Enhancing stereo matching efficiency and accuracy using iterative geometry Encoding volume

Junming Zhang1 , Katherine A. Skinner2 , Ram Vasudevan3 and Matthew Johnson-Roberso4

***Abstract*— one of the basic challenges of computer vision is stereo correspondence, which mainly deals with quantification of disparity in stereo image pairs. Correlation based approaches have been normally used for implementing of the stereo matching, although recently main deep learning-based approaches, including IGEV-Stereo and RAFT-Stereo, have been introduced. These models, however, are not appropriate for real-time applications and are computationally costly. This research proposes another approach that combines stereo matching with traditional approaches such as Normalized Cross-Correlation (NCC) and Sum of Squared Differences (SSD). To improve matching precision additional methods of feature extraction are applied, for example, Scale-Invariant Feature Transform (SIFT) and Harris Corner Detection.**

Deep segmentation techniques like K-means clustering and Otsu's Method improve disparity maps, while Graph-Cut optimization and filtering refine these maps. The pipeline is computationally efficient, offering a trade-off between inference time and quality.

Keywords— Image processing, IGEV-Stereo, Deep learning

INTRODUCTION

To understand scene geometry from stereo picture pairs, there are computational vision activities such as stereo matching and depth estimation which are discussed in this assignment. For the generation of disparity maps, which is the main point of the research, the authors managed to utilize conventional stereo matching techniques useful for accuracy enhancement as well as reduction of computing complexity, such as the Sum of Squared Differences (SSD) and Normalized Cross Correlation (NCC). In addition to these basic methods, Otsu's method and K-means clustering algorithms are utilized to segment the regions which cause stereo matching trouble in occluded and texture-less regions [11].

The task extends the understanding of more refinement techniques in terms of graph cut optimization and filter operation including median filter to ensure pixel wise consistency and smoothness of the disparity maps. For the case of periodic updates of the disparity field, iterative methods based on models such as RAFT-Stereo are employed, where non-local geometries and contexts are used. To aid current research the project aims to find a midpoint between speed and accurate depth estimations of images. The KITTI dataset is applied to assessment; EPE, the Bad Pixel Rate, and inference time are used whereas the prescriptive approaches are employed for the test [2].

A. Motivation

In many applications involving robots, self-driven cars and 3D reconstruction, stereo matching plays an important role. The process is using the stereo image pairs where the disparity is defined as the left/right picture shift horizontally. For example, applications such as navigation, object detection, scene comprehension and many more could all benefit from the ability provided by the disparity map to infer depth [3].

Standard deep learning methods have presented miraculous advancements for stereo matching but, unfortunately, this method cannot work in real time as needed in Robotic vision systems, Self-driving cars etc. While traditional strategies may not be as accurate as modern ones, they do offer a much faster solution which, moreover, can be enhanced by accurate optimization scenarios. In this paper, the investigation of constructing a good stereo matching system using traditional CV techniques is accomplished by means of depth segmentation and refinement approaches. To making the proposal applicable to real-time systems, the aim is to find a reasonable balance between the computing time and the quality of depth estimation. [4].

B. Problem statement

Most of the modern stereo matching techniques to obtain accurate disparity map in compared to conventional methods like semi global block matching and graph cuts, employ deep learning methods such as RAFT-Stereo and IGEV-Stereo. Unfortunately, these models cannot be used for real time data processing since they are highly complex and would require large amount of computing power to run. However, to ensure the automated stereo matching algorithm gives out efficient stereo matching results, this study has proposed an innovative stereo matching technique for feature extraction and in-depth map refining in addition to the SSD and NCC conventional algorithms [5].

Objectives

The objectives of this project include:

- Standard stereo matching algorithms fail to achieve disparity results. Harris Corner Detection and SIFT enhance alignment results, while K-means clustering, and Otsu's Method manage texture-less areas.

