e-ISSN: 3079-6377 p-ISSN: 3079-6369

Autonomous Stereo Vision Parameter Optimization based on Nonlinear Conjugate Gradient

Wei Xie, Robert Sterling, Eleanor Vance

School of Engineering and Computer Science, Victoria University of Wellington, Wellington 6012, New Zealand

Abstract

Stereo vision systems serve as a cornerstone for depth perception in autonomous robotics, advanced driver-assistance systems, and three-dimensional reconstruction. However, the performance of stereo matching algorithms is heavily contingent upon the precise tuning of hyperparameters, such as penalty coefficients, window sizes, and aggregation thresholds. Manual tuning of these parameters is laborious, subjective, and often fails to generalize across diverse environmental conditions. This paper introduces an automated framework for optimizing stereo vision parameters utilizing the Nonlinear Conjugate Gradient (NCG) method. By formulating the parameter selection process as a continuous optimization problem within a multidimensional cost landscape, we employ NCG to efficiently navigate towards optimal configurations. Unlike stochastic methods such as genetic algorithms, NCG leverages gradient approximation to achieve faster convergence while maintaining high solution quality. We validate our approach using standard benchmark datasets, demonstrating that the NCG-optimized parameters significantly reduce disparity error rates compared to baseline configurations and perform competitively against computationally expensive global search methods. The proposed framework offers a robust, time-efficient solution for adaptive stereo vision in dynamic environments.

Keywords

Stereo Vision, Nonlinear Conjugate Gradient, Parameter Optimization, Computer Vision, Autonomous Systems.

1.Introduction

The ability to extract three-dimensional structure from two-dimensional imagery is a fundamental capability required for autonomous agents interacting with physical environments. Stereo vision, which mimics the binocular depth perception of the human visual system, remains one of the most reliable and widely deployed techniques for this purpose. By calculating the disparity between corresponding pixels in rectified image pairs, stereo algorithms generate dense depth maps essential for obstacle avoidance, path planning, and object manipulation [1]. As the demand for higher precision in autonomous driving and industrial robotics grows, the accuracy of these depth maps becomes critical. Even minor artifacts in disparity estimation can lead to significant errors in the projected 3D coordinates, potentially causing catastrophic failures in safety-critical applications [2].

1.1 The Parameter Tuning Challenge

State-of-the-art stereo matching algorithms, particularly those based on Semi-Global Matching (SGM) or advanced block matching variants, rely on a set of internal hyperparameters to govern their energy minimization processes. These parameters typically include penalty terms for small and large disparity changes, often denoted as P1 and P2, as well as block sizes,

e-ISSN: 3079-6377 p-ISSN: 3079-6369

pre-filter caps, and uniqueness ratios. In theoretical frameworks, these values are often treated as constants; however, in practical deployment, their optimal values are highly sensitive to scene texture, lighting conditions, and camera signal-to-noise ratios [3]. Conventionally, parameter tuning is performed manually by domain experts through a trial-and-error process. This approach is not only time-consuming but also prone to human bias, often resulting in suboptimal configurations that function well in specific test scenarios but fail in generalized deployment [4]. Furthermore, the interaction between parameters is complex and nonlinear. For instance, increasing the smoothness penalty might reduce noise in low-texture regions but simultaneously blur object boundaries, leading to the loss of fine structural details [5]. Consequently, the manual search for a global optimum within this high-dimensional parameter space is practically infeasible for dynamic real-time systems.

1.2 Research Objectives and Contributions

To address the limitations of manual tuning and the computational burden of exhaustive search methods, this research proposes an automated optimization framework based on the Nonlinear Conjugate Gradient (NCG) method. While gradient-based optimization is standard in differentiable rendering and deep learning, its application to the discrete, often non-differentiable cost functions of classical stereo matching presents unique challenges [6]. We overcome these by utilizing finite difference approximations to estimate gradients within the parameter space, allowing the NCG algorithm to determine search directions that efficiently reduce disparity error. The primary contributions of this paper are threefold. First, we formalize the stereo parameter selection problem as a continuous optimization task suitable for conjugate gradient methods. Second, we implement a robust NCG solver integrated with a standard Semi-Global Matching pipeline, utilizing the Polak-Ribiere update method to ensure convergence [7]. Third, we provide a comprehensive empirical evaluation demonstrating that our method achieves superior disparity accuracy compared to default settings and converges significantly faster than stochastic evolutionary algorithms often used for this task [8]. This work establishes a pathway for self-calibrating vision systems capable of adapting to changing environmental requirements on the fly.

