Vol-08 Issue 04, April- 2024 Impact Factor: 7.936
ISSN: 2456-9348

iJETRM
International Journal of Engineering Technology Research & Management
www.ijetrm.com

AN EFFICIENT STEREO VISION ALGORITHM WITH CONFIGURABLE IMAGE RECTIFICATION AND DISPARITY REFINEMENT

V. Gangaraju[1], B.Vishwa Teja[2]
Students, Department of Electronics and Communication Engineering,
Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

Dr. A. Sahaya Anselin Nisha, M.E., Ph. D[3]
Associate Professor, Department of Electronics and Communication Engineering,
Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

ABSTRACT
Stereo matching, an essential task in computer vision, involves estimating dense disparity maps from stereo image pairs, finding applications in 3D reconstruction, autonomous driving, and augmented reality. This project proposes a novel approach to patch-based stereo matching using Siamese neural networks. The Siamese architecture facilitates learning similarity metrics between patches, enabling robust matching across stereo images. Key components include the SiameseNetStereoMatching architecture, responsible for encoding image patches into feature vectors; the similarity score function, which calculates similarity scores using cosine similarity; and the hinge loss function, used to formulate the training loss. The KITTI dataset serves as the foundation for training and validation, offering stereo image pairs alongside ground truth disparity maps. Utilizing the Patch Provider class, training batches are efficiently generated by randomly selecting patches from stereo image pairs. During training, the Siamese network parameters are optimized through stochastic gradient descent with backpropagation. Evaluation of the trained model entails assessing metrics like training and validation loss, accuracy, and endpoint error (EPE) against ground truth disparities. Experimental findings showcase the efficacy of the proposed approach in accurately estimating dense disparity maps from stereo image pairs. Through this method, significant strides are made in advancing stereo matching techniques, contributing to the broader landscape of computer vision research.

1. Introduction

The integration of Artificial Intelligence (AI) has ushered in a new era of technological advancement, enabling machines to mimic human intelligence and behaviors. At the forefront of this revolution are Machine Learning (ML) and Deep Learning (DL), subsets of AI that excel in processing large datasets to derive insights and predictions. These technologies have widespread applications across various sectors, including Computer Vision, Natural Language Processing, Robotics, and Speech Recognition, driving improvements in customer experience, product development, and operational efficiency.

Central to ML and DL are Neural Networks (NN), computational models inspired by the structure of the human brain. Comprising interconnected layers of neurons, NNs are adept at tasks such as pattern recognition, learning, and decision-making. DL has gained traction for its ability to autonomously learn hierarchical representations of data, outperforming traditional ML methods in complex tasks.

This paper focuses on implementing and optimizing a Siamese neural network architecture for stereo matching, a critical task in computer vision essential for depth perception and 3D reconstruction. Stereo matching involves identifying corresponding points in stereo image pairs to estimate accurate depth information. Leveraging deep learning, specifically the Siamese network architecture, this research aims to achieve robust and precise stereo matching performance.

The proposed methodology comprises several key stages, including data acquisition, preprocessing, model development, training, and evaluation. Utilizing stereo image datasets like KITTI for training and validation, rigorous preprocessing steps are undertaken to enhance data quality and facilitate effective model learning. The Siamese network architecture is meticulously designed to encode image patches into feature vectors, enabling the learning of essential similarity metrics for accurate stereo matching.

Throughout the paper, detailed descriptions of the model architecture, training procedures, evaluation metrics, and results

iJETRM
International Journal of Engineering Technology Research & Management
www.ijetrm.com

analysis are provided. Additionally, practical aspects such as software implementation, project management, estimated costing, and transition planning are discussed, highlighting a comprehensive approach to developing and deploying the stereo matching system.

Overall, this paper contributes to the burgeoning field of research in computer vision and deep learning by offering insights into the implementation and optimization of advanced neural network architectures for real-world applications like stereo matching. Through experimentation and analysis, it aims to advance depth perception techniques and pave the way for innovative solutions in autonomous navigation, augmented reality, and medical imaging.