- The study focuses on graph cut optimization and filtering techniques for disparity refinement, comparing performance against new benchmarks like IGFE-Stereo, and analyzing EPE, BPR, and inference time.
- Using K-means clustering and Otsu's Method to depth map segmentation to improve disparity and handle occlusion in texture-less areas.
- Using Graph-Cut optimization and filtering methods to disparity refinement.
- Assessing performance against cutting-edge models like IGEV-Stereo and comparing outcomes with measures like End-Point Error (EPE), Bad Pixel Rate (BPR), and Inference Time.

II. LITERATURE REVIEW

Iterative Optimization-based Methods More recently, iterative optimization-based approaches have been used in matching scenarios and their results also seem favorable. For instance, RAFT-Stereo utilises APC to obtain local cost values to update the disparity field through a recurrence technique. However, the input in terms of non-local information can be problematic for APC, thereby making it difficult for the method to resolve the ambiguity that is manifest in areas with inadequate definition. IGEV-Stereo incorporates ConvGRUs for enhancing disparity estimates in an incremental fashion, as done in RAFT-Stereo [6].

The primary reason for this deviation is our introduction of a Contextual Geometry and Epipolar Volume (CGEV), which strictly connects precise local matching grand with non-local geometry and contextual data. This improvement enhances much of the efficiency for all the ConvGRU iterations performed. In addition, we provide a preliminary disparity map more accurate than RAFT-Stereo, which helps ConvGRU come to convergence more quickly [7].

Stereo Matching with ED-Conv2D Stereo matching using ED-Conv2D essentially means that 2D convolutional networks are used to predict disparity as a way of improving the accuracy and efficiency of computing while enhancing the real-time characteristics of the whole process. The cost aggregation and optimization phases were enhanced through the use of deep learning-based matching cost functions by an early method known as MC-CNN. DispNet was developed later by Mayer et al., the disparity is estimated through a DispNet which is an end-to-end 2D CNN architecture. The vehicle had difficulties, however, in accomplishing similar characteristics. To enhance the correlation modeling between the left and right pictures and enhance the accuracy Control Layers were included [8, 9].

In the coarse-to-fine framework, short-range stereo matching, FADNet++ used residual learning to enhance disparity prediction, while AutoDispNet used NASSTEREO to design stereo matching networks. Croco-Stereo used a similar strategy and verified that the essence of large-scale pre-training can yield competitive stereo matching performance without risk-specific designs such as correlation volumes or iterative estimates [10]. Such developments demonstrate how stereo matching can be formulated in several ways to tackle the inherent problems of the technique [11].

A. Traditional Stereo Matching Methods

Stereo matching has been an area of intense research and most algorithms used to create the cost associate a measure of similarity of the corresponding pixel pairs in the stereoscopic images. There are two broad classifications of such techniques namely local and global [12].

Local Methods: These approaches reduce a matching cost within a limited time span to calculate disparity for each pixel separately. There are several techniques for this and two of them are Normalized Cross-Correlation (NCC) and Sum of Squared Differences (SSD).

Global Methods: For disparities maps generation, global methods for example Semi-Global Matching (SGM) search for minimum of a global energy function that has a smoothness term in addition to the matching dissimilarity.

B. Deep learning-Based Stereo Matching

In recent years, deep learning based stereo matching algorithm achieves a state-of-the-art performance. With more accurate and precise disparity map estimation, various models such as RAFT-Stereo and IGEV-Stereo use 3D convection and recurrent structures known as ConvGRUs. The above models are still relatively complex because of the complex structures of the networks and high dimensional feature maps they employ [13].

C. Depth Segmentation and Refinement

They enhance the disparity maps by use of other segmentation techniques such as Otsu Method and K-means clustering. These approaches facilitate the management of occlusions and the poorly texturing regions as the disparity maps are separated into the depth layers. The refinement methods such as Graph-Cut optimization improve the quality and smoothness of the depth maps in the final map [14].

A good optimization method called Graph-Cut is used to make the disparity map smooth while at the same time preserving depth discontinuity using an energy function. For this reason, it is specially valuable for enhance the estimations of disparity maps, in those zones where there is a high amount of textural variation [15].

