Efficient Multiview Image Compression Using Disparity Estimation

Zi Jie Xie Xiaotao Wu

Github Link: https://github.com/DuckWu/EECE-5698-Final-Project

Abstract

This project investigates a rate-distortion optimization (RDO) framework applied to disparity maps obtained from stereo image pairs. By treating disparity map processing as a compression problem, we quantize local regions (blocks) and select the best representation to minimize a Lagrangian cost function that balances distortion against coding rate. We incorporate entropy-based measures to approximate coding costs and adapt parameters such as the quantization step size and Lagrange multiplier. Experiments on a stereo dataset with available ground truth show that the proposed RDO approach yields improved quality—measured by Peak Signal-to-Noise Ratio (PSNR)—over a baseline non-optimized disparity map. These results highlight the potential of RDO for efficient storage, transmission, and further processing of disparity information.

1 Introduction

Stereo imaging systems capture the same scene from two slightly different viewpoints, enabling the estimation of disparity maps that relate corresponding pixels between views. Disparity maps are crucial for many computer vision tasks, including 3D reconstruction, depth estimation, and object recognition. However, raw disparity maps can be large and contain noise or inconsistencies.

While post-processing techniques can smooth or refine disparity maps, their focus typically lies on improving perceptual quality rather than explicitly balancing fidelity against bitrate or storage costs. Inspired by compression theory and video coding standards, this project introduces a rate-distortion optimization (RDO) paradigm into disparity map refinement. By considering both distortion (e.g., mean squared error with respect to ground truth) and a rate-like measure (approximated through entropy), we find an optimal quantized representation for each block of the disparity map. This approach can be viewed as a step towards more efficient coding of disparity maps, beneficial in scenarios where they must be stored, transmitted, or integrated into bandwidth-constrained systems.

The primary contributions of this project include:

- Implemented a block-based RDO framework on disparity maps to achieve a better quality-compression trade-off.
- Evaluated the performance on a known stereo dataset with ground truth disparity, demonstrating improvements in PSNR after optimization.
- Investigated the impact of parameters like the quantization steps and Lagrange multiplier (λ) on the achieved rate-distortion balance.

Through these contributions, this project advances the field of multi-view image compression by addressing key challenges and leveraging disparity information for enhanced performance.

2 Related Work

- **Rate-distortion optimization for disparity map compression**

  Approaches like the one presented by Lee and Ho in [3] have focused on formulating disparity map compression as a rate-distortion optimization (RDO) problem. These methods treat the disparity map as a signal to be compressed, aiming to minimize distortion relative to a ground truth reference while reducing the coding rate. RDO-based techniques often rely on block-based operations, quantization, and entropy-based coding decisions, which are closely aligned with the approach used in our project. However, a major limitation of such methods is their reliance on a well-estimated and noise-free disparity input. When disparity maps contain inaccuracies, noise, or uncertainty (such as along depth boundaries or in occlusion regions), the RDO process may produce less effective results. Additionally, the computational complexity and the need for careful parameter tuning—like selecting appropriate quantization steps and adjusting the Lagrange multiplier—pose significant challenges. While the rate-distortion optimization framework proposed by Lee and Ho has inspired parts of our work, future efforts will need to focus on improving robustness to estimation errors and addressing complexity issues to make it more practical for real-world applications.

- **Efficient Depth Map Coding for 3D Video Systems Using Multi-View and Epipolar Constraints**

  In their work, Hu et al. [4] look at ways to make depth map compression more efficient by using the geometric relationships between multiple camera views. They rely on epipolar constraints—basically the straight-line connections between matching points in different camera views—to help align and compress depth information more effectively. This reduces the amount of repeated or unnecessary data stored across multiple views.

  However, one major drawback is that this approach depends heavily on having accurate camera calibration and a solid understanding of the scene's geometry. If the calibration is off or if the geometric relationships aren't stable, the compression efficiency might not improve as much as expected. Also, putting this method into existing video pipelines can be tricky and might require more processing power. While the idea shows that understanding the scene's geometry can lead to better depth map compression, making the method robust and easy to integrate into various setups still needs more work.

3 Background

The increasing prevalence of stereo and multi-view imaging systems has brought attention to the challenges of efficiently handling the large amounts of data they produce. Such systems capture scenes from multiple vantage points, enabling depth perception and three-dimensional analysis. A central component in this process is the concept of a *disparity map*, which encodes the relative shift of corresponding points between stereo image pairs. By determining how much a feature in one image is displaced in the other, the disparity map effectively encodes depth information, enabling 3D reconstruction and scene understanding.