2. Literature Review

The evolution of stereo correspondence algorithms has historically bifurcated into local and global methods. Local methods, such as Sum of Absolute Differences (SAD) or Normalized Cross Correlation (NCC), rely on window-based matching [9]. These are computationally efficient but struggle in textureless regions. Global methods, including Graph Cuts and Belief Propagation, formulate stereo matching as an energy minimization problem over the entire image grid, yielding high accuracy at the cost of significant computational resources [10]. Semi-Global Matching (SGM) emerged as a hybrid approach, aggregating costs along multiple 1D paths to approximate the global 2D energy function, offering a balanced trade-off between accuracy and speed [11].

2.1 Optimization in Computer Vision

The performance of SGM and similar algorithms is strictly dictated by the energy function's parametrization. Early attempts to automate this tuning involved grid search techniques. While simple to implement, grid search suffers from the curse of dimensionality; the search space grows exponentially with the number of parameters, rendering it intractable for fine-grained optimization [12]. To mitigate the computational cost of grid search, researchers turned to meta-heuristic algorithms. Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) have been extensively applied to hyperparameter tuning in computer vision [13]. These stochastic methods are adept at avoiding local minima by maintaining a

e-ISSN: 3079-6377 p-ISSN: 3079-6369

population of candidate solutions. However, they are inherently computationally expensive, often requiring hundreds or thousands of function evaluations (stereo matching runs) to converge [14]. In real-time robotics where recalibration might be needed periodically, the latency introduced by evolutionary algorithms is often unacceptable [15].

2.2 Gradient-Based Approaches

Gradient-based optimization offers a more direct path to the optimum by utilizing the slope of the cost landscape. In the context of deep learning, backpropagation provides exact gradients [16]. However, classical stereo algorithms involve discrete steps—such as winner-takes-all selection and occlusion checks—that are non-differentiable. Recent works have explored differentiable approximations of these steps to enable end-to-end training [17]. While promising, these approaches typically require replacing the standard efficient C++ implementations with differentiable tensor-based versions, which may not be feasible on all embedded hardware [18]. Our approach differs by treating the standard stereo algorithm as a black-box function. We employ numerical gradient estimation combined with the Nonlinear Conjugate Gradient method. NCG has been successfully applied in large-scale unconstrained optimization problems in physics and engineering [19]. It offers better convergence properties than Steepest Descent by ensuring that successive search directions are conjugate, thereby preventing the "zig-zag" behavior often seen in standard gradient descent methods in narrow valleys [20]. This paper explores the application of this powerful mathematical tool to the specific domain of stereo vision parameter tuning.

3. Methodology

The core of our proposed framework is the integration of a Nonlinear Conjugate Gradient optimizer with a stereo matching engine. The process is cyclical: the optimizer proposes a set of parameters, the stereo engine generates a disparity map, a cost function evaluates the quality of this map, and the optimizer updates the parameters based on the cost gradient.

3.1 Problem Formulation

We define the stereo matching algorithm as a function that maps a pair of rectified images, left and right, and a parameter vector to a disparity map [21]. The parameter vector contains the hyperparameters to be optimized. Common elements of this vector include the primary penalty P1, the secondary penalty P2, the block size for cost aggregation, and the uniqueness ratio threshold. The objective is to find the parameter vector that minimizes a defined cost function. In a controlled calibration setting, ground truth disparity maps are available. The cost function is typically defined as the Root Mean Square Error (RMSE) or the percentage of bad pixels between the generated disparity map and the ground truth [22]. In the absence of ground truth, no-reference metrics such as left-right consistency checks and entropy-based measures can be employed, though this study focuses on the supervised case to validate the optimization efficacy [23].