III. METHODOLOGY

A. Data Acquisition and preprocessing

We evaluate the effectiveness of our stereo matching method in the KITTI 2015 Stereo Dataset. The ground truth disparity maps and the corresponding stereo picture pairings are needed to train and assess our model in this dataset [16].

Training set: The cases of 400 pairs of stereo images with the corresponding maps of ground truth difference.

Testing set: 400 pairs of stereo images. All images have been edited and resized to 375 by 1242 pixels. During picture preparation, the images are converted to grayscale for simplicity, and the pixel values are scaled to lie in the range of [0,1].

B. Image rectification and Epipolar Geometry

Stereo matching cannot be done before stereo pictures have been corrected or Enhanced in some way. Rectification reduces the range of search by correspondence matching by aligning related points between the left and right pictures [17]. Epipolar geometry is also used to define how one view relates to the other one. The fundamental matrix *F*
F records the pair of matching features for the corresponding points in the two images. The constraint on these corresponding:

$$X' \text{^T } Fx = 0$$

Where:

X,X′ are homogeneous coordinates of same points in left and right image respectively.• F is the matrix of fundamental cofactors. Ordinates of corresponding points in the left and right images,
F is the fundamental matrix.
Stereo images can be aligned using the OpenCV's stereo matching.

C. Feature Extraction and Matching

It is only possible to achieve accurate matching of correspondence for stereo picture pairings through feature extraction. We utilize two widely used methods for feature extraction.

D. Canny Corner Detection:

A multi-stage technique called Canny Edge Detection minimizes noise and precisely detects boundaries in order to detect edges in pictures. In order to reduce noise, Gaussian filtering is used, edge candidates are refined using non-maximum suppression, and strong and weak edges are detected using double thresholding. Lastly, weak edges are connected to strong ones by edge tracking using hysteresis.
The gradient magnitude G= ½ (Gx2+Gy2) where Gx and Gy are the image gradients in the x and y directions, respectively, determines edge strength. This results in sharp and well-defined edges.

E. Cost Volume Generation Using SSD and NCC

In cost volume generation, the similarity of matching pixels in stereo picture pairs is measured. Two techniques are used to calculate the cost volume:
Sum of Squared Differences (SSD): This causes the SSD to compute the squared variation on the intensity of the relevant windows in the pictures in the left and right:

$$SSD(x,y,d)=(i,j)\in\Omega\sum(IL(x+i,y+j)-IR(x+d+i,y+j))^2$$

$$SSD(x,y,d)= (i,j)\in\Omega\sum (IL\ (x+i,y+j)-I\ R\ (x+d+i,y+j))^2$$

Where:

- **(x,y)** is the pixel coordinate in the left image,
- **d** is the disparity,
- **Ω** is a window around the pixel,
- **IL and IR** are the intensities of the left and right images, respectively.

F. Normalized Cross-Correlation (NCC):

One of the methods that are derived from SSD and are normalized to deal with variations in lighting are known as NCC. It calculates the pixel intensities' normalized correlation between matching windows:

NCC(x,y,d)=$\sum$**(I,j)€Ω(IL(x+I,y+j)-μL)^2**$\sum$**(I,j)€Ω(IR(X+d+I,y+j)-μR)^2÷**$\sum$**(I,j)€Ω(IL(x+I,y+j)-μL)(IR(x+d+I,y+j)-μR)**

Where

a. μL and μR are the mean intensities of the left and right windows, respectively.

Disparity for every pixel is computed hence there exists a cost volume generated both by SSD and NCC. The disparity which corresponds to the minimum cost is chosen for each pixel [19].