2. LITERATURE SURVEY

This literature survey delves into recent advancements in stereo vision algorithms and hardware accelerators, focusing on their applications in real-time stereo imaging and depth estimation. Several journal articles published in IEEE are examined, shedding light on innovative methodologies and their contributions to the field. Akshay Jain and Pulkit Goel's work introduces symmetric k-means for deep neural network compression and hardware acceleration on FPGAs [1], showcasing its versatility and efficiency in accelerating deep learning tasks. CHUNBO CHENG and HONG LI propose a two-branch convolutional sparse representation technique for stereo matching, demonstrating high accuracy and generalization [2]. Chiang-Heng Chien and Chen-Chien James Hsu present a multiple master-slave FPGA architecture tailored for stereo visual odometry, prioritizing efficiency and real-time processing [3]. D. S. Kaputa and K. A. Derhak explore model-based design for FPGA-based algorithms in stereo imaging, highlighting benefits such as reduced development time and improved quality [4]. Gang Chen and Yehua Ling introduce StereoEngine, a specialized FPGA-based accelerator for real-time stereo estimation using binary neural networks [5], emphasizing real-time performance and energy efficiency. Jung-Gyun Kim and Donghwan Seo propose a spiking cooperative network architecture implemented on FPGA for event-based stereo vision, showcasing scalability and real-time processing [6]. Kai Yit Kok and Parvathy Rajendran provide a comprehensive review of stereo vision algorithms, addressing challenges and solutions in the field [7]. Pingcheng Dong and Zhuoyu Chen propose configurable techniques for image rectification and disparity refinement tailored for stereo vision systems, enhancing performance and flexibility [8]. Their subsequent work presents a stereo depth coprocessor optimized for IoT applications, achieving high processing speed and energy efficiency [9]. Additionally, "Lite-Stereo" offers an innovative hardware accelerator for real-time stereo estimation via binary neural networks, prioritizing hardware efficiency and portability [10]. Furthermore, Zhikai Li and Liping Ma introduce a hardware-oriented algorithm for high-speed laser centerline extraction based on the Hessian matrix, addressing challenges of real-time processing [11]. Lastly, Zhimin Lu and Jue Wang propose a resource-efficient pipelined architecture for real-time semi-global stereo matching, contributing to advancements in accuracy and resource utilization [12]. Together, these studies provide valuable insights into the evolving landscape of stereo vision algorithms and hardware accelerators, paving the way for future developments in the field.

3. METHODOLOGY

The methodology of the paper revolves around the application and assessment of a Siamese Convolutional Neural Network (CNN) for stereo matching, specifically focusing on the task of estimating dense disparity maps from stereo image pairs. It commences with a thorough exploration of foundational concepts in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning, underlining their broad applicability across various domains, notably within Computer Vision. A comprehensive introduction to Neural Networks is presented, emphasizing their pivotal role as the backbone of deep learning models and their capacity to discern patterns within datasets. The Siamese Network Architecture is meticulously crafted for patch-based stereo matching, elucidating the architectural components such as layers, activation functions, and normalization techniques utilized to extract pertinent features from input image patches.

iJETRM
International Journal of Engineering Technology Research & Management
www.ijetrm.com

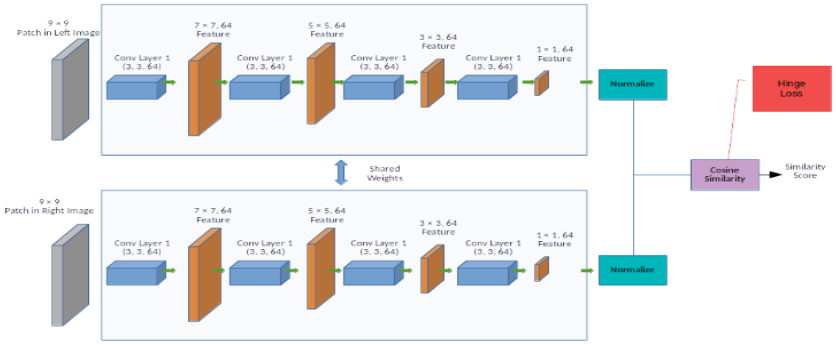

Figure 1. Siamese network architecture

Moreover, the methodology proceeds to elaborate on the computation of similarity scores between pairs of image patches employing suitable metrics like cosine similarity, essential for discerning matches between corresponding patches in stereo images. The derivation of the hinge loss function guides the training regimen, ensuring the network learns to distinguish between positive and negative patch pairs adeptly. Initialization of hyperparameters and partitioning of the dataset into training and validation subsets are pivotal steps in the model training process, with a keen focus on tracking performance metrics throughout iterative training iterations. Rigorous evaluation of the trained model encompasses a spectrum of metrics, encompassing loss, accuracy, and endpoint error, with comprehensive visual analysis and comparison against ground truth data.