3.1 Disparity Estimation and Representation

Disparity estimation typically involves matching features across rectified stereo image pairs. Rectification ensures that corresponding points lie on the same horizontal scanline, simplifying the search for matches. Various algorithms exist for this step, ranging from classical methods—such as block matching and semi-global matching—to advanced techniques utilizing machine learning or deep neural networks. The resulting disparity map assigns a disparity value to each pixel, with larger values indicating closer objects in the scene.

Once computed, disparity maps serve as crucial inputs for tasks like 3D modeling, object recognition, and visual navigation. However, these maps can be large and noisy, requiring substantial storage or bandwidth if transmitted. Hence, effectively representing and refining disparity data remains an essential problem.

3.2 Rate-Distortion Concepts

In image and video compression, the process of reducing data size while maintaining acceptable quality is guided by the principles of *rate-distortion optimization (RDO)*. The concept stems from information theory and is applied in modern codecs, where encoding decisions are made to minimize distortion at a given bitrate or, equivalently, to achieve the best possible quality for a desired compression level. The *distortion* typically measures how far the reconstructed data deviates from a reference, while the *rate* reflects how many bits are required to store or transmit the data.

Balancing rate and distortion involves tuning parameters—such as quantization levels or coding modes—to achieve an optimal trade-off. A higher compression ratio may come at the cost of visual quality, whereas striving for negligible quality loss may yield larger file sizes. By treating disparity maps as signals to be compressed, these RDO principles can be applied to choose the best representation for each region, ensuring efficient and high-quality storage or transmission.

3.3 Applying RDO to Disparity Maps

Adopting RDO in the context of disparity maps means going beyond simple smoothing or filtering. Instead, the disparity map is considered as a source that can be encoded using a Lagrangian cost function:

$$J = D + \lambda R,$$

where D is the distortion relative to a ground truth map, R is an estimate of the coding rate, and λ is a Lagrange multiplier that controls the balance between quality and compressibility. This formulation allows the encoder to make informed decisions about quantization and representation at the block level, potentially improving both the fidelity and the storage efficiency of the resulting disparity map.

Such an approach is not only beneficial for compression; it can also enhance subsequent tasks that rely on accurate and consistent disparity data. High-quality, compact representations of disparity maps facilitate faster transmission and more efficient downstream processing, making RDO-based techniques a promising avenue for improving the overall utility of stereo and multi-view imaging systems.

4 Methodology

This section outlines the steps taken to compute, optimize, and evaluate the disparity maps using a rate-distortion optimization (RDO) framework. By carefully balancing the fidelity of the disparity values against an entropy-based measure of complexity, we aim to produce representations that are both visually coherent and more compact for storage or transmission.

4.1 Disparity Map Computation

The process begins with a pair of stereo images, denoted as $I_{\text{left}}(x, y)$ and $I_{\text{right}}(x, y)$.

Figure 1: left

Figure 2: right

In order to compute an accurate disparity map, it is essential to ensure that the stereo images are rectified so that corresponding points appear on the same horizontal lines. The rectification process uses the intrinsic and extrinsic camera parameters from the calibration data.

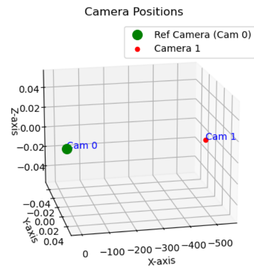

Figure 3: Camera Position

Using Semi-Global Block Matching algorithm, we compute a preliminary disparity map $D(x, y)$:

$$D(x, y) = \arg\min_{d} \ C(x, y, d),$$

where $C(x, y, d)$ is a cost function measuring the dissimilarity between a patch centered at (x, y) in the left image and a corresponding patch at $(x-d, y)$ in the right image. After optimization via semi-global methods, this yields an initial disparity map which may contain noise or inaccuracies, especially in regions of low texture or near object boundaries.

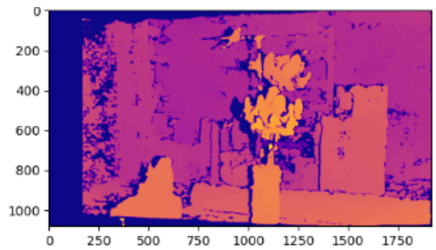

Figure 4: Disparity Map

4.2 Block-Based Quantization and Rate-Distortion Framework

To apply RDO, we first partition the raw disparity map into non-overlapping blocks of size $B \times B$ (e.g., 16×16 pixels). For each block D_b, we consider a set of quantization step sizes $\mathcal{Q} = \{q_1, q_2, \ldots, q_n\}$, each of which maps the original disparity values to a coarser set of levels. Given a quantization step $q \in \mathcal{Q}$, we obtain a quantized block $D_b^{(q)}$:

$$D_b^{(q)}(x, y) = \text{round}\left(\frac{D_b(x, y)}{q}\right) \cdot q.$$

Larger q values lead to fewer distinct disparity levels and thus potentially lower complexity, but may introduce greater distortion.