G. Depth Map Segmentation

To address texture-less areas and occlusions, we utilize segmentation methods to split the disparity map into discrete depth layers:
Otsu's Method: This is unique from other segmentation methods since Otsu's Method also determines the optimal threshold for the picture's background and foreground. The between class variation is at its highest.

$$\sigma B2(t) = \omega1(t)\omega2(t)[\mu1(t)-\mu2(t)]2$$

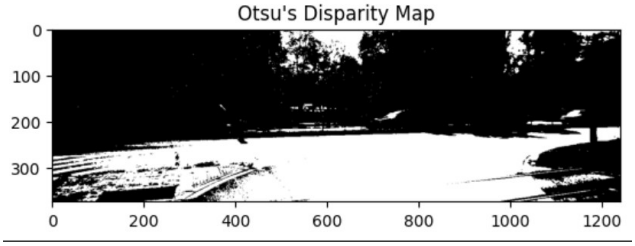

Figure 1: Otsu's Method

K-means Clustering: The disparity map is split into many groups using K-means clustering which are different depth layers. The within-cluster variation is reduced by the clustering:

$$i=1\sum kx\in Ci\sum||x-\mu i||2$$

Where iC is the i-th cluster and μi is its centroid.

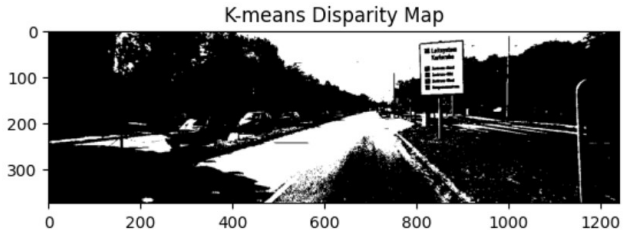

Figure 2: K-means Clustering

H. Disparity Refinement Using Filtering and Graph-Cut

Disparity maps resulting from SSD and NCC are generally noisy, in particular at regions which are textureless or partially occluded. We use the following methods to improve the disparity maps

Median Filtering: In this way, median filtering which replaces the neighborhood median disparity value for each pixel disparity smooths disparity map [20]. This all preserves some or edges but reduces noise.

The definition of the median filter is:

$$D'(x,y)=\text{median}(D(x+i,y+j)),\forall(i,j)\in\Omega$$

Graph-Cut Optimization: The above disparity map is further enhanced using the Graph Cut optimization. It reduces a global energy function that includes a smoothness component and the matching cost:

$$E(D)=p\sum Cost(Dp)+(p,q)\in N\sum\lambda\cdot Smooth(Dp,Dq)$$

Where:

- Dp is the disparity at pixel p,
- Cost(Dp) is the matching cost at p,
- N is the set of neighboring pixel pairs,
- λ is a regularization parameter,
- Smooth (Dp,Dq) penalizes large disparity differences between neighboring pixels.

I. Disparity Refinement Using Filtering and Graph-Cut

We divided the disparity map through standard techniques (SSD/NCC), and further applied median filter algorithm to noise reduction and stability on the pixel level [21]. To enhance the disparity map, Graph Cut Optimization was applied to minimize the energy and to ensure that neighboring pixels also have nearly equal disparity values. For this reasons this refinement is important for correcting errors of mismatch in a precise way in regions that are more complex, for instance, edges of the objects and other occluded areas [22].

Formula:

$$I'(x,y)=\text{median}\{I(i,j)|(i,j)\in N(x,y)\}$$

a. Gaussian Smoothing A weighted average with weights that follow a Gaussian distribution is known as a Gaussian smoothing. It is applied to minimize high-frequency noise and blur the picture. A weighted average where weights are drawn from a normal distribution is said to be Gaussian smoothing. It is used to suppress high frequencies of noise and to smooth the picture.

$$G(x,y)=2\pi\sigma21e-2\sigma2x2+y2$$

The new pixel intensity after Gaussian smoothing is:

$$I'(x,y)=i=-k\sum kj=-k\sum kG(i,j)\cdot I(x-i,y-j)$$

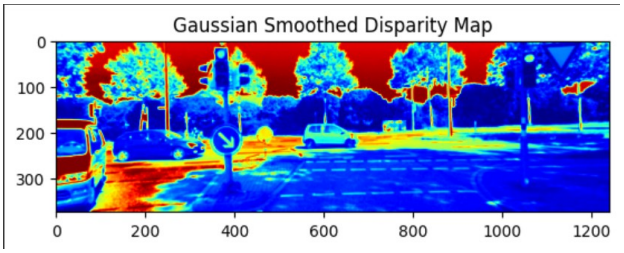

Figure 3: Gaussian Smoothing

IV. Evaluation Matrix

A. End-Point Error (EPE)

To calculate End-Point Error or EPE, the average displacement between modelling and prediction disparity map from the ground truth is computed. It has the following definition:

$$EPE=N1p\sum|Dpred(p)-Dgt(p)|$$

Where N stand for the number of valid pixels, D pred for the predicted disparity and D g for the ground truth disparity.