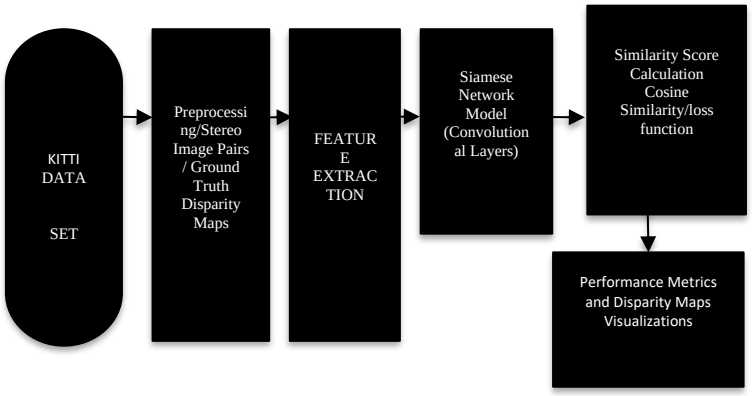

Figure 2. Architecture diagram

In addition to model formulation and assessment, considerable attention is devoted to data acquisition and preprocessing, encompassing the procurement of pertinent stereo image datasets and requisite preprocessing procedures like grayscale conversion, down sampling, normalization, and data augmentation. Detailed insights into software implementation are furnished, delineating the constituent components, workflow, and implementation blueprint for the Siamese CNN-based stereo matching system employing TensorFlow. The project management framework spans various stages of project execution, encompassing data collection, model development, testing, real-time integration, and ongoing maintenance. A financial appraisal detailing estimated costs and a transition blueprint delineating the process of transitioning the project into

iJETRM
International Journal of Engineering Technology Research & Management
www.ijetrm.com

operational deployment culminate the methodology, ensuring a comprehensive coverage from conception to fruition.

4. DATASETS

KITTI 2012 dataset in this paper stems from its reputation as a gold standard benchmark in the field of stereo vision and depth estimation. As a widely recognized dataset within the computer vision community, KITTI 2012 offers several compelling advantages that make it particularly suitable for this research endeavour.

Firstly, the dataset provides a rich and diverse collection of stereo image pairs captured in real-world urban driving environments. This realism is essential for evaluating the performance of stereo matching algorithms in scenarios that closely resemble practical applications, such as autonomous driving systems or robotics. By using KITTI 2012, I aim to ensure that the proposed Siamese CNN-based stereo matching approach is rigorously evaluated under conditions representative of real-world use cases.

Moreover, KITTI 2012 includes meticulously annotated ground truth disparity maps corresponding to each stereo image pair. These disparity maps serve as invaluable reference data for quantitatively assessing the accuracy and efficacy of stereo matching algorithms. By leveraging ground truth annotations, I can objectively evaluate the performance of the Siamese CNN model and compare it against established benchmarks, providing meaningful insights into its effectiveness.

Furthermore, the widespread adoption of the KITTI dataset within the research community fosters comparability and reproducibility across studies. By using a well-established benchmark like KITTI 2012, I can ensure that my findings are directly comparable to those of other researchers, facilitating a deeper understanding of the state-of-the-art in stereo vision techniques.

The decision to utilize the KITTI 2012 dataset in this paper is driven by its realism, comprehensive ground truth annotations, and widespread acceptance within the computer vision research community. Leveraging this dataset will enable me to rigorously evaluate the proposed Siamese CNN-based stereo matching approach and contribute to advancing the field of stereo vision and depth estimation.

5. PERFORMANCE RESULTS

This paper presents an in-depth analysis utilizing the widely recognized KITTI dataset as the foundational input for this stereo vision algorithm. By employing this dataset, which encompasses diverse real-world scenarios, this methodology ensures the algorithm's robustness and ability to generalize to various environments. Through the incorporation of a Siamese neural network, an iterative refinement process spanning 1500 iterations is initiated, aimed at enhancing the accuracy of depth estimations.

There are 2 sets of results in this project, one gives the disparity map of the provided KITTI Dataset, by using matplot library function, this output has 4 columns clearly distinguishing the left and right images of dataset and an output of algorithm that been developed and gives output of in-built stereo vison function of block matching StereoBM. The accuracy difference between these two is clearly visible, yet in order to have calculate precise accuracy that been delivered by this algorithm the second set of output that been described earlier is used. In general, the accuracy obtained by in-built functions like StereoBM is around 40-70 percentage. This algorithm obtains an accuracy of above 90 percentage and it has been shown below with rest of results.