4.3 Distortion and Rate Measures

To judge the quality of a quantized block, we define a distortion measure based on the Mean Squared Error (MSE) relative to the ground truth disparity map $D_{\text{gt}}(x, y)$:

$$D = \text{MSE}(D_b^{(q)}, D_{\text{gt}}) = \frac{1}{B^2} \sum_{(x,y)\in b} [D_b^{(q)}(x, y) - D_{\text{gt}}(x, y)]^2.$$

Next, we need a rate measure R that approximates how efficiently we can encode the quantized block. Instead of performing a full encoding, we compute the block's entropy as a proxy for rate. Let $p(v)$ represent the normalized frequency of a quantized disparity value v in the block. The entropy H is defined as:

$$H = -\sum_v p(v) \log_2 p(v).$$

This entropy serves as a stand-in for coding complexity (higher entropy suggests more bits needed). We can use $R = H$ as our rate metric.

4.4 Rate-Distortion Optimization

We combine D and R into a Lagrangian cost function:

$$J(q) = D + \lambda R,$$

where λ is the Lagrange multiplier that controls the trade-off between preserving detail (low D) and reducing complexity (low R). For each block, we choose the quantization step $\hat{q}$ that minimizes the cost:

$$\hat{q} = \arg\min_{q\in\mathcal{Q}} [D + \lambda R].$$

Selecting $\hat{q}$ block by block yields an optimized disparity map that, on average, should be closer to the ground truth at a lower entropy level than the non-optimized version.

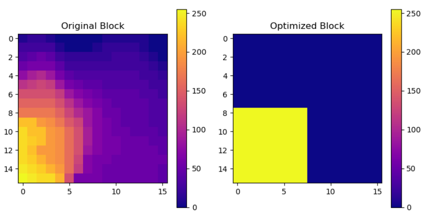

Figure 5: Block Comparison

4.5 Parameter Selection

The choice of block size B, quantization candidates $\mathcal{Q}$, and λ can significantly affect results:

- **Block Size (B):** Larger blocks may smooth out local details but capture global structure. Smaller blocks allow finer control, especially in areas with complex disparity gradients.
- **Quantization Steps ($\mathcal{Q}$):** A diverse set of quantization steps (e.g., $\{1, 2, 4, 8\}$) provides flexibility. Too coarse a step might oversimplify disparities, while too fine a step might fail to reduce entropy meaningfully.
- **Lagrange Multiplier (λ):** Increasing λ places greater emphasis on lowering entropy. A higher λ leads to a smoother, more compressible disparity map, potentially at the cost of accuracy.

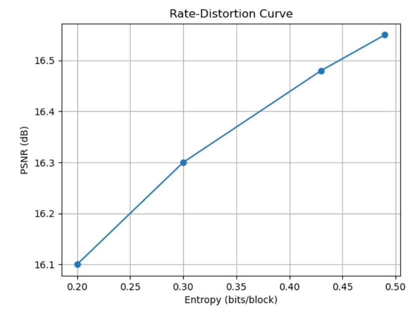

Figure 6: RD Curve

4.6 Implementation Details and Execution

For implementation, we rely on Python, OpenCV, and NumPy. The workflow proceeds as follows:

1. **Disparity Computation:** Use OpenCV's StereoSGBM to obtain $D(x, y)$ from the stereo images.
2. **Block Partitioning:** Divide $D(x, y)$ into $B \times B$ blocks.
3. **Quantization and Evaluation:** For each block, quantize using each $q \in \mathcal{Q}$, compute D, R, and $J(q)$.
4. **Mode Selection:** Choose $\hat{q}$ that minimizes $J(q)$ for the block.
5. **Reconstruction:** Assemble the optimized blocks into the final disparity map.

This approach leverages established rate-distortion concepts to refine disparity maps, potentially enabling more efficient storage and transmission in scenarios where bandwidth or memory is limited.

5 Experiments/Results

We evaluated our RDO approach using a stereo dataset with known ground truth. Starting from an initial disparity map obtained via StereoSGBM, we applied our RDO framework, tuning parameters like the Lagrange multiplier (λ) and the quantization steps.

The final comparison figure compares the ground truth disparity, the original map, and the RDO-optimized result. While the original map shows noise and artifacts, the optimized version appears smoother and slightly more uniform. Quantitatively, the original map achieved about 16.48 dB PSNR, while the optimized map reached approximately 16.57 dB. Although modest, this improvement demonstrates that RDO can yield measurable gains in fidelity.