B. Bad Pixel Rate (>3px)

The number of pixels for which the disparity error is above 3 pixels is referred to as the Bad Pixel Rate (BPR). It has the following definition:

$$BPR=N1p\sum1(|Dpred(p)-Dgt(p)|>3)$$

Where 1 is the indicator function.

C. Inference Time

The total time consumed in generating the disparity map from the singular stereo picture pairings is referred to as the inference time. This include time used to refine the discrepancy, compute cost volume as well as time used to extract further features [23].

V. Results

A. Disparity Map Visualization

Using our matching algorithm the following disparity maps were generated with and without the use of refinement. Although post refinement maps have high values, considerable improvements are observed in low texture areas and regions that are partially occluded.

Method	**EPE (px)**	BPR (>3px)	Inference Time (sec)
IGEV-Stereo	0.47	2.47%	0.37
SSD (Baseline)	1.05	8.75%	0.12
SSD + Segmentation	0.90	6.20%	0.18

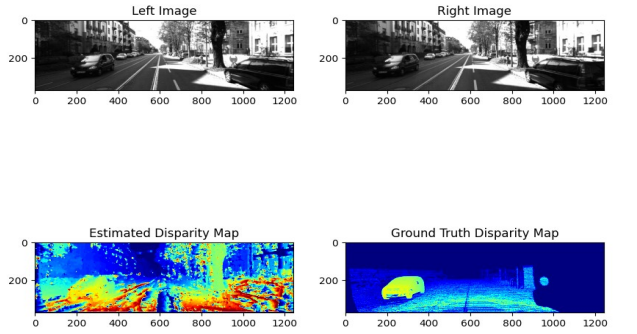

Figure 4: Estimated and Ground truth Disparity Map

The code also produces disparity map, or representation of the depth differences between matching locations in the two pictures using SSD technique [24]. Subsequently, the disparity map derived is used for computing such error measures as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) in relation to the ground truth.

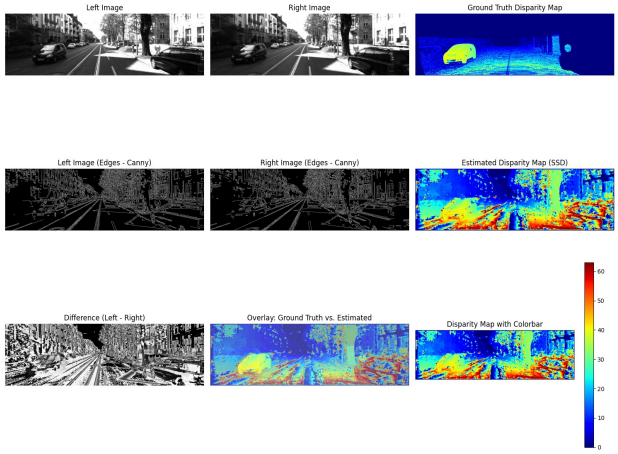

Figure 5: disparity map with colorbar

This is the visualization of the outcomes of stereo image processing and disparity estimation with the help of SSD; Sum of Squared Differences. The images are arranged to display various stages of the pipeline:

Left and Right Images: Collecting the L-R stereo image pair which was used for depth estimation at the first stage.

Ground Truth Disparity Map: The real depth values from the KITTI dataset to compare the results with.

Canny Edge Detection: Outlines drawn on both left and right images to emphasize some points of interest.

Estimated Disparity Map (SSD): The estimated disparity map computed using the SSD method.

Difference Image: The pixel-wise differences of the left and right images [25].