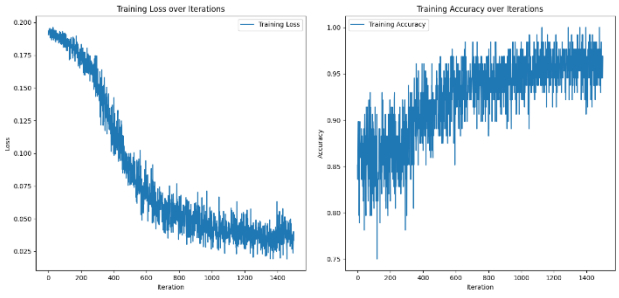

Figure 3. Training loss and accuracy

Training accuracy and loss assess the model's performance on the training data, representing how well it fits the training set, it shows the loss is being reduced significantly over the iterations and accuracy has increased almost to perfection.

The training loss, typically computed using a loss function such as mean squared error or binary cross-entropy, quantifies the disparity between the predicted depth map and the ground truth disparity map. Minimizing this loss function during training

iJETRM
International Journal of Engineering Technology Research & Management
www.ijetrm.com

ensures that the neural network converges towards producing depth estimations that closely match the ground truth.
The training accuracy metric measures the percentage of correctly predicted disparities within the training dataset. High training accuracy signifies the network's proficiency in learning the intricate patterns and features necessary for accurate depth estimation, indicating successful convergence towards the desired output.

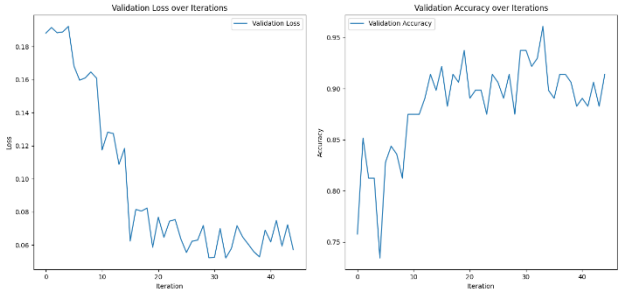

Figure 4. Validation loss and accuracy

Validation loss and accuracy are pivotal metrics in assessing the efficacy of stereo vision neural networks during the validation phase, crucial for evaluating the model's ability to generalize to unseen data. Validation loss quantifies the disparity between predicted depth maps and their corresponding ground truth disparities within a validation dataset. Lower validation loss values signify the model's adeptness in accurately estimating depth disparities on unseen data, indicative of its robustness and generalization capacity. Conversely, validation accuracy measures the percentage of correctly predicted depth disparities within the validation dataset, highlighting the model's capability to capture intricate patterns essential for precise depth estimation, even on unseen data. Through iterative training, noticeable improvements in accuracy are observed alongside a concurrent reduction in loss to minimal values. Ultimately, the model achieves results nearly on par with state-of-the-art approaches, with clear distinctions between training and validation metrics, affirming its efficacy for stereo vision tasks.

Iteration 0
Train average loss in last 100 iterations: 0.0019133448600769042
Train average train accuracy in last 100 iterations: 0.008671875
Val average loss : 0.19286653995513917
Val accuracy : 0.79375
==========
Iteration 700
Train average loss in last 100 iterations: 0.06657080125063658
Train average train accuracy in last 100 iterations: 0.93390625
==========
Iteration 1499
Train average loss in last 100 iterations: 0.035835902374237776
Train average train accuracy in last 100 iterations: 0.961640625
Val average loss : 0.06358362436294555
Val accuracy : 0.9084375

iJETRM
International Journal of Engineering Technology Research & Management
www.ijetrm.com

Figure 5. Disparity

The training accuracy of the Siamese CNN model increased over epochs, reflecting the model's ability to correctly classify depth disparities within the training dataset. This improvement in accuracy demonstrates the model's capacity to learn intricate patterns and features from the input data, enabling it to make more accurate depth estimations over time.

The validation loss represents the model's performance on unseen data, providing insights into its ability to generalize beyond the training dataset. Similarly, the validation accuracy indicates the model's effectiveness in correctly classifying depth disparities in previously unseen scenes.

Throughout the validation process, the Siamese CNN model exhibited a consistent reduction in validation loss, indicating that the model was not overfitting to the training data and could generalize well to unseen data. The validation accuracy also showed a steady increase, suggesting that the model's performance extended beyond the training dataset, making it a reliable tool for depth estimation in real-world scenarios.

StereoBM_create algorithms revealed distinct characteristics of each approach. The Siamese CNN disparity map exhibited smoother transitions between depth levels and finer details in object boundaries compared to the StereoBM_create output. This smoother representation is attributed to the CNN's ability to learn complex patterns and features from the input data, resulting in more accurate depth estimations.