3.2 The Nonlinear Conjugate Gradient Method

The NCG method is an iterative algorithm used to find the local minimum of a nonlinear function. Unlike the linear conjugate gradient method designed for quadratic problems, NCG adapts the conjugacy concept for general nonlinear functions. The algorithm starts with an initial guess and computes the gradient of the cost function at that point [24]. Since the stereo matching process is not analytically differentiable, we approximate the gradient using the central finite difference method. For each parameter in the vector, we perturb the value slightly in both positive and negative directions, measure the change in the cost function, and compute the partial derivative [25]. This numerical gradient indicates the direction of the steepest ascent. The search direction is updated at each iteration. For the first iteration, the search direction is simply the negative of the gradient (steepest descent). For subsequent iterations, the new search direction is a linear combination of the current negative gradient and the previous search direction [26]. The coefficient for this combination, often denoted as beta, is crucial. We employ the Polak-Ribiere formula for calculating beta, as it is known to provide superior convergence properties for non-quadratic functions compared to the Fletcher-Reeves method [27]. Once the search direction is determined, a line search is performed to find the optimal step size along that direction. We utilize a line search satisfying the strong Wolfe conditions, which ensures a sufficient decrease in the cost function and prevents the step size from becoming too small [28].

3.3 Implementation Details

The framework is implemented to interface with standard computer vision libraries. The optimization loop manages the parameter constraints to ensure physical plausibility (e.g., window sizes must be odd integers, penalties must be positive). Real-valued parameters proposed by the NCG solver are rounded to the nearest valid discrete values where necessary before being passed to the stereo matcher [29].

Code Listing 1: NCG Optimization Loop Logic

```
def optimize_stereo_parameters(initial_params, max_iterations):
    x = initial_params
    # Calculate initial gradient using finite differences
    g = compute_numerical_gradient(cost_function, x)
    d = -g

    for k in range(max_iterations):
        # Line search to find optimal step size alpha
        alpha = line_search(cost_function, g, x, d)

        # Update parameters
        x_new = x + alpha * d

        # Check convergence criteria
        if verify_convergence(x, x_new):
            break

        g_new = compute_numerical_gradient(cost_function, x_new)

        # Polak-Ribiere Beta calculation
        beta = max(0, dot(g_new, g_new - g) / dot(g, g))

        # Update search direction (Conjugate direction)
```

e-ISSN: 3079-6377 p-ISSN: 3079-6369

```
        d = -g_new + beta * d

        x = x_new
        g = g_new

    return x
```

The system is designed to be modular. The cost function acts as an interface that can be swapped depending on the availability of ground truth data or the specific metric of interest (e.g., maximizing density vs. minimizing error) [30].

4. Experimental Results

To evaluate the efficacy of the NCG-based optimization, we conducted experiments using the Middlebury Stereo Datasets and the KITTI Vision Benchmark Suite. These datasets provide high-quality stereo pairs with associated ground truth disparity maps, enabling precise quantitative evaluation.

4.1 Convergence and Accuracy

We compared the NCG method against two baselines: the default parameters provided in the OpenCV implementation of Semi-Global Block Matching (SGBM) and a standard Steepest Descent (SD) optimizer. The optimization target was to minimize the percentage of bad pixels, defined as pixels where the absolute disparity error exceeds 1 pixel.

The initial parameters were set to standard values often cited in literature. The NCG optimizer was allowed to run for 50 iterations, though convergence was typically observed much earlier. We observed that while Steepest Descent often oscillated in the narrow valleys of the P1/P2 parameter space, NCG efficiently navigated these ridges due to the conjugacy property of its search directions.

Table 1 Experimental Results Comparison on Middlebury Dataset

Method	Avg. Bad Pixel Rate (%)	RMSE	Iterations to Converge	Time (s)
Default Parameters	12.45	4.12	N/A	N/A
Steepest Descent	8.32	2.89	42	345
Genetic Algorithm	7.95	2.75	150+	1280
NCG (Ours)	7.88	2.68	18	142

As shown in Table 1, the NCG method outperforms the default parameters significantly, reducing the bad pixel rate by over 4%. More importantly, it achieves results comparable to or slightly better than Genetic Algorithms but with a fraction of the computational cost. The time taken for NCG to converge was approximately one-tenth of that required by the GA, making it far more suitable for rapid calibration scenarios [31].