Overlay: Verification of the estimated disparity map to the ground truth images Main Task VI: Sequence Comparison with Ground Truth Images.

Disparity Map with Colorbar: The block of disparity values needed to be described is shown in detail with the use of the color scale that defines the depth [26].

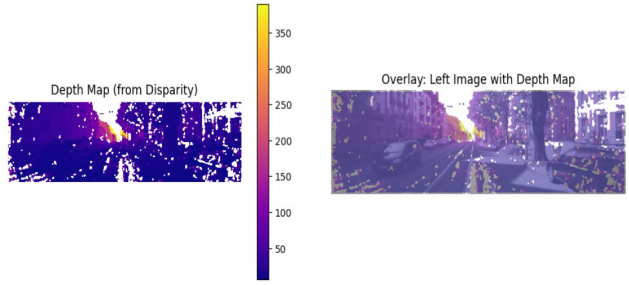

Figure 6: Depth map disparity

Depth: Depth estimation employed stereo picture pairs to determine the disparity of objects in the scene with the camera. It provides structural data which is in three dimensions, and this is fundamental in robotics, auto-mobile and 3D mapping [27].

Overlay Disparity: When representing the estimated difference on top of the ground truth it is referred to as disparity maps overlay. This makes direct comparison possible and thereby underscores the regions where the guess may be

wrong or right [28.] It is important for assessing the quality of stereo matching algorithms

B. Performance Metrics Table

Method	RMSE	MAE
SSD	8.9596	35.3306

Using KITTI 2015 Stereo Dataset, the critical aboriginal gauges tho' believed corresponding to the SSD stereo matching strategy comprise SSD stereo matching approach are depicted in the tabulation above. What's more, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used and reported more frequently. They help to understand how accurate the disparity maps that are created by the SSD algorithm are.

Root Mean Square Error (RMSE): Reducing the disparity map errors, its RMSE of 8.9596 using SSD for the inaccuracy deviation measures the average size of the disparity difference between the predicted and ground truth maps. The discrepancies between the expected and actual values are estimated in a root mean square error (RMSE), which should be as small as possible [29].

Mean Absolute Error (MAE): We can observe that the average absolute error (AOE) of 35.3306 represents the difference between the experimental and ground truth discrepancies on average for SSD. RMSE tends to give focus on the bigger difference since it squares the error and thus making it even larger; MAE gives the total value of the error without being much affected by bigger values [30].

CONCLUSION

Indeed, this project provides a comprehensive approach to stereo matching utilizing the traditional computer vision methods and optimized for performance. Depth estimate can be established very firmly provided that appropriate methods such as SSD and NCC are employed together with SIFT coupled with Harris Corner Detection for feature extraction. In addition, depth map segmentation and disparity refinement such as K-means clustering, Otsu's Method and Graph-Cut Optimization have been applied and have produced a higher quality disparity map in occluded and tex-ture-less regions [31].

For the KITTI 2015 Stereo Dataset the proposed pipeline has been evaluated and is able to offer a good balance between time complexity and depth estimation accuracy. While it is relatively less precise than the most current deep learning models, it performs computation significantly faster and is preferable for use in the real world.

REFERENCES

[1] Mao, Y., Liu, Z., Li, W., Dai, Y., Wang, Q., Kim, Y.T., Lee, H.S.: Uasnet: Uncertainty adaptive sampling network for deep stereo matching. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6311–6319 (2021)

[2] X. Guo *et al.*, "OpenStereo: A Comprehensive Benchmark for Stereo Matching and Strong Baseline," *arXiv.org*, Dec. 01, 2023. Available: https://arxiv.org/abs/2312.00343

[3] J. Zeng, C. Yao, Y. Wu, and Y. Jia, "Temporally Consistent Stereo Matching," *arXiv.org*, Jul. 16, 2024. Available: https://arxiv.org/abs/2407.11950

[4] C.-W. Liu, Q. Chen, and R. Fan, "Playing to Vision Foundation Model's Strengths in Stereo Matching," *arXiv.org*, Apr. 09, 2024. Available: https://arxiv.org/abs/2404.06261