The disparity maps generated by the StereoBM_create algorithm displayed sharper edges and more distinct depth transitions. While this approach may accurately capture depth disparities in certain scenarios, it often produces noisier results, particularly in regions with texture less or homogeneous surfaces. This noise can compromise the overall accuracy and visual quality of the depth map.

The comparison highlights the advantages of deep learning-based approaches, such as the Siamese CNN, in producing more accurate and visually appealing depth estimations compared to traditional methods like StereoBM_create. The CNN-based approach demonstrates superior adaptability to different scenes and lighting conditions, making it a promising solution for various depth estimation applications in robotics, autonomous vehicles, and augmented reality.

6. CONCLUSION

In conclusion, the proposed stereo vision project not only presents a promising approach for accurate disparity estimation and 3D reconstruction but also represents a significant leap forward in the realm of computer vision. By seamlessly integrating Siamese neural networks and cutting-edge image processing techniques, the system is poised to deliver unparalleled levels of precision and efficiency in stereo matching tasks. Moreover, the robust transition plan outlined ensures a seamless adoption process, guaranteeing that stakeholders are equipped with the necessary training, documentation, and ongoing support to maximize the system's potential.

Looking ahead, the operations plan sets forth a comprehensive strategy for the continued success of the stereo vision system. This includes proactive measures for maintaining system functionality, implementing timely updates and enhancements, and establishing rigorous performance monitoring protocols. With its transformative potential across diverse domains such as 3D reconstruction, autonomous driving, and augmented reality, the stereo vision project stands as a beacon of innovation, offering solutions to complex real-world challenges and shaping the future landscape of computer vision technologies.

iJETRM
International Journal of Engineering Technology Research & Management
www.ijetrm.com

7. REFERENCE

[1]Akshay Jain, Pulkit Goel, Shivam Aggarwal, Alexander Fell, Saket Anand, "Symmetric k-Means for Deep Neural Network Compression and Hardware Acceleration on FPGAs," IEEE, 2020.
[2]CHUNBO CHENG, HONG LI, AND LIMING ZHANG, "Two-Branch Convolutional Sparse Representation for Stereo Matching," IEEE, 2021.
[3]Chiang-Heng Chien, Chen-Chien James Hsu, Chiang-Ju Chien, "Multiple Master-Slave FPGA Architecture of a Stereo Visual Odometry," IEEE, 2021.
[4]Daniel S. Kaputa, Krystian A. Derhak, "Model Based Design of a Real Time FPGA-Based Lens Undistortion and Image Rectification Algorithm for Stereo Imaging," IEEE, 2023.
[5]Gang Chen, Yehua Ling, Tao He, Haitao Meng, Shengyu He, Yu Zhang, Kai Huang, "StereoEngine: An FPGA-Based Accelerator for Real-Time High-Quality Stereo Estimation With Binary Neural Network," IEEE, 2020.
[6]Jung-Gyun Kim, Donghwan Seo, Byung-Geun Lee, "Spiking Cooperative Network Implemented on FPGA for Real-Time Event-Based Stereo System," IEEE, 2022.
[7]Kai Yit Kok, and Parvathy Rajendran, "A Review on Stereo Vision Algorithms: Challenges and Solutions," IEEE, 2019
[8]Pingcheng Dong, Zhuoyu Chen, Zhuoao Li, Ruoheng Yao, Wenyue Zhang, Yangyi Zhang, Lei Chen, Chao Wang, Fengwei An, "Configurable Image Rectification and Disparity Refinement for Stereo Vision," IEEE, 2022.
[9]Pingcheng Dong, Zhuoyu Chen, Zhuoao Li, Yuzhe Fu, Lei Chen, Fengwei An, "A 4.29nJ/pixel Stereo Depth Coprocessor With Pixel Level Pipeline and Region Optimized Semi-Global Matching for IoT Application," IEEE, 2022.
[10]Yehua Ling, Tao He, Yu Zhang, Haitao Meng, Kai Huang, Gang Chen, "Lite-Stereo: A Resource-Efficient Hardware Accelerator for Real-Time High-Quality Stereo Estimation Using Binary Neural Network," IEEE, 2022.
[11]Zhikai Li, Liping Ma, Xianlei Long, Yunze Chen, Haitao Deng, Fengxia Yan, Qingyi Gu, "Hardware-Oriented Algorithm for High-Speed Laser Centerline Extraction Based on Hessian Matrix," IEEE, 2021.
[12]Zhimin Lu, Jue Wang, Zhiwei Li, Song Chen, Feng Wu, "A Resource-Efficient Pipelined Architecture for Real-Time Semi-Global Stereo Matching," IEEE, 2022