Examining the rate-distortion trade-off revealed the expected trend: increasing entropy improved PSNR. This indicates that the optimized disparity map retained more detail with higher entropy, leading to better fidelity to the ground truth. The results demonstrate the efficacy of the rate-distortion framework in balancing compression and quality. Future work could focus on exploring more refined quantization steps, dynamic λ adjustment, or adaptive strategies to further enhance the rate-distortion trade-off and achieve greater flexibility in optimization.

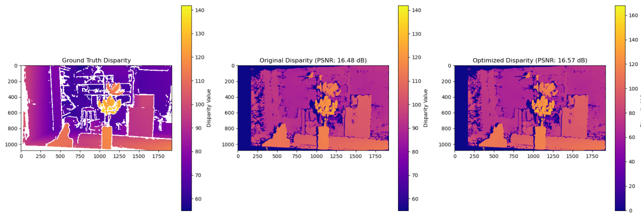

Figure 7: Final Comparison

6 Challenges

During the development and evaluation of the RDO-based disparity optimization framework, we encountered several technical challenges that influenced both the implementation details and the resulting image quality.

6.1 Parameter Sensitivity

One of the primary difficulties was selecting appropriate values for the Lagrange multiplier (λ) and defining suitable quantization steps. Initial experiments with coarse quantization levels and high λ values resulted in overly smoothed disparity blocks, while finer steps and lower λ settings provided more detail but diminished rate savings. To address this issue, iterative experimentation was conducted, gradually refining λ and introducing an expanded set of quantization candidates. This iterative approach helped balance the trade-off between minimizing distortion and reducing entropy.

6.2 Handling Irregular Blocks

Regions near edges and occlusions presented significant challenges due to their complex disparity patterns, which did not quantize neatly. To manage this, smaller block sizes were introduced, allowing the optimization to focus on localized complexity. This adjustment helped preserve structural details in challenging areas while still enabling RDO to smooth out and compress planar or uniform regions effectively.

6.3 Summary

These solutions not only facilitated the current study but also provide a foundation for future enhancements to the RDO approach for disparity maps.

7 Conclusion/Future Improvement

In this project, we explored the application of rate-distortion optimization (RDO) principles to the processing of disparity maps derived from stereo images. By formulating the optimization as a balance between minimizing distortion and reducing coding entropy, we managed to enhance the overall quality of the disparity map. The optimized map exhibited smoother regions and reduced high-frequency noise, especially in more uniform, planar areas. This reduction in unnecessary detail not only improved the measured PSNR but also holds promise for more efficient compression and transmission, as a cleaner disparity representation can be encoded with fewer bits.

However, the improvements observed are influenced significantly by parameter selection. Adjusting the Lagrange multiplier (λ) and refining the set of quantization steps are key levers that can produce more substantial gains. Striking the right balance here is critical—finer quantization steps may yield better fidelity but at the expense of increased computational complexity. Moreover, certain challenging regions, such as textured boundaries, occlusions, and areas with complex depth gradients, remain difficult to optimize. This indicates that while the current approach shows potential, it may benefit from integrating additional spatial context or more sophisticated quantization strategies to handle diverse scene content.

For future work, several avenues can be explored. Adaptive parameter selection methods, perhaps guided by local image statistics or machine learning models, could tailor the quantization steps and λ values more intelligently. Incorporating spatial or temporal coherence—for instance, by considering neighboring blocks or multiple frames from a stereo video—might also help maintain structural details in complex regions. Furthermore, extending the methodology to multi-view datasets or employing advanced transforms could unlock even greater efficiency and quality improvements. Ultimately, refining the RDO framework and expanding its applicability across various imaging scenarios could establish it as a valuable tool for robust, high-quality disparity representation and compression.

References

[1] Y. Chen, S. Chen, Y. Wang, H. Schwarz, D. Marpe, and T. Wiegand, "3D-HEVC-based Multiview Video Coding With Depth Maps," in *2013 IEEE International Conference on Image Processing*, Melbourne, Australia: IEEE, 2013, pp. 523–527, doi: 10.1109/ICIP.2013.6738089.

[2] C. Lee and Y.-S. Ho, "Rate-Distortion Optimization for Disparity Map Compression," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt: IEEE, 2009, pp. 3337–3340, doi: 10.1109/ICIP.2009.5414092.

[3] X. Hu, K. Müller, P. Merkle, A. Smolic, and T. Wiegand, "Efficient Depth Map Coding for 3D Video Systems Using Multi-View and Epipolar Constraints," in *2011 18th IEEE International Conference on Image Processing*, Brussels, Belgium: IEEE, 2011, pp. 773–776, doi: 10.1109/ICIP.2011.6116688.