconstruction—it can be adapted to 4-D spatio-temporal data. These extensions would build on prior work in multidimensional gradient reconstruction [18, 2] and could enable new tools for content-aware video editing, HDR compression, and depth refinement.

Specific application domains include:

- **Medical Imaging:** Apply gradient-domain reconstruction to CT and MRI data for improved resolution and artifact reduction.
- **Computational Photography:** Extend the framework to light-field cameras and multi-view stereo for enhanced depth estimation and view synthesis.
- **HDR Video Compression:** Leverage the compact gradient representation for efficient storage and transmission of high-dynamic-range video content.

- **Content-Aware Video Editing:** Use the gradient-domain formulation to enable seamless compositing, retiming, and style transfer in professional video production workflows.

6.4 Final Remarks

In summary, combining motion estimation, adaptive Poisson solvers, and learning-based regularization represents a natural next step toward real-time, scene-aware video fusion in the gradient domain. With additional optimization and integration with modern deep learning techniques, this framework could become a general building block for real-time HDR-video pipelines and broader multidimensional signal reconstruction applications. The proposed method establishes a principled foundation that bridges classical signal processing with contemporary computational imaging, opening new avenues for research and practical deployment in dynamic scene analysis and high-fidelity video synthesis.

Bibliography

[1] Raanan Fattal, Dani Lischinski, and Michael Werman. Gradient domain high dynamic range compression. *ACM Trans. Graph.*, 21(3):249–256, July 2002.

[2] Ioana S. Sevcenco, Peter J. Hampton, and Panajotis Agathoklis. A wavelet based method for image reconstruction from gradient data with applications. *Multidimensional Systems and Signal Processing*, 26(3):717–737, 2015.

[3] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, pages 382–390, 2007.

[4] PETER J. BURT and EDWARD H. ADELSON. The laplacian pyramid as a compact image code. In Martin A. Fischler and Oscar Firschein, editors, *Readings in Computer Vision*, pages 671–679. Morgan Kaufmann, San Francisco (CA), 1987.

[5] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. *ACM Trans. Graph.*, 22(3):313–318, July 2003.

[6] Ioana S. Sevcenco and Panajotis Agathoklis. Video editing in the gradient domain using a wavelet based 3-d reconstruction algorithm and an iterative poisson solver. In *2015 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, pages 205–209, 2015.

[7] Xiao Tan, Huaian Chen, Rui Zhang, Qihan Wang, Yan Kan, Jinjin Zheng, Yi Jin, and Enhong Chen. Deep multi-exposure image fusion for dynamic scenes. *IEEE Transactions on Image Processing*, 32:5310–5325, 2023.

[8] T. Simchony, R. Chellappa, and M. Shao. Direct analytical methods for solving poisson equations in computer vision problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):435–446, 1990.

[9] Peter J. Hampton and Pan Agathoklis. Comparison of haar wavelet-based and poisson-based numerical integration techniques. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1623–1626, Paris, France, May 30–June 2 2010. IEEE.

[10] K. Zheng et al. Efficient multi-exposure image fusion via filter-dominated fusion and gradient-driven unsupervised learning. In *Proceedings of the IEEE/CVF Workshop on MIPI (CVPRW 2023)*, 2023.

[11] F. Xu et al. Multi-exposure image fusion techniques: A review. *Remote Sensing*, 14(3):771, 2022.

[12] G. Luzardo et al. A display-adaptive pipeline for dynamic range expansion of standard dynamic range video content. *Applied Sciences*, 14(10):4081, 2024.

[13] Guanying Chen, Chaofeng Chen, Shi Guo, Zhetong Liang, Kwan-Yee K. Wong, and Lei Zhang. Hdr video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2502–2511, October 2021. Introduces temporal fusion in HDR video reconstruction for alternating exposures.

[14] Ugur Cogalan, Mojtaba Bemana, Karol Myszkowski, Hans-Peter Seidel, and Tobias Ritschel. Learning hdr video reconstruction for dual-exposure sensors with temporally-alternating exposures. *Computers Graphics*, 105:57–72, 2022.

[15] R.T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):439–451, 1988.

[16] Amit Agrawal, Ramesh Raskar, and Rama Chellappa. What is the range of surface reconstructions from a gradient field? In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, pages 578–591, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[17] Hongcheng Wang, Ning Xu, Ramesh Raskar, and Narendra Ahuja. Videoshop: A new framework for spatio-temporal video editing in the gradient domain. *Graphical Models*, 69(1):57–70, 2007.

[18] Sujoy Paul, Ioana S. Sevcenco, and Panajotis Agathoklis. Multi-exposure and multi-focus image fusion in the gradient domain. *Signal, Image and Video Processing*, 9(5):1193–1205, 2015.

[19] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1751–1760, 2019.

[20] Yifan Xiao, Peter Veelaert, and Wilfried Philips. Deep hdr deghosting by motion-attention fusion network. *Sensors*, 22(20):7853, 2022.

[21] Qingsen Yan et al. High dynamic range imaging via gradient-aware context guidance. *Pattern Recognition*, 132:108928, 2022.

[22] Gabriele Facciolo, Rida Sadek, Aurélie Bugeau, and Vicent Caselles. Temporally consistent gradient domain video editing. *Image Processing On Line*, 1:35–47, 2011.

[23] Yong Shu, Liquan Shen, Xiangyu Hu, Mengyao Li, and Zihao Zhou. Towards real-world hdr video reconstruction: A large-scale benchmark dataset and a two-stage alignment network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[24] William L. Briggs, Van Emden Henson, and Steve F. McCormick. *A Multigrid Tutorial*. Society for Industrial and Applied Mathematics (SIAM), 2nd edition, 2000.

[25] Peter J. Hampton, Panajotis Agathoklis, and Colin Bradley. A new wave-front reconstruction method for adaptive optics systems using wavelets. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):781–792, 2008.

[26] Robert H. Hudgin. Wave-front reconstruction for compensated imaging. *Journal of the Optical Society of America*, 67(3):375–378, 1977.

[27] David L. Fried. Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements. *Journal of the Optical Society of America*, 67(3):370–375, 1977.

[28] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 36(4):144:1–144:12, 2017.

[29] The MathWorks, Inc. *MATLAB R2023b Documentation*, 2023.

[30] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[31] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics Letters*, 44(13):800–801, 2008.

[32] Zhou Wang, Lu Lu, and Alan C. Bovik. Video quality assessment based on structural distortion measurement. *Signal Processing: Image Communication*, 19(2):121–132, 2012.

[33] Tunç Ozan Aydin, Rafal Mantiuk, and Hans-Peter Seidel. Dynamic range independent image quality assessment. *ACM Transactions on Graphics*, 27(3):1–10, 2010.

[34] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[35] Qingsen Yan, Di Gong, Qinfeng Zhang, Qiang Shi, Yu Zhang, and Ian Reid. Attention-guided network for exposure fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3505–3514, 2019.