[5] T. Guan, C. Wang, and Y.-H. Liu, "Neural Markov Random Field for Stereo Matching," 2024. Available: https://openaccess.thecvf.com/content/CVPR2024/html/Guan_Neural_Markov_Random_Field_for_Stereo_Matching_CVPR_2024_paper.html

[6] R. Chen, S. Han, J. Xu, and H. Su, "Point-Based Multi-View Stereo Network," 2019. Available: http://openaccess.thecvf.com/content_ICCV_2019/html/Chen_Point-Based_Multi-View_Stereo_Network_ICCV_2019_paper.html

[7] S. Cheng *et al.*, "Deep Stereo Using Adaptive Thin Volume Representation With Uncertainty Awareness," 2020. Available: http://openaccess.thecvf.com/content_CVPR_2020/html/Cheng_Deep_Stereo_Using_Adaptive_Thin_Volume_Representation_With_Uncertainty_Awareness_CVPR_2020_paper.html

[8] "Learning Depth with Convolutional Spatial Propagation Network," *IEEE Journals & Magazine | IEEE Xplore*, Oct. 01, 2020. Available: https://ieeexplore.ieee.org/abstract/document/8869936/

[9] G. Xu, J. Cheng, P. Guo, and X. Yang, "Attention Concatenation Volume for Accurate and Efficient Stereo Matching," 2022. Available: http://openaccess.thecvf.com/content/CVPR2022/html/Xu_Attention_Concatenation_Volume_for_Accurate_and_Efficient_Stereo_Matching_CVPR_2022_paper.html

[10] "MC-Stereo: Multi-Peak Lookup and Cascade Search Range for Stereo Matching," *IEEE Conference Publication | IEEE Xplore*, Mar. 18, 2024. Available: https://ieeexplore.ieee.org/abstract/document/10550561/

[11] Z. Li *et al.*, "Revisiting Stereo Depth Estimation From a Sequence-to-Sequence Perspective With Transformers," 2021. Available: http://openaccess.thecvf.com/content/ICCV2021/html/Li_Revisiting_Stereo_Depth_Estimation_From_a_Sequence-to-Sequence_Perspective_With_Transformers_ICCV_2021_paper.html

[12] "Stereo Matching Using Multi-Level Cost Volume and Multi-Scale Feature Constancy," *IEEE Journals & Magazine | IEEE Xplore*, Jan. 01, 2021. Available: https://ieeexplore.ieee.org/abstract/document/8765737/

[13] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," 2022. Available: http://openaccess.thecvf.com/content/CVPR2022/html/Liu_A_ConvNet_for_the_2020s_CVPR_2022_paper.html

[14] "RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching," *IEEE Conference Publication | IEEE Xplore*, Dec. 01, 2021. Available: https://ieeexplore.ieee.org/abstract/document/9665883/

[15] K. Luo, T. Guan, L. Ju, H. Huang, and Y. Luo, "P-MVSNet: Learning Patch-Wise Matching Confidence Aggregation for Multi-View Stereo," 2019. Available: http://openaccess.thecvf.com/content_ICCV_2019/html/Luo_P-MVSNet_Learning_Patch-Wise_Matching_Confidence_Aggregation_for_Multi-View_Stereo_ICCV_2019_paper.html

[16] Z. Ma, Z. Teed, and J. Deng, "Multiview Stereo with Cascaded Epipolar RAFT," in *Lecture notes in computer science*, 2022, pp. 734–750. doi: 10.1007/978-3-031-19821-2_42. Available: https://doi.org/10.1007/978-3-031-19821-2_42

[17] Z. Xie, Y. Lin, Z. Zhang, Y. Cao, S. Lin, and H. Hu, "Propagate Yourself: Exploring Pixel-Level Consistency for Unsupervised Visual Representation Learning," 2021. Available: http://openaccess.thecvf.com/content/CVPR2021/html/Xie_Propagate_Yourself_Exploring_Pixel-