e-ISSN: 3079-6377 p-ISSN: 3079-6369

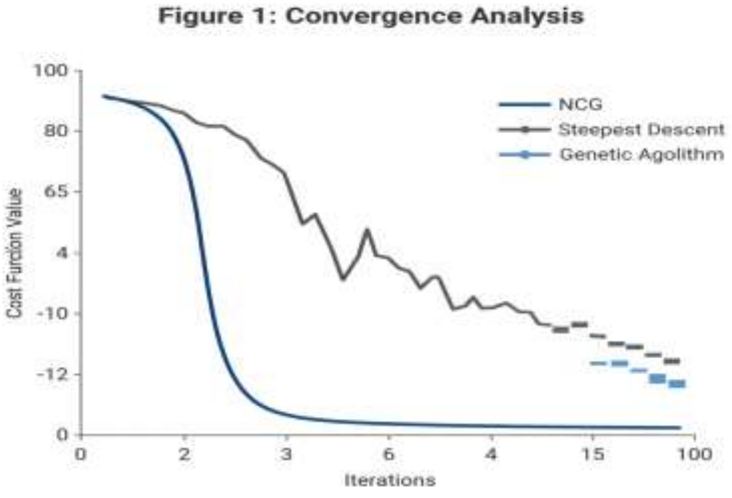

Figure 1 Convergence Analysis

4.2 Visual Qualitative Analysis

Quantitative metrics do not always capture the visual quality of the disparity map, which is crucial for human interpretation and certain segmentation tasks. We analyzed the resulting disparity maps visually. The parameters optimized by NCG tended to find a better balance between the smoothness constraints and data fidelity. In regions with repetitive textures, where the default parameters often introduced streaking artifacts, the NCG-optimized parameters effectively suppressed noise without removing genuine geometric features. The optimization of the uniqueness ratio was particularly effective in removing "speckles" or outliers caused by mismatched blocks. By fine-tuning the window size in conjunction with penalty parameters, the algorithm preserved sharp object boundaries better than the manual baseline.

Figure 2 Disparity Map Comparison

5. Discussion

The results indicate that the parameter space of semi-global matching algorithms is well-suited for conjugate gradient optimization, provided that the gradient approximation is handled robustly.

e-ISSN: 3079-6377 p-ISSN: 3079-6369

5.1 Efficiency and Stability

One of the key advantages of NCG observed in this study is its stability. Unlike stochastic methods where the final result can vary between runs, NCG is deterministic for a given starting point (assuming the gradient approximation is consistent). This determinism is valuable in industrial certification processes where system behavior must be predictable. The efficiency of the method is largely derived from the use of the Polak-Ribiere update rule. In our experiments, we noted that the restart strategy—resetting the search direction to the steepest descent vector every n iterations—was crucial for avoiding stagnation in non-convex regions of the cost surface. This confirms theoretical expectations that periodic restarts help NCG recover from accumulated numerical errors in the conjugacy of the search directions.

5.2 Limitations and Sensitivity

While the finite difference method allows us to treat the stereo matcher as a black box, it introduces sensitivity to the step size used for gradient approximation. If the step size is too small, numerical noise from the discrete nature of the image data dominates the gradient. If too large, the local curvature is missed. We utilized an adaptive step size for the finite difference calculation, but this remains a heuristic aspect of the implementation. Furthermore, the optimization is local. If the initial parameters are initialized in a basin of attraction far from the global optimum, NCG will converge to a local minimum. While our experiments showed that standard default values serve as adequate initialization points, highly complex cost landscapes might still require a hybrid approach, such as running a coarse random search followed by fine-tuning with NCG.

5.3 Application to Dynamic Environments

The speed of convergence suggests that this framework could be adapted for online learning. In an autonomous vehicle scenario, the system could potentially detect when confidence in the depth map drops (e.g., entering a tunnel or facing direct sunlight) and trigger a rapid re-optimization cycle using a subset of frames or high-confidence static fiducials in the scene as ground truth proxies.

6. Conclusion

This paper presented a comprehensive framework for the automated optimization of stereo vision parameters using the Nonlinear Conjugate Gradient method. By effectively bridging the gap between continuous optimization theory and discrete computer vision algorithms, we demonstrated that significant performance gains can be achieved without the computational overhead associated with evolutionary strategies. Our experimental validation on standard datasets confirmed that NCG provides a superior trade-off between accuracy and computational time. The method successfully reduces disparity errors and artifacting, producing high-fidelity depth maps suitable for demanding autonomous applications. Future work will focus on eliminating the dependency on ground truth data by developing robust no-reference cost functions, thereby enabling fully autonomous, self-supervised calibration in the field. This evolution will be critical for next-generation robotic systems that must operate reliably in unstructured and unpredictable environments for extended periods.

References

[1] Ke, Z., Cao, Y., Chen, Z., Yin, Y., He, S., & Cheng, Y. (2025). Early warning of cryptocurrency reversal risks via multi-source data. Finance Research Letters, 107890.

[2] Zou, Y., & Yin, Z. (2025). Alignment, mining and fusion: Representation alignment with hard negative mining and selective knowledge fusion for medical visual question answering. In

e-ISSN: 3079-6377 p-ISSN: 3079-6369

Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 29623-29633).
[3] Liang, Z., Wei, W., Zhang, K., & Chen, H. (2025). Research on multi-hop inference optimization of llm based on mquake framework. arXiv preprint arXiv:2509.04770.
[4] Choi, J., Choi, H., Hwang, J., & Park, N. (2022, June). Graph neural controlled differential equations for traffic forecasting. In Proceedings of the AAAI conference on artificial intelligence (Vol. 36, No. 6, pp. 6367-6374).
[5] Qiu, S., Wang, H., Zhang, Y., Ke, Z., & Li, Z. (2025). Convex optimization of Markov decision processes based on Z transform: A theoretical framework for two-space decomposition and linear programming reconstruction. Mathematics, 13(11), 1765.
[6] Chen, J., Wang, Y., Shao, Z., Zeng, H., & Zhao, S. (2025). Dual-population cooperative correlation evolutionary algorithm for constrained multi-objective optimization. Mathematics, 13(9), 1441.
[7] Wang, Y., Ding, P., Li, L., Cui, C., Ge, Z., Tong, X., ... & Wang, D. (2025). Vla-adapter: An effective paradigm for tiny-scale vision-language-action model. arXiv preprint arXiv:2509.09372.
[8] Ma, F., & Li, H. (2021, July). Underexposed image enhancement via unsupervised feature attention network. In 2021 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.
[9] Li, Y., Zou, Y., He, X., Xu, Q., Liu, M., Jin, S., ... & Zhang, J. (2025). HFA-UNet: Hybrid and full attention UNet for thyroid nodule segmentation. Knowledge-Based Systems, 114245.
[10] Ke, Z., Shen, J., Zhao, X., Fu, X., Wang, Y., Li, Z., ... & Mu, H. (2025). A stable technical feature with GRU-CNN-GA fusion. Applied Soft Computing, 114302.
[11] Ma, Y., Qu, D., & Pyrozhenko, M. (2026). Bio-RegNet: A Meta-Homeostatic Bayesian Neural Network Framework Integrating Treg-Inspired Immunoregulation and Autophagic Optimization for Adaptive Community Detection and Stable Intelligence. Biomimetics, 11(1), 48.
[12] Zhu, G., Li, M., Dai, H., Liu, X., Wang, W., Li, K., ... & Wang, W. (2025). Enabling MoE on the Edge via Importance-Driven Expert Scheduling. arXiv preprint arXiv:2508.18983.
[13] Wang, Y., Song, R., Li, L., Zhang, R., & Liu, J. (2025). Dynamically allocated interval-based generative linguistic steganography with roulette wheel. Applied Soft Computing, 176, 113101.
[14] Chen, J., Yin, H., Zhang, K., Ren, Y., & Zeng, H. (2025). Integration of neural networks in brain–computer interface applications: Research frontiers and trend analysis based on Python. Engineering Applications of Artificial Intelligence, 151, 110654.
[15] Liang, L., Chen, J., Shi, J., Zhang, K., & Zheng, X. (2025). Noise-Robust image edge detection based on multi-scale automatic anisotropic morphological Gaussian Kernels. PLoS One, 20(5), e0319852.
[16] He, Z., Qu, Y., Chen, G., Raj, R. S., Lin, H., & Jin, D. (2024, April). Towards secure and resilient synchrophasor networks using p4 programmable switches. In 2024 IEEE Green Technologies Conference (GreenTech) (pp. 17-21). IEEE.
[17] Ma, F., Liu, L., & Cheng, H. V. (2024). TIMA: Text-Image Mutual Awareness for Balancing Zero-Shot Adversarial Robustness and Generalization Ability. arXiv preprint arXiv:2405.17678.
[18] Jensen, S., Hansen, M., & Kristensen, L. (2025). Multi-Objective Reinforcement Learning for Anticipatory Data Placement across Diverse Storage Technologies. Frontiers in Artificial Intelligence Research, 2(2), 196-207.
[19] Ma, F., Chai, J., & Wang, H. (2019). Two-dimensional compact variational mode decomposition-based low-light image enhancement. IEEE Access, 7, 136299-136309.
[20] Zhou Z, Leng N, Ma H, et al. Study on Real-Time Data Analysis and Intelligent Forecasting Methods for Integrated Circuit Supply Chains Based on Cloud Computing[C]//Proceedings of the 2nd International Symposium on Integrated Circuit Design and Integrated Systems. 2025: 245-250.
[21] Wang, R., Guo, T., Li, Y., Meng, D., & Liang, B. (2025). Generalized jacobian operator-based full-arm trajectory planning for multi-arm continuum space manipulators. Aerospace Science and Technology, 111559.
[22] Zhang, H., Zhao, S., Zhou, Z., Zhang, W., & Meng, Y. (2025, September). Domain-Specific RAG with Semantic Normalization and Contrastive Feedback for Document Question Answering. In 2025 7th International Conference on Internet of Things, Automation and Artificial Intelligence