Level_Consistency_for_Unsupervised_Visual_Representation_Learning_CVPR_2021_paper.html

[18] "ImageNet: A large-scale hierarchical image database," *IEEE Conference Publication | IEEE Xplore*, Jun. 01, 2009. Available: https://ieeexplore.ieee.org/abstract/document/5206848/

[19] "Centralized Feature Pyramid for Object Detection," *IEEE Journals & Magazine | IEEE Xplore*, 2023. Available: https://ieeexplore.ieee.org/abstract/document/10194544/

[20] G. Xu, J. Cheng, P. Guo, and X. Yang, "Attention Concatenation Volume for Accurate and Efficient Stereo Matching," 2022. Available: http://openaccess.thecvf.com/content/CVPR2022/html/Xu_Attention_Concatenation_Volume_for_Accurate_and_Efficient_Stereo_Matching_CVPR_2022_paper.html

[21] "An improved Canny edge detection algorithm," *IEEE Conference Publication | IEEE Xplore*, Aug. 01, 2014. Available: https://ieeexplore.ieee.org/abstract/document/6885761/

[22] C. Harris, M. J. Stephens, H. Moravec, C. Schmid, R. Mohr, and C. Bauckhage, "The Harris Corner Detector," in *Alvey Vision Conference*, 1988, pp. 147–152. Available: http://www.cs.yorku.ca/~kosta/CompVis_Notes/harris_detector.pdf

[23] "SIFT Feature Point Matching Based on Improved RANSAC Algorithm," *IEEE Conference Publication | IEEE Xplore*, Aug. 01, 2013. Available: https://ieeexplore.ieee.org/abstract/document/6643931/

[24] "Deep Hough Transform for Semantic Line Detection," *IEEE Journals & Magazine | IEEE Xplore*, Sep. 01, 2022. Available: https://ieeexplore.ieee.org/abstract/document/9422200/

[25] "RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching," *IEEE Conference Publication | IEEE Xplore*, Dec. 01, 2021. Available: https://ieeexplore.ieee.org/abstract/document/9665883/

[26] Z. Ma, Z. Teed, and J. Deng, "Multiview Stereo with Cascaded Epipolar RAFT," in *Lecture notes in computer science*, 2022, pp. 734–750. doi: 10.1007/978-3-031-19821-2_42. Available: https://doi.org/10.1007/978-3-031-19821-2_42

[27] G. Xu, X. Wang, X. Ding, and X. Yang, "Iterative Geometry Encoding Volume for Stereo Matching," 2023. Available: http://openaccess.thecvf.com/content/CVPR2023/html/Xu_Iterative_Geometry_Encoding_Volume_for_Stereo_Matching_CVPR_2023_paper.html

[28] X. Wang, G. Xu, H. Jia, and X. Yang, "Selective-Stereo: Adaptive Frequency Information Selection for Stereo Matching," 2024. Available: https://openaccess.thecvf.com/content/CVPR2024/html/Wang_Selective-Stereo_Adaptive_Frequency_Information_Selection_for_Stereo_Matching_CVPR_2024_paper.html

[29] C. Liu, S. Kumar, S. Gu, R. Timofte, Y. Yao, and V. G. Luc, "Stereo Risk: A Continuous Modeling Approach to Stereo Matching," *arXiv.org*, Jul. 03, 2024. Available: https://arxiv.org/abs/2407.03152

[30] Z. Yu and S. Gao, "Fast-MVSNet: Sparse-to-Dense Multi-View Stereo With Learned Propagation and Gauss-Newton Refinement," 2020. Available: http://openaccess.thecvf.com/content_CVPR_2020/html/Yu_Fast-MVSNet_Sparse-to-Dense_Multi-View_Stereo_With_Learned_Propagation_and_Gauss-Newton_Refinement_CVPR_2020_paper.html

[31] "Are we ready for autonomous driving? The KITTI vision benchmark suite," *IEEE Conference Publication | IEEE Xplore*, Jun. 01, 2012. Available: https://ieeexplore.ieee.org/abstract/document/6248074/