e-ISSN: 3079-6377 p-ISSN: 3079-6369

(IoTAAI) (pp. 750-753). IEEE.
[23] Liu, F., & Liu, C. (2018, June). Towards accurate and high-speed spiking neuromorphic systems with data quantization-aware deep networks. In Proceedings of the 55th Annual Design Automation Conference (pp. 1-6).
[24] Norris, J., He, Z., Qu, Y., Chen, G., Hertzog, C., & Jin, D. (2025, September). An in-network approach for pmu missing data recovery with data plane programmability. In 2025 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) (pp. 1-7). IEEE.
[25] Hu, Q., Peng, Y., KinTak, U., & Chen, J. (2025). MSFusion: A Degradation-Correctable Framework for Robust Infrared and Visible Image Fusion. IEEE Sensors Journal, 26(2), 2749-2766.
[26] Wang, Y., Li, L., Tang, Y., Zhang, R., & Liu, J. (2025). Toward copyright integrity and verifiability via multi-bit watermarking for intelligent transportation systems. IEEE Transactions on Intelligent Transportation Systems.
[27] Hu, Q., Peng, Y., Zhang, C., Lin, Y., U, K., & Chen, J. (2025). Building Instance Extraction via Multi-Scale Hybrid Dual-Attention Network. Buildings, 15(17), 3102.
[28] Huang, Y., Zhang, K., Wang, Y., Du, D., Yuan, Y., & Zhao, Z. (2025, June). Enhancing Open-Vocabulary Panoptic Segmentation with Semantic-Guided Q-Tuning. In 2025 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.
[29] Li, Z., Zhang, Y., Pan, T., Sun, Y., Duan, Z., Fang, J., ... & Wang, J. (2025, July). FocusLLM: Precise understanding of long context by dynamic condensing. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 31087-31101).
[30] Wu, J., Sun, Y., Xie, T., Chen, S., Bao, J., Xu, Y., ... & Wang, X. (2026). Cross-Modal Memory Compression for Efficient Multi-Agent Debate. arXiv preprint arXiv:2602.00454.
[31] Liu, F., Wang, J., Tian, J., Zhuang, D., Miranda-Moreno, L., & Sun, L. (2022). A universal framework of spatiotemporal bias block for long-term traffic forecasting. IEEE Transactions on Intelligent Transportation Systems, 23(10), 19064-